On Time-Varying Factor Models: Estimation and Testing

Liangjun Su\textsuperscript{a} and Xia Wang\textsuperscript{b}

\textsuperscript{a} School of Economics, Singapore Management University, Singapore
\textsuperscript{b} School of Management, University of Chinese Academy of Sciences, Beijing

April 17, 2015

Abstract

Conventional 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 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.

JEL Classification: C12, C14, C33, C38.

Key Words: Factor model, Information criterion, Local principal component, Local smoothing, Structural change, Test, Time-varying parameter.
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 random 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 estimated factors 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.

In recent years, more and more research has focused on structural changes in factor loadings. Stock and Watson (2008) examine the forecasting reliability when there exists a structural break in the factor loadings. Breitung and Eickmeier (2011) propose three statistics (namely, LR, LM and Wald statistics) to test for structural breaks in factor loadings. Chen et al. (2014) propose a two-stage procedure to detect big breaks in factor loadings by testing the parameter stability in a 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 (2014) propose a joint test for structural breaks in factor loadings based on the second moments of the estimated factors. Cheng et al. (2014) consider the case that 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 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 $L_2$-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. By 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 fixed unknown number in our setup. Unlike the existing tests, such as Breitung and Eickmeier (2011), Chen et al. (2014) and Han and Inoue (2014), 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 separately identifiable. Moreover, unlike the existing tests for unknown break date, namely the supremum-type tests of Breitung and Eickmeier (2011), Chen et al. (2014), and Han and Inoue (2014), 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 end of the sample under the alternative.

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 are relegated to the appendix. Further technical details are contained on the online supplementary appendix.

NOTATION. For an $m \times n$ real matrix $A$, we denote its transpose as $A'$, its Frobenius norm as $\|A\|$.
(≡ \[\text{tr}(AA')^{1/2}\]), its spectral norm as \(\|A\|_{sp} \equiv \sqrt{\mu_1(A^TA)}\) and its Moore-Penrose generalized inverse as \(A^+\), where \(\equiv\) means “is defined as” and \(\mu_s(\cdot)\) denotes the \(s\)th largest eigenvalue of a real symmetric matrix by counting eigenvalues of multiplicity multiple times. Note that the two norms are equal when \(A\) is a vector. We will frequently use the submultiplicative property of these norms and the fact that \(\|A\|_{sp} \leq \|A\| \leq \|A\|_{sp}\text{rank}(A)^{1/2}\). We also use \(\mu_{\text{max}}(B)\) and \(\mu_{\text{min}}(B)\) to denote the largest and smallest eigenvalues of a symmetric matrix \(B\), respectively. We use \(B > 0\) to denote that \(B\) is positive definite.

Let \(P_A \equiv A (A'A)^+ A'\) and \(M_A \equiv I_m - P_A\), where \(I_m\) denotes an \(m \times m\) identity matrix. The operator \(\rightarrow\) denotes convergence in probability, \(\Rightarrow\) convergence in distribution, and plim probability limit. We use \((N, T) \rightarrow \infty\) to denote that \(N\) and \(T\) pass to infinity jointly.

## 2 Factor Model with Time-varying Factor Loadings

Let \(\{X_{it}, i = 1, 2, \ldots, N; t = 1, 2, \ldots, T\}\) be an \(N\)-dimensional time series with \(T\) observations. The index \(i\) represents the \(i\)th cross section unit in panel data set or the \(i\)th random variable in multiple time series data set. We assume that \(X_{it}\) admits the following time-varying factor model with \(R\) latent common factors \(F_t = [F_{1t}, \ldots, F_{Rt}]'\):

\[
X_{it} = \lambda_{it}'F_t + e_{it}, \tag{2.1}
\]

where the idiosyncratic errors \(\{e_{it}\}\) is assumed to be weakly dependent over both cross sectional unit \(i\) and time period \(t\). Furthermore, \(F_t\) satisfies \(E[F_tF_t'] = \Sigma_F\) for some positive definite covariance matrix \(\Sigma_F\).

Our model given by (2.1) generalizes Stock and Watson’s (2002) and Bai’s (2003) dynamic factor models by allowing for structural changes in factor loadings.\(^1\) Specifically, we consider smooth structural changes. This is in contrast to the factor models with structural breaks that have recently been studied in the literature; see, e.g., Breitung and Eickmeier (2011), Chen et al. (2014), and Han and Inoue (2014). Because the driving forces of structural changes including preference changes, technological progress, policy changes usually accrue gradually over a long period of time, it seems more realistic to assume smooth structural changes rather than abrupt changes in reality. More importantly, the factor model with abrupt changes could be regarded as the time-invariant factor model with more latent factors. By using more factors, one can approximate the true model well and yield reasonable economic analysis and forecasting (see Breitung and Eickmeier 2011, Chen et al. 2014). However, this is not the case for factor models with smooth structural changes. In our model with time-varying factor loading, the conventional PCA will result in inconsistent estimators and forecasts even if we use more factors.

To avoid model misspecification and to allow our model to capture various kinds of time-varying factor loadings, we use a nonparametric local smoothing method to estimate \(\lambda_{it}\). Specifically, we follow the

---

\(^1\)Stock and Watson (2002) also consider a time-varying factor model with a stochastic drift in the factor loadings: \(\lambda_{it} = \lambda_{i,t-1} + g_{iT}\zeta_{it}\). They assume that \(g_{iT}\) is a scalar and small with \(g_{iT} = O_p(T^{-1})\) and \(\zeta_{it}\) is an \(R \times 1\) random vector such that \(\lambda_{it} - \lambda_{i0} = O_p(T^{-1/2})\), and demonstrate that such small instability does not affect the consistency of the estimated factors. Del Negro and Otkok (2008) 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.
nonparametric literature on time-varying models (see, e.g., Cai 2007, Robinson 2012, Chen et al. 2012, Su et al. 2015) and model \( \lambda_{it} \) as a nonrandom function of \( t/T \):

\[
\lambda_{it} = \lambda_i(t/T).
\]

where \( \lambda_i(\cdot) \) is an unknown piece-wise smooth function of \( t/T \) on \((0, 1]\) for each \( i \). The specification that \( \lambda_i(t/T) \) 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 \( \lambda_i(t/T) \) for each \( i \) at some fixed point \( t/T \) by increasing the amount of data on which it depends. For more discussion, see Robinson (1989, 1991).

3 Estimation

In this section, we introduce the local version of PCA to estimate the time-varying factor loadings \( \lambda_{it} \) and the factors \( F_t \). We also establish the asymptotic distributions of these estimators and propose a BIC-type information criterion to determine the number of factors.

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

\[
\lambda_{it} = \lambda_i \left( \frac{t}{T} \right) \approx \lambda_i \left( \frac{r}{T} \right) = \lambda_{ir} \text{ when } \frac{t}{T} \approx \frac{r}{T}.
\]

It follows that

\[
X_{it} \approx \lambda_{ir} F_t + e_{it} \text{ when } \frac{t}{T} \approx \frac{r}{T}.
\]

To estimate \( \{\lambda_{ir}\}_{i=1}^N \) and \( \{F_t\}_{t=1}^T \), we can consider the following local weighted least squares (WLS) problem:

\[
\min_{\{\lambda_{ir}\}_{i=1}^N, \{F_t\}_{t=1}^T} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T [X_{it} - \lambda_{ir} F_t]^2 K_h \left( \frac{t-r}{T} \right),
\]

where \( K_h(x) = h^{-1} K(x/h) \) with kernel \( K : \mathbb{R} \to \mathbb{R}^+ \) and bandwidth \( h = h(T, N) \). As we now argue, the solution to the above minimization problem can be obtained via PCA.

To see this, multiplying both sides of (3.1) by \( k_{h, tr} = h^{-1} K((t-r)/(Th)) \) yields the transformed model

\[
k_{h, tr} X_{it} \approx k_{h, tr} \lambda_{ir} F_t + k_{h, tr} e_{it}.
\]

Define the \( T \times N \) matrices \( X^{(r)} = (X_1^{(r)}, \ldots, X_N^{(r)}) \) and \( e^{(r)} = (e_1^{(r)}, \ldots, e_N^{(r)}) \), where \( X_i^{(r)} = (k_{h, tr} X_{i1}, \ldots, k_{h, tr} X_{iT})' \) and \( e_i^{(r)} = (k_{h, tr} e_{i1}, \ldots, k_{h, tr} e_{iT})' \). Let \( F^{(r)} = (k_{h, tr} F_1, \ldots, k_{h, tr} F_T)' \) and \( \Lambda^{(r)} = (\lambda_{ir}, \ldots, \lambda_{Nr})' \), which are \( T \times R \) and \( N \times R \) matrices, respectively. In matrix notation, the transformed model (3.3) can be written as

\[
X^{(r)} \approx F^{(r)} \Lambda^{(r)}' + e^{(r)}.
\]
Then the minimization problem in (3.2) becomes

$$
\min_{F^{(r)}, \Lambda^{(r)}} \text{tr} \left[ \left( X^{(r)} - F^{(r)} \Lambda^{(r)} \right) \left( X^{(r)} - F^{(r)} \Lambda^{(r)} \right)^\top \right]
$$

By concentrating out $\Lambda^{(r)} = X^{(r)'(r)' F^{(r)} F^{(r)}_T} = X^{(r)' F^{(r)}/T}$ under the restriction $F^{(r)' F^{(r)}/T} = I_R$, the objective function becomes

$$
\text{tr} \left[ X^{(r)' X^{(r)}} \right] - T^{-1} \text{tr} \left[ F^{(r)' X^{(r)} F^{(r)}} \right].
$$

Then we can consider maximizing $\text{tr}[F^{(r)' X^{(r)} F^{(r)}]}$ subject to $F^{(r)' F^{(r)}/T} = I_R$. This is the conventional principal component problem. The estimated factor matrix, denoted by $\hat{F}^{(r)} = (\hat{F}_1^{(r)}, ..., \hat{F}_T^{(r)})'$, is $\sqrt{T}$ times eigenvectors corresponding to the $R$ largest eigenvalues of the $T \times T$ matrix $X^{(r)' X^{(r)}/T} = \sum_{i=1}^N X_i^{(r)' X_i^{(r)'}/T}$, and $\hat{\Lambda}^{(r)} = (\hat{F}^{(r)} \hat{F}^{(r)')-1} \hat{F}^{(r)' X^{(r)}/T}$, $r = 1, 2, ..., T$, are the estimators of the corresponding time-varying factor loadings. We use $\hat{\lambda}_{ir}$ to denote the $i$th column of $\hat{\Lambda}^{(r)}$. 

It is well known that a local constant estimator may suffer from boundary problem. When the kernel function $K(\cdot)$ has compact support $[-1, 1]$, the boundary regions for our local WLS problem are given by $[0, h] \cup [1 - h, 1]$. Even though the length of these regions is shrinking to zero as $h \to 0$, there are still a significant amount of data falling into these regions in finite samples. To avoid the boundary problem, we apply the following boundary kernel (see, Hong and Li 2005, Li and Racine 2006, p.31):

$$
k_{h, tr}^* = h^{-1}K^*_r \left( \frac{t-r}{Th} \right) = \begin{cases} h^{-1}K \left( \frac{t-r}{Th} \right) / \int_{-r/Th}^{1} K(u)du, & \text{if } r \in [0, [Th]] \\ h^{-1}K \left( \frac{t-r}{Th} \right), & \text{if } r \in [[Th], T - [Th]] \\ h^{-1}K \left( \frac{t-r}{Th} \right) / \int_{1}^{(1-r/T)/h} K(u)du, & \text{if } r \in (T - [Th], T] \end{cases}
$$

where $[a]$ denote the integer part of $a$. 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 $k_{h, tr} = K_h \left( \frac{t-r}{Th} \right)$ 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 $F_t^{(r)} \equiv k_{h, tr}^{1/2} F_t$. To obtain a consistent estimator of $F_t$ after suitable rotation, we consider a two-stage estimation procedure. Based on the consistent estimates of $\lambda_{it}$’s obtained in the first stage, we can obtain the consistent estimators of $F_t$, $t = 1, 2, ..., T$, in the second stage, by minimizing the following least squares problem:

$$
\min_{F_t \in \mathbb{R}^h} \sum_{i=1}^N \left[ X_{it} - \hat{\lambda}_{it}^t F_t \right]^2, \quad t = 1, 2, ..., T.
$$

The solution to the above problem is: $\hat{F}_t = \left( \sum_{i=1}^N \hat{\lambda}_{it} \hat{\lambda}_{it}^t \right)^{-1} \left( \sum_{i=1}^N \hat{\lambda}_{it} X_{it} \right)$ for $t = 1, 2, ..., T$.

### 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.
Let $\Lambda_r = (\lambda_{1r}, ..., \lambda_{Nr})^\prime$ for $r = 1, ..., T$. Let $\gamma_N(s, t) = N^{-1}E(e'_s e_t)$, $\gamma_{N,F}(s, t) = N^{-1}E(F_s e'_t e_t)$, $\gamma_{N,FF}(s, t) = N^{-1}E(F_s e'_t F'_t)$, and $\zeta_{st} = N^{-1}[e'_s e_t - E(e'_s e_t)]$. Define

$$\varpi_{NT,1}(r) = \frac{\sqrt{h}}{\sqrt{NT}} F^{(r)} e^{(r)} \Lambda_r = \frac{\sqrt{h}}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} k_{ht,ir} F_i e_{it} \lambda'_i,$$

$$\varpi_{NT,2}(r, t) = \frac{\sqrt{h}}{\sqrt{NT}} [F^{(r)} e^{(r)} e_t - E(F^{(r)} e^{(r)} e_t)] = \frac{\sqrt{h}}{\sqrt{NT}} \sum_{s=1}^{N} \sum_{i=1}^{T} k_{hs,ir} [F_s e_{is} e_{it} - E(F_s e_{is} e_{it})].$$

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

**Assumption A.1.**

(i) $E(e_{it}) = 0$ and $\max_{i,t} E(e_{it}^2) < \infty$.

(ii) $\max_{t} E(F_t) = \Sigma_F > 0$ for some $R \times R$ matrix $\Sigma_F$.

(iii) $\lambda_{it}$ are nonrandom such that $\max_{i,t} \| \lambda_{it} \| \leq \overline{\lambda} < \infty$ and $N^{-1} \Lambda_r \Lambda_r = \Sigma_{\Lambda_r} + O(N^{-1/2})$ for some $R \times R$ positive definite matrix $\Sigma_{\Lambda_r}$ and for all $r$.

(iv) $\max_{t} \sum_{i=1}^{T} \| \text{Cov}(F_{t,m} F_{t,n}, F_{t,n} F_{s,n}) \| \leq C$ for $m, n = 1, ..., R$, where $F_{t,m}$ denotes the $m$th element of $F_t$.

(v) $\max_{t} \sum_{i=1}^{T} \| \gamma(s, t) \| \leq C$ and $\max_{t} \sum_{i=1}^{T} \| \gamma(s, t) \| \leq C$ for $\gamma = \gamma_N, \gamma_{N,F}$, and $\gamma_{N,FF}$.

(vi) $\max_{t, s, t \leq T} E(N^{1/2} \zeta_{st})^4 \leq C$ and $\max_{r, t} E(N^{-1/2} \Lambda'_r e_t)^4 \leq C$.

(vii) $\varpi_{NT,1}(r) \sim O_p(1)$ and $\max_{t} E(\varpi_{NT,2}(r, t))^2 \leq C$ for each $r$.

(viii) For all $r$, the eigenvalues of the $R \times R$ matrix $\Sigma_{\Lambda_r}^{1/2} \Sigma_F \Sigma_{\Lambda_r}^{1/2}$ are distinct.

**Assumption A.2.**

(i) $N^{-1/2} \Lambda'_r e_t \overset{d}{\longrightarrow} N(0, \Sigma_{\Lambda_r})$ for each $r, t$, where $\Sigma_{\Lambda_r} = \lim_{N \to \infty} N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_{ir} \lambda'_{jr} \times E(e_{it} e_{jt})$.

(ii) $\frac{\sqrt{T}}{\sqrt{N}} \sum_{s=1}^{T} k_{hs,ir} F_s e_{is} \overset{d}{\longrightarrow} N(0, \Omega_{i,r})$, where

$$\Omega_{i,r} = \lim_{T \to \infty} \left[ \frac{h}{T} \sum_{s=1}^{T} k_{hs,ir}^2 \text{Cov}(F_s F'_t e^2_{is}) + \frac{T}{h} \sum_{s=1}^{T} \sum_{t=1}^{T} k_{hs,ir} k_{ht,ir} \text{Cov}(F_s F'_t e_{is} e_{it}) \right].$$

**Assumption A.3**

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

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

A.1 mainly imposes moment conditions on the error terms, factors, factor loadings, and their interactions. They are widely used in the literature; see, e.g., Bai and Ng (2002) and Bai (2003). Note that we follow Stock and Watson (2002), Bai (2003), and Breitung and Eickmeier (2011) and assume that $E[F_t F'_t]$ is homogeneous over $t$ in A.1(ii), which facilitates the derivation of the asymptotic results significantly. With more complicated and lengthy arguments, we can allow for time-varying covariance for the factor loadings. 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 is 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 Davydenko inequality,
we can argue that the limit $\Omega_{i,r}$ in (3.5) exists. Without further assumptions, we cannot simplify it. If $E(F_tF_t^\prime e_{it}e_{it}^\prime) = \Omega$, for each $s$ and $\{e_{it}\}$ is a martingale difference sequence (m.d.s. hereafter) with respect to $\mathcal{F}_{it}$, the sigma-field generated from $\{e_{i,t-1}, e_{i,t-2}, ..., F_t, F_{t-1}, ...\}$, then we can readily show that $\Omega_{i,r} = \Phi_i \lim_{T \to \infty} \frac{1}{T^2} \sum_{t=1}^{T} K^*_t \left( \frac{F_t}{\sqrt{T}} \right)^{\prime} = \Phi_i \int_{-1}^{1} K(u)^2 du$ if $r \in [[Th], T - [Th]]$. 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 $r = 1, 2, ..., T$, we can show that there exists an invertible matrix $H^{(r)}$ such that $\hat{F}_t^{(r)}$ is a consistent estimator of $H^{(r)}F_t^{(r)}$ and $\hat{\lambda}_{ir}$ is a consistent estimator of $H^{(r)-1}\lambda_{ir}$.

The following theorem reports the asymptotic distribution of $\hat{F}_t^{(r)}$.

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

$$K^*_t \left( \frac{r-t}{Th} \right)^{-1/2} \sqrt{N h} \left[ \hat{F}_t^{(r)} - H^{(r)}F_t^{(r)} \right] \overset{d}{\to} N \left( 0, V_r^{-1}Q_r^\prime \Gamma_rQ_r^{(r)-1} \right),$$

where $H^{(r)} = \left( (N^{-1}\Lambda_r^\prime \Lambda_r) (T^{-1}F^{(r)} \hat{F}^{(r)}) \right) V_{NT}^{(r)-1}$, $V_{NT}^{(r)-1}$ denotes the $R \times R$ diagonal matrix of the first $R$ largest eigenvalues of $(NT)^{-1}X^{(r)}X^{(r)}^\prime$, $V_r$ is the diagonal matrix consisting of the eigenvalues of $\Sigma_{\Lambda_r}^{1/2} \Sigma_F^{1/2} \Sigma_{\Lambda_r}^{1/2}$ in descending order with $\Upsilon_r$ being the corresponding (normalized) eigenvector matrix ($\Upsilon_r^\prime \Upsilon_r = I_R$), and $Q_r = V_r^{1/2} \Upsilon_r^{-1} \Sigma_{\Lambda_r}^{-1/2}$.

Theorem 3.1 establishes the asymptotic normality of $\hat{F}_t^{(r)}$. We note that $\hat{F}_t^{(r)}$ is a consistent estimator for the transformed latent factor $F_t^{(r)} = \hat{k}_{i,r}^{1/2}F_t$ pre-multiplied by a transformation matrix $H^{(r)}$. Since we allow cross sectional dependence in the error terms, the limiting distribution depends on the cross-section correlation structure among $\{e_{it}\}$. In the case where $e_{it}$ is uncorrelated over $i$, we have $\Gamma_{rt} = \lim_{N \to \infty} N^{-1} \sum_{i=1}^{N} \lambda_{ir} \lambda_{ir}^\prime \sigma_{it}^2$ with $\sigma_{it}^2 = E[e_{it}^2]$. In addition, if $\sigma_{it}^2 = \sigma_i^2$ for each $i$, then we have $\Gamma_{rt} = \Sigma_{\Lambda_r} \sigma_i^2$.

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 $i = 1, 2, ..., N$ and $r = 1, 2, ..., T$, we have:

$$\sqrt{Th} \left( \hat{\lambda}_{ir} - H^{(r)-1}\lambda_{ir} \right) \overset{d}{\to} N \left( 0, (Q_r^{(r)-1})^{-1} \Omega_rQ_r^{(r)-1} \right).$$

Theorem 3.2 establishes the asymptotic normality of $\hat{\lambda}_{ir}$. When $\{e_{it}, \mathcal{F}_{it}\}$ is an m.d.s., the asymptotic variance can be simplified, leading to

$$\sqrt{Th} \left( \hat{\lambda}_{ir} - H^{(r)-1}\lambda_{ir} \right) \overset{d}{\to} N \left( 0, \int_{-1}^{1} K(u)^2 du \left( Q_r^{(r)-1} \right)^\prime \Omega_rQ_r^{(r)-1} \right).$$
when \( r \in [\lfloor Th \rfloor, T - \lfloor Th \rfloor] \).

As mentioned above, Theorem 3.1 only establishes asymptotic distribution for the transformed common factor \( F_t^{(r)} \). Since economists are usually interested in the estimation of the latent factor \( F_t \) itself, 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 Assumptions A.1, A.2(i) and A.3 hold. Then, for each \( t = 1, 2, \ldots, T \) we have

\[
\sqrt{N} \left[ \hat{F}_t - H^{(t)'} F_t \right] \xrightarrow{d} N \left( 0, (\Sigma_{\Lambda_t}^{-1} Q_t^{-1})' \Gamma_{tt} \Sigma_{\Lambda_t}^{-1} Q_t^{-1} \right).
\]

**Remark.** Interestingly, although the convergence rates of \( \hat{F}_t^{(r)} \) 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 can not derive the conventional optimal bandwidth in terms of minimizing the asymptotic mean square 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 \( \hat{h} \) by solving the following minimization problem:

\[
\min_{h} CV(h) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[ X_{it} - \hat{\lambda}_{it}^{(-i)'} \hat{F}_t^{(-i)} \right]^2,
\]

where \( \hat{\lambda}_{it}^{(-i)} \) and \( \hat{F}_t^{(-i)} \) are the analogue of \( \hat{\lambda}_{it} \) and \( \hat{F}_t \) by leaving the \( i \)th cross-sectional unit out in the local PCA procedure. But a rigorous study of the asymptotic behavior of \( \hat{h} \) would demand higher order asymptotics, which goes beyond the scope of the current paper.

### 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 \( \hat{\lambda}_{it} (R) \) denote the local PCA estimators of the factors and factor loadings by assuming \( R \) factors in the model and using the following normalization rule

\[
N^{-1} \Lambda'_t \Lambda_t = I_R 	ext{ and } T^{-1} F^{(r)'} F^{(r)} \text{ is a diagonal matrix},
\]

where we make the dependence of the \( R \times 1 \) vectors \( \hat{F}_t (R) \) and \( \hat{\lambda}_{it} (R) \) on \( R \) explicit. Let \( \hat{\Lambda}_t^{(R)} = (\hat{\lambda}_1^{(R)}, \ldots, \hat{\lambda}_R^{(R)})' \) and \( \hat{\Lambda}_t^{(R)} = (NT)^{-1} X^{(r)'} X^{(r)} \hat{\Lambda}_t^{(R)} \) for \( r = 1, \ldots, T \). Let \( \hat{\Lambda}_t^{(R)} \) denote the transpose of the \( i \)th row of \( \hat{\Lambda}_t^{(R)} \). Define

\[
V \left( R, \left\{ \hat{\Lambda}_t^{(R)} \right\} \right) = \min_{F = (\hat{F}_t^{(1)}, \ldots, \hat{F}_t^{(R)})'} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[ X_{it} - \hat{F}_t^{(R)} \hat{\lambda}_{it}^{(R)} \right]^2.
\]
Motivated by Bai and Ng (2002), we propose the following BIC-type information criterion to determine $R_0$:

$$IC(R) = \ln V \left( R, \left\{ \hat{A}(R) \right\} \right) + \rho_{NT} R,$$

(3.6)

where $\rho_{NT}$ 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.4.** (i) $\|e\|_{\mathrm{sp}} = O_P \left( N^{1/2} + T^{1/2} \right)$.  
(ii) $\max_{t,s} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left| e_{it}e_{is} - E(e_{it}e_{is}) \right| \right\} = O_P(N^{-1/2} \ln(T)^{1/2})$,

(iii) $\max_{t,r} \left\{ \frac{1}{T} \sum_{t=1}^{T} k_{h, tr} F_t e_{it} \right\} = O_P((T h)^{-1/2} \ln(NT)^{1/2})$ and $\max_r \left\{ \frac{1}{T} \sum_{t=1}^{T} k_{h, tr} \| F_t \|^2 - E \| F_t \|^2 \right\}$

$= O_P(T^{-1/2} \ln(T)^{1/2})$.

(iv) $\max_{t, s} E \left\{ N^{-1/2} k_{h, tr} e_{il} F_t \right\}^4 \leq C$ and $\max_{s, t} E \left\{ N^{-1/2} [F_s e_{il} e_{it} F_t - E(F_s e_{il} e_{it} F_t)] \right\}^2 \leq C$.

(v) $\max_{r} E \left\{ \frac{1}{(NT)^{1/2}} \sum_{i=1}^{N} \sum_{t=1}^{T} k_{h, tr} \left[ F_t e_{it} e_{ir} F_r - E(F_t e_{it} e_{ir} F_r) \right] \right\}^2 \leq C$.

**Assumption A.5.** As $(N, T) \to \infty$, $\rho_{NT} \to 0$ and $\rho_{NT} C_{NT}^2 \to \infty$ where $C_{NT} = \min(\sqrt{T h}, \sqrt{N})$.

The conditions in A.4 can be verified under some primitive conditions that are used in the factor literature. For example, Moon and Weidner (2014) demonstrate that A.4(i) can be satisfied for various error processes; Su et al. (2015) verify similar conditions to those in A.4(ii)-(v) under some mixing conditions. The conditions on $\rho_{NT}$ in A.5 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 3.4** Suppose that Assumptions A.1 and A.3-A.5 hold. Then

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

Theorem 3.4 indicates the class of information criteria defined by $IC(R)$ in (3.6) can consistently estimate $R_0$. To implement the information criterion, one still needs to choose the penalty coefficient $\rho_{NT}$. Following the lead of Bai and Ng (2002), we suggest setting $\rho_{NT} = \frac{N + T h}{NT h} \ln \left( \frac{NT h}{N + T h} \right)$ or $\rho_{NT} = \frac{N + T h}{NT h} C_{NT}^2$ with $C_{NT} = \min(\sqrt{T h}, \sqrt{N})$ 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

$$
H_0 : \lambda_{it} = \lambda_{i0} \text{ for } i = 1, 2, \ldots, N \text{ and } t = 1, 2, \ldots, T,
$$

and the alternative hypothesis is

$$
H_1 : \lambda_{it} \neq \lambda_{i0} \text{ for some non-negligible values of } (i, t),
$$

where $\lambda_{i0}$ is an unknown factor loading matrix, and $\lambda_{it} = \lambda_i(t/T)$ is an unknown piece-wise smooth function on $(0, 1]$ for each $i$ with a finite number of discontinuities under $H_1$.

Obviously, under the null hypothesis, $\lambda_{it}$ 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. Nevertheless, it is well known that factor models may exhibit structural changes over time. For this reason, much recent research has focused on testing for structural changes in factor models; see Breitung and Eickmeier (2011), Chen et al. (2014), Cheng et al. (2014), and Han and Inoue (2014). These authors aim at testing the existence of a single structural change in the factor loadings by using some supremum-type test statistics. However, usually no prior information about the structural change alternative is available in practice. It is extremely restrictive to assume only a single sudden structural break 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 $\lambda_{it}$ 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 $e_{it}^\dagger = e_{it} + (\lambda_{it} - \lambda_{i0})' F_t$. Let $X_t \equiv (X_{1t}, \ldots, X_{Nt})'$, $e_t \equiv (e_{1t}, \ldots, e_{Nt})'$, $e_t^\dagger \equiv (e_{1t}^\dagger, \ldots, e_{Nt}^\dagger)'$, $F \equiv (F_1, \ldots, F_T)'$, and $\Lambda_0 \equiv (\lambda_{10}, \ldots, \lambda_{N0})'$. Let $X \equiv (X_1', \ldots, X_T')$, $e \equiv (e_1', \ldots, e_T')'$, $e^\dagger \equiv (e_1^\dagger, \ldots, e_T^\dagger)'$. Then we can rewrite (2.1) in matrix form

$$
X = FA_0 + e^\dagger.
$$

(4.3)

Then we can rewrite (2.1) in matrix form

$$
X = FA_0 + e^\dagger.
$$

(4.3)

The conventional PCA method solves the following minimization problem

$$
\min_{F, \Lambda} \text{tr} (X - FA') (X - FA')' = \sum_{i=1}^{N} \sum_{i=1}^{N} (X_{it} - \lambda_i' F_t)^2
$$

11
under certain identification restrictions. In this paper, we follow Bai (2003) and consider the following identification restrictions: $T^{-1}F'F = 1_R$ and $\Lambda'\Lambda$ is a diagonal matrix. Let $\hat{F}_t$ and $\hat{\lambda}_{0t}$ be the principal component estimators of $F_t$ and $\lambda_{0t}$, respectively under the above identification restrictions. Let $\tilde{F} = (\tilde{F}_1, ..., \tilde{F}_T)'$ and $\hat{\lambda}_0 = (\hat{\lambda}_{10}, ..., \hat{\lambda}_{N0})'$. It is well known that $\tilde{F}$ is $\sqrt{T}$ times eigenvectors corresponding to the $R$ largest eigenvalues of the $T \times T$ matrix $XX'$, and $\hat{\lambda}_0 = (\tilde{F}'\tilde{F})^{-1}\tilde{F}'X = T^{-1}\tilde{F}'X$.

Given the estimates $(\hat{\lambda}_0', \hat{F}_t)$ of the common components $\lambda_{0t}'F_t$ under $H_0$ and those $(\hat{\lambda}_{0t}', \hat{F}_t)$ 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:

$$
\hat{M} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( \hat{\lambda}_{0t}' \hat{F}_t - \hat{\lambda}_{0t}' \tilde{F}_t \right)^2.
$$

We will show that after being appropriately rescaled and centered, $\hat{M}$ 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 $\|A\|_r = \{E\|A\|^r\}^{1/r}$ for $r \geq 1$. We add the following assumptions.

**Assumption A.6.** (i) For each $i = 1, 2, ..., N$, the process $\{e_{it}, t = 1, 2, \ldots\}$ is an m.d.s. such that $E(e_{it}|F_{NT,i-1}) = 0 \forall t$, where $F_{NT,i-1} = \{F_t, F_{t-1}, ..., e_{t-1}, e_{t-2}, \ldots\}$.

(ii) For each $i = 1, 2, ..., N$, the process $\{(e_{it}, F_t), t = 1, 2, \ldots\}$ is strong mixing with mixing coefficients $\alpha_i(s) \cdot \alpha(\cdot) \equiv \max_i \alpha_i(\cdot)$ satisfies $\sum_{s=1}^{\infty} \alpha(s)^{\delta/(2+\delta)} \leq C < \infty$ for some $\delta > 0$. In addition, there exists an integer $T_0 \in [1, T)$ such that $T^{-2} \max(T_0^2, T_0 h^{-1}, T_0^2 h^{-2}) \to 0$ and $N^2 Th^2 \alpha(T_0)^{\delta/(1+\delta)} \to 0$ as $(N, T) \to \infty$.

(iii) $\max_{i,t} \|F_t e_{it}\|_{8+4\delta} \leq C$ and $\max_{i,t} \|e_{it}\|_{8+4\delta} \leq C$.

(iv) $\max_{t \neq r} \|N^{-1/2} F_t e_{is} F_s'\|_4 \leq C$ and $E(e_{is} e_{js} F_s' F_s) = \tau_{ij,s}$ satisfies $\frac{1}{NT} \sum_{i=1}^{N} \sum_{s=1}^{N} \sum_{t=1}^{T} |\tau_{ij,s}| \leq C$.

**Assumption A.7.** (i) $\max_{s,t} \|N^{-1/2} A_{s}'e_{it}\| = O_P(N^{-1/2}(\ln T)^{1/2})$ and $\max_{s,t} \|N^{-1/2} A_{s}' e_{is} F_s\| = O_P(N^{-1/2}(\ln T)^{1/2})$.

(ii) $\max_r \|\varpi_{NT,1}(r)\| = O_P((\ln T)^{1/2})$ and $\max_{r,t} \|\varpi_{NT,2}(r, t)\| = O_P((\ln T)^{1/2})$.

(iii) $\max_r \left\| \frac{1}{N^2} \sum_{t=1}^{T} \lambda_{s}' e_{is}' F_t' (r')_{t} \right\| = \max_r \left\| \frac{1}{N^2} \sum_{t=1}^{T} k_{h,t}^{-1} \lambda_{s}' e_{is} F_t' \right\| = O_P ((NTh)^{-1/2}(\ln T)^{1/2})$.

Assumption A.6(i) assumes that the process $\{e_{it}, t = 1, 2, \ldots\}$ is an m.d.s. with respect to the filter $\{F_{NT,i}\}$. 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. A.6(ii) requires the process $\{(e_{it}, F_t), t = 1, 2, \ldots\}$ 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 \( \lim \sup_N \frac{1}{N} \sum_{i=1}^{N} \sum_{s=1}^{\infty} \alpha_i(s) \delta^{(2+\delta)} \leq C < \infty \). If the processes are strong mixing with a geometric rate (e.g., \( \alpha(s) = \rho^s \) for some \( \rho \in [0,1) \)), then the conditions on \( \alpha(\cdot) \) can be all met by specifying \( T_0 = \lfloor C_0 \ln T \rfloor \) for some sufficiently large positive constant \( C_0 \). A.6(iii) assumes some moment conditions on \( F_t e_{it} \) and \( e_{it} \), which, in conjunction with A.6(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(vi), A.6(iv) controls the cross-sectional dependence among \( \{F_t e_{it}, \ i = 1, 2, ..., N\} \).

Under A.6(iii), this condition becomes redundant if we would assume independence of \( e_t = (e_{i1}, ..., e_{iT})' \) across \( i \) conditional on the factors. A.7 imposes conditions on the uniform probability order of some summation objects. Again, these conditions can be easily verified by using Bernstein-type exponential inequality for independent or strong mixing processes.

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

**Assumption A.3.** (ii’) As \((N,T) \to \infty, h \to 0, Th^2 \to \infty, Nh^2 \to \infty, Th/N \to 0, Th/\ln T \to \infty, \text{ and } N^3 T^{-2} h^{-1} (\ln T)^{-2} \to \infty.\)

Let \( V_{NT} \) denote the \( R \times R \) diagonal matrices of the first \( R \) largest eigenvalues of \( (NT)^{-1} XX' \) in decreasing order and \( H = (N^{-1} \Lambda_0' \Lambda_0) (T^{-1} F' F) V_{NT}^{-1} \). Let \( k_{h, st} = h^{-1} K_s' \left( \frac{s-t}{2h} \right) \), \( \tilde{k}_{st} = \tilde{K} \left( \frac{s-t}{2h} \right) \) with \( \tilde{K} (u) = \int_{1}^{1} K (v) K (u-v) dv \) being the two-fold convolution kernel of \( K(\cdot) \). For example, if we use the Epanechnikov kernel \( K(u) = 0.75(1-u^2) 1 \{ |u| \leq 1 \} \) with \( 1 \{ \cdot \} \) being the usual indicator function, then \( \tilde{K}(u) = (\frac{3}{4} - \frac{1}{4}u^2 + \frac{3}{8} |u|^3 - \frac{3}{160} |u|^5) 1 \{ |u| \leq 2 \} \).

Let \( L_{st} = k_{h, st} H^{(1)} H^{(1)'} - HH' \). Define

\[
\mathbb{B}_{NT} = \frac{1}{N^{1/2} T^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} (F_{it}' L_{st} F_s)^2 e_{is},
\]

\[
V_{NT} = 2T^{-2} N^{-1} h^{-1} \sum_{1 \leq r \neq s \leq T} \tilde{k}_{sr}^2 E \left[ (F_{s}' H_0 \Sigma F_0 H'' F_{r})^2 (e_r e_s)^2 \right],
\]

where \( \tilde{\Sigma}_F = H_0' \Sigma F H_0, H_0 = Q^{-1} = (V^{-1/2} \Upsilon \Sigma_{\lambda_0}^{-1/2})^{-1} \) denotes the probability limit of \( H \) under \( \mathbb{H}_0 \), \( V \) is an \( R \times R \) diagonal matrix containing the \( R \) largest eigenvalues of \( \Sigma_{\lambda_0}^{-1/2} \Sigma F \Sigma_{\lambda_0}^{-1/2} \) in decreasing order, \( \Upsilon \) is the corresponding eigenvector matrix such that \( \Upsilon' \Upsilon = I_R \), and \( \Sigma_{\lambda_0} \) is the probability limits of \( N^{-1} \Lambda_0' \Lambda_0 \) under \( \mathbb{H}_0 \).

The following theorem states the asymptotic null distribution of our test statistic.

**Theorem 4.1** Suppose that Assumptions A.1, A.3(i) and (ii’), and A.6-A.7 hold. Then the test statistic \( J_{NT} \equiv V_{NT}^{-1/2} \left( T N^{-1/2} h^{1/2} \tilde{M} - \mathbb{B}_{NT} \right) \xrightarrow{d} N(0,1) \) under \( \mathbb{H}_0 \).

The proof of the above theorem is quite involved and is relegated to Appendix B. We make some remarks. First, each of the four terms, namely, \( \hat{\lambda}_{it}, \tilde{F}_t, \hat{\lambda}_{i0}, \) and \( \tilde{F}_t, \) in the definition of \( \tilde{M} \) contributes to the asymptotic distribution of \( J_{NT} \). We need to study the asymptotic expansion for each of these four estimators. Second, after some tedious calculations, we can demonstrate that under \( \mathbb{H}_0, T N^{-1/2} h^{1/2} \tilde{M} - \mathbb{B}_{NT} = \sum_{s=2}^{T} Z_{NT, s} + o_p(1) \), where

\[
Z_{NT, s} = 2T^{-1} N^{-1/2} h^{-1/2} \sum_{r=1}^{s-1} \sum_{i=1}^{N} \tilde{k}_{sr} F_{is}' H_0 \Sigma F_0 H'' F_r e_{is} e_{ir}.
\]
Under the m.d.s. condition in Assumption A.6(i), one can verify that \( E(Z_{NT,s} \mid F_{NT,s-1}) = 0 \) and resort to a martingale central limit theorem (e.g., Pollard 1984, p.171) to derive the asymptotic distribution of \( J_{NT} \). Difficulty arrives when we try to verify the Lyapunov condition via the fourth order moment of \( Z_{NT,s} \) because we do not assume cross-sectional independence among \( e_i = (e_{i1}, ..., e_{iT})' \) conditional on the factors. The strong mixing condition in A.6(ii) and the moment conditions in A.6(iii)-(iv) greatly facilitate the verification of the Lyapunov condition. Third, despite the assumed m.d.s. condition, the variance term \( \mathbb{V}_{NT} \) still takes the form of a double U-statistic that involves two summations over each of the individual and time dimensions.

To implement the test, we need to estimate both the asymptotic bias \( \hat{B}_{NT} \) and the asymptotic variance \( \hat{V}_{NT} \). The consistent estimators for \( B_{NT} \) and \( \mathbb{V}_{NT} \) are respectively given by

\[
\hat{B}_{NT} = \frac{h^{1/2}}{TN^{1/2}} \sum_{i=1}^{N} \frac{1}{T} \sum_{s=1}^{T} \left( k_{i,s} \hat{F}_i \hat{F}'_i - \hat{F}'_i \hat{F}_i \right) \hat{e}^2_{is}, \quad \text{and}
\]

\[
\hat{V}_{1NT} = 2T^{-2}N^{-1}h^{-1} \sum_{1 \leq s \neq r \leq T} \hat{k}_{sr} \left( \hat{F}'_s \hat{F}'_r \right)^2 \left( \hat{e}'_s \hat{e}'_r \right)^2,
\]

where \( \hat{e}_{is} = X_{is} - \hat{\lambda}_{is}' \hat{F}_s \). Then we consider the feasible test statistic:

\[
\hat{J}_{NT} = \hat{V}_{1NT}^{-1/2} \left( TN^{1/2}h^{1/2} \hat{M} - \hat{B}_{NT} \right).
\]

The following theorem establishes the consistency of \( \hat{B}_{NT} \) and \( \hat{V}_{1NT} \) and the asymptotic normality of \( \hat{J}_{NT} \).

**Theorem 4.2** Suppose that Assumptions A.1, A.3(i) and (ii*), and A.6-A.7 hold. Then under \( \mathbb{H}_0 \), \( \hat{B}_{NT} = B_{NT} + o_P(1) \), \( \hat{V}_{1NT} = V_{1NT} + o_P(1) \), and \( \hat{J}_{NT} \overset{d}{\rightarrow} N(0,1) \).

Theorem 4.2 indicates that our test statistic \( \hat{J}_{NT} \) is asymptotically pivotal under \( \mathbb{H}_0 \). We can compare the value of \( \hat{J}_{NT} \) to the critical value \( z_\alpha \), 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 \( \hat{J}_{NT} > z_\alpha \).

### 4.4 Asymptotic local power

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

\( \mathbb{H}_1(a_{NT}) : \lambda_{it} = \lambda_{i0} + a_{NT} g_i \left( \frac{t}{T} \right) \) for each \( i \) and \( t \),

where \( a_{NT} \to 0 \) as \( (N, T) \to \infty \), it controls the speed at which the local alternative converges to the null hypothesis, and \( g_i \left( \frac{t}{T} \right) \) is a vector-valued piecewise smooth function with a finite number of discontinuity points. Noting that \( \lambda_{i0} + a_{NT} g_i \left( \frac{t}{T} \right) = (\lambda_{i0} + c_{i,NT}) + a_{NT}[g_i \left( \frac{t}{T} \right) - c_{i,NT}/a_{NT}] \) for any \( c_{i,NT} = O(a_{NT}) \), below we will assume that

\[
\int_0^1 g_i(u) \, du = 0
\]
for location normalization purpose. With this normalization, both \( \lambda_{i0} \) and \( g_i(\cdot) \) can be dependent on the sample sizes \( N \) and \( T \). But for notational simplicity, we continue to write them as \( \lambda_{i0} \) and \( g_i(\cdot) \) instead of \( \lambda_{i0,NT} \) and \( g_{i,NT}(\cdot) \).

Let \( g_{it} = g_i(\frac{t}{T}) \), \( g_{it}^j = F_i^j g_i(\frac{t}{T}) \), and \( g_i^j = (g_{i1},...,g_{iN})' \). Define

\[
\Pi_1 = \lim_{(N,T) \to \infty} T^{-1} \sum_{t=1}^{T} \text{tr} \left[ (H_0^{-1})' V_0^{-1} H_0^{-1} \left( N^{-1} A_0 g_i^j \right) \left( N^{-1} g_i^j A_0 \right) (H_0^{-1})' V_0^{-1} H_0^{-1} \Sigma_0 \right],
\]

\[
\Pi_2 = \lim_{(N,T) \to \infty} (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \text{tr} (\Sigma_F g_{it} g_{it}').
\]

(4.5)

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

**Assumption A.8.**

(i) For each \( i = 1,2,...,N \), \( g_i(\cdot) \) is piecewise continuous with a finite number of discontinuous points on \( (0,1] \).

(ii) \( \max_{1 \leq i \leq N} \max_{1 \leq t \leq T} \| \lambda_{i0} \| \leq \hat{c}_\lambda < \infty \).

(iii) \( \max_{1 \leq i \leq T} \left\| \hat{\beta}_i \sum_{s=1}^{N} \sum_{j=1}^{T} \hat{\beta}_{is} r_s g_{it} \right\| = O_P((NTh/\ln(NT))^{-1/2}).

(iv) The limits \( \Pi_1 \) and \( \Pi_2 \) defined in (4.5) exist and \( \Pi_1 + \Pi_2 > 0 \).

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 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), and A.6-A.8 hold. Then under \( \mathbb{H}_1(a_{NT}) \) with \( a_{NT} = T^{-1/2} N^{-1/4} h^{-1/4} \), \( \hat{\mathbb{B}}_{NT} = \mathbb{B}_{NT} + o_P(1) \), \( \hat{\beta}_{NT} = \beta_{NT} + o_P(1) \), and \( \hat{J}_{NT} \stackrel{d}{\to} N(\pi_0,1) \), where \( \pi_0 = (\Pi_1 + \Pi_2)/\mathbb{V}_0^{1/2} \) and \( \mathbb{V}_0 = \lim_{(N,T) \to \infty} \mathbb{V}_{NT} \).

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 we allow the existence of a finite number of unknown discontinuity points in 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.

#### 4.5 Asymptotic global power

To study the asymptotic global power property of our test, we define

\[
F_T = \left\{ \tilde{F} : \tilde{F}' \tilde{F} = I_R \right\} \quad \text{and} \quad A_N = \left\{ \tilde{A} : \tilde{A}' \tilde{A} = \text{diagonal matrix} \right\},
\]

where \( \tilde{A} = (\tilde{\lambda}_1,...,\tilde{\lambda}_N)' \) and \( \tilde{F} = (\tilde{F}_1,...,\tilde{F}_T)' \).

**Assumption A.9.** There exists \( \xi_{AF} > 0 \) such that \( \text{plim}_{(N,T) \to \infty} \inf_{(\tilde{A},\tilde{F}) \in A_N \times F_T} 1 \overline{N}T \sum_{i=1}^{N} \sum_{t=1}^{T} (\lambda_{it} F_t - \tilde{\lambda}_{it} \tilde{F}_t)^2 \geq \xi_{AF} \).
Assumption A.9 is intuitively clear: in the spaces of factors and factor loadings such that the normalization rules in \( F_t \) and \( \Lambda_N \) are satisfied, we cannot find any time-invariant factor loadings \( \tilde{\lambda} \)'s and the associated factors \( \tilde{F}_t \)'s such that \( \tilde{\lambda}_t \tilde{F}_t \) converges to the true common component \( \lambda_t F_t \) in the sense of mean square error. If A9 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 and A.9 hold. Then under the global alternative \( H_1 \), \( P(J \geq c_{NT}) \rightarrow 1 \) as \( (N, T) \rightarrow \infty \) for any positive sequence \( c_{NT} \) that is \( o(TN^{1/2}h^{1/2}) \).

Theorem 4.4 implies that \( J \) is consistent and divergent to infinity at the rate \( TN^{1/2}h^{1/2} \). Note that A.6-A.8 are not required here as there is no need to derive the asymptotic distribution of \( J \) 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 \( \hat{\sigma}_i^2 = T^{-1} \sum_{t=1}^T \hat{e}_{it}^2 \), where \( \hat{e}_{it} = X_{it} - \hat{\lambda}^{\top}_i \tilde{F}_t \), and \( \tilde{F}_t \) and \( \hat{\lambda}_i \) are the estimates of the factors and factor loadings under the null. Let \( e_{it} = \hat{\sigma}_i \hat{\varepsilon}_{it} \) with \( \varepsilon_{it} \) being IID \( N(0, 1) \) over both \( i \) and \( t \). Then one can generate the bootstrap resamples via \( X_{it}^* = \hat{\lambda}^{\top}_i \tilde{F}_t + e_{it} \) 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 \( \{e_{it}^*\} \) 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 \( \hat{e}_t = (\hat{e}_{1t}, ..., \hat{e}_{Nt})' \) and \( \hat{\Sigma} = T^{-1} \sum_{t=1}^T \hat{e}_t \hat{e}_t' \). Let \( \hat{\sigma}_{ij} \) denote the \((i,j)\)th element of \( \hat{\Sigma} \). Define the shrinkage version of \( \hat{\Sigma} \) as \( \tilde{\Sigma} \) whose \((i,j)\)th element is given by

\[
\tilde{\sigma}_{ij} = \hat{\sigma}_{ij}^0 (1 - \epsilon)^{|j-i|} \quad \text{for } i, j = 1, ..., N,
\]

where \( \epsilon \) is a small positive number (e.g., 0.01) to ensure the maximum absolute column/row sum norm of \( \tilde{\Sigma} \) to be stochastically bounded provided \( \max_{i,j} |\hat{\sigma}_{ij}^0| \) is. By construction, \( \tilde{\Sigma} \) is also symmetric and positive semi-definite. The stochastic boundedness of \( \max_{i,j} |\hat{\sigma}_{ij}| \) is sufficient but not necessary for the justification of the asymptotic validity of our bootstrap procedure below.
1. Estimate the restricted model \( X_{i t} = \lambda_{i0} F_t + e_{i t} \) by the PCA method and the unrestricted model \( X_{i t} = \lambda_{i0} F_t + e_{i t} \) by the local PCA method to obtain the two sets of estimates \( \{\tilde{\lambda}_{i0}, \tilde{F}_t\} \) and \( \{\hat{\lambda}_{i t}, \hat{F}_t\} \). Based on these estimates, construct the test statistic \( \tilde{J}_{NT} \) as in Section 4.2.

2. For \( i = 1, \ldots, N \) and \( t = 1, 2, \ldots, T \), obtain the bootstrap error \( e^*_{i t} = \tilde{\Sigma}^{1/2}\zeta_t \), where \( \zeta_t = (\zeta_{1 t}, \ldots, \zeta_{N t})' \) with \( \zeta_{i t} \) being IID \( N(0, 1) \) across \( i \) and \( t \). Generate \( X^*_{i t} = \tilde{\lambda}_{i0} \tilde{F}_t + e^*_{i t} \).

3. Use \( \{X^*_{i t}\} \) to run the restricted and unrestricted models to obtain the bootstrap versions \( \{\tilde{\lambda}^*_{i0}, \tilde{F}^*_t\} \) and \( \{\hat{\lambda}^*_{i t}, \hat{F}^*_t\} \) of \( \{\tilde{\lambda}_{i0}, \tilde{F}_t\} \) and \( \{\hat{\lambda}_{i t}, \hat{F}_t\} \), respectively. Calculate the bootstrap test statistic \( \tilde{J}^*_{NT} \), the bootstrap version of \( \tilde{J}_{NT} \).

4. Repeat steps 2 and 3 for \( B \) times and index the bootstrap test statistics as \( \{\tilde{J}^*_{NT,l}\}_{l=1}^B \). The bootstrap \( p \)-value is calculated by \( p^* = B^{-1} \sum_{l=1}^B 1\{\tilde{J}^*_{NT,l} > \tilde{J}_{NT}\} \).

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

\begin{theorem}
Suppose that Assumptions A.1, A.3(i) and (ii*) and A.6-A.7 hold. Suppose that (i) \( \max_{i,j} |\hat{\sigma}_{ij}^0| = O_P(\zeta_{NT}) \) with \( \zeta_{NT} = o(T^{1/2}) \), (ii) \( \frac{1}{T} \sum_{t=1}^T \left\| \tilde{F}_t \right\|^8 = O_P(1) \) and (iii) \( \frac{1}{N} \sum_{i=1}^N \left\| \tilde{\lambda}_0 \right\|^8 = O_P(1) \). Then \( \tilde{J}^*_{NT} \overset{D}{\to} \mathcal{N}(0, 1) \) in probability, where \( \overset{D}{\to} \) denotes weak convergence under the bootstrap probability measure conditional on the observed sample \( X \).
\end{theorem}

Theorem 4.5 shows that the bootstrap provides an asymptotic valid approximation to the limit null distribution of \( \tilde{J}_{NT} \). This holds as long as we generate the bootstrap data by imposing the null hypothesis. If the null hypothesis does not hold in the observed sample, then we expect \( \tilde{J}_{NT} \) to explode at the rate \( T^{1/2}N^{1/4}h^{1/4} \), which delivers the consistency of the bootstrap-based test \( \tilde{J}_{NT} \). 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 apply arguments as used in the proof of Lemma B.7(i) to demonstrate that \( \frac{1}{T} \sum_{t=1}^T \left\| \tilde{F}_t \right\|^8 = O_P(1) + O_P(T^3(N^{-4} + T^{-4})) = O_P(1) \) and similarly \( \frac{1}{N} \sum_{i=1}^N \left\| \tilde{\lambda}_0 \right\|^8 = O_P(1) \) provided \( T^3N^{-4} + N^3T^{-4} = O(1) \).

5 Monte Carlo Study

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

5.1 Data generating process

We generate data under the framework of large model with \( R = 2 \) common factors:

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

**DGP 1: (IID)**

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

**DGP 2: (Heteroskedasticity)**

\[ \lambda_{it} = \lambda_{i0} \sim \text{IID } N(0, I_2), \ e_{it} = \sigma_i v_{it}, \text{ where } \sigma_i \sim \text{IID } U(0.5, 1.5) \text{ and } v_{it} \sim \text{IID } N(0, 1). \]

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

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

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

\[
\begin{align*}
\lambda_{it,k} &= \begin{cases} 
\lambda_{i0,k}, & \text{for } t = 1, 2, ..., T/2 \\
\lambda_{i0,k} + b, & \text{for } t = T/2 + 1, ..., T 
\end{cases}, \quad \lambda_{i0,k} \sim \text{IID } N(1, 1), \ k = 1, 2; \\
 e_{it} &= \sigma_i v_{it}, \text{ where } \sigma_i \sim \text{IID } U(0.5, 1.5), v_{it} \sim \text{IID } N(0, 1), \text{ and } b = 1, 2, 4.
\end{align*}
\]

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

\[
\begin{align*}
\lambda_{i0,t} &= \begin{cases} 
\lambda_{i0,1} + 0.5b, & \text{for } 0.6T < t \leq 0.8T \\
\lambda_{i0,1} - 0.5b, & \text{for } 0.2T < t \leq 0.4T \\
\lambda_{i0,1}, & \text{otherwise}
\end{cases}, \quad \lambda_{i0,1} \sim \text{IID } N(1, 1), \lambda_{i0,2} = \lambda_{i0,2} \sim \text{IID } N(0, 1), \\
 e_{it} &\sim \text{IID } N(0, 1), \text{ and } b = 1, 2, 4.
\end{align*}
\]

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

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

\[ e_{it} \sim \text{IID } N(0, 1). \]

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

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

\[ e_{it} \sim \text{IID } N(0, 1). \]

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

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

\[ e_{t} = (e_{1t}, ..., e_{Nt})' \sim \text{IID } N(0, \Sigma_e), t = 1, 2, ..., T, \text{ where } \Sigma_e = (c_{ij})_{i,j=1, ..., N} \text{ with } c_{ij} = 0.5^{i-j}. \]

Here, \( G(z; \kappa, \gamma) = \{1 + \exp[-\kappa \sum_{l=1}^{p} (z - \gamma_l)]\}^{-1} \) denotes the Logistic function with tuning parameter \( \kappa \) and location parameter \( \gamma = (\gamma_1, ..., \gamma_p)' \).

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. Specifically, DGP 2 and 3 examine the performance of our test and the information criterion under heteroskedasticity and cross sectional dependence. DGP 4-8 describe various time-varying factor loadings. DGP 4 and 5 exhibit single and four sudden structural breaks, respectively. DGP 6-7 describe two kinds of smooth structural changes. Particularly, the factor loadings generated in DGP 6 are monotonic functions while the factor loadings given in DGP 7 are smooth.
transition functions with multiple regime shifts. DGP 8 considers the process with smooth structural changes and cross sectional dependence.

5.2 Determination of the number of factors

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

\[
IC_{h1}(R) = \ln V \left( R, \left\{ \lambda^{(R)}_i \right\} \right) + R \left( \frac{N + Th}{NTh} \right) \ln \left( \frac{NTh}{N + Th} \right),
\]

\[
IC_{h2}(R) = \ln V \left( R, \left\{ \lambda^{(R)}_i \right\} \right) + R \left( \frac{N + Th}{NTh} \right) \ln C_{NT}^2, \quad C_{NT} = \min \left\{ \sqrt{Th}, \sqrt{N} \right\}.
\]

For comparison purpose, we also consider Bai and Ng’s (2002) four information criteria (namely, \(PC_{p1}, PC_{p2}, IC_{p1}, \) and \(IC_{p2}\)), and Ahn and Horenstein’s (2013) two criterion functions (\(ER\) for eigenvalue ratio and \(GR\) for growth ratio). In addition, we implement Onatski’s (2009) sequential testing procedure (\(Ona\)) to determine the number of factors.

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. Our local PCA involves nonparametric estimation. We use the Epanechnikov kernel and Silverman’s rule of thumb (RoT) to set the bandwidth as \(h = (2.35/\sqrt{12})T^{-1/3}N^{-1/10}\). [Note that \(\{t/T\}_{t=1}^T\) behaves like a uniform random variable on \([0, 1]\) and thus has variance \(1/12\).] We also try the Uniform kernel and the Quartic kernel, and the RoT bandwidth with different tuning parameters. Our simulation studies show that the choice of kernel function and the bandwidth has 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 \(\hat{R} = 1\) 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 4-8 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 4-8.

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 dependent error terms
Table 1: Comparison of various information criteria in determining the number of factors: DGPs 1-5

<table>
<thead>
<tr>
<th>DGP (N, T)</th>
<th>IC_{h1}</th>
<th>IC_{h2}</th>
<th>PC_{p1}</th>
<th>PC_{p2}</th>
<th>IC_{p1}</th>
<th>IC_{p2}</th>
<th>Ona</th>
<th>ER</th>
<th>GR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (100,100)</td>
<td>2.02</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
</tr>
<tr>
<td>(100,200)</td>
<td>2.02</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
</tr>
<tr>
<td>(200,100)</td>
<td>2.02</td>
<td>2.01</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
</tr>
<tr>
<td>(200,200)</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
</tr>
</tbody>
</table>

| 2 (100,100) | 2.01 | 2.01 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 |
| (100,200) | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 |
| (200,100) | 2.02 | 2.01 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 |
| (200,200) | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 |

| 3 (100,100) | 2.02 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 |
| (100,200) | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 |
| (200,100) | 2.02 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 |
| (200,200) | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 |

| 4 b=1 (100,100) | 2.03 | 2.00 | 3.24 | 3.02 | 3.00 | 3.00 | 2.39 | 1.67 | 2.01 |
| (100,200) | 2.00 | 2.00 | 3.19 | 3.03 | 3.00 | 3.00 | 2.68 | 1.77 | 2.01 |
| (200,100) | 2.00 | 2.00 | 3.00 | 3.00 | 3.00 | 3.00 | 2.92 | 1.50 | 2.06 |
| (200,200) | 2.00 | 2.00 | 3.00 | 3.00 | 3.00 | 3.00 | 2.97 | 1.59 | 2.14 |
| b=2 (100,100) | 2.21 | 2.08 | 3.10 | 3.00 | 3.00 | 3.00 | 2.49 | 1.01 | 2.32 |
| (100,200) | 2.00 | 2.00 | 3.19 | 3.03 | 3.00 | 3.00 | 2.68 | 1.77 | 2.01 |
| (200,100) | 2.13 | 2.08 | 3.00 | 3.00 | 3.00 | 3.00 | 2.98 | 1.01 | 2.63 |
| (200,200) | 2.02 | 2.01 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 1.01 | 3.00 |
| b=4 (100,100) | 2.66 | 2.48 | 3.00 | 3.00 | 3.00 | 3.00 | 2.87 | 1.00 | 2.08 |
| (100,200) | 2.53 | 2.32 | 3.00 | 3.00 | 3.00 | 3.00 | 2.99 | 1.00 | 2.84 |
| (200,100) | 2.59 | 2.48 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 1.00 | 2.64 |
| (200,200) | 2.39 | 2.30 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 1.00 | 3.00 |

| 5 b=1 (100,100) | 2.00 | 2.00 | 2.68 | 2.34 | 2.12 | 2.12 | 2.01 | 2.00 | 2.00 |
| (100,200) | 2.00 | 2.00 | 2.90 | 2.72 | 2.36 | 2.36 | 2.02 | 2.00 | 2.00 |
| (200,100) | 2.00 | 2.00 | 2.64 | 2.26 | 2.26 | 2.26 | 2.60 | 2.00 | 2.00 |
| (200,200) | 2.00 | 2.00 | 2.95 | 2.78 | 2.78 | 2.78 | 2.23 | 2.00 | 2.00 |
| b=2 (100,100) | 2.01 | 2.00 | 3.00 | 3.00 | 3.00 | 3.00 | 2.76 | 2.10 | 2.30 |
| (100,200) | 2.00 | 2.00 | 3.00 | 3.00 | 3.00 | 3.00 | 2.96 | 2.30 | 2.53 |
| (200,100) | 2.00 | 2.00 | 3.00 | 3.00 | 3.00 | 3.00 | 2.98 | 2.45 | 2.65 |
| (200,200) | 2.00 | 2.00 | 3.00 | 3.00 | 3.00 | 3.00 | 2.84 | 2.94 | 2.94 |
| b=4 (100,100) | 2.32 | 2.21 | 3.00 | 3.00 | 3.00 | 3.00 | 2.97 | 2.30 | 2.30 |
| (100,200) | 2.23 | 2.12 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 |
| (200,100) | 2.21 | 2.16 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 |
| (200,200) | 2.16 | 2.11 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 |

Note: (i) IC_{h1} and IC_{h2} denote the information criteria proposed in this paper; (ii) PC_{p1}, PC_{p2}, IC_{p1}, and IC_{p2} denote Bai and Ng’s (2002) information criteria; (iii) Ona denotes the results of Onatski’s (2009) test; (iv) ER and GR denote Ahn and Horenstein’s (2013) criteria. Numbers in the main entries are the results based on 1000 replications.
<table>
<thead>
<tr>
<th>DGP</th>
<th>( (N,T) )</th>
<th>( IC_{h1} )</th>
<th>( IC_{h2} )</th>
<th>( PC_{p1} )</th>
<th>( PC_{p2} )</th>
<th>( IC_{p1} )</th>
<th>( IC_{p2} )</th>
<th>( Ona )</th>
<th>( ER )</th>
<th>( GR )</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( b=1 )</td>
<td>( (100,100) )</td>
<td>2.00</td>
<td>2.00</td>
<td>2.12</td>
<td>2.02</td>
<td>2.00</td>
<td>2.00</td>
<td>2.01</td>
<td>2.00</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>( (100,200) )</td>
<td>2.00</td>
<td>2.00</td>
<td>2.09</td>
<td>2.01</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>( (200,100) )</td>
<td>2.00</td>
<td>2.00</td>
<td>2.15</td>
<td>2.05</td>
<td>2.01</td>
<td>2.01</td>
<td>2.11</td>
<td>2.21</td>
<td>2.20</td>
</tr>
<tr>
<td></td>
<td>( (200,200) )</td>
<td>2.00</td>
<td>2.00</td>
<td>2.37</td>
<td>2.07</td>
<td>2.08</td>
<td>2.08</td>
<td>2.18</td>
<td>2.18</td>
<td>2.00</td>
</tr>
<tr>
<td>( b=2 )</td>
<td>( (100,100) )</td>
<td>2.00</td>
<td>2.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>2.96</td>
<td>2.93</td>
<td>1.91</td>
</tr>
<tr>
<td></td>
<td>( (100,200) )</td>
<td>2.00</td>
<td>2.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>2.96</td>
<td>2.93</td>
<td>1.91</td>
</tr>
<tr>
<td></td>
<td>( (200,100) )</td>
<td>2.00</td>
<td>2.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>2.96</td>
<td>2.93</td>
<td>1.91</td>
</tr>
<tr>
<td></td>
<td>( (200,200) )</td>
<td>2.00</td>
<td>2.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
<td>2.96</td>
<td>2.93</td>
<td>1.91</td>
</tr>
<tr>
<td>( b=4 )</td>
<td>( (100,100) )</td>
<td>2.07</td>
<td>2.04</td>
<td>3.98</td>
<td>3.89</td>
<td>3.83</td>
<td>3.83</td>
<td>1.51</td>
<td>1.01</td>
<td>2.27</td>
</tr>
<tr>
<td></td>
<td>( (100,200) )</td>
<td>2.01</td>
<td>2.01</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
<td>1.45</td>
<td>1.00</td>
<td>2.55</td>
</tr>
<tr>
<td></td>
<td>( (200,100) )</td>
<td>2.03</td>
<td>2.02</td>
<td>3.99</td>
<td>3.97</td>
<td>3.96</td>
<td>3.96</td>
<td>2.76</td>
<td>1.09</td>
<td>2.70</td>
</tr>
<tr>
<td></td>
<td>( (200,200) )</td>
<td>2.00</td>
<td>2.00</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
<td>2.35</td>
<td>1.04</td>
<td>2.94</td>
</tr>
</tbody>
</table>

Note: See the note in Table 1.
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). This is as expected for DGP 4 because we have a single break there and the break only occurs for one of the two common factors. 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 6-8 are factor models with smooth structural changes in factor loadings and/or cross sectionally dependent errors. As shown in Table 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 \((b = 1)\).

### 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) and Han and Inoue (2014) for a single structural break with an unknown break date in factor loadings.

It is well known that a nonparametric test that relies on the asymptotic normal approximation may perform poorly in finite samples. To conquer this problem we consider the wild bootstrap procedure proposed in Section 4.6. Since the bootstrap procedure is rather time consuming, we generate 500 data sets in this subsection and set the bootstrap replication number \(B\) to be 200. As in the previous subsection, we use the Epanechnikov kernel and the rule of thumb 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 and Han and Inoue’s (2014) sup-LM and sup-Wald tests. We follow these papers to set the trimming parameter \(\tau = 0.15\). The tests of Chen et al. (2014) and Han and Inoue (2014) 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 (2014), while the bootstrap critical values are applied to check the performance of our test.
Table 3: Size of various tests for DGPs 1-3

<table>
<thead>
<tr>
<th>DGP</th>
<th>N</th>
<th>T</th>
<th>J_{NT}</th>
<th>H_{LM}</th>
<th>H_{IW}</th>
<th>CDG_{LM}</th>
<th>CDG_{IW}</th>
<th>BE_{LM}</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>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>100</td>
<td>5.4</td>
<td>3.6</td>
<td>7.2</td>
<td>2.0</td>
<td>2.2</td>
<td>8.2</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</td>
<td>5.0</td>
<td>6.0</td>
<td>11.6</td>
<td>1.4</td>
<td>5.6</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>100</td>
<td>5.2</td>
<td>3.2</td>
<td>6.0</td>
<td>0.0</td>
<td>1.6</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>3.8</td>
<td>8.8</td>
<td>7.0</td>
<td>14.4</td>
<td>2.0</td>
<td>7.8</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>100</td>
<td>6.8</td>
<td>3.0</td>
<td>8.4</td>
<td>0.4</td>
<td>1.8</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</td>
<td>5.8</td>
<td>11.0</td>
<td>3.2</td>
<td>8.2</td>
<td>2.6</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>100</td>
<td>7.4</td>
<td>12.8</td>
<td>2.8</td>
<td>6.8</td>
<td>0.4</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>6.2</td>
<td>11.6</td>
<td>3.6</td>
<td>9.6</td>
<td>2.2</td>
<td>7.4</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>100</td>
<td>6.0</td>
<td>11.0</td>
<td>3.6</td>
<td>8.2</td>
<td>0.2</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</td>
<td>4.2</td>
<td>8.6</td>
<td>7.0</td>
<td>12.8</td>
<td>1.8</td>
<td>6.6</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>100</td>
<td>4.8</td>
<td>11.2</td>
<td>3.2</td>
<td>6.0</td>
<td>0.2</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>4.2</td>
<td>8.6</td>
<td>7.8</td>
<td>13.2</td>
<td>1.8</td>
<td>6.8</td>
</tr>
</tbody>
</table>

Note: (i) $J_{NT}$ denote the results of our test based on the bootstrap $p$-values; (ii) $H_{LM}$ and $H_{IW}$ denote Han and Inoue’s (2014) sup-LM and sup-Wald tests; (iii) $CDG_{LM}$ and $CDG_{IW}$ denote Chen et al.’s (2014) sup-LM and sup-Wald tests; (iv) $BE_{LM}$ denotes Breitung and Eickmeier’s (2011) variable-specific sup-LM test. The entries report the average percentage of rejection for various tests.

Table 3 reports the empirical sizes of various tests at both 10% and 5% levels. As shown in the table, our test has reasonable sizes using the bootstrap $p$-values. Han and Inoue’s (2014) sup-LM test delivers reasonable size and their sup-Wald test tends to under-reject the null hypothesis. Chen et al.’s (2014) sup-LM test also has reasonable size, but their sup-Wald test tends to over-reject the null hypothesis. In addition, Breitung and Eickmeier’s (2011) variable-specific sup-LM test suffers from slight underrejection for DGPs 1-3.

Table 4 reports the empirical powers of various tests for DGPs 4-8 at the 5% and 10% significance levels. 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 4-8 and the simulation results are consistent with our theoretical prediction that our test is able to detect both a finite number of sudden structural breaks and smooth structural changes. Second, the other tests are all designed to test for a one-time abrupt structural change in DGP 4. As expected, they all have power against DGP 4 despite the fact their power is not as great as that of our test. Third, for the other DGPs, all of Han and Inoue’s (2014) sup-LM and sup-Wald tests, Chen et al.’s (2014) sup-LM and sup-Wald tests, and Breitung and Eickmeier’s (2011) have lower power than our test too. In particular, these tests have little or low power in detecting deviations from the null in DGPs 5 and 7 but 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 (2014) 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.

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
1959:I-2006:IV. 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 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 details of the data description and processing, one can refer
to Stock and Watson (2009).

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 $\Pi_1$, $\Pi_2$, $\Pi_3$, and
$I_C$, Onatski’s (2009) testing procedure, Ahn and Horenstein’s (2013) criterion functions $ER$ and $GR$
and our information criterion proposed in Section 3.3 to determine the number of common factors. The
results are reported in Table 5. According to the table, we report the test results for the cases of one to
two common factors respectively in the following context.

Table 6 reports the results of the test and the corresponding critical values at the 5% and 10% sig-
nificance levels. Our test rejects the null hypothesis of time-invariant factor loadings for all the cases of
1–5 common factors. In contrast, Han and Inoue’s (2014) sup-LM and sup-Wald tests cannot reject the
null for any case at 5% significance level, while Chen et al.’s (2014) results are mixed, and they can only
reject the null for $R = 5$ at 5% significance level when using the sup-Wald test. This is consistent with the
results of our simulation studies that suggest the tests of Han and Inoue (2014) and Chen et al. (2014)
have relatively low power.

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. Figure 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

---

2The dataset is publicly available on Professor Mark W. Watson’s website http://www.princeton.edu/~mwatson/publi.html.
<table>
<thead>
<tr>
<th>DGP</th>
<th>N</th>
<th>T</th>
<th>( J_{NT} )</th>
<th>( H_{LM} )</th>
<th>( H_{W} )</th>
<th>( CDG_{LM} )</th>
<th>( CDG_{W} )</th>
<th>( BE_{LM} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td>4, ( b = 1 )</td>
<td>100</td>
<td>100</td>
<td>99.0</td>
<td>99.6</td>
<td>39.6</td>
<td>62.6</td>
<td>21.0</td>
<td>43.8</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>95.2</td>
<td>98.0</td>
<td>92.6</td>
<td>97.0</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>100</td>
<td>99.4</td>
<td>99.6</td>
<td>41.2</td>
<td>65.0</td>
<td>22.2</td>
<td>46.0</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>96.0</td>
<td>98.6</td>
<td>93.4</td>
<td>98.0</td>
</tr>
<tr>
<td>4, ( b = 2 )</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>70.8</td>
<td>85.6</td>
<td>52.0</td>
<td>73.8</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>99.6</td>
<td>100</td>
<td>99.4</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>100</td>
<td>99.4</td>
<td>99.6</td>
<td>41.2</td>
<td>65.0</td>
<td>22.2</td>
<td>46.0</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>96.0</td>
<td>98.6</td>
<td>93.4</td>
<td>98.0</td>
</tr>
<tr>
<td>5, ( b = 1 )</td>
<td>100</td>
<td>100</td>
<td>96.0</td>
<td>97.8</td>
<td>4.2</td>
<td>10.4</td>
<td>0.6</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>9.6</td>
<td>21.8</td>
<td>8.6</td>
<td>19.4</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>100</td>
<td>99.4</td>
<td>99.8</td>
<td>3.8</td>
<td>9.6</td>
<td>0.6</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>8.0</td>
<td>19.8</td>
<td>7.4</td>
<td>17.4</td>
</tr>
<tr>
<td>5, ( b = 2 )</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>6.0</td>
<td>16.2</td>
<td>1.4</td>
<td>7.2</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>29.4</td>
<td>48.6</td>
<td>26.8</td>
<td>42.0</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>5.4</td>
<td>14.4</td>
<td>1.2</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>28.8</td>
<td>46.0</td>
<td>25.8</td>
<td>41.2</td>
</tr>
<tr>
<td>6, ( b = 1 )</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>85.2</td>
<td>94.8</td>
<td>67.0</td>
<td>88.6</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>9.6</td>
<td>21.8</td>
<td>8.6</td>
<td>19.4</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>100</td>
<td>99.4</td>
<td>99.8</td>
<td>3.8</td>
<td>9.6</td>
<td>0.6</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>8.0</td>
<td>19.8</td>
<td>7.4</td>
<td>17.4</td>
</tr>
<tr>
<td>6, ( b = 2 )</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>6.0</td>
<td>16.2</td>
<td>1.4</td>
<td>7.2</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>29.4</td>
<td>48.6</td>
<td>26.8</td>
<td>42.0</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>5.4</td>
<td>14.4</td>
<td>1.2</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>28.8</td>
<td>46.0</td>
<td>25.8</td>
<td>41.2</td>
</tr>
<tr>
<td>7, ( b = 1 )</td>
<td>100</td>
<td>100</td>
<td>95.0</td>
<td>97.8</td>
<td>11.8</td>
<td>19.6</td>
<td>0.2</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>23.6</td>
<td>36.0</td>
<td>2.0</td>
<td>4.6</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>100</td>
<td>99.4</td>
<td>99.4</td>
<td>10.8</td>
<td>19.8</td>
<td>0.2</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>23.6</td>
<td>37.0</td>
<td>2.0</td>
<td>4.8</td>
</tr>
<tr>
<td>7, ( b = 2 )</td>
<td>100</td>
<td>100</td>
<td>97.4</td>
<td>98.6</td>
<td>19.6</td>
<td>33.0</td>
<td>0.2</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>43.4</td>
<td>52.8</td>
<td>1.6</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>100</td>
<td>99.8</td>
<td>99.8</td>
<td>19.4</td>
<td>33.0</td>
<td>0.2</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>43.2</td>
<td>53.8</td>
<td>1.6</td>
<td>4.4</td>
</tr>
<tr>
<td>8, ( b = 1 )</td>
<td>100</td>
<td>100</td>
<td>92.2</td>
<td>96.0</td>
<td>82.0</td>
<td>94.8</td>
<td>66.2</td>
<td>88.0</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</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>94.2</td>
<td>96.8</td>
<td>78.0</td>
<td>93.6</td>
<td>69.2</td>
<td>87.0</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>100</td>
<td>100</td>
<td>99.4</td>
<td>99.8</td>
<td>99.8</td>
<td>99.8</td>
</tr>
<tr>
<td>8, ( b = 2 )</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>86.8</td>
<td>96.8</td>
<td>67.6</td>
<td>87.6</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>200</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>82.6</td>
<td>94.4</td>
<td>68.8</td>
<td>87.2</td>
</tr>
</tbody>
</table>

Note: See the note in Table 3.
Table 5: 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>(\alpha, \beta, \gamma)</td>
<td>(\theta_1, \theta_2)</td>
<td>(\rho_1, \rho_2)</td>
<td>(\nu)</td>
</tr>
</tbody>
</table>

Note: See note in Table 1.

Table 6: Tests of structural changes in the U.S. economy

<table>
<thead>
<tr>
<th>(J_{NT})</th>
<th>5%</th>
<th>10%</th>
<th>sup-LM</th>
<th>sup-Wald</th>
<th>5%</th>
<th>10%</th>
<th>sup-LM</th>
<th>sup-Wald</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R = 1)</td>
<td>5.40</td>
<td>2.82</td>
<td>2.12</td>
<td>7.03</td>
<td>7.51</td>
<td>8.85</td>
<td>7.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R = 2)</td>
<td>23.90</td>
<td>10.94</td>
<td>9.84</td>
<td>13.04</td>
<td>13.61</td>
<td>14.15</td>
<td>12.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R = 3)</td>
<td>31.48</td>
<td>16.35</td>
<td>15.30</td>
<td>17.05</td>
<td>17.85</td>
<td>20.26</td>
<td>18.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R = 4)</td>
<td>30.44</td>
<td>23.14</td>
<td>22.43</td>
<td>24.31</td>
<td>24.22</td>
<td>27.03</td>
<td>26.62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R = 5)</td>
<td>35.50</td>
<td>26.20</td>
<td>25.65</td>
<td>31.79</td>
<td>31.12</td>
<td>35.06</td>
<td>32.51</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: (i) Under \(J_{NT}\) 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). Bold elements denote significance at the 5% nominal level.

Figure 1: Plots of estimated time invariant factor loadings (dashed line), time varying factor loadings (solid line) 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.
loadings may have 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 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.