On Time-Varying Factor Models: Estimation and Testing*

Liangjun Su^a and Xia Wang^b

^a School of Economics, Singapore Management University, Singapore

^b School of Economics and Management, University of Chinese Academy of Sciences, Beijing

December 3, 2016

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 and uniform convergence of the estimated factors and factor loadings in the standard large N and large T framework. We also propose a BIC-type information criterion to determine the number of factors, which can be used in models with either time-varying or time-invariant factor models. Based on the comparison between the estimates of the common components under the null hypothesis of no structural changes and those under the alternative, we propose a consistent test for structural changes in factor loadings. We establish the null distribution, the asymptotic local power property, and the consistency of our test. Simulations are conducted to evaluate both our nonparametric estimates and test statistic. We also apply our test to investigate Stock and Watson's (2009) U.S. macroeconomic data set and find strong evidence of structural changes in the factor loadings.

---

*The authors thank the co-editor Oliver Linton, an associate editor, and three anonymous referees for their constructive comments and suggestions. They also express their sincere appreciation to Serena Ng for discussions on the subject matter. They thank the conference and seminar participants at the Princeton-QUT-SJTU-SMU Frontiers in Econometrics Conference, the SNU Workshop on Advances in Microeconometrics, the International Symposium on Recent Developments in Econometric Theory with Applications in Xiamen University, UC San Diego, University of Washington in Seattle, and National University of Singapore for their valuable comments. Su gratefully acknowledges the Singapore Ministry of Education for Tier-2 Academic Research Fund (AcRF) under grant number MOE2012-T2-2-021 and the funding support provided by the Lee Kong Chian Fund for Excellence. Wang acknowledges financial supports from the National Science Foundation of China (No. 71401160) and the Ministry of Education of Humanities and Social Sciences Project of China (No. 14YJC790120). Address Correspondence to: Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903; E-mail: ljsu@smu.edu.sg, Phone: +65 6828 0386.

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 these variables and the latent common factors, are usually assumed to be fixed over a long period of time in the conventional factor models (e.g., Stock and Watson 2002, Bai and Ng 2002, Bai 2003). Stock and Watson (2002, 2009) argue that when the factor loadings undergo *small* instabilities, the factor estimates obtained via the conventional principal component analysis (PCA) are still consistent. However, since macroeconomic datasets typically span a long time period, it is restrictive to assume that the factor loadings are time-invariant or undergo negligible changes during the whole sampling period. In fact, there exist various driving forces such as institutional switching, economic transition, preference changes and technological progress that may influence the relationship between random variables significantly. By ignoring potentially significant structural changes in factor loadings, the estimated common factors might not converge to the desired object and forecasting and inference based on them can be misleading or unreliable. In addition, even if one concerns only the common component, which is equal to the product of factor loadings and the common factors, one may get misleading results.

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 to test for structural breaks in factor loadings based on the idea of Andrews (1993). Chen et al. (2014) propose a two-stage procedure to detect *big* breaks in factor loadings by testing the parameter stability in the regression of one of the estimated factors on the remaining estimated factors. Corradi and Swanson (2014) propose a test to check structural stability of both factor loadings and factor-augmented forecasting regression coefficients. Han and Inoue (2015) propose a test for structural breaks in the factor loadings by checking whether the second moments of the estimated factors exhibit structural changes. Yamamoto and Tanaka (2015) propose a modified version of Breitung and Eickmeier's (2011) test to ensure that it is robust to the non-monotonic power problem.

Cheng et al. (2016) consider the case where both the factor loadings and the number of factors may change simultaneously at a time point. These studies provide appropriate econometric tools to examine the problem of structural breaks in factor loadings. However, all these papers focus on the case of one-time abrupt structural changes. The analyses may be inappropriate if, for example, such driving forces of structural changes as preference changes, technological progress and policy changes, play a role gradually over a long period of time, or some abrupt policy changes also take a period of time to take effect. Indeed, as Hansen (2001) points out, “it may seem unlikely that a structural break could be immediate and might seem more reasonable to allow a structural change to take a period of time to take effect”. Hence, it seems more realistic to assume smooth changes rather than abrupt changes. To the best of our knowledge, Bates et al. (2013) is the only paper that allows for smooth changes in factor loading. By controlling the magnitude of instabilities to be “small”, they show that the principal component (PC) estimators of factors are still consistent. In fact, changes in comovement induced by technological progress and other forces are gradual but fundamental. As a result, we can neither assume the structural changes to be negligible nor check the instabilities of factor loadings under the framework of abrupt structural changes.

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

More importantly, we propose an L_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 unknown

finite number in our setup. Unlike the existing tests, such as Breitung and Eickmeier (2011), Chen et al. (2014), Han and Inoue (2015), and Yamamoto and Tanaka (2015), which check the stability of the moments of factor loadings or common factors, our test compares the estimates of the common components because it is well known that the latent factors and the factor loadings are not 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), Han and Inoue (2015), and Yamamoto and Tanaka (2015), no trimming of the boundary regions near the starting or ending of period is required for our test. In other words, we allow the breaks to occur near the beginning and the ending of the sample under the alternative.

After the first version of the paper was circulated, we found that Motta et al. (2011, MHV hereafter) consider a similar model to ours. They propose a time-varying PC regression (PCR) estimator by approximating the kernel-weighted objective function by something that can be estimated via the usual PC procedure. They establish the consistency of Rodríguez-Poo and Linton's (2001) nonparametric estimator of the time-varying covariance matrix $\text{Var}(X_t)$ when the elements of $X_t = (X_{1t}, \dots, X_{Nt})'$ exhibit the time varying factor structure in (2.1) below. Based on this, they also establish the convergence rates for their PCR estimates of the factors, factor loadings, and the common components. Nevertheless, MHV do not study the asymptotic distributions of their estimators. Neither do they consider a specification test for the widely used time-invariant factor models.

The rest of this paper is organized as follows. In Section 2, we introduce our factor models with time-varying factor loadings. In Section 3, we propose the local PCA procedure and develop the asymptotic normality for the estimated common factors and factor loadings. In Section 4, we construct our test statistic for time-varying factor loadings, derive the asymptotic distribution of our test and investigate the asymptotic power properties. In Section 5, we study the finite sample performance of our estimation and test via simulation. Section 6 provides an empirical study. Section 7 concludes. All proofs are relegated to the online supplementary appendix.

NOTATION. For an $m \times n$ real matrix A , we denote its transpose as A' , its Frobenius norm as $\|A\|$ ($\equiv [\text{tr}(AA')]^{1/2}$), its spectral norm as $\|A\|_{\text{sp}}$ ($\equiv \sqrt{\mu_1(A'A)}$) 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 $\|A\|_{\text{sp}} \leq \|A\| \leq \|A\|_{\text{sp}} \text{rank}(A)^{1/2}$. We also use $\mu_{\max}(B)$ and $\mu_{\min}(B)$ 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 $M_A \equiv \mathbb{I}_m - P_A$, where $\mathbb{I}_m$ denotes an $m \times m$ identity matrix. The operator $\xrightarrow{P}$ denotes convergence in probability, $\xrightarrow{d}$ 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 $\{X_{it}, i = 1, 2, \dots, N; t = 1, 2, \dots, T\}$ be an N -dimensional time series with T observations. The index i represents the i th cross sectional unit in a panel data set or the i th random variable in a multiple time series data set. We assume that X_{it} admits the following time-varying factor model with R latent common factors $F_t = (F_{1t}, \dots, F_{Rt})'$:

$$X_{it} = \lambda'_{it} F_t + e_{it}, \quad (2.1)$$

where the idiosyncratic error $\{e_{it}\}$ is assumed to be weakly dependent over both cross sectional unit i and time period t .

Our model given by (2.1) generalizes Stock and Watson's (2002) and Bai's (2003) factor models by allowing for structural changes in factor loadings.¹ To allow the model to capture various kinds of time-varying factor loadings, we follow the nonparametric literature on time-varying models (e.g., Cai, 2007; Robinson, 2012; Chen and Hong, 2012; Chen et al. 2012; Su et al., 2016) and model λ_{it} as a nonrandom function of t/T :

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

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).

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

$$X_{it} = m(Z_{it})' F_t + e_{it},$$

where Z_{it} is a $d \times 1$ vector of observable random variables, $m(\cdot)$ is a smooth function, the first element of factor vector F_t is given by 1 and the other elements of F_t is unobserved. They apply

---

¹Stock 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 ζ_{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 Otrok (2009) propose a dynamic factor model with time-varying factor loadings and stochastic volatility in both the latent factors and idiosyncratic components, and estimate the model via a Gibbs sampling procedure.

B-splines to approximate the unknown coefficient function $m(\cdot)$ and propose a Newton-Raphson algorithm to estimate the factors and factor loadings. Compared with Park et al.'s (2009) model, our model does not rely on the availability of Z_{it} and hence is more flexible to capture various kinds of smooth structural changes in factor loadings.

As in the conventional factor models, λ_{it} and F_t are not separately identifiable. At each time point t , we have $\lambda'_{it}F_t = (H_t^{-1}\lambda_{it})'(H'_tF_t)$ for any $R \times R$ nonsingular matrix H_t and we need to impose R^2 restrictions in order to identify λ_{it} and F_t . Let $\Lambda_t = (\lambda_{1t}, \dots, \lambda_{Nt})'$ and $F = (F_1, \dots, F_T)'$. One set of identification conditions would be $F'F/T = \mathbb{I}_R$ and $\Lambda'_t\Lambda_t/N$ a diagonal matrix with descending diagonal elements (c.f. Bai and Ng (2002) and Bai (2003)).

3 Estimation

In this section, we introduce the local version of PCA to estimate the time-varying factor loadings λ_{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 $r \in \{1, 2, \dots, T\}$. Under the assumption that $\lambda_i : [0, 1] \rightarrow \mathbb{R}$ 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

$$X_{it} \approx \lambda'_{ir}F_t + e_{it} \text{ when } \frac{t}{T} \approx \frac{r}{T}. \quad (3.1)$$

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

$$\min_{\{\lambda_{ir}\}_{i=1}^N, \{F_t\}_{t=1}^T} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T [X_{it} - \lambda'_{ir}F_t]^2 K_h\left(\frac{t-r}{T}\right) \quad (3.2)$$

subject to certain identification restrictions to be specified later on. Here, $K_h(x) = h^{-1}K(x/h)$, $K : \mathbb{R} \rightarrow \mathbb{R}^+$ is a kernel function, and $h = h(T, N)$ is a bandwidth parameter. 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

$$k_{h,tr}^{1/2}X_{it} \approx k_{h,tr}^{1/2}\lambda'_{ir}F_t + k_{h,tr}^{1/2}e_{it} \text{ when } \frac{t}{T} \approx \frac{r}{T}. \quad (3.3)$$

Define the $T \times N$ matrices $X^{(r)} = (X_1^{(r)}, \dots, X_N^{(r)})$ and $e^{(r)} = (e_1^{(r)}, \dots, e_N^{(r)})$, where $X_i^{(r)} = (k_{h,1r}^{1/2}X_{i1}, \dots, k_{h,Tr}^{1/2}X_{iT})'$ and $e_i^{(r)} = (k_{h,1r}^{1/2}e_{i1}, \dots, k_{h,Tr}^{1/2}e_{iT})'$. Let $F^{(r)} = (k_{h,1r}^{1/2}F_1^0, \dots, k_{h,Tr}^{1/2}F_T^0)'$.

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

$$\min_{F^{(r)}, \Lambda^{(r)}} \text{tr} \left[\left(X^{(r)} - F^{(r)} \Lambda^{(r)'} \right) \left(X^{(r)} - F^{(r)} \Lambda^{(r)'} \right)' \right]. \quad (3.4)$$

We will consider the following identification restrictions:

$$F^{(r)'} F^{(r)} / T = \mathbb{I}_R \text{ and } \Lambda_r' \Lambda_r = \text{diagonal matrix}. \quad (3.5)$$

By concentrating out $\Lambda_r = X^{(r)'} F^{(r)} (F^{(r)'} F^{(r)})^{-1} = X^{(r)'} F^{(r)} / T$ under the restriction $F^{(r)'} F^{(r)} / T = \mathbb{I}_R$, the objective function becomes

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

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

Clearly, if the columns of $\hat{F}^{(r)} / \sqrt{T}$ are the eigenvectors associated with the $X^{(r)} X^{(r)'}$, so are the columns of $-\hat{F}^{(r)} / \sqrt{T}$. In fact, $\hat{F}_t^{(r)}$ is only able to identify the transformed factor $F_t^{(r)} = k_{h,tr}^{1/2} F_t$ up to the rotation (i.e., $H^{(r)'} F_t^{(r)}$ in Theorem 3.1 below) and up to sign (c.f., Rodríguez-Poo and Linton, 2001, p.1295); $\hat{\lambda}_{ir}$ is able to identify λ_{ir} up to the rotation (i.e., $H^{(r)-1} \lambda_{ir}$ Theorem 3.2 below) and up to sign. Similarly, the rotation matrix $H^{(r)}$ is identified up to sign. In practice, we recommend to determine the sign of $\hat{\Lambda}_r = (\hat{\lambda}_{1r}, \dots, \hat{\lambda}_{Nr})'$ by fixing the signs of its first row's elements; in simulations, we can enforce their consistency with the signs of the corresponding elements in λ_{1r} .

It is well known that a local constant estimator may suffer from boundary bias problem when we estimate a fixed conditional mean object. As we shall see, because the factors and factor loadings are not separately identified in our model, we are only interested in estimating a rotational version of them and our nonparametric kernel estimators for either object does not have the usual boundary bias issue. Similar phenomenon occurs in MHV when they consider their PCR estimator of factors and factor loadings in a locally stationary model. Despite this observation, we find that the use of a boundary kernel helps us to obtain some uniform results. Specifically, suppose that the true factor F_t^0 satisfies $E(F_t^0 F_t^{0'}) = \Sigma_F$ and the kernel function $K(\cdot)$ is continuously differentiable

probability density function (PDF) with compact support $[-1, 1]$. We frequently call upon the uniform order of the following object

$$\Delta(r) \equiv \frac{1}{Th} \sum_{t=1}^T K\left(\frac{t-r}{Th}\right) E(F_t^0 F_t^{0'}) - \Sigma_F.$$

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

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

Note that $k_{h,tr}^*$ coincides with $k_{h,tr}$ in the interior region but not in the boundary regions. By using this boundary kernel to replace $k_{h,tr} = K_h\left(\frac{t-r}{T}\right)$ 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 estimators of λ_{it} 's obtained in the first stage, we can obtain the consistent estimators of F_t , $t = 1, 2, \dots, T$, in the second stage, by considering the following least squares problem:

$$\min_{F_t \in \mathbb{R}^R} \sum_{i=1}^N \left[X_{it} - \hat{\lambda}_{it}' F_t \right]^2, \quad t = 1, 2, \dots, T.$$

The solution to the above problem is: $\hat{F}_t = \left(\sum_{i=1}^N \hat{\lambda}_{it} \hat{\lambda}_{it}' \right)^{-1} \left(\sum_{i=1}^N \hat{\lambda}_{it} X_{it} \right)$ for $t = 1, 2, \dots, T$.

3.2 Limiting distributions of the estimated factors and factor loadings

In this subsection, we establish the asymptotic distributions of the estimated common factors and time-varying factor loadings. It is worth mentioning that the factors and factor loadings that appear in our assumptions denote the true values F_t^0 and λ_{it}^0 . But for notational simplicity, we suppress the superscript 0 hereafter.

Let $e_t = (e_{1t}, \dots, e_{Nt})$, $\gamma_N(s, t) = N^{-1}E(e'_s e_t)$, $\gamma_{N,F}(s, t) = N^{-1}E(F_s e'_s e_t)$, $\gamma_{N,FF}(s, t) = N^{-1}E(F_s e'_s \times e_t F'_t)$, and $\zeta_{st} = N^{-1}[e'_s e_t - E(e'_s e_t)]$. Define

$$\begin{aligned}\varpi_{NT,1}(r) &= \frac{h^{1/2}}{\sqrt{NT}} F^{(r)'} e^{(r)} \Lambda_r = \frac{h^{1/2}}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T k_{h,tr} F_t e_{it} \lambda'_{ir}, \\ \varpi_{NT,2}(r, t) &= \frac{h^{1/2}}{\sqrt{NT}} [F^{(r)'} e^{(r)} e_t - E(F^{(r)'} e^{(r)} e_t)] = \frac{h^{1/2}}{\sqrt{NT}} \sum_{s=1}^T \sum_{i=1}^N k_{h,sr} [F_s e_{is} e_{it} - E(F_s e_{is} e_{it})].\end{aligned}$$

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 $\max_{i,t} E(e_{it}^8) < \infty$.

(ii) $\max_t E\|F_t\|^8 < \infty$ and $E(F_t F'_t) = \Sigma_F > 0$ for some $R \times R$ matrix Σ_F .

(iii) λ_{it} are nonrandom such that $\max_{i,t} \|\lambda_{it}\| \leq \bar{c}_\lambda < \infty$ and $N^{-1} \Lambda'_r \Lambda_r = \Sigma_{\Lambda_r} + O(N^{-1/2})$ for some $R \times R$ positive definite matrix Σ_{Λ_r} and for all r .

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

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

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

(vii) $\varpi_{NT,1}(r) = O_P(1)$ 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) $N^{-1/2} \Lambda'_r e_t \xrightarrow{d} N(0, \Gamma_{rt})$ for each r, t , where $\Gamma_{rt} = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sum_{j=1}^N \lambda_{ir} \lambda'_{jr} E(e_{it} e_{jt})$.

(ii) $\frac{\sqrt{h}}{\sqrt{T}} \sum_{s=1}^T k_{h,sr} F_s e_{is} \xrightarrow{d} N(0, \Omega_{i,r})$, where

$$\Omega_{i,r} = \lim_{T \rightarrow \infty} \left[\frac{h}{T} \sum_{s=1}^T k_{h,sr}^2 E(F_s F'_s e_{is}^2) + \frac{2h}{T} \sum_{s=1}^{T-1} \sum_{t=s+1}^T k_{h,sr} k_{h,tr} E(F_s F'_t e_{is} e_{it}) \right]. \quad (3.6)$$

Assumption A.3 (i) The kernel function $K : \mathbb{R} \rightarrow \mathbb{R}^+$ is a symmetric continuously differentiable PDF function with compact support $[-1, 1]$.

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

A.1 mainly imposes moment conditions on the errors, factors, factor loadings, and their interactions. They are widely used in the literature; see, e.g., Bai and Ng (2002) and Bai (2003). In particular, A.1(i), (ii), (iii), and (viii) correspond to Assumptions C.1, A, B, and G in Bai (2003), respectively. The main difference is that we require $E(F_t F'_t) = \Sigma_F$ in A.1(ii) for the reasons to be explained below and $N^{-1} \Lambda'_r \Lambda_r = \Sigma_{\Lambda_r} + O(N^{-1/2})$ in A.1(iii) to simplify proofs in subsequent sections. Han and Inoue (2015) also assume the latter conditions. A.1(iv)-(vi) restrict the time and cross-sectional dependence for the idiosyncratic errors and the weak dependence between factors

and errors, which are in the same spirit as Assumptions C.2-5, D and E in Bai (2003). A.1(vii) is a kernel-weighted version of Assumptions F.1-F.2 in Bai (2003). Please note that Assumption F.1 in Bai (2003) can not be justified if F_s is random and $E\{F_s[e_{ks}e_{kt} - E(e_{ks}e_{kt})]\} \neq 0$, and one has to replace $F_s[e_{ks}e_{kt} - E(e_{ks}e_{kt})]$ by $F_se_{ks}e_{kt} - E(F_se_{ks}e_{kt})$ in their conditions. A similar remark holds for Assumption 6(a) in Han and Inoue (2015).

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

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

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

A.2(i) extends Assumption F.3 in Bai (2003) to allow for factor loadings to be time-varying and Assumption A.2(ii) is a kernel-weighted version of Assumption F in Bai (2003). Both parts are used to establish the asymptotic normality of our local PCA estimators and can be verified under some primitive conditions. For example, by the central limit theorem (CLT hereafter) for strong mixing processes (e.g., White 2001, Theorem 5.20), one can readily verify A.2(ii). Using Davydov inequality, we can argue that the limit $\Omega_{i,r}$ in (3.6) exists. Without further assumptions, we cannot simplify it. If $E(F_sF_s'e_{is}^2) = \Omega_i$ 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 $\{e_{i,t-1}, e_{i,t-2}, \dots, F_t, F_{t-1}, \dots\}$, then we can readily show that $\Omega_{i,r} = \Phi_i \lim_{T \rightarrow \infty} \frac{1}{Th} \sum_{s=1}^T K_r^*\left(\frac{s-r}{Th}\right)^2 = \Phi_i \int_{-1}^1 K(u)^2 du$ 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 $r = 1, 2, \dots, T$, 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(i) and A.3 hold. Then, for each $t = 1, 2, \dots, T$ and $r = 1, 2, \dots, T$ such that $|r - t| \leq Th$, we have:*

$$K_r^* \left(\frac{r - t}{Th} \right)^{-1/2} \sqrt{Nh} \left[\hat{F}_t^{(r)} - H^{(r)'} F_t^{(r)} \right] \xrightarrow{d} N \left(0, V_r^{-1} Q_r \Gamma_{rt} Q_r' V_r^{-1} \right),$$

where $H^{(r)} = (N^{-1} \Lambda_r' \Lambda_r) (T^{-1} F^{(r)'} \hat{F}^{(r)}) V_{NT}^{(r)-1}$, $V_{NT}^{(r)}$ denotes the $R \times R$ diagonal matrix of the first R largest eigenvalues of $(NT)^{-1} X^{(r)} X^{(r)'}$, V_r is the diagonal matrix consisting of the eigenvalues of $\Sigma_{\Lambda_r}^{1/2} \Sigma_F \Sigma_{\Lambda_r}^{1/2}$ in descending order with Υ_r being the corresponding (normalized) eigenvector matrix ($\Upsilon_r' \Upsilon_r = \mathbb{I}_R$), and $Q_r = V_r^{1/2} \Upsilon_r^{-1} \Sigma_{\Lambda_r}^{-1/2}$.

Theorem 3.1 establishes the asymptotic normality of $\hat{F}_t^{(r)}$ and it is a local version of Theorem 1 in Bai (2003). The latter theorem in Bai (2003) reports the asymptotic distribution of the factor estimator in a conventional time-invariant factor model under the identification restrictions: $F'F/T = \mathbb{I}_R$ and $\Lambda' \Lambda / N =$ a diagonal matrix with descending diagonal elements. Our Theorem 3.1 reports the asymptotic distribution of our local PCA estimator of the transformed factor under the local identification restrictions: $F^{(r)'} F^{(r)} / T = \mathbb{I}_R$ and $\Lambda_r' \Lambda_r / N =$ a diagonal matrix with descending diagonal elements. The major difference lies in two aspects. First, all objects that appear in our theorem is a local version of the corresponding objects in Bai (2003). Second, the convergence rate of our local PCA estimator $\hat{F}_t^{(r)}$ is $\sqrt{Nh}$ in contrast with the parametric $\sqrt{N}$ -rate in Theorem 1 of Bai (2003).

We note that $\hat{F}_t^{(r)}$ is a consistent estimator for the transformed latent factor $F_t^{(r)} = k_{h,tr}^{1/2} F_t$ 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 $\Gamma_{rt} = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \lambda_{ir} \lambda_{ir}' \sigma_{it}^2$ with $\sigma_{it}^2 = E[e_{it}^2]$. In addition, if $\sigma_{it}^2 = \sigma_t^2$ for each i , then we have $\Gamma_{rt} = \Sigma_{\Lambda_r} \sigma_t^2$.

The asymptotic distribution of $\hat{\lambda}_{it}$ is reported in the next theorem.

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

$$\sqrt{Th} \left[\hat{\lambda}_{ir} - H^{(r)-1} \lambda_{ir} \right] \xrightarrow{d} N \left(0, (Q_r')^{-1} \Omega_{ir} Q_r^{-1} \right).$$

Theorem 3.2 establishes the asymptotic normality of $\hat{\lambda}_{ir}$ and is a local version of Theorem 2 in Bai (2003). The latter theorem in Bai (2003) reports the asymptotic distribution of the estimator of the factor loadings in the conventional time-invariant factor model under the aforementioned identification restrictions. Our Theorem 3.2 reports the asymptotic distribution of our local PCA estimator of the factor loadings under the corresponding local identification restrictions. As expected, all objects that appear in our theorem is a local version of the corresponding objects in

Bai (2003), and our local PCA estimator of the factor loadings converge to a rotational version of the true factor loadings at the nonparametric $\sqrt{Th}$ -rate, in contrast with the parametric $\sqrt{T}$ -rate in Theorem 2 of Bai (2003).

When $\{e_{it}, \mathcal{F}_{it}\}$ is an m.d.s., the asymptotic variance can be simplified, leading to

$$\sqrt{Th} \left(\hat{\lambda}_{ir} - H^{(r)-1} \lambda_{ir} \right) \xrightarrow{d} N \left(0, \int_{-1}^1 K(u)^2 du (Q_r^{-1})' \Omega_i Q_r^{-1} \right)$$

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 $t = 1, 2, \dots, T$ we have*

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

Theorem 3.3 reports the asymptotic distribution of the second-stage estimator of the factor. This result is not directly comparable with any result in Bai (2003) because no second-stage refitting is needed for the conventional time-invariant factor models. Interestingly, although the convergence rates of $\hat{F}_t^{(r)}$ 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 squared error of the nonparametric estimates. In practice, we suggest using some data-driven methods to choose the bandwidth. For example, one can use the cross-validation method to choose the bandwidth $\hat{h}$ by solving the following minimization problem:

$$\min_h CV(h) = \frac{1}{NT} \sum_{i=1}^N \sum_{r=1}^T \left[X_{ir} - \hat{\lambda}_{ir}^{(-r)'} \hat{F}_r^{(-r)} \right]^2,$$

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

On the surface, Theorem 3.3 does not appear very useful because the rotation matrix $H^{(t)}$ is t -dependent. This will create some problem that does not occur in the conventional time-invariant factor model. In the latter case, the factor estimator $\hat{F}_t$ is consistent with $H' F_t$ for a rotation

matrix H that does not depend on t , and we can continue to use $\hat{F}_t$ to do out-of-sample forecast as in a diffusion index model (e.g., Stock and Watson, 1998, 2002). For our time-varying factor model, the rotational matrix is t -dependent; see a similar result in Motta, Hafner, and von Sachs (2011, MHV). Nevertheless, the result in Theorem 3.3 continues to be useful in several important scenarios. First, if we are only interested in the common component, $\lambda'_{it}F_t$, the asymptotics for its estimate will not depend on any rotation matrix. See Theorem 3.4 below. Second, the rotation matrix does not play any essential role when we consider a specification test for the null hypothesis of time-invariant factor loadings. In the next section, we consider such a test based on the comparison of the estimate $\tilde{\lambda}'_{i0}\tilde{F}_t$ of the common component $\lambda'_{it}F_t$ under the null and that $(\hat{\lambda}'_{it}\hat{F}_t)$ under the alternative. Third, even if the rotation matrix $H^{(t)}$ is t -dependent, the estimates $\hat{F}_t$ can still be used for out-of-sample forecasting when the factors have time-varying coefficients. To see this, we observe that $H^{(t)}$ can be written as $H(t/T)$ when the factor loadings λ_{it} are smooth and written as $\lambda_i(t/T)$. For example, we can consider the following 1-step-ahead forecasting model

$$y_{t+1} = \alpha'_t F_t + \varepsilon_{t+1},$$

where $\alpha_t = \alpha(t/T)$ is time-varying. When $\hat{F}_t - H^{(t)'}F_t = o_P(1)$, we have

$$\begin{aligned} y_{t+1} &= \alpha'_t F_t + \varepsilon_{t+1} \\ &= \alpha'_t \left(H^{(t)-1} \right)' \hat{F}_t + \varepsilon_{t+1} + \alpha'_t \left(H^{(t)-1} \right)' \left[H^{(t)'} F_t - \hat{F}_t \right] \\ &= \alpha'_t \left(\bar{H}^{(t)-1} \right)' \hat{F}_t + \varepsilon_{t+1} + o_P(1) \\ &= \bar{\alpha}(t/T)' \hat{F}_t + \varepsilon_{t+1} + o_P(1), \end{aligned}$$

where $\bar{H}^{(t)} = \bar{H}(t/T)$ is the probability limit of $H^{(t)} = H(t/T)$ and $\bar{\alpha}(t/T) \equiv \bar{H}(t/T)^{-1} \alpha(t/T)$ can be consistently estimated by regressing y_{t+1} on $\hat{F}_t$. Then a one-step-ahead forecast can be made for y_{t+1} based on the consistent estimate of $\bar{\alpha}(t/T)$, say by the local linear kernel estimation. A rigorous study of such a forecasting model is left for future research.

Following Bai (2003), we can also derive the limit theory of the estimated common components from Theorems 3.2-3.3. Define

$$C_{it}^0 = \lambda'_{it}F_t \text{ and } \hat{C}_{it} = \hat{\lambda}'_{it}\hat{F}_t.$$

The following theorem studies the asymptotic distribution of $\hat{C}_{it}$.

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

$$\left(\frac{1}{N} V_{1it} + \frac{1}{Th} V_{2it} \right)^{-1/2} \left(\hat{C}_{it} - C_{it}^0 \right) \xrightarrow{d} N(0, 1),$$

where $V_{1it} = \lambda'_{it} \Sigma_{\Lambda_t}^{-1} \Gamma_{tt} \Sigma_{\Lambda_t}^{-1} \lambda_{it}$ and $V_{2it} = F_t' \Sigma_F^{-1} \Omega_{i,t} \Sigma_F^{-1} F_t$.

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

$$\frac{C_{NT}}{\left(\frac{C_{NT}^2}{N}V_{1it} + \frac{C_{NT}^2}{Th}V_{2it}\right)^{1/2}} \left(\hat{C}_{it} - C_{it}^0\right) \xrightarrow{d} N(0, 1)$$

and the denominator in the above expression is bounded from both above and below, we can conclude that the convergence rate of $\hat{C}_{it}$ is given by C_{NT} . This rate is different from $\min\{\sqrt{N}, \sqrt{T}\}$ -convergence rate for the estimator of the common component in the conventional time-invariant factor model. To make inference, it is standard to obtain the consistent estimates of V_{1it} and V_{2it} . To save space, we omit the details.

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

Assumption A.4. (i) $\|e\|_{\text{sp}} = O_P(N^{1/2} + T^{1/2})$ and $\max_{t,s} \left| \frac{1}{N} \sum_{i=1}^N [e_{it}e_{is} - E(e_{it}e_{is})] \right| = O_P(N^{-1/2}(\ln T)^{1/2})$.

(ii) $\max_r \left| \frac{1}{N} \Lambda_r' \Lambda_r - \Sigma_{\Lambda_r} \right| = o(1)$, and the eigenvalues of Σ_{Λ_r} are bounded below from 0 and above from infinity uniformly in r .

(iii) $\max_{i,r} \left\| \frac{1}{T} \sum_{t=1}^T k_{h,tr} F_t e_{it} \right\| = O_P((Th)^{-1/2}(\ln(NT))^{1/2})$, $\max_r \left\| \frac{1}{T} \sum_{t=1}^T k_{h,tr} (F_t F_t' - \Sigma_F) \right\| = O_P(T^{-1/2}(\ln T)^{1/2})$, and $\max_r \left| \frac{1}{T} \sum_{t=1}^T k_{h,tr} [\|F_t\| - E\|F_t\|] \right| = O_P(T^{-1/2}(\ln T)^{1/2})$.

(iv) $\max_{s,t} \|N^{-1} \Lambda_s' e_t\| = O_P(N^{-1/2}(\ln T)^{1/2})$ and $\max_s \|N^{-1} \Lambda_s' e_s F_s\| = O_P(N^{-1/2}(\ln T)^{1/2})$.

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

(vi) $\max_r \left\| \frac{1}{NT} \sum_{t=1}^T \Lambda_r' e_t^{(r)} F_t^{(r)'} \right\| = \max_r \left\| \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T k_{h,tr} \lambda_{ir} e_{it} F_t' \right\| = O_P((NTh)^{-1/2}(\ln T)^{1/2})$.

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

The following theorem studies the uniform convergence of $\hat{\lambda}_{ir}$, $\hat{F}_t$, and $\hat{C}_{it}$.

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

- (i) $\max_{i,t} \left\| \hat{\lambda}_{it} - H^{(t)-1} \lambda_{it} \right\| = O_P \left((Th/\ln T)^{-1/2} \right),$
- (ii) $\max_t \left\| \hat{F}_t - H^{(t)'} F_t \right\| = O_P \left((N/\ln T)^{-1/2} \right),$
- (iii) $\max_{i,t} \left| \hat{C}_{it} - C_{it}^0 \right| = O_P \left((N/\ln T)^{-1/2} \right) + o_P \left((Th/\ln T)^{-1/2} T^{1/8} \right).$

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

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 $R_{\max}$. 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

$$N^{-1} \Lambda_r' \Lambda_r = \mathbb{I}_R \text{ and } T^{-1} F^{(r)'} F^{(r)} \text{ is a diagonal matrix with descending diagonal elements,} \quad (3.7)$$

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

$$V \left(R, \left\{ \check{\Lambda}_r^{(R)} \right\} \right) = \min_{\check{F} = (\check{F}_1', \dots, \check{F}_T')'} \frac{1}{NT} \sum_{i=1}^N \sum_{r=1}^T \left[X_{ir} - \check{F}_r' \check{\lambda}_{ir}(R) \right]^2. \quad (3.8)$$

As Bai and Ng (2002, p.199) remark, the above sum of squared residuals (SSR) does not depend on which estimate of the factor loadings is used because they span the same vector space for any localization point r . We use the normalization rule in (3.7) instead of the one used before because we find that in this case it is relatively easier to study the asymptotic properties of

$\hat{\Lambda}_r^{(R)}$, $\check{\Lambda}_r^{(R)}$, and the associated rotational matrix $H^{(r,R)}$ (see Lemma A.7 for the definition) than using the previous rule, regardless of the value of R relative to the true value R_0 .² We define $\check{\Lambda}_r^R = (NT)^{-1} X^{r'} X^r \hat{\Lambda}_r^R$ and use it to obtain for $V(\cdot, \cdot)$ for two reasons. First, $\hat{\Lambda}_r^R$ always has full rank R because $N^{-1} \hat{\Lambda}_r^{R'} \hat{\Lambda}_r^R = \mathbb{I}_R$ by normalization. In contrast, the asymptotic rank of $\check{\Lambda}_r^R$ is given by $\min(R, R_0)$ (see Lemma ??) and hence $\check{\Lambda}_r^R$ is informative on R_0 when $R > R_0$. Second, we study the asymptotic behavior of $\hat{\Lambda}_r^R$ in Lemma ?? via the study of $\check{\Lambda}_r^R$. Note that Lu and Su (2016) apply a similar trick in their shrinkage estimation of panel data models with a multi-factor error structure.

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

$$IC(R) = \ln V\left(R, \left\{\check{\Lambda}_r^{(R)}\right\}\right) + \rho_{NT} R, \quad (3.9)$$

where ρ_{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.5. (i) $\max_{s,t} E \left\| N^{-1/2} \Lambda'_s e_t F'_t \right\|^4 \leq C$ and $\max_{s,t} E \left\| N^{-1/2} [F_s e'_s e_t F'_t - E(F_s e'_s e_t F'_t)] \right\|^2 \leq C$.

(ii) $\max_r E \left\| \frac{h^{1/2}}{(NT)^{1/2}} \sum_{i=1}^N \sum_{t=1}^T k_{h,tr} [F_t e_{it} e_{ir} F'_r - E(F_t e_{it} e_{ir} F'_r)] \right\|^2 \leq C$.

Assumption A.6. As $(N, T) \rightarrow \infty$, $\rho_{NT} \rightarrow 0$ and $\rho_{NT} C_{NT}^2 \rightarrow \infty$ where $C_{NT} = \min(\sqrt{Th}, \sqrt{N})$.

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

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

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

Theorem 3.6 indicates the class of information criteria defined by $IC(R)$ in (3.9) can consistently estimate R_0 . To implement the information criterion, one needs to choose the penalty coefficient ρ_{NT} . Following the lead of Bai and Ng (2002), we suggest setting $\rho_{NT} = \frac{N+Th}{NTh} \ln\left(\frac{NTh}{N+Th}\right)$ or $\rho_{NT} = \frac{N+Th}{NTh} \ln C_{NT}^2$ with $C_{NT} = \min\{\sqrt{Th}, \sqrt{N}\}$ and evaluate the performance of these two information criteria in our simulation studies.

---

²By using the normalization rule in (3.7), we only need to study the asymptotic properties $\check{\Lambda}_r^{(R)}$ and $\hat{\Lambda}_r^{(R)}$ but not those of the first-stage factor estimators. This greatly facilitates the proofs of Lemmas A.7-A.9 used in the proof of Theorem 3.4 below.

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

$$\mathbb{H}_0 : \lambda_{it} = \lambda_{i0} \text{ for } i = 1, 2, \dots, N \text{ and } t = 1, 2, \dots, T, \quad (4.1)$$

and the alternative hypothesis is

$$\mathbb{H}_1 : \lambda_{it} \neq \lambda_{i0} \text{ for some } (i, t), \quad (4.2)$$

where λ_{i0} is an unknown vector of factor loadings. We allow $\lambda_{it} = \lambda_i(t/T)$ to be a piece-wise smooth function on $(0, 1]$ for each i with a finite number of discontinuities under $\mathbb{H}_1$.

Obviously, under the null hypothesis, λ_{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. Motivated by the potential presence of structural changes in factor models, Breitung and Eickmeier (2011), Chen et al. (2014), Han and Inoue (2015), and Yamamoto and Tanaka (2015) propose various tests for the existence of a single structural change in the factor loadings. In contrast, we do not impose any essential restriction on the alternative. The alternative (4.2) allows for a finite number of abrupt structural breaks. More importantly, by assuming λ_{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 $\mathbb{H}_0$, 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 $\mathbb{H}_1$, 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 $\mathbb{H}_0$ by measuring the squared distance between the estimates of the common components under $\mathbb{H}_0$ and those under $\mathbb{H}_1$.

Let $e_{it}^\dagger = e_{it} + (\lambda_{it} - \lambda_{i0})' F_t$. Let $X_t \equiv (X_{1t}, \dots, X_{Nt})'$, $e_t \equiv (e_{1t}, \dots, e_{Nt})'$, $e_t^\dagger \equiv (e_{1t}^\dagger, \dots, e_{Nt}^\dagger)'$, $F \equiv (F_1, \dots, F_T)'$, and $\Lambda_0 \equiv (\lambda_{10}, \dots, \lambda_{N0})'$. Let $X = (X_1', \dots, X_T')'$, $e \equiv (e_1', \dots, e_T')'$, $e^\dagger \equiv$

$(e_1^\dagger, \dots, e_T^\dagger)'$. Then we can rewrite (2.1) in matrix form

$$X = F\Lambda'_0 + e^\dagger. \quad (4.3)$$

The conventional PCA method solves the following minimization problem

$$\min_{F, \Lambda} \text{tr} (X - F\Lambda') (X - F\Lambda')' = \sum_{i=1}^N \sum_{t=1}^T (X_{it} - \lambda'_i F_t)^2$$

under certain identification restrictions. In this paper, we follow Bai (2003) and consider the following identification restrictions: $T^{-1}F'F = \mathbb{I}_R$ and $\Lambda'\Lambda$ is a diagonal matrix. Let $\tilde{F}_t$ and $\tilde{\lambda}_{i0}$ be the principal component estimators of F_t and λ_{i0} , respectively under the above identification restrictions. Let $\tilde{F} = (\tilde{F}_1, \dots, \tilde{F}_T)'$ and $\tilde{\Lambda}_0 = (\tilde{\lambda}_{1,0}, \dots, \tilde{\lambda}_{N,0})'$. 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 $\tilde{\Lambda}'_0 = (\tilde{F}'\tilde{F})^{-1}\tilde{F}'X = T^{-1}\tilde{F}X$.

Given the estimates $\tilde{\lambda}'_{i0}\tilde{F}_t$ of the common components $\lambda'_{it}F_t$ under $\mathbb{H}_0$ and those $(\hat{\lambda}'_{it}\hat{F}_t)$ under $\mathbb{H}_1$, 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:

$$\hat{M} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left(\hat{\lambda}'_{it}\hat{F}_t - \tilde{\lambda}'_{i0}\tilde{F}_t \right)^2. \quad (4.4)$$

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 $\mathbb{H}_0$. Let $\|A\|_q = \{E \|A\|^q\}^{1/q}$ for $q \geq 1$. We add the following assumptions.

Assumption A.7. (i) For each $i = 1, 2, \dots, N$, the process $\{e_{it}, t = 1, 2, \dots\}$ is an m.d.s. such that $E(e_{it}|\mathcal{F}_{NT,t-1}) = 0 \forall t$, where $\mathcal{F}_{NT,t-1} = \{F_t, F_{t-1}, \dots, e_{t-1}, e_{t-2}, \dots\}$.

(ii) For each $i = 1, 2, \dots, N$, the process $\{(e_{it}, F_t), t = 1, 2, \dots\}$ is strong mixing with mixing coefficients $\alpha_i(\cdot)$. $\alpha(\cdot) \equiv \max_i \alpha_i(\cdot)$ satisfies $\sum_{s=1}^\infty \alpha(s)^{\delta/(2+\delta)} \leq C < \infty$ for some $\delta > 0$. In addition, there exists an integer $T_0 \in [1, T)$ such that $T^{-2} \max(T_0^4, T_0^3 h^{-1}, T_0^2 h^{-2}) \rightarrow 0$ and $N^2 T h^2 \alpha(T_0)^{\delta/(1+\delta)} \rightarrow 0$ as $(N, T) \rightarrow \infty$.

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

Assumption A.7 are new in the literature on testing for structural breaks in large dimensional factor models. Previous authors only consider testing for the presence of a one-time structural

break under the alternative. To avoid the comparison between two large-dimensional factor loadings matrices, they reduce the *infinite-dimensional* problem to a *finite-dimensional* one in different ways. For example, Chen et al. (2014) run the regression of one estimated factor on the remaining ones and then test for the structural changes in such a linear regression by constructing the sup-Wald and sup-LM statistics of Andrews (1993); Han and Inoue (2015) construct their sup-Wald and sup-LM statistic by comparing the pre- and post- break subsample second moments of the estimated factors. In either case, the test statistics have the same asymptotic distribution as the conventional sup-Wald statistic of Andrews (1993). In contrast, we are dealing with the *infinite dimensional* parameter problem directly through the construction of an L_2 -distance statistic in (4.4). Our test is a nonparametric test and we have to analyze much more complicated objects than those in Chen et al. (2014) and Han and Inoue (2015). This explains why our assumptions are also distinct from those in the early literature.

A.7(i) assumes that the process $\{e_{it}, t = 1, 2, \dots\}$ is an m.d.s. with respect to the filter $\{\mathcal{F}_{NT,t}\}$ and it allows cross-sectional dependence among the error terms. This assumption is essential for the establishment of the asymptotic distribution of our test statistic under the null hypothesis and a sequence of Pitman local alternatives. It is possible to allow for both serial dependence and cross-sectional dependence in the error terms. But that will substantially complicate the asymptotic analysis and we are not sure how to estimate the asymptotic variance of our raw test statistic in this case. A.7(ii) requires the process $\{(e_{it}, F_t), t = 1, 2, \dots\}$ 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 $\limsup_N \frac{1}{N} \sum_{i=1}^N \sum_{s=1}^{\infty} \alpha_i(s)^{\delta/(2+\delta)} \leq C < \infty$. 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.7(ii), reflects the usual tradeoff between the dependence and moment conditions: a smaller value of δ requires faster decay in the mixing coefficients but less stringent moment conditions. Like A.1(vi), A.7(iv) controls the cross-sectional dependence among $\{F_t e_{it}, i = 1, 2, \dots, N\}$. Under A.7(iii), this condition becomes redundant if we would assume independence of $e_{i\cdot} = (e_{i1}, \dots, e_{iT})'$ across i conditional on the factors.

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

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

Let V_{NT} denote the $R \times R$ diagonal matrices of the first R largest eigenvalues of $(NT)^{-1} X X'$ in decreasing order and $H = (N^{-1} \Lambda_r' \Lambda_r)(T^{-1} F' \tilde{F}) V_{NT}^{-1}$. Let $k_{h,st} = h^{-1} K_t^* \left(\frac{s-t}{Th} \right)$, $\bar{k}_{st} = \bar{K} \left(\frac{s-t}{Th} \right)$ with $\bar{K}(u) = \int_{-1}^1 K(v) K(u-v) dv$ being the two-fold convolution kernel of $K(\cdot)$. For example, if we use the Epanechnikov kernel $K(u) = 0.75(1-u^2)\mathbf{1}\{|u| \leq 1\}$ with $\mathbf{1}\{\cdot\}$ being the usual indicator

function, then $\bar{K}(u) = \left(\frac{3}{5} - \frac{3}{4}u^2 + \frac{3}{8}|u|^3 - \frac{3}{160}|u|^5\right) \mathbf{1}\{|u| \leq 2\}$. Let $L_{st} = k_{h,st}H^{(t)}H^{(t)'} - HH'$. Define

$$\begin{aligned}\mathbb{B}_{NT} &= \frac{h^{1/2}}{N^{1/2}T^2} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T (F_t' L_{st} F_s)^2 e_{is}^2, \\ \mathbb{V}_{NT} &= 2T^{-2}N^{-1}h^{-1} \sum_{1 \leq r \neq s \leq T} \bar{k}_{sr}^2 E \left[(F_s' H_0 \bar{\Sigma}_F H_0' F_r)^2 (e_r' e_s)^2 \right],\end{aligned}$$

where $\bar{\Sigma}_F = H_0' \Sigma_F H_0$, $H_0 = Q^{-1} \equiv (V^{1/2} \Upsilon' \Sigma_{\Lambda_0}^{-1/2})^{-1}$ 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, Υ is the corresponding eigenvector matrix such that $\Upsilon' \Upsilon = \mathbb{I}_R$, and Σ_{Λ_0} is the probability limits of $N^{-1} \Lambda_r' \Lambda_r$ 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 (iv*), A.4, and A.7 hold. Then we have: $J_{NT} \equiv \mathbb{V}_{NT}^{-1/2} (TN^{1/2}h^{1/2}\hat{M} - \mathbb{B}_{NT}) \xrightarrow{d} N(0, 1)$ under $\mathbb{H}_0$.*

We make some remarks. First, each of the four terms, namely, $\hat{\lambda}_{it}$, $\hat{F}_t$, $\tilde{\lambda}_{i0}$, and $\tilde{F}_t$, in the definition of $\hat{M}$ 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 $\mathbb{H}_0$, $TN^{1/2}h^{1/2}\hat{M} - \mathbb{B}_{NT} = \sum_{s=2}^T Z_{NT,s} + o_P(1)$, where

$$Z_{NT,s} = 2T^{-1}N^{-1/2}h^{-1/2} \sum_{r=1}^{s-1} \sum_{i=1}^N \bar{k}_{sr} F_s' H_0 \bar{\Sigma}_F H_0' F_r e_{is} e_{ir}.$$

Under the m.d.s. condition in Assumption A.7(i), one can verify that $E(Z_{NT,s} | \mathcal{F}_{NT,s-1}) = 0$ and resort to a martingale CLT (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 $e_i = (e_{i1}, \dots, e_{iT})'$ conditional on the factors. The strong mixing condition in A.7(ii) and the moment conditions in A.7(iii)-(iv) greatly facilitate the verification of the Lyapunov condition. Third, despite the assumed m.d.s. condition, the variance $\mathbb{V}_{NT}$ 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 $\mathbb{B}_{NT}$ and the asymptotic variance $\mathbb{V}_{NT}$. The consistent estimators for $\mathbb{B}_{NT}$ and $\mathbb{V}_{NT}$ are respectively given by

$$\begin{aligned}\hat{\mathbb{B}}_{NT} &= \frac{h^{1/2}}{T^2 N^{1/2}} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T \left(k_{h,st} \hat{F}_s' \hat{F}_t - \tilde{F}_s' \tilde{F}_t \right)^2 \hat{e}_{is}^2, \text{ and} \\ \hat{\mathbb{V}}_{1NT} &= 2T^{-2}N^{-1}h^{-1} \sum_{1 \leq s \neq r \leq T} \bar{k}_{sr}^2 \left(\hat{F}_s' \hat{\Sigma}_F \hat{F}_r \right)^2 (\hat{e}_r' \hat{e}_s)^2,\end{aligned}$$

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

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

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

Theorem 4.2 *Suppose that Assumptions A.1, A.3(i) and (ii*), A.4, and A.7 hold. Then under $\mathbb{H}_0$, $\hat{\mathbb{B}}_{NT} = \mathbb{B}_{NT} + o_P(1)$, $\hat{\mathbb{V}}_{1NT} = \mathbb{V}_{1NT} + o_P(1)$, and $\hat{J}_{NT} \xrightarrow{d} N(0, 1)$.*

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_α , the upper α -percentile of the $N(0, 1)$ distribution, as the test is one-sided, and reject the null at α 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:

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

where $a_{NT} \rightarrow 0$ as $(N, T) \rightarrow \infty$, 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 λ_{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 λ_{i0} and $g_i(\cdot)$ instead of $\lambda_{i0,NT}$ and $g_{i,NT}(\cdot)$.

Let $g_{it} = g_i \left(\frac{t}{T} \right)$, $g_{it}^\dagger = F_t' g_i \left(\frac{t}{T} \right)$, and $g_t^\dagger = (g_{1t}^\dagger, \dots, g_{Nt}^\dagger)'$. Define

$$\begin{aligned} \Pi_1 &= \text{plim}_{(N,T) \rightarrow \infty} T^{-1} \sum_{t=1}^T \text{tr} \left[\left(N^{-1} \Lambda_0' g_t^\dagger \right) \left(N^{-1} g_t^\dagger \Lambda_0 \right) (H_0^{-1})' V_0^{-1} H_0^{-1} \Sigma_{\Lambda_0} (H_0^{-1})' V_0^{-1} H_0^{-1} \right], \\ \Pi_2 &= \lim_{(N,T) \rightarrow \infty} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \text{tr} (\Sigma_F g_{it} g_{it}'). \end{aligned} \quad (4.5)$$

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

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

- (ii) $\max_{1 \leq r \leq T} \left\| \frac{1}{NT} \sum_{s=1}^T \sum_{i=1}^N k_{h, sr} F_s e_{is} g'_{ir} \right\| = O_P((NTh/\ln(NT))^{-1/2})$.
- (iii) The limits Π_1 and Π_2 defined in (4.5) exist.

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

The following theorem studies the asymptotic local power property of $\hat{J}_{NT}$.

Theorem 4.3 *Suppose that Assumptions A.1, A.3(i) and (ii*), A.4, and A.7-A.8 hold. Then under $\mathbb{H}_1(a_{NT})$ with $a_{NT} = T^{-1/2}N^{-1/4}h^{-1/4}$, $\hat{\mathbb{B}}_{NT} = \mathbb{B}_{NT} + o_P(1)$, $\hat{\mathbb{V}}_{NT} = \mathbb{V}_{NT} + o_P(1)$, and $\hat{J}_{NT} \xrightarrow{d} N(\pi_0, 1)$, where $\pi_0 = (\Pi_1 + \Pi_2)/\mathbb{V}_0^{1/2}$ and $\mathbb{V}_0 = \lim_{(N,T) \rightarrow \infty} \mathbb{V}_{NT}$.*

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 N and T enter a_{NT} differently. An intuitive reason for this is that the cross sectional and time dimensions play distinct roles when i and t enter the null hypothesis $\mathbb{H}_0$ asymmetrically. Similar local alternative rates have been found in the literature. For example, Su and Chen (2013) consider testing slope homogeneity in linear panel data models with interactive fixed effects and they find that their LM-type statistic has nontrivial power against *parametric* local alternatives converging to the null at the rate $T^{-1/2}N^{-1/4}$. This rate is the same as that obtained by Pesaran and Yamagata (2008) for testing slope homogeneity in linear panel data models with the usual one-way error component. It is similar to our rate $T^{-1/2}N^{-1/4}h^{-1/4}$ for our *nonparametric* local alternative. Note that we allow the existence of a finite number of unknown 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.

In order for our test to have non-trivial power against the local alternatives, we need $\pi_0 > 0$. This requires that the factor loadings should not be time-varying only for an asymptotically negligible set of individuals. Let $\mathcal{N} = \{1, 2, \dots, N\}$. Let $|\cdot|$ denote the cardinality of a set $\cdot$. Define a subset of $\mathcal{N}$:

$$\mathcal{S}_N = \{i \in \mathcal{N} : \lambda_{it} = \lambda_{i0} \text{ for all } t\}.$$

Let $\mathcal{S}_N^c = \mathcal{N} \setminus \mathcal{S}_N$, the complement of $\mathcal{S}_N$ relative to $\mathcal{N}$. It is easy to verify that if $|\mathcal{S}_N^c|/N = o(1)$, then $\Pi_1 = \Pi_2 = \pi_0 = 0$ and thus our test does not have power in this case. Similar phenomenon occurs in Su and Chen's (2013) test for slope homogeneity where they require the degree of heterogeneity be sufficiently large. Similarly, the nonzero π_0 also requires that g_{it} should not be nonzero for an asymptotically negligible number of periods. Let $\mathcal{T} = \{1, 2, \dots, T\}$. For each $i \in \mathcal{N}$, define a subset of $\mathcal{T}$:

$$\mathcal{S}_{Ti} = \{t \in \mathcal{T} : g_{it} = 0\}.$$

Let $\mathcal{S}_{Ti}^c = T \setminus \mathcal{S}_{Ti}$. If $\max_{1 \leq i \leq N} |\mathcal{S}_{Ti}^c|/T = o(1)$, we can also verify that $\Pi_1 = \Pi_2 = \pi_0 = 0$, implying that our test does not have power in this case either. In general, as long as at least a fixed proportional of individuals $\mathcal{N}$ either undergo a one-time or multiple times of abrupt structural change, or undergo a nonshrinking proportion of T periods of smooth structural changes, $\pi_0 > 0$ and our test has asymptotic power to detect them.

4.5 Asymptotic global power

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

$$\mathbf{F}_T = \left\{ \check{F} : \check{F}'\check{F}/T = \mathbb{I}_R \right\} \text{ and } \mathbf{\Lambda}_N = \left\{ \check{\Lambda} : \check{\Lambda}'\check{\Lambda} = \text{diagonal matrix} \right\},$$

where $\check{\Lambda} = (\check{\lambda}_1, \dots, \check{\lambda}_N)'$ and $\check{F} = (\check{F}_1, \dots, \check{F}_T)'$.

Assumption A.9. There exists $\underline{\mathcal{L}}_{\Lambda F} > 0$ such that $\text{plim}_{(N,T) \rightarrow \infty} \inf_{(\check{\Lambda}, \check{F}) \in \mathbf{\Lambda}_N \times \mathbf{F}_T} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (\lambda'_{it} F_t - \check{\lambda}'_i \check{F}_t)^2 \geq \underline{\mathcal{L}}_{\Lambda F}$.

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

Theorem 4.4 *Suppose that Assumptions A.1, A.3-A.4 and A.9 hold. Then under the global alternative $\mathbb{H}_1$, $P(\hat{J} \geq c_{NT}) \rightarrow 1$ 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 $\hat{J}$ 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 $\hat{J}$ 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 $\tilde{\sigma}_i^2 = T^{-1} \sum_{t=1}^T \tilde{e}_{it}^2$, where $\tilde{e}_{it} = X_{it} - \tilde{\lambda}'_{i0} \tilde{F}_t$, and $\tilde{F}_t$ and $\tilde{\lambda}_{i0}$ are the estimates of the factors and factor loadings under the null. Let $e_{it}^* = \tilde{\sigma}_i \varsigma_{it}$ with ς_{it} being IID

$N(0, 1)$ over both i and t . Then one can generate the bootstrap resamples via $X_{it}^* = \tilde{\lambda}_{i0}' \tilde{F}_t + e_{it}^*$ 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 $\tilde{e}_t = (\tilde{e}_{1t}, \dots, \tilde{e}_{Nt})'$ and $\tilde{\Sigma}^0 = T^{-1} \sum_{t=1}^T \tilde{e}_t \tilde{e}_t'$. Let $\tilde{\sigma}_{ij}^0$ denote the (i, j) th element of $\tilde{\Sigma}^0$. Define the shrinkage version of $\tilde{\Sigma}^0$ as $\tilde{\Sigma}$ whose (i, j) th element is given by

$$\tilde{\sigma}_{ij} = \tilde{\sigma}_{ij}^0 (1 - \epsilon)^{|j-i|} \text{ for } i, j = 1, \dots, N,$$

where ϵ 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 $\max_{i,j} |\tilde{\sigma}_{ij}^0|$ is. By construction, $\tilde{\Sigma}$ is also symmetric and positive semi-definite. The stochastic boundedness of $\max_{i,j} |\tilde{\sigma}_{ij}^0|$ is sufficient but not necessary for the justification of the asymptotic validity of our bootstrap procedure below:

1. Estimate the restricted model $X_{it} = \lambda_{i0}' F_t + e_{it}^\dagger$ by the PCA method and the unrestricted model $X_{it} = \lambda_{it}' F_t + e_{it}$ by the local PCA method to obtain the two sets of estimates $\{\tilde{\lambda}_{i0}, \tilde{F}_t\}$ and $\{\hat{\lambda}_{it}, \hat{F}_t\}$. Based on these estimates, construct the test statistic $\hat{J}_{NT}$ as in Section 4.2.
2. For $i = 1, 2, \dots, N$ and $t = 1, 2, \dots, T$, obtain the bootstrap error $e_t^* = \tilde{\Sigma}^{1/2} \varsigma_t$, where $\varsigma_t = (\varsigma_{1t}, \dots, \varsigma_{Nt})'$ with ς_{it} being IID $N(0, 1)$ across i and t . Generate $X_{it}^* = \tilde{\lambda}_{i0}' \tilde{F}_t + e_{it}^*$.
3. Use $\{X_{it}^*\}$ to run the restricted and unrestricted models to obtain the bootstrap versions $\{\tilde{\lambda}_{i0}^*, \tilde{F}_t^*\}$ and $\{\hat{\lambda}_{it}^*, \hat{F}_t^*\}$ of $\{\tilde{\lambda}_{i0}, \tilde{F}_t\}$ and $\{\hat{\lambda}_{it}, \hat{F}_t\}$, respectively. Calculate the bootstrap test statistic $\hat{J}_{NT}^*$, the bootstrap version of $\hat{J}_{NT}$.
4. Repeat steps 2 and 3 for B times and index the bootstrap test statistics as $\{\hat{J}_{NT,l}^*\}_{l=1}^B$. The bootstrap p -value is calculated by $p^* \equiv B^{-1} \sum_{l=1}^B \mathbf{1}\{\hat{J}_{NT,l}^* > \hat{J}_{NT}\}$.

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

Theorem 4.5 *Suppose that Assumptions A.1, A.3(i) and (ii*), A.4, and A.7 hold. Suppose that (i) $\max_{i,j} |\tilde{\sigma}_{ij}^0| = O_P(\zeta_{NT})$ with $\zeta_{NT} = o(T^{1/2})$, (ii) $\frac{1}{T} \sum_{t=1}^T \|\tilde{F}_t\|^8 = O_P(1)$ and (iii) $\frac{1}{N} \sum_{i=1}^N \|\tilde{\lambda}_{i0}\|^8 = O_P(1)$. Then $\hat{J}_{NT}^* \xrightarrow{D^*} N(0, 1)$ in probability, where $\xrightarrow{D^*}$ denotes weak convergence under the bootstrap probability measure conditional on the observed sample X .*

Theorem 4.5 shows that the bootstrap provides an asymptotic valid approximation to the limit null distribution of $\hat{J}_{NT}$. This holds because we generate the bootstrap data by imposing the null hypothesis. If the null hypothesis does not hold in the observed sample, then we expect $\hat{J}_{NT}$ to explode at the rate $T^{1/2}N^{1/4}h^{1/4}$, which delivers the consistency of the bootstrap-based test $\hat{J}_{NT}^*$. 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 ??(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 $\frac{1}{N} \sum_{i=1}^N \left\| \tilde{\lambda}_{i0} \right\|^8 = O_P(1)$ 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 tests through Monte Carlo simulations.

5.1 Data generating process

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

$$X_{it} = \lambda'_{it} F_t + e_{it},$$

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)

$$\lambda_{it} = \lambda_{i0} \sim \text{IID } N(0, \mathbb{I}_2), e_{it} \sim \text{IID } N(0, 1).$$

DGP 2: (Heteroskedasticity)

$$\lambda_{it} = \lambda_{i0} \sim \text{IID } N(0, \mathbb{I}_2), e_{it} = \sigma_i v_{it}, \text{ where } \sigma_i \sim \text{IID } U(0.5, 1.5) \text{ and } v_{it} \sim \text{IID } N(0, 1).$$

DGP 3: (Cross sectional dependence)

$$\lambda_{it} = \lambda_{i0} \sim \text{IID } N(0, \mathbb{I}_2), e_t = (e_{1t}, \dots, e_{Nt})' \sim \text{IID } N(0, \Sigma_e), t = 1, 2, \dots, T, \text{ where } \Sigma_e = (c_{ij})_{i,j=1,\dots,N} \text{ with } c_{ij} = 0.5^{|i-j|}.$$

DGP 4: (Single structural break)

$$\lambda_{it,k} = \begin{cases} \lambda_{i0,k}, & \text{for } t = 1, 2, \dots, T/2 \\ \lambda_{i0,k} + b, & \text{for } t = T/2 + 1, \dots, T \end{cases}, \quad \lambda_{i0,k} \sim \text{IID } N(1, 1), k = 1, 2;$$

$$e_{it} = \sigma_i v_{it}, \text{ where } \sigma_i \sim \text{IID } U(0.5, 1.5), v_{it} \sim \text{IID } N(0, 1), \text{ and } b = 1, 2, 4.$$

DGP 5: (Multiple structural breaks)

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

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

DGP 6: (Smooth structural changes I)

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

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

DGP 7: (Smooth structural changes II)

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

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

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

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

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

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

$\lambda_{it,k} = \begin{cases} \lambda_{i0,k}, & \text{for } t = 1, 2, \dots, T/2 \\ \lambda_{i0,k} + b, & \text{for } t = T/2 + 1, \dots, T \end{cases}$, $\lambda_{i0,k} \sim \text{IID } N(1, 1)$, $k = 1, 2$;

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

and $b = 1, 2, 4$.

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

$\lambda_{it,1} = \begin{cases} \lambda_{i0,1} + 0.5b, & \text{for } 0.6T < t \leq 0.8T \\ \lambda_{i0,1} - 0.5b, & \text{for } 0.2T < t \leq 0.4T \\ \lambda_{i0,1}, & \text{otherwise} \end{cases}$, $\lambda_{i0,1} \sim \text{IID } N(1, 1)$, $\lambda_{it,2} = \lambda_{i0,2} \sim \text{IID } N(0, 1)$,

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

and $b = 1, 2, 4$.

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

5.2 Determination of the number of factors

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

$$\begin{aligned} IC_{h1}(R) &= \ln V \left(R, \left\{ \check{\Lambda}_r^{(R)} \right\} \right) + R \left(\frac{N + Th}{NTh} \right) \ln \left(\frac{NTh}{N + Th} \right), \\ IC_{h2}(R) &= \ln V \left(R, \left\{ \check{\Lambda}_r^{(R)} \right\} \right) + R \left(\frac{N + Th}{NTh} \right) \ln C_{NT}^2, \quad C_{NT} = \min \left\{ \sqrt{Th}, \sqrt{N} \right\}. \end{aligned}$$

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. We also redo the simulations by regenerating the factor loading randomly for each of the replications. The results are quite similar to those reported here and hence are omitted. Our local PCA involves nonparametric estimation. We use the Epanechnikov kernel and Silverman's rule of thumb (RoT) to set the bandwidth as $h = (2.35/\sqrt{12})T^{-1/5}N^{-1/10}$.³ We also try the uniform kernel and the quartic kernel, and the RoT bandwidth with different tuning parameters. Our simulation studies show that the choices of kernel function and the bandwidth have little impact on the performance of our information criteria. Each series is demeaned and standardized to have unit variance.

We use two measures to evaluate the information criteria, i.e., the average number of common factors and the empirical probability of correct selection over 1000 replications. Bai and Ng (2002) apply the former measure. However, this measure can be misleading. 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 4-10 are the time-varying factor models with abrupt or smooth structural changes, where the value of b indicates the magnitude of structural changes. To check the sensitivity of the information criteria to the magnitude of structural changes, we consider $b = 1, 2, 4$ for DGPs 4-10.

---

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

Table 1: Comparison of various information criteria in determining the number of factors: DGPs 7-10

DGP	(N, T)	Average number of factors										Empirical probability of correct selection								
		IC_{h1}	IC_{h2}	PC_{p1}	PC_{p2}	IC_{p1}	IC_{p2}	Ona	ER	GR	IC_{h1}	IC_{h2}	PC_{p1}	PC_{p2}	IC_{p1}	IC_{p2}	Ona	ER	GR	
1	(100,100)	2.02	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	.983	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
	(100,200)	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	.983	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
	(200,100)	2.02	2.01	2.00	2.00	2.00	2.00	2.00	2.00	2.00	.980	.992	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
	(200,200)	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
2	(100,100)	2.01	2.01	2.00	2.00	2.00	2.00	2.00	2.00	2.00	.988	.999	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
	(100,200)	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
	(200,100)	2.02	2.01	2.00	2.00	2.00	2.00	2.00	2.00	2.00	.983	.993	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
	(200,200)	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
3	(100,100)	2.02	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	.985	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
	(100,200)	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
	(200,100)	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
	(200,200)	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
4	$b=1$	(100,100)	2.02	2.00	3.24	3.02	3.00	2.95	2.39	1.67	2.01	.981	.998	.000	.000	.003	.051	.546	.666	.963
		(100,200)	2.00	2.00	3.19	3.03	3.00	3.00	2.68	1.77	2.01	1.00	1.00	.000	.000	.000	.000	.324	.768	.988
		(200,100)	2.00	2.00	3.00	3.00	3.00	2.99	2.92	1.50	2.06	.998	.999	.000	.000	.000	.008	.075	.495	.890
		(200,200)	2.00	2.00	3.00	3.00	3.00	3.00	2.97	1.59	2.14	1.00	1.00	.000	.000	.000	.000	.029	.585	.860
	$b=2$	(100,100)	2.21	2.08	3.10	3.00	3.00	3.00	2.49	1.01	2.32	.796	.918	.000	.000	.000	.000	.024	.011	.241
		(100,200)	2.00	2.00	3.19	3.03	3.00	3.00	2.68	1.77	2.01	1.00	1.00	.000	.000	.000	.000	.324	.768	.988
		(200,100)	2.13	2.08	3.00	3.00	3.00	3.00	2.98	1.01	2.63	.881	.926	.000	.000	.000	.000	.000	.006	.106
		(200,200)	2.02	2.01	3.00	3.00	3.00	3.00	3.00	1.01	3.00	.981	.995	.000	.000	.000	.000	.000	.000	.007
	$b=4$	(100,100)	2.66	2.48	3.00	3.00	3.00	3.00	2.87	1.00	2.08	.358	.524	.000	.000	.000	.000	.000	.000	.009
		(100,200)	2.53	2.32	3.00	3.00	3.00	3.00	2.99	1.00	2.84	.471	.680	.000	.000	.000	.000	.000	.000	.000
		(200,100)	2.59	2.48	3.00	3.00	3.00	3.00	3.00	1.00	2.64	.429	.523	.000	.000	.000	.000	.000	.000	.003
		(200,200)	2.39	2.30	3.00	3.00	3.00	3.00	3.00	1.00	3.00	.607	.702	.000	.000	.000	.000	.000	.000	.000
5	$b=1$	(100,100)	2.00	2.00	2.68	2.34	2.12	2.02	2.01	2.00	2.00	.997	1.00	.317	.664	.879	.977	.992	1.00	1.00
		(100,200)	2.00	2.00	2.90	2.72	2.36	2.17	2.02	2.00	2.00	1.00	1.00	.099	.282	.638	.833	.980	1.00	1.00
		(200,100)	2.00	2.00	2.64	2.46	2.26	2.13	2.26	2.00	2.00	.999	1.00	.362	.542	.739	.870	.736	1.00	1.00
		(200,200)	2.00	2.00	2.95	2.79	2.78	2.45	2.23	2.00	2.00	1.00	1.00	.047	.214	.225	.547	.770	1.00	1.00
	$b=2$	(100,100)	2.01	2.00	3.00	3.00	3.00	2.96	2.76	2.10	2.30	.992	.999	.000	.001	.005	.044	.243	.829	.701
		(100,200)	2.00	2.00	3.00	3.00	3.00	3.00	2.96	2.30	2.53	1.00	1.00	.000	.000	.000	.000	.042	.701	.475
		(200,100)	2.00	2.00	3.00	3.00	3.00	3.00	3.00	2.45	2.65	.999	1.00	.000	.001	.001	.001	.001	.493	.350
		(200,200)	2.00	2.00	3.00	3.00	3.00	3.00	3.00	2.84	2.94	1.00	1.00	.000	.000	.000	.000	.000	.161	.056
	$b=4$	(100,100)	2.32	2.21	3.00	3.00	3.00	3.00	3.00	2.97	3.00	.693	.795	.000	.000	.000	.000	.000	.006	.000
		(100,200)	2.23	2.12	3.00	3.00	3.00	3.00	3.00	3.00	3.00	.771	.884	.000	.000	.000	.000	.000	.000	.000
		(200,100)	2.21	2.16	3.00	3.00	3.00	3.00	3.00	3.00	3.00	.805	.842	.000	.000	.000	.000	.000	.000	.000
		(200,200)	2.16	2.11	3.00	3.00	3.00	3.00	3.00	3.00	3.00	.837	.891	.000	.000	.000	.000	.000	.000	.000
6	$b=1$	(100,100)	2.00	2.00	2.12	2.02	2.00	2.00	2.01	2.00	2.00	.998	1.00	.881	.985	.999	1.00	.991	1.00	1.00
		(100,200)	2.00	2.00	2.09	2.01	2.00	2.00	2.00	2.00	2.00	1.00	1.00	.907	.994	1.00	1.00	1.00	1.00	1.00
		(200,100)	2.00	2.00	2.15	2.05	2.01	2.00	2.21	2.00	2.00	.997	.999	.851	.948	.986	.998	.795	1.00	1.00
		(200,200)	2.00	2.00	2.37	2.07	2.08	2.08	2.18	2.00	2.00	1.00	1.00	.635	.934	.922	.922	.819	1.00	1.00
	$b=2$	(100,100)	2.00	2.00	3.00	3.00	3.00	2.97	2.93	1.91	2.33	1.00	1.00	.000	.000	.003	.035	.072	.700	.656
		(100,200)	2.00	2.00	3.00	3.00	3.00	3.00	2.96	2.12	2.46	1.00	1.00	.000	.000	.000	.000	.035	.779	.543
		(200,100)	2.00	2.00	3.00	3.00	3.00	3.00	3.09	2.11	2.35	1.00	1.00	.000	.000	.000	.000	.041	.788	.651
		(200,200)	2.00	2.00	3.00	3.00	3.00	3.00	3.09	2.17	2.44	1.00	1.00	.000	.000	.000	.000	.008	.830	.558
	$b=4$	(100,100)	2.07	2.04	3.98	3.89	3.83	3.45	1.51	1.01	2.27	.935	.958	.000	.000	.000	.000	.000	.000	.000
		(100,200)	2.01	2.00	4.00	4.00	4.00	4.00	1.45	1.00	2.55	.995	.998	.000	.000	.000	.000	.000	.000	.000
		(200,100)	2.03	2.02	3.99	3.97	3.96	3.88	2.76	1.09	2.70	.971	.978	.000	.000	.000	.000	.000	.000	.000
		(200,200)	2.00	2.00	4.00	4.00	4.00	4.00	2.35	1.04	2.94	.998	.999	.000	.000	.000	.000	.000	.000	.000

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 2: Comparison of various information criteria in determining the number of factors: DGPs 7-10

Average number of factors											Empirical probability of correct selection									
DGP	(N, T)	IC _{h1}	IC _{h2}	PC _{p1}	PC _{p2}	IC _{p1}	IC _{p2}	Ona	ER	GR	IC _{h1}	IC _{h2}	PC _{p1}	PC _{p2}	IC _{p1}	IC _{p2}	Ona	ER	GR	
7																				
b=1	(100,100)	2.01	2.00	2.80	2.33	2.10	2.01	2.00	2.00	2.00	.995	1.00	.224	.671	.904	.989	1.00	1.00	1.00	
	(100,200)	2.00	2.00	2.92	2.71	2.27	2.09	2.01	2.00	2.00	1.00	1.00	.088	.291	.726	.907	.994	1.00	1.00	
	(200,100)	2.00	2.00	2.59	2.38	2.18	2.08	2.18	2.00	2.00	1.00	1.00	.409	.616	.825	.917	.821	.999	1.00	
	(200,200)	2.00	2.00	2.94	2.68	2.66	2.31	2.09	2.00	2.00	1.00	1.00	.063	.318	.338	.688	.909	.999	1.00	
b=2	(100,100)	2.01	2.00	3.18	3.00	2.97	2.82	2.17	1.69	2.00	.990	.998	.000	.002	.030	.177	.777	.687	.987	
	(100,200)	2.00	2.00	3.11	3.02	3.00	3.00	2.50	1.74	2.00	1.00	1.00	.000	.000	.000	.002	.435	.735	.994	
	(200,100)	2.00	2.00	3.00	3.00	2.98	2.93	2.86	1.61	2.04	1.00	1.00	.002	.004	.025	.067	.132	.613	.931	
	(200,200)	2.00	2.00	3.00	3.00	3.00	3.00	2.89	1.62	2.09	1.00	1.00	.000	.000	.000	.000	.110	.623	.914	
b=4	(100,100)	2.05	2.02	3.04	3.00	3.00	3.00	1.90	1.03	1.80	.948	.978	.000	.001	.001	.005	.189	.025	.665	
	(100,200)	2.02	2.00	3.01	3.00	3.00	3.00	2.60	1.00	1.98	.985	.998	.000	.000	.000	.000	.024	.004	.656	
	(200,100)	2.01	2.01	3.00	3.00	3.00	3.00	2.97	1.01	2.12	.988	.995	.000	.001	.001	.001	.002	.012	.499	
	(200,200)	2.00	2.00	3.00	3.00	3.00	3.00	2.99	1.00	2.71	.999	1.00	.000	.000	.000	.000	.002	.002	.261	
8																				
b=1	(100,100)	2.01	2.00	2.12	2.01	2.00	2.00	2.01	2.00	2.00	.995	1.00	.889	.994	1.00	1.00	.995	1.00	1.00	
	(100,200)	2.00	2.00	2.11	2.02	2.00	2.00	2.00	2.00	2.00	1.00	1.00	.888	.983	1.00	1.00	.999	1.00	1.00	
	(200,100)	2.00	2.00	2.11	2.03	2.01	2.00	2.18	2.00	2.00	.997	.999	.891	.967	.994	.998	.819	1.00	1.00	
	(200,200)	2.00	2.00	2.36	2.07	2.08	2.01	2.00	2.00	1.00	1.00	.641	.935	.923	.993	.917	.801	1.00	1.00	
b=2	(100,100)	2.00	2.00	3.00	3.00	3.00	2.97	2.93	1.90	2.31	1.00	1.00	.000	.000	.002	.029	.067	.728	.678	
	(100,200)	2.00	2.00	3.00	3.00	3.00	3.00	2.96	2.10	2.47	1.00	1.00	.000	.000	.000	.000	.034	.753	.532	
	(200,100)	2.00	2.00	3.00	3.00	3.00	3.00	3.11	2.08	2.32	1.00	1.00	.000	.000	.000	.001	.030	.816	.678	
	(200,200)	2.00	2.00	3.00	3.00	3.00	3.00	3.09	2.19	2.47	1.00	1.00	.000	.000	.000	.000	.006	.803	.526	
b=4	(100,100)	2.05	2.04	3.98	3.88	3.81	3.45	1.54	1.04	2.25	.946	.963	.000	.000	.000	.000	.000	.000	.000	
	(100,200)	2.01	2.01	4.00	4.00	3.99	3.96	1.50	1.01	2.49	.990	.995	.000	.000	.000	.000	.000	.000	.000	
	(200,100)	2.04	2.03	3.99	3.97	3.95	3.86	2.74	1.08	2.68	.965	.972	.000	.000	.000	.000	.000	.000	.000	
	(200,200)	2.01	2.00	4.00	4.00	4.00	4.00	2.34	1.04	2.94	.993	.995	.000	.000	.000	.000	.000	.000	.000	
9																				
b=1	(100,100)	2.01	2.00	3.06	3.00	3.00	2.96	2.44	1.65	2.02	.988	.998	.000	.001	.005	.040	.496	.653	.945	
	(100,200)	2.00	2.00	3.03	3.00	3.00	3.00	2.78	1.71	2.01	1.00	1.00	.000	.000	.000	.000	.243	.708	.980	
	(200,100)	2.01	2.00	3.00	3.00	3.00	3.00	2.97	1.48	2.15	.994	.997	.000	.000	.001	.003	.025	.474	.811	
	(200,200)	2.00	2.00	3.00	3.00	3.00	3.00	3.00	1.51	2.38	1.00	1.00	.000	.000	.000	.000	.002	.497	.622	
b=2	(100,100)	2.18	2.07	3.00	3.00	3.00	3.00	2.77	1.01	2.60	.818	.929	.000	.000	.000	.000	.003	.008	.155	
	(100,200)	2.05	2.02	3.00	3.00	3.00	3.00	2.98	1.00	2.95	.946	.981	.000	.000	.000	.000	.000	.000	.037	
	(200,100)	2.13	2.09	3.00	3.00	3.00	3.00	3.00	1.02	2.92	.870	.915	.000	.000	.000	.000	.000	.002	.040	
	(200,200)	2.04	2.02	3.00	3.00	3.00	3.00	3.00	1.13	3.00	.962	.980	.000	.000	.000	.000	.000	.000	.000	
b=4	(100,100)	2.65	2.50	3.00	3.00	3.00	3.00	2.99	1.00	2.62	.360	.498	.000	.000	.000	.000	.000	.000	.002	
	(100,200)	2.48	2.32	3.00	3.00	3.00	3.00	3.00	1.00	2.99	.522	.677	.000	.000	.000	.000	.000	.000	.000	
	(200,100)	2.59	2.50	3.00	3.00	3.00	3.00	3.00	1.00	2.94	.436	.509	.000	.000	.000	.000	.000	.000	.000	
	(200,200)	2.42	2.34	3.00	3.00	3.00	3.00	3.00	1.00	3.00	.576	.657	.000	.000	.000	.000	.000	.000	.000	
10																				
b=1	(100,100)	2.01	2.00	2.71	2.34	2.14	2.02	2.00	2.00	2.00	.993	1.00	.299	.661	.858	.977	.999	1.00	1.00	
	(100,200)	2.00	2.00	2.91	2.71	2.35	2.13	2.02	2.00	2.00	1.00	1.00	.095	.290	.654	.872	.983	1.00	1.00	
	(200,100)	2.00	2.00	2.65	2.45	2.22	2.11	2.27	2.00	2.00	.999	1.00	.352	.554	.776	.893	.728	1.00	1.00	
	(200,200)	2.00	2.00	2.96	2.80	2.79	2.46	2.21	2.00	2.00	1.00	1.00	.039	.198	.211	.540	.795	1.00	1.00	
b=2	(100,100)	2.01	2.00	3.00	3.00	2.99	2.95	2.79	2.12	2.30	.990	.997	.000	.001	.006	.048	.217	.813	.698	
	(100,200)	2.00	2.00	3.00	3.00	3.00	3.00	2.96	2.26	2.52	1.00	1.00	.000	.000	.000	.000	.044	.732	.480	
	(200,100)	2.00	2.00	3.00	3.00	3.00	3.00	2.99	2.45	2.67	1.00	1.00	.000	.000	.000	.000	.006	.492	.333	
	(200,200)	2.00	2.00	3.00	3.00	3.00	3.00	3.00	2.85	2.95	1.00	1.00	.000	.000	.000	.000	.000	.146	.052	
b=4	(100,100)	2.33	2.22	3.00	3.00	3.00	3.00	3.00	2.96	2.99	.684	.776	.000	.000	.000	.000	.000	.005	.002	
	(100,200)	2.25	2.13	3.00	3.00	3.00	3.00	3.00	3.00	3.00	.754	.875	.000	.000	.000	.000	.000	.000	.000	
	(200,100)	2.20	2.15	3.00	3.00	3.00	3.00	3.00	3.00	3.00	.815	.856	.000	.000	.000	.000	.000	.000	.000	
	(200,200)	2.20	2.13	3.00	3.00	3.00	3.00	3.00	3.00	3.00	.803	.870	.000	.000	.000	.000	.000	.000	.000	

Note: See the note in Table 1.

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

5.3 Performance of the test

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

Since our bootstrap testing procedure is rather time consuming, we generate 500 data sets in this subsection and set the bootstrap replication number B to be 200. As in the previous subsection, we use the Epanechnikov kernel and the RoT bandwidth $h = (2.35/\sqrt{12})T^{-1/5}N^{-1/10}$. In addition to our test, we also consider Breitung and Eickmeier's (2011) sup-LM variable-specific test, Chen et al.'s (2014) sup-LM and sup-Wald tests, Han and Inoue's (2015) sup-LM and sup-Wald tests, and Yamamoto and Tanaka's (2015) sup-Wald test. We follow these papers to set

the trimming parameter $\tau = 0.15$, which implies that the single break can only occur on the time interval $[[0.15T], [0.85T]]$. The tests of Chen et al. (2014) and Han and Inoue (2015) involve the long run variance estimation. We set the time-lag truncation parameter as $m = \lfloor T^{1/5} \rfloor$ and choose the Bartlett kernel. The critical values presented in Andrews (1993) are applied for the tests of Breitung and Eickmeier (2011), Chen et al. (2014) and Han and Inoue (2015), while the bootstrap critical values are applied to check the performance of our test.

In the literature, the number of common factors is either determined by Bai and Ng's (2002) information criteria (e.g., Han and Inoue, 2015) or specified by some fixed numbers, which may be equal to, less than, or greater than the correct number of factors (e.g., Chen et al., 2014). In this paper, we apply some information criteria to determine the number of factors used for our simulations. Specifically, we apply IC_{h1} and IC_{h2} given in Section 5.2 to determine the number of factors used for our tests. For the tests of Breitung and Eickmeier (2011), Chen et al. (2014), Han and Inoue (2015), and Yamamoto and Tanaka (2015), the number of factors is determined by Bai and Ng's (2002) information criteria IC_{p1} and IC_{p2} ; the results based on these two information criteria are quite similar and we only report the results using IC_{p1} to save space. In addition, we also implement our tests and those of Breitung and Eickmeier (2011), Chen et al. (2014), and Han and Inoue (2015) by specifying the number of factors as the correct number of common factors.⁴ The results are reported in the online supplementary material.

Table 3 reports the empirical sizes of various tests at both 5% and 10% significance levels. As shown in the table, our tests based on IC_{h1} and IC_{h2} have reasonable size. Han and Inoue's (2015) sup-LM test delivers reasonable size but their sup-Wald test tends to under-reject the null hypothesis. Chen et al.'s (2014) sup-LM test has reasonable size, but their sup-Wald test tends to over-reject the null hypothesis in DGP 2. Breitung and Eickmeier's (2011) variable-specific sup-LM test suffers from slight under-rejection for DGPs 1-3. The size of Yamamoto and Tanaka's (2015) test is generally well behaved.

Tables 4 and 5 report the empirical powers of various tests for DGPs 4-10 at the 5% and 10% significance levels when the number of factors are determined from the data. To save space, we only report the results for $b = 1$ and 2. We summarize some important findings. First, our $\hat{J}_{NT}$ test is powerful in detecting all the forms of time-varying factor loadings given by DGPs 4-10 and the simulation results are consistent with our theoretical conclusion that our test is able to detect both a finite number of sudden structural breaks and smooth structural changes. Second, Breitung and Eickmeier's (2011) test is powerless for DGPs 4-10, which is consistent with Yamamoto and Tanaka's (2015) argument that when the number of factors is estimated from the data, Breitung

---

⁴Yamamoto and Tanaka (2015) argue that when the number of factors is estimated from the data, Breitung and Eickmeier's (2011) test tends to exhibit non-monotonic power due to the presence of estimated spurious factors under the alternative. So they propose a test that is robust to this non-monotonic power problem. But their test reduces to Breitung and Eickmeier's (2011) test when we fix the number of factors at the true number.

Table 3: Size of various tests for DGPs 1-3

DGP	N	T	J_{h1}		J_{h2}		HI_{LM}		HI_W		CDG_{LM}		CDG_W		BE_{LM}		YT	
			5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
1	100	100	6.4	12.8	6.0	11.6	3.6	7.2	0.4	2.0	2.2	8.2	5.8	11.0	2.9	6.7	5.9	10.5
	100	200	5.0	8.8	5.4	10.2	6.0	11.6	1.4	5.6	4.8	10.2	6.0	10.2	3.6	7.7	6.2	11.7
	200	100	8.0	11.4	7.2	10.6	3.2	6.0	0.0	1.6	2.8	8.6	6.4	12.4	2.9	6.6	5.6	10.5
	200	200	4.4	11.4	4.4	10.6	7.0	14.4	2.0	7.8	5.8	11.6	6.6	12.2	3.6	7.7	6.0	11.4
2	100	100	5.4	12.0	4.4	11.2	3.0	8.4	0.4	1.8	3.2	9.8	6.2	14.4	2.9	6.6	6.1	11.0
	100	200	6.8	13.4	7.6	12.8	3.2	8.2	2.6	7.8	4.4	11.2	5.4	14.2	3.7	7.8	6.4	11.8
	200	100	8.0	13.2	7.0	12.4	2.8	6.8	0.4	2.0	3.6	9.2	7.4	15.2	2.8	6.4	6.2	11.2
	200	200	4.8	10.2	5.6	9.8	3.6	9.6	2.2	7.4	3.4	9.2	5.8	13.2	3.6	7.7	6.3	11.7
3	100	100	7.8	12.0	5.2	11.0	3.6	8.2	0.2	2.2	2.8	7.4	5.4	10.4	2.7	6.4	6.4	11.3
	100	200	5.0	10.4	5.4	11.0	7.0	12.8	1.8	6.6	4.6	8.8	5.0	10.8	3.4	7.5	6.5	11.9
	200	100	7.4	12.2	6.2	10.8	3.2	6.0	0.2	2.0	3.0	7.2	5.6	10.8	2.8	6.3	5.8	10.7
	200	200	5.8	11.0	5.8	11.0	7.8	13.2	1.8	6.8	6.4	11.6	7.2	13.0	3.4	7.4	7.4	12.9

Note: (i) J_{h1} and J_{h2} denote the results of our test (J_{NT}) based on the bootstrap p-values by determining the number of common factors using our information criteria IC_{h1} and IC_{h2} proposed in Section 5.2; (ii) HI_{LM} and HI_W denote Han and Inoue's (2015) sup-LM and sup-Wald tests by determining the number of common factors using the information criteria IC_{p1} proposed by Bai and Ng (2002); (iii) CDG_{LM} and CDG_W denote Chen et al.'s (2014) sup-LM and sup-Wald tests by determining the number of common factors using the information criteria IC_{p1} proposed by Bai and Ng (2002); (iv) BE_{LM} denotes Breitung and Eickmeier's (2011) variable-specific sup-LM test by determining the number of common factors using the information criteria IC_{p1} proposed by Bai and Ng (2002); (v) YT denote Yamamoto and Tanaka's (2015) test by determining the number of common factors using the information criteria IC_{p1} proposed by Bai and Ng (2002). The main entries report the average percentage of rejection.

and Eickmeier’s (2011) test tends to exhibit non-monotonic power. Third, the tests of Chen et al. (2014), Han and Inoue (2015) and Yamamoto and Tanaka (2015) are all designed to test for a one-time abrupt structural change in DGPs 4 and 9. As expected, they all have power against DGPs 4 and 9. Fourth, for the other DGPs, all of Han and Inoue’s (2015) sup-LM and sup-Wald tests, Chen et al.’s (2014) sup-LM and sup-Wald tests, and Yamamoto and Tanaka’s (2015) test have lower power than our test too. In particular, the tests of Chen et al. (2014) and Han and Inoue (2015) have very low power in detecting deviations from the null in DGPs 5, 7 and 10, especially when the sample size is small. But these tests have reasonable power against DGPs 6 and 8. It is easy to explain why some of these other tests have power against DGPs 6 and 8. Note that in these two DGPs, the factor loadings are monotonic functions of the time ratio t/T for each i . If we apply the PCA method to estimate the factor model, the estimated factors would exhibit a trend with increasing volatilities. Since Han and Inoue’s (2015) test checks the time invariance property of the second order moments of the common factors, it is possible to capture such smooth structural changes in DGPs 6 and 8. Similarly, Chen et al.’s (2014) test is based on the regression of one of the estimated factors on the remaining estimated factors, and their LM and Wald test statistics will not have the usual asymptotic distribution when one estimated factor exhibits trending behavior.

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.⁵ By excluding the first two quarters, which is missing when computing the first and second differences, we get a total of $T = 190$ quarterly observations. Also, we follow the suggestion of Stock and Watson (2009) to delete some high level aggregates related by identities to the lower level sub-aggregates and end up with $N = 109$ time series. For 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

---

⁵The dataset is publicly available on Professor Mark W. Watson’s website <http://www.princeton.edu/~mwatson/publi.html>.

Table 4: Power of tests under DGPs 4-7

DGP	N	T	J_{h1}		J_{h2}		HI_{LM}		HI_W		CDG_{LM}		CDG_W		BE_{LM}		YT	
			5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
4, $b = 1$	100	100	99.8	100	100	100	73.4	94.2	64.4	90.4	99.2	99.6	99.8	99.8	5.3	10.5	98.2	98.8
	100	200	100	100	100	100	100	100	100	100	100	100	100	100	10.4	16.9	99.9	100
	200	100	100	100	100	100	76.2	96.0	69.8	94.2	100	100	100	100	3.8	8.3	99.1	99.4
	200	200	100	100	100	100	100	100	100	100	100	100	100	100	6.3	12.1	100	100
4, $b = 2$	100	100	100	100	100	100	80.4	96.6	73.0	94.4	99.2	100	100	100	4.3	8.9	99.5	99.7
	100	200	100	100	100	100	100	100	100	100	100	100	100	100	8.0	14.1	100	100
	200	100	100	100	100	100	81.2	97.4	73.0	95.4	99.8	100	100	100	3.5	7.7	99.9	99.9
	200	200	100	100	100	100	100	100	100	100	100	100	100	100	5.7	11.3	100	100
5, $b = 1$	100	100	94.2	96.6	94.6	96.6	4.0	10.2	0.4	3.2	5.2	13.6	12.2	23.0	6.2	11.5	20.1	27.9
	100	200	100	100	100	100	24.6	42.2	18.4	35.2	40.6	49.8	46.0	54.0	11.1	17.4	44.7	53.4
	200	100	98.8	99.6	98.6	100	3.0	7.8	0.6	2.6	4.2	16.6	16.0	28.8	5.1	10.1	24.3	30.7
	200	200	100	100	100	100	47.8	71.0	31.6	57.8	74.6	81.2	78.4	83.4	5.8	11.0	65.4	71.2
5, $b = 2$	100	100	94.2	96.6	94.6	96.6	2.0	7.4	0.0	1.0	6.0	29.6	44.6	65.0	3.3	7.4	68.5	77.8
	100	200	100	100	100	100	64.2	86.6	44.2	72.8	88.4	96.2	98.2	99.4	4.6	9.7	94.5	97.0
	200	100	98.8	99.0	98.6	100	1.6	5.6	0.0	0.4	9.4	32.6	45.4	68.6	3.1	7.2	73.8	81.9
	200	200	100	100	100	100	65.8	88.8	46.4	75.4	92.4	98.2	99.2	100	4.2	9.1	96.2	97.9
6, $b = 1$	100	100	93.8	97.8	94.0	98.2	83.4	94.0	65.4	86.2	34.2	50.4	71.2	76.2	14.9	24.2	49.2	58.8
	100	200	100	100	100	100	100	100	100	100	65.4	72.2	78.8	83.0	34.9	46.1	75.7	81.7
	200	100	98.6	99.2	98.0	99.4	83.2	94.8	67.2	89.0	36.8	50.0	59.8	68.0	15.5	24.9	53.2	62.5
	200	200	100	100	100	100	100	100	100	100	60.8	68.2	66.6	71.8	34.1	45.0	76.6	82.4
6, $b = 2$	100	100	99.6	100	99.6	100	80.0	96.0	29.4	67.0	98.8	99.4	100	100	9.4	17.3	99.1	99.5
	100	200	100	100	100	100	100	100	100	100	100	100	100	100	23.2	34.1	100	100
	200	100	100	100	100	100	79.6	95.8	30.6	70.0	98.6	100	100	100	9.8	17.9	99.2	99.5
	200	200	100	100	100	100	100	100	100	100	100	100	100	100	23.9	35.0	100	100
7, $b = 1$	100	100	95.0	98.6	94.6	97.6	12.0	20.2	0.2	1.4	7.0	13.2	3.4	8.4	4.3	8.8	16.4	22.2
	100	200	100	100	100	100	31.6	46.0	1.6	4.6	27.4	34.4	6.4	15.4	7.4	13.1	33.5	39.5
	200	100	99.6	100	99.6	100	10.0	20.8	0.0	1.4	8.2	15.6	3.2	8.8	4.2	8.7	20.4	24.7
	200	200	100	100	100	100	50.8	65.8	1.0	3.6	63.0	67.8	10.6	27.0	5.6	10.7	62.9	66.2
7, $b = 2$	100	100	99.8	100	99.8	100	15.0	29.0	0.0	0.0	36.2	58.0	1.6	5.8	3.1	7.0	90.2	92.7
	100	200	100	100	100	100	64.4	81.6	0.6	3.0	96.2	99.6	8.6	26.2	4.7	9.2	98.7	99.2
	200	100	100	100	100	100	17.6	29.4	0.0	0.0	39.2	62.6	2.6	7.0	3.0	6.7	92.7	94.2
	200	200	100	100	100	100	72.0	86.2	0.6	4.0	97.8	99.8	9.0	27.2	4.0	8.5	98.3	98.9

Note: See the note in Table 3.

Table 5: Power of tests under DGPs 8-10

DGP	N	T	J_{h1}		J_{h2}		HI_{LM}		HI_W		CDG_{LM}		CDG_W		BE_{LM}		YT	
			5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
8, $b = 1$	100	100	92.6	96.0	93.2	95.6	81.4	95.0	68.6	88.8	39.2	53.2	71.4	78.0	14.8	23.9	50.4	60.0
	100	200	100	100	100	100	99.8	100	100	100	62.6	70.6	76.2	81.4	35.0	46.0	74.4	80.5
	200	100	97.0	98.8	97.0	98.8	78.0	92.8	67.6	86.8	39.2	53.2	60.2	69.0	15.4	24.9	53.6	62.9
	200	200	100	100	100	100	100	100	99.6	99.8	56.0	65.8	63.0	70.8	33.9	44.5	74.7	80.6
8, $b = 2$	100	100	99.6	99.8	99.2	99.8	78.6	94.0	24.4	61.2	98.0	100	100	100	9.5	17.4	99.1	99.5
	100	200	100	100	100	100	100	100	100	100	100	100	100	100	23.1	33.9	100	100
	200	100	100	100	100	100	76.4	94.6	26.2	67.8	98.2	99.8	99.8	100	9.6	17.7	99.1	99.5
	200	200	100	100	100	100	100	100	100	100	100	100	100	100	23.9	35.0	100	100
9, $b = 1$	100	100	99.8	99.8	99.8	99.8	71.0	93.4	64.0	90.0	99.2	99.8	99.8	99.8	4.7	9.7	98.2	98.8
	100	200	100	100	100	100	100	100	100	100	100	100	100	100	8.6	14.8	100	100
	200	100	100	100	100	100	79.0	96.6	68.8	94.6	99.6	100	100	100	3.4	7.7	99.4	99.7
	200	200	100	100	100	100	100	100	100	100	100	100	100	100	5.1	10.4	100	100
9, $b = 2$	100	100	98.4	99.2	98.8	99.2	77.8	96.2	71.6	95.2	98.8	100	100	100	3.5	7.9	99.7	99.8
	100	200	100	100	100	100	100	100	100	100	100	100	100	100	5.9	11.5	100	100
	200	100	98.0	98.8	98.2	98.6	83.4	97.4	75.2	95.6	98.8	99.6	100	100	3.0	7.0	99.9	99.9
	200	200	100	100	100	100	100	100	100	100	100	100	100	100	4.3	9.2	100	100
10, $b = 1$	100	100	93.8	96.4	94.0	96.6	3.6	9.2	0.4	3.0	6.2	14.6	13.2	21.6	6.3	11.4	23.0	30.7
	100	200	100	100	100	100	27.0	42.8	16.6	33.2	39.2	48.0	43.8	51.2	10.6	17.1	43.4	51.7
	200	100	98.0	98.8	98.2	98.8	2.0	7.2	0.4	2.8	5.6	16.4	13.0	26.2	5.3	10.3	25.4	33.1
	200	200	100	100	100	100	48.6	74.2	34.8	56.8	77.8	84.2	81.8	85.2	5.4	10.6	68.2	73.9
10, $b = 2$	100	100	99.4	99.4	99.8	99.8	1.6	6.2	0.0	0.0	6.0	25.0	46.4	68.0	3.2	7.4	69.8	78.9
	100	200	100	100	100	100	62.0	85.8	43.0	72.2	86.6	94.6	97.8	99.6	4.4	9.4	93.8	96.5
	200	100	100	100	100	100	2.0	5.8	0.0	0.4	9.4	30.8	43.2	69.0	3.1	7.1	73.7	81.9
	200	200	100	100	100	100	63.4	90.8	45.8	75.2	91.4	98.0	98.8	100	4.1	8.8	95.8	97.7

Note: See the note in Table 3.

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

Number of selected factors	1	3	4	5
Criterion functions	Ona, ER, GR	IC_{h1}, IC_{h2}	$PC_{p2}, IC_{p1}, IC_{p2}$	PC_{p1}

Note: See the note in Table 1.

Table 7: Tests of structural changes in the U.S. economy

	Our test: bootstrap			Han and Inoue (2015)				Chen et al. (2014)				BE (2011)	
	J_{NT}	5%	10%	sup-LM	sup-Wald	5%	10%	sup-LM	sup-Wald	5%	10%	5%	10%
$R = 1$	5.40	2.82	2.12	7.03	7.51	8.85	7.17	—	—	—	—	.2844	.3945
$R = 2$	23.90	10.94	9.84	13.04	13.61	14.15	12.27	2.75	3.46	8.85	7.17	.4037	.4587
$R = 3$	31.48	16.35	15.30	17.05	17.85	20.26	18.12	7.03	11.54	11.79	11.01	.4771	.5872
$R = 4$	30.44	23.14	22.43	24.31	24.22	27.03	24.62	9.96	11.44	14.15	12.27	.4862	.5872
$R = 5$	35.50	26.20	25.65	31.79	31.12	35.06	32.51	12.60	54.92	16.45	14.31	.4679	.5596

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) except for the Breitung and Eickmeier’s (2011) test; (iii) Under 5% and 10% of BE (2011) are the empirical rejection frequencies of Breitung and Eickmeier’s (2011) variable-specific sup-LM test by using 5% and 10% asymptotic critical values respectively. Bold elements denote significance at the 5% nominal level.

criteria $PC_{p1}, PC_{p2}, IC_{p1}$, and IC_{p2} , Onatski’s (2009) testing procedure, Ahn and Horenstein’s (2013) criterion functions ER and GR and our information criterion proposed in Section 3.3 to determine the number of common factors. The results are reported in Table 6. According to the table, we report the test results for the cases of one to five common factors respectively in the following context .

Table 7 reports the results of the tests and the corresponding critical values at the 5% and 10% significance 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 (2015) sup-LM and sup-Wald tests cannot reject the null for any case at the 5% significance level, while Chen et al.’s (2014) results are mixed, and they can only reject the null for $R = 5$ at the 5% significance level when using the sup-Wald test. This is consistent with the results of our simulation studies that suggest the tests of Han and Inoue (2015) and Chen et al. (2014) have relatively low power when structural changes are not a one-time break. In addition, Breitung and Eickmeier’s (2011, BE) variable-specific sup-LM test reject the null of time-invariant factor loadings for about half of the variables.

Our empirical result suggests the existence of possible smooth or sudden structural changes in U.S. economy. We now estimate the common factors and the time-varying factor loadings by using our local principal component approach proposed in Section 2 by assuming 3 common factors. 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

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

(right panel) corresponding to the three common factors selected by our information criteria. From this figure, we can see that the estimated factor loadings show significant time-varying features. The finding of time-varying factor loadings has some important implications. For example, most of the existing studies estimate the common factors under the framework of time-invariant factor loadings and then forecast some key variables based on the estimated common factors. We may provide more reliable forecasts by accommodating the documented time-varying features of factor loadings by using a local version of the principal component method.

7 Conclusion

Conventional factor models assume that factor loadings are fixed over a long horizon of time, which appears restrictive and unrealistic in empirical applications. In this paper, we introduce a time-varying factor model where factor loadings are allowed to change smoothly over time and propose a local version of the PCA method to estimate the latent factors and time-varying factor loadings simultaneously. We establish the limiting distributions of the estimated factors and factor loadings in the standard large N and large T framework. We also propose a BIC-type information criterion to determine the number of common factors for time-varying factor models. Our information criterion works no matter whether the factor loadings are time-invariant or time-varying and it is extremely useful when structural changes are suspected.

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

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

one believes that the changes are abrupt. The estimation of factor models with multiple breaks is certainly an interesting and challenging issue. We are exploring these topics in ongoing work.

References

- Ahn S., Horenstein A., 2013. Eigenvalue ratio test for the number of factors. *Econometrica* 81, 1203-1227.
- Andrews, D. W. K., 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821-856.
- Bai, J., 2003. Inferential theory for factor models of large dimensions. *Econometrica* 71, 135-172.
- Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica* 90, 191-221.
- Bai, J., Ng, S., 2013. Principal components estimation and identification of static factors. *Journal of Econometrics* 176 18-29.
- Bates, B.J., Plagborg-Møller, M., Stock, J.H., Watson, M.W., 2013. Consistent factor estimation in dynamic factor models with structural instability. *Journal of Econometrics* 177, 289-304.
- Breitung, J. and Eickmeier, S., 2011. Testing for structural breaks in dynamic factor models. *Journal of Econometrics* 163, 71-84.
- Cai, Z., 2007. Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics* 136, 163-188.
- Chen, B., Hong, Y., 2012. Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica* 80, 1157-1183.
- Chen, J., Gao, J., Li, D., 2012. Semiparametric testing for smooth structural changes in panel data models. *Journal of Econometrics* 171, 71-85.
- Chen, L., Dolado, J., Gonzalo, J., 2014. Detecting big structural breaks in large factor models. *Journal of Econometrics* 180, 30-48.
- Corradi, V. and Swanson, N.R., 2014. Testing for structural stability of factor augmented forecasting models. *Journal of Econometrics* 182, 100-118.
- Cheng, X., Liao, Z. and Schorfheide, F., 2016. Shrinkage estimation of high-dimensional factor models with structural instabilities. *Review of Economic Studies* 83, 1511-1543.
- Del Negro, M. D., Otrok, C. 2009. Dynamic factor models with time-varying parameters: measuring changes in international business cycles. Staff Report No. 326, Federal Reserve Bank of New York.
- Motta, G., Hafner, C. M., von Sachs, R., 2011. Local stationary factor models: identification and non-parametric estimation. *Econometric Theory* 27, 1279-1319.
- Han, X., Inoue, A., 2015. Tests for parameter instability in dynamic factor models. *Econometric Theory* 31, 1117-1152.
- Hansen, B., 2001. The new econometrics of structural change: dating breaks in U.S. labor productivity. *Journal of Economic Perspectives* 15, 117-128.
- Hong, Y., Li, H., 2005. Nonparametric specification testing for continuous-time models with applications to term structure of interest rate. *Review of Financial Studies* 18, 37-84.
- Li, Q., Racine, J.S., 2007. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.
- Lu, X., Su, L., 2016. Shrinkage estimation of dynamic panel data models with interactive fixed effects. *Journal of Econometrics* 190, 148-175.
- Moon, H., Weidner, M., 2015. Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica* 83, 1543-1579.

- Onatski, A., 2009. Testing hypotheses about the number of factors in large factor models. *Econometrica* 77, 1447-1479.
- Park, B., Mammen E., Härdle W., Borak S., 2009. Time series modelling with semiparametric factor dynamics. *Journal of the American Statistical Association*, 104, 284-298.
- Pesaran, M. H., Yamagata, T., 2008. Testing slope homogeneity in large panels. *Journal of Econometrics* 142, 50-93.
- Pollard, D., 1984. *Convergence of Stochastic Processes*, Springer-Verlag.
- Robinson, P.M., 1989. Nonparametric estimation of time-varying parameters. In Hackl, P. (eds), *Statistical Analysis and Forecasting of Economic Structural Change*, pp. 253-264. Berlin: Springer.
- Robinson, P.M., 1991. Time-varying nonlinear regression. In Hackl, P. (eds), *Economic Structural Change Analysis and Forecasting*, pp.179-190. Berlin: Springer.
- Robinson, P.M., 2012. Nonparametric trending regression with cross-sectional dependence. *Journal of Econometrics* 169, 4-14.
- Rodríguez-Poo, J. M., Linton, O. 2001. Nonparametric factor analysis of residual time series. *Test* 10, 161-182.
- Stock, J.H., Watson, M.W., 1998. Diffusion indexes. NBER Working Paper 6702.
- Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97, 1167-1179.
- Stock, J.H., Watson, M.W., 2009. Forecasting in dynamic factor models subject to structural instability. In Castle, J., Shephard, N. (Eds.), *The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry*, pp.1-57, Oxford University Press.
- Su, L., Chen, Q., 2013. Testing homogeneity in panel data models with interactive fixed effects. *Econometric Theory* 29, 1079-1135.
- Su, L., Jin, S., Zhang, Y., 2015. Specification test for panel data models with interactive fixed effects. *Journal of Econometrics* 185, 230-258.
- Su, L., Wang, X., Jin, S., 2016. Sieve estimation of time-varying panel data models with latent structures. Working Paper, Singapore Management University.
- White, H., 2001. *Asymptotic Theory for Econometricians*. Emerald, UK.
- Yamamoto, Y., Tanaka, S., 2015. Testing for factor loading structural change under common breaks. *Journal of Econometrics* 189, 187-206.