On time-varying factor models: Estimation and testing

Liangjun Su a,∗, Xia Wang b

a School of Economics, Singapore Management University, Singapore
b Lingnan (University) College, Sun Yat-Sen University, Guangzhou, PR China

A R T I C L E   I N F O

Article history:
Received 14 May 2015
Received in revised form 28 April 2016
Accepted 2 December 2016
Available online 20 February 2017

JEL classification:
C12
C14
C33
C38

Keywords:
Factor model
Information criterion
Local principal component
Local smoothing
Structural change
Test
Time-varying parameter

A B S T R A C T

Conventionally, factor models assume that factor loadings are fixed over a long horizon of time, which appears overly restrictive and unrealistic in applications. In this paper, we introduce a time-varying factor model where factor loadings are allowed to change smoothly over time. We propose a local version of the principal component method to estimate the latent factors and time-varying factor loadings simultaneously. We establish the limiting distributions and uniform convergence of the estimated factors and factor loadings in the standard large-N and large-T framework. We also propose a BIC-type information criterion to determine the number of factors, which can be used in models with either time-varying or time-invariant factor models. Based on the comparison between the estimates of the common components under the null hypothesis of no structural changes and those under the alternative, we propose a consistent test for structural changes in factor loadings. We establish the null distribution, the asymptotic local power property, and the consistency of our test. Simulations are conducted to evaluate both our nonparametric estimates and test statistic. We also apply our test to investigate Stock and Watson’s (2009) U.S. macroeconomic data set and find strong evidence of structural changes in the factor loadings.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Factor models provide a flexible way to summarize information from large data sets and have received extensive attention recently. In a factor model, a few latent common factors drive the comovement of a large dimensional vector of time series variables. Although economists realize that the relationships between economic and financial variables may suffer from structural changes over time, the factor loadings, which capture the relationships between these variables and the latent common factors, are usually assumed to be fixed over a long period of time in the conventional factor models (e.g., Stock and Watson, 2002; Bai and Ng, 2002; Bai, 2003). Stock and Watson (2002, 2009) argue that when the factor loadings undergo small instabilities, the factor estimates obtained via the conventional principal component analysis (PCA) are still consistent. However, since macroeconomic datasets typically span a long time period, it is restrictive to assume that the factor loadings are time-invariant or undergo negligible changes during the whole sampling period. In fact, there exist various driving forces such as institutional switching, economic transition, preference changes and technological progress that may influence the relationship between random variables significantly. By ignoring potentially significant structural changes in factor loadings, the estimated common factors might not converge to the desired object and forecasting and inference based on them can be misleading or unreliable. In addition, even if one concerns only the common component, which is equal to the product of factor loadings and the common factors, one may get misleading results.

http://dx.doi.org/10.1016/j.jeconom.2016.12.004
0304-4076/© 2017 Elsevier B.V. All rights reserved.
In recent years, more and more research has focused on structural changes in factor loadings. Stock and Watson (2009) examine the forecasting reliability when there exists a structural break in the factor loadings. Breitung and Eickmeier (2011) propose three statistics to test for structural breaks in factor loadings based on the idea of Andrews (1993). Chen et al. (2014) propose a two-stage procedure to detect big breaks in factor loadings by testing the parameter stability in the regression of one of the estimated factors on the remaining estimated factors. Corradi and Swanson (2014) propose a test to check structural stability of both factor loadings and factor-augmented forecasting regression coefficients. Han and Inoue (2015) propose a test for structural breaks in the factor loadings by checking whether the second moments of the estimated factors exhibit structural changes. Yamamoto and Tanaka (2015) propose a modified version of Breitung and Eickmeier’s (2011) test to ensure that it is robust to the non-monotonic power problem. Cheng et al. (2016) consider the case where both the factor loadings and the number of factors may change simultaneously at a time point. These studies provide appropriate econometric tools to examine the problem of structural breaks in factor loadings. However, all these papers focus on the case of one-time abrupt structural changes. The analyses may be inappropriate if, for example, such driving forces of structural changes as preference changes, technological progress and policy changes, play a role gradually over a long period of time, or some abrupt policy changes also take a period of time to take effect. Indeed, as Hansen (2001) points out, “it may seem unlikely that a structural break could be immediate and might seem more reasonable to allow a structural change to take a period of time to take effect”. Hence, it seems more realistic to assume smooth changes rather than abrupt changes. To the best of our knowledge, Bates et al. (2013) is the only paper that allows for smooth changes in factor loading. By controlling the magnitude of instabilities to be “small”, they show that the principal component (PC) estimators of factors are still consistent. In fact, changes in comovement induced by technological progress and other forces are gradual but fundamental. As a result, we can neither assume the structural changes to be negligible nor check the instabilities of factor loadings under the framework of abrupt structural changes. 

In this paper, we shall model and test smooth structural changes in factor loadings under the local smoothing framework. Specifically, we assume that economic structures undergo gradual but fundamental changes over a long horizon of time, i.e., although the factor loadings change smoothly, the cumulative changes over the entire time period are too large to be ignored. We think that such a situation is realistic in economic and financial analysis as the driving forces such as globalization, preference changes, and technological progress, may all induce evolutionary changes and their accumulative effects cannot be simply ignored. In this case, Stock and Watson’s (2002, 2009) conclusion about small instabilities of factor loadings will fail and the conventional PCA will yield inconsistent estimates of common factors and factor loadings. To conquer the problem, we propose a local version of PCA to estimate the latent factors and the time-varying factor loadings simultaneously. We establish the limiting distributions of the estimated factors and factor loadings under the standard large N and large T framework. We also propose a BIC-type information criterion to determine the number of common factors. Our information criterion extends that of Bai and Ng (2002) and can be applied even when we have a fixed number of breaks in the factor models. So it is robust to the presence of structural breaks in factor models. 

More importantly, we propose an \( L_1 \)-distance-based test statistic to check the stability of factor loadings. The basic idea is to estimate the time-varying factor loadings and the latent common factors by the local version of PCA, and compare the fitted values of the common components with those estimated by the conventional PCA method based on the whole sample. Construction, our test is able to capture both smooth and abrupt structural changes in factor loadings, where the number of abrupt changes is usually assumed to be one in the literature but can be any unknown finite number in our setup. Unlike the existing tests, such as Breitung and Eickmeier (2011), Chen et al. (2014), Han and Inoue (2015), and Yamamoto and Tanaka (2015), which check the stability of the moments of factor loadings or common factors, our test compares the estimates of the common components because it is well known that the latent factors and the factor loadings are not separated identifiably. Moreover, unlike the existing tests for unknown break date, namely the supremum-type tests of Breitung and Eickmeier (2011), Chen et al. (2014), Han and Inoue (2015), and Yamamoto and Tanaka (2015), no trimming of the boundary regions near the starting or ending of period is required for our test. In other words, we allow the breaks to occur near the beginning and the ending of the sample under the alternative.

After the first version of the paper was circulated, we found that Motta et al. (2011), MHV (hereafter) consider a similar model to ours. They propose a time-varying PC regression (PCR) estimator by approximating the kernel-weighted objective function by something that can be estimated via the usual PC procedure. They establish the consistency of Rodriguez-Poo and Linton’s (2001) nonparametric estimator of the time-varying covariance matrix \( \text{Var}(X_t) \) when the elements of \( X_t = (X_{i1}, \ldots, X_{iT})' \) exhibit the time-varying factor structure in (2.1). Based on this, they also establish the convergence rates for their PCR estimates of the factors, factor loadings, and the common components. Nevertheless, MHV do not study the asymptotic distributions of their estimators. Neither do they consider a specification test for the widely used time-invariant factor models.

The rest of this paper is organized as follows. In Section 2, we introduce our factor models with time-varying factor loadings. In Section 3, we propose the local PCA procedure and develop the asymptotic normality for the estimated common factors and factor loadings. In Section 4, we construct our test statistic for time-varying factor loadings, derive the asymptotic distribution of our test and investigate the asymptotic power properties. In Section 5, we study the finite sample performance of our estimation and test via simulation. Section 6 provides an empirical study. Section 7 concludes. All proofs and technical lemmas are relegated to the online supplementary appendix (see Appendix A).
in a multiple time series data set. We assume that $X_t$ admits the following time-varying factor model with $R$ latent common factors

$$F_t = (F_{t1}, \ldots, F_{tR})',$$

and the idiosyncratic error $e_t$ is assumed to be weakly dependent over both cross-sectional unit $i$ and time period $t$.

Our model given by (2.1) generalizes Stock and Watson's (2002) and Bai's (2003) factor models by allowing for structural changes in factor loadings. To allow the model to capture various kinds of time-varying factor loadings, we follow the nonparametric literature on time-varying models [e.g., Cai, 2007; Robinson, 2012; Chen and Hong, 2012; Chen et al., 2012; Su et al., 2016] and model $\lambda_t$ as a nonrandom function of $t/T$:

$$\lambda_t = \lambda(t/T),$$

where $\lambda(\cdot)$ is an unknown piece-wise smooth function of $t/T$ on $(0, 1]$ for each $i$. The specification that $\lambda(t/T)$ is a function of ratio $t/T \in (0, 1]$ rather than time index $t$ is a commonly used scaling scheme in the literature. An intuitive explanation to this requirement is that the increasingly intensive sampling of data points ensures consistent estimation of $\lambda_t(t/T)$ for each $i$ at some fixed point $t/T$ by increasing the amount of data on which it depends. For more discussion, see Stock and Watson (1988, 1991).

Since we model the factor loadings as a function of scaled time, our factor model can be regarded as a semiparametric factor model. In a similar spirit to ours, Park et al.'s (2009) study a semiparametric factor model

$$X_t = (Z_t) F_t + \varepsilon_t,$$

where $Z_t$ is a $d \times 1$ vector of observable random variables, $m(\cdot)$ is a smooth function, the first element of factor vector $F_t$ is given by 1 and the other elements of $F_t$ is unobserved. They apply B-splines to approximate the unknown coefficient function $m(\cdot)$ and propose a Newton–Raphson algorithm to estimate the factors and factor loadings. Compared with Park et al.'s (2009) model, our model does not rely on the availability of $Z_t$ and hence is more flexible to capture various kinds of smooth structural changes in factor loadings.

As in the conventional factor models, $\lambda_t$ and $F_t$ are not separately identifiable. At each time point $t$, we have $\lambda_t F_t = (H_t F_{t1})'$ (or $H_t F_t$ or $H_t F_t'$) for any $R \times R$ nonsingular matrix $H_t$ and we need to impose $R^2$ restrictions in order to identify $\lambda_t$ and $F_t$. Let $A_t = (A_{t1}, \ldots, A_{tR})'$ and $F_t = (F_{t1}, \ldots, F_{tR})'$. One set of identification conditions would be $F_t F_t' = I_R$ and $A_t A_t / N = F_t F_t' = I_R$. The objective function becomes

$$\min_{P, A_t} \text{tr} \left[ (X(t) - F(t) A_t A_t')' (X(t) - F(t) A_t A_t') \right].$$

By optimizing $A_t$ under the restriction $F_t F_t' = I_R$, the objective function becomes

$$\text{tr} \left[ (X(t) X(t))' (X(t) X(t)) \right].$$

Then we can consider maximizing $\text{tr}[F(t) X(t) X(t) F(t)']$ subject to $F(t) F(t)' = I_R$. This is the conventional PCA problem. The estimated factor matrix, denoted by $\hat{F}(t) = \left(\hat{F}_{t1}, \ldots, \hat{F}_{tR}\right)$, is $\sqrt{T}$ times eigenfactors corresponding to the $R$ largest eigenvalues of $T \times T$ matrix $X(t) X(t)'$. Then, to estimate the loading corresponding to each factor, we use $\hat{\lambda}_{ir}$ to denote the $ith$ column of $\hat{A}_t$.

Clearly, if the columns of $\hat{F}(t) / \sqrt{T}$ are the eigenvectors associated with the $X(t) X(t)'$, so are the columns of $-\hat{F}(t) / \sqrt{T}$. In fact, $\hat{F}_{t1}$ is only able to identify the transformed factor $F_{t1} = \hat{k}_{m, n} F_{t1}$ up to the rotation (i.e., $H(t) F_{t1}$ in Theorem 3.1) and up to sign (c.f., Rodriguez-Poo and Linton, 2001, p. 1295); $\lambda_{ir}$ is able to identify $\hat{\lambda}_{ir}$ up to the rotation (i.e., $H(t - 1)_{m, n}$) and up to sign. Similarly, the rotation matrix $H(t)$ is identified up to sign. In practice, we recommend to determine the sign of $\hat{\lambda}_{ir}$ by $\hat{\lambda}_{ir} = \left(\hat{\lambda}_{ir1}, \ldots, \hat{\lambda}_{iR}\right)$.

3.1. Local principal component analysis

For the moment, fix $r \in \{1, 2, \ldots, T\}$. Under the assumption that $\lambda_i : [0, 1] \to \mathbb{R}$ is a smooth function, we have

$$\lambda_{ir} = \lambda_i \left( \frac{t}{T} \right) \approx \hat{\lambda}_i \left( \frac{r}{T} \right) = \hat{\lambda}_{ir}$$

when $T \approx r / T$.\footnote{Stock and Watson (2002) also consider a time-varying factor model with a stochastic drift in the factor loadings: $\lambda_t = \lambda_{ir} + \beta_{ir} \varepsilon_t$. They assume that $\beta_{ir}$ is a scalar and small with $\varepsilon_t = O(t^{-1/2})$, and demonstrate that such small instability does not affect the consistency of the estimated factors. Del Negro and Otrok (2009) propose a dynamic factor model with time-varying factor loadings and stochastic volatility in both the latent factors and idiosyncratic components, and estimate the model via a Gibbs sampling procedure.}

It follows that

$$\lambda_{ir} \approx \lambda_{ir}^2 F_{ir} + \varepsilon_{ir}$$

where $\lambda_{ir} \approx \lambda_{ir}^2 F_{ir}$ is unobserved. They apply $B$-splines to approximate the unknown coefficient function $m(\cdot)$ and propose a Newton–Raphson algorithm to estimate the factors and factor loadings. Compared with Park et al.'s (2009) model, our model does not rely on the availability of $Z_t$ and hence is more flexible to capture various kinds of smooth structural changes in factor loadings.
by fixing the signs of its first row’s elements; in simulations, we can enforce their consistency with the signs of the corresponding elements in \( \lambda_i \).

It is well known that a local constant estimator may suffer from boundary bias problem when we estimate a fixed conditional mean object. As we shall see, because the factors and factor loadings are not separately identified in our model, we are only interested in estimating a rotational version of them and our nonparametric kernel estimators for either object does not have the usual boundary bias issue. Similar phenomenon occurs in MHV when they consider their PCR estimator of factors and factor loadings in a locally stationary model. Despite this observation, we find that the use of a boundary kernel helps us to obtain some uniform results. Specifically, suppose that the true factor \( F_0 \) satisfies \( E (F_0^2) = \Sigma_t \) and the kernel function \( K(\cdot) \) is continuously differentiable probability density function (PDF) with compact support \([-1, 1] \). We frequently call upon the uniform order of the following kernel:

\[
\Delta (r) = \frac{1}{Th} \sum_{i=1}^{T} K \left( \frac{t - r}{Th} \right) E (F_0^2) - \Sigma_t.
\]

Let \([a]\) denote the integer part of \( a \). Noting that \( \frac{1}{Th} \sum_{i=1}^{T} K \left( \frac{t - r}{Th} \right) = 1 + O \left( \frac{1}{Th} \right) \) uniformly in \( r \in \{[[Th], T - [Th]]\} \) by the error analysis in Riemann sum approximation of a definite integral, we have \( \max \{\| K(\cdot) \| : r \in [[Th], T - [Th]] \} \| \Delta (r) \| = O \left( \frac{1}{Th} \right) \). But a similar conclusion does not hold if \( r \) lies in the boundary region, namely, \( r \in \{1, [Th] \} \cup (T - [Th], T) \). For example, for \( r \in \{1, [Th] \} \), we have \( \frac{1}{Th} \sum_{i=1}^{T} K \left( \frac{t - r}{Th} \right) = \frac{1}{Th} \sum_{i=1}^{T} K \left( \frac{t - r}{Th} \right)/ \right) \right) \), which is typically strictly less than \( r \) in large samples unless \( r/Th \rightarrow 0 \). In this case, the order of \( \| \Delta (r) \| \) will be \( O(1) \) instead of \( O \left( \frac{1}{Th} \right) \) when \( r \) lies in the two tails of the \([1, T] \) interval. In order to control the order of \( \Delta (r) \) uniformly well, we follow Hong and Li (2005) and Li and Racine (2006, p. 31) to apply the following boundary kernel:

\[
k_{h, tr}^* = h^{-1}K^* \left( \frac{t - r}{Th} \right) =
\begin{cases} 
    h^{-1}K^\left( \frac{t - r}{Th} \right) / \int_{-1}^{-1} K(u) du, & \text{if } r \in \{1, [Th] \}, \\
    h^{-1}K^\left( \frac{t - r}{Th} \right), & \text{if } r \in ([Th], T - [Th]), \\
    h^{-1}K^\left( \frac{t - r}{Th} \right) / \int_{-h/Th}^{r} K(u) du, & \text{if } r \in (T - [Th], T].
\end{cases}
\]

Note that \( k_{h, tr}^* \) coincides with \( k_{h, tr} \) in the interior region but not in the boundary regions. By using this boundary kernel to replace \( k_{h, tr} \) in (3.2)-(3.4), we obtain the estimators to be analyzed below. But for notational simplicity, we will use \( k_{h, tr} \) to denote \( k_{h, tr}^* \) hereafter.

The estimator \( \hat{F}_t^{(r)} \) is only consistent for a rotational version of the weighted factor \( F_t^{(r)} = \lambda_t^inh_{h, tr} \). To obtain a consistent estimator of \( F_t \) after suitable rotation, we consider a two-stage estimation procedure. Based on the consistent estimators of \( \lambda_t \)’s obtained in the first stage, we can obtain the consistent estimators of \( F_t \), \( t = 1, 2, \ldots, T \), in the second stage, by considering the following least squares problem:

\[
\min_{\lambda_t} \sum_{t=1}^{T} \left[ X_t - \lambda_t F_t \right]^2, \quad t = 1, 2, \ldots, T.
\]

The solution to the above problem is:

\[
\hat{F}_t = \left( \sum_{i=1}^{N} \hat{\lambda}_i \hat{X}_i \right)^{-1} \left( \sum_{i=1}^{N} \hat{\lambda}_i \hat{X}_i \right) \quad \text{for } t = 1, 2, \ldots, T.
\]

3.2 Limiting distributions of the estimated factors and factor loadings

In this subsection, we establish the asymptotic distributions of the estimated common factors and time-varying factor loadings. It is worth mentioning that the factors and factor loadings that appear in our assumptions denote the true values \( F_0 \) and \( \lambda_0 \). But for notational simplicity, we suppress the superscript \( 0 \) hereafter.

Let \( e_t = (e_{t1}, \ldots, e_{TN}) \), \( \gamma_{N, s}(t) = N^{-1}E (e_{ts} e_t^\prime) \), \( \gamma_{N,F}(s, t) = N^{-1}E (F_t e_s e_t^\prime) \), and \( \xi_{ts} = N^{-1}[e_{ts} e_t^\prime - E (e_{ts} e_t^\prime)] \). Define

\[
\sigma_{NT,1}(r) = \frac{h^{1/2}}{\sqrt{NT}} F_0^{\prime}(e^\prime) \Lambda_r, \quad \sigma_{NT,2}(r, t) = \frac{h^{1/2}}{\sqrt{NT}} [F_0^{\prime}(e^\prime) e_t - E (F_0^{\prime}(e^\prime) e_t)]
\]

Let \( C < \infty \) denote a positive constant that may vary from case to case. We make the following assumptions.

**Assumption A.1.**

(i) \( E (e_{ts}) = 0 \) and \( \max_t E (e_{ts}^2) < \infty \).

(ii) \( \max_t E (F_t^{2}) < \infty \) and \( E (F_t^{2}) = \Sigma_t > 0 \) for some \( R \times R \) matrix \( \Sigma_t \).

(iii) \( \Lambda_r \) are nonrandom such that \( \max_t \| \Lambda_t \| \leq \xi_0 < \infty \) and \( N^{-1} \Lambda_t^\prime \Lambda_t = \Sigma_t + O \left( N^{-1/2} \right) \) for some \( R \times R \) positive definite matrix \( \Sigma_t \) and for all \( r \).

(iv) \( \max \sum_{s=1}^{T} [\text{Cov} (F_s, F_{ts}, F_s, F_{ts})] \leq C \) for \( n, m, n, m \) and \( \max \sum_{s=1}^{T} [\| \gamma(s, t) \|] \leq C \) for \( \gamma = \gamma_{N, s}, \gamma_{N, F}, \gamma_{N, FF} \).

(vi) \( \max 1 \leq s, t \leq 1 \) \( E (N^{-1/2} \xi_{ts}) \leq C \) and \( \max_s E (N^{-1/2} \Lambda_t^\prime e_t^2) \leq C \).

(vii) \( \sigma_{NT,1}(r) = 0_{p} (1) \) and \( \max_t E (\sigma_{NT,2}(r, t), t) \leq C \) for each \( r \).

(viii) For all \( r \), the eigenvalues of the \( R \times R \) matrix \( \Sigma_t^{1/2} \Sigma_t^{-1/2} \) are distinct.

**Assumption A.2.**

(i) \( N^{-1/2} \Lambda_t e_t \to N (0, \Gamma_t) \) for each \( r, t \), where \( \Gamma_t = \lim_{N \to \infty} N^{-1} \sum_{s=1}^{N} \sum_{i=1}^{N} \lambda_{hi} \Lambda_t^\prime e_s e_t^\prime \).

(ii) \( \frac{\sqrt{N}^{-1} \sum_{t=1}^{T} k_{h, tr} F_t e_t^\prime}{d} \to N (0, \Omega_t) \), where

\[
\Omega_t = \lim_{T \to \infty} \left[ \frac{h}{T} \sum_{i=1}^{T} k_{h, tr} F_t e_t^\prime \right] + \frac{2h^2}{T} \sum_{i=1}^{T} \sum_{t=1}^{T} k_{h, tr} k_{h, ti} E (F_t F_s e_t^\prime e_s^\prime) \right].
\]

**Assumption A.3.**

(i) The kernel function \( K : \mathbb{R} \to \mathbb{R}^+ \) is a symmetric continuously differentiable PDF function with compact support \([-1, 1]\).

(ii) As \( (N, T) \to \infty, h \to 0, Th^2 \to \infty, Nh^2 \to \infty, \text{Th}/N \to 0, \text{Th}/N^{1/2} \to \infty \).

A.1 mainly imposes moment conditions on the errors, factors, factor loadings, and their interactions. They are widely used in the literature; see, e.g., Bai and Ng (2002) and Bai (2003). In particular, A.1(i), (ii), (iii), and (viii) correspond to Assumptions
C.1, A, B, and G in Bai (2003), respectively. The main difference is that we require $E(F_r F_r') = \Sigma_F$ in A.1(iii) for the reasons to be explained below and $N^{-1/2}A_r A_r = \Sigma_+ + O(N^{-1/2})$ in A.1(iii) to simplify proofs in subsequent sections. Han and Inoue (2015) also assume the latter conditions. A.1(iv)-(vi) restrict the time and cross-sectional dependency for the idiosyncratic errors and the weak dependence between factors and errors, which are in the same spirit as Assumptions C.2-5, D and E in Bai (2003). A.1(vii) is a kernel-weighted version of Assumptions F.1-F.2 in Bai (2003). Please note that Assumption F.1 in Bai (2003) cannot be justified if $F_t$ is random and $E[F(t)\{e_{r,t} - E(e_{r,t})\}] \neq 0$, and one has to replace $F_r [e_{r,th} - E(e_{r,th})]$ by $F_r e_{r,th} - E(F_r e_{r,th})$ in their conditions. A similar remark holds for Assumption 6(a) in Han and Inoue (2015).

Note that we follow Stock and Watson (2002), Breitung and Eickmeier (2011), Motta et al. (2011), and Han and Inoue (2015) and assume that $E(F_t F_t')$ is homogeneous over $t$ in A.1(ii). The latter assumption is made for convenience as it greatly facilitates the derivation of the asymptotic results. It is not as restrictive as it appears in our model. For instance, if $E(F_t F_t') = \Sigma_F$, we could rewrite the common component as

$$\lambda_{nr}^* F_t = \left(\Sigma_F^{-1/2}\Sigma_{r,t}^1\lambda_{r}^* F_t\right)^{1/2} \Sigma_{r,t}^{-1/2} F_t = \lambda_{nr}^* F_t',$$

where $\lambda_{nr}^* = \Sigma_F^{-1/2}\Sigma_{r,t}^1\lambda_{r}^* F_t$ and $r_t = \Sigma_{F,t}^{1/2}\Sigma_{r,t}^{-1/2} F_t$ satisfies $E(F_t F_t') = \Sigma_F$ by construction. Just because the common factors $F_t$ and the factor loadings $\lambda_{r}^*$ are not separately identifiable and we allow for the factor loadings to be time-varying, it does not lose any generality to assume that $E(F_t F_t') = \Sigma_F$ for each $t$. Similarly, following Bai (2003) and Breitung and Eickmeier (2011), we assume that the factor loadings are nonrandom in A.1(iii) because they are treated as functions of time.

A.2(i) extends Assumption F.3 in Bai (2003) to allow for factor loadings to be time-varying and Assumption A.2(ii) is a kernel-weighted version of Assumption Fin Bai (2003). Both parts are used to establish the asymptotic normality of our local PCA estimators and can be verified under some primitive conditions. For example, by the central limit theorem (CLT hereafter) for strong mixing processes (e.g., White, 2001, Theorem 5.20), one can readily verify A.2(ii). Using Davydov inequality, we can argue that the limit $\Omega_{s,t}$ in (3.6) exists. Without further assumptions, we cannot simplify it. If $E(F_t F_t'^*) = \Sigma_F$ for each $s$ and $\{e_{t}\}$, a martingale difference sequence (m.d.s. hereafter) with respect to $F_{t,s}$, the sigma-field generated from $\{e_{t-1}, e_{t-2}, \ldots, F_{t-1}, F_{t-1}, \ldots\}$, then we can readily show that $\Omega_{s,t} = \Phi_t \lim_{t \to \infty} \frac{1}{\sqrt{N}} \sum_{s=t}^{T} K^* \left(\frac{t-s}{\sqrt{N}}\right) = \Phi_t \int_{-1}^{1} K(u^2) du$ if $r \in \{[T], T - [T]\}$. A.3 imposes regularity conditions on the kernel function and bandwidth.

Under these regularity conditions, we now establish the asymptotic distributions for latent factors and time-varying factor loadings estimated via our local PCA method. As is well known, latent common factors and factor loadings are not separately identifiable. However, they can be identified up to an invertible $R \times R$ matrix transformation. Since our local PCA method can be regarded as a conventional PCA method in any small interval around the fixed time ratio $r/T$ for $r = 1, 2, \ldots, T$, we can show that there exists an invertible matrix $H(t)$ such that $H(t)$ is a consistent estimator of $H(t)$ and $\hat{\lambda}_{nr}$ is a consistent estimator of $H(t)$.

The following theorem reports the asymptotic distribution of $\hat{\lambda}_{nr}$.

**Theorem 3.1.** Suppose that Assumptions A.1, A.2(i) and A.3 hold. Then, for each $t = 1, 2, \ldots, T$ and $r = 1, 2, \ldots, T$ such that $|r - t| \leq T$, we have:

$$K_t^* \left(\frac{T-t}{T}\right)^{-1/2} \sqrt{N} \left[\tilde{F}_{r,t}^1 - H(t)F_{r,t}^1\right] \overset{d}{\rightarrow} N(0, \Lambda_{nr}^1 \Lambda_{nr}^{-1})$$

where $H(t) = (N^{-1} A_r A_r)(T^{-1} P(t) P(t)' V(t) V(t)^{-1})$, $V(t)$ denotes the R \times R diagonal matrix of the first R largest eigenvalues of $(NT)^{-1}X(t)X(t)^\prime$, $X(t)$ is the diagonal matrix consisting of the eigenvalues of $\Sigma_{\lambda,\lambda}^{1/2} \Sigma_{r,t}^{1/2}$ in descending order with $\gamma_t$ being the corresponding (normalized) eigenvector matrix ($\gamma_t' \gamma_t = 1$), and $Q_t = \gamma_t' \gamma_t^{-1} \Sigma_{\lambda,\lambda}^{-1/2}$.

**Theorem 3.1** establishes the asymptotic normality of $\hat{\lambda}_{nr}$ and it is a local version of Theorem 1 in Bai (2003). The latter theorem in Bai (2003) reports the asymptotic distribution of the factor estimator in a conventional time-invariant factor model under the identification restrictions: $F(t)T = T_0$ and $A_r A_r/N = diagonal$ matrix with descending diagonal elements. Our Theorem 3.1 reports the asymptotic distribution of our local PCA estimator of the transformed factor under the local identification restrictions: $F(t)T = T_0$ and $A_r A_r/N = diagonal$ matrix with descending diagonal elements. The major difference lies in two aspects. First, all objects that appear in our local version are local versions of the corresponding objects in Bai (2003). Second, the convergence rate of our local PCA estimator $\hat{F}_t$ is $\sqrt{N}$ in contrast with the parametric $\sqrt{N}$-rate in Theorem 1 of Bai (2003).

We note that $\hat{F}_t$ is a consistent estimator for the transformed latent factor $F(t)^1 = k_{1,t}^1 F_t^1$ pre-multiplied by a transformation matrix $H(t)$. Since we allow cross sectional dependence in the error terms, the limiting distribution depends on the cross-section correlation structure among $\{e_{i}\}$. In the case where $e_{i}$ is uncorrelated over $i$, we have $\hat{\Omega}_t = \lim_{t \to \infty} N^{-1} \sum_{i=1}^{N} (\lambda_{ir}^* \sigma_{ir})^2 = E(\sigma_{ir}^2)$. In addition, if $\sigma_{ir}^2 = \sigma_{i}^2$ for each $i$, then we have $\hat{\Omega}_t = \sum_{i} \sigma_{i}^2$. The asymptotic distribution of $\hat{\lambda}_{ir}$ is reported in the next theorem.

**Theorem 3.2.** Suppose that Assumptions A.1, A.2(ii) and A.3 hold. Then, for each $i = 1, 2, \ldots, N$ and $r = 1, 2, \ldots, T$, we have:

$$\sqrt{N} \left[\hat{\lambda}_{ir} - H(t)^i -1 \lambda_{ir}\right] \overset{d}{\rightarrow} N(0, (Q_i^{-1})^\prime \Omega_{ir} Q_i^{-1})$$

**Theorem 3.2** establishes the asymptotic normality of $\hat{\lambda}_{ir}$ and is a local version of Theorem 2 in Bai (2003). The latter theorem in Bai (2003) reports the asymptotic distribution of the estimator of the factor loadings in the conventional time-invariant factor model under the aforementioned identification restrictions. Our Theorem 3.2 reports the asymptotic distribution of our local PCA estimator of the factor loadings under the corresponding local identification restrictions. As expected, all objects that appear in our theorem are local versions of the corresponding objects in Bai (2003), and our local PCA estimator of the factor loadings converges to a rotational version of the true factor loadings at the nonparametric $\sqrt{N}$-rate in contrast with the parametric $\sqrt{T}$-rate in Theorem 2 of Bai (2003).

When $\{e_{i}, F_i\}$ is an m.d.s., the asymptotic variance can be simplified, leading to

$$\sqrt{N} \left[\hat{\lambda}_{ir} - H(t)^i -1 \lambda_{ir}\right] \overset{d}{\rightarrow} N(0, \int_{-1}^{1} K(u^2) du (Q_i^{-1})^\prime \Omega_{ir} Q_i^{-1})$$

when $r \in \{[T], T - [T]\}$. As mentioned above, Theorem 3.1 only establishes asymptotic distribution for the transformed common factor $F_t^1$. Since economists are usually interested in the estimation of the latent factor $F_t$, which are particularly useful in economic modeling and forecasting, it is desirable to establish asymptotic distribution for the estimator of $F_t$ after suitable rotation.

**Theorem 3.3.** Suppose that Assumption A.1, A.2(ii) and A.3 hold. Then, for each $t = 1, 2, \ldots, T$ we have

$$\sqrt{N} \left[\hat{F}_t - H(r)F_t\right] \overset{d}{\rightarrow} N(0, (\Sigma_{\lambda,\lambda}^{-1})^\prime \Gamma_{tt} \Sigma_{\lambda,\lambda}^{-1})$$
Theorem 3.3 reports the asymptotic distribution of the second-stage estimator of the factor. This result is not directly comparable with any result in Bai (2003) because no second-stage refitting is needed for the conventional time-invariant factor models. Interestingly, although the convergence rates of $\hat{F}_t$ and $\hat{\lambda}_{it}$ depend on the bandwidth $h$, the estimated factor $\hat{F}_t$ could achieve the usual parametric $\sqrt{N}$-rate of convergence. In addition, even though we apply the nonparametric local smoothing method, we do not have the usual asymptotic bias–variance tradeoff for the estimators of either the factors or the factor loadings because neither estimators possess the usual asymptotic bias terms. As a result, we cannot derive the conventional optimal bandwidth in terms of minimizing the asymptotic mean squared error of the nonparametric estimates. In practice, we suggest using some data-driven methods to choose the bandwidth. For example, one can use the cross-validation method to choose the bandwidth $h$ by solving the following minimization problem:

$$
\min_h \text{CV}(h) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( \bar{X}_{it} - \hat{\lambda}_{it} \hat{F}_t \right)^2,
$$

where $\hat{\lambda}_{it}$ and $\hat{F}_t$ are the analogue of $\hat{\lambda}_{it}$ and $\hat{F}_t$ by leaving the $r$th time series observation out in the local PCA procedure. But a rigorous study of the asymptotic behavior of $h$ would demand higher order asymptotic theory, which goes beyond the scope of the current paper.

On the surface, Theorem 3.3 does not appear very useful because the rotation matrix $H\hat{F}_t$ is $t$-dependent. This will create some problem that does not occur in the conventional time-invariant factor model. In the latter case, the factor estimator $\hat{F}_t$ is consistent with $H\hat{F}_t$, for a rotation matrix $H$ that does not depend on $r$, and we can continue to use $\hat{F}_t$ to do out-of-sample forecast as in a diffusion index model (e.g., Stock and Watson, 1998, 2002). For our time-varying factor model, the rotational matrix is $t$-dependent; see a similar result in Motta et al. (2011), MHV). Nevertheless, the result in Theorem 3.3 continues to be useful in several important scenarios. First, if we are only interested in the common component, $\hat{\lambda}_{it} \hat{F}_t$, the asymptotics for its estimate will not depend on any rotation matrix. See Theorem 3.4. Second, the rotation matrix does not play any essential role when we consider a specification test for the null hypothesis of time-invariant factor loadings. In the next section, we consider such a test based on the comparison of the estimate $\hat{\lambda}_{it} \hat{F}_t$ of the common component $\lambda_{it} \hat{F}_t$ under the null and that $\hat{\lambda}_{it} \hat{F}_t$ under the alternative. Third, even if the rotation matrix $H\hat{F}_t$ is $t$-dependent, the estimates $\hat{F}_t$ can still be used for out-of-sample forecasting when the factors have time-varying behavior. To see this, we observe that $H\hat{F}_t$ can be written as $H(t/T)$ when the factor loadings $\lambda_{it}$ are smooth and written as $\lambda_t(t/T)$. For example, we can consider the following 1-step-ahead forecasting model

$$
y_{t+1} = \alpha_t \hat{F}_t + \epsilon_{t+1},
$$

where $\alpha_t = \alpha(t/T)$ is time-varying. When $\hat{F}_t = H\hat{F}_t = o_p(1)$, we have

$$
y_{t+1} = \alpha_t \hat{F}_t + \epsilon_{t+1} = \alpha_t \hat{F}_t + \epsilon_{t+1} = \alpha_t \hat{F}_t = o_p(1).
$$

Theorem 3.4. Suppose that Assumptions A.1–A.3 hold. Then, for each $i = 1, 2, \ldots, N$ and $r = 1, 2, \ldots, T$, we have:

$$
\frac{1}{N} \left( \frac{1}{N} \text{V}_{it} + \frac{1}{Th} \text{V}_{2it} \right)^{-1/2} \left( \hat{C}_t - \hat{C}_0 \right) \rightarrow_d N(0, 1),
$$

where $\text{V}_{it} = \lambda_{it}^2 \Sigma_{\lambda}^{-1} F_t \Sigma_{\lambda}^{-1} \lambda_{it}$ and $\text{V}_{2it} = F_t \Sigma_{\lambda}^{-1} \Sigma_{\lambda}^{-1} F_t$.

Theorem 3.4 parallels Theorem 3 in Bai (2003). Let $C_{NT} = \min(\sqrt{N}, \sqrt{T})$. Noting that

$$
\frac{c_{NT}^2}{N} \text{V}_{it} + \frac{c_{NT}^2}{T} \text{V}_{2it} \rightarrow_{d} N(0, 1)
$$

and the denominator in the above expression is bounded from both above and below, we can conclude that the convergence rate of $\hat{C}_t$ is given by $C_{NT}$. This rate is different from $\min(\sqrt{N}, \sqrt{T})$-convergence rate for the estimator of the common component in the conventional time-invariant factor model. To make inference, it is standard to obtain the consistent estimates of $\text{V}_{it}$ and $\text{V}_{2it}$. To save space, we omit the details.

To study the uniform convergence of $\hat{\lambda}_{it}, \hat{F}_t$, and $\hat{C}_t$, we add the following assumption.

Assumption A.4. (i) $\|\epsilon\|_p = o_p\left(N^{1/2} + T^{1/2}\right)$ and $\max_{t,s} \left\| \sum_{i=1}^{N} \left( e_{it} e_{is} - E(e_{it} e_{is}) \right) \right\| = o_p\left(N(T)^{1/2} + (\ln NT)^{1/2}\right)$.

(ii) $\max \left\| \hat{\lambda}_{it}^\prime \hat{\lambda}_{it} - \lambda_{it}^\prime \lambda_{it} \right\| = o(1)$, and the eigenvalues of $\Sigma_{\lambda}$ are bounded below from $0$ and above from infinity uniformly in $t$.

(iii) $\max_{i,t} \left\| \alpha_{it}^\prime \alpha_{it} - \alpha_t^\prime \alpha_t \right\| = o(1)$,

(iv) $\max_{i,t} \left\| \alpha_t^\prime \alpha_t \right\| = o_p\left(N(T)^{1/2}\right)$, and $\max_{i,t} \left\| \alpha_t^\prime \alpha_t \right\| = o_p\left(NT^{1/2}\right)$;

(v) $\max \left\| \Sigma_{\lambda t} \right\| = o_p((\ln T)^{1/2})$.

The conditions in A.4 can be verified under some primitive conditions that are used in the factor literature. For example, Moon and Weidner (2015) demonstrate that the first part of A.4(i) can be satisfied for various processes; Su et al. (2015) verify similar conditions to those in A.4(ii)–(vi) under some mixing conditions. Even though the first part of A.4(i) is not directly assumed in Bai and Ng (2002), BN hereafter), their proof of Lemma 4 contains some errors. In an online erratum, they make the correction by assuming that $\frac{1}{\sqrt{\ln N}} \epsilon_t$ has the largest eigenvalue that is $o_p\left(N^{-1/2} + T^{1/2}\right)$, which is equivalent to requiring that $\|\epsilon\|_p = o_p\left(N^{1/2} + T^{1/2}\right)$. The second part of A.4(i) is essentially the same as Assumption C.5 in BN except the logarithm term which is required in order to justify the uniformity in $(t, s)$. A.4(ii) strengthens A.1(iii) as we now need to $1/\alpha_t^\prime \alpha_t$ and its limit to behave well uniformly in $t$. A.4(iii)–(vi) imposes conditions on the uniform probability order of some summation objects. The first part of A.4(iii) corresponds to
and extends Assumption D in BN because we need the condition to hold uniformly in \((i, r)\). The second and third parts of Assumption A.4(iii) is new but can be verified under weak data dependence conditions and some moment conditions. Similarly, one can also verify the conditions on A.4(iv)–(vi). In the online supplementary appendix (see Appendix A), we verify A.4(vi) under some primitive conditions.

The following theorem studies the uniform convergence of \(\hat{\lambda}_n, \hat{F}_t, \) and \(C_{it} \).

**Theorem 3.5.** Suppose that Assumptions A.1 and A.3–A.4 hold. Then

(i) \(\max_{i,t} ||\hat{\lambda}_{it} - H^{-1}(\cdot)\lambda_{it}|| = O_p((\ln N / \ln T)^{-1/2}),\)

(ii) \(\max_{i,t} ||\hat{F}_t - H^{(i)}F_t|| = O_p((N / \ln T)^{-1/2}),\)

(iii) \(\max_{i,t} ||\hat{C}_{it} - C_{it}|| = O_p((N / \ln T)^{-1/2}) + o_p((\ln T / \ln T)^{-1/2}/T^{1/8}).\)

**Theorem 3.5** establishes the uniform convergence rates for our estimators of the factor loadings, factors, and common components. Bai (2003) establishes the uniform convergence rate for the PCA estimator of the factors only in a conventional time-invariant factor model, which is given by \(O_p((N^{-1/2}T^{1/2} + T^{-1/2}) = O_p(N^{-1/2}T^{1/2})\) by Assumption A.3(ii). Apparently, our rate significantly improves over his rate despite our utilization of nonparametric method. The first and second terms in the convergence rate of \(C_{it}\) signify the contributions of the factor estimate and factor loadings estimate, respectively. \(T^{1/8}\) in the second term reflects the probability order of \(\max_{i,t} ||\hat{F}_t||\), which is \(O_p(T^{1/8})\) under Assumption A.1(ii) and by Markov inequality. Similar object does not appear in the first term because we assume that the factor loadings \(\lambda_n\) are uniformly bounded in Assumption A.1(iii). In the special case where the factors are uniformly bounded, the second term \(o_p((\ln T / \ln T)^{-1/2}/T^{1/8})\) can be replaced by \(O_p((\ln T / \ln T)^{-1/2})\).

3.3. Determination of the number of factors

In the above analysis, we assume that the number of factors, \(R\), is known. In practice, one has to determine \(R\) from the data. Here we assume that the true value of \(R\), denoted as \(R_0\), is bounded from above by a finite integer \(R_{\text{max}}\). We propose a BIC-type information criterion to determine \(R_0\).

Let \(\hat{F}_t (R)\) and \(\hat{\lambda}_n (R)\) denote the local PCA estimators of the factors and factor loadings by assuming \(R\) factors in the model and using the following normalization rule

\[
N^{-1} \hat{A}_t A_t = I_R 	ext{ and } T^{-1} F^{(1)} F^{(1)^T} \text{ is a diagonal matrix with descending diagonal elements,}
\]

where we make the dependence of the \(R \times 1\) vectors \(\hat{F}_t (R)\) and \(\hat{\lambda}_n (R)\) on \(R\) explicit. Let \(\hat{\lambda}_t (R) = (\hat{\lambda}_t (R), \ldots, \hat{\lambda}_t (R))'\) and \(\hat{\lambda}_t (R) = (NT)^{-1} X^{(t)} X^{(t)^T} \hat{\lambda}_t (R)\) for \(r = 1, \ldots, T\). Let \(\lambda_n (R)\) denote the transpose of the \(i\)th row of \(\hat{\lambda}_t (R)\). Define

\[
V (R, \{\hat{\lambda}_t (R)\}) = \min_{F_t = (F_{t1}, \ldots, F_{tR})} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left| X_{it} - F_t \hat{\lambda}_n (R) \right|^2.
\]

As Bai and Ng (2002, p. 199) remark, the above sum of squared residuals (SSR) does not depend on which estimate of the factor loadings is used because they span the same vector space for any localization point \(r\). We use the normalization rule in (3.7) instead of the one used before because we find that in this case it is relatively easier to study the asymptotic properties of \(\lambda_n (R)\), \(\lambda_t (R)\), and the associated rotational matrix \(H^{(t)} (R)\) (see Lemma A.8 in the online supplement for the definition (see Appendix A)) than using the previous rule, regardless of the value of \(R\) relative to the true value \(R_0\). We define \(\tilde{\lambda}_n (R) = (NT)^{-1} X^{(i)} X^{(i)^T} \hat{\lambda}_n (R)\) and use it to obtain for \(V (\cdot, \cdot)\) for two reasons. First, \(\tilde{\lambda}_n (R)\) always has full rank \(R\) because \(N^{-1} \hat{A}_t A_t = I_R\) by normalization. In contrast, the asymptotic rank of \(\hat{\lambda}_n (R)\) is given by \(\min (R, R_0)\) (see Lemma A.8) and hence \(\tilde{\lambda}_n (R)\) is informative on \(R_0\) when \(R > R_0\). Second, we study the asymptotic behavior of \(\tilde{\lambda}_n (R)\) in Lemma A.9 via the study of \(\hat{\lambda}_n (R)\). Note that Lu and Su (2016) apply a similar trick in their shrinkage estimation of panel data models with a multi-factor error structure.

Motivated by Bai and Ng (2002), we propose the following BIC-type information criterion to determine \(R_0\):

\[
IC (R) = \ln N + \left( \{ \lambda_t (R) \} + p_R \right) R,
\]

where \(p_R\) plays the role of \(\ln (NT)/(NT)\) in the case of BIC and \(2/(NT)\) in the case of AIC. Let \(\hat{R} = \arg \min_R IC (R)\).

We add the following two assumptions.

Assumption A.5(i) \(\max_{i,t} E \left\| N^{-1/2} \hat{A}_t e_t F_t \right\|^4 \leq C\) and \(\max_{i,t} E \left\| N^{-1/2} (F_t e_t e_t F_t - E (F_t e_t e_t F_t)) \right\|^2 \leq C\).

(ii) \(\max_{t} E \left\| \frac{1}{(NT)^{1/2}} \sum_{i=1}^{N} \sum_{t=1}^{T} k_{ith} \left( F_t e_t e_t F_t - E (F_t e_t e_t F_t) \right) \right\| \leq C\).

Assumption A.6. As \((N, T) \rightarrow \infty, R \rightarrow 0\) and \(\rho_{NT} C_{NT} \rightarrow \infty\) where \(C_{NT} = \min(\sqrt{NT}, \sqrt{N})\).

Assumptions A.5(1)–(2) are new and needed in the proof of Lemma A.9 in the appendix (see Appendix A). The conditions on \(\rho_{NT}\) in A.6 are typical conditions in order to estimate the number of factors consistently. The penalty coefficient \(\rho_{NT}\) has to shrink to zero at an appropriate rate to avoid both overfitting and underfitting.

**Theorem 3.6.** Suppose that Assumptions A.1 and A.3–A.6 hold. Then

\[ P \left( \hat{R} = R_0 \right) \rightarrow 1 \text{ as } (N, T) \rightarrow \infty. \]

**Theorem 3.6** indicates the class of information criteria defined by \(IC (R)\) in (3.9) can consistently estimate \(R_0\). To implement the information criterion, one needs to choose the penalty coefficient \(\rho_{NT}\). Following the lead of Bai and Ng (2002), we suggest setting \(\rho_{NT} = \frac{k + \ln h}{N h R_0} \ln \left( \frac{N h R_0}{k + \ln h} \right)\) or \(\rho_{NT} = \frac{k + \ln h}{N h R_0} \ln C_{NT}\) with \(C_{NT} = \min(\sqrt{NT}, \sqrt{N})\) and evaluate the performance of these two information criteria in our simulation studies.

4. Testing for the constancy of factor loadings over time

In this section, we propose a formal test for the constancy of factor loadings over time and study its asymptotic properties under a sequence of Pitman local alternatives.

4.1. The hypotheses

The null hypothesis of time-invariant factor loadings could be written as

\[ H_0 : \lambda_{it} = \lambda_{it} \text{ for } i = 1, 2, \ldots, N \text{ and } t = 1, 2, \ldots, T, \]

and the alternative hypothesis is

\[ H_1 : \lambda_{it} \neq \lambda_{it} \text{ for some } (i, t). \]
where $\lambda_{\theta}$ is an unknown vector of factor loadings. We allow $\lambda_{t} = \lambda_{t}(t/T)$ to be a piece-wise smooth function on $[0, 1]$ for each $i$ with a finite number of discontinuities under $H_{1}$.

Obviously, under the null hypothesis, $\lambda_{t}$ is time-invariant and the model (2.1) degenerates to the conventional factor model as studied by Stock and Watson (2002), Bai and Ng (2002), and Bai (2003), among others. Motivated by the potential presence of structural changes in factor models, Breitung and Eckmeier (2011), Chen et al. (2014), Han and Inoue (2015), and Yamamoto and Tanaka (2015) propose various tests for the existence of a single structural change in the factor loadings. In contrast, we do not impose any essential restriction on the alternative. The alternative (4.2) allows for a finite number of abrupt structural breaks. More importantly, by assuming $\lambda_{t}$ to be a piece-wise smooth function under the alternative, we also allow for smooth structural changes in the factor loadings. This type of alternative seems more reasonable and realistic than the single structural break alternative given the fact that the driving forces of structural changes such as preference changes, technological progress and policy modifications accrue gradually in a long period of time.

4.2. Test statistic

Under $H_{0}$, we can follow Bai and Ng (2002) and Bai (2003) to apply the conventional PCA method to estimate the common factors and time-invariant factor loadings. Under $H_{1}$, we can apply the local PCA method to estimate the common factors and time-varying factor loadings. Then, we can construct a quadratic test statistic to check $H_{0}$ by measuring the squared distance between the estimates of the common components under $H_{0}$ and those under $H_{1}$.

Let $e_{t} = e_{t} + (\lambda_{t} - \lambda_{0})^\prime F_{t}$. Let $X_{t} = (X_{1t}, \ldots, X_{Nt})^\prime$, $e_{t} = (e_{1t}, \ldots, e_{Nt})^\prime$, $F = (F_{1}, \ldots, F_{N})^\prime$, and $A_{0} = (\lambda_{10}, \ldots, \lambda_{N0})^\prime$. Let $X = (X_{1}, \ldots, X_{N})^\prime$, $e = (e_{1}, \ldots, e_{N})^\prime$, $e^{1} = (e_{1}^{1}, \ldots, e_{N}^{1})^\prime$. Then we can rewrite (2.1) in matrix form

$$X = FF^\prime e + e^{1}.$$  (4.3)

The conventional PCA method solves the following minimization problem

$$\min_{F} \text{tr} \left( X - FF^\prime \right) \left( X - FF^\prime \right) = \sum_{t=1}^{T} \sum_{t=1}^{N} (X_{it} - \lambda_{0}^{\prime} F_{it})^{2},$$

under certain identification restrictions. In this paper, we follow Bai (2003) and consider the following identification restrictions: $T^{-1/2}F = T_{r}$ and $A^\prime A$ is a diagonal matrix. Let $F_{t}$ and $\tilde{A}_{t}$ be the principal component estimators of $F_{t}$ and $\lambda_{0}$, respectively under the above identification restrictions. Let $\tilde{F} = (F_{t}, \ldots, F_{T})^\prime$ and $A_{0} = (\lambda_{10}, \ldots, \lambda_{N0})^\prime$. It is well known that $F$ is $\sqrt{T}$ times eigenvectors corresponding to the $k$ largest eigenvalues of the $T \times T$ matrix $XX^\prime$, and $\tilde{A}_{0} = (\widetilde{F} F)^{-1/2} X = T^{-1/2} FX$.

Given the estimates $\lambda_{0}^{\prime} \tilde{F}_{t}$ of the common components $\lambda_{0}^{\prime} F_{t}$ under $H_{0}$ and those $(\lambda_{0}^{\prime} \tilde{F}_{t})$ under $H_{1}$, we propose a quadratic form statistic to check the null hypothesis of time-invariant factor loadings based on the comparison of the two sets of estimates:

$$\hat{M} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\lambda_{0}^{\prime} \tilde{F}_{t} - \lambda_{0}^{\prime} \tilde{F}_{t})^{2}. $$  (4.4)

We will show that after being appropriately rescaled and centered, $M$ follows the standard normal distribution under the null hypothesis and has non-trivial power to detect a sequence of Pitman local alternatives that converge to the null at a suitable rate.

4.3. Asymptotic null distribution

In this subsection, we study the asymptotic distribution of $\hat{M}$ under $H_{0}$. Let $\|A\|_{q} = E \|A^{\prime} \|^{q/2}$ for $q \geq 1$. We add the following assumptions.

**Assumption A.7.** (i) For each $i = 1, 2, \ldots, N$, the process $(e_{it}, t = 1, 2, \ldots)$ is an m.d.s. such that $E (e_{it} T_{iT}) = 0$ where $T_{iT} = (F_{i}, T_{i1} = (F_{1}, 1, \ldots, e_{i1}, e_{i2}, \ldots)$. (ii) For each $i = 1, 2, \ldots, N$, the process $(e_{it}, F_{i}, t = 1, 2, \ldots)$ is strong mixing with mixing coefficients $\alpha_{i}(\cdot), \alpha_{i}(\cdot) \equiv \max; \alpha_{i}(\cdot)$ satisfies $\sum_{i=1}^{\infty} \alpha_{i}(\cdot)^{p} \leq C < \infty$ for some $\delta > 0$. In addition, there exists an integer $T_{0} \in [1, T]$ such that $T^{-2} \max(T_{0}, T_{0}^{-1} h^{-1}, T_{0}^{-2} h^{-2}) \to 0$ and $N^{2} T^{-1} \max(T_{0}, T_{0}^{-1} h^{-1}, T_{0}^{-2} h^{-2}) \to 0$ as $(N, T) \to \infty$. (iii) $\max_{i,t} \|t_{it} e_{it}^{2}\|_{4,44} \leq C$ and $\max_{i,t} \|t_{it} e_{it}^{2}\|_{4,44} \leq C$.

Assumption A.7 are new in the literature on testing for structural breaks in large dimensional factor models. Previous authors only consider testing for the presence of a one-time structural break under the alternative. To avoid the comparison between two large-dimensional factor loadings matrices, they reduce the infinite-dimensional problem to a finite-dimensional one in different ways. For example, Chen et al. (2014) run the regression of one estimated factor on the remaining ones and then test for the structural changes in such a linear regression by constructing the sup-Wald and sup-LM statistics of Andrews (1993); Han and Inoue (2015) construct their sup-Wald and sup-LM statistic by comparing the pre- and post-break subsample second moments of the estimated factors. In either case, the test statistics have the same asymptotic distribution as the conventional sup-Wald statistic of Andrews (1993). In contrast, we are dealing with the infinite dimensional parameter problem directly through the construction of an $L_{2}$-distance statistic in (4.4).

Our test is a nonparametric test and we have to analyze much more complicated objects than those in Chen et al. (2014) and Han and Inoue (2015). This explains why our assumptions are also different from those in the early literature.

A.7(i) assumes that the process $(e_{it}, t = 1, 2, \ldots)$ is an m.d.s. with respect to the filter $T_{iT}$ and it allows cross-sectional dependence among the error terms. This assumption is essential for the establishment of the asymptotic distribution of our test statistic under the null hypothesis and a sequence of Pitman local alternatives. It is possible to allow for both serial dependence and cross-sectional dependence in the error terms. But that will substantially complicate the asymptotic analysis and we are not sure how to estimate the asymptotic variance of our raw test statistic in this case.

A.7(ii) requires the process $(e_{it}, F_{i}, t = 1, 2, \ldots)$ to be strong mixing with some algebraic mixing rate. With more complicated notation, one can allow different individual time series to have different mixing rates and then relax the summability mixing condition to $\lim_{T \to \infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \alpha_{i}(\cdot)^{p} \leq C < \infty$. If the processes are strong mixing with a geometric rate (e.g., $\alpha(s) = \rho^{s}$ for some $\rho \in [0, 1]$), then the conditions on $\alpha(\cdot)$ can be all met by specifying $T_{0} = [C_{0} \log T]$ for some sufficiently large positive constant $C_{0}$. A.6(iii) assumes some moment conditions on $F_{it}$ and $\lambda_{0}$, which, in conjunction with A.7(ii), reflects the usual tradeoff between the dependence and moment conditions: a smaller value of $\delta$ requires faster decay in the mixing coefficients but less stringent moment conditions. Like A.1(iv), A.7(iv) controls the cross-sectional dependence among $(F_{it}, e_{it}, i = 1, 2, \ldots, N)$. Under A.7(iii), this condition becomes redundant if we would assume...
independence of $e_i = (e_{i1}, \ldots, e_{ip})$ across $i$ conditional on the factors.

In addition, we need to strengthen A.3(ii) to the following assumptions:

**Assumption A.3. (iii)** As $(N, T) \to \infty$, $h \to 0$, $Th^2 \to \infty$, $Nh^2 \to \infty$, $Th/N \to 0$, $Th/\ln T \to \infty$, and $N^2T^{-1}h^{-1}(\ln T)^{-2} \to \infty$.

Let $V_{NT}$ denote the $R \times R$ diagonal matrix of the first $R$ largest eigenvalues of $(NT)^{-1}XX'$ in decreasing order and $H = (N^{-1}A'_iA_i)(T^{-1}F)\bar{V}_{NT}^{-1}$. Let $\hat{k}_{h,t} = h^{-1}k^* (\frac{t}{\sqrt{m}})$, $\hat{\kappa} t = \tilde{K} (\frac{t}{\sqrt{m}})$ with $K(u) = \int_0^1 K(u) (u-u') du$ being the two-fold convolution kernel of $K(\cdot)$. For example, if we use the Epanechnikov kernel $K(u) = 0.75(1-u^2)I(|u|\leq 1)$ with $I\{\cdot\}$ being the usual indicator function, then $K(u) = \frac{1}{\sqrt{3}} - \frac{3}{4}u^2 + \frac{3}{8}|u| - \frac{3}{320}|u|^3 I(|u|\leq 2)$. Let $L_t = \delta_t H(t)\bar{H}(t)$. Define

$$
\bar{B}_{NT} = \frac{h^{1/2}}{NT^{1/2}} \sum_{i=1}^T \sum_{t=1}^T \sum_{\tau \neq \nu \leq T} (F'_iL_tF_\tau)^2 e_i^2 e_\tau^2,
$$

$$
\sqrt{V}_{NT} = 2T^{-2}N^{-1}h^{-1} \sum_{1 \leq \tau \neq \nu \leq T} \hat{k}^2 \hat{\kappa} (F'_iH_0\hat{\Sigma}_T H'_\tau F_\nu)^2 (e'_i e'_\tau)^2,
$$

where $\hat{e}_i = X_i - \hat{\lambda}_i \hat{F}_i$. Then we consider the feasible test statistic:

$$
\hat{J}_{NT} = \hat{\psi}_{NT}^{-1/2} \left( TN^{1/2}h^{1/2}\hat{M} - \hat{\mu}_{NT} \right).
$$

The following theorem establishes the consistency of $\hat{B}_{NT}$ and $\sqrt{V}_{NT}$ and the asymptotic normality of $\hat{J}_{NT}$.

**Theorem 4.2.** Suppose that Assumptions A.1, A.3(i) and (iii*), A.4, and A.7 hold. Then under $H_0$, $\hat{B}_{NT} = B_N + o_P(1)$, $\sqrt{V}_{NT} = V_N + o_P(1)$, and $\hat{J}_{NT} \to N(0,1)$.

**Theorem 4.2** indicates that our test statistic $\hat{J}_{NT}$ is asymptotically pivotal under $H_0$. We can compare the value of $\hat{J}_{NT}$ to the critical value $z_n$, the upper $\alpha$-percentile of the $N(0,1)$ distribution, as the test is one-sided, and reject the null at $\alpha$ significance level when $\hat{J}_{NT} > z_n$.

### 4.4. Asymptotic local power

To study the asymptotic local power property of our test, we consider the following sequence of local alternatives:

$$
H_1(NT) : \lambda_i = \lambda_{io} + o_N(\frac{g_i (\frac{i}{T})}{T})
$$

for each $i$ and $t$.

where $a_N \to 0$ as $(N, T) \to \infty$, it controls the speed at which the local alternative converges to the null hypothesis, and $g_1 (\cdot)$ is a vector-valued piecewise smooth function with a finite number of discontinuity points. Noting that $\lambda_{io} + o_N(\frac{g_i (\cdot)}{T}) = (\lambda_{io} + c_{i,NT} / a_N)$ for any $c_{i,NT} = o(a_N)$, below we will assume that

$$
\int_0^1 g_i(u) du = 0
$$

for location normalization purpose. With this normalization, both $\lambda_{io}$ and $g_i (\cdot)$ can be dependent on the sample sizes $N$ and $T$. But for notational simplicity, we continue to write them as $\lambda_{io}$ and $g_i (\cdot)$ instead of $\lambda_{io,NT}$ and $g_i(\cdot)$.

Let $g_{ia} = g_i (\frac{\cdot}{T})$, $g^a_{ia} = F'_i g_{ia} (\frac{\cdot}{T})$, and $g^a_i = (g^a_{i1}, \ldots, g^a_{ip})$. Define

$$
\Pi_1 = \lim_{(N,T) \to \infty} (NT)^{-1} \sum_{i=1}^T \text{tr} \left( \left( N^{-1} A'_i g^a_i \right) \left( N^{-1} g^a_i A_0 \right) \right) \times \left( (H_0^{-1})' \hat{V}'_N H_0^{-1} \hat{\Sigma}_T \left( H_0^{-1} \right)' \hat{V}'_N H_0^{-1} \right).
$$

$$
\Pi_2 = \lim_{(N,T) \to \infty} (NT)^{-1} \sum_{i=1}^T \sum_{\tau \neq \nu} \text{tr} \left( \hat{\Sigma}' T g_{ia} g_{\tau a} \right). \quad (4.5)
$$

To study the asymptotic power property of $\hat{J}_{NT}$, we impose the following assumption:

**Assumption A.8.** (i) For each $i = 1, 2, \ldots, N$, $g_i (\cdot)$ is piecewise continuous with a finite number of discontinuous points on $[0,1]$.

(ii) $\max_{1 \leq \tau \leq T} \left( \frac{1}{NT} \sum_{i=1}^T \sum_{\tau \neq \nu} k_{h,t} F'_i g_{ia} g_{\tau a} \right) = o_P \left( \left( NTH/\ln (NT) \right)^{-1/2} \right)$.

(iii) The limits $\Pi_1$ and $\Pi_2$ defined in (4.5) exist.

Assumption A.8 allows the factor loadings to change smoothly over time or abruptly at a finite number of unknown discontinuity points. In either case, we assume that the factor loadings are uniformly bounded in A.1(iii) to facilitate the asymptotic analysis.

The following theorem studies the asymptotic local power property of $\hat{J}_{NT}$.
Theorem 4.3. Suppose that Assumptions A.1, A.3(i) and (ii*), A.4, and A.7–A.8 hold. Then under $\mathbb{H}_1$ (null) with $a_{NT} = T^{-1/2}N^{-1/4}h^{-1/4}$, $\hat{\beta}_{NT} = \beta_{NT} + o_p(1)$, $\hat{\gamma}_{NT} = \gamma_{NT} + o_P(1)$, and $\hat{\beta}_{NT} \overset{d}{\rightarrow} N(\theta_{0}, 1)$, where $\theta_{0} = (\Pi_1 + \Pi_2) / \sqrt{\Pi_0}$ and $\Pi_0 = \lim_{N(T) \to \infty} \Pi_{NT}$.

Theorem 4.3 implies that our test has nontrivial asymptotic power against the class of local alternatives that deviate from the null hypothesis at the rate $a_{NT} = T^{-1/2}N^{-1/4}h^{-1/4}$. Note that $N$ and $T$ enter $a_{NT}$ differently. An intuitive reason for this is that the cross sectional and time dimensions play distinct roles when $i$ and $t$ enter the null hypothesis $H_{0N}$ asymmetrically.

Similar local alternative rates have been found in the literature. For example, Su and Chen (2013) consider testing slope homogeneity in linear panel data models with interactive fixed effects and they find that their LM-type statistic has nontrivial power against parametric local alternatives converging to the null at the rate $T^{-1/2}N^{-1/4}$. This rate is the same as that obtained by Pesaran and Yamagata (2008) for testing slope homogeneity in linear panel data models with the usual one-way error component. It is similar to our rate $T^{-1/2}N^{-1/4}h^{-1/4}$ for our nonparametric local alternative. Note that we allow the existence of a finite number of unknown factor loadings. As a result, our test has power against not only the smooth structural changes in factor loadings but also a finite number of abrupt changes.

In order for our test to have non-trivial power against the local alternatives, we need $\theta_{0} > 0$. This requires that the factor loadings should not be time-varying only for an asymptotically negligible set of individuals. $\mathcal{N} = \{1, 2, \ldots, N\}$. Let $|$ denote the cardinality of a set $\mathcal{N}$. Define a subset of $\mathcal{N}$: $\delta_{NT} = \{i \in \mathcal{N}: \lambda_{it} = \theta_{0} \text{ for all } t\}$. Let $\delta_{NT} = \mathcal{N} \setminus \delta_{NT}$, the complement of $\delta_{NT}$ relative to $\mathcal{N}$. It is easy to verify that if $|\delta_{NT}| / N = 0$ (1), then $\Pi_1 = \Pi_2 = \pi_0 = 0$ and thus our test does not have power in this case. Similar phenomenon occurs in Su and Chen’s (2013) test for slope homogeneity where they require the degree of heterogeneity to be sufficiently large. Similarly, the nonzero $\pi_0$ also requires that $\delta_{NT}$ should not be nonzero for an asymptotically negligible number of periods. Let $\mathcal{T} = \{1, 2, \ldots, T\}$. For each $i \in \mathcal{N}$, define a subset of $\mathcal{T}$: $\delta_{IT} = \{t \in \mathcal{T}: g_{it} = 0\}$. Let $\delta_{IT} = \mathcal{T} \setminus \delta_{IT}$. If $\max_{i \leq N \mid \delta_{IT}} |\lambda_{it}| / T = 0$ (1), we can also verify that $\Pi_1 = \Pi_2 = \pi_0 = 0$. Implying that our test does not have power in this case either. In general, as long as at least a fixed proportion of individuals $\mathcal{N}$ either undergo a one-time or multiple times of abrupt structural change, or undergo a nonshrinking proportion of $T$ periods of smooth structural changes, $\pi_0 > 0$ and our test has asymptotic power to detect them.

4.5. Asymptotic global power

To study the asymptotic global power property of our test, we define $F_T = \left[ \tilde{F}, \tilde{F}/T = \Pi_{IT} \right]$ and $A_N = \left\{ \hat{\lambda}: \hat{\lambda}' \hat{\lambda} = \text{diagonal matrix} \right\}$, where $\hat{\lambda} = (\tilde{\lambda}_1, \ldots, \tilde{\lambda}_N)'$ and $\tilde{F} = (\tilde{F}_1, \ldots, \tilde{F}_T)'$.

Assumption A.9. There exists $\epsilon_{AF} > 0$ such that $\lim_{(N, T) \to \infty} \inf_{(\lambda, F) \in A_N \times F_T} \frac{1}{T} T \sum_{i=1}^{T} \left| \sum_{n=1}^{N} (\lambda_{it} F_{it} - \hat{\lambda}_{it} \tilde{F}_{it}) \right|^2 \geq \epsilon_{AF}$.

Assumption A.9 is intuitively clear: in the spaces of factors and factor loadings such that the normalization rules in $F_T$ and $A_N$ are satisfied, we cannot find any time-invariant factor loadings $\tilde{\lambda}_i$’s and the associated factors $\tilde{F}_i$’s such that $\tilde{\lambda}' \tilde{F}_i$ converges to the true common component $\lambda_{it} F_{it}$ in the sense of mean square error. If A.9 is violated, then we can approximate the time-varying factor model by a time-invariant factor model so that the instability of the factor loadings has to be small and asymptotically negligible.

Theorem 4.4. Suppose that Assumptions A.1, A.3–A.4 and A.9 hold. Then under the global alternative $\mathbb{H}_1$, $P(\hat{\lambda} \geq \epsilon_{AF}) \to 1$ as $N, T \to \infty$ for any positive sequence $c_{NT}$ that is $o(T^{1/2}h^{1/2})$.

Theorem 4.4 implies that $\hat{\lambda}$ is consistent and divergent to infinity at the rate $T^{1/2}h^{1/2}$. Note that A.6–A.8 are not required here as there is no need to derive the asymptotic distribution of $\hat{\lambda}$ or to study the consistency of the bias or variance estimator.

4.6. A bootstrap version of our test

It is well known that a kernel-based nonparametric test may not exhibit good size in finite samples because its asymptotic null distribution may not approximate its finite sample distribution well when the null hypothesis is satisfied in the real data. Therefore it is worthwhile to propose a bootstrap procedure to improve the finite sample performance of our test.

There are various ways to conduct the bootstrap. One simple way is to adopt the standard wild bootstrap method. To do so, let $\hat{\sigma}_{i}^{2} = T^{-1} \sum_{t=1}^{T} e_{it}^{2}$, where $e_{it} = X_{it} - \hat{\mu}_{it} \hat{\beta}_{it}$, and $\hat{\mu}_{it}$ and $\hat{\beta}_{it}$ are the estimates of the factors and factor loadings under the null. Let $\hat{e}_{it}^{\circ} = \hat{\mu}_{it} g_{it}$ with $g_{it}$ being IID $N(0, 1)$ over both $i$ and $t$. Then one can generate the bootstrap resamples via $X_{it}^{\circ} = \hat{\beta}_{it} \hat{e}_{it}^{\circ} + e_{it}^{\circ}$ and obtain the bootstrap test statistics and $p$-values as usual. One can justify the asymptotic validity of this method under very weak conditions despite the fact that the bootstrap error terms $\{e_{it}^{\circ}\}$ fail to capture the potential cross sectional dependence structure in the original error terms $\{e_{it}\}$. Preliminary simulations suggest this method works fairly well if either $\{e_{it}\}$ do not exhibit cross-sectional dependence or only exhibit fairly weak cross-sectional dependence. In the presence of moderate or strong cross sectional dependence in the error terms, tests based on this standard wild bootstrap method tend to be oversized.

For the above reason, we propose an alternative bootstrap procedure that tries to mimic the cross-sectional dependence in $\{e_{it}\}$. Let $\hat{\epsilon}_{it} = (\hat{\epsilon}_{it}, \ldots, \hat{\epsilon}_{it})'$ and $\Sigma^{0} = T^{-1} \sum_{t=1}^{T} \hat{\epsilon}_{it} \hat{\epsilon}_{it}'$. Let $\hat{\sigma}_{ij}^{\circ}$ denote the $(i, j)$th element of $\Sigma^{0}$. Define the shrinkage version of $\Sigma^{0}$ as $\Sigma^{\circ}$ whose $(i, j)$th element is given by $\hat{\sigma}_{ij}^{\circ} = \hat{\sigma}_{ij}^{0} (1 - e)^{|i-j|}$ for $i, j = 1, \ldots, N$, where $e$ is a small positive number (e.g., 0.01) to ensure the maximum absolute column/row sum norm of $\Sigma$ to be stochastically bounded provided $\max_{i} |\hat{\sigma}_{ij}^{0}|$ is. By construction, $\Sigma$ is also symmetric and positive semi-definite. The stochastic boundedness of $\max_{i} |\hat{\sigma}_{ij}^{0}|$ is sufficient but not necessary for the justifiability of the asymptotic validity of our bootstrap procedure below:

1. Estimate the restricted model $X_{it} = \lambda_{it} \beta_{it} + e_{it}^{\circ}$ by the PCA method and the unrestricted model $\hat{X}_{it} = \hat{\mu}_{it} \hat{\beta}_{it} + e_{it}$ by the local PCA method to obtain the two sets of estimates $\{\hat{\lambda}_{it}, \hat{\beta}_{it}\}$ and $\{\hat{\lambda}_{it}, \hat{\beta}_{it}\}$. Based on these estimates, construct the test statistic $J_{NT}$ as in Section 4.2.

2. For $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$, obtain the bootstrap error $e_{it}^{\circ} = \Sigma^{1/2} \hat{\epsilon}_{it}$, where $\hat{\epsilon}_{it} = (\hat{\epsilon}_{it}, \ldots, \hat{\epsilon}_{it})'$ with $\hat{\epsilon}_{it}$ being IID $N(0, 1)$ across $i$ and $t$. Generate $X_{it}^{\circ} = \hat{\mu}_{it} \hat{\beta}_{it} + e_{it}^{\circ}$.

3. Use $X_{it}^{\circ}$ to run the restricted and unrestricted models to obtain the bootstrap versions $\{\hat{\lambda}_{it}^{\circ}, \hat{\beta}_{it}^{\circ}\}$ and $\{\hat{\lambda}_{it}^{\circ}, \hat{\beta}_{it}^{\circ}\}$ of $\{\hat{\lambda}_{it}, \hat{\beta}_{it}\}$ and $\{\hat{\lambda}_{it}, \hat{\beta}_{it}\}$, respectively. Calculate the bootstrap test statistic $J_{NT}^{\circ}$, the bootstrap version of $J_{NT}$.
4. Repeat steps 2 and 3 for $B$ times and index the bootstrap test statistics as $(\hat{f}_{NT}^{\text{IB}})_{j=1}^{B}$. The bootstrap $p$-value is calculated by $p^* = B^{-1} \sum_{j=1}^{B} 1[\hat{f}_{NT}^{\text{IB}}>\hat{f}_{NT}]$.

The following theorem establishes the asymptotic validity of the above bootstrap method.

**Theorem 4.5.** Suppose that Assumptions A.1, A.3(i) and (ii*), A.4, and A.7 hold. Suppose that (i) $\max_{ij} |\tilde{a}_{ij}^0|$ = $O_p(\zeta_N)$ with $\zeta_N = O(T^{1/2})$. (ii) $\frac{1}{T} \sum_{t=1}^{T} f_t^8 = O_p(1)$ and (iii) $\frac{1}{N} \sum_{i=1}^{N} |\tilde{a}_{i0}|^8 = O_p(1)$. Then $\hat{f}_{NT}^{\text{IB}} \overset{D}{\to} N(0,1)$ in probability, where $D \to$ denotes weak convergence under the bootstrap probability measure conditional on the observed sample $X$.

**Theorem 4.5** shows that the bootstrap provides an asymptotic valid approximation to the limit null distribution of $\hat{f}_{NT}$. This holds because we generate the bootstrap data by imposing the null hypothesis. If the null hypothesis does not hold in the observed sample, then we expect $\hat{f}_{NT}$ to explode at the rate $T^{1/2}N^{1/4}h^{1/4}$, which delivers the consistency of the bootstrap-based test $\hat{f}_{NT}^{\text{IB}}$.

The extra conditions (i)-(iii) in the above theorem can be easily verified if the original data satisfies either the null hypothesis or the local alternative studied above. For example, in this case we can argue as usual in the proof of Lemma B.7(i) to demonstrate that $\frac{1}{T} \sum_{t=1}^{T} f_t^8 = O_p(1) + O_p(T^3(N^{-4} + T^{-4})) = O_p(1)$ and similarly $\frac{1}{N} \sum_{i=1}^{N} |\tilde{a}_{i0}|^8 = O_p(1)$ provided $T^3N^{-4} + T^3N^{-4} = O(1)$.

5. Monte Carlo study

In this section, we study the finite sample performance of our nonparametric estimates and tests through Monte Carlo simulations.

5.1. Data generating process

We generate data with $R = 2$ common factors:

$$X_{it} = \lambda_{i0} F_t + e_{it},$$ where $F_t = (F_{1t}, F_{2t})', F_{1t} = 0.6F_{1t-1} + u_{1t}, u_{1t}$ are IID $N(0, 1 - 0.6^2)$, $F_{2t} = 0.3F_{2t-1} + u_{2t}, u_{2t}$ are IID $N(0, 1 - 0.3^2)$ and independent of $u_{1t}$. We consider the following setups for the factor loadings $\lambda_{i0} = (\lambda_{i0,1}, \lambda_{i0,2})'$ and the error terms $e_{it}$:

**DGP 1:** (IID)

$$\lambda_{i0} \sim \text{IID} N(0, I_2), e_{it} \sim \text{IID} N(0, 1).$$

**DGP 2:** (Heteroskedasticity)

$$\lambda_{i0} \sim \text{IID} N(0, I_2), e_{it} = \sigma_i u_{it}, \sigma_i \sim \text{IID} U(0.5, 1.5) \text{ and } u_{it} \overset{\text{iid}}{\sim} N(0, 1).$$

**DGP 3:** (Cross sectional dependence)

$$\lambda_{i0} \sim \text{IID} N(0, I_2), e_{it} = (e_{it}, \ldots, e_{it})' \sim \text{IID} N(0, \Sigma_e), t = 1, 2, \ldots, T, \Sigma_e = (\sigma_{ij})_{j=1,...,N} \text{ with } \sigma_{ij} = 0.5^{i-j}.$$ 

**DGP 4:** (Single structural break)

$$\lambda_{i0,k} = \begin{cases} \lambda_{i0,k} + b, & \text{for } t = 1, 2, \ldots, T/2 \\ \lambda_{i0,k}, & \text{for } t = T/2 + 1, \ldots, T \end{cases}, \lambda_{i0,0} \sim \text{IID} N(1, 1), k = 1, 2; e_{it} = \sigma_i u_{it}, \sigma_i \sim \text{IID} U(0.5, 1.5), u_{it} \sim \text{IID} N(0, 1), \text{ and } b = 1, 2, 4.$$

**DGP 5:** (Multiple structural breaks)

$$\lambda_{i0,1} = \begin{cases} \lambda_{i0,1} + 0.5b, & \text{for } 0.67 < t \leq 0.87 \\ \lambda_{i0,1} - 0.5b, & \text{for } 0.27 < t \leq 0.47 \\ \lambda_{i0,1}, & \text{otherwise} \end{cases} \sim \text{IID} N(1, 1),$$

$$\lambda_{i0,2} = \lambda_{i0,2} \sim \text{IID} N(0, 1), e_{i0} \sim \text{IID} N(0, 1), \text{ and } b = 1, 2, 4.$$ 

**DGP 6:** (Smooth structural changes I)

$$\lambda_{i0,1} = \lambda_{i0,1} \sim \text{IID} N(0, 1), \lambda_{i0,2} = bG(10t/T; 2, 5/N + 2), e_{i0} \sim \text{IID} N(0, 1), \text{ where } b = 1, 2, 4.$$ 

**DGP 7:** (Smooth structural changes II)

$$\lambda_{i0,1} = \mu_i + bG(10t/T; 0.1, (2, 4, 6, 8)'), \mu_i \sim \text{IID} N(0, 1), \lambda_{i0,2} = \lambda_{i0,2} \sim \text{IID} N(0, 1), e_{i0} \sim \text{IID} N(0, 1).$$

**DGP 8:** (Smooth structural changes I + cross sectional dependence)

$$\lambda_{i0,1} = \lambda_{i0,1} \sim \text{IID} N(0, 1), \lambda_{i0,2} = bG(10t/T; 2, 5/N + 2), \text{ where } b = 1, 2, 4; e_{i0} \sim \text{IID} N(0, 1).$$

**DGP 9:** (Single structural break + cross sectional dependence)

$$\lambda_{i0,k} = \begin{cases} \lambda_{i0,k} + b, & \text{for } t = 1, 2, \ldots, T/2 \\ \lambda_{i0,k}, & \text{for } t = T/2 + 1, \ldots, T \end{cases} \sim \text{IID} N(1, 1), k = 1, 2; e_{i0} = (e_{i0}, \ldots, e_{i0})' \sim \text{IID} N(0, \Sigma_e), t = 1, 2, \ldots, T, \Sigma_e = (\sigma_{ij})_{j=1,...,N} \text{ with } \sigma_{ij} = 0.5^{i-j}.\text{ and } b = 1, 2, 4.$$ 

**DGP 10:** (Multiple structural breaks + cross sectional dependence)

$$\lambda_{i0,1} = \begin{cases} \lambda_{i0,1} + 0.5b, & \text{for } 0.67 < t \leq 0.87 \\ \lambda_{i0,1} - 0.5b, & \text{for } 0.27 < t \leq 0.47 \\ \lambda_{i0,1}, & \text{otherwise} \end{cases} \sim \text{IID} N(1, 1),$$

$$\lambda_{i0,2} = \lambda_{i0,2} \sim \text{IID} N(0, 1), e_{i0} = (e_{i0}, \ldots, e_{i0})' \sim \text{IID} N(0, \Sigma_e), t = 1, 2, \ldots, T, \Sigma_e = (\sigma_{ij})_{j=1,...,N} \text{ with } \sigma_{ij} = 0.5^{i-j}.\text{ and } b = 1, 2, 4.$$ 

Here, $G(z; \kappa, \gamma) = \{1 + \exp[-\kappa \sum (z - \gamma_j)]\}^{-1}$ denotes the Logistic function with tuning parameter $\kappa$ and location parameter $\gamma = (\gamma_1, \ldots, \gamma_j)$. DGP 1–3 satisfy the null hypothesis of time-invariant factor loadings, and are used to study the size of our test and the performance of our information criteria to determine the number of factors under the framework of time-invariant factor models. Note that we allow for cross sectional heteroskedasticity in DGP 2 and cross sectional dependence in DGP 3. DGP 4–10 describe various time-varying factor loadings. DGP 4 and 5 exhibit single and four sudden structural breaks, respectively. DGPS 6–7 exhibit smooth structural changes: the factor loadings generated in DGP 6 are monotonic functions while those in DGP 7 are smooth transition functions with multiple regime shifts. DGPS 8, 9, and 10 parallel DGPS 6, 4, and 5 but allow for cross sectional dependence.

5.2. Determination of the number of factors

In this subsection, we evaluate the information criteria to determine the number of common factors. In particular, we consider the following two information criteria:

$$IC_{11}(R) = \ln V(R, \{A(r)_{11}^T\}) + R \left(\frac{N + Th}{NTh} \right) \ln \left(\frac{NTh}{N + Th}\right),$$

$$IC_{12}(R) = \ln V(R, \{A(r)_{11}^T\}) + R \left(\frac{N + Th}{NTh} \right) \ln C_{NT},$$

$$C_{NT} = \min \left\{\sqrt{T}, \sqrt{N}\right\}.$$ For comparison purpose, we also consider Bai and Ng’s (2002) four information criteria (namely, $PC_{11}$, $PC_{12}$, $K_{21}$, and $I_{22}$), and Ahn and Horenstein’s (2013) two criterion functions (ER for eigenvalue ratio and GR for growth ratio). In addition, we implement Onatski’s
For each DGP, we simulate 1000 data sets with sample sizes $N, T = 100, 200$. Since the factor loadings are assumed to be nonrandom, we generate them once and fix them across the 1000 replications. We also redo the simulations by regenerating the factor loading randomly for each of the replications. The results are quite similar to those reported here and hence are omitted. Our local PCA involves nonparametric estimation. We use the Epanechnikov kernel and Silverman’s rule of thumb (RoT) to set the bandwidth as $h = (2.35/\sqrt{12})T^{-1/5}N^{-1/10}$ \(^3\). We also try the uniform kernel and the quartic kernel, and the RoT bandwidth with different tuning parameters. Our simulation studies show that the choices of kernel function and the bandwidth have little impact on the performance of our information criteria. Each series is demeaned and standardized to have unit variance.

We use two measures to evaluate the information criteria, i.e., the average number of common factors and the empirical probability of correct selection over 1000 replications. Bai and Ng (2002) apply the former measure. However, this measure can be misleading, e.g., the average number of common factors and the empirical impact on the performance of our information criteria. Each series of tuning parameters. Oursimulation studies show that the choices of kernel function and the bandwidth have little impact on the performance of our information criteria. Each series is demeaned and standardized to have unit variance.

We use two measures to evaluate the information criteria, i.e., the average number of common factors and the empirical probability of correct selection over 1000 replications. Bai and Ng (2002) apply the former measure. However, this measure can be misleading. For example, when the true number of factors is $R = 2$ but the information criteria select $R = 1$ or 3 with equal chance, the average number of selected factors can be still 2. Hence, we also report the empirical probability of correct selection to evaluate the information criteria comprehensively.

Tables 1 and 2 report the average number of common factors and the empirical probability of correct selection over 1000 replications of various information criteria in determining the number of common factors. DGPs 1–3 satisfy the null hypothesis of time-invariant factor loadings and allow us to compare the performance of these information criteria for the conventional factor models. DGPs 4–10 are the time-varying factor models with abrupt or smooth structural changes, where the value of $b$ indicates the magnitude of structural changes. To check the sensitivity of the information criteria to the magnitude of structural changes, we consider $b = 1, 2, 4$ for DGPs 4–10.

As shown in the tables, our information criteria work fairly well for all the DGPs under investigation. For the conventional factor models with IID, heteroskedastic, and cross-sectionally independent error terms in DGPs 1–3, respectively, the information criteria proposed by Bai and Ng (2002), Onatski (2009) and Ahn and Horenstein (2013) could select the true number of factors accurately. Our information criteria are slightly less accurate than the others when the sample size is small, but it is as good as them when the sample sizes are large (e.g., $(N, T) = (200, 200)$). The less accuracy of our information criteria can be attributed to the use of nonparametric estimation in our local PCA procedure. DGPs 4 and 5 are factor models with single and four abrupt structural breaks, respectively. We can see that all of Bai and Ng’s (2002) four information criteria have the tendency to choose 3 common factors, which is larger than the true number of factors (2 here). Onatski’s (2009) testing procedure also tends to choose 3 common factors except for the case of DGP 5 with $b = 1$, which is merely acceptable with larger than 70% correct selection probability. Ahn and Horenstein’s (2013) ER and GR criterion functions perform well for the case of DGP 5 with $b = 1$, but they still suffer from severe over- or under-selection for other cases. In contrast, although our information criteria are proposed for smooth structural changes, they still work well for small and moderate magnitude ($b = 1, 2$) of abrupt structural breaks. Although they tend to choose factors slightly more than necessary for $b = 4$, the results are still acceptable and much better than those of other information criteria. DGPs 6–7 are factor models with smooth structural changes in factor loadings and cross-sectionally independent errors. As shown in Tables 1 and 2, our information criteria give precise estimates of the number of common factors for all cases. However, the criteria proposed by Bai and Ng (2002), Onatski (2009) and Ahn and Horenstein (2013) work poorly except for the case of small structural changes ($b = 1$). Similarly, when the error terms are cross-sectionally dependent in DGPs 8–10, we can also see our information criteria outperform the others in Table 2.

5.3. Performance of the test

In this subsection, we study the finite sample performance of our test for time-varying factor loadings. We also compare our test with the tests of Breitung and Eickmeier (2011), Chen et al. (2014), Han and Inoue (2015) and Yamamoto and Tanaka (2015) for a single structural break with an unknown break date in factor loadings.

Since our bootstrap testing procedure is rather time consuming, we generate 500 data sets in this subsection and set the bootstrap replication number $B$ to 200. As in the previous subsection, we use the Epanechnikov kernel and the RoT bandwidth $h = (2.35/\sqrt{12})T^{-1/5}N^{-1/10}$. In addition to our test, we also consider Breitung and Eickmeier’s (2011) sup-LM variable-specific test, Chen et al.’s (2014) sup-LM and sup-Wald tests, Han and Inoue’s (2015) sup-LM and sup-Wald tests, and Yamamoto and Tanaka’s (2015) sup-Wald test. We follow these papers to set the trimming parameter $\tau = 0.15$, which implies that the single break can only occur on the time interval $[0.15T, 0.85T]$. The tests of Chen et al. (2014) and Han and Inoue (2015) involve the long run variance estimation. We set the time-lag truncation parameter as $m = [T^{1/5}]$ and choose the Bartlett kernel. The critical values presented in Andrews (1993) are applied for the tests of Breitung and Eickmeier (2011), Chen et al. (2014) and Han and Inoue (2015), while the bootstrap critical values are applied to check the performance of our test.

In the literature, the number of common factors is either determined by Bai and Ng’s (2002) information criteria (e.g., Han and Inoue, 2015) or specified by some fixed numbers, which may be equal to, less than, or greater than the correct number of factors (e.g., Chen et al., 2014). In this paper, we apply some information criteria to determine the number of factors used for our simulations. Specifically, we apply $IC_{B1}$ and $IC_{B2}$ given in Section 5.2 to determine the number of factors used for our tests. For the tests of Breitung and Eickmeier (2011), Chen et al. (2014), Han and Inoue (2015), and Yamamoto and Tanaka (2015), the number of factors is determined by Bai and Ng’s (2002) information criteria $IC_{A1}$ and $IC_{A2}$: the results based on these two information criteria are quite similar and we only report the results using $IC_{B1}$ to save space. In addition, we also implement our tests and those of Breitung and Eickmeier (2011), Chen et al. (2014), and Han and Inoue (2015) by specifying the number of factors as the correct number of common factors. The results are reported in the online supplementary material (see Appendix A).

Table 3 reports the empirical sizes of various tests at both 5% and 10% significance levels. As shown in the table, our tests based on $IC_{B1}$ and $IC_{B2}$ have reasonable size. Han and Inoue’s (2015) sup-LM test delivers reasonable size but their sup-Wald test tends to be conservative.

\(^3\) Note that $[t/T]_{t \in T}$ behaves like a uniform random variable on $[0, 1]$ and thus has variance $1/12$. 

\(^4\) Yamamoto and Tanaka (2015) argue that when the number of factors is estimated from the data, Breitung and Eickmeier’s (2011) test tends to exhibit non-monotonic power due to the presence of estimated spurious factors under the alternative. So they propose a test that is robust to this non-monotonic power problem. But their test reduces to Breitung and Eickmeier’s (2011) test when we fix the number of factors at the true number.
under-reject the null hypothesis. Chen et al.’s (2014) sup-LM test has reasonable size, but their sup-Wald test tends to over-reject the null hypothesis in DGP 2. Breitung and Eickmeier’s (2011) variable-specific sup-LM test suffers from slight under-rejection for DGPs 1–3. The size of Yamamoto and Tanaka’s (2015) test is generally well behaved.

Tables 4 and 5 report the empirical powers of various tests for DGPs 4–10 at the 5% and 10% significance levels when the number of factors are determined from the data. To save space, we only report the results for \( b = 1 \) and 2. We summarize some important findings. First, our \( J_{NT} \) test is powerful in detecting all the forms of time-varying factor loadings given by DGPs 4–10 and the simulation results are consistent with our theoretical conclusion that our test is able to detect both a finite number of sudden structural breaks and smooth structural changes. Second, Breitung and Eickmeier’s (2011) test is powerless for DGPs 4–10, which is consistent with Yamamoto and Tanaka’s (2015) argument that when the number of factors is estimated from the data, Breitung and Eickmeier’s (2011) test tends to exhibit non-monotonic power. Third, the tests of Chen et al. (2014), Han and Inoue (2015) and Yamamoto and Tanaka (2015) are all designed to test for a one-time abrupt structural change in DGPs 4 and 9. As expected, they all have power against DGPs 4 and 9. Fourth, for the other DGPs, all of Han and Inoue’s (2015) sup-LM and sup-Wald tests, Chen et al.’s (2014) sup-LM and sup-Wald tests, and Yamamoto and Tanaka’s (2015) test have lower power than our test too. In particular, the
tests of Chen et al. (2014) and Han and Inoue (2015) have very low power in detecting deviations from the null in DGPs 7, 10, especially when the sample size is small. But these tests have reasonable power against DGPs 6 and 8. It is easy to explain why some of these other tests have power against DGPs 6 and 8. Note that in these two DGPs, the factor loadings are monotonic functions of the time ratio \( t/T \) for each \( i \). If we apply the PCA method to estimate the factor model, the estimated factors would exhibit a trend with increasing volatilities. Since Han and Inoue’s (2015) test checks the time invariance property of the second order moments of the common factors, it is possible to capture such smooth structural changes in DGPs 6 and 8. Similarly, Chen et al.’s (2014) test is based on the regression of one of the estimated factors on the remaining estimated factors, and their LM and Wald test statistics will not have the usual asymptotic distribution when one estimated factor exhibits trending behavior.

### Table 2
Comparison of various information criteria in determining the number of factors: DGPs 7–10.

<table>
<thead>
<tr>
<th>DGP</th>
<th>((N, T))</th>
<th>Average number of factors</th>
<th>Empirical probability of correct selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b = 1)</td>
<td>(100,100)</td>
<td>2.01 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
<tr>
<td></td>
<td>(100,200)</td>
<td>2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
<tr>
<td></td>
<td>(200,100)</td>
<td>2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
<tr>
<td></td>
<td>(200,200)</td>
<td>2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
<tr>
<td></td>
<td>(100,100)</td>
<td>2.01 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
<tr>
<td></td>
<td>(100,200)</td>
<td>2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
<tr>
<td></td>
<td>(200,100)</td>
<td>2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
<tr>
<td></td>
<td>(200,200)</td>
<td>2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
<tr>
<td>(b = 2)</td>
<td>(100,100)</td>
<td>2.01 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
<tr>
<td></td>
<td>(100,200)</td>
<td>2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
<tr>
<td></td>
<td>(200,100)</td>
<td>2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
<tr>
<td></td>
<td>(200,200)</td>
<td>2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
<tr>
<td>(b = 3)</td>
<td>(100,100)</td>
<td>2.01 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
<tr>
<td></td>
<td>(100,200)</td>
<td>2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
<tr>
<td></td>
<td>(200,100)</td>
<td>2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
<tr>
<td></td>
<td>(200,200)</td>
<td>2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
<tr>
<td>(b = 4)</td>
<td>(100,100)</td>
<td>2.01 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
<tr>
<td></td>
<td>(100,200)</td>
<td>2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
<tr>
<td></td>
<td>(200,100)</td>
<td>2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
<tr>
<td></td>
<td>(200,200)</td>
<td>2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2.00</td>
<td><strong>Ona ER GR</strong></td>
</tr>
</tbody>
</table>

Note: See the note in Table 1.

### 6. An application to Stock and Watson’s (2009) U.S. macroeconomic data set

In this section, we apply our approach to check whether the U.S. economy suffers from structural changes. The data set, constructed by Stock and Watson (2009), consists of 144 quarterly time series, spanning 1959Q1-2006Q4.5 By excluding the first two quarters, which is missing when computing the first and second differences, we get a total of \( T = 190 \) quarterly observations. Also, we follow the suggestion of Stock and Watson (2009) to delete some high level aggregates related by identities to the lower level sub-aggregates and end up with \( N = 109 \) time series. For

---

5 The dataset is publicly available on Professor Mark W. Watson’s website http://www.princeton.edu/~mwatson/publi.html.
some time series that are available monthly, we take averages over the quarter to get the corresponding quarterly data. Following the literature, we transform the data by taking the first or second order (log-)difference and removing outliers. All the data have been standardized to have zero mean and unit variance. For the order (log-)difference and removing outliers. All the data have

Note: (i) \( J_{h1} \) and \( J_{h2} \) denote the results of our test \( J_{h} \) based on the bootstrap p-values by determining the number of common factors using our information criteria \( IC_{h1} \) and \( IC_{h2} \) proposed in Section 5.2; (ii) \( H_{ILM} \) and \( H_{IW} \) denote Han and Inoue's (2015) sup-LM and sup-Wald tests by determining the number of common factors using the information criteria \( IC_{p} \) proposed by Bai and Ng (2002); (iii) \( CDG_{p} \) denote Chen et al.'s (2014) sup-LM and sup-Wald tests by determining the number of common factors using the information criteria \( IC_{p} \) proposed by Bai and Ng (2002); (iv) \( BE_{LM} \) denotes Breitung and Eickmeier's (2011) variable-specific sup-LM test by determining the number of common factors using the information criteria \( IC_{p} \) proposed by Bai and Ng (2002); (v) \( YT \) denotes Yamamoto and Tanaka's (2015) test by determining the number of common factors using the information criteria \( IC_{p} \) proposed by Bai and Ng (2002). The main entries report the average percentage of rejection.

Table 4

<table>
<thead>
<tr>
<th>DGP</th>
<th>N</th>
<th>T</th>
<th>( J_{h1} )</th>
<th>( J_{h2} )</th>
<th>( H_{ILM} )</th>
<th>( H_{IW} )</th>
<th>( CDG_{LM} )</th>
<th>( CDG_{GW} )</th>
<th>( BE_{LM} )</th>
<th>YT</th>
</tr>
</thead>
<tbody>
<tr>
<td>4, b = 1</td>
<td>100</td>
<td>100</td>
<td>99.8</td>
<td>100</td>
<td>100</td>
<td>73.4</td>
<td>94.2</td>
<td>64.4</td>
<td>90.4</td>
<td>99.2</td>
</tr>
<tr>
<td>5, b = 1</td>
<td>100</td>
<td>100</td>
<td>94.2</td>
<td>96.6</td>
<td>96.6</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Note: See the note in Table 3.

Table 6

<table>
<thead>
<tr>
<th>DGP</th>
<th>N</th>
<th>T</th>
<th>( J_{h1} )</th>
<th>( J_{h2} )</th>
<th>( H_{ILM} )</th>
<th>( H_{IW} )</th>
<th>( CDG_{LM} )</th>
<th>( CDG_{GW} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>100</td>
<td>6.4</td>
<td>12.8</td>
<td>6.0</td>
<td>11.6</td>
<td>3.6</td>
<td>7.2</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>200</td>
<td>5.0</td>
<td>8.8</td>
<td>5.4</td>
<td>10.2</td>
<td>6.0</td>
<td>11.6</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>200</td>
<td>6.8</td>
<td>13.4</td>
<td>7.6</td>
<td>12.8</td>
<td>3.2</td>
<td>6.8</td>
</tr>
</tbody>
</table>

Note: We first determine the appropriate number of common factors. The maximum number of common factors is set to be 8 in this empirical study. Other presettings such as the kernel and bandwidth are the same as in the simulation section. We use Bai and Ng's (2002) information criteria \( PC_{p1}, PC_{p2}, IC_{p1}, \) and \( IC_{p2}, \) Onatski's (2009) testing procedure, Ahn and Horenstein's (2013) criterion functions \( ER \) and \( GR \) and our information criterion \( IC_{h1} \) and \( IC_{h2} \) proposed in Section 3.3 to determine the number of common factors. The results are reported in Table 6. According to the table, we report the test results for the cases of one to five common factors respectively in the following context.
mixed, and they can only reject the null for $R$ at the 5% significance level, while Chen et al.'s (2014) results are sup-LM and sup-Wald tests cannot reject the null for any case at the 5% significance level, while Chen et al.'s (2014) results are mixed, and they can only reject the null for $R = 5$ at the 5% significance level when using the sup-Wald test. This is consistent with the results of our simulation study that suggests that the tests of Han and Inoue (2015) and Chen et al. (2014) have relatively low power when structural changes are not a one-time break. In addition, Breitung and Eickmeier’s (2011, BE) variable-specific sup-LM test rejects the null of time-invariant factor loadings for about half of the variables.

Our empirical result suggests the existence of possible smooth or sudden structural changes in U.S. economy. We now estimate the common factors and the time-varying factor loadings by using our local principal component approach proposed in Section 2 by assuming 3 common factors. Fig. 1 plots the estimated time-varying factor loadings and their 90% confidence bands for real personal consumption expenditures (left panel) and industrial production index of durable goods (right panel) corresponding to the three common factors selected by our information criteria. From this figure, we can see that the estimated factor loadings show significant time-varying features. The finding of time-varying factor loadings has some important implications. For example, most of the existing studies estimate the common factors under the framework of time-invariant factor loadings and then forecast some key variables based on the estimated common factors. We may provide more reliable forecasts by accommodating the documented time-varying features of factor loadings by using a local version of the principal component method.

### 7. Conclusion

Conventional factor models assume that factor loadings are fixed over a long horizon of time, which appears restrictive and unrealistic in empirical applications. In this paper, we introduce a time-varying factor model where factor loadings are allowed to change smoothly over time and propose a local version of the PCA method to estimate the latent factors and time-varying factor loadings simultaneously. We establish the limiting distributions of the estimated factors and factor loadings in the standard large $N$ and large $T$ framework. We also propose a BIC-type information

---

**Table 5**

<table>
<thead>
<tr>
<th>DGP</th>
<th>$N$</th>
<th>$T$</th>
<th>$J_{B1}$</th>
<th>$J_{B2}$</th>
<th>$H_{LM}$</th>
<th>$H_{W}$</th>
<th>$CDG_{LM}$</th>
<th>$CDG_{W}$</th>
<th>$BE_{LM}$</th>
<th>$YT$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td>8, $b = 1$</td>
<td>100</td>
<td>100</td>
<td>92.6</td>
<td>96.0</td>
<td>93.2</td>
<td>95.6</td>
<td>81.4</td>
<td>95.0</td>
<td>68.6</td>
<td>88.8</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99.8</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99.6</td>
<td>99.8</td>
<td>98.0</td>
<td>92.8</td>
</tr>
<tr>
<td>8, $b = 2$</td>
<td>100</td>
<td>100</td>
<td>99.6</td>
<td>99.8</td>
<td>99.2</td>
<td>99.8</td>
<td>98.6</td>
<td>94.0</td>
<td>24.4</td>
<td>61.2</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>76.4</td>
<td>94.6</td>
<td>26.2</td>
<td>67.8</td>
</tr>
<tr>
<td>9, $b = 1$</td>
<td>100</td>
<td>100</td>
<td>99.8</td>
<td>99.8</td>
<td>99.8</td>
<td>99.8</td>
<td>98.0</td>
<td>100</td>
<td>98.0</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>79.0</td>
<td>96.6</td>
<td>68.8</td>
<td>94.6</td>
</tr>
<tr>
<td>9, $b = 2$</td>
<td>100</td>
<td>100</td>
<td>98.4</td>
<td>99.2</td>
<td>98.8</td>
<td>99.2</td>
<td>77.8</td>
<td>96.2</td>
<td>71.6</td>
<td>95.2</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>98.6</td>
<td>98.2</td>
<td>75.2</td>
<td>95.6</td>
</tr>
<tr>
<td>10, $b = 1$</td>
<td>100</td>
<td>100</td>
<td>93.8</td>
<td>96.4</td>
<td>94.0</td>
<td>96.6</td>
<td>3.6</td>
<td>9.2</td>
<td>0.4</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>27.0</td>
<td>42.8</td>
<td>16.6</td>
<td>33.2</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>48.6</td>
<td>74.2</td>
<td>34.8</td>
<td>56.8</td>
</tr>
<tr>
<td>10, $b = 2$</td>
<td>100</td>
<td>100</td>
<td>99.4</td>
<td>99.4</td>
<td>99.8</td>
<td>99.8</td>
<td>1.6</td>
<td>6.2</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>62.0</td>
<td>85.8</td>
<td>43.0</td>
<td>72.2</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>63.4</td>
<td>90.8</td>
<td>45.8</td>
<td>75.2</td>
</tr>
</tbody>
</table>

Note: See the note in Table 3.

**Table 6**

Tests of structural changes in the U.S. economy.

<table>
<thead>
<tr>
<th>Number of selected factors</th>
<th>1</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion functions</td>
<td>On, ER, GR</td>
<td>$IC_{B1, B2}$, $PC_{p1, p2}$</td>
<td>$IC_{p1}$</td>
<td>$IC_{p2}$</td>
</tr>
<tr>
<td>Note: See the note in Table 1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 7**

Tests of structural changes in the U.S. economy.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$R = 1$</td>
<td>5.40</td>
<td>2.82</td>
<td>2.12</td>
</tr>
<tr>
<td>$R = 2$</td>
<td>23.90</td>
<td>10.94</td>
<td>9.84</td>
</tr>
<tr>
<td>$R = 3$</td>
<td>31.48</td>
<td>16.35</td>
<td>15.30</td>
</tr>
<tr>
<td>$R = 4$</td>
<td>30.44</td>
<td>23.14</td>
<td>24.31</td>
</tr>
<tr>
<td>$R = 5$</td>
<td>35.50</td>
<td>26.20</td>
<td>25.65</td>
</tr>
</tbody>
</table>

Note: (i) Under $IC_{B1}$ and sup-LM and sup-Wald are the values of the corresponding test statistics; (ii) Under 5% and 10% are the corresponding bootstrap critical values (our test, 500 bootstrap resamples) or asymptotic critical values (Han and Inoue’s and Chen et al.’s tests) except for the Breitung and Eickmeier’s (2011) test; (iii) Under 5% and 10% of BE (2011) are the empirical rejection frequencies of Breitung and Eickmeier’s (2011) variable-specific sup-LM test by using 5% and 10% asymptotic critical values respectively. Bold elements denote significance at the 5% nominal level.
criterion to determine the number of common factors for time-varying factor models. Our information criterion works no matter whether the factor loadings are time-invariant or time-varying and it is extremely useful when structural changes are suspected.

More importantly, we propose an $L_2$-distance-based test statistic to check the stability of factor loadings. By construction, our test can capture both smooth and abrupt structural changes in factor loadings and one does not need to know the number of breaks in the data. Monte Carlo studies demonstrate the excellent performance of the BIC-type information criterion in determining the number of common factors, and the reasonable size and excellent power of our test in checking the time-invariance of factor loadings. In an application to Stock and Watson’s (2009) U.S. macroeconomic data set, we find significant evidence against the time-invariant factor loadings imposed by the conventional factor models.

There are several interesting topics for further research. First, as a referee points out, one potential application of our time-varying factor model is out-of-sample forecasting based on augmented factor models. In the presence of time-varying factor loadings, we could apply our local PCA method to estimate the factors. Despite the fact that such factor estimates are only consistent with the true factors up to an invertible rotation matrix which is also $t$-dependent, the estimated factors can be used in conjunction with a functional coefficient forecasting model where the coefficients of the factors are specified as a function of the scaled time $t$. Our preliminary simulation results suggest that this can help to improve the performance of out-of-sample forecasts. Second, when our test rejects the null, one can estimate the number of breaks and break points if one believes that the changes are abrupt. The estimation of factor models with multiple breaks is certainly an interesting and challenging issue. We are exploring these topics in ongoing work.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jeconom.2016.12.004.

References

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 90, 191–221.
Moon, H., Weidner, M., 2015. Linear regression for panel with unknown number of factors as interactive fixed effects. Econometrica 83, 1543–1579.


