Distinguishing Time-varying Factor Models*

Zhonghao Fu[a,b], Liangjun Su[c], Xia Wang[d]

[a] School of Economics, Fudan University, China

[b] Shanghai Institute of International Finance and Economics, China

[c] School of Economics and Management, Tsinghua University, China

[d] School of Economics, Renmin University of China, China

Abstract

Time-varying factor models have been widely used to model changing relationships among economic and financial variables. The existing literature usually specifies the time-varying factor loadings as deterministic functions of time or unit root processes. This paper proposes two consistent tests to distinguish these two specifications based on a randomization approach. By setting the null hypothesis as either specification, we show that the proposed test statistics follow an asymptotic chi-squared distribution under the respective null hypotheses and diverge to infinity in probability under the respective alternatives. Simulation studies show that both test statistics perform reasonably well in finite samples. We apply the proposed tests to the U.S. macroeconomic and global macroeconomic and financial datasets. The results suggest that the time-varying factor loadings as deterministic functions of time should be adopted for these two applications.

JEL Classification: C12, C14, C22.

Key Words: Nonnested models; Randomized tests; Smooth function; Time-varying factor models

---

*Fu acknowledges financial support from the National Science Foundation of China (NSFC) (No.71973030). Su acknowledges the financial support from the NSFC (No.72133002). Wang acknowledges financial supports from NSFC (No.71873151). Address correspondence to: Xia Wang, School of Economics, Renmin University of China, Beijing, 100872, China. Email: *wxia@ruc.edu.cn*.

1 Introduction

Factor models have attracted great attention in analyzing large-dimensional macroeconomic and financial datasets. In a factor model, a few latent common factors drive the comovement of a large-dimensional vector of time series variables, and the factor loadings capture the relationships between large-dimensional variables and latent common factors. The conventional factor models (e.g., Stock and Watson, 2002; Bai and Ng, 2002; Bai, 2003) assume the factor loadings to be time-invariant for a long period. Recently, more and more studies have realized that the factor loadings may be time-varying due to various forces such as economic transition, institutional switching, preference changes, and technological innovations. With growing empirical evidence of the widespread time-varying economic relationships, macroeconomists and financial economists have paid more and more attention to the time-varying factor models and their related models.

A vast literature on time-varying factor models specifies the time-varying factor loadings as either deterministic functions of time or stochastic processes, mostly stationary VAR processes or unit root processes. Bates et al. (2013) consider estimating approximate factor models with temporal instability in the factor loadings. They show that for the time-varying factor model, the common factors can still be consistently estimated by the principal components analysis (PCA) in terms of the mean square error convergence under certain conditions on structural instability. Mikkelsen et al. (2019) specify the factor loadings to evolve as stationary vector autoregressive (VAR) processes and propose a two-step maximum likelihood estimator for time-varying factor loadings. They note that the PCA approach can still deliver consistent estimators for the common factors under certain conditions when the time-varying factor loadings follow stationary VAR processes. Thus, one can ignore the stationary stochastic time-varying behavior in factor loadings and use the conventional PCA procedure to estimate the common factors. Nevertheless, if the time-varying factor loadings evolve as deterministic functions of time with a sufficient magnitude of variation or as the usual unit root processes, the PCA will result in inconsistent estimation. Consequently, existing tests for structural changes constructed based on common factors, such as Chen et al. (2014), Han and Inoue (2015), and Cheng et al. (2016), can distinguish a factor model with stationary VAR factor loadings from that with the factor loadings evolving as deterministic functions of time or unit root processes. However, they cannot distinguish the factor loadings with unit root processes from those as deterministic functions of time despite the broad applications of the latter two specifications in empirical studies.

Factor models with a random-walk type of factor loadings have been widely adopted in empirical applications. Stock and Watson (2002) assume the time-varying factor loadings to follow unit root processes with innovations of order $O_P(T^{-1})$, where T is the sample size of time. Such small innovations can be regarded as temporal instability (e.g., Stock and Watson, 1996, 1998). Eickmeier et al. (2015) consider a time-varying factor-augmented vector autoregressive (FAVAR) model where the factor loadings are assumed to be varying as random walk processes with non-asymptotically negligible innovations and apply the model to a large dataset of U.S. macroeconomic variables. Del Negro and Otrok (2008) assume that the factor loadings evolve as random walk processes and use the time-varying factor model to study the evolution of international business cycles in the post-Bretton Woods period. Baumeister et al. (2013) and Korobilis (2013) also specify the factor loadings as random walk processes when studying the transmission mechanism of the U.S. monetary policy. Mumtaz and Musso (2021) use a dynamic factor model with a random-walk type of factor loadings to extract global, regional, and country-specific uncertainty. Regarding estimation, a common approach is to estimate the time-varying factor model with random walk factor loadings under the Bayesian framework, say, via a Gibbs sampling procedure. However, this procedure does not yield consistent estimates for the time-varying factor loadings.

An alternative specification for the time-varying factor loadings regards the factor loadings as piecewise smooth functions of time. For example, Breitung and Eickmeier (2011) investigate the consequences of structural breaks in the factor loadings for the specification and estimation of the factor model based on the PCA and propose three statistics to test for structural breaks in factor loadings. Su and Wang (2017) introduce a time-varying factor model in which the factor loadings are allowed to change smoothly over time and propose a local version of the PCA to estimate the common factors and factor loadings. Ma et al. (2020) propose a high-dimensional alpha test for the time-varying factor models with high-dimensional assets, where the time-varying factor loadings are specified as smooth functions of time. Fu et al. (2022a) propose a time-varying FAVAR model with both the factor loadings and the regression coefficients being smooth functions of time.

Although both the stochastic and deterministic specifications for factor loadings can model time-varying relationships among large-dimensional variables, their interpretations and implications are quite distinct. Cogley and Sargent (2001) point out that the fluctuations in the parameters of a reduced-form economic system may result from evolving beliefs of the policymaker, which leads to stochastic evolution. The evolution in beliefs itself is a potential product of the interaction of model misspecification by policymakers and the effects of policies on the economy. In contrast, the deterministic time-varying coefficients arise when either smooth or abrupt structural changes exist. The economic motivation behind it is the changing structure of an economic relationship, and such a structural change only depends on time. In addition, the estimation procedures for these two types of specifications are also different. For example, most existing literature estimates the time-varying factor models with stochastic factor loadings under the Bayesian framework and proposes several sampling procedures. In contrast, nonparametric kernel or sieve methods are typically employed to estimate the time-varying factor models when the factor loadings are smooth functions of time. To draw a reliable conclusion, one should specify an appropriate structure for the time-varying factor loadings. To the best of our knowledge, no formal test exists to distinguish these two different specifications for time-varying factor loadings despite their broad applications in empirical research.

This paper proposes specification tests to distinguish the deterministic time-varying and unit-root type factor loadings. Since these two types of models are nonnested, one cannot employ some existing testing procedures that typically work for nested models. Instead, we adopt the randomization approach pioneered by Pearson (1950) to construct our tests. The randomization approach has a long history in the literature. The idea behind it is simple. When a statistic only has one realization based on the sample observations, one can inject randomness into it to gauge its asymptotic properties. For instance, Corradi and Swanson (2006) propose a randomized test statistic to test for appropriate data transformations. Under the proper transformation, they show that the test statistic follows a well-defined asymptotic distribution using the added random variables conditioning on the sample but diverges to infinity under the improper transformations. Based on a similar idea but under a different framework, Bandi and Corradi (2014) adopt the randomization approach to propose nonparametric tests for nonstationarity, which are robust to nonlinear dynamics. Trapani (2018) estimates the number of common factors in a static factor model by using a randomization approach to test the magnitude order of the sample eigenvalues of the data's covariance matrix sequentially. Barigozzi and Trapani (2020) use the randomization approach to monitor the structural stability of a static factor model.

To distinguish the two types of time-varying factor loadings, we construct two statistics that exhibit different orders of magnitude under the deterministic time-varying and unit-root type factor loadings. Using the randomization approach, we can show that the randomized test statistics follow an asymptotic chi-squared distribution with one degree of freedom (χ_1^2) under the respective null hypotheses of either specification and diverge to infinity under the respective alternative hypotheses. The merit of such an approach is three-fold. First, we avoid direct Bayesian or nonparametric estimation of the unknown time-varying factor model and do not need to consider consistent estimation for the true number of factors under either specification. Second, the test statistic is easy to construct and asymptotically pivotal. Third, we can construct two test statistics for either specification as the null hypothesis so that we can formally distinguish the two nonnested specifications.

Admittedly, a well-known drawback of the randomization approach is that it relies on the generated randomness under which different researchers may draw different conclusions with small positive probability using the same dataset. However, such an issue does not only exist for the randomization approach. For instance, sample conditioning is also applied to bootstrap tests. Different researchers may obtain distinct bootstrap quantiles even with the same sample and the same value of the actual test statistic. Despite this, we note that a substantial difference remains between the bootstrap resampling and the randomization approach. For the bootstrap method, if the sample size is sufficiently large, all researchers should reject the null hypothesis at $100\alpha\%$ of the cases under the null hypothesis when the asymptotic level of the test is α. In contrast, for the randomization approach, we should expect that $100\alpha\%$ of the researchers shall reject the null hypothesis when it is true, and the asymptotic level is chosen to be α, given the same dataset. As a result, the interpretations of committing Type I errors are different for the bootstrap and randomized tests.

The remainder of the paper is organized as follows. Section 2 introduces the hypotheses of interest and the test statistics. Section 3 investigates the asymptotic properties of our test statistics under the null and alternative hypotheses. Section 4 discusses some possible extensions of our test. Section 5 reports Monte Carlo simulation results. Section 6 provides empirical studies on the U.S. macroeconomic dataset and the global economic and financial dataset. Final remarks are given in Section 7. All proofs are relegated to the Appendix.

2 The Hypotheses and Test Statistics

In this section, we first introduce the hypotheses of interest and then propose the corresponding test statistics.

2.1 The Hypotheses of Interest

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 time-varying factor model:

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

where $F_t = (F_{1t}, ..., F_{Rt})'$ is an $R \times 1$ vector of unobserved common factors with R being the true number of common factors, λ_{it} is an $R \times 1$ vector of time-varying factor loadings, and ε_{it} is the idiosyncratic error such that $E(\varepsilon_{it}|F_t, \lambda_{it}) = 0$.

In the literature, there are two widely adopted specifications for the time-varying factor loadings λ_{it}. One of them specifies λ_{it} as a deterministic function of time:

$$\lambda_{it} = \lambda_i(t/T), \tag{2.2}$$

where $\lambda_i : (0, 1] \to \mathbb{R}$ is a smooth function of the rescaled time index t/T with countably many discontinuity points defined on the unit interval $(0, 1]$. Related works include Breitung and Eickmeier (2011), Chen et al. (2014), Han and Inoue (2015), and Su and Wang (2017).

The other specification assumes that λ_{it} follows a unit root process:

$$\lambda_{it} = \mu_i + \lambda_{i(t-1)} + \nu_{it}, \tag{2.3}$$

where μ_i is an $R \times 1$ drift term and ν_{it} is an $R \times 1$ martingale difference sequence for each i with $E(\nu_{it}) = 0$ and $\text{var}(\nu_{it}) = \Pi_i$. Related works include Stock and Watson (2002), Banerjee et al. (2008), Del Negro and Otrok (2008), Bates et al. (2013), and Eickmeier et al. (2015).

These two specifications have distinct economic interpretations. The specification given by (2.2) is closely related to the literature on structural breaks in factor models. Related literature includes Breitung and Eickmeier (2011), Chen et al. (2014), Han and Inoue (2015), Su and Wang (2017, 2020), and Fu et al. (2022b), who test for structural breaks in factor models. Structural breaks imply that the time-varying feature of the factor loadings only depends on time t deterministically. For example, the abrupt structural breaks considered by Breitung and Eickmeier (2011), Chen et al. (2014), and Han and Inoue (2015) assume that λ_{it} is a step function of the rescaled time index t/T for each i. The smooth structural changes studied by Su and Wang (2017, 2020) and Fu et al. (2022b) allow λ_{it} to be a smooth function of t/T for each i. In contrast, the specification given by (2.3) assumes that the time-varying feature of the factor loadings is stochastic, which relates to the literature on random coefficient models in time series regression; see, e.g., Rosenberg (1973) and Cooley and Prescott (1976). Recently, several empirical studies, such as Del Negro and Otrok (2008), Baumeister et al. (2013), Korobilis (2013), Eickmeier et al. (2015), and Mumtaz and Musso (2021), model the time-varying factor loadings as random walk processes.

In empirical applications, a natural question arises: which specification, (2.2) or (2.3), should one adopt for a particular study? The primary interest of this paper is to propose formal tests to distinguish these two specifications. Specifically, we consider the following two hypotheses

$$\mathbb{H}_1 : \lambda_{it} = \lambda_i \left(t/T\right),$$

where $\lambda_i\left(\cdot\right)$ is a smooth function of the rescaled time index t/T with countably many discontinuity points defined on the unit interval $(0, 1]$, and

$$\mathbb{H}_2 : \lambda_{it} = \lambda_{i(t-1)} + \nu_{it},$$

in which the drift term is set as zero. At this moment, we focus on the unit root process without a drift. We will consider the case with a drift term in Section 4.2.

2.2 Test Statistics

For an $m \times n$ real matrix A, we denote A' as its transpose and $\|A\| \equiv \mathrm{tr}(AA')^{1/2}$ as its Frobenius norm, where $\equiv$ signifies definitional relationship and $\mathrm{tr}(\cdot)$ is the usual trace operator. Let $\mathbb{I}(\cdot)$ denote the indicator function and the operator $\stackrel{a.s.}{\to}$ denote the almost sure convergence as the sample size $T \to \infty$.

In this section, we propose two statistics to test $\mathbb{H}_1$ against $\mathbb{H}_2$ and $\mathbb{H}_2$ against $\mathbb{H}_1$, respectively. In this way, one can formally distinguish these two specifications. To achieve this, we first show the impact of these two specifications on the factor models. Specifically, we consider the sum of the sample eigenvalues of the $T \times T$ matrix $XX'/(NT)$, where X is a $T \times N$ matrix with the (t, i)th element given by X_{it}.

Let $\hat{\phi}_{jNT}$ be the jth largest sample eigenvalue of $XX'/(NT)$. Then, we have

$$\hat{D} \equiv \sum_{j=1}^{T} \hat{\phi}_{jNT} = \mathrm{tr}\left(\frac{XX'}{NT}\right) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(F_t'\lambda_{it} + \varepsilon_{it})^2$$

$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} F_t'\lambda_{it}\lambda_{it}'F_t + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}^2 + \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} F_t'\lambda_{it}\varepsilon_{it}.$$

Under $\mathbb{H}_1$, it is straightforward to show that $\hat{D} = O_p(1)$ under certain regularity conditions on F_t and λ_{it} (e.g., Su and Wang, 2017). Intuitively, the deterministic time-varying features of λ_{it} do not affect the typical order of the sample eigenvalues. However, when $\mathbb{H}_2$ holds, we can show that $\hat{D} = O_p(T)$ due to the explosive features of λ_{it}. This implies that we can distinguish $\mathbb{H}_1$ and $\mathbb{H}_2$ by examining the order of magnitude of the sample eigenvalues.

However, since one can only observe one sample in practice, it is infeasible to test the order of magnitude of $\hat{D}$ using conventional tests. For this reason, we adopt the randomization approach pioneered by Pearson (1950), in which randomization is employed in conjunction with sample conditioning. Specifically, we add randomness to the basic statistic $\hat{D}$ and then derive the asymptotic results conditional on the observed sample. Besides that, we show that the established results hold for all samples, save for a zero-measure set. Thus, we need to derive the almost sure convergence results for $\hat{D}$ under $\mathbb{H}_1$ and $\mathbb{H}_2$, respectively.

Now we show how to test $\mathbb{H}_1$ and $\mathbb{H}_2$ against each other using the randomization approach. Let $\mathcal{Y}_{NT} \equiv \mathcal{Y}_{NT}(\omega)$ be a statistic based on the sample path ω, e.g., $\mathcal{Y}_{NT}(\omega) = T\hat{D}(\omega)^{-1}$, where $\hat{D}(\omega) = \hat{D}$. Suppose we have $\mathcal{Y}_{NT} \overset{a.s.}{\to} \infty$ under the null hypothesis and $\mathcal{Y}_{NT} \overset{a.s.}{\to} y$ for some constant $y > 0$ under the alternative hypothesis. Consider the following procedure:

- Step 1. Generate an i.i.d. random sample $\{\xi_m\}_{m=1}^{M}$ with a common distribution $G(\cdot) = P(\xi_m \leq \cdot)$ such that $G(0) \neq 0$ or 1, and define

$$V_{m,NT}(\omega) = \mathcal{Y}_{NT}(\omega)\xi_m.$$

- Step 2. For any u drawn from a cumulative distribution function (CDF) $\Phi(\cdot)$ with a compact support $\mathbb{U}$, construct

$$Z_{NT}(u,\omega) = \frac{1}{\sqrt{M}}\sum_{m=1}^{M}\frac{\mathbb{I}(V_{m,NT}(\omega) \leq u) - G(0)}{\sqrt{G(0)[1-G(0)]}}.$$

- Step 3. Compute the test statistic

$$S_{NT}(\omega) = \int_{\mathbb{U}} |Z_{NT}(u,\omega)|^2 \, \mathrm{d}\Phi(u).$$

Without loss of generality, we will focus on the case where $G(\cdot)$ is the CDF of the standard normal distribution. With the added randomness via $\{\xi_m\}_{m=1}^M$, under certain regularity conditions for M, we can show that $S_{NT}(\omega)$ converges in distribution to a chi-squared distribution with one degree of freedom almost surely conditional on the sample under the null hypothesis and diverges to infinity at the rate $\sqrt{M}$ under the alternative hypothesis. We now provide some intuitions. Suppose we generate $\{\xi_m\}_{m=1}^M$ using independent standard normal distributions. Then, conditional on the sample, $V_{m,NT}(\omega) \sim N(0, \mathcal{Y}_{NT}^2(\omega))$ for each $m = 1, 2, ..., M$. Note that $\mathcal{Y}_{NT} \overset{a.s.}{\to} \infty$ as $T \to \infty$ under the null hypothesis. Let $\Omega_1 = \{\omega : \mathcal{Y}_{NT}(\omega) \to \infty \text{ as } T \to \infty\}$. Then, $P(\Omega_1) = 1$ and we can restrict our attention to the case where $\omega \in \Omega_1$. Under the null hypothesis, we have that $V_{m,NT}(\omega)$ diverges to $+\infty$ or $-\infty$ for each m with probability approaching (w.p.a.) 1/2, where $G(0) = 1 - G(0) = 1/2$. For any fixed u, we have that the Bernoulli random variable $\mathbb{I}(V_{m,NT}(\omega) \leq u)$ should approach 1 or 0 w.p.a. 1/2, and thus it will have an asymptotic mean 1/2. For this reason, a central limit theorem (CLT) holds such that $Z_{NT}(u, \omega)$ converges in distribution to a standard normal distribution conditional on the sample ω for each fixed u. Therefore, we expect $S_{NT}(\omega)$ to converge to a chi-square distribution with one degree of freedom conditional on the sample ω. On the other hand, when the alternative hypothesis holds, $\mathcal{Y}_{NT} \overset{a.s.}{\to} y$ as $T \to \infty$ for some fixed constant y. Then, $V_{m,NT}(\omega)$ follows a normal distribution with mean zero and variance y^2 conditional on the sample. For any fixed $u \neq 0$, we can show that asymptotic mean of the Bernoulli random variable $\mathbb{I}(V_{m,NT}(\omega) \leq u)$ will differ from 1/2 depending on the location of u. Hence, the mean of $Z_{NT}(u, \omega)$ deviates from 0 conditional on the sample, and it diverges to infinity at the rate $\sqrt{M}$ for each $u \neq 0$. This mechanism ensures the asymptotic power of the test statistic $S_{NT}(\omega)$ under the alternative hypothesis.

The above discussion illustrates how and why one can use the randomization approach to test hypotheses when the order of a statistic varies under the null and alternative hypotheses. This motivates us to use $\hat{D}$ as the basic statistic, given that it has different orders of magnitude under $\mathbb{H}_1$ and $\mathbb{H}_2$. Under certain regularity conditions stated in Section 3.1, we can show that $\hat{D} = O_{a.s.}(1)$ under $\mathbb{H}_1$ and $\hat{D} = O_{a.s.}(T)$ under $\mathbb{H}_2$.

When testing $\mathbb{H}_1$ against $\mathbb{H}_2$, i.e., $\mathbb{H}_1$ is the null and $\mathbb{H}_2$ is the alternative, we let $\mathcal{Y}_{NT}(\omega) = T\hat{D}(\omega)^{-1}$ such that $\mathcal{Y}_{NT} \overset{a.s.}{\to} \infty$ under the null hypothesis and $\mathcal{Y}_{NT} = O_{a.s.}(1)$ under the alternative hypothesis. We then construct a test statistic by

$$S_{NT}^{(1)}(\omega) = S_{NT}(\omega), \tag{2.4}$$

following the three steps described above with $\mathcal{Y}_{NT}(\omega) = T\hat{D}(\omega)^{-1}$. When testing $\mathbb{H}_2$ against $\mathbb{H}_1$, we let $\mathcal{Y}_{NT}(\omega) = \hat{D}(\omega)$. Then the conditions that $\mathcal{Y}_{NT} \overset{a.s.}{\to} \infty$ under the null hypothesis and $\mathcal{Y}_{NT} = O_{a.s.}(1)$ under the alternative hypothesis are also satisfied. Following the same three steps, we define

$$S_{NT}^{(2)}(\omega) = S_{NT}(\omega) \tag{2.5}$$

with $\mathcal{Y}_{NT}(\omega) = \hat{D}(\omega)$.

3 Asymptotic Results

Let P^* represent the probability law governing $\{\xi_m\}_{m=1}^M$, conditional on the sample. Let $\overset{d^*}{\to}$ and E^* denote convergence in distribution and expectation operator respectively under the probability law P^*. Furthermore, we use the notation a.s.-ω to denote conditional on the sample and for all samples except a set of measure zero.

3.1 Assumptions

In this subsection, we first introduce some basic assumptions and then derive the asymptotic distributions of our tests under the respective null hypotheses and study the power properties under the respective alternative hypotheses. Since we do not need to consistently estimate the number of factors, or the factors and factor loadings under either $\mathbb{H}_1$ or $\mathbb{H}_2$, we do not require $N \to \infty$. We mainly rely on the large T asymptotics and allow N to be either divergent or fixed as $T \to \infty$. When both N and T diverge to infinity, we use $(N,T) \to \infty$ to signify that they pass to infinity jointly.

Let $\varsigma_{NT} = \sum_{i=1}^N \sum_{t=1}^T \left[\varepsilon_{it}^2 - E\left(\varepsilon_{it}^2\right)\right]$. Let $\Lambda_t = (\lambda_{1t}, ..., \lambda_{Nt})'$ and $\Sigma_{\Lambda_t} = N^{-1}\Lambda_t'\Lambda_t$. Let $\max_{i,t} = \max_{1\le i\le N}\max_{1\le t\le T}$, and define $\max_i$ and $\max_t$ analogously. Let C be a generic positive constant whose value may vary across places. For an $R \times R$ matrix A, we use $A > 0$ and $A \ge 0$ to denote that A is positive definite and positive semi-definite, respectively.

Assumption A.1 [Factors]: (i) $\{F_t\}_{t=1}^T$ is an $R \times 1$ weakly stationary time series process with $E(F_tF_t') = \Sigma_F > 0$; (ii) $\max_t E\|F_t\|^4 \le C$, and (iii) $\max_{r,l,r_1,l_1}\max_t \sum_{s=1}^T |E\{[F_{rt}F_{lt} - E(F_{rt}F_{lt})]$ $[F_{r_1s}F_{l_1s} - E(F_{r_1s}F_{l_1s})]\}| \le C$.

Assumption A.2 [Error Terms]: (i) $\max_{i,t} E|\varepsilon_{it}|^4 \le C$; (ii) $(NT)^{-1}\sum_{t=1}^T\sum_{i=1}^N E(\varepsilon_{it}^2) \to \sigma_\varepsilon^2$ as $T \to \infty$ or $(N,T) \to \infty$; and (iii) $(NT)^{-1}E\left[\varsigma_{NT}^2\right] \le C$.

Assumption A.3 [Deterministic Time-varying Factor Loadings]: (i) $\{\lambda_{it}\}_{i=1,t=1}^{N,T}$ are nonrandom such that $\max_{i,t} \|\lambda_{it}\| \leq C$; (ii) $T^{-1}\sum_{t=1}^{T}\Sigma_{\Lambda_t} \to \Sigma_\Lambda \geq 0$ as $T \to \infty$, and (iii) $(NT)^{-1}E\left|\sum_{i=1}^{N}\sum_{t=1}^{T}F_t'\lambda_{it}\varepsilon_{it}\right|^2 \leq C$.

Assumption A.4 [Unit Root Factor Loadings]: (i) $\lambda_{it} = \lambda_{i(t-1)} + \nu_{it}$, where $E(\nu_{it}) = 0$, $\max_{i,t} E\|\nu_{it}\|^4 < \infty$, and $\max_i E\|\lambda_{i0}\|^4 < \infty$; (ii) v_{is} and F_t are independent for all (i,t,s); (iii) for each $i = 1,\ldots,N$, there exists an R-dimensional standard Wiener process $W_i(t)$ such that $\max_t N^{-1}\sum_{i=1}^{N}\left\|\lambda_{it} - \Sigma_i^{1/2}W_i(t)\right\|^2 = o_{a.s.}\left(T^{1-2\epsilon_0}\right)$ for some $\epsilon_0 \in (0,1/2)$ and some covariance matrix $\Sigma_i \geq 0$; and (iv) $N^{-1}\sum_{i=1}^{N}\|\Sigma_i\|^2 \leq C$, $N^{-1}\sum_{i=1}^{N}\Sigma_i \to \Sigma > 0$ if $N \to \infty$, and $\Sigma \equiv N^{-1}\sum_{i=1}^{N}\Sigma_i > 0$ if N is fixed.

Assumption A.1 imposes conditions on the latent common factors. Following Stock and Watson (2002), Han and Inoue (2015), and Su and Wang (2017), we assume that $E(F_tF_t') = \Sigma_F$ is homogeneous over t. This suggests 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. It is well known that the 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 one with stationary common factors and time-varying factor loadings. Even if there is no structural change in factor loadings or the second moment of common factors, we can always write $\lambda_{i0}'F_t = \lambda_{i0}'\mathcal{Q}(t/T)^{-1}\mathcal{Q}(t/T)F_t = \breve{\lambda}_{it}'\breve{F}_t$ for any nonsingular square matrix $\mathcal{Q}(t/T)$ with $\breve{\lambda}_{it} \equiv \left[\mathcal{Q}(t/T)^{-1}\right]'\lambda_{i0}$ and $\breve{F}_t \equiv \mathcal{Q}(t/T)F_t$ being the time-varying factor loadings and common factors with time-varying second moments. Assumption A.1(i) rules out this possibility. Assumption A.1(ii) is a moment restriction commonly adopted in the factor model analysis. Assumption A.1(iii) imposes weak dependence conditions on the process $\{F_t\}$, which can be verified under various weak dependence conditions, including strong mixing, mixingale, near-Epoch dependence, and weak functional dependence; see the Section A.3.2 in the online supplement of Barigozzi and Trapani (2022).

Assumption A.2 provides regularity conditions on the error terms. Assumption A.2(i) imposes a moment restriction on the error process that is weaker than that in classical literature on factor model analysis, such as Bai and Ng (2002) and Bai (2003). Assumption A.2(ii) allows for both time-series and cross-sectional heteroskedasticity and requires the average of unconditional variances to converge to a fixed constant. Assumption A.2(iii) imposes weak dependence conditions along both the cross-section and time dimensions of the error terms.

Assumption A.3 states the conditions on the time-varying factor loadings under $\mathbb{H}_1$. Assumption A.3(i) requires that the factor loadings be nonrandom and uniformly bounded. These conditions are similar to those in Su and Wang (2017) and can be relaxed at the cost of more lengthy arguments. Note that we do not need to estimate the time-varying factor loadings consistently. Thus, we avoid imposing regularity conditions such as those in the smooth structural change as in Su and Wang (2017) or abrupt structural breaks as in Breitung and Eickmeier (2011), Chen et al. (2014), and Han and Inoue (2015). Furthermore, Assumption A.3(i) also covers a special case that $\lambda_{it} = \lambda_{i0}$, which is a constant parameter vector over time. Assumption A.3(ii) is weak, and it does not require the limit matrix Σ_Λ to be positive definite. This implies that we allow for weak factors under $\mathbb{H}_1$. Like Assumption A.2(iii), Assumption A.3(iii) imposes weak dependence conditions on $\{F_t'\lambda_{it}\varepsilon_{it}\}$ along both the cross-section and time dimensions.

Assumption A.4 provides the regularity conditions on the time-varying factor loadings under $\mathbb{H}_2$. Assumption A.4(i) depicts a unit root process without drift. The reason that we do not consider a drift term is two-fold. One is that when a drift term is present, the data matrix will exhibit apparent explosive features. So one may directly distinguish $\mathbb{H}_1$ from $\mathbb{H}_2$ by simply plotting the time series of X_{it}. Second, we can show that both our tests $S_{NT}^{(1)}(\omega)$ and $S_{NT}^{(2)}(\omega)$ are valid when a drift term exists. We will elaborate this issue in Section 4.2. Assumption A.4(ii) imposes independence between v_{is} and F_t. Assumption A.4(iii) implies that the innovation process $\{v_{it}\}$ obeys a version of functional central limit theorem where Σ_i is associated with the long-run variance of $\{v_{it}\}$. It can be verified under various primitive weak-dependence conditions on $\{v_{it}\}$ along the time dimension. For example, when $\{v_{it}, t \geq 1\}$ is a linear process as specified in Phillips and Solo (1992) satisfying certain regularity conditions, one can apply Corollary 3.7 in Liu and Lin (2009) and show that $\max_t \left\|\lambda_{it} - \Sigma_i^{1/2} W_i(t)\right\| = o_{a.s.}\left(T^{1/3}\log\log T\right)$ for each i. Similarly, when $\{v_{it}, t \geq 1\}$ is a mixingale process satisfying certain regularity conditions, one can apply Theorem 2 in Eberlein (1986) and show that $\max_t \left\|\lambda_{it} - \Sigma_i^{1/2} W_i(t)\right\| = o_{a.s.}\left(T^{1/2-\epsilon_0}\right)$ for some $\epsilon_0 > 0$. Assumption A.4(iii) restricts that the latter order also holds when one averages over i, which is automatically satisfied if N is fixed. Note that we allow $\{v_{it}\}$ to be correlated over i so that $\{W_i(t)\}$ can be dependent over i. In addition, we do not assume Σ_i to be positive definite for each i but do require that the cross-sectional average of Σ_i be asymptotically positive definite in Assumption A.4(iv).

3.2 Asymptotic results

To establish the asymptotic results for the proposed test statistics $S_{NT}^{(1)}(\omega)$ and $S_{NT}^{(2)}(\omega)$, we need to show the strong convergence results for $\hat{D}$ under both $\mathbb{H}_1$ and $\mathbb{H}_2$.

Proposition 1. *(i) Suppose Assumptions A.1 to A.3 hold. Then* $\hat{D} = D_1 + o_{a.s.}(1)$, *where* $D_1 = \mathrm{tr}\left(\Sigma_\Lambda \Sigma_F\right) + \sigma_\varepsilon^2 > 0$.
(ii) Suppose Assumptions A.1, A.2, and A.4 hold. Then $T^{-1}\hat{D} = D_2 + o_{a.s.}(1)$, *where* $D_2 = \frac{1}{2}\mathrm{tr}\left(\Sigma \Sigma_F\right) > 0$.

We need to establish the almost sure convergence because the asymptotic distributions of our randomized test statistics are derived under the probability law of $\{\xi_m\}_{m=1}^M$ conditional on the observed sample. Thus, we need to ensure that the obtained asymptotic results hold for all sample paths except for those with measure zero. With Proposition 1, we now provide the asymptotic results for the test statistics $S_{NT}^{(1)}(\omega)$ and $S_{NT}^{(2)}(\omega)$, respectively.

When testing $\mathbb{H}_1$ against $\mathbb{H}_2$, $\mathcal{Y}_{NT}(\omega) = T\hat{D}^{-1}(\omega)$. Recall that $\Omega_1 \equiv \{\omega : \mathcal{Y}_{NT}(\omega) \to \infty\}$ and let $\Omega_2 \equiv \{\omega : \mathcal{Y}_{NT}(\omega) \to D_2^{-1}\}$. Proposition 1 implies that $P\left(\Omega_1\right) = 1$ under Assumptions A.1–A.3, and $P\left(\Omega_2\right) = 1$ under Assumptions A.1, A.2, and A.4. Let $\lfloor \cdot \rfloor$ denote the integer part of $\cdot$.

Theorem 1. *Let* $M = \lfloor T^a \rfloor$ *with* $0 < a < 2$.
(i) $S_{NT}^{(1)}(\omega) \xrightarrow{d^*} \chi_1^2$, *a.s.-*$\omega \in \Omega_1$ *as* $T \to \infty$ *or* $(N, T) \to \infty$ *under Assumptions A.1–A.3.*
(ii) For any nonrandom positive sequence $c_M = o(M)$, $P^*[S_{NT}^{(1)}(\omega) > c_M] \to 1$ *a.s.-*$\omega \in \Omega_2$ *as* $T \to \infty$ *or* $(N, T) \to \infty$ *under Assumptions A.1, A.2, and A.4.*

The limiting distribution should be understood in the following sense: conditional on the sample $\omega \in \Omega_1$ with $P\left(\Omega_1\right) = 1$, $S_{NT}^{(1)}(\omega)$ has a well-defined limiting distribution in terms of the law governing the added randomness under the null hypothesis $\mathbb{H}_1$, and it diverges under the alternative hypothesis $\mathbb{H}_2$. As explained by Corradi and Swanson (2006), the notion of size is different from the standard one: classically, the significance level α of a test means that if a researcher applies the test B times under the null hypothesis, then (s)he will falsely reject the null hypothesis $\lfloor \alpha B \rfloor$ times on average. In contrast, in the context of randomized tests, α means that out of J researchers who apply the test, about $\lfloor \alpha J \rfloor$ of them will reject the null when this is true. Still, in our article, we obtain a test statistic which, for a given level α, rejects the null with probability α when it is true and with probability 1 when it is false as $T \to \infty$ or $(N, T) \to \infty$.

It is worth mentioning the impact of M on the asymptotic results. Theorem 1(ii) indicates that $S^{(1)}_{NT}(\omega)$ diverges to infinity at the rate M. Although the asymptotic power of the test is positively related to M, we note that a large M may distort the finite sample size performance. Note that we require $M \to \infty$, but at a slower rate than that of T^2. This condition ensures that the test statistic follows an asymptotic chi-squared distribution under $\mathbb{H}_1$ and diverges under $\mathbb{H}_2$. Furthermore, we can show that the asymptotic power of the test statistic under the alternative hypothesis depends positively on a location parameter $\int_{\mathbb{U}} |F(u) - G(0)|^2 \mathrm{d}\Phi(u)$, where $F(\cdot)$ is the limit CDF of $V_{m,NT}(\omega)$ under $\mathbb{H}_2$. Note that $F(\cdot)$ is the CDF of $N\left(0, D_2^{-1}\right)$ when $G(\cdot)$ is the CDF of the standard normal distribution. In this case, the larger $|u|$ is, the larger $|F(u) - 1/2|^2$ will be. In practice, one can choose a discrete random variable to generate u, e.g., a Bernoulli random variable taking values on $-c$ and c with probability $1/2$ for some $c > 0$. Choosing a large c is beneficial for the power performance of the test. However, it may harm the finite sample size performance if c is too large. For details, please see the proof of Theorem 1.

Analogously, we provide the asymptotic results for the test statistic $S^{(2)}_{NT}(\omega)$.

Theorem 2. *Let* $M = \lfloor T^a \rfloor$ *with* $0 < a < 2$.
(i) $S^{(2)}_{NT}(\omega) \xrightarrow{d^*} \chi^2_1$, *a.s.-*$\omega \in \Omega_2$ *as* $T \to \infty$ *or* $(N,T) \to \infty$ *under Assumptions A.1, A.2, and A.4.*
(ii) For any nonrandom positive sequence $c_M = o(M)$, $P^*[S^{(2)}_{NT}(\omega) > c_M] \to 1$ *a.s.-*$\omega \in \Omega_1$ *as* $T \to \infty$ *or* $(N,T) \to \infty$ *under Assumptions A.1–A.3.*

The intuition for the results in Theorem 2 is similar to that of Theorem 1. To put it simply, we use the fact that $\hat{D}$ diverges to infinity at the rate T almost surely under $\mathbb{H}_2$ but converges to a fixed constant under $\mathbb{H}_1$. Following analogous steps as in the proof of Theorem 1, we can establish Theorem 2 straightforwardly. Note that both test statistics follow an asymptotic pivotal chi-squared distribution under the respective null hypotheses. Hence, one can use the asymptotic critical values without relying on any data-dependent resampling method. In the simulation studies, we examine the finite sample performance of both tests using the asymptotic critical values.

The proposed randomization tests are especially suitable for the goal of this paper. Using the randomization approach, one does not need to estimate the factor model since the key input for both test statistics is $\hat{D}$, which is the trace of $(NT)^{-1}XX'$. The key difference between $\mathbb{H}_1$ and $\mathbb{H}_2$ is captured by the distinct orders of magnitude of $\hat{D}$ under the two specifications. The randomization approach enables us to test the order of $\hat{D}$ within a single observed sample using the generated randomness. As a result, we avoid the issue of consistent estimation for the common factors or factor loadings under either type of specifications. Moreover, one does not need to worry about estimating the number of common factors consistently under either $\mathbb{H}_1$ or $\mathbb{H}_2$. Although Su and Wang (2017) provide a consistent estimator for the number of common factors under $\mathbb{H}_1$ with additional smooth conditions, consistent estimation for the number of common factors under $\mathbb{H}_2$ is void in the literature. Besides, as mentioned above, we do not need $N \to \infty$, and when both N and T diverge to infinity, there is no need to specify any requirement on the relative rates at which N and T diverge to infinity.

Furthermore, with slight modifications to the basic statistic of the randomized test, we can not only test $\mathbb{H}_1$ against $\mathbb{H}_2$ but also do the opposite. In this sense, our tests are comprehensive. It can provide a clear conclusion for the practitioners when adopting a suitable specification for time-varying factor loadings.

4 Discussions

In this section, we discuss some empirically relevant issues associated with the proposed randomization tests. One issue is that one needs to conduct a certain transformation of the data matrix before applying the proposed tests. The other issue is that when the factor loadings are time-varying in a deterministic way but with weak factors, or the factor loadings follow unit root processes with a drift, we show that the proposed tests are still valid.

4.1 Scale invariance

Note that the finite sample performance of each test relies on the realized value of $\hat{D}$. With one observed sample, one can only have one observation for $\hat{D}$, which can be sensitive to the scale of data. For instance, we can generate a factor model under $\mathbb{H}_1$ but amplify the generated data using a large constant. Then, our test $S_{NT}^{(1)}(\omega)$ may falsely reject the null hypothesis since $\hat{D}$ behaves like an unbounded statistic when inflated by a large constant in a finite sample. To avoid such a drawback, we suggest that a practitioner conducts a rescaling of the data before applying the proposed tests. This is a common practice for tests based on randomization. For example, when monitoring structural changes in a factor model, Barigozzi and Trapani (2020) sequentially examine the order of magnitude of the $(r+1)$-th sample eigenvalue, where r is the estimated number of common factors in the stable period. To ensure scale invariance, they consider the ratio of the $(r+1)$-th sample eigenvalue to the sum of all sample eigenvalues. When testing for the proper transformation of data, Corradi and Swanson (2006) rescale the basic statistic using the first observation to achieve invariance to scalar multiplication of the data.

In this paper, we adopt a similar approach to that of Corradi and Swanson (2006) to ensure scale invariance. Specifically, we first compute the cross-sectional average and sample standard deviation of the first time series vector observation:

$$\bar{X}_1 = \frac{1}{N}\sum_{i=1}^{N} X_{i1}, \text{ and } \hat{\sigma}_1 = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)^2}.$$

Then, we standardize the data matrix by $\tilde{X}_{it} = (X_{it} - \bar{X}_1)/\hat{\sigma}_1$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, and then compute $\tilde{D} = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}^2$ using the transformed data $\{\tilde{X}_{it}\}_{i=1,t=1}^{N,T}$. Using such a procedure, the observed value of $\tilde{D}$ is invariant to level shifting or scalar multiplication of the data.

More importantly, we can show that the order of magnitude of $\tilde{D}$ is identical to that of $\hat{D}$ computed based on the untransformed data. Using straightforward algebra, we can show that the order of magnitude of $\tilde{D}$ is asymptotically equivalent to $(\hat{D} - L)/\hat{\sigma}_1^2$, where L is a location term that depends on $\bar{X}_1$. Under Assumptions 1–3, we have $\bar{X}_1 = O_{a.s.}(1)$ and $\hat{\sigma}_1 = O_{a.s.}(1)$. Thus, $\tilde{D} = O_{a.s.}(1)$. While under Assumptions 1, 2, and 4, we can still show that $\bar{X}_1 = O_{a.s.}(1)$ and $\hat{\sigma}_1 = O_{a.s.}(1)$ since the first observation of a unit root process does not exhibit any explosive feature. For this reason, the asymptotic leading term of $\tilde{D}$ is proportional to that of $\hat{D}$, which diverges to infinity almost surely. Hence, we conclude that under either $\mathbb{H}_1$ or $\mathbb{H}_2$, such a data transformation will not change the order of magnitude of $\hat{D}$. In this way, the test statistic is invariant to level shifting or scalar multiplication of the data. Simulation studies demonstrate the excellent finite sample performance of the proposed scale invariance transformation approach.

4.2 Weak factors and unit root with a drift

In this subsection, we show that the proposed tests are still valid when the factors are weak under $\mathbb{H}_1$ or when the unit root process under $\mathbb{H}_2$ has a drift.

First, we consider the presence of weak factors under $\mathbb{H}_1$. As mentioned in the remark on Assumption A.3(ii), we allow Σ_Λ to be positive semi-definite or singular, which corresponds to the presence of weak factors under $\mathbb{H}_1$. In the extreme case where $\Sigma_\Lambda = 0$, we have $D_1 = \sigma_\varepsilon^2 > 0$, which suffices to ensure $T\hat{D}$ to be $O_{a.s.}(T)$ under $\mathbb{H}_1$. As a result, the conclusions in Theorems 1 and 2 remain true in such an extreme case.

Now, we consider the presence of a drift term in the unit root process under $\mathbb{H}_2 : \lambda_{it} = \mu_i + \lambda_{i(t-1)} + \nu_{it}$, where $\mu_i \neq 0$ for some $i = 1, \ldots, N$. Let $\bar{\Sigma}_\mu = N^{-1} \sum_{i=1}^N \mu_i \mu_i'$. Furthermore, let $\Sigma_\mu = \bar{\Sigma}_\mu$ when N is fixed and $\Sigma_\mu = \lim_{N\to\infty} \bar{\Sigma}_\mu$ when $N \to \infty$. For simplicity, we consider the case where $\Sigma_\mu > 0$. We first consider testing $\mathbb{H}_1$ against $\mathbb{H}_2$ using $S_{NT}^{(1)}(\omega)$. Noting that $\hat{D} \overset{a.s.}{\to} D_1$ under $\mathbb{H}_1$, we have $\mathcal{Y}_{NT} = T\hat{D}^{-1} \overset{a.s.}{\to} \infty$ under the null hypothesis. As a result, the CLT for $Z_{NT}^0(u, \omega)$ defined in the proof of Theorem 1 still holds and $S_{NT}^{(1)}(\omega)$ converges in distribution to χ_1^2. Hence, the test $S_{NT}^{(1)}$ displays the correct asymptotic size using the asymptotic critical values. In contrast, when $\mathbb{H}_2$ holds with a drift term, following the proof of Proposition 1, we can show that

$$T^{-2}\hat{D} = T^{-2}\hat{D}_1 + o_{a.s.}(1) \overset{a.s.}{\to} D_3 > 0 \tag{4.1}$$

where

$$\begin{aligned} T^{-2}\hat{D}_1 &= \frac{1}{NT^3} \sum_{i=1}^N \sum_{t=1}^T F_t' \lambda_{it} \lambda_{it}' F_t = \frac{1}{NT^3} \sum_{i=1}^N \sum_{t=1}^T t^2 F_t' \mu_i \mu_i' F_t + o_{a.s.}(1) \\ &= \operatorname{tr}\left(\bar{\Sigma}_\mu \Sigma_F\right) \frac{1}{T^3} \sum_{t=1}^T t^2 + o_{a.s.}(1) \overset{a.s.}{\to} \frac{1}{3} \operatorname{tr}\left(\Sigma_\mu \Sigma_F\right) \equiv D_3. \end{aligned}$$

It follows that $\mathcal{Y}_{NT} = T\hat{D}^{-1} = O_{a.s.}\left(T^{-1}\right)$. Then, $V_{m,NT}(\omega) = \mathcal{Y}_{NT}(\omega)\xi_m$ converges to zero for each m and each $\omega \in \Omega_3 = \{\omega : T\mathcal{Y}_{NT}(\omega) \to D_3^{-1}\}$. For each fixed $u \neq 0$, the Bernoulli random variable $\mathbb{I}(V_{m,NT}(\omega) \leq u)$ should be 0 or 1 w.p.a. 0 or 1 depending on the sign of u. Thus, we can show that $\mathcal{Z}_{NT}(u, \omega)$ will not converge to a normal distribution since the asymptotic mean of $\mathbb{I}(V_{m,NT}(\omega) \leq u)$, which equals 0 or 1 depending on the sign of u, is different from $G(0)$. Such a mechanism ensures that $S_{NT}^{(1)}(\omega)$ has power against $\mathbb{H}_2$ in which the unit root factor loading process has a drift term. In particular, the result in Theorem 1(ii) still holds.

Next, we investigate the test of $\mathbb{H}_2$ against $\mathbb{H}_1$ using $S_{NT}^{(2)}(\omega)$ when the unit-root type factor loading process under $\mathbb{H}_2$ has a drift term. When $\Sigma_\mu > 0$, $\mathcal{Y}_{NT} = \hat{D}$ diverges to infinity almost surely at the rate T^2 under the null hypothesis by (4.1). Following the arguments used in the proof of Theorem 1, we can show that $S_{NT}^{(2)}(\omega)$ still follows an asymptotic chi-squared distribution with one degree of freedom conditional on the sample ω. Under the alternative hypothesis, $\mathcal{Y}_{NT} = \hat{D} \overset{a.s.}{\to} D_1 > 0$, and the result in Theorem 2(ii) continues to hold.

In sum, the above two paragraphs indicate that the theory developed in Theorems 1 and 2 continues to hold when there is a drift term in the unit root process for the time-varying factor loadings under $\mathbb{H}_2$. This is also the case when we have weak factors under $\mathbb{H}_1$ and a unit root process with a drift term under $\mathbb{H}_2$. Consequently, the results in Theorems 1 and 2 are robust to the presence of weak factors under $\mathbb{H}_1$ or/and a unit root process with a drift term under $\mathbb{H}_2$.

The above discussions imply that as long as the orders of magnitude differ under two hypotheses, $\mathbb{H}_1$ and $\mathbb{H}_2$, it is possible to propose a pair of randomization tests to distinguish them. For example, we can consider the specification of a unit-root type factor loading process without a drift under $\mathbb{H}_2$ versus the specification of a unit-root factor loading process with a drift under $\mathbb{H}_{2'}$. Since $\hat{D}$ diverges to infinity at rates T and T^2 under $\mathbb{H}_2$ and $\mathbb{H}_{2'}$, respectively, one can test $\mathbb{H}_2$ against $\mathbb{H}_{2'}$ by using $\mathcal{Y}_{NT} = T^2\hat{D}^{-1}$ in the testing procedure and test $\mathbb{H}_{2'}$ against $\mathbb{H}_2$ by using $\mathcal{Y}_{NT} = T^{-1}\hat{D}$ in the testing procedure. Then the resulting test statistics have the same asymptotic properties as stated in Theorems 1 and 2.

5 Monte Carlo Studies

In this section, we conduct simulation studies to examine the finite sample performance of the proposed tests.

5.1 Data generating processes

We generate the data under the following large-dimensional factor model with $R = 2$ common factors:

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

for $i = 1, ..., N$ and $t = 1, ..., T$, where $F_t \equiv (F_{1t}, F_{2t})'$, with $F_{1t} = 0.2 + 0.6F_{1(t-1)} + z_{1t}$, $z_{1t} \sim i.i.d.N(0, 1 - 0.6^2)$, and $F_{2t} = 0.2 + 0.3F_{2(t-1)} + z_{2t}$, $z_{2t} \sim i.i.d.N(0, 1 - 0.3^2)$.

Let $\lambda_{i0} = (\lambda_{i0,1}, \lambda_{i0,2})'$ with $\lambda_{i0,1} \sim i.i.d.N(0, 1)$ and $\lambda_{i0,2} \sim i.i.d.U(0, 1)$ for $i = 1, ..., N$. Set $b_1 = 0.2$ and $b_2 = 0.5$, which are used to define the magnitude of structural changes below. Let $\mathcal{G}(y; \varsigma, \gamma) = [1 + e^{-\varsigma \prod_{l=1}^{p}(y-\gamma_l)}]^{-1}$ be the logistic function with the scale parameter ς and location parameter vector $\gamma = (\gamma_1, ..., \gamma_p)'$. To examine the size and power performance of the proposed tests, we consider the following designs for the factor loading $\lambda_{it} \equiv (\lambda_{it,1}, \lambda_{it,2})'$:

DGP 1: $\lambda_{it} = \lambda_{i0}$ for all $t = 1, \ldots, T$;

DGP 2: $\lambda_{it,k} = \begin{cases} \lambda_{i0,k}, & \text{for } 1 \le t \le 0.5T \\ \lambda_{i0,k} + b_k, & \text{for } 0.5T < t \le T \end{cases}$, for $k = 1, 2$;

$$\text{DGP 3: } \lambda_{it,1} = \begin{cases} \lambda_{i0,1}, & \text{for } 0.1T < t \leq 0.2T \text{ and } 0.7T < t \leq 0.8T \\ \lambda_{i0,1} + b_1, & \text{for } 0.4T < t \leq 0.5T \\ \lambda_{i0,1} - b_2, & \text{otherwise} \end{cases}, \text{ and}$$

$$\lambda_{it,2} = \begin{cases} \lambda_{i0,2}, & \text{for } 1 \leq t \leq 0.6T \\ \lambda_{i0,2} + b_2, & \text{for } 0.6T < t \leq T \end{cases};$$

$$\text{DGP 4: } \lambda_{it,k} = b_k \lambda_{i0,k} g_{k,t} \text{ for } k = 1, 2, \text{ where } \begin{cases} g_{1,t} = g_1(t/T) = 3[1 - e^{-3(t/T)^2}] \\ g_{2,t} = g_2(t/T) = [1 + e^{20(t/T-0.5)}]^{-1} \end{cases};$$

$$\text{DGP 5: } \lambda_{it,1} = \lambda_{i0,k} + b_k h_{k,t} \text{ for } k = 1, 2, \text{ where } \begin{cases} h_{1,t} = h_1(t/T) = [2 - e^{-4(t/T-0.5)^2}] \\ h_{2,t} = h_2(t/T) = \sin(2\pi t/T) \end{cases};$$

$$\text{DGP 6: } \lambda_{it,k} = \begin{cases} N^{-0.2}\lambda_{i0,k}, & \text{for } 1 \leq t \leq 0.5T \\ N^{-0.2}(\lambda_{i0,k} + b_k), & \text{for } 0.5T < t \leq T \end{cases}, \text{ for } k = 1, 2;$$

DGP 7: $\lambda_{it} = \lambda_{i(t-1)} + \nu_{it}$, where $\nu_{it} \sim i.i.d.N(0, I_2)$.

DGP 8: $\lambda_{it} = \mu_i + \lambda_{i(t-1)} + \nu_{it}$, where $\nu_{it} \sim i.i.d.N(0, I_2)$ and $\mu_i = (\mu_{i1}, \mu_{i2})'$ with $\mu_{ik} \sim i.i.d.U(0, 1)$ for $k = 1, 2$.

For each DGP, we consider five cases of error terms $\{\varepsilon_{it}\}$: (i) the i.i.d. case, where $\varepsilon_{it} \sim i.i.d.N(0, 1)$; (ii) the heteroskedastic case, where ε_{it} follows independent $N(0, \sigma_{it}^2)$ distribution with $\sigma_{it} = v_t w_i$, where $v_t = 3\left[1 + \exp\left(-10\,(t/T - 0.5)^2\right)\right]$ for $t = 1, ..., T$, and $w_i \sim i.i.d.U(0.5, 1.5)$ for $i = 1, ..., N$; (iii) the cross-sectionally dependent case, where $\varepsilon_t \sim i.i.d.N(0, \Sigma_\varepsilon)$; (iv) the serially dependent case, where $\varepsilon_{it} = 0.5\varepsilon_{i(t-1)} + v_{it}$ with $v_{it} \sim i.i.d.N(0, 1)$; (v) the cross-sectionally and serially dependent case, where $\varepsilon_t = 0.5\varepsilon_{t-1} + v_t, v_t \sim i.i.d.N(0, \Sigma_v)$. Note that we let $\Sigma_\varepsilon = \Sigma_v = (c_{ij})_{i,j=1,...,N}$ with $c_{ij} = 0.5^{|i-j|}$ for cases (iii) and (v).

DGPs 1–6 satisfy $\mathbb{H}_1$. DGP 1 depicts a time-invariant factor model, which can be viewed as a special case of $\mathbb{H}_1$. We note that there exists intensive literature on testing constancy of factor loadings in a static factor model, e.g., Breitung and Eickmeier (2011), Chen et al. (2014), Han and Inoue (2015), Su and Wang (2017, 2020), and Fu et al. (2022b). DGPs 2–6 describe various specifications of time-varying factor loadings. DGP 2 admits a single abrupt structural break in the factor loadings. Such a type of factor loadings has been widely adopted in the related literature, e.g., Breitung and Eickmeier's (2011) equation (4), Han and Inoue's (2015) DGP.A1, Su and Wang's (2017) DGP.4, and Su and Wang's (2020) DGP.P1. DGP 3 allows for multiple abrupt structural breaks, adopted by Su and Wang (2017, 2020) and Fu et al. (2022b). DGP 4 depicts monotonic smooth structural changes, in which the first and second factor loadings $\lambda_{it,1}$ and $\lambda_{it,2}$ are monotonically increasing and decreasing functions of the rescaled time index t/T, respectively. DGP 5 describes non-monotonically changing time-varying factor loadings. Moreover, DGP 6 allows us to examine the size and power performance of the proposed tests when the time-varying factor loadings are weak since the sample covariance matrix of the factor loadings shrinks to 0 as the sample size grows before and after the break. DGPs 7 and 8 describe time-varying factor models with a unit root type of factor loadings. DGP 7 does not contain a drift term, which satisfies $\mathbb{H}_2$. DGP 8 does not satisfy either $\mathbb{H}_1$ or $\mathbb{H}_2$ since it depicts a unit root process with a drift. However, as we have explained in Section 4.2, we can show that both tests $S_{NT}^{(1)}(\omega)$ and $S_{NT}^{(2)}(\omega)$ are asymptotically valid under DGP 8.

5.2 Simulation results

When testing $\mathbb{H}_1$ against $\mathbb{H}_2$, we use DGPs 1–6 to examine the size performance and DGPs 7–8 to examine the power performance of $S_{NT}^{(1)}(\omega)$. In contrast, when testing $\mathbb{H}_2$ against $\mathbb{H}_1$, we use DGPs 7–8 to examine the size performance and DGPs 1–6 to examine the power performance of $S_{NT}^{(2)}(\omega)$. We set the number of replications to be 1000. We consider various combinations of sample sizes $N = 100, 200$ and $T = 100, 200$. We generate $\{\xi_m\}_{m=1}^{M}$ using independent standard normal distributions, adopt a binary distribution for $\Phi(\cdot)$, which has probability mass 1/2 at $\sqrt{2}$ and $-\sqrt{2}$, and set $M = \lfloor T^a \rfloor$ with $a = 0.8, 1$, and 1.2. We construct the test statistics $S_{NT}^{(1)}(\omega)$ and $S_{NT}^{(2)}(\omega)$ according to the steps described in Section 2.2. We use the asymptotic critical values of χ_1^2 and examine the size and power performance of the proposed tests at the 5% and 10% significance levels.

Tables 1–3 report the size and power performance of $S_{NT}^{(1)}(\omega)$ when testing $\mathbb{H}_1$ against $\mathbb{H}_2$. Note that DGPs 1–6 satisfy $\mathbb{H}_1$, describing various types of deterministic time-varying factor loadings. The empirical rejection rates of $S_{NT}^{(1)}(\omega)$ are close to the corresponding nominal levels. The size performance usually improves with T but not N since the asymptotic theory of our tests is derived based on large T. DGPs 7 and 8 satisfy $\mathbb{H}_2$ without and with a drift. We note that the power performance of $S_{NT}^{(1)}(\omega)$ increases with T since M only depends on T. Furthermore, comparing the power performance of $S_{NT}^{(1)}(\omega)$ under various values of a, we note that the empirical rejection rates increase with a. This is consistent with the established asymptotic results.

Tables 4–6 show the finite sample power and size performance of $S_{NT}^{(2)}(\omega)$ when testing $\mathbb{H}_2$ against $\mathbb{H}_1$. Now, we have that DGPs 7 and 8 satisfy the null hypothesis, but DGPs 1–6 depict time-varying factor loadings under the alternative hypothesis. For each fixed a, we observe that $S_{NT}^{(2)}(\omega)$ exhibits reasonable size performance under DGPs 7 and 8 and satisfactory power performance under DGPs 1–6. Although the considered sample sizes are relatively small, the empirical rejection rates of $S_{NT}^{(2)}(\omega)$ are close to unity under various values of a. This shows the excellent power of the proposed test. Furthermore, $S_{NT}^{(2)}(\omega)$ displays robustness to a unit root process with or without a drift. The empirical rejection rates of $S_{NT}^{(2)}(\omega)$ are close to the nominal levels under both DGPs 7 and 8. Analogous to the results of $S_{NT}^{(1)}(\omega)$ in testing $\mathbb{H}_1$ against $\mathbb{H}_2$, a larger value of a can incur better power performance but an oversize issue. This is consistent with the established asymptotic results.

We have also examined the finite sample performance of the proposed tests under $M = \lfloor T^a \rfloor$ with $a = 1.5$. The power performance of both $S_{NT}^{(1)}(\omega)$ and $S_{NT}^{(2)}(\omega)$ improve since the asymptotic power of the proposed tests is positively related to M. However, the size performance of both tests is not comparable to those under a smaller a. The oversize issue arises since the asymptotic higher-order terms of the proposed tests shrink to zero at a slower rate with a larger M.

6 Empirical Applications

In this section, we consider two empirical applications of the proposed tests in macroeconomics.

6.1 U.S. macroeconomic dataset

We now use the proposed tests to examine the time-varying feature of the U.S. macroeconomic dataset. Most existing studies, such as Stock and Watson (2009), Su and Wang (2017), and Mikkelsen et al. (2019), construct a factor model for this dataset and document that the factor loadings are time-varying. Su and Wang (2017) find strong evidence of structural changes in factor loadings and model the time-varying factor loadings as unknown piece-wise smooth functions of the rescaled time. Mikkelsen et al. (2019) propose a time-varying factor model where the factor loadings evolve as VAR processes and then apply the model to the dataset of Stock and Watson (2009). For the majority of the variables, they find evidence of time-varying factor loadings and show that a large increase in the in-sample fit of the common component can be obtained by incorporating time-varying factor loadings. Furthermore, several empirical studies apply the time-varying factor model to extract valuable information based on the U.S. macroeconomic dataset and explore various interesting applications. For example, Baumeister et al. (2013) and Korobilis (2013) extract common factors

Table 1: Empirical rejection rates of $S_{NT}^{(1)}(\omega)$ with $M = \lfloor T^{0.8} \rfloor$

N	T	DGP 1		DGP 2		DGP 3		DGP 4		DGP 5		DGP 6		DGP 7		DGP 8	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
i.i.d. errors																	
100	100	6.2	11.4	4.3	9.9	3.9	9.6	5.8	10.1	4.4	9.3	4.3	9.5	77.0	86.0	100	100
200	100	4.7	8.0	4.8	9.6	4.5	9.9	6.0	11.8	3.4	8.4	5.8	12.1	76.4	85.2	100	100
100	200	5.0	9.7	5.3	9.6	6.0	10.2	5.8	10.9	5.7	10.8	4.3	9.7	87.4	92.3	100	100
200	200	6.5	12.6	4.4	11.0	3.9	9.3	4.7	9.8	5.2	10.0	4.7	9.4	86.9	91.5	100	100
heteroskedastic errors																	
100	100	4.8	10.5	4.8	10.1	4.7	9.2	5.2	9.8	5.5	10.3	4.2	10.0	73.2	83.5	100	100
200	100	4.0	8.9	4.1	9.1	4.1	8.7	3.5	8.4	4.5	10.2	4.7	9.0	73.8	82.4	100	100
100	200	5.6	11.7	5.1	10.2	5.5	11.1	3.8	8.1	5.3	10.7	5.5	10.7	88.5	92.6	100	100
200	200	4.2	8.9	4.7	11.1	5.1	10.4	4.6	10.1	5.0	10.6	5.0	10.0	85.5	92.6	100	100
cross-sectionally dependent errors																	
100	100	3.0	10.2	4.2	10.7	3.8	8.3	3.8	9.5	4.4	9.9	4.6	8.9	75.7	85.4	100	100
200	100	3.8	9.1	4.8	8.9	5.6	10.9	5.2	9.9	5.2	11.4	5.2	10.0	76.1	84.2	100	100
100	200	4.6	9.1	3.5	9.1	6.6	11.7	4.3	8.3	4.7	9.7	6.6	12.2	87.0	92.3	100	100
200	200	5.3	11.2	4.4	8.4	3.8	9.8	5.7	10.3	3.8	10.2	5.5	10.3	86.5	92.3	100	100
serially dependent errors																	
100	100	5.2	9.7	4.2	7.4	4.5	9.2	4.4	9.2	4.6	10.6	5.3	9.9	74.8	82.5	100	100
200	100	3.5	7.7	4.5	8.3	5.0	10.0	4.5	10.3	4.4	9.1	5.1	10.1	73.9	83.5	100	100
100	200	3.8	9.3	4.7	9.2	3.3	8.3	5.6	10.0	4.0	8.0	5.8	10.3	80.8	89.4	100	100
200	200	5.0	9.7	5.7	12.1	4.6	10.6	4.5	8.6	3.4	8.4	5.3	9.9	85.0	91.1	100	100
cross-sectionally and serially dependent errors																	
100	100	3.2	7.6	3.2	8.4	4.8	9.9	4.5	9.5	4.8	9.0	5.0	10.8	71.4	82.5	100	100
200	100	5.2	9.8	5.5	9.8	4.7	9.3	4.8	9.1	4.0	8.9	5.5	10.8	76.0	84.2	100	100
100	200	6.7	11.7	5.4	9.2	3.8	8.9	5.1	9.7	5.4	10.4	5.1	10.4	85.4	91.9	100	100
200	200	3.9	8.1	5.6	9.9	5.4	9.9	4.9	9.8	4.0	9.1	4.4	8.9	85.3	91.7	100	100

Note: The entries report the percentage of rejections over 1000 replications.

Table 2: Empirical rejection rates of $S_{NT}^{(1)}(\omega)$ with $M = T$

N	T	DGP 1		DGP 2		DGP 3		DGP 4		DGP 5		DGP 6		DGP 7		DGP 8	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
i.i.d. errors																	
100	100	5.2	10.0	5.4	9.5	5.5	10.7	4.4	9.8	5.5.	11.6	6.3	11.7	91.1	95.5	100	100
200	100	3.9	7.3	5.3	11.2	5.5	9.5	5.2	9.4	4.8	10.1	5.6	9.7	93.5	96.3	100	100
100	200	4.6	8.2	5.6	9.2	4.7	9.5	6.2	10.3	4.7	8.8	5.1	10.1	97.4	98.4	100	100
200	200	4.8	9.4	4.0	8.8	5.9	10.2	6.1	10.2	4.7	11.1	5.4	9.6	97.1	98.9	100	100
heteroskedastic errors																	
100	100	6.2	10.2	3.7	9.4	5.0	10.5	4.9	10.4	4.0	9.9	4.8	9.4	91.9	95.8	100	100
200	100	4.2	9.8	5.1	9.9	3.9	9.6	5.1	9.7	6.6	11.1	5.5	11.2	92.0	95.2	100	100
100	200	4.6	9.4	4.5	9.7	5.8	10.3	4.7	9.5	6.4	11.2	4.8	9.1	97.8	99.3	100	100
200	200	5.6	11.0	4.3	10.0	5.4	9.4	6.4	10.4	4.1	8.8	5.6	9.7	96.4	98.3	100	100
cross-sectionally dependent errors																	
100	100	4.3	9.0	6.4	9.8	3.7	8.8	5.5	10.2	4.6	10.2	5.0	9.8	92.5	95.0	100	100
200	100	4.2	9.0	3.5	8.7	4.9	9.7	5.1	10.7	5.5	10.3	5.9	10.8	92.0	95.6	100	100
100	200	5.7	10.1	4.4	8.7	5.6	9.6	4.8	9.7	4.8	8.7	5.6	11.1	98.2	99.3	100	100
200	200	5.2	10.6	5.0	11.4	4.9	8.5	5.8	10.4	3.9	8.1	5.9	10.0	98.0	99.1	100	100
serially dependent errors																	
100	100	5.7	10.1	5.2	11.0	5.3	9.7	5.0	10.0	4.5	10.1	5.8	9.5	89.2	93.6	100	100
200	100	4.7	9.2	5.4	10.3	4.9	10.9	5.8	10.2	4.4	9.1	6.1	10.6	91.7	95.5	100	100
100	200	5.8	10.9	4.3	8.5	4.7	10.2	6.2	10.4	5.7	11.7	5.1	9.9	97.0	98.7	100	100
200	200	6.1	10.7	5.0	8.6	4.4	9.0	4.1	8.4	5.2	8.9	4.7	8.9	96.8	98.2	100	100
cross-sectionally and serially dependent errors																	
100	100	4.2	10.0	4.4	10.1	4.0	9.8	4.2	10.1	6.6	9.9	5.1	10.2	92.9	96.3	100	100
200	100	5.0	9.9	3.6	7.9	4.3	9.0	4.8	10.0	6.3	9.7	6.1	11.6	90.6	94.5	100	100
100	200	4.1	8.5	5.3	10.5	4.0	8.7	5.8	10.3	5.2	10.3	6.2	10.7	96.4	98.4	100	100
200	200	3.9	8.8	4.6	9.5	3.5	8.8	5.2	9.5	5.1	8.5	6.4	9.6	97.8	98.6	100	100

Note: The main entries report the percentage of rejections over 1000 replications.

Table 3: Empirical rejection rates of $S_{NT}^{(1)}(\omega)$ with $M = \lfloor T^{1.2} \rfloor$

N	T	DGP 1		DGP 2		DGP 3		DGP 4		DGP 5		DGP 6		DGP 7		DGP 8	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
i.i.d. errors																	
100	100	5.7	10.9	5.9	11.7	4.5	11.4	6	10.2	5.0	10.0	5.7	11.0	98.4	99.3	100	100
200	100	5.1	9.0	5.8	10.8	6.6	10.8	4.2	8.1	5.1	11.0	4.4	9.5	98.3	99.0	100	100
100	200	5.5	10.6	4.7	9.5	4.7	9.7	4.3	9.0	4.2	9.7	5.5	10.6	99.9	100	100	100
200	200	5.1	10.2	4.1	8.3	4.5	9.0	4.8	10.0	5.0	10.0	4.7	8.8	99.8	100	100	100
heteroskedastic errors																	
100	100	5.7	11.5	5.4	10.8	4.9	10.1	6.4	12.2	5.9	12.0	5.5	9.7	97.6	98.6	100	100
200	100	5.5	11.9	4.9	11.6	6.2	10.8	5.1	11.6	5.9	11.6	5.9	10.5	97.5	99.1	100	100
100	200	4.6	10.5	5.9	11.4	5.6	10.3	4.9	9.9	5.0	11.3	5.6	10.0	99.7	99.9	100	100
200	200	5.9	11.4	5.5	11.2	6.7	12.7	4.5	8.6	5.1	9.4	5.9	11.2	100	100	100	100
cross-sectionally dependent errors																	
100	100	4.0	10.1	5.1	11.7	7.7	13.8	4.5	10.1	6.2	10.9	5.9	9.9	97.8	98.7	100	100
200	100	4.6	10.4	6.2	10.7	5.7	11.1	5.9	10.4	4.6	9.9	5.7	10.2	98.2	98.9	100	100
100	200	5.3	8.6	5.4	9.9	5.1	9.5	5.8	10.0	4.9	11.3	6.0	11.8	99.9	99.9	100	100
200	200	5.0	10.3	5.5	10.9	4.9	12.2	5.2	9.3	5.1	9.8	4.9	9.6	100	100	100	100
serially dependent errors																	
100	100	6.6	12.3	4.9	10.1	5.2	10.5	5.4	11.1	5.9	12.0	5.3	10.4	97.9	98.9	100	100
200	100	4.9	10.0	5.8	10.9	6.8	13.0	5.7	10.8	5.4	10.9	5.4	10.4	97.7	99.2	100	100
100	200	4.9	10.2	4.9	10.9	4.4	8.2	5.5	10.4	6.2	11.5	5.1	10.9	99.5	99.9	100	100
200	200	3.8	8.9	4.5	9.6	4.7	9.4	4.4	9.3	5.1	9.7	4.4	8.4	99.6	100	100	100
cross-sectionally and serially dependent errors																	
100	100	4.7	9.4	4.6	11	5.3	11.1	5.4	9.6	5.6	11.5	5.7	10.8	98.7	99.4	100	100
200	100	5.1	10.4	5.8	11.2	5.8	12.1	5.3	11.4	5.3	10.8	4.7	10.7	98.2	99.1	100	100
100	200	6.0	10.3	6.5	10.9	4.3	9.6	4.4	9.4	5.4	11.3	5.4	10.2	99.6	99.9	100	100
200	200	5.2	10.2	5.4	10.3	3.3	8.5	6.2	11.3	5.6	10.7	5.7	10.4	99.6	100	100	100

Note: The main entries report the percentage of rejections over 1000 replications.

Table 4: Empirical rejection rates of $S_{NT}^{(2)}(\omega)$ with $M = \lfloor T^{0.8} \rfloor$

N	T	DGP 1		DGP 2		DGP 3		DGP 4		DGP 5		DGP 6		DGP 7		DGP 8	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
i.i.d. errors																	
100	100	96.3	98.6	95.3	98.8	94.6	98.4	100	100	93.9	97.6	100	100	4.9	10.4	5.8	10.4
200	100	97.4	99.4	92.9	97.7	94.4	97.7	100	100	94.4	98.0	100	100	5.4	10.9	3.2	8.8
100	200	99.9	100	99.0	99.8	99.2	99.9	100	100	99.0	99.8	100	100	4.5	9.9	5.4	10.2
200	200	99.9	99.9	99.1	99.9	99.4	99.8	100	100	99.6	99.8	100	100	3.7	8.7	5.6	8.7
heteroskedastic errors																	
100	100	88.2	95.2	83.6	92.4	86.4	93.6	95.7	99	87.6	95.4	97.2	99.4	4.2	9.9	4.7	10.7
200	100	90.0	95.7	86.4	92.9	86.0	93.1	97.5	99.2	86.8	95.5	98.6	99.6	5.2	9.5	3.4	9.5
100	200	98.8	99.7	97.2	99.2	96.4	99.1	99.5	100	97.0	99.0	99.9	100	4.2	10.3	4.3	7.4
200	200	98.7	99.7	97.4	99.6	96.5	99.3	100	100	97.8	99.3	100	100	4.4	9.8	4.2	8.0
cross-sectionally dependent errors																	
100	100	96.2	98.6	91.7	97.1	93.4	95.9	99.9	100	93.0	97.4	100	100	4.3	10.3	5.8	12.0
200	100	96.4	98.9	94.7	97.7	94.0	98.1	100	100	95.7	98.4	100	100	4.2	8.2	4.5	8.9
100	200	99.6	99.9	98.8	99.6	99.2	99.9	100	100	99.4	99.8	100	100	4.5	9.9	4.8	9.5
200	200	99.9	100	99.6	99.9	99.3	99.8	100	100	99.7	99.9	100	100	3.8	8.3	5.6	10.3
serially dependent errors																	
100	100	99.1	99.8	97.5	98.9	97.0	98.7	100	100	98.0	99.8	100	100	4.9	10.1	4.1	9.7
200	100	99.2	99.8	97.9	99.6	97.9	99.2	100	100	98.2	99.8	100	100	6.2	10.3	4.7	10.1
100	200	100	100	99.4	100	100	100	100	100	99.6	100	100	100	3.9	8.6	6.0	8.9
200	200	100	100	100	100	99.8	100	100	100	99.9	100	100	100	3.4	8.5	6.2	9.2
cross-sectionally and serially dependent errors																	
100	100	98.9	99.6	96.4	98.9	95.9	98.6	99.9	100	97.5	99.1	100	100	5.8	11.0	4.9	9.1
200	100	99.0	99.8	97.7	99.5	97.6	99.3	100	100	98.3	99.6	100	100	5.7	11.1	5.4	11.4
100	200	100	100	99.7	100	99.8	100	100	100	100	100	100	100	4.5	9.4	5.8	9.2
200	200	99.9	100	100	100	99.9	100	100	100	99.9	100	100	100	5.2	10.5	4.5	8.4

Note: The main entries report the percentage of rejections over 1000 replications.

Table 5: Empirical rejection rates of $S_{NT}^{(2)}(\omega)$ with $M = T$

N	T	DGP 1		DGP 2		DGP 3		DGP 4		DGP 5		DGP 6		DGP 7		DGP 8	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
i.i.d. errors																	
100	100	99.9	100	99.9	100	100	100	100	100	99.9	99.9	100	100	5.6	12.4	5.8	10.2
200	100	100	100	99.9	100	100	100	100	100	100	100	100	100	5.4	12.0	5.9	9.5
100	200	100	100	100	100	100	100	100	100	100	100	100	100	4.1	10.1	6.2	12.3
200	200	100	100	100	100	100	100	100	100	100	100	100	100	5.3	10.4	5.6	12.5
heteroskedastic errors																	
100	100	99.8	100	99.2	99.9	99.3	99.9	100	100	99.7	99.9	100	100	5.9	12.9	6.5	9.3
200	100	100	100	99.7	100	99.8	100	100	100	99.8	99.9	100	100	5.9	11.7	4.5	8.1
100	200	100	100	100	100	100	100	100	100	100	100	100	100	5.4	9.0	6.4	10.8
200	200	100	100	100	100	100	100	100	100	100	100	100	100	4.7	9.1	5.4	9.8
cross-sectionally dependent errors																	
100	100	100	100	99.6	99.9	99.8	100	100	100	100	100	100	100	5.5	12.1	4.7	8.7
200	100	100	100	99.9	100	100	100	100	100	100	100	100	100	5.4	11.1	5.7	9.4
100	200	100	100	100	100	100	100	100	100	100	100	100	100	5.6	11.2	6.0	11.8
200	200	100	100	100	100	100	100	100	100	100	100	100	100	4.2	8.0	5.9	9.5
serially dependent errors																	
100	100	100	100	100	100	100	100	100	100	100	100	100	100	6.5	12.4	6.6	10.4
200	100	100	100	100	100	100	100	100	100	100	100	100	100	4.9	10.7	5.7	10
100	200	100	100	100	100	100	100	100	100	100	100	100	100	5.0	9.0	5.0	9.5
200	200	100	100	100	100	100	100	100	100	100	100	100	100	5.7	11.4	4.9	9.8
cross-sectionally and serially dependent errors																	
100	100	100	100	100	100	100	100	100	100	99.9	100	100	100	6.0	11.7	5.7	9.4
200	100	100	100	99.9	100	100	100	100	100	100	100	100	100	5.0	12.1	4.8	9.5
100	200	100	100	100	100	100	100	100	100	100	100	100	100	4.4	10.8	5.7	10.7
200	200	100	100	100	100	100	100	100	100	100	100	100	100	5.3	9.2	6.4	11.4

Note: The main entries report the percentage of rejections over 1000 replications.

Table 6: Empirical rejection rates of $S_{NT}^{(2)}(\omega)$ with $M = \lfloor T^{1.2} \rfloor$

N	T	DGP 1		DGP 2		DGP 3		DGP 4		DGP 5		DGP 6		DGP 7		DGP 8	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
i.i.d. errors																	
100	100	100	100	100	100	100	100	100	100	100	100	100	100	8.6	15.7	4.8	8.7
200	100	100	100	100	100	100	100	100	100	100	100	100	100	7.7	14.5	3.4	7.7
100	200	100	100	100	100	100	100	100	100	100	100	100	100	6.2	11.7	5.2	11.3
200	200	100	100	100	100	100	100	100	100	100	100	100	100	5.9	12.5	5.2	9.6
heteroskedastic errors																	
100	100	100	100	100	100	100	100	100	100	100	100	100	100	6.7	14.3	4.1	9.8
200	100	100	100	100	100	100	100	100	100	100	100	100	100	6.8	13.6	4.1	9.7
100	200	100	100	100	100	100	100	100	100	100	100	100	100	6.3	12.2	5.4	10.7
200	200	100	100	100	100	100	100	100	100	100	100	100	100	6.3	13.3	4.9	10.8
cross-sectionally dependent errors																	
100	100	100	100	100	100	100	100	100	100	100	100	100	100	8.4	16.6	4.6	9.9
200	100	100	100	100	100	100	100	100	100	100	100	100	100	7.1	14.2	5.9	11.1
100	200	100	100	100	100	100	100	100	100	100	100	100	100	5.8	10.6	4.3	9.5
200	200	100	100	100	100	100	100	100	100	100	100	100	100	6.0	11.9	6.1	11.9
serially dependent errors																	
100	100	100	100	100	100	100	100	100	100	100	100	100	100	9.3	17.4	4.1	9.9
200	100	100	100	100	100	100	100	100	100	100	100	100	100	7.0	14.8	4.5	9.7
100	200	100	100	100	100	100	100	100	100	100	100	100	100	6.4	12.8	4.8	9.1
200	200	100	100	100	100	100	100	100	100	100	100	100	100	5.4	12.9	4.4	10.2
cross-sectionally and serially dependent errors																	
100	100	100	100	100	100	100	100	100	100	100	100	100	100	7.2	13.8	5.6	11.3
200	100	100	100	100	100	100	100	100	100	100	100	100	100	8.8	17.6	5.4	10.2
100	200	100	100	100	100	100	100	100	100	100	100	100	100	6.5	13.0	4.3	9.2
200	200	100	100	100	100	100	100	100	100	100	100	100	100	6.4	12.2	4.6	8.9

Note: The main entries report the percentage of rejections over 1000 replications.

from the U.S. macroeconomic dataset. They specify the factor loadings as random walk processes and then study the transmission mechanism of the U.S. monetary policy. Eickmeier et al. (2015) extract common factors from a large balanced dataset containing 316 quarterly U.S. time series using the time-varying factor model and further identify monetary policy shocks and their transmission to the economy. They specify both the factor loadings and VAR coefficients as random walk processes. Nevertheless, although these studies find strong evidence of time-varying factor loadings, whether the time-varying factor loadings are random walk processes or deterministic functions of time remains an open question.

Now we use our tests to examine the specification of the time-varying factor loadings for the U.S. macroeconomic dataset. Specifically, we use the dataset first constructed by Stock and Watson (2012) and then extended by Cheng et al. (2016). The dataset consists of $N = 102$ time series of monthly macroeconomic and financial indicators spanning 1985:M1 to 2013:M1 with $T = 337$. To implement our tests, we follow the pre-settings in Section 5 to generate $\{\xi_m\}_{m=1}^{M}$ using independent standard normal distributions, adopt a binary distribution for $\Phi(\cdot)$, which has probability mass 1/2 at $\sqrt{2}$ and $-\sqrt{2}$, and set $M = \lfloor T^a \rfloor$ with $a = 0.8, 1, 1.2$, and 1.5, respectively.

Before applying our tests, we use the tests of Chen et al. (2014) and Han and Inoue (2015) to first test for structural changes in the factor loadings. Given that their tests require the estimated common factors, we set the selected number of common factors to be 4, which is determined using Bai and Ng's (2002) information criteria IC_{p1} and IC_{p2}. Both Chen et al.'s (2014) sup-LM and sup-Wald tests and Han and Inoue's (2015) sup-LM and sup-Wald tests reject the null hypothesis of no structural changes in factor loadings at the 5% significance level. As mentioned above, if the time-varying factor loadings evolve as stationary VAR processes, the tests of Chen et al. (2014) and Han and Inoue (2015) will not reject the null hypothesis. Hence, we find significant evidence that the time-varying factor loadings do not evolve as stationary stochastic processes. Now, we use our tests to investigate whether the time-varying factor loadings follow deterministic functions of time or random walk processes.

Table 7 reports the values of our test statistics and the corresponding P-values under different choices of M. It shows that the test using $S_{NT}^{(1)}(\omega)$ cannot reject the null hypothesis of deterministic-varying factor loadings. In contrast, the test using $S_{NT}^{(2)}(\omega)$ significantly rejects the null hypothesis of the unit-root type factor loadings. Hence, we find strong evidence that the time-varying factor loadings do not follow unit root processes. It implies that deterministic functions of time should be

Table 7: Empirical results for the U.S. macroeconomic dataset

	Testing $\mathbb{H}_1$ against $\mathbb{H}_2$		Testing $\mathbb{H}_2$ against $\mathbb{H}_1$	
	$S_{NT}^{(1)}(\omega)$	P-value	$S_{NT}^{(2)}(\omega)$	P-value
$a = 0.8$	0.0606	0.8055	56.5152	0.0000
$a = 1$	0.0526	0.8185	159.3474	0.0000
$a = 1.2$	0.0627	0.8022	442.9943	0.0000
$a = 1.5$	0.2850	0.5935	2200.5943	0.0000

Note: Entries under "$S_{NT}^{(1)}(\omega)$" and "$S_{NT}^{(2)}(\omega)$" are the values of the corresponding test statistics. Entries under under "P-value" are the corresponding asymptotic P-values based on $\chi^2(1)$ distribution.

adopted when modeling the time-varying features of the factor loadings in this application, which is consistent with Su and Wang's (2017) results. Furthermore, we note that the conclusion is robust to the choices of M. Although the P-values under $S_{NT}^{(1)}(\omega)$ decrease with increasing values of M, the conclusion remains unchanged.

6.2 Global macroeconomic and financial dataset

In this subsection, we apply our tests to examining the time-varying feature of the global macroeconomic and financial dataset used by Mumtaz and Musso (2019). They propose a dynamic factor model with time-varying parameters and use this model to extract global, region-specific, and country-specific uncertainty. The time-varying factor loadings have been directly specified as random walk processes in their model.

Now, we adopt our tests to investigate the proper specification of time-varying factor loadings for this dataset constructed by global macroeconomic and financial variables. The dataset consists of quarterly data, spanning the first quarter of 1960 to the fourth quarter of 2016 for 22 OECD countries. For each of the 22 countries, the dataset contains 20 variables, ranging from real economic activity, consumer prices, labor market variables, interest rates, credit market variables, money, international trade variables, and exchange rates. For the details of the data description, please see Mumtaz and Musso (2019).

Table 8 reports the values of test statistics and the corresponding P-values under different choices of M when applying our tests to the global macroeconomic and financial dataset. Similar to the results in Table 7, Table 8 shows that the specification under $\mathbb{H}_1$ is supported by both $S_{NT}^{(1)}(\omega)$ and $S_{NT}^{(2)}(\omega)$. Hence, we suggest using deterministic functions of time when modeling the time-varying

Table 8: Empirical results for the global macroeconomic and financial dataset

	Testing $\mathbb{H}_1$ against $\mathbb{H}_2$		Testing $\mathbb{H}_2$ against $\mathbb{H}_1$	
	$S_{NT}^{(1)}(\omega)$	P-value	$S_{NT}^{(2)}(\omega)$	P-value
$a = 0.8$	0.6579	0.4173	29.3947	0.0000
$a = 1$	1.0088	0.3152	108.5859	0.0000
$a = 1.2$	0.1446	0.7038	320.8957	0.0000
$a = 1.5$	2.0918	0.1481	1497.7760	0.0000

Note: See the notes in Table 7.

factor loadings in this application. Furthermore, we note that the conclusion is also robust to various choices of M.

7 Conclusion

Time-varying factor models have attracted great attention in analyzing large-dimensional macroeconomic and financial datasets. Most existing literature on time-varying factor models specifies the time-varying factor loadings as either deterministic functions of time or stochastic processes, mostly stationary VAR processes and unit root processes. Different specifications of time-varying factor loadings lead to distinct estimation procedures and economic implications. The existing literature has paid considerable attention to testing for structural changes in a factor model. The related works, such as Chen et al. (2014), Han and Inoue (2015), and Cheng et al. (2016), can distinguish the factor model with stationary VAR factor loadings from the model with deterministic-varying or unit-root type factor loadings. However, no formal tests exist to distinguish deterministic time-varying factor loadings from unit-root type factor loadings.

This paper fills in the gap in the literature by proposing two tests that can test the null hypothesis of either the deterministic time-varying factor loadings or unit-root type factor loadings against each other. Both proposed tests are based on the randomization approach, which is especially suitable for the current setting since consistent estimation for the unit-root type time-varying factor loadings is infeasible. As a result, we do not need to consistently estimate the number of common factors for the two non-nested models. Furthermore, the proposed test statistics are easy to compute and asymptotically pivotal. Monte Carlo studies demonstrate that both proposed tests have reasonable size and excellent power in distinguishing these two specifications of time-varying factor loadings. Our empirical studies suggest that specifying the factor loadings as deterministic functions of time is appropriate for both the U.S. macroeconomic dataset and Mumtaz and Musso's (2019) global macroeconomic and financial dataset.

Mathematical Appendix

In this appendix we prove the main results of the paper. Let C_0, C_1 and C_2 be generic constants that may vary over places. Let $\sum_{t,s=1}^{T} = \sum_{t=1}^{T}\sum_{s=1}^{T}$ and $\sum_{r,r_1,l,l_1=1}^{R} = \sum_{r=1}^{R}\sum_{r_1=1}^{R}\sum_{l=1}^{R}\sum_{l_1=1}^{R}$.

A Proof of Proposition 1

To prove Proposition 1, we need the following lemma.

Lemma 1. *Consider a multi-index partial sum process* $U_{S_1,\ldots,S_h} = \sum_{i_1=1}^{S_1}\cdots\sum_{i_h=1}^{S_h}\xi_{i_1\cdots i_h}$. *Assume that* $E\left|U_{S_1,\ldots,S_h}\right|^p \leq C_0\prod_{j=1}^{h} S_j^{d_j}$ *where* $p \geq 1$ *and* $d_j \geq 1$ *for all* $1 \leq j \leq h$. *Then*

$$\lim \sup_{\min(S_1,\ldots,S_h)\to\infty} \frac{U_{S_1,\ldots,S_h}}{\prod_{j=1}^{h} S_j^{d_j/p}\left(\log S_j\right)^{1+\frac{1}{p}+\epsilon}} = 0 \ a.s.$$

for all $\epsilon > 0$.

For a proof of the above lemma, see Lemma A.1 in Massacci and Trapani (2021).

Given that $\hat{\phi}_{jNT}$ is the jth largest eigenvalue of the $T \times T$ matrix $XX'/(NT)$,

$$\begin{aligned}
\hat{D} \equiv \sum_{j=1}^{T}\hat{\phi}_{jNT} &= \operatorname{tr}\left(\frac{XX'}{NT}\right) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(F_t'\lambda_{it}+\varepsilon_{it})^2 \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}F_t'\lambda_{it}\lambda_{it}'F_t + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}^2 + \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}F_t'\lambda_{it}\varepsilon_{it} \\
&\equiv \hat{D}_1 + \hat{D}_2 + 2\hat{D}_3.
\end{aligned}$$

We first prove (i) by showing that under $\mathbb{H}_1$, we have: (i.a) $\hat{D}_1 \overset{a.s.}{\to} \operatorname{tr}[\Sigma_\Lambda\Sigma_F]$, (i.b) $\hat{D}_2 \overset{a.s.}{\to} \sigma_\varepsilon^2$, and (i.c) $\hat{D}_3 \overset{a.s.}{\to} 0$. We first show (i.a). Recall that $\Sigma_{\Lambda_t} = N^{-1}\Lambda_t'\Lambda_t$. Let e_r denote the rth column of I_R. Then

$$\begin{aligned}
\hat{D}_1 - E(\hat{D}_1) &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\lambda_{it}'\left(F_tF_t' - \Sigma_F\right)\lambda_{it} = \frac{1}{T}\sum_{t=1}^{T}\operatorname{tr}\left[\Sigma_{\Lambda_t}\left(F_tF_t' - \Sigma_F\right)\right] \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{r=1}^{R}e_r'\Sigma_{\Lambda_t}\left(F_tF_t' - \Sigma_F\right)e_r = \frac{1}{T}\sum_{t=1}^{T}\xi_{1t},
\end{aligned}$$

where $\xi_{1t} \equiv \sum_{r=1}^{R} e_r' \Sigma_{\Lambda_t} \left(F_t F_t' - \Sigma_F\right) e_r$. By the uniform boundedness of Σ_{Λ_t} under Assumption A.3(i), we can readily apply Assumption A.1(iii) and show that

$$E\left|\sum_{t=1}^{T} \xi_{1t}\right|^2 = \sum_{t=1}^{T}\sum_{s=1}^{T} E\left(\xi_{1t}\xi_{1s}\right) \leq C_0 \sum_{t,s=1}^{T} \sum_{r,r_1,l,l_1=1}^{R} \left|E\left\{f_{rl,t} f_{r_1 l_1,s}\right\}\right| \leq C_1 T,$$

where $f_{rl,t} = F_{rt}F_{lt} - E\left(F_{rt}F_{lt}\right)$. Then by Lemma 1,

$$\frac{1}{T}\sum_{t=1}^{T} \xi_{1t} = o_{a.s.}\left(T^{-1/2}(\log T)^{\frac{3}{2}+\epsilon}\right)$$

for all $\epsilon > 0$. In addition, $E(\hat{D}_1) = \frac{1}{T}\sum_{t=1}^{T}\text{tr}[\Sigma_{\Lambda_t}\Sigma_F] \to \text{tr}[\Sigma_\Lambda \Sigma_F] \geq 0$ by Assumptions A.1(i) and A.3(ii). It follows that $\hat{D}_1 = \text{tr}[\Sigma_\Lambda\Sigma_F] + o_{a.s.}(1)$. Next, we show (i.b). Note that

$$\hat{D}_2 - E(\hat{D}_2) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\varepsilon_{it}^2 - E\left(\varepsilon_{it}^2\right)\right] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi_{2,it},$$

where $\xi_{2,it} = \varepsilon_{it}^2 - E\left(\varepsilon_{it}^2\right)$. Under Assumption A.2(iii), we have $E\left|\sum_{i=1}^{N}\sum_{t=1}^{T}\xi_{2,it}\right|^2 \leq CNT$. Then by Lemma 1,

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\xi_{2,it} = o_{a.s.}\left((NT)^{-1/2}(\log N \log T)^{\frac{3}{2}+\epsilon}\right)$$

for all $\epsilon > 0$. This, along with the fact that $E(\hat{D}_2) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} E\left(\varepsilon_{it}^2\right) \to \sigma_\varepsilon^2 > 0$ under Assumption A.2(ii), implies that $\hat{D}_2 = \sigma_\varepsilon^2 + o_{a.s.}(1)$. Similarly, noting that $E\left\|\sum_{i=1}^{N}\sum_{t=1}^{T} F_t'\lambda_{it}\varepsilon_{it}\right\|^2 \leq CNT$ under Assumption A.3(iii), we can apply Lemma 1 to obtain

$$\hat{D}_3 = o_{a.s.}\left((NT)^{-1/2}(\log N \log T)^{\frac{3}{2}+\epsilon}\right)$$

for all $\epsilon > 0$. Then (i.c) holds. In sum, we have

$$\hat{D}_1 = D_1 + \sigma_\varepsilon^2 + o_{a.s.}(1).$$

Now, we show (ii) by showing that under $\mathbb{H}_2$, we have: (ii.a) $T^{-1}\hat{D}_1 \overset{a.s.}{\to} \frac{1}{2}\text{tr}(\Sigma\Sigma_F)$, (ii.b) $T^{-1}\hat{D}_2 \overset{a.s.}{\to} 0$, and (ii.c) $T^{-1}\hat{D}_3 \overset{a.s.}{\to} 0$. Noting that

$$\begin{aligned}
\lambda_{it}\lambda_{it}' &= \left[\lambda_{it} - \Sigma_i^{1/2}W_i(t) + \Sigma_i^{1/2}W_i(t)\right]\left[\lambda_{it} - \Sigma_i^{1/2}W_i(t) + \Sigma_i^{1/2}W_i(t)\right]' \\
&= \Sigma_i^{1/2}W_i(t)W_i(t)'\Sigma_i^{1/2} + \Sigma_i^{1/2}W_i(t)\left[\lambda_{it} - \Sigma_i^{1/2}W_i(t)\right]' \\
&\quad + \left[\lambda_{it} - \Sigma_i^{1/2}W_i(t)\right]W_i(t)'\Sigma_i^{1/2} + \left[\lambda_{it} - \Sigma_i^{1/2}W_i(t)\right]\left[\lambda_{it} - \Sigma_i^{1/2}W_i(t)\right]' \\
&\equiv d_{it,1} + d_{it,2} + d_{it,3} + d_{it,4},
\end{aligned}$$

we have

$$T^{-1}\hat{D}_1=\frac{1}{NT^2}\sum_{t=1}^{N}\sum_{t=1}^{T}F_t'\lambda_{it}\lambda_{it}'F_t=\sum_{\ell=1}^{4}\frac{1}{NT^2}\sum_{t=1}^{N}\sum_{t=1}^{T}F_t'd_{it,1\ell}F_t\equiv\sum_{\ell=1}^{4}\hat{D}_{1,\ell}.$$

Let $D_{1,1N}(t)=\frac{1}{N}\sum_{i=1}^{N}\Sigma_i^{1/2}W_i(t)W_i(t)'\Sigma_i^{1/2}$. Then

$$\hat{D}_{1,1}=\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{tr}\left(F_t'\Sigma_i^{1/2}W_i(t)W_i(t)'\Sigma_i^{1/2}F_t\right)=\frac{1}{T^2}\sum_{t=1}^{T}\mathrm{tr}\left[D_{1,1N}(t)F_tF_t'\right].$$

By the independence between $\{v_{is}\}$ and $\{F_t\}$ under Assumption A.4(ii) and the fact that $E\left[W_i(t)W_i(t)'\right]=tI_R$, we have

$$\begin{aligned}E(\hat{D}_{1,1}) &= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{tr}\left(\Sigma_i^{1/2}E\left[W_i(t)W_i(t)'\right]\Sigma_i^{1/2}\Sigma_F\right)\\ &= \frac{1}{N}\sum_{i=1}^{N}\mathrm{tr}\left(\Sigma_i\Sigma_F\right)\frac{1}{T^2}\sum_{t=1}^{T}t\to\frac{1}{2}\mathrm{tr}\left(\Sigma\Sigma_F\right)>0.\end{aligned}$$

In addition,

$$\begin{aligned}\hat{D}_{1,1}-E(\hat{D}_{1,1}) &= \frac{1}{T^2}\sum_{t=1}^{T}\mathrm{tr}\left\{D_{1,1N}(t)F_tF_t'-E\left[D_{1,1N}(t)F_tF_t'\right]\right\}\\ &= \frac{1}{T^2}\sum_{t=1}^{T}\mathrm{tr}\left\{\left[D_{1,1N}(t)-E\left(D_{1,1N}(t)\right)\right]\left(F_tF_t'-\Sigma_F\right)\right\}\\ &\quad+\frac{1}{T^2}\sum_{t=1}^{T}\mathrm{tr}\left\{E\left(D_{1,1N}(t)\right)\left(F_tF_t'-\Sigma_F\right)\right\}\\ &\quad+\frac{1}{T^2}\sum_{t=1}^{T}\mathrm{tr}\left\{\left[D_{1,1N}(t)-E\left(D_{1,1N}(t)\right)\right]\Sigma_F\right\}\\ &= \frac{1}{T^2}\sum_{t=1}^{T}\xi_{3Nt}+\frac{1}{T^2}\sum_{t=1}^{T}\xi_{4Nt}+\frac{1}{T^2}\sum_{t=1}^{T}\xi_{5Nt},\end{aligned}$$

where $\xi_{3Nt}=\mathrm{tr}\{[D_{1,1N}(t)-E(D_{1,1N}(t))](F_tF_t'-\Sigma_F)\}$, $\xi_{4Nt}=\mathrm{tr}\{E(D_{1,1N}(t))(F_tF_t'-\Sigma_F)\}$, and $\xi_{5Nt}=\mathrm{tr}\{[D_{1,1N}(t)-E(D_{1,1N}(t))]\Sigma_F\}$. Note that

$$\begin{aligned}E\left|\sum_{t=1}^{T}\xi_{3Nt}\right|^2 &= \sum_{t,s=1}^{T}E\left(\xi_{3Nt}\xi_{3Ns}\right)\\ &= \sum_{t,s=1}^{T}\sum_{r,l=1}^{R}E\{e_r'\left[D_{1,1N}(t)-E\left(D_{1,1N}(t)\right)\right]\left(F_tF_t'-\Sigma_F\right)e_re_l'\left(F_sF_s'-\Sigma_F\right)\\ &\quad\times\left[D_{1,1N}(s)-E\left(D_{1,1N}(s)\right)\right]e_l\}\end{aligned}$$

$$
\begin{aligned}
&= \sum_{t,s=1}^{T}\sum_{r,l=1}^{R} \operatorname{tr}\{E\left[\left(F_tF_t'-\Sigma_F\right)e_re_l'\left(F_sF_s'-\Sigma_F\right)\right] \\
&\quad \times E\left\{\left[D_{1,1N}(s)-E\left(D_{1,1N}(s)\right)\right]e_le_r'\left[D_{1,1N}(t)-E\left(D_{1,1N}(t)\right)\right]\right\}\} \\
&\leq C_1\sum_{t,s=1}^{T}\sum_{r,l=1}^{R} s\cdot t\left\|E\left[\left(F_tF_t'-\Sigma_F\right)e_re_l'\left(F_sF_s'-\Sigma_F\right)\right]\right\| \\
&\leq C_1\sum_{s=1}^{T}s^2\sum_{r,l=1}^{R}\sum_{t=1}^{T}\left\|E\left[\left(F_tF_t'-\Sigma_F\right)e_re_l'\left(F_sF_s'-\Sigma_F\right)\right]\right\| \\
&\leq C_2T^3,
\end{aligned}
$$

where the second inequality holds by the Cauchy-Schwarz inequality (viz., $st \leq (s^2+t^2)/2$) and the last inequality holds by Assumption A.1(iii). In addition, the first inequality in the above derivation holds because

$$
\begin{aligned}
&\left\|E\left\{\left[D_{1,1N}(s)-E\left(D_{1,1N}(s)\right)\right]e_le_r'\left[D_{1,1N}(t)-E\left(D_{1,1N}(t)\right)\right]\right\}\right\|^2 \\
\leq\ & E\left\|\frac{1}{N}\sum_{i=1}^{N}\Sigma_i^{1/2}\left[W_i(s)W_i(s)'-sI_R\right]\Sigma_i^{1/2}e_l\right\|^2 E\left\|e_r'\frac{1}{N}\sum_{i=1}^{N}\Sigma_i^{1/2}\left[W_i(t)W_i(t)'-tI_R\right]\Sigma_i^{1/2}\right\|^2 \\
\leq\ & \frac{1}{N}\sum_{i=1}^{N}\left\|\Sigma_i\right\|^2E\left\|W_i(s)W_i(s)'-sI_R\right\|^2\frac{1}{N}\sum_{i=1}^{N}\left\|\Sigma_i\right\|^2E\left\|W_i(t)W_i(t)'-tI_R\right\|^2 \\
\leq\ & \left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\Sigma_i\right\|^2\right\}^2\max_i E\left\|W_i(s)W_i(s)'-sI_R\right\|^2\max_i E\left\|W_i(t)W_i(t)'-tI_R\right\|^2 \\
\leq\ & C_1s^2t^2,
\end{aligned}
$$

where we use the fact that $E\left\|W_i(s)\right\|^4 = E\left(\Sigma_{r=1}^{R}W_{i,r}^2(s)\right)^2 \leq R\Sigma_{r=1}^{R}E[W_{i,r}^4(s)] = 3R^2s^2$ with $W_{i,r}(s)$ denoting the r-th element of $W_i(s)$. Note that we allow for full, strong, or weak dependence of $W_i(s)$ over i. Then by Lemma 1, $\frac{1}{T^2}\sum_{t=1}^{T}\xi_{3Nt} = o_{a.s.}\left(T^{-1/2}(\log T)^{3/2+\epsilon}\right)$ for any $\epsilon > 0$. Analogously, we can show that $\frac{1}{T^2}\sum_{t=1}^{T}\xi_{kNt} = o_{a.s.}(1)$ for $k = 4, 5$. It follows that $\hat{D}_{1,1} = \frac{1}{2}\operatorname{tr}(\Sigma\Sigma_F) + o_{a.s.}(1)$. Next,

$$
\begin{aligned}
\hat{D}_{1,4} &= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}F_t'\left[\lambda_{it}-\Sigma_i^{1/2}W_i(t)\right]\left[\lambda_{it}-\Sigma_i^{1/2}W_i(t)\right]'F_t \\
&\leq \max_t\frac{1}{NT}\sum_{i=1}^{N}\left\|\lambda_{it}-\Sigma_i^{1/2}W_i(t)\right\|^2\frac{1}{T}\sum_{t=1}^{T}\left\|F_t\right\|^2 \\
&= o_{a.s.}\left(T^{-2\epsilon_0}\right)O_{a.s.}(1) = o_{a.s.}(1),
\end{aligned}
$$

where we use Assumption A.4(iii) and the fact that $\frac{1}{T}\sum_{t=1}^{T}F_tF_t' = \Sigma_F + o_{a.s.}(1)$ under Assumption

A.1 by a simple application of Lemma 1. Now, by the Cauchy-Schwarz inequality,

$$\left|\hat{D}_{1,2}\right| = \left|\hat{D}_{1,3}\right| \leq \hat{D}_{1,1}^{1/2}\hat{D}_{1,4}^{1/2} = O_{a.s.}(1)\, o_{a.s.}(1) = o_{a.s.}(1).$$

In sum, we have shown that $T^{-1}\hat{D}_1 = \frac{1}{2}\mathrm{tr}(\Sigma\Sigma_F) + o_{a.s.}(1)$.

Next, by the result in (i.b), $T^{-1}\hat{D}_2 = O_{a.s.}\left(T^{-1}\right)$ and thus (ii.b) follows. For (ii.c), we have by the Cauchy-Schwarz inequality that

$$T^{-1}\left|\hat{D}_3\right| \leq \left\{T^{-1}\hat{D}_1\right\}^{1/2}\left\{T^{-1}\hat{D}_2\right\}^{1/2} = O_{a.s.}(1)\, O_{a.s.}\left(T^{-1/2}\right) = o_{a.s.}(1).$$

In sum, we have $T^{-1}\hat{D} = \frac{1}{2}\mathrm{tr}(\Sigma\Sigma_F) + o_{a.s.}(1)$ under $\mathbb{H}_2$.

B Proof of Theorem 1

When testing $\mathbb{H}_1$ against $\mathbb{H}_2$, we define $\mathcal{Y}_{NT} = T\hat{D}^{-1}$. Given Proposition 1, we have $\mathcal{Y}_{NT} = \bar{O}_{a.s.}(T)$ under $\mathbb{H}_1$ and $\mathcal{Y}_{NT} = \bar{O}_{a.s.}(1)$ under $\mathbb{H}_2$, where $\bar{O}_{a.s.}(\cdot)$ denote the exact order. Without loss of generality, we assume that $\{\xi_m\}_{m=1}^M \sim i.i.d.N(0,1)$ with $G(0) = 1/2$.

(i) Note that conditioning on the sample ω associated with $\hat{D}(\omega)$,

$$V_{m,NT}(\omega) \sim N(0, \mathcal{Y}_{NT}^2(\omega)),$$

for each $m = 1, \ldots, M$. Let $\Omega_1 \equiv \{\omega : \mathcal{Y}_{NT}(\omega) \to \infty\}$. With the almost sure convergence by Proposition 1, we have $P(\Omega_1) = 1$ under Assumptions A.1–A.3. Fix some $u > 0$.

$$\begin{aligned}
P^*(V_{m,NT}(\omega) \leq u) &= P^*(V_{m,NT}(\omega) \leq 0) + P^*(0 < V_{m,NT}(\omega) \leq u) \\
&= \frac{1}{2} + \frac{1}{\sqrt{2\pi}\mathcal{Y}_{NT}(\omega)}\int_0^u e^{-\frac{x^2}{2\mathcal{Y}_{NT}^2(\omega)}}\mathrm{d}x \\
&= \frac{1}{2} + O_{a.s.}(T^{-1}).
\end{aligned}$$

Similar arguments apply to the case when $u < 0$, i.e.,

$$\begin{aligned}
P^*(V_{m,NT}(\omega) \leq u) &= P^*(V_{m,NT}(\omega) \leq 0) - P^*(u < V_{m,NT}(\omega) \leq 0) \\
&= \frac{1}{2} - \frac{1}{\sqrt{2\pi}\mathcal{Y}_{NT}(\omega)}\int_u^0 e^{-\frac{x^2}{2\mathcal{Y}_{NT}^2(\omega)}}\mathrm{d}x \\
&= \frac{1}{2} - O_{a.s.}(T^{-1}).
\end{aligned}$$

Conditioning on the sample, we have

$$E^*\left[\mathbb{I}(V_{m,NT}(\omega) \leq u)\right] = P^*(V_{m,NT}(\omega) \leq u) = \frac{1}{2} + O_{a.s.}(T^{-1}).$$

We then decompose

$$\begin{aligned} Z_{NT}(u,\omega) &= Z_{NT}^0(u,\omega) + 2\sqrt{M}\left\{E^*\left[\mathbb{I}(V_{m,NT}(\omega) \le u)\right] - \frac{1}{2}\right\} \\ &= Z_{NT}^0(u,\omega) + O_{a.s}(M^{1/2}T^{-1}). \end{aligned}$$

where $Z_{NT}^0(u,\omega) = \frac{2}{\sqrt{M}}\sum_{m=1}^M \left\{\mathbb{I}(V_{m,NT}(\omega) \le u) - E^*\left[\mathbb{I}(V_{m,NT}(\omega) \le u)\right]\right\}$. It follows that $E^*[Z_{NT}(u,\omega)]$ $= o_{a.s.}(1)$ since $M^{1/2}T^{-1} \to 0$ as $(M,T) \to \infty$. Notice that $\mathbb{I}(V_{m,NT}(\omega) \le u)$ is a binary variable for any u. We have

$$E^*\left[\mathbb{I}(V_{m,NT}(\omega) \le u)\right]^2 = E^*[\mathbb{I}(V_{m,NT}(\omega) \le u)] = \frac{1}{2} + O_{a.s.}(T^{-1}),$$

and

$$\begin{aligned} \text{Var}^*\left[\mathbb{I}(V_{m,NT}(\omega) \le u)\right] &= E^*[\mathbb{I}(V_{m,NT}(\omega) \le u)]\left\{1 - E^*[\mathbb{I}(V_{m,NT}(\omega) \le u)]\right\} \\ &= \left[\frac{1}{2} + O_{a.s.}(T^{-1})\right]\left[\frac{1}{2} - O_{a.s.}(T^{-1})\right] \\ &= \frac{1}{4} + O_{a.s.}(T^{-1}). \end{aligned}$$

It follows that

$$\begin{aligned} \text{Var}^*\left[\frac{2}{\sqrt{M}}\sum_{m=1}^M \left[\mathbb{I}(V_{m,NT}(\omega) \le u)\right]\right] &= \frac{4}{M}\sum_{m=1}^M \text{Var}^*\left[\mathbb{I}(V_{m,NT}(\omega) \le u)\right] \\ &= 4\left[\frac{1}{4} + O_{a.s.}(T^{-1})\right] = 1 + O_{a.s.}(T^{-1}). \end{aligned}$$

Given that $\mathbb{I}(V_{m,NT}(\omega) \le u)$, $m = 1, \ldots, M$, is an independent and identically distributed (i.i.d.) triangular array, using the CLT for i.i.d. triangular array, we have

$$Z_{NT}(u,\omega) \xrightarrow{d^*} N(0,1),$$

as $M \to \infty$ for each fixed u.

Now, we show the convergence holds uniformly in $\mathbb{U}$. Consider $u_1 \in \mathbb{U}$ and $u_2 \in \mathbb{U}$ with $u_1 < u_2$. Then

$$\begin{aligned} E^*\left(\left|Z_{NT}^0(u_2,\omega) - Z_{NT}^0(u_1,\omega)\right|^2\right) &= \text{Var}^*\left(Z_{NT}^0(u_2,\omega) - Z_{NT}^0(u_1,\omega)\right) \\ &= \text{Var}^*\left(\left|\frac{2}{\sqrt{M}}\sum_{m=1}^M \mathbb{I}(u_1 < V_{m,NT}(\omega) \le u_2)\right|\right) \\ &= \frac{4}{M}\sum_{m=1}^M \text{Var}^*\left[\mathbb{I}(u_1 < V_{m,NT}(\omega) \le u_2)\right] \end{aligned}$$

$$
\begin{aligned}
&\leq 4E^*\left[\mathbb{I}(u_1 < V_{m,NT}(\omega) \leq u_2)\right] \\
&= 4\left[P^*(V_{m,NT}(\omega) \leq u_2) - P^*(V_{m,NT}(\omega) \leq u_1)\right] \\
&= O_{a.s.}(T^{-1}).
\end{aligned}
$$

Similarly, we can show that

$$
\begin{aligned}
E^*\left[Z^0_{NT}(u_1,\omega)Z^0_{NT}(u_2,\omega)\right] &= \frac{4}{M}\sum_{m=1}^{M} \mathrm{Cov}^*\left[\mathbb{I}(V_{m,NT}(\omega) \leq u_1), \mathbb{I}(V_{m,NT}(\omega) \leq u_2)\right] \\
&= 4\{E^*\left[\mathbb{I}(V_{m,NT}(\omega) \leq u_1)\mathbb{I}(V_{m,NT}(\omega) \leq u_2)\right] \\
&\quad -E^*\left[\mathbb{I}(V_{m,NT}(\omega) \leq u_1)\right]E^*\left[\mathbb{I}(V_{m,NT}(\omega) \leq u_2)\right]\} \\
&\leq 4\left\{\left[\frac{1}{2} + O_{a.s.}(T^{-1})\right] - \left[\frac{1}{2} + O_{a.s.}(T^{-1})\right]\left[\frac{1}{2} + O_{a.s.}(T^{-1})\right]\right\} \\
&= 1 + O_{a.s.}(T^{-1}).
\end{aligned}
$$

Then by the continuous mapping theorem, we have

$$
S_{NT}(\omega) = \int_{\mathbb{U}} \left|Z^0_{NT}(u,\omega)\right|^2 \mathrm{d}\Phi(u) + o_{a.s.}(1) \overset{d^*}{\to} \chi^2_1,
$$

conditional on the sample path $\omega \in \Omega_1$.

(ii) Under $\mathbb{H}_2$, Proposition 1(ii) implies that $\mathcal{Y}_{NT} = T\hat{D}^{-1} \overset{a.s.}{\to} D_2^{-1}$. Then $P(\Omega_2) = 1$, where $\Omega_2 \equiv \{\omega : \mathcal{Y}_{NT}(\omega) \to D_2^{-1}\}$. Conditioning on the sample path ω,

$$
V_{m,NT}(\omega) \overset{d^*}{\to} N(0, D_2^{-2})
$$

as $(N,T) \to \infty$ for each m. Let $F(u)$ be the CDF of a $N(0, D_2^{-2})$ random variable and $F_{NT}(u,\omega) \equiv P^*(V_{m,NT}(\omega) \leq u)$ be the CDF of $V_{m,NT}(\omega)$ conditional on the sample path ω. Then

$$
\begin{aligned}
Z_{NT}(u,\omega) &= \frac{2}{\sqrt{M}}\sum_{m=1}^{M}\left[\mathbb{I}(V_{m,NT}(\omega) \leq u) - F_{NT}(u,\omega)\right] + 2\sqrt{M}\left[F_{NT}(u,\omega) - F(u)\right] \\
&\quad +2\sqrt{M}\left[F(u) - \frac{1}{2}\right] \\
&\equiv A_1(u,\omega) + A_2(u,\omega) + A_3(u,\omega), \text{ say.}
\end{aligned}
$$

Conditioning on the sample path ω, for each u, it is straightforward to show that $E^*[A_1(u,\omega)] = 0$ given that $E^*\left[\mathbb{I}(V_{m,NT}(\omega) \leq u)\right] = F_{NT}(u,\omega)$, and

$$
E^*\left[A_1^2(u,\omega)\right] = \frac{4}{M}\sum_{m=1}^{M} \mathrm{Var}^*\left[\mathbb{I}(V_{m,NT}(\omega) \leq u)\right] = 4F_{NT}(u,\omega)[1 - F_{NT}(u,\omega)].
$$

Using similar arguments, we can show that $A_1(u,\omega)$ is asymptotically tight on $\mathbb{U}$. Thus, $\sup_{u\in\mathbb{U}}|A_1(u,\omega)| = O_{p^*}(1)$ for each $\omega \in \Omega_2$. Next, consider $A_2(u,\omega)$. Under Proposition 1(ii), $T^{-1}\hat{D} - D_2 = o_{a.s.}(1)$. By the Berry-Esséen Theorem (see, e.g., Theorem 25.7 in Davidson (1994)),

$$\sup_{u\in\mathbb{U}} |F_{NT}(u,\omega) - F(u)| = O(M^{-1/2}) \text{ a.s.-}\omega.$$

Thus, $\sup_{u\in\mathbb{U}}|A_2(u,\omega)| = O_{a.s}(1)$. Furthermore, since $F(0) = 1/2$, $A_3(u,\omega) \neq 0$ for $u \neq 0$. Then, it follows that $Z_{NT}(u,\omega)$ diverges to infinity at the rate $\sqrt{M}$ for all $\omega \in \Omega_2$ with $P(\Omega_2) = 1$. Then $P^*\left[S_{NT}^{(1)}(\omega) > c_M\right] \to 1$ a.s.-$\omega \in \Omega_2$ as $T \to \infty$ or $(N,T) \to \infty$ for any $c_M = o(M)$.

C Proof of Theorem 2

When testing $\mathbb{H}_2$ against $\mathbb{H}_1$, we define $\mathcal{Y}_{NT} = \hat{D}$. Given Proposition 1, we have $\mathcal{Y}_{NT} = \bar{O}_{a.s.}(T)$ under $\mathbb{H}_2$ and $\mathcal{Y}_{NT} = \bar{O}_{a.s.}(1)$ under $\mathbb{H}_1$. Without loss of generality, we still assume that $\{\xi_m\}_{m=1}^M \sim i.i.d.N(0,1)$. Note that we have the similar orders of magnitude of $\mathcal{Y}_{NT}$ under the null and alternative hypotheses as in Theorem 1. Hence, following analogous arguments as used in the proof of Theorem 1, we can establish the desired results.

References

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

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

Bandi, F. M., & Corradi, V. (2014). Nonparametric nonstationarity tests. *Econometric Theory* 30(1), 127-149.

Banerjee, A., Marcellino, M., & Masten, I. (2008). Forecasting macroeconomic variables using diffusion indexes in short samples with structural change. In Forecasting in the presence of structural breaks and model uncertainty. Emerald Group Publishing Limited.

Barigozzi, M., & Trapani, L. (2020). Sequential testing for structural stability in approximate factor models. *Stochastic Processes and Their Applications* 130(8), 5149-5187.

Barigozzi, M., & Trapani, L. (2022). Testing for common trends in non-stationary large datasets. *Journal of Business & Economic Statistics* 40(3), 1107-1122.

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

Baumeister, C., Liu, P., & Mumtaz, H. (2013). Changes in the effects of monetary policy on disaggregate price dynamics. *Journal of Economic Dynamics and Control*, 37(3), 543-560.

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

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

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

Cogley, T., & Sargent, T. J. (2001). Evolving post-world war II US inflation dynamics. *NBER Macroeconomics Annual* 16, 331-373.

Cooley, T. F., & Prescott, E. C. (1976). Estimation in the presence of stochastic parameter variation. *Econometrica* 44(1), 167-184.

Corradi, V., & Swanson, N. R. (2006). The effect of data transformation on common cycle, cointegration, and unit root tests: monte carlo and a simple test, *Journal of Econometrics* 132, 195-229.

Davidson, J. (1994). Stochastic Limit Theory, Oxford University Press, Oxford.

Del Negro, M., & Otrok, C. (2008). Dynamic factor models with time-varying parameters: measuring changes in international business cycles. *FRB of New York Staff Report*, (326).

Eberlein, E. (1986). On strong invariance principles under dependence assumptions. *Annals of Probability* 14(1), 260-270.

Eickmeier, S., Lemke, W., & Marcellino, M. (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., Su, L., & Wang, X. (2022a). Estimation and inference on time-varying FAVAR models. Working paper.

Fu, Z., Hong, Y., & Wang, X. (2022b). Testing for structural changes in large dimensional factor models via discrete Fourier transform. *Journal of Econometrics*, forthcoming https://doi.org/10.1016/j.jecon

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

Korobilis, D. (2013). Assessing the transmission of monetary policy using time-varying parameter dynamic factor models. *Oxford Bulletin of Economics and Statistics* 75(2), 157-179.

Liu, W. & Lin, Z. (2009). Strong approximation for a class of stationary processes. *Stochastic Processes and Their Applications* 119(1), 249-280.

Massacci, D., & Trapani, L. (2021). High dimensional threshold regression with common stochastic trends, Working Paper, School of Economics, University of Nottingham.

Ma, S., Lan, W., Su, L., & Tsai, C. L. (2020). Testing alphas in conditional time-varying factor models with high-dimensional assets. *Journal of Business & Economic Statistics* 38(1), 214-227.

Mikkelsen, J. G., Hillebrand, E., & Urga, G. (2019). Consistent estimation of time-varying loadings in high-dimensional factor models. *Journal of Econometrics* 208, 535-562.

Mumtaz, H., & Musso, A. (2021). The evolving impact of global, region-specific, and country-specific uncertainty. *Journal of Business & Economic Statistics* 39(2), 466-481.

Pearson. E. S. (1950). On questions raised by the combination of tests Based on discontinuous distributions. *Biometrica* 37, 383-398.

Phillips, P. C., & Solo, V. (1992). Asymptotics for linear processes. *Annals of Statistics* 20(2), 971-1001.

Rosenberg, B. (1973). Random coefficients models: the analysis of a cross section of time series by stochastically convergent parameter regression. *Annals of Economic and Social Measurement* 2(4), 399-428.

Stock, J. H., & Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics* 14(1), 11-30.

Stock, J. H., & Watson, M. W. (1998). Median unbiased estimation of coefficient variance in a time-varying parameter model. *Journal of the American Statistical Association* 93(441), 349-358.

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

Stock, J.H., & Watson, M.W. (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*, 1-57, Oxford University Press, Oxford.

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

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

Su, L. & Wang, X. (2020). Testing for structural changes in factor models via a nonparametric regression. *Econometric Theory* 36, 1127-1158.

Trapani, L. (2018). A randomized sequential procedure to determine the number of factors. *Journal of the American Statistical Association* 113(523), 1341-1349.