On Time-Varying Factor Models: Estimation and Inference*

Liangjun Su[a] and Xia Wang[b]

[a] School of Economics, Singapore Management University, Singapore

[b] School of Management, University of Chinese Academy of Sciences, Beijing

March 6, 2015

Abstract

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

JEL Classification: C12, C14, C33, C38.

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

*Su gratefully acknowledges the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021. 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.

1 Introduction

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

In recent years, more and more research has focused on structural changes in factor loadings. Stock and Watson (2008) examine the forecasting reliability when there exists a structural break in the factor loadings. Breitung and Eickmeier (2011) propose three statistics (namely, LR, LM and Wald statistics) to test for structural breaks in factor loadings. Chen et al. (2014) propose a two-stage procedure to detect *big* breaks in factor loadings by testing the parameter stability in a regression of one of the estimated factors on the remaining estimated factors. Corradi and Swanson (2014) propose a test to check structural stability of both factor loadings and factor-augmented forecasting regression coefficients. Han and Inoue (2014) propose a joint test for structural breaks in factor loadings based on the second moments of the estimated factors. Cheng et al. (2014) consider the case that both the factor loadings and the number of factors may change simultaneously at a time point. These studies provide appropriate econometric tools to examine the problem of structural breaks in factor loadings. However, all these papers focus on the case of one-time abrupt structural changes. The analyses may be inappropriate if, for example, such driving forces of structural changes as preference changes, technological progress and policy changes, play a role gradually over a long period of time, or some abrupt policy changes also take a period of time to take effect. Indeed, as Hansen (2001) points out, "it may seem unlikely that a structural break could be immediate and might seem more reasonable to allow a structural change to take a period of time to take effect". Hence, it seems more realistic to assume smooth changes rather than abrupt changes. To the best of our knowledge, Bates et al. (2013) is the only paper that allows for smooth changes in factor loading. By controlling the magnitude of instabilities to be "small", they show that the principal component estimators of factors are still consistent. In fact, changes in comovement induced by technological progress and other forces are gradual but fundamental. As a result, we can neither assume the structural changes to be negligible nor check the instabilities of factor loadings under the framework of abrupt structural changes.

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

More importantly, we propose an L_2-distance-based test statistic to check the stability of factor loadings. The basic idea is to estimate the time-varying factor loadings and the latent common factors by the local version of PCA, and compare the fitted values of the common components with those estimated by the conventional PCA method based on the whole sample. By construction, our test is able to capture both smooth and abrupt structural changes in factor loadings, where the number of abrupt changes is usually assumed to be one in the literature but can be any fixed unknown number in our setup. Unlike the existing tests, such as Breitung and Eickmeier (2011), Chen et al. (2014) and Han and Inoue (2014), which check the stability of the moments of factor loadings or common factors, our test compares the estimates of the common components because it is well known that the latent factors and the factor loadings are not separately identifiable. Moreover, unlike the existing tests for unknown break date, namely the supremum-type tests of Breitung and Eickmeier (2011), Chen et al. (2014), and Han and Inoue (2014), no trimming of the boundary regions near the starting or ending of period is required for our test. In other words, we allow the breaks to occur near the end of the sample under the alternative.

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

NOTATION. For an $m \times n$ real matrix A, we denote its transpose as A', its Frobenius norm as $\|A\|$ ($\equiv [\text{tr}(AA')]^{1/2}$), 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 sth 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 $(N, T) \to \infty$ to denote that N and T pass to infinity jointly.

2 Factor Model with Time-varying Factor Loadings

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

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

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

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

To avoid model misspecification and to allow our model to capture various kinds of time-varying factor loadings, we use a nonparametric local smoothing method to estimate λ_{it}. Specifically, we follow the nonparametric literature on time-varying models (see, e.g., Cai 2007, Robinson 2012, Chen et al. 2012, Su et al. 2015) and model λ_{it} as a nonrandom function of t/T:

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

where $\lambda_i(\cdot)$ is an unknown piece-wise smooth function of t/T on $(0, 1]$ for each i. The specification that $\lambda_i(t/T)$ is a function of ratio $t/T \in (0, 1]$ rather than time index t is a commonly used scaling scheme in the literature. The intuitive explanation to this requirement is that the increasingly intensive sampling of data points ensures consistent estimation of $\lambda_i(t/T)$ for each i at some fixed point t/T by increasing the amount of data on which it depends. For more discussion, see Robinson (1989, 1991).

3 Estimation

In this section, we introduce the local version of PCA to estimate the time-varying factor loadings 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

Assuming that the time-varying factor loadings $\lambda_{ir} = \lambda_i(r/T)$ change smoothly over the ratio r/T for each i, we can consider the local constant approximation of $\lambda_i\left(t/T\right)$ around the fixed time point r/T as follows

$$\lambda_i\left(\frac{t}{T}\right) \approx \alpha_{ir} = \lambda_i\left(\frac{r}{T}\right).$$

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

$$\min_{\{\alpha_{ir}\}_{i=1}^N, \{F_t\}_{t=1}^T} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \left[X_{it} - \alpha_{ir}' F_t\right]^2 K_h\left(\frac{t-r}{T}\right), \tag{3.1}$$

where α_{ir} is an $R \times 1$ parameter vector, $K_h(x) = h^{-1}K(x/h)$ with kernel $K : \mathbb{R} \to \mathbb{R}^+$ and bandwidth $h = h(T, N)$. We note that, the local constant estimators $\{\hat{\alpha}_{ir}\}$ could be regarded as the OLS estimators of the transformed model

$$\tilde{k}_{tr}^{1/2} X_{it} \approx \tilde{k}_{tr}^{1/2} \alpha_{ir}' F_t + \tilde{k}_{tr}^{1/2} e_{it}, \tag{3.2}$$

where $\tilde{k}_{tr} = h^{-1}K((t-r)/(Th))$. Define the $T \times N$ matrices $X^{(r)} = (X_1^{(r)}, X_2^{(r)}, \ldots, X_N^{(r)})$ and $e^{(r)} = (e_1^{(r)}, e_2^{(r)}, \ldots, e_N^{(r)})$, where $X_i^{(r)} = (\tilde{k}_{1r}^{1/2} X_{i1}, \ldots, \tilde{k}_{Tr}^{1/2} X_{iT})'$ and $e_i^{(r)} = (\tilde{k}_{1r}^{1/2} e_{i1}, \ldots, \tilde{k}_{Tr}^{1/2} e_{iT})'$. Let $F^{(r)} = (\tilde{k}_{1r}^{1/2} F_1, \tilde{k}_{2r}^{1/2} F_2, \ldots, \tilde{k}_{Tr}^{1/2} F_T)'$ and $\alpha^{(r)} = (\alpha_{1r}, \alpha_{2r}, \ldots, \alpha_{Nr})'$, which are $T \times R$ and $N \times R$ matrices, respectively. In matrix notation, the transformed model (3.2) can be written as

$$X^{(r)} \approx F^{(r)} \alpha^{(r)\prime} + e^{(r)}, \quad r = 1, 2, \ldots, T.$$

Then the minimization problem in (3.1) becomes

$$\min_{F^{(r)},\alpha^{(r)}} \operatorname{tr}\left[\left(X^{(r)}-F^{(r)}\alpha^{(r)\prime}\right)\left(X^{(r)}-F^{(r)}\alpha^{(r)\prime}\right)'\right] \tag{3.3}$$

By concentrating out $\alpha^{(r)} = X^{(r)\prime}F^{(r)}(F^{(r)\prime}F^{(r)})^{-1} = X^{(r)\prime}F^{(r)}/T$ under the restriction $F^{(r)\prime}F^{(r)}/T = \mathbb{I}_R$, the objective function becomes

$$\operatorname{tr}\left[X^{(r)\prime}X^{(r)}\right] - T^{-1}\operatorname{tr}\left[F^{(r)\prime}X^{(r)}X^{(r)\prime}F^{(r)}\right].$$

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

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

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

where $\lfloor a \rfloor$ denote the integer part of a. By using this boundary kernel, we define the weight $k_{h,tr} = K_h^*\left(\frac{t-r}{T}\right)$ and use it to replace $K_h\left(\frac{t-r}{T}\right)$ in (3.1) in the following analysis. Note that $k_{h,tr}$ coincides with $\tilde{k}_{tr}$ in the interior region but not in the boundary regions. By replacing $\tilde{k}_{tr}$ with $k_{h,tr}$ in (3.2) and (3.3), we obtain the estimators (still denoted as $\hat{F}_t^{(r)}$ and $\hat{\lambda}_{ir}$) to be analyzed below.

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 $F_t, t = 1, 2, \ldots, T$, in the second stage, by minimizing the following least squares problem:

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

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 $t = 1, 2, \ldots, T$.

3.2 Limiting distributions of the estimated factors and factor loadings

In this subsection, we establish the asymptotic distributions of the estimated common factors and time-varying factor loadings.

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

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

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

(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 $N^{-1}\Lambda'_r \Lambda_r = \Sigma_{\Lambda_r} + O\left(N^{-1/2}\right)$ 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 $m, n = 1, ..., R$, where $F_{t,m}$ denotes the mth 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) $\max_{1 \leq s,t \leq T} E\left|N^{1/2}\zeta_{st}\right|^4 \leq C$ and $\max_{r,t} E\left\|N^{-1/2}\Lambda'_r e_t\right\|^4 \leq C$.

(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

$$\Gamma_{rt} = \lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_{ir}\lambda'_{jr} E(e_{it}e_{jt}).$$

(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

$$\Omega_{i,r} = \lim_{T\to\infty} \left[\frac{h}{T} \sum_{s=1}^{T} k_{h,sr}^2 E\left(F_s F'_s e_{is}^2\right) + \frac{2h}{T} \sum_{s=1}^{T-1} \sum_{t=s+1}^{T} k_{h,sr} k_{h,tr} E(F_s F'_t e_{is} e_{it}) \right]. \quad (3.4)$$

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

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

A.1 mainly imposes moment conditions on the error terms, factors, factor loadings, and their interactions. They are widely used in the literature; see, e.g., Bai and Ng (2002) and Bai (2003). Note that we follow Stock and Watson (2002), Bai (2003), and Breitung and Eickmeier (2011) and assume that $E[F_t F'_t]$ is homogeneous over t in A.1(ii), which facilitates the derivation of the asymptotic results significantly. With more complicated and lengthy arguments, we can allow for time-varying covariance for the factor loadings. Similarly, following Bai (2003) and Breitung and Eickmeier (2011), we assume that the factor loadings are nonrandom in A.1(iii) because they are treated as functions of time. A.2 is used to establish the asymptotic normality of our local PCA estimators and can be verified under some primitive conditions. For example, by the central limit theorem (CLT hereafter) for strong mixing processes (e.g., White 2001, Theorem 5.20), one can readily verify A.2(ii). Using Davydov inequality, we can argue that the limit $\Omega_{i,r}$ in (3.4) exists. Without further assumptions, we cannot simplify it. If $E\left(F_s F_s' e_{is}^2\right) = \Omega_i$ for each s and $\{e_{it}\}$ is an m.d.s. with respect to $\mathcal{F}_{it}$, the sigma-field generated from $\{e_{i,t-1}, e_{i,t-2}, ..., F_t, F_{t-1}, ...\}$, then we can readily show that $\Omega_{i,r} = \Phi_i \lim_{T\to\infty} \frac{1}{Th} \sum_{s=1}^{T} K_r^* \left(\frac{s-r}{Th}\right)^2 = \Phi_i \int_{-1}^{1} K(u)^2 \, du$ if $r \in [\lfloor Th \rfloor, T - \lfloor Th \rfloor]$. A.3 imposes regularity conditions on the kernel function and bandwidth.

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

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

where $H^{(r)} = (N^{-1}\Lambda_r'\Lambda_r)(T^{-1}F^{(r)\prime}\hat{F}^{(r)})V_{NT}^{(r)-1}$, $V_{NT}^{(r)}$ denotes the $R \times R$ diagonal matrix of the first R largest eigenvalues of $(NT)^{-1} X^{(r)} X^{(r)\prime}$, 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 $\hat{F}_t^{(r)}$. 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 $H^{(r)\prime}$. Since we allow cross sectional dependence in the error terms, the limiting distribution depends on the cross-section correlation structure among $\{e_{it}\}$. In the case where e_{it} is uncorrelated over i, we have $\Gamma_{rt} = \lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} \lambda_{ir}\lambda_{ir}'\sigma_{it}^2$ 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 $i = 1, 2, \ldots, N$ and $r = 1, 2, \ldots, T$, we have:*

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

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

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

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

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

Remark. Interestingly, although the convergence rates of $\hat{F}_t^{(r)}$ and $\hat{\lambda}_{it}$ depend on the bandwidth h, the estimated factor $\hat{F}_t$ could achieve the usual parametric $\sqrt{N}$-rate of convergence. In addition, even though we apply the nonparametric local smoothing method, we do not have the usual asymptotic bias-variance tradeoff for the estimators of either the factors or the factor loadings because neither estimators possess the usual asymptotic bias terms. As a result, we can not derive the conventional optimal bandwidth in terms of minimizing the asymptotic mean square error of the nonparametric estimates. In practice, we suggest using some data-driven methods to choose the bandwidth. For example, one can use the cross-validation method to choose the bandwidth $\hat{h}$ by solving the following minimization problem:

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

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

3.3 Determination of the number of factors

In the above analysis, we assume that the number of factors, R, is known. In practice, one has to determine R from the data. Here we assume that the true value of R, denoted as R_0, is bounded from above by a finite integer R_{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

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

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

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

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

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

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.4. (i) $\|e\|_{\text{sp}} = O_P\left(N^{1/2} + T^{1/2}\right)$.

(ii) $\max_{t,s} \left|\frac{1}{N} \sum_{i=1}^{N} \left[e_{it} e_{is} - E\left(e_{it} e_{is}\right)\right]\right| = O_P(N^{-1/2} (\ln T)^{1/2})$,

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

(iv) $\max_{s,t} \left\|N^{-1/2} \Lambda_s' e_t F_t'\right\|_4 \le C$ and $\max_{s,t} \left\|N^{-1/2} \left[F_s e_s' e_t F_t' - E\left(F_s e_s' e_t F_t'\right)\right]\right\|_2 \le C$.

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

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

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

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

Theorem 3.4 indicates the class of information criteria defined by $IC(R)$ can consistently estimate R_0. To implement the information criterion, one still 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.

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

$$\mathbb{H}_0 : \lambda_{it} = \lambda_{i0} \text{ for } i = 1, 2, \ldots, N \text{ and } t = 1, 2, \ldots, T, \tag{4.1}$$

and the alternative hypothesis is

$$\mathbb{H}_1 : \lambda_{it} \neq \lambda_{i0} \text{ for some non-negligible values of } (i, t), \tag{4.2}$$

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

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. Nevertheless, it is well known that factor models may exhibit structural changes over time. For this reason, much recent research has focused on testing for structural changes in factor models; see Breitung and Eickmeier (2011), Chen et al. (2014), Cheng et al. (2014), and Han and Inous (2014). These authors aim at testing the existence of a single structural change in the factor loadings by using some supremum-type test statistics. However, usually no prior information about the structural change alternative is available in practice. It is extremely restrictive to assume only a single sudden structural break in the factor loadings. In contrast, we do not impose any essential restriction on the alternative. The alternative (4.2) allows for a finite number of abrupt structural breaks. More importantly, by assuming 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 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 $X_t \equiv (X_{1t}, \ldots, X_{Nt})'$, $e_t \equiv (e_{1t}, \ldots, e_{Nt})'$, $e_t^{\dagger} \equiv (e_{1t}^{\dagger}, ..., e_{Nt}^{\dagger})'$, $F \equiv (F_1, \ldots, F_T)'$, and $\Lambda_0 \equiv (\lambda_{10}, ..., \lambda_{N0})'$. Let $X = (X_1', ..., X_T')'$, $e \equiv (e_1', \ldots, e_T')'$, $e^{\dagger} \equiv (e_1^{\dagger}, ..., e_T^{\dagger})'$.

Then we can rewrite (2.1) in matrix form

$$X = F\Lambda_0' + e^{\dagger}. \tag{4.3}$$

The conventional PCA method solves the following minimization problem

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

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 $\tilde{F} = (\tilde{F}_1, ..., \tilde{F}_T)'$ and $\tilde{\Lambda}_0 = (\tilde{\lambda}_{1,0}, ..., \tilde{\lambda}_{N0})'$. It is well known that $\tilde{F}$ is $\sqrt{T}$ times eigenvectors corresponding to the R largest eigenvalues of the $T \times T$ matrix XX', and $\tilde{\Lambda}_0' = (\tilde{F}'\tilde{F})^{-1}\tilde{F}'X = T^{-1}\tilde{F}X$.

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:

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

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 $\|A\|_r = \{E\|A\|^r\}^{1/r}$ for $r \geq 1$. We add the following assumptions.

Assumption A.6. (i) For each $i = 1, 2, \ldots, N$, the process $\{e_{it}, t = 1, 2, \ldots\}$ is a martingale difference sequence (m.d.s. hereafter) such that $E\left(e_{it}|\mathcal{F}_{NT,t-1}\right) = 0\, \forall t$, where $\mathcal{F}_{NT,t-1} = \{F_t, F_{t-1}, \ldots, e_{t-1}, e_{t-2}, \ldots\}$.

(ii) For each $i = 1, 2, \ldots, N$, the process $\{(e_{it}, F_t), t = 1, 2, ...\}$ is strong mixing with mixing coefficients $\alpha_i(\cdot s)$. $\alpha(\cdot) \equiv \max_i \alpha_i(\cdot)$ 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 $T^{-2}\max(T_0^4, T_0^3h^{-1}, T_0^2h^{-2}) \to 0$ and $N^2Th^2\alpha(T_0)^{\delta/(1+\delta)} \to 0$ as $(N, T) \to \infty$.

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

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

Assumption A.7. (i) $\max_{s,t}\left\|N^{-1}\Lambda_s'e_t\right\| = O_P(N^{-1/2}(\ln T)^{1/2})$ and $\max_s\left\|N^{-1}\Lambda_s'e_sF_s\right\| = O_P(N^{-1/2}(\ln T)^{1/2})$,

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

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

A.6(i) assumes that the process $\{e_{it}, t = 1, 2, ...\}$ is a m.d.s. with respect to the filter $\{\mathcal{F}_{NT,t}\}$. This assumption is essential for the establishment of the asymptotic distribution of our test statistic under the null hypothesis and a sequence of Pitman local alternatives. A.6(ii) requires the process $\{(e_{it}, F_t), t = 1, 2, ...\}$ 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.6(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.6(iv) controls the cross-sectional dependence among $\{F_t e_{it},\ i = 1, 2, ..., N\}$. Under A.6(iii), this condition becomes redundant if we would assume independence of $e_{i\cdot} = (e_{i1}, ..., e_{iT})'$ across i conditional on the factors. A.7 imposes conditions on the uniform probability order of some summation objects. Again, these conditions can be easily verified by using Bernstein-type exponential inequality for independent or strong mixing processes.

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

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

Let V_{NT} denote the $R \times R$ diagonal matrices of the first R largest eigenvalues of $(NT)^{-1} XX'$ in decreasing order and $H = (N^{-1}\Lambda_r' \Lambda_r)(T^{-1} F' \tilde{F}) V_{NT}^{-1}$. 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 $L_{st} = k_{h,st} H^{(t)} H^{(t)\prime} - HH'$. Define

$$\mathbb{B}_{NT} = \frac{h^{1/2}}{N^{1/2} T^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(F_t' L_{st} F_s\right)^2 e_{is}^2,$$

$$\mathbb{V}_{NT} = 2T^{-2} N^{-1} h^{-1} \sum_{1 \leq r \neq s \leq T} \bar{k}_{sr}^2 E\left[\left(F_s' H_0 \bar{\Sigma}_F H_0' F_r\right)^2 \left(e_r' e_s\right)^2 \right],$$

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 (ii*), and A.6-A.7 hold. Then the test statistic* $J_{NT} \equiv \mathbb{V}_{NT}^{-1/2} \left(T N^{1/2} h^{1/2} \hat{M} - \mathbb{B}_{NT} \right) \xrightarrow{d} N(0,1)$ *under* $\mathbb{H}_0$.

The proof of the above theorem is quite involved and is relegated to Appendix B. We make some remarks. First, each of the four terms, namely, $\hat{\lambda}_{it}$, $\hat{F}_t$, $\tilde{\lambda}_{i0}$, 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.6(i), one can verify that $E(Z_{NT,s}|\mathcal{F}_{NT,s-1}) = 0$ and resort to a martingale central limit theorem (e.g., Pollard 1984, p.171) to derive the asymptotic distribution of J_{NT}. Difficulty arrives when we try to verify the Lyapunov condition via the fourth order moment of $Z_{NT,s}$ because we do not assume cross-sectional independence among $e_{i\cdot} = (e_{i1}, ..., e_{iT})'$ conditional on the factors. The strong mixing condition in A.6(ii) and the moment conditions in A.6(iii)-(iv) greatly facilitate the verification of the Lyapunov condition. Third, despite the assumed m.d.s. condition, the variance term $\mathbb{V}_{NT}$ still takes the form of a double U-statistic that involves two summations over each of the individual and time dimensions.

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

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

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*), and A.6-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:

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

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

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

for location normalization purpose. With this normalization, both λ_{i0} and $g_i\left(\cdot\right)$ can be dependent on the sample sizes N and T. But for notational simplicity, we continue to write them as λ_{i0} and $g_i\left(\cdot\right)$ instead of $\lambda_{i0,NT}$ and $g_{i,NT}\left(\cdot\right)$.

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

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

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

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

(ii) $\max_{1\le i\le N} \max_{1\le t\le T} \|\lambda_{it}\| \le \bar{c}_\lambda < \infty$.

(iii) $\max_{1\le r\le T} \left\| \frac{1}{NT} \sum_{s=1}^{T} \sum_{i=1}^{N} k_{h,sr} F_s e_{is} g_{ir}' \right\| = O_P\left((NTh/\ln(NT))^{-1/2}\right)$.

(iv) The limits Π_1 and Π_2 defined in (4.5) exist and $\Pi_1 + \Pi_2 > 0$.

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

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

Theorem 4.3 *Suppose that Assumptions A.1, A.3(i) and (ii*), and A.6-A.8 hold. Then under* $\mathbb{H}_1\left(a_{NT}\right)$ *with* $a_{NT} = T^{-1/2} N^{-1/4} h^{-1/4}$, $\hat{\mathbb{B}}_{NT} = \mathbb{B}_{NT} + o_P\left(1\right)$, $\hat{\mathbb{V}}_{NT} = \mathbb{V}_{NT} + o_P\left(1\right)$, *and* $\hat{J}_{NT} \xrightarrow{d} N(\pi_0, 1)$, *where* $\pi_0 = \left(\Pi_1 + \Pi_2\right)/\mathbb{V}_0^{1/2}$ *and* $\mathbb{V}_0 = \lim_{(N,T)\to\infty} \mathbb{V}_{NT}$.

Theorem 4.3 implies that our test has nontrivial asymptotic power against the class of local alternatives that deviate from the null hypothesis at the rate $a_{NT} = T^{-1/2} N^{-1/4} h^{-1/4}$. Note that we allow the existence of a finite number of unknown discontinuity points in factor loadings. As a result, our test has power against not only the smooth structural changes in factor loadings but also a finite number of abrupt changes.

4.5 Asymptotic global power

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

$$\boldsymbol{F}_T = \left\{ \breve{F} : \breve{F}'\breve{F} = \mathbb{I}_R \right\} \text{ and } \boldsymbol{\Lambda}_N = \left\{ \breve{\Lambda} : \breve{\Lambda}'\breve{\Lambda} = \text{diagonal matrix} \right\},$$

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

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

Assumption A.9 is intuitively clear: in the spaces of factors and factor loadings such that the normalization rules in $\boldsymbol{F}_T$ and $\boldsymbol{\Lambda}_N$ are satisfied, we cannot find any time-invariant factor loadings $\breve{\lambda}_i$'s and the associated factors $\breve{F}_t$'s such that $\breve{\lambda}_i'\breve{F}_t$ converges to the true common component $\lambda_{it}'F_t$ in the sense of mean square error.

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

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 $e_{it}^{\star} = \tilde{\sigma}_i \varsigma_{it}$ with ς_{it} being IID $N(0, 1)$ over both i and t. Then one can generate the bootstrap resamples via $X_{it}^{*} = \tilde{\lambda}_{i0}'\tilde{F}_t + e_{it}^{\star}$ 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 $\{e_{it}^{\star}\}$ 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 $\tilde{e}_t = (\tilde{e}_{1t}, ..., \tilde{e}_{Nt})'$ 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

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

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 $\max_{i,j} \left|\tilde{\sigma}_{ij}^0\right|$ is. By construction, $\tilde{\Sigma}$ is also symmetric and positive semi-definite. The stochastic boundedness of $\max_{i,j} \left|\tilde{\sigma}_{ij}^0\right|$ 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 $i = 1, ..., N$ and $t = 1, 2, \ldots, T$, obtain the bootstrap error $e_t^* = \tilde{\Sigma}^{1/2} \varsigma_t$, where $\varsigma_t = (\varsigma_{1t}, \ldots, \varsigma_{Nt})'$ 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*) and A.6-A.7 hold. Suppose that (i)* $\max_{i,j} \left|\tilde{\sigma}_{ij}^0\right| = O_P(\zeta_{NT})$ *with* $\zeta_{NT} = o\left(T^{1/2}\right)$, *(ii)* $\frac{1}{T}\sum_{t=1}^T \left\|\tilde{F}_t\right\|^8 = O_P(1)$ *and (iii)* $\frac{1}{N}\sum_{i=1}^N \left\|\tilde{\lambda}_{i0}\right\|^8 = O_P(1)$. *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 as long as we generate the bootstrap data by imposing the null hypothesis. If the null hypothesis does not hold in the observed sample, then we expect $\hat{J}_{NT}$ 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 B.7(i) to demonstrate that $\frac{1}{T}\sum_{t=1}^T \left\|\tilde{F}_t\right\|^8 = O_P(1) + O_P\left(T^3 C_{0NT}^{-8}\right) = O_P(1)$ and similarly $\frac{1}{N}\sum_{i=1}^N \left\|\tilde{\lambda}_{i0}\right\|^8 = O_P(1)$.

5 Monte Carlo Study

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

5.1 Data generating process

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

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

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)

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

DGP 2: (Heteroskedasticity)

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

DGP 3: (Cross sectional dependence)

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

DGP 4: (Single structural break)

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

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

DGP 5: (Multiple structural breaks)

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

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

DGP 6: (Smooth structural changes 1)

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

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

DGP 7: (Smooth structural changes 2)

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

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

DGP 8: (Smooth structural changes 1 + cross sectional dependence)

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

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

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

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

5.2 Determination of the number of factors

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

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

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. Our local PCA involves nonparametric estimation. We use the Epanechnikov kernel and Silverman's rule of thumb (RoT) to set the bandwidth as $h=(2.35/\sqrt{12})T^{-1/5}N^{-1/10}$. [Note that $\{t/T\}_{t=1}^{T}$ behaves like a uniform random variable on $[0,1]$ and thus has variance 1/12.] We also try the Uniform kernel and the Quartic kernel, and the RoT bandwidth with different tuning parameters. Our simulation studies show that the choice of kernel function and the bandwidth has little impact on the performance of our information criteria. Each series is demeaned and standardized to have unit variance.

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

Tables 1 and 2 report the average number of common factors and the empirical probability of correct selection over 1000 replications of various information criteria in determining the number of common factors. DGPs 1-3 satisfy the null hypothesis of time-invariant factor loadings and allow us to compare the performance of these information criteria for the conventional factor models. DGPs 4-8 are the time-varying factor models with abrupt or smooth structural changes, where the value of b measure the magnitude of structural changes. To check the sensitivity of the information criteria to the magnitude of structural changes, we consider the case $b = 1, 2, 4$ for DGPs 4-8.

As shown in the tables, our information criteria work fairly well for all the DGPs under investigation. For the conventional factor models with IID, heteroskedastic, and cross sectionally dependent error terms 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 multiple 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 factors. Onatski's (2009) testing procedure also tends to choose 3 common factors except for the case of DGP 5 with $b = 1$, which is merely acceptable with larger than 70% correct selection probability. Ahn and Horenstein's (2013) ER and GR criterion functions perform well for the case of DGP 5 with $b = 1$, but they still suffer from severe over- or under- selection for other cases. In contrast, although our information criteria are proposed for smooth structural changes, they still work well for small and moderate magnitude ($b = 1, 2$) of abrupt structural breaks. Although they tend to slight overparameterize for $b = 4$, the results are still acceptable and much better than those of other information criteria. DGPs 6-8 are factor models with smooth structural changes in factor loadings and/or cross sectionally dependent errors. As shown in Table 2, our information criteria give precise estimates of the number of common factors for all cases. However, the criteria proposed by Bai and Ng (2002), Onatski (2009) and Ahn and Horenstein (2013) work poorly except for the case of small structural changes ($b = 1$).

5.3 Performance of the test

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

It is well known that a nonparametric test that relies on the asymptotic normal approximation may perform poorly in finite samples. To conquer this problem we consider the wild bootstrap procedure proposed in Section 4.6. Since the bootstrap procedure is rather time consuming, we generate 500 data sets in this subsection and set the bootstrap replication number B to be 200. As in the previous subsection, we

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

		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
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	3.00	2.39	1.67	2.01	.981	.998	.000	.000	.003	.003	.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	3.00	2.92	1.50	2.06	.998	.999	.000	.000	.000	.000	.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	2.12	2.01	2.00	2.00	.997	1.00	.317	.664	.879	.879	.992	1.00	1.00
	(100,200)	2.00	2.00	2.90	2.72	2.36	2.36	2.02	2.00	2.00	1.00	1.00	.099	.282	.638	.638	.980	1.00	1.00
	(200,100)	2.00	2.00	2.64	2.46	2.26	2.26	2.26	2.00	2.00	.999	1.00	.362	.542	.739	.739	.736	1.00	1.00
	(200,200)	2.00	2.00	2.95	2.79	2.78	2.78	2.23	2.00	2.00	1.00	1.00	.047	.214	.225	.225	.770	1.00	1.00
b=2	(100,100)	2.01	2.00	3.00	3.00	3.00	3.00	2.76	2.10	2.30	.992	.999	.000	.001	.005	.005	.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

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 6-8

		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
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	.999	.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	2.01	2.21	2.00	2.00	.997	.999	.851	.948	.986	.986	.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	3.00	2.93	1.91	2.33	1.00	1.00	.000	.000	.003	.003	.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	3.83	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	3.96	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
7																			
b=1	(100,100)	2.01	2.00	2.80	2.33	2.10	2.10	2.00	2.00	2.00	.995	1.00	.224	.671	.904	.904	1.00	1.00	1.00
	(100,200)	2.00	2.00	2.92	2.71	2.27	2.27	2.01	2.00	2.00	1.00	1.00	.088	.291	.726	.726	.994	1.00	1.00
	(200,100)	2.00	2.00	2.59	2.38	2.18	2.18	2.18	2.00	2.00	1.00	1.00	.409	.616	.825	.825	.821	.999	1.00
	(200,200)	2.00	2.00	2.94	2.68	2.66	2.66	2.09	2.00	2.00	1.00	1.00	.063	.318	.338	.338	.909	.999	1.00
b=2	(100,100)	2.01	2.00	3.18	3.00	2.97	2.97	2.17	1.69	2.00	.990	.998	.000	.002	.030	.030	.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	.000	.435	.735	.994
	(200,100)	2.00	2.00	3.00	3.00	2.98	2.98	2.86	1.61	2.04	1.00	1.00	.002	.004	.025	.025	.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	.001	.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	.882	.991	.999	.999	.995	1.00	1.00
	(100,200)	2.00	2.00	2.11	2.11	2.00	2.00	2.00	2.00	2.00	1.00	1.00	.895	.985	.999	.999	.999	1.00	1.00
	(200,100)	2.00	2.00	2.11	2.03	2.01	2.01	2.18	2.00	2.00	.997	.999	.895	.970	.993	.993	.819	1.00	1.00
	(200,200)	2.00	2.00	2.38	2.07	2.08	2.20	2.00	2.00	1.00	1.00	.619	.931	.917	.917	.917	.801	1.00	1.00
b=2	(100,100)	2.00	2.00	3.00	3.00	3.00	3.00	2.93	1.90	2.31	1.00	1.00	.000	.000	.005	.005	.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	.000	.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	3.86	3.78	3.78	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	3.99	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	3.95	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

Note: See the note in Table 1.

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

DGP	N	T	J_{NT}		HI_{LM}		HI_W		CDG_{LM}		CDG_W		BE_{LM}	
			5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
1	100	100	5.4	11.4	3.6	7.2	0.4	2.0	2.2	8.2	5.8	11.0	2.8	6.5
	100	200	5.0	10.0	6.0	11.6	1.4	5.6	4.8	10.2	6.0	10.2	3.5	7.4
	200	100	5.2	11.2	3.2	6.0	0.0	1.6	2.8	8.6	6.4	12.4	2.7	6.2
	200	200	3.8	8.8	7.0	14.4	2.0	7.8	5.8	11.6	6.6	12.2	3.4	7.5
2	100	100	6.8	14.2	3.0	8.4	0.4	1.8	3.2	9.8	6.2	14.4	2.9	6.6
	100	200	5.8	11.0	3.2	8.2	2.6	7.8	4.4	11.2	5.4	14.2	3.7	7.8
	200	100	7.4	12.8	2.8	6.8	0.4	2.0	3.6	9.2	7.4	15.2	2.8	6.4
	200	200	6.2	11.6	3.6	9.6	2.2	7.4	3.4	9.2	5.8	13.2	3.6	7.7
3	100	100	6.0	11.0	3.6	8.2	0.2	2.2	2.8	7.4	5.4	10.4	2.7	6.4
	100	200	4.2	8.6	7.0	12.8	1.8	6.6	4.6	8.8	5.0	10.8	3.4	7.5
	200	100	4.8	11.2	3.2	6.0	0.2	2.0	3.0	7.2	5.6	10.8	2.8	6.3
	200	200	4.2	8.6	7.8	13.2	1.8	6.8	6.4	11.6	7.2	13.0	3.4	7.4

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

use the Epanechnikov kernel and the rule of thumb bandwidth $h = (2.35/\sqrt{12})T^{-1/5}N^{-1/10}$. In addition to our test, we also consider Breitung and Eickmeier's (2011) sup-LM variable-specific test, Chen et al. (2014) sup-LM and sup-Wald tests and Han and Inoue's (2014) sup-LM and sup-Wald tests. We follow these papers to set the trimming parameter $\tau = 0.15$. The tests of Chen et al. (2014) and Han and Inoue (2014) involve the long run variance estimation. We set the time-lag truncation parameter as $m = \lfloor T^{1/5} \rfloor$ 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 (2014), while the bootstrap critical values are applied to check the performance of our test.

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

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

6 An Application to Stock and Watson's (2009) U.S. Macroeconomic Data Set

In this section, we apply our approach to check whether the U.S. economy suffers from structural changes. The data set, constructed by Stock and Watson (2009), consists of 144 quarterly time series, spanning 1959:I-2006:IV. By excluding the first two quarters, which will be missing when computing 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 first or second order (log-)difference and removing outliers. All the data have been standardized to have zero mean and unit variance. For the details of the data description and processing, one can refer to Stock and Watson (2009).

We first determine the appropriate number of common factors. The maximum number of common factors is set to be 8 in this empirical study. Other presettings such as the kernel and bandwidth are the same as in the simulation section. We use Bai and Ng's (2002) information criteria $PC_{p1}, PC_{p2}, IC_{p1}$, 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 5. According to the table, we report the test results for the cases of one to five common factors respectively in the following context .

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

Table 4: Power of Tests Under DGPs 4-8

DGP	N	T	J_{NT}		HI_{LM}		HI_W		CDG_{LM}		CDG_W		BE_{LM}	
			5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
4, $b = 1$	100	100	99.0	99.6	39.6	62.6	21.0	43.8	5.2	10.6	31.4	42.8	34.9	41.5
	100	200	100	100	95.2	98.0	92.6	97.0	8.0	15.0	33.2	42.8	48.5	54.2
	200	100	99.4	99.6	41.2	65.0	22.2	46.0	6.6	11.6	30.2	41.2	35.7	43.2
	200	200	100	100	96.0	98.6	93.4	98.0	7.8	17.0	33.2	43.2	52.9	59.0
4, $b = 2$	100	100	100	100	70.8	85.6	52.0	73.8	5.2	11.0	63.4	72.0	43.9	50.1
	100	200	100	100	99.6	100	99.4	100	8.8	17.0	65.6	72.4	57.6	63.1
	200	100	100	100	71.4	86.2	51.2	75.0	6.8	13.8	61.4	69.0	46.7	53.6
	200	200	100	100	99.8	100	99.2	100	9.8	19.2	68.6	75.4	63.7	69.0
5, $b = 1$	100	100	96.0	97.8	4.2	10.4	0.6	3.8	5.4	11.8	9.4	18.6	7.3	12.9
	100	200	100	100	9.6	21.8	8.6	19.4	12.2	22.8	17.0	27.4	16.5	23.7
	200	100	99.4	99.8	3.8	9.6	0.6	3.8	4.0	12.6	9.0	19.6	7.1	12.7
	200	200	100	100	8.0	19.8	7.4	17.4	9.2	18.6	13.8	23.2	16.3	23.8
5, $b = 2$	100	100	100	100	6.0	16.2	1.4	7.2	10.4	20.4	25.8	37.6	15.2	22.7
	100	200	100	100	29.4	48.6	26.8	42.0	32.6	49.2	50.6	64.2	32.1	39.7
	200	100	100	100	5.4	14.4	1.2	7.4	9.4	18.0	24.2	34.8	16.8	24.7
	200	200	100	100	28.8	46.0	25.8	41.2	26.6	40.4	45.4	58.0	35.4	43.5
6, $b = 1$	100	100	100	100	85.2	94.8	67.0	88.6	28.4	41.8	90.2	93.4	53.1	62.6
	100	200	100	100	100	100	100	100	56.0	63.0	97.4	97.8	76.7	82.3
	200	100	100	100	84.6	95.2	68.6	89.2	32.6	47.0	80.4	85.0	53.9	63.1
	200	200	100	100	100	100	100	100	68.2	74.2	90.4	92.4	77.3	82.6
6, $b = 2$	100	100	100	100	88.2	96.6	66.8	88.4	8.4	15.4	98.6	99.0	85.2	89.3
	100	200	100	100	100	100	100	100	13.4	22.6	99.6	99.8	96.2	97.9
	200	100	100	100	87.2	96.4	67.2	89.4	6.6	15.0	98.6	99.4	85.3	89.5
	200	200	100	100	100	100	100	100	11.8	20.4	99.8	100	96.7	98.3
7, $b = 1$	100	100	95.0	97.8	11.8	19.6	0.2	1.6	4.4	10.0	3.4	8.4	4.5	9.2
	100	200	100	100	23.6	36.0	2.0	4.6	6.4	12.8	2.6	6.2	9.0	15.4
	200	100	99.4	99.4	10.8	19.8	0.2	1.6	3.6	8.6	3.2	7.8	4.7	9.6
	200	200	100	100	23.6	37.0	2.0	4.8	5.8	10.8	2.4	6.6	10.3	17.1
7, $b = 2$	100	100	97.4	98.6	19.6	33.0	0.2	1.4	5.4	10.4	2.4	7.8	6.2	11.6
	100	200	100	100	43.4	52.8	1.6	4.4	8.0	14.4	2.4	5.6	14.5	22.2
	200	100	99.8	99.8	19.4	33.0	0.2	1.2	4.8	10.8	2.4	7.0	6.9	12.7
	200	200	100	100	43.2	53.8	1.6	4.4	7.0	12.2	2.2	4.4	16.7	24.6
8, $b = 1$	100	100	92.2	96.0	82.0	94.8	66.2	88.0	39.2	54.8	70.8	79.4	14.8	24.1
	100	200	100	100	99.8	100	100	100	61.2	69.6	77.4	82.0	34.7	45.6
	200	100	94.2	96.8	78.0	93.6	69.2	87.0	38.4	53.2	60.0	70.0	15.4	25.0
	200	200	100	100	100	100	99.4	99.8	53.8	65.0	62.0	69.8	37.2	48.3
8, $b = 2$	100	100	100	100	86.8	96.8	67.6	87.6	6.8	14.0	98.8	99.2	52.7	62.1
	100	200	100	100	99.8	100	100	100	10.4	19.8	98.8	99.6	76.4	81.9
	200	100	100	100	82.6	94.4	68.8	87.2	9.8	17.4	97.6	98.6	52.9	62.3
	200	200	100	100	100	100	99.4	99.8	9.6	15.2	99.8	99.8	77.3	82.6

Note: See the note in Table 3.

Table 5: 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 note in Table 1.

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

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

Note: (i) Under J_{NT} and sup-LM and sup-Wald are the values of the corresponding test statistics; (ii) Under 5% and 10% are the corresponding bootstrap critical values (our test, 500 bootstrap resamples) or asymptotic critical values (Han and Inoue's and Chen et al.'s tests). Bold elements denote significance at the 5% nominal level.

the results of our simulation studies that the tests of Han and Inoue (2014) and Chen et al. (2014) have relatively low power.

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

7 Conclusion

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

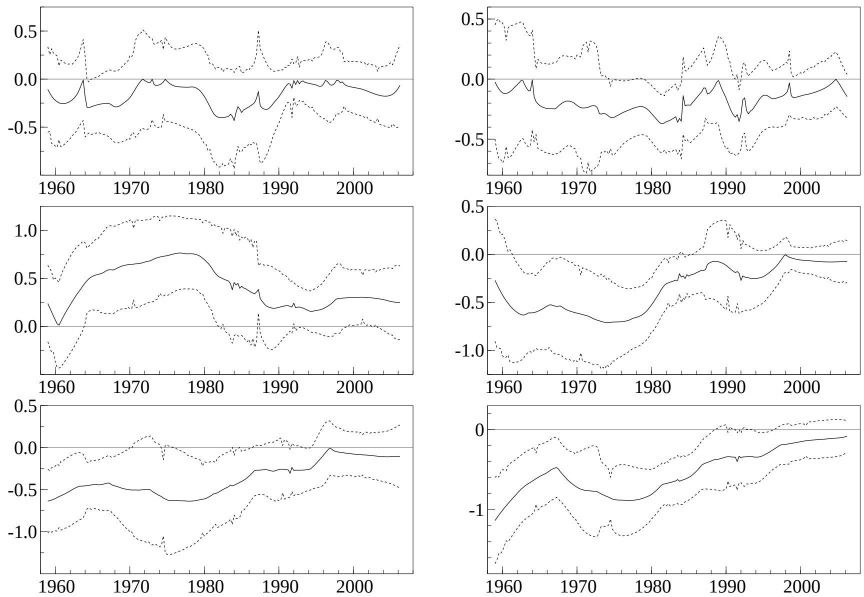

Figure 1: Plots of estimated factor loadings and their 90% confidential intervals for real personal consumption expenditures (left panel) and industrial production index of durable goods (right panel) corresponding to 3 common factors

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.

Mathematical Appendix

This appendix provides the proofs of theorems in Sections 3 and 4. Recall that $V_{NT}^{(r)}$ and V_{NT} denote the $R \times R$ diagonal matrices of the first R largest eigenvalues of $(NT)^{-1} X^{(r)} X^{(r)\prime}$ (for $r = 1, ..., T$) and $(NT)^{-1} XX'$ in decreasing order, respectively. Let $H^{(r)} = (N^{-1}\Lambda_r'\Lambda_r)(T^{-1}F^{(r)\prime}\hat{F}^{(r)})V_{NT}^{(r)-1}$ and $H = (N^{-1}\Lambda_r'\Lambda_r)(T^{-1}F^{'}\tilde{F})V_{NT}^{-1}$. Let $C_{NT} = \min\{\sqrt{Th}, \sqrt{N}\}$.

A Proofs of Theorems in Section 3

We first state two lemmas that are useful in proving the main results in this paper. The proofs of these lemmas are available in the online supplementary material.

Lemma A.1 *Suppose that Assumptions A.1 and A.3 hold. Then*

(i) $T^{-1}\hat{F}^{(r)\prime}\left[(NT)^{-1} X^{(r)} X^{(r)\prime}\right]\hat{F}^{(r)} = V_{NT}^{(r)} = V_r + O_P\left(C_{NT}^{-1}\right)$,

(ii) $T^{-1}\hat{F}^{(r)\prime}F^{(r)} = Q_r + O_P\left(C_{NT}^{-1}\right)$,

(iii) $H^{(r)} = Q_r^{-1} + O_P\left(C_{NT}^{-1}\right)$,

where 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, and* $Q_r = V_r^{1/2}\Upsilon_r^{-1}\Sigma_{\Lambda_r}^{-1/2}$.

Lemma A.2 *Suppose that Assumptions A.1 and A.3 hold. Then*

(i) $\frac{1}{T}\left\|\hat{F}^{(r)} - F^{(r)}H^{(r)}\right\|^2 = \frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_t^{(r)} - H^{(r)\prime}F_t^{(r)}\right\|^2 = O_P(C_{NT}^{-2})$,

(ii) $\frac{1}{T}\left\|\left(\hat{F}^{(r)} - F^{(r)}H^{(r)}\right)' F^{(r)}H^{(r)}\right\| = O_P(C_{NT}^{-2})$,

(iii) $\frac{1}{T}\left\|\left(\hat{F}^{(r)} - F^{(r)}H^{(r)}\right)' \hat{F}^{(r)}\right\| = O_P(C_{NT}^{-2})$.

Proof of Theorem 3.1. Noting that $(NT)^{-1}X^{(r)}X^{(r)\prime}\hat{F}^{(r)} = \hat{F}^{(r)}V_{NT}^{(r)}$ and $X_{it}^{(r)} = \lambda_{ir}'F_t^{(r)} + e_{it}^{(r)}$, we can decompose $\hat{F}_t^{(r)} - H^{(r)\prime}F_t^{(r)}$ as follows:

$$\begin{aligned}
\hat{F}_t^{(r)} - H^{(r)\prime}F_t^{(r)} &= V_{NT}^{(r)-1}\frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^{(r)}X_s^{(r)\prime}X_t^{(r)} - H^{(r)\prime}F_t^{(r)} \\
&= V_{NT}^{(r)-1}\frac{1}{NT}\sum_{s=1}^{T}\hat{F}_s^{(r)}\left[\Lambda_r F_s^{(r)} + e_s^{(r)}\right]'\left[\Lambda_r F_t^{(r)} + e_t^{(r)}\right] - H^{(r)\prime}F_t^{(r)} \\
&= V_{NT}^{(r)-1}\left\{\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s^{(r)}E(e_s^{(r)\prime}e_t^{(r)}/N) + \frac{1}{T}\sum_{s=1}^{T}\hat{F}_s^{(r)}\left[e_s^{(r)\prime}e_t^{(r)}/N - E(e_s^{(r)\prime}e_t^{(r)}/N)\right]\right. \\
&\quad \left. + \frac{1}{T}\sum_{s=1}^{T}\hat{F}_s^{(r)}F_s^{(r)\prime}\Lambda_r'e_t^{(r)}/N + \frac{1}{T}\sum_{s=1}^{T}\hat{F}_s^{(r)}F_t^{(r)\prime}\Lambda_r'e_s^{(r)}/N\right\} \\
&\equiv A_1(t,r) + A_2(t,r) + A_3(t,r) + A_4(t,r), \quad \text{say.} \qquad \text{(A.1)}
\end{aligned}$$

Note that $V_{NT}^{(r)-1}$ is well defined by Lemma A.1(i) and Assumptions A1(ii)-(iii). By Lemmas A.3(i)-(iii) below $\sqrt{Nh}A_j(t,r) = o_P(1)$, $j = 1,2,4$. It suffices to prove the theorem by showing that $K_r^*\left(\frac{t-r}{Th}\right)^{-1/2}\sqrt{Nh}A_3(t,r) \xrightarrow{d} N\left(0, V_r^{-1}Q_r\Gamma_{rt}Q_r'V_r^{-1}\right)$.

Observe that $\sqrt{Nh}A_3(t,r) = V_{NT}^{(r)-1}\left[\frac{1}{T}\sum_{s=1}^T \hat{F}_s^{(r)}F_s^{(r)'}\right]\left[\frac{h^{1/2}}{\sqrt{N}}\sum_{i=1}^N \lambda_{ir}e_{it}^{(r)}\right]$. By Lemmas A.1(i)-(ii), $V_{NT}^{(r)} \xrightarrow{p} V_r$ and $\frac{1}{T}\sum_{s=1}^T \hat{F}_s^{(r)}F_s^{(r)'} \xrightarrow{p} Q_r$. By Assumption A.2(i), $K_r^*\left(\frac{t-r}{Th}\right)^{-1/2}\frac{h^{1/2}}{\sqrt{N}}\sum_{i=1}^N \lambda_{ir}e_{it}^{(r)} = N^{-1/2}\sum_{i=1}^N \lambda_{ir}e_{it} \xrightarrow{d} N(0,\Gamma_{rt})$. Then by Slutsky theorem, $K_r^*\left(\frac{t-r}{Th}\right)^{-1/2}\sqrt{Nh}A_3(t,r) \xrightarrow{d} N(0, V_r^{-1}Q_r\Gamma_{rt} Q_r'V_r^{-1})$. This completes the proof of Theorem 3.1. ■

Lemma A.3 *Suppose that Assumptions A.1 and A.3 hold. Then*
(i)$\sqrt{Nh}\left[A_1(t,r) + A_2(t,r)\right] = o_P(1)$,
(ii) $\sqrt{Nh}A_4(t,r) = o_P(1)$.

Proof of Theorem 3.2. Noting that $\hat{\Lambda}_r' = T^{-1}\hat{F}^{(r)'}X^{(r)}$, $T^{-1}\hat{F}^{(r)'}\hat{F}^{(r)} = \mathbb{I}_R$, and $X^{(r)} = F^{(r)}\Lambda_r' + e^r$, we have

$$\begin{aligned}
\hat{\lambda}_{ir} &= \frac{1}{T}\hat{F}^{(r)'}X_i^{(r)} = \frac{1}{T}\hat{F}^{(r)'}\left(F^{(r)}\lambda_{ir} + e_i^{(r)}\right) \\
&= \frac{1}{T}\hat{F}^{(r)'}F^{(r)}\lambda_{ir} + \frac{1}{T}\hat{F}^{(r)'}e_i^{(r)} \\
&= H^{(r)-1}\lambda_{ir} + \frac{1}{T}H^{(r)'}F^{(r)'}e_i^{(r)} + \frac{1}{T}\left(\hat{F}^{(r)} - F^{(r)}H^{(r)}\right)'e_i^{(r)} - \frac{1}{T}\hat{F}^{(r)'}\left(\hat{F}^{(r)}H^{(r)-1} - F^{(r)}\right)\lambda_{ir} \\
&\equiv H^{(r)-1}\lambda_{ir} + D_1(i,r) + D_2(i,r) - D_3(i,r), \quad \text{say.}
\end{aligned} \tag{A.2}$$

By Lemmas A.4(i)-(ii) below, $\sqrt{Th}D_l(i,r) = o_P(1)$ for $l = 2$, 3. By Lemma A.1(iii), $H^{(r)} \xrightarrow{p} Q_r^{-1}$. By Assumption A.2(ii), $\frac{\sqrt{h}}{\sqrt{T}}\sum_{s=1}^T k_{h,sr}F_se_{is} \xrightarrow{d} N\left(0,\Omega_{i,r}\right)$. It follows that $\sqrt{Th}D_1(i,r) = \frac{1}{\sqrt{Th}}H^{(r)'}\sum_{s=1}^T K_r^*\left(\frac{s-r}{Th}\right)$ $\times F_se_{is} \xrightarrow{d} N(0, \left(Q_r^{-1}\right)'\Omega_{i,r}Q_r^{-1})$. This completes the proof of Theorem 3.2. ■

Lemma A.4 *Suppose that Assumptions A.1 and A.3 hold. Then*
(i) $\sqrt{Th}D_2(i,r) = o_P(1)$,
(ii) $\sqrt{Th}D_3(i,r) = o_P(1)$.

To prove Theorem 3.3, we need another lemma.

Lemma A.5 *Suppose that Assumptions A.1 and A.3 hold. Then*
(i) $\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T}\left(\hat{F}^{(t)} - F^{(t)}H^{(t)}\right)'e_i^{(t)}\right\|^2 = O\left(T^{-2}h^{-2} + N^{-1}T^{-1}h^{-1} + N^{-3/2}\right)$ *for* $t = 1,2,...,T$;
(ii) $\frac{1}{N}\left\|\hat{\Lambda}_t - \Lambda_tH^{(t)-1\prime}\right\|^2 = \frac{1}{N}\sum_{i=1}^N\left\|\hat{\lambda}_{it} - H^{(t)-1}\lambda_{it}\right\|^2 = O_P(C_{NT}^{-2})$ *for* $t = 1,2,...,T$.

Proof of Theorem 3.3. Noting that $X_{it} = \hat{\lambda}_{it}' H^{(t)\prime} F_t + e_{it} + \left[\hat{\lambda}_{it} - H^{(t)-1}\lambda_{it}\right]' H^{(t)\prime} F_t$, we have

$$\begin{aligned}
\hat{F}_t - H^{(t)'} F_t &= \left(\frac{1}{N}\sum_{i=1}^{N} \hat{\lambda}_{it}\hat{\lambda}_{it}'\right)^{-1} \left(\frac{1}{N}\sum_{i=1}^{N} \hat{\lambda}_{it} X_{it}\right) - H^{(t)'} F_t \\
&= \hat{S}_{\lambda,t}^{-1} H^{(t)-1} \frac{1}{N}\sum_{i=1}^{N} \lambda_{it} e_{it} + \hat{S}_{\lambda,t}^{-1} \frac{1}{N}\sum_{i=1}^{N} \left(\hat{\lambda}_{it} - H^{(t)-1}\lambda_{it}\right) e_{it} \\
&+ \hat{S}_{\lambda,t}^{-1} \frac{1}{N}\sum_{i=1}^{N} \hat{\lambda}_{it} \left[\hat{\lambda}_{it} - H^{(t)-1}\lambda_{it}\right]' H^{(t)\prime} F_t \\
&\equiv A_1(t) + A_2(t) + A_3(t), \text{ say,} \qquad \text{(A.3)}
\end{aligned}$$

where $\hat{S}_{\lambda,t} = \frac{1}{N}\sum_{i=1}^{N} \hat{\lambda}_{it}\hat{\lambda}_{it}'$. By Lemmas A.6(i)-(iii) below $\sqrt{N} A_l(t) = o_P(1)$ for $l = 2, 3$. Then by Lemmas A.1(iii) and A.6(i), and Assumption A.2(i), $\sqrt{N}\left[\hat{F}_t - H^{(t)'} F_t\right] = (Q_t \Sigma_{\Lambda_t} Q_t')^{-1} Q_t \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \lambda_{it} e_{it} + o_P(1) \xrightarrow{d} N\left(0, (Q_t \Sigma_{\Lambda_t} Q_t')^{-1} Q_t \Gamma_{tt} Q_t' (Q_t \Sigma_{\Lambda_t} Q_t')^{-1}\right) = N\left(0, \left(\Sigma_{\Lambda_t}^{-1} Q_t^{-1}\right)' \Gamma_{tt} \Sigma_{\Lambda_t}^{-1} Q_t^{-1}\right)$. This completes the proof of Theorem 3.3. ■

Lemma A.6 *Suppose that Assumptions A.1 and A.3 hold. Then for $t = 1, 2, \ldots, T$,*
(i) $\hat{S}_{\lambda,t} = \frac{1}{N}\sum_{i=1}^{N} \hat{\lambda}_{it}\hat{\lambda}_{it}' = Q_t \Sigma_{\Lambda_t} Q_t' + o_P(1)$,
(ii) $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \left[\hat{\lambda}_{it} - H^{(t)-1}\lambda_{it}\right] e_{it} = o_P(1)$,
(iii) $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \hat{\lambda}_{it} \left[\hat{\lambda}_{it} - H^{(t)-1}\lambda_{it}\right]' H^{(t)\prime} F_t = o_P(1)$.

To prove Theorem 3.4, we need three lemmas. More precisely, Lemmas A.7 and A.8 are used in the proof of Lemma A.9, which in turn is used to prove Theorem 3.4.

Lemma A.7 *Suppose that Assumptions A.1 and A.3-A.4 hold. Then for any $R \geq 1$, there exist $R_0 \times R$ matrices $\left\{H^{(r,R)} \equiv (NT)^{-1} F^{(r)\prime} F^{(r)} \Lambda_r' \hat{\Lambda}_r^{(R)}\right\}$ with rank$(H^{(r,R)}) = \min\{R, R_0\}$ such that*
(i) $(NT)^{-1}\sum_{r=1}^{T} ||\breve{\Lambda}_r^{(R)} - \Lambda_r H^{(r,R)}||^2 = O_P\left(C_{NT}^{-2}\right)$,
(ii) $\max_{1\leq r\leq T} N^{-1}||\breve{\Lambda}_r^{(R)} - \Lambda_r H^{(r,R)}||^2 = O_P\left(C_{NT}^{-2}\ln T\right)$,
(iii) $\max_{1\leq r\leq T}\left\|N^{-1}\breve{\Lambda}_r^{(R)\prime}\breve{\Lambda}_r^{(R)} - N^{-1}H^{(r,R)\prime}\Lambda_r'\Lambda_r H^{(r,R)}\right\| = O_P(C_{NT}^{-1}(\ln T)^{1/2})$.

Lemma A.8 *Suppose that Assumptions A.1 and A.3-A.4 hold. Let $H^{(t,R)}$ be as defined in Lemma A.7 with Moore-Penrose generalized inverse $H^{(r,R)+} = \begin{pmatrix} H^{(r,R)+}(1) \\ H^{(r,R)+}(2) \end{pmatrix}$, where $H^{(r,R)+}(1)$ and $H^{(r,R)+}(2)$ are $R_0 \times R_0$ and $(R - R_0) \times R_0$ matrices, respectively. Let $V_{NT}^{(r,R)}$ denote an $R \times R$ diagonal matrix consisting of the R largest eigenvalues of the $N \times N$ matrix $(NT)^{-1} X^{(r)\prime} X^{(r)}$ where the eigenvalues are ordered in decreasing order along the main diagonal line. Write $\hat{\Lambda}_r^{(R)} = [\hat{\Lambda}_r^{(R)}(1), \hat{\Lambda}_r^{(R)}(2)]$ and $H^{(r,R)} = [H^{(r,R)}(1), H^{(r,R)}(2)]$, where $\hat{\Lambda}_r^{(R)}(1)$, $\hat{\Lambda}_r^{(R)}(2)$, $H^{(r,R)}(1)$, and $H^{(r,R)}(2)$ are $N \times R_0$, $N \times (R - R_0)$, $R_0 \times R_0$, and $R_0 \times (R - R_0)$ matrices, respectively. Write $V_{NT}^{(r,R)} =$ diag$\left(V_{NT}^{(r,R)}(1), V_{NT}^{(r,R)}(2)\right)$, where $V_{NT}^{(r,R)}(1)$ denotes the upper left $R_0 \times R_0$ submatrix of $V_{NT}^{(r,R)}(1)$. Then*
(i) $\max_{1\leq r\leq T} N^{-1}\left\|\hat{\Lambda}_r^{(R)}(1) - \Lambda_r H^{(r,R)}(1) V_{NT}^{(r,R)}(1)^{-1}\right\|^2 = O_P\left(C_{NT}^{-2}\ln T\right)$ *and* $\max_{1\leq r\leq T}\left\|H^{(r,R)}(2)\right\|^2 = O_P\left(T^{-1}h^{-1}\ln T + N^{-1}h^{-1}\right)$,

(ii) $\max_{1\leq r\leq T}\left\Vert H^{(r,R)+}(1)\right\Vert = O_P(1)$ *and* $\max_{1\leq r\leq T}\left\Vert H^{(r,R)+}(2)\right\Vert = O_P(T^{-1/2}h^{-1/2}(\ln T)^{1/2}+N^{-1/2}h^{-1/2})$,

(iii) $(NT)^{-1}\sum_{r=1}^{T}F_r'H^{(r,R)+\prime}(\breve{\Lambda}_r^{(R)}-\Lambda_r H^{(r,R)})'e_r = O_P\left(C_{NT}^{-2}\right)$,

(iv) $(NT)^{-1}\sum_{r=1}^{T}||(\breve{\Lambda}_r^{(R)}-\Lambda_r H^{(r,R)})H^{(r,R)+}F_r||^2 = O_P\left(C_{NT}^{-2}\right)$.

Lemma A.9 *Suppose that Assumptions A.1 and A.3-A.4 hold. Let* $H^{(t,R)}$ *be as defined in Lemma A.7. Then*

(i) $V\left(R,\{\breve{\Lambda}_r^{(R)}\}\right)-V\left(R,\{\Lambda_r H^{(r,R)}\}\right)=O_P\left(C_{NT}^{-1}(\ln T)^{1/2}\right)$ *for each* R *with* $1\leq R\leq R_0$,

(ii) there exists a $c_R>0$ *such that plim inf*$_{(N,T)\to\infty}\left[V\left(R,\{\Lambda_r H^{(r,R)}\}\right)-V\left(R,\{\Lambda_r\}\right)\right]\geq c_R$ *for each* R *with* $1\leq R<R_0$,

(iii) $V\left(R,\{\breve{\Lambda}_r^{(R)}\}\right)-V\left(R_0,\{\breve{\Lambda}_r^{(R_0)}\}\right)=O_P\left(C_{NT}^{-2}\right)$ *for each* R *with* $R\geq R_0$.

Proof of Theorem 3.4. The proof is analogous to that of Corollary 1 in Bai and Ng (2002). For notational simplicity, let $V(R)=V(R,\{\breve{\Lambda}_r^{(R)}\})$ for all R. Note that $IC(R)-IC(R_0)=\ln[V(R)/V(R_0)]+(R-R_0)\rho_{NT}$. We discuss two cases: (1) $R<R_0$, and (2) $R>R_0$.

In case (1), by Lemmas A.9(i) and (ii), $V(R)/V(R_0)>1+\epsilon_0$ and hence $\ln[V(R)/V(R_0)]\geq\epsilon_0/2$ for some $\epsilon_0>0$ w.p.a.1. This, in conjunction with the fact that $(R-R_0)\rho_{NT}\to 0$ under our assumption, implies that $IC(R)-IC(R_0)\geq\epsilon_0/4$ w.p.a.1. It follows that

$$P(IC(R)-IC(R_0)>0)\to 1 \text{ for any } R<R_0 \text{ as } (N,T)\to\infty.$$

In case (2), we apply Lemma A.9(iii) and Assumption A.5 to obtain

$$\begin{aligned}P(IC(R)-IC(R_0)>0) &= P(\ln[V(R)/V(R_0)]+(R-R_0)\rho_{NT}>0)\\ &= P\left(O_P(1)+(R-R_0)\rho_{NT}C_{NT}^2>0\right)\to 1 \text{ for any } R>R_0 \text{ as } (N,T)\to\infty.\end{aligned}$$

Consequently, the minimizer of $IC(R)$ can only be achieved at $R=R_0$ w.p.a.1. That is, $P(\hat{R}=R_0)\to 1$ for any $R\in[1,R_{\max}]$ as $(N,T)\to\infty$. ■

B Proofs of Theorems in Section 4

To proceed, we need to introduce some notations and lemmas. Let $e_{it}^{\dagger}=e_{it}+a_{NT}F_t'g_{it}$ and $e_t^{\dagger}=(e_{1t}^{\dagger},...,e_{Nt}^{\dagger\prime})'$. Then $X_{it}=F_t'\lambda_{it}+e_{it}=F_t'\lambda_{i0}+e_{it}^{\dagger}$ and $X_t=\Lambda_0F_t+e_t^{\dagger}$. As in (A.1) we can decompose $\tilde{F}_t-H'F_t=V_{NT}^{-1}\frac{1}{NT}\sum_{s=1}^{T}\tilde{F}_sX_s'X_t-H'F_t$ as follows:

$$\begin{aligned}\tilde{F}_t-H'F_t &= V_{NT}^{-1}\frac{1}{NT}\sum_{s=1}^{T}\tilde{F}_s\left[\Lambda_0F_s+e_s^{\dagger}\right]'\left[\Lambda_0F_t+e_t^{\dagger}\right]-H'F_t\\ &= V_{NT}^{-1}\left\{\frac{1}{T}\sum_{s=1}^{T}\tilde{F}_sE(e_s^{\dagger\prime}e_t^{\dagger}/N)+\frac{1}{T}\sum_{s=1}^{T}\tilde{F}_s\left[e_s^{\dagger\prime}e_t^{\dagger}/N-E(e_s^{\dagger\prime}e_t^{\dagger}/N)\right]\right.\\ &\quad\left.+\frac{1}{T}\sum_{s=1}^{T}\tilde{F}_sF_s'\Lambda_0'e_t^{\dagger}/N+\frac{1}{T}\sum_{s=1}^{T}\tilde{F}_sF_t'\Lambda_0'e_s^{\dagger}/N\right\}\\ &\equiv A_1^0(t)+A_2^0(t)+A_3^0(t)+A_4^0(t),\quad\text{say.}\end{aligned}\tag{B.1}$$

Using $X_{it} = F_t'\lambda_{i0} + e_{it}^{\dagger}$ and by Bai (2003, p.165), we have

$$\tilde{\lambda}_{i0} - H^{-1}\lambda_{i0} = H'\frac{1}{T}\sum_{s=1}^{T} F_s e_{is}^{\dagger} + \frac{1}{T}\sum_{s=1}^{T}\left[\tilde{F}_s - HF_s\right] e_{is}^{\dagger} - \frac{1}{T}\tilde{F}'\left[\tilde{F}H^{-1} - F\right]\lambda_{i0}$$
$$\equiv D_1^0(i) + D_2^0(i) - D_3^0(i), \text{ say.} \tag{B.2}$$

Let V_0, H_0, Q_0, and Σ_{Λ_0} be the probability limits of $V_{NT}^{(r)}$, $H^{(r)}$, $\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s^{(r)}F_s^{(r)\prime}$, and $N^{-1}\Lambda_r'\Lambda_r$ under $\mathbb{H}_0$, respectively. Note that they are also the limits under $\mathbb{H}_1(a_{NT})$, and $H_0 = Q_0^{-1}$. Let $S_{\lambda,0} = Q_0\Sigma_{\Lambda_0}Q_0'$. To prove Theorems 4.1 and 4.3, we need three lemmas.

Lemma B.1 *Suppose that Assumptions A.1, A.3(i) and (ii*), and A.6-A.7 hold. Then under* $\mathbb{H}_1(a_{NT})$ *with* $a_{NT} = T^{-1/2}N^{-1/4}h^{-1/4}$. *Then*

(i) $\max_r \left\|V_{NT}^{(r)} - V_0\right\| = O_P(C_{NT}^{-1}(\ln T)^{1/2})$,

(ii) $\max_r \left\|H^{(r)} - H_0\right\| = O_P(C_{NT}^{-1}(\ln T)^{1/2})$,

(iii) $\max_r \frac{1}{T}\left\|\left(\hat{F}^{(r)} - F^{(r)}H^{(r)}\right)\right\|^2 = O_P\left(T^{-1}h^{-1} + N^{-1}\ln T\right)$,

(iv) $\max_r \left\|\frac{1}{T}\left(\hat{F}^{(r)} - F^{(r)}H^{(r)}\right)' F^{(r)}H^{(r)}\right\| = O_P\left(T^{-1}h^{-1} + N^{-1}\ln T\right)$,

(v) $\max_{i,r} \left\|\frac{1}{T}\left(\hat{F}^{(r)} - F^{(r)}H^{(r)}\right)' e_i^{(r)}\right\| = O_P\left(T^{-1}h^{-1} + N^{-1}\ln T\right)$,

(vi) $\max_r \left\|\hat{S}_{\lambda,r} - S_{\lambda,0}\right\| = O_P(C_{NT}^{-1}(\ln T)^{1/2})$,

(vii) $\max_r \frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\lambda}_{ir} - H^{(r)-1}\lambda_{ir}\right\|^2 = O_P\left(C_{NT}^{-2}\ln(T)\right)$,

(viii) $\max_t \left\|\frac{1}{N}\sum_{i=1}^{N}[\hat{\lambda}_{ir} - H^{(r)-1}\lambda_{ir}]\lambda_{ir}'H^{(r)-1}\right\|^2 = O_P\left(C_{NT}^{-2}\ln(T)\right)$,

(ix) $\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{j=1}^{N}(\hat{\lambda}_{jt} - H^{(t)-1}\lambda_{jt})e_{jt}\right\|^2 = O_P\left(T^{-2}h^{-2} + N^{-2}\ln T\right)$,

(x) $\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_t - H^{(t)\prime}F_t\right\|^2 = O_P\left(N^{-1}\right)$.

Lemma B.2 *Suppose that Assumptions A.1, A.3(i) and (ii*) and A.6-A.7 hold. Then under* $\mathbb{H}_1(a_{NT})$ *with* $a_{NT} = T^{-1/2}N^{-1/4}h^{-1/4}$,

(i) $\frac{1}{T}\left\|\tilde{F} - FH\right\|^2 = O_P\left(C_{0NT}^{-2}\right)$,

(ii) $\frac{1}{T}(\tilde{F} - FH)'FH = O_P\left(C_{0NT}^{-2}\right) + o_P(a_{NT})$,

(iii) $\frac{1}{T}(\tilde{F} - FH)'\tilde{F} = O_P\left(C_{0NT}^{-2}\right) + o_P(a_{NT})$,

(iv) $\frac{1}{T}(\tilde{F}'\tilde{F} - H'F'FH) = O_P\left(C_{0NT}^{-2}\right) + o_P(a_{NT})$,

(v) $V_{NT} = V_0 + O_P\left(C_{0NT}^{-1}\right)$,

(vi) $H = Q_0^{-1} + O_P\left(C_{0NT}^{-1}\right)$,

(vii) $\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\lambda}_{i0} - H^{-1}\lambda_{i0}\right\|^2 = O_P\left(C_{0NT}^{-2}\right)$,

(viii) $\max_r \left\|\left(H^{(r)-1}\right)'\hat{S}_{\lambda,r}^{-1}H^{(r)-1} - \left(H^{-1}\right)'V_{NT}^{-1}(\frac{1}{T}\tilde{F}'F)\right\| = O_P(C_{NT}^{-1}(\ln T)^{1/2})$.

Lemma B.3 *Let* $R_{\Lambda}(i,t) = D_2(i,t) - D_3(i,t)$, $R_F(t) = A_2(t) + A_3(t)$, $R_{\Lambda}^0(i) = D_2^0(i) - D_3^0(i)$, *and* $R_F^0(t) = A_1^0(t) + A_2^0(t) + A_4^0(t)$. *Suppose that Assumptions A.1, A.3(i) and (ii*) and A.6-A.7 hold. Then under* $\mathbb{H}_1(a_{NT})$ *with* $a_{NT} = T^{-1/2}N^{-1/4}h^{-1/4}$,

(i) $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|R_{\Lambda}(i,t)\|^2 = O_P\left(T^{-2}h^{-2}+N^{-2}(\ln T)^2\right)$,
(ii) $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|R_F(t)\|^2 = O_P\left(C_{NT}^{-4}(\ln T)^2\right)$,
(iii) $\frac{1}{N}\sum_{i=1}^{N}[\left\|R_{\Lambda}^0(i)\right\|^2 = O_P\left(C_{0NT}^{-4}\right)+o_P\left(a_{NT}^2\right)$,
(iv) $\frac{1}{T}\sum_{t=1}^{T}\left\|R_F^0(t)\right\|^2 = O_P\left(C_{0NT}^{-4}\right)+o_P\left(a_{NT}^2\right)$.

In addition, we need the following lemma from Sun and Chiang (1997).

Lemma B.4 *Let $\{V_t, t\geq 1\}$ be a strong mixing process with mixing coefficient $\alpha(\cdot)$. Let $G_{t_1,\ldots,t_m}$, denote the distribution function of $(V_{t_1},\ldots,V_{t_m})$. For any integer $m>1$ and integers $(t_1,\ldots,t_m)$ such that $1\leq t_1<t_2<\ldots<t_m$, let ϑ be a Borel measurable function such that max$\{\int|\vartheta(v_1,\ldots,v_m)|^{1+\widetilde{\eta}}$ $dG_{t_1,\ldots,t_j}(v_1,\ldots,v_j)\,dG_{t_j+1,\ldots,t_m}(v_{j+1},\ldots,v_m), \int|\vartheta(v_1,\ldots,v_m)|^{1+\widetilde{\eta}}dG_{t_1,\ldots,t_m}\}\leq M$ for some $\widetilde{\eta}>0$. Then $|\int\vartheta(v_1,\ldots,v_m)\,dG_{t_1,\ldots,t_m}(v_1,\ldots,v_m)-\int\vartheta(v_1,\ldots,v_m)\,dG_{t_1,\ldots,t_j}(v_1,\ldots,v_j)\,dG_{t_j+1,\ldots,t_m}(v_{j+1},\ldots,v_m)|\leq$ $4M^{1/(1+\widetilde{\eta})}\alpha(t_{j+1}-t_j)^{\widetilde{\eta}/(1+\widetilde{\eta})}$.*

Proof of Theorem 4.1. For the convenience of proving Theorem 4.3 below, we prove that under $\mathbb{H}_1(a_{NT})$ with $a_{NT}=T^{-1/2}N^{-1/4}h^{-1/4}$,

$$J_{NT}=TN^{1/2}h^{1/2}\hat{M}-\mathbb{B}_{NT}-\Pi\xrightarrow{d}N(0,\mathbb{V}_0),$$

where $\Pi=\Pi_1+\Pi_2$ and $\mathbb{V}_0=\lim_{(N,T)\to\infty}\mathbb{V}_{NT}$. In the special case where $\mathbb{H}_0$ holds, we see that $\Pi=0$ and the result in Theorem 4.1 holds.

Using

$$\hat{\lambda}_{it}'\hat{F}_t=(\hat{\lambda}_{it}-H^{(t)-1}\lambda_{it})'H^{(t)'}F_t+\lambda_{it}'(H^{(t)-1})'(\hat{F}_t-H^{(t)\prime}F_t)+(\hat{\lambda}_{it}-H^{(t)-1}\lambda_{it})'(\hat{F}_t-H^{(t)\prime}F_t)+\lambda_{it}'F_t$$

and

$$\tilde{\lambda}_{i0}'\tilde{F}_t=(\tilde{\lambda}_{i0}-H^{-1}\lambda_{i0})'H'F_t+\lambda_{i0}'\left(H^{-1}\right)'\left(\tilde{F}_t-H'F_t\right)+(\tilde{\lambda}_{i0}-H^{-1}\lambda_{i0})'\left(\tilde{F}_t-H'F_t\right)+\lambda_{i0}'F_t,$$

we have $\hat{\lambda}_{it}'\hat{F}_t-\tilde{\lambda}_{i0}'\tilde{F}_t=d_{1it}+d_{2it}+d_{3it}$, where

$$\begin{aligned}
d_{1it} &= F_t'H^{(t)}(\hat{\lambda}_{it}-H^{(t)-1}\lambda_{it})-F_t'H(\tilde{\lambda}_{i0}-H^{-1}\lambda_{i0})+\lambda_{it}'(H^{(t)-1})'(\hat{F}_t-H^{(t)\prime}F_t)-\lambda_{i0}'(H^{-1})'(\tilde{F}_t-H'F_t),\\
d_{2it} &= (\lambda_{it}-\lambda_{i0})'F_t, \text{ and}\\
d_{3it} &= (\hat{\lambda}_{it}-H^{(t)-1}\lambda_{it})'(\hat{F}_t-H^{(t)\prime}F_t)-(\tilde{\lambda}_{i0}-H^{-1}\lambda_{i0})'(\tilde{F}_t-H'F_t).
\end{aligned}$$

As we shall see, d_{1it} contributes to the asymptotic bias, variance, and local power of our test statistic, d_{2it} only contributes to the asymptotic local power and is vanishing under $\mathbb{H}_0$, and d_{3it} collects the second order term in the expansion of $\hat{\lambda}_{it}'\hat{F}_t-\tilde{\lambda}_{i0}'\tilde{F}_t$ and is asymptotically negligible. Then

$$\begin{aligned}
TN^{1/2}h^{1/2}\hat{M} &= N^{-1/2}h^{1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}(d_{1it}+d_{2it}+d_{3it})^2\\
&= N^{-1/2}h^{1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(d_{1it}^2+d_{2it}^2+d_{3it}^2+2d_{1it}d_{2it}+2d_{1it}d_{3it}+2d_{2it}d_{3it}\right)\\
&\equiv M_1+M_2+M_3+2M_4+2M_5+2M_6, \quad \text{say.} \qquad \text{(B.3)}
\end{aligned}$$

We prove the theorem by showing that under $\mathbb{H}_1(a_{NT})$, (i) $M_1 - \mathbb{B}_{NT} - \Pi_1 \xrightarrow{d} N(0, \mathbb{V}_0)$, (ii) $M_2 = \Pi_2 + o_P(1)$, and (iii) $M_j = o_P(1)$ for $j = 3, 4, 5, 6$. To save space, we only prove (i) here and relegate the proofs of (ii) and (iii) to Lemma B.6 below.

To prove (i), let $R_\Lambda(i,t)$, $R_F(t)$, $R^0_\Lambda(i)$, and $R^0_F(t)$ be defined as in the statement of Lemma B.3. Then by (A.2), (A.3), (B.1), and (B.2), we have

$$\hat{\lambda}_{it} - H^{(t)-1}\lambda_{it} = \frac{1}{T}H^{(t)\prime}\sum_{s=1}^{T}k_{h,st}F_s e_{is} + R_\Lambda(i,t), \quad \hat{F}_t - H^{(t)\prime}F_t = \hat{S}^{-1}_{\lambda,t}H^{(t)-1}\frac{1}{N}\sum_{i=1}^{N}\lambda_{it}e_{it} + R_F(t),$$
$$\tilde{F}_t - H'F_t = V^{-1}_{NT}\left(\frac{1}{T}\tilde{F}'F\right)\Lambda_0' e^\dagger_t/N + R^0_F(t), \text{ and } \tilde{\lambda}_{i0} - H^{-1}\lambda_{i0} = \frac{1}{T}H'\sum_{s=1}^{T}F_s e^\dagger_{is} + R^0_\Lambda(i),$$

where apparently $R_\Lambda(i,t)$, $R_F(t)$, $R^0_\Lambda(i)$, and $R^0_F(t)$ represent the smaller order (remainder) terms in each of the above four asymptotic expansions. Using $e^\dagger_{is} = e_{is} + a_{NT}g^\dagger_{is}$, $\lambda_{it} = \lambda_{i0} + a_{NT}g_{it}$ and the above expressions, we further decompose d_{1it} as follows

$$d_{1it} = d_{1it,1} - d_{1it,2} + d_{1it,3} + d_{1it,4},$$

where

$$d_{1it,1} = F_t'\frac{1}{T}\sum_{s=1}^{T}\left[k_{h,st}H^{(t)}H^{(t)\prime} - HH'\right]F_s e_{is},$$
$$d_{1it,2} = a_{NT}\left[F_t'HH'\frac{1}{T}\sum_{s=1}^{T}F_s g^\dagger_{is} + \lambda_{i0}'\left(H^{-1}\right)'V^{-1}_{NT}\left(\frac{1}{T}\tilde{F}'F\right)\Lambda_0' g^\dagger_t/N\right],$$
$$d_{1it,3} = \lambda_{i0}'\left[\left(H^{(t)-1}\right)'\hat{S}^{-1}_{\lambda,t}H^{(t)-1} - \left(H^{-1}\right)'V^{-1}_{NT}\left(\frac{1}{T}\tilde{F}'F\right)\right]\Lambda_0' e_t/N$$
$$+a_{NT}\lambda_{i0}'\left(H^{(t)-1}\right)'\hat{S}^{-1}_{\lambda,t}H^{(t)-1}\frac{1}{N}\sum_{i=1}^{N}g_{it}e_{it} + a_{NT}g_{it}'\left(H^{(t)-1}\right)'\hat{S}^{-1}_{\lambda,t}H^{(t)-1}\frac{1}{N}\sum_{i=1}^{N}\lambda_{it}e_{it},$$
$$d_{1it,4} = F_t'H^{(t)}R_\Lambda(i,t) - F_t'HR^0_\Lambda(i) + \lambda_{it}'\left(H^{(t)-1}\right)'R_F(t) - \lambda_{i0}'\left(H^{-1}\right)'R^0_F(t).$$

It will be clear that $d_{1it,1}$ contributes to the asymptotic bias and variance of the test statistic, $d_{1it,2}$ contributes to the asymptotic local power, and $d_{1it,3}$ and $d_{1it,4}$ are asymptotically negligible. With these notations, we can decompose M_1 as follows:

$$M_1 = N^{-1/2}h^{1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(d_{1it,1} - d_{1it,2} + d_{1it,3} + d_{1it,4}\right)^2$$
$$= N^{-1/2}h^{1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}(d^2_{1it,1} + d^2_{1it,2} + d^2_{1it,3} + d^2_{1it,4} - 2d_{1it,1}d_{1it,2} + 2d_{1it,1}d_{1it,3} + 2d_{1it,2}d_{1it,4}$$
$$-2d_{1it,2}d_{1it,3} - 2d_{1it,2}d_{1it,4} + 2d_{1it,3}d_{1it,4})$$
$$\equiv M_{1,1} + M_{1,2} + M_{1,3} + M_{1,4} - 2M_{1,5} + 2M_{1,6} + 2M_{1,7} - 2M_{1,8} - 2M_{1,9} + 2M_{1,10}, \text{ say.}$$

We prove (i) by showing that (i1) $M_{1,1} - \mathbb{B}_{NT} \xrightarrow{d} N(0, \mathbb{V}_0)$, (i2) $M_{1,2} = \Pi_1 + o_P(1)$, (i3) $M_{1,l} = o_P(1)$ for $l = 3, 4, ..., 10$.

First, we prove (i1) $M_{1,1} - \mathbb{B}_{NT} \xrightarrow{d} N(0, \mathbb{V}_0)$**.** Let $L_{st} = k_{h,st} H^{(t)} H^{(t)\prime} - HH'$ and $\bar{L}_{st} = (k_{h,st} - 1) H_0 H_0'$. We further decompose $M_{1,1}$ as follows:

$$\begin{aligned}
M_{1,1} &= N^{-1/2} h^{1/2} \sum_{i=1}^{N} \sum_{t=1}^{T} d_{1it,1}^2 = N^{-1/2} h^{1/2} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \frac{1}{T} \sum_{s=1}^{T} F_t' L_{st} F_s e_{is} \right\}^2 \\
&= \frac{h^{1/2}}{N^{1/2} T^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(F_t' L_{st} F_s \right)^2 e_{is}^2 + \frac{h^{1/2}}{N^{1/2} T^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{1 \le s \ne r \le T} F_t' L_{st} F_s F_r' L_{rt} F_t e_{is} e_{ir} \\
&= \frac{h^{1/2}}{N^{1/2} T^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(F_t' L_{st} F_s \right)^2 e_{is}^2 + \frac{h^{1/2}}{N^{1/2} T^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{1 \le s \ne r \le T} F_t' \bar{L}_{st} F_s F_r' \bar{L}_{rt} F_t e_{is} e_{ir} \\
&\quad + \frac{2h^{1/2}}{N^{1/2} T^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{1 \le s \ne r \le T} F_t' \left(L_{st} - \bar{L}_{st} \right) F_s F_r' \bar{L}_{rt} F_t e_{is} e_{ir} \\
&\quad + \frac{h^{1/2}}{N^{1/2} T^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{1 \le s \ne r \le T} F_t' \left(L_{st} - \bar{L}_{st} \right) F_s F_r' \left(L_{rt} - \bar{L}_{rt} \right) F_t e_{is} e_{ir} \\
&\equiv M_{1,1}^{(1)} + M_{1,1}^{(2)} + M_{1,1}^{(3)} + M_{1,1}^{(4)}, \text{ say.}
\end{aligned}$$

Apparently, $M_{1,1}^{(1)} = \frac{h^{1/2}}{N^{1/2} T^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(F_t' L_{st} F_s \right)^2 e_{is}^2 = \mathbb{B}_{NT}$. Using $\bar{L}_{st} = (k_{h,st} - 1) H_0 H_0'$, we can further decompose $M_{1,1}^{(2)}$ as follows

$$\begin{aligned}
M_{1,1}^{(2)} &= \frac{h^{1/2}}{N^{1/2} T} \sum_{i=1}^{N} \sum_{1 \le r \ne s \le T} \operatorname{tr} \left(F_s F_r' \frac{1}{T} \sum_{t=1}^{T} \bar{L}_{rt} F_t F_t' \bar{L}_{st} \right) e_{is} e_{ir} \\
&= \frac{h^{1/2}}{N^{1/2} T} \sum_{i=1}^{N} \sum_{1 \le r \ne s \le T} \operatorname{tr} \left(F_s F_r' H_0 H_0' \frac{1}{T} \sum_{t=1}^{T} k_{h,st} k_{h,rt} F_t F_t' H_0 H_0' \right) e_{is} e_{ir} \\
&\quad - \frac{2h^{1/2}}{N^{1/2} T} \sum_{i=1}^{N} \sum_{1 \le r \ne s \le T} \operatorname{tr} \left(F_s F_r' H_0 H_0' \frac{1}{T} \sum_{t=1}^{T} k_{h,st} F_t F_t' H_0 H_0' \right) e_{is} e_{ir} \\
&\quad + \frac{h^{1/2}}{N^{1/2} T} \sum_{i=1}^{N} \sum_{1 \le r \ne s \le T} \operatorname{tr} \left(F_s F_r' H_0 H_0' \frac{1}{T} \sum_{t=1}^{T} F_t F_t' H_0 H_0' \right) e_{is} e_{ir} \\
&\equiv M_{1,1}^{(2,1)} - 2 M_{1,1}^{(2,2)} + M_{1,1}^{(2,3)}.
\end{aligned}$$

We shall focus on the analysis of $M_{1,1}^{(2,1)}$ as by analogous arguments we can readily show that $M_{1,1}^{(2,l)} = O_P\left(h^{1/2}\right)$ for $l = 2, 3$. For $M_{1,1}^{(2,1)}$, we make further decomposition:

$$\begin{aligned}
M_{1,1}^{(2,1)} &= \frac{2h^{-1/2}}{N^{1/2} T} \sum_{i=1}^{N} \sum_{1 \le r < s \le T} \operatorname{tr} \left(F_s F_r' H_0 H_0' \Sigma_F H_0 H_0' \right) \bar{k}_{sr} e_{is} e_{ir} \\
&\quad + \frac{2h^{-1/2}}{N^{1/2} T} \sum_{i=1}^{N} \sum_{1 \le r < s \le T} \operatorname{tr} \left(F_s F_r' H_0 H_0' \left(\frac{h}{T} \sum_{t=1}^{T} k_{h,st} k_{h,rt} F_t F_t' - \bar{k}_{sr} \Sigma_F \right) H_0 H_0' \right) e_{is} e_{ir} \\
&\equiv M_{1,1}^{(2,1,a)} + M_{1,1}^{(2,1,b)}, \text{ say,}
\end{aligned}$$

where $\bar{k}_{sr} = \bar{K}\left(\frac{s-r}{Th}\right)$ and $\bar{K}(u) = \int_{-1}^{1} K(v) K(u-v) dv$. Let $Z_{NT,s} = 2T^{-1}N^{-1/2}h^{-1/2}\sum_{r=1}^{s-1}\bar{k}_{sr}F_s'H_0$ $\times\bar{\Sigma}_F H_0' F_r e_s' e_r$ with $\bar{\Sigma}_F = H_0'\Sigma_F H_0$. Then $M_{1,1}^{(2,1,a)} = \sum_{s=2}^{T} Z_{NT,s}$ and $E(Z_{NT,s}|\mathcal{F}_{NT,s-1}) = 0$. By the martingale central limit theorem (e.g., Pollard, 1984, p.171), it suffices to prove $\mathbb{V}_{NT}^{-1/2}M_{1,1}^{(2,1,a)} \xrightarrow{d} N(0,1)$ by showing that

$$\mathcal{Z} \equiv \sum_{s=2}^{T} E\left(Z_{NT,s}^4|\mathcal{F}_{NT,s-1}\right) = o_P(1) \text{ and } \sum_{s=2}^{T} Z_{NT,s}^2 - \mathbb{V}_{NT} = o_P(1). \tag{B.4}$$

First, we verify the first part of (B.4). Observing that $\mathcal{Z} \geq 0$, it suffices to show $\mathcal{Z} = o_P(1)$ by showing that $E(\mathcal{Z}) = o(1)$ by Markov inequality. Letting $\phi_{sr} = F_s'H_0\bar{\Sigma}_F H_0' F_r e_s' e_r$, we have

$$\begin{aligned}
E(\mathcal{Z}) &= \sum_{s=2}^{T} E\left\{\left[\frac{2}{TN^{1/2}h^{1/2}}\sum_{r=1}^{s-1}\bar{k}_{sr}\phi_{sr}\right]^4\right\} \\
&= \frac{16}{T^4N^2h^2}\sum_{s=2}^{T} E\left[\sum_{r=1}^{s-1}\bar{k}_{sr}^4\phi_{sr}^4 + 2\sum_{1\leq r_1<r_2\leq s-1}\bar{k}_{sr_1}^2\bar{k}_{sr_2}^2\phi_{sr_1}^2\phi_{sr_2}^2 + 4\sum_{t=1}^{s-1}\sum_{1\leq r_1<r_2\leq s-1}\bar{k}_{st}^2\bar{k}_{sr_1}\bar{k}_{sr_2}\phi_{st}^2\phi_{sr_1}\phi_{sr_2}\right. \\
&\quad \left. +4\sum_{1\leq r_1<r_2\leq s-1, 1\leq t_1<t_2\leq s-1}\bar{k}_{r_1s}\bar{k}_{r_2s}\bar{k}_{t_1s}\bar{k}_{t_2s}\phi_{sr_1}\phi_{sr_2}\phi_{st_1}\phi_{st_2}\right] \\
&\equiv \mathcal{Z}_1 + \mathcal{Z}_2 + \mathcal{Z}_3 + \mathcal{Z}_4, \text{ say.}
\end{aligned}$$

Noting that $\max_{r<s}\left\|N^{-1/2}\phi_{sr}\right\|_4^4 \leq C < \infty$ under Assumption A.6(iv), we can readily bound $\mathcal{Z}_1$, $\mathcal{Z}_2$, and $\mathcal{Z}_3$ as follows:

$$\begin{aligned}
\mathcal{Z}_1 &\leq T^{-2}h^{-1}\max_{r<s}\left\|N^{-1/2}\phi_{sr}\right\|_4^4\left(\frac{16}{T^2h}\sum_{s=2}^{T}\sum_{r=1}^{s-1}\bar{k}_{sr}^4\right) = O\left(T^{-2}h^{-1}\right), \\
\mathcal{Z}_2 &\leq T^{-1}\max_{r<s}\left\|N^{-1/2}\phi_{sr}\right\|_4^4\left(\frac{32}{T^3h^2}\sum_{s=2}^{T}\sum_{1\leq r_1<r_2\leq s-1}\bar{k}_{sr_1}^2\bar{k}_{sr_2}^2\right) = O\left(T^{-1}\right), \\
\mathcal{Z}_3 &\leq h\max_{r<s}\left\|N^{-1/2}\phi_{sr}\right\|_4^4\left(\frac{32}{T^4h^3}\sum_{s=2}^{T}\sum_{t=1}^{s-1}\sum_{1\leq r_1<r_2\leq s-1}\bar{k}_{st}^2\bar{k}_{sr_1}\bar{k}_{sr_2}\right) = O(h).
\end{aligned}$$

To study $\mathcal{Z}_4$, let $A = H_0\bar{\Sigma}_F H_0' = \{a_{mn}\}$. Then $\phi_{sr} = F_s'H_0\bar{\Sigma}_F H_0' F_r e_s' e_r = \sum_{m=1}^{R}\sum_{n=1}^{R} a_{mn}F_{s,m}F_{r,n}e_s'e_r$ and $\mathcal{Z}_4 = \sum_{1\leq m_1,m_2,m_3,m_4\leq R}\sum_{1\leq n_1,n_2,n_3,n_4\leq R}\left(\Pi_{l=1}^4 a_{m_l n_l}\right)\mathcal{Z}_{4NT}(m_{1:4}, n_{1:4})$, where

$$\begin{aligned}
\mathcal{Z}_{4NT}(m_{1:4}, n_{1:4}) &= \frac{64}{T^4N^2h^2}\sum_{t_0=3}^{T}\sum_{\substack{1\leq t_1<t_2\leq t_0-1 \\ 1\leq t_3<t_4\leq t_0-1}}\sum_{1\leq i_1,i_2,i_3,i_4\leq N}\bar{k}_{t_1t_0}\bar{k}_{t_2t_0}\bar{k}_{t_3t_0}\bar{k}_{t_4t_0}E\left[\varkappa_{mn}(i_{1:4}, t_{0:4})\right] \\
\varkappa_{mn}(i_{1:4}, t_{0:4}) &= \Pi_{l=1}^4\left(F_{t_0,m_l}F_{t_l,n_l}e_{i_lt_0}e_{i_lt_l}\right).
\end{aligned}$$

Since R is fixed and a_{mn}'s are finite, $\mathcal{Z}_4 = o(1)$ provided that $\mathcal{Z}_{4NT}(m_{1:4}, n_{1:4}) = o(1)$ $\forall m_{1:4} = (m_1, ..., m_4)$ and $n_{1:4} = (n_1, ..., n_4)$. Let $\#\mathcal{A}$ denote the cardinality of a set $\mathcal{A}$ and $\mathcal{S}_1 = \{t_0, t_1, t_2, t_3, t_4\}$ .We consider three cases: (1) $\#\mathcal{S}_1 = 5$, (2) $\#\mathcal{S}_1 = 4$, and (3) $\#\mathcal{S}_1 = 3$. We use $\mathcal{Z}_{4NT}^{(l)}$ to denote $\mathcal{Z}_{4NT}(m_{1:4}, n_{1:4})$ when the time indices in the summation are restricted to satisfy the condition in case (l) for $l = 1, 2, 3$. Note that $E\left|\varkappa_{mn}(i_{1:4}, t_{0:4})\right|^{1+\delta/2} \leq C < \infty$ by Assumption A.6(iii). Apparently, in case (3) we must have $t_1 = t_3$ and $t_2 = t_4$ and it is easy to obtain

$$\begin{aligned}\mathcal{Z}_{4NT}^{(3)} &= \frac{64}{T^4N^2h^2}\sum_{t_0=3}^{T}\sum_{1\leq t_1<t_2\leq t_0-1}\bar{k}_{t_1t_0}^2\bar{k}_{t_2t_0}^2 \\ &\quad\times E\left[F_{t_0,m_1}F_{t_1,n_1}F_{t_0,m_2}F_{t_2,n_2}F_{t_0,m_3}F_{t_1,n_3}F_{t_0,m_4}F_{t_2,n_4}(e_{t_0}'e_{t_1})^2(e_{t_0}'e_{t_2})^2\right] \\ &\leq \max_{m,n}\max_{t_1<t_0}\left\|N^{-1/2}F_{t_0,m}e_{t_0}'e_{t_1}F_{t_1,n}\right\|_4^4\frac{64}{T^4h^2}\sum_{t_0=3}^{T}\sum_{1\leq t_1<t_2\leq t_0-1}\bar{k}_{t_1t_0}^2\bar{k}_{t_2t_0}^2 = O\left(T^{-1}\right) = o(1).\end{aligned}$$

Let T_0 be as given in Assumption A.6(ii). In case (1), we consider three subcases: (1a) there exists at least one time index $l \in \mathcal{S}_1$ such that $|l - s| > T_0$ for all $s \in \mathcal{S}_1$ with $s \neq l$, and (1b) all the remaining cases. We use $\mathcal{Z}_{4NT}^{(1v)}$ to denote $\mathcal{Z}_{4NT}^{(1)}$ when the time indices in the summation are restricted to satisfy the condition in subcase $(1v)$ for $v = a, b$. In subcase (1a), we can readily apply Lemma B.4 and Assumptions A.6(i)-(iii) to obtain

$$\mathcal{Z}_{4NT}^{(1a)} \leq \frac{CN^2}{T^4h^2}\sum_{t_0,t_1,t_2,t_3,t_4 \text{ are all distinct}}\bar{k}_{t_1t_0}\bar{k}_{t_2t_0}\bar{k}_{t_3t_0}\bar{k}_{t_4t_0}\alpha(T_0)^{\delta/(2+\delta)} = O\left(N^2Th^2\alpha(T_0)^{\delta/(2+\delta)}\right) = o(1).$$

In case (1b), we have

$$\begin{aligned}\mathcal{Z}_{4NT}^{(1b)} &\leq \max_{m,n}\max_{t_0,t_1}\left\|N^{-1/2}F_{t_0,m}e_{t_0}'e_{t_1}F_{t_1,n}\right\|_4^4\frac{64}{T^4h^2}\sum_{\substack{1\leq t_1<t_2\leq t_0-1,1\leq t_3<t_4\leq t_0-1\\ t_0,t_1,t_2,t_3,t_4 \text{ satisfies condition in case (1b)}}}\bar{k}_{t_1t_0}\bar{k}_{t_2t_0}\bar{k}_{t_3t_0}\bar{k}_{t_4t_0} \\ &= O\left(T^{-3}T_0^4 + T^{-3}T_0^3h^{-1} + T^{-3}T_0^2h^{-2}\right) = o(1)\end{aligned}$$

as the total number of terms in the last summation is of order $O\left(TT_0^4\right)$. So $\mathcal{Z}_{4NT}^{(1)} = o(1)$. Similarly, we can show that $\mathcal{Z}_{4NT}^{(2)} = o(1)$. Thus $\mathcal{Z}_4 = o(1)$ and $E(\mathcal{Z}) = o(1)$, implying $\mathcal{Z} = o_P(1)$.

To verify the second part of (B.4), it suffices to show (I) $\sum_{s=2}^{T}E\left(Z_{NT,s}^2\right) = \mathbb{V}_{NT} + o(1)$, and (II) $\text{Var}\left(\sum_{s=2}^{T}Z_{NT,s}^2\right) = o_P(1)$ by Chebyshev inequality. These two claims can be easily proved if we also assume independence of $\{e_{i\cdot} = (e_{i1}, ..., e_{iT})'\}$ across i conditional on the factors. Here we prove them without imposing such cross-sectional independence conditions. We first prove (I). Observe that

$$\begin{aligned}\sum_{s=2}^{T}E\left(Z_{NT,s}^2\right) &= 4T^{-2}N^{-1}h^{-1}\sum_{s=2}^{T}\sum_{r=1}^{s-1}\bar{k}_{sr}^2E(\phi_{sr})^2 + 4T^{-2}N^{-1}h^{-1}\sum_{s=2}^{T}\sum_{1\leq r_1\neq r_2\leq s-1}\bar{k}_{sr_1}\bar{k}_{sr_2}E\left(\phi_{sr_1}\phi_{sr_2}\right) \\ &= \mathbb{V}_{NT} + o(1),\end{aligned}$$

provided $b_{NT} \equiv T^{-2}N^{-1}h^{-1}\sum_{s=2}^{T}\sum_{1\leq r_1\neq r_2\leq s-1}\bar{k}_{sr_1}\bar{k}_{sr_2}E\left(\phi_{sr_1}\phi_{sr_2}\right) = o(1)$. For notational simplicity, we assume for the moment that $R = 1$ so that each term in the product $F_s'AF_t$ becomes a scalar. [Otherwise, we need to utilize $F_s'AF_r = \sum_{m=1}^{R}\sum_{n=1}^{R}a_{mn}F_{s,m}F_{s,n}$ as in the analysis of $\mathcal{Z}_4$ above.] In this case, we have

$$b_{NT} = A^2 T^{-2} N^{-1} h^{-1} \sum_{r_3=3}^{T} \sum_{1\le r_1<r_2\le r_3-1} \sum_{i=1}^{N}\sum_{j=1}^{N} \bar{k}_{r_3 r_1}\bar{k}_{r_3 r_2} E(F_{r_3}F_{r_3}e_{ir_3}e_{jr_3}\ F_{r_1}F_{r_2}e_{ir_1}e_{jr_2}).$$

Let $\mathcal{S}_2 = \{r_1, r_2, r_3\}$. We consider three cases: (1) $|r_3 - r_2| > T_0$, (2) $|r_3 - r_2| \le T_0$ and $|r_2 - r_1| > T_0$, and (3) $|r_3 - r_2| \le T_0$ and $|r_1 - r_1| \le T_0$. We use $b_{NT}^{(l)}$ to denote b_{NT} when the time indices are restricted to case (l) for $l = 1, 2, 3$. In case (1), we apply Lemma B.4 and the fact that $E(F_{r_1}F_{r_2}e_{ir_1}e_{jr_2}) = 0$ for $r_1 < r_2$ under Assumption A.6(i) to obtain

$$\left|b_{NT}^{(1)}\right| \le CT^{-2}N^{-1}h^{-1} \sum_{r_1<r_2<r_3} \sum_{i=1}^{N}\sum_{j=1}^{N} \bar{k}_{r_3 r_1}\bar{k}_{r_3 r_2}\alpha(T_0)^{\delta/(1+\delta)} = O\left(NTh\alpha(T_0)^{\delta/(1+\delta)}\right) = o(1).$$

In case (2), we apply Lemma B.4 and the fact that $E(F_{r_1}e_{ir_1}) = 0$ to obtain

$$\left|b_{NT}^{(2)}\right| \le CT^{-2}N^{-1}h^{-1} \sum_{r_1<r_2<r_3} \sum_{i=1}^{N}\sum_{j=1}^{N} \bar{k}_{r_3 r_1}\bar{k}_{r_3 r_2}\alpha(T_0)^{\delta/(1+\delta)} = O\left(NTh\alpha(T_0)^{\delta/(1+\delta)}\right) = o(1).$$

In case (3), we have

$$\begin{aligned}
\left|b_{NT}^{(3)}\right| &= T^{-2}N^{-1}h^{-1} \sum_{r_1<r_2<r_3,\ \text{case (3)}} \bar{k}_{r_3 r_1}\bar{k}_{r_3 r_2} \left|E\left(F_{r_3}F_{r_3}e_{r_1}' e_{r_3} e_{r_2}' e_{r_3} F_{r_1}F_{r_2}\right)\right| \\
&\le \max_{m,n}\max_{r<s}\left\|N^{-1/2}F_rF_s e_r' e_s\right\|_2^2 T^{-2}h^{-1} \sum_{r_1<r_2<r_3,\ \text{case (3)}} \bar{k}_{r_3 r_1}\bar{k}_{r_3 r_2} = O\left(T^{-1}T_0^2 h\right) = o(1),
\end{aligned}$$

where we use the fact that the total number of terms in the summation over the three time indices for $b_{NT}^{(3)}$ are of order $O\left(TT_0^2\right)$. In sum, we have shown that $b_{NT} = o(1)$ and $\sum_{s=2}^{T} E\left(Z_{NT,s}^2\right) = \mathbb{V}_{NT} + o(1)$.

Now, we want to prove (II) by showing that $E\left(\sum_{s=2}^{T} Z_{NT,s}^2\right)^2 = \mathbb{V}_{NT}^2 + o(1)$. Noting that

$$\begin{aligned}
E\left(\sum_{s=2}^{T} Z_{NT,s}^2\right)^2 &= E\left(\sum_{s=2}^{T}\left[\frac{2}{TN^{1/2}h^{1/2}}\sum_{r=1}^{s-1}\bar{k}_{sr}\phi_{sr}\right]^2\right)^2 = \frac{16}{T^4N^2h^2}E\left(\sum_{s=2}^{T}\sum_{r_1=1}^{s-1}\sum_{r_2=1}^{s-1}\bar{k}_{sr}\bar{k}_{sr_1}\phi_{sr}\phi_{sr_1}\right)^2 \\
&= \frac{16}{T^4N^2h^2}E\left(\sum_{s=2}^{T}\sum_{r=1}^{s-1}\bar{k}_{sr}^2\phi_{sr}^2\right)^2 + \frac{16}{T^4N^2h^2}E\left(\sum_{s=2}^{T}\sum_{1\le r_1\ne r_2\le s-1}\bar{k}_{sr_1}\bar{k}_{sr_2}\phi_{sr_1}\phi_{sr_2}\right)^2 \\
&\quad + \frac{32}{T^4N^2h^2}E\left[\left(\sum_{s=2}^{T}\sum_{r=1}^{s-1}\bar{k}_{sr}^2\phi_{sr}^2\right)\sum_{s=2}^{T}\sum_{1\le r_1\ne r_2\le s-1}\bar{k}_{sr_1}\bar{k}_{sr_2}\phi_{sr_1}\phi_{sr_2}\right] \\
&\equiv b_{1NT} + b_{2NT} + b_{3NT},\ \text{say},
\end{aligned}$$

it suffices to show that (a) $b_{1NT} = \mathbb{V}_{NT}^2 + o_P(1)$ and (b) $b_{2NT} = o_P(1)$ because then $b_{3NT} \le 2\{b_{1NT}b_{2NT}\}^{1/2}$ $= o_P(1)$ by CS inequality. Note that

$$\begin{aligned}
b_{1NT} &= \frac{16}{T^4N^2h^2}\sum_{1\le r_1<s_1\le T, 1\le r_1<s_1\le T}\bar{k}_{s_1r_1}^2\bar{k}_{s_2r_2}^2 E\left(\phi_{s_1r_1}^2\phi_{s_2r_2}^2\right), \\
\mathbb{V}_{NT}^2 &= \frac{16}{T^4N^2h^2}\sum_{1\le r_1<s_1\le T, 1\le r_1<s_1\le T}\bar{k}_{s_1r_1}^2\bar{k}_{s_2r_2}^2 E\left(\phi_{s_1r_1}^2\right)E\left(\phi_{s_2r_2}^2\right).
\end{aligned}$$

Let $\mathcal{S}_3 = \{r_1, s_1, r_2, s_2\}$. We consider two cases: (1) for each $t \in \mathcal{S}_3$, $|t-l| > T_0$ for all $l \in \mathcal{S}_3$ with $l \neq t$, and (2) all the other remaining cases. Let $\mathcal{S}_{3,1}$ and $\mathcal{S}_{3,2}$ denote the subsets of $\mathcal{S}_3$ corresponding to these two cases, respectively. For $l = 1, 2$, let $a_{1NT}(l)$ and $\mathbb{V}_{NT}^2(l)$ to denote a_{1NT} and $\mathbb{V}_{NT}^2$ when the time indices are restricted to lie in $\mathcal{S}_{3,l}$, respectively. Note that $b_{1NT} = b_{1NT}(1) + b_{1NT}(2)$ and $\mathbb{V}_{NT}^2 = \mathbb{V}_{NT}^2(1) + \mathbb{V}_{NT}^2(2)$. In case (2), we have

$$b_{1NT}(2) \leq \max_{s<r} \left\| N^{-1}\phi_{sr}^2 \right\|_2^2 \frac{16}{T^4h^2} \sum_{\substack{1\leq r_1<s_1\leq T, 1\leq r_1<s_1\leq T,\\ \text{case (2)}}} \bar{k}_{s_1r_1}^2 \bar{k}_{s_2r_2}^2 = O\left(T_0T^{-1} + T^{-2}h^{-2}\right) = o(1),$$

$$\mathbb{V}_{NT}^2(2) \leq \max_{s<r} \left[E\left(N^{-1}\phi_{sr}^2\right)\right]^2 \frac{16}{T^4h^2} \sum_{\substack{1\leq r_1<s_1\leq T, 1\leq r_1<s_1\leq T\\ \text{case (2)}}} \bar{k}_{s_1r_1}^2 \bar{k}_{s_2r_2}^2 = O\left(T_0T^{-1} + T^{-2}h^{-2}\right) = o(1),$$

where we use the fact that there are at most T^3T_0 terms in the above displayed summations. In case (1), we consider six subcases: (1a) $r_1 < s_1 < r_2 < s_2$, (1b) $r_2 < s_2 < r_1 < s_1$, (1c) $r_1 < r_2 < s_1 < s_2$, (1d) $r_2 < r_1 < s_1 < s_2$, (1e) $r_1 < r_2 < s_2 < s_1$, and (1f) $r_2 < r_1 < s_2 < s_1$. We use $a_{1NT}(1, v)$ and $\mathbb{V}_{NT}^2(1, v)$ to denote $a_{1NT}(1)$ and $\mathbb{V}_{NT}^2(1)$, respectively, when the summation over the time indices are restricted to satisfy the conditions in subcase $(1v)$ for $v = a, b, c, d, e, f$. First, we study subcase (1a). By Lemma B.4,

$$\begin{aligned}
b_{1NT}(1,a) &= \frac{16}{T^4N^2h^2} \sum_{r_1<s_1<r_2<s_2, \mathcal{S}_{3,1}} \bar{k}_{s_1r_1}^2 \bar{k}_{s_2r_2}^2 E\left(\phi_{s_1r_1}^2 \phi_{s_2r_2}^2\right) \\
&= \frac{16}{T^4N^2h^2} \sum_{r_1<s_1<r_2<s_2, \mathcal{S}_{3,1}} \sum_{i_1,j_1,i_2,j_2} \bar{k}_{s_1r_1}^2 \bar{k}_{s_2r_2}^2 E(f_{s_1r_1}^2 e_{i_1s_1}e_{i_1r_1}e_{j_1s_1}e_{j_1r_1} f_{s_2r_2}^2 e_{i_2s_2}e_{i_2r_2}e_{j_2s_2}e_{j_2r_2}) \\
&\leq \frac{16}{T^4N^2h^2} \sum_{r_1<s_1<r_2<s_2, \mathcal{S}_{3,1}} \sum_{i_1,j_1,i_2,j_2} \bar{k}_{s_1r_1}^2 \bar{k}_{s_2r_2}^2 \{E\left(f_{s_1r_1}^2 e_{i_1s_1}e_{i_1r_1}e_{j_1s_1}e_{j_1r_1}\right) \\
&\quad \times E\left(f_{s_2r_2}^2 e_{i_2s_2}e_{i_2r_2}e_{j_2s_2}e_{j_2r_2}\right) + C\alpha(T_0)^{\delta/(2+\delta)}\} \\
&= \frac{16}{T^4N^2h^2} \sum_{r_1<s_1<r_2<s_2, \mathcal{S}_{3,1}} \sum_{i_1,j_1,i_2,j_2} \bar{k}_{s_1r_1}^2 \bar{k}_{s_2r_2}^2 E\left(\phi_{s_1r_1}^2\right) E\left(\phi_{s_2r_2}^2\right) + O(N^2\alpha(T_0)^{\delta/(2+\delta)}) \\
&= \mathbb{V}_{NT}^2(1,a) + o(1),
\end{aligned}$$

where $(f_{sr} = F_s'AF_r$, $\sum_{i_1,j_1,i_2,j_2}$ denotes $\sum_{i_1=1}^N \sum_{j_1=1}^N \sum_{i_2=1}^N \sum_{j_2=1}^N$, $\sum_{r_1<s_1<r_2<s_2,\mathcal{S}_{3,1}}$ indicates the summation is done over the four time indices satisfying the condition in case (1). By the same token, $b_{1NT}(1,b) = \mathbb{V}_{NT}^2(1,b) + o(1)$. Now, consider subcase (1c). As above, we also assume here that $R = 1$

so that each term in $F_s' A F_t$ is a scalar. By applying Lemma B.4 three times, we have

$$
\begin{aligned}
b_{1NT}(1,c) &= \frac{16}{T^4N^2h^2} \sum_{r_1<r_2<s_1<s_2,\mathcal{S}_{3,1}} \bar{k}_{s_1r_1}^2 \bar{k}_{s_2r_2}^2 E\left(\phi_{s_1r_1}^2 \phi_{s_2r_2}^2\right) \\
&= \frac{16}{T^4N^2h^2} \sum_{r_1<r_2<s_1<s_2,\mathcal{S}_{3,1}} \sum_{i_1,j_1,i_2,j_2} \bar{k}_{s_1r_1}^2 \bar{k}_{s_2r_2}^2 \\
&\quad \times E\left(F_{s_1}^2 F_{s_2}^2 F_{r_1}^2 F_{r_2}^2 e_{i_1s_1} e_{i_1r_1} e_{j_1s_1} e_{j_1r_1} e_{i_2s_2} e_{i_2r_2} e_{j_2s_2} e_{j_2r_2}\right) \\
&\leq \frac{16A^4}{T^4N^2h^2} \sum_{r_1<r_2<s_1<s_2,\mathcal{S}_{3,1}} \sum_{i_1,j_1,i_2,j_2} \bar{k}_{s_1r_1}^2 \bar{k}_{s_2r_2}^2 \{E\left(F_{r_1}^2 F_{r_2}^2 e_{i_1r_1} e_{j_1r_1} e_{i_2r_2} e_{j_2r_2}\right) \\
&\quad \times E\left(F_{s_1}^2 F_{s_2}^2 e_{i_1s_1} e_{j_1s_1} e_{i_2s_2} e_{j_2s_2}\right) + C\alpha\left(T_0\right)^{\delta/(2+\delta)}\} \\
&\leq \frac{16A^4}{T^4N^2h^2} \sum_{r_1<r_2<s_1<s_2,\mathcal{S}_{3,1}} \sum_{i_1,j_1,i_2,j_2} \bar{k}_{s_1r_1}^2 \bar{k}_{s_2r_2}^2 \{E\left(F_{r_1}^2 e_{i_1r_1} e_{j_1r_1}\right) E\left(F_{r_2}^2 e_{i_2r_2} e_{j_2r_2}\right) \\
&\quad \times E\left(F_{s_1}^2 e_{i_1s_1} e_{j_1s_1}\right) E\left(F_{s_2}^2 e_{i_2s_2} e_{j_2s_2}\right) + 2C\alpha\left(T_0\right)^{\delta/(2+\delta)}\} \\
&= \frac{16A^4}{T^4N^2h^2} \sum_{r_1<r_2<s_1<s_2,\mathcal{S}_{3,1}} \sum_{i_1,j_1,i_2,j_2} \bar{k}_{s_1r_1}^2 \bar{k}_{s_2r_2}^2 E\left(F_{r_1}^2 e_{i_1r_1} e_{j_1r_1}\right) E\left(F_{r_2}^2 e_{i_2r_2} e_{j_2r_2}\right) \\
&\quad \times E\left(F_{s_1}^2 e_{i_1s_1} e_{j_1s_1}\right) E\left(F_{s_2}^2 e_{i_2s_2} e_{j_2s_2}\right) + o(1).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathbb{V}_{NT}^2(1,c) &= \frac{16}{T^4N^2h^2} \sum_{r_1<r_2<s_1<s_2,\mathcal{S}_{3,1}} \bar{k}_{s_1r_1}^2 \bar{k}_{s_2r_2}^2 E\left(\phi_{s_1r_1}^2\right) E\left(\phi_{s_2r_2}^2\right) \\
&= \frac{16A^4}{T^4N^2h^2} \sum_{r_1<r_2<s_1<s_2,\mathcal{S}_{3,1}} \sum_{i_1,j_1,i_2,j_2} \bar{k}_{s_1r_1}^2 \bar{k}_{s_2r_2}^2 E\left(F_{r_1}^2 F_{r_2}^2 e_{i_1r_1} e_{j_1r_1} e_{i_2r_2} e_{j_2r_2}\right) \\
&\quad \times E\left(F_{s_1}^2 F_{s_2}^2 e_{i_1s_1} e_{j_1s_1} e_{i_2s_2} e_{j_2s_2}\right) \\
&\leq \frac{16A^4}{T^4N^2h^2} \sum_{r_1<r_2<s_1<s_2,\mathcal{S}_{3,1}} \sum_{i_1,j_1,i_2,j_2} \bar{k}_{s_1r_1}^2 \bar{k}_{s_2r_2}^2 \{E\left(F_{r_1}^2 e_{i_1r_1} e_{j_1r_1}\right) E\left(F_{r_2}^2 e_{i_2r_2} e_{j_2r_2}\right) \\
&\quad \times E\left(F_{s_1}^2 e_{i_1s_1} e_{j_1s_1}\right) E\left(F_{s_2}^2 e_{i_2s_2} e_{j_2s_2}\right) + C\alpha\left(T_0\right)^{\delta/(2+\delta)}\} \\
&= \frac{16A^4}{T^4N^2h^2} \sum_{r_1<r_2<s_1<s_2,\mathcal{S}_{3,1}} \sum_{i_1,j_1,i_2,j_2} \bar{k}_{s_1r_1}^2 \bar{k}_{s_2r_2}^2 E\left(F_{r_1}^2 e_{i_1r_1} e_{j_1r_1}\right) E\left(F_{r_2}^2 e_{i_2r_2} e_{j_2r_2}\right) \\
&\quad \times E\left(F_{s_1}^2 e_{i_1s_1} e_{j_1s_1}\right) E\left(F_{s_2}^2 e_{i_2s_2} e_{j_2s_2}\right) + o(1).
\end{aligned}
$$

It follows that $b_{1NT}(1,c) = \mathbb{V}_{NT}^2(1,c) + o(1)$. Analogously, we can show that $b_{1NT}(1,v) = \mathbb{V}_{NT}^2(1,v) + o(1)$ for $v = d, e, f$. Consequently, we have $b_{1NT}(1) = \mathbb{V}_{NT}^2(1) + o(1)$ and $b_{1NT} = \mathbb{V}_{NT}^2 + o(1)$. Using arguments as used in the analysis of b_{1NT} and Lemma B.4, we can also show that

$$
\begin{aligned}
b_{2NT} &= \frac{16}{T^4N^2h^2} \sum_{s_1=2}^{T} \sum_{s_2=2}^{T} \sum_{1\leq r_1\neq r_2\leq s_1-1} \sum_{1\leq r_3\neq r_4\leq s_2-1} \bar{k}_{s_1r_1} \bar{k}_{s_1r_2} \bar{k}_{s_2r_3} \bar{k}_{s_2r_4} E\left(\phi_{s_1r_1} \phi_{s_1r_2} \phi_{s_2r_3} \phi_{s_2r_4}\right) \\
&= O\left(T^{-1}h^{-2} + N^2Th^2\alpha\left(T_0\right)^{\delta/(2+\delta)} + T^{-2}T_0^4 + T^{-2}T_0^3h^{-1} + T^{-2}T_0^2h^{-2}\right) = o(1).
\end{aligned}
$$

It follows that $E\left(\sum_{s=2}^{T}Z_{NT,s}^{2}\right)^{2}=\mathbb{V}_{NT}^{2}+o\left(1\right)$ and $\operatorname{Var}\left(\sum_{s=2}^{T}Z_{NT,s}^{2}\right)=o_{P}\left(1\right).$ Then the second part of (B.4) follows by Chebyshev inequality. In addition, by straightfoward moment calculations, we can show that $M_{1,1}^{(2,1,b)}=o_{P}\left(1\right).$ It follows that $M_{1,1}^{(2)}-\mathbb{B}_{NT}\overset{d}{\to}N\left(0,\mathbb{V}_{0}\right).$

Now, using $L_{st}-\bar{L}_{st}=k_{h,st}\left(H^{(t)}H^{(t)\prime}-H_{0}H_{0}^{\prime}\right)-(HH^{\prime}-H_{0}H_{0}^{\prime})$, we decompose $M_{1,1}^{(3)}$ as follows

$$\begin{aligned}
M_{1,1}^{(3)} &= \frac{2h^{1/2}}{N^{1/2}T^{2}}\sum_{t=1}^{T}\sum_{1\leq s\neq r\leq T}k_{h,st}F_{t}^{\prime}\left(H^{(t)}H^{(t)\prime}-H_{0}H_{0}^{\prime}\right)F_{s}F_{r}^{\prime}\bar{L}_{rt}F_{t}e_{s}^{\prime}e_{r}\\
&\quad+\frac{2h^{1/2}}{N^{1/2}T^{2}}\sum_{t=1}^{T}\sum_{1\leq s\neq r\leq T}F_{t}^{\prime}(HH^{\prime}-H_{0}H_{0}^{\prime})F_{s}F_{r}^{\prime}\bar{L}_{rt}F_{t}e_{s}^{\prime}e_{r}\equiv M_{1,1}^{(3,1)}+M_{1,1}^{(3,2)},\text{ say.}
\end{aligned}$$

For the first term, by Lemma B.1(ii) and letting $b_{4NT}\equiv\frac{1}{T}\sum_{t=1}^{T}\left\Vert\frac{h^{1/2}}{N^{1/2}T}\sum_{1\leq s\neq r\leq T}k_{h,st}\left(k_{h,rt}-1\right)F_{s}F_{r}^{\prime}e_{s}^{\prime}e_{r}\right\Vert^{2}$, we have

$$\begin{aligned}
\left|M_{1,1}^{(3,1)}\right| &\leq \frac{2}{T}\sum_{t=1}^{T}\left\Vert\frac{h^{1/2}}{N^{1/2}T}\sum_{1\leq s\neq r\leq T}k_{h,st}\left(k_{h,rt}-1\right)H_{0}H_{0}^{\prime}F_{t}F_{t}^{\prime}\left(H^{(t)}H^{(t)\prime}-H_{0}H_{0}^{\prime}\right)F_{s}F_{r}^{\prime}e_{s}^{\prime}e_{r}\right\Vert\\
&\leq 2\left\Vert H_{0}\right\Vert^{2}\max_{t}\left\Vert H^{(t)}H^{(t)\prime}-H_{0}H_{0}^{\prime}\right\Vert\frac{1}{T}\sum_{t=1}^{T}\left\Vert F_{t}\right\Vert^{2}\left\Vert\frac{h^{1/2}}{N^{1/2}T}\sum_{1\leq s\neq r\leq T}k_{h,st}\left(k_{h,rt}-1\right)F_{s}F_{r}^{\prime}e_{s}^{\prime}e_{r}\right\Vert\\
&\leq 2\left\Vert H_{0}\right\Vert^{2}\max_{t}\left\Vert H^{(t)}H^{(t)\prime}-H_{0}H_{0}^{\prime}\right\Vert\left\{\frac{1}{T}\sum_{t=1}^{T}\left\Vert F_{t}\right\Vert^{4}\right\}^{1/2}\left\{b_{4NT}\right\}^{1/2}\\
&= O_{P}\left(\left(Th/\ln T\right)^{-1/2}\right)O_{P}\left(1\right)O_{P}\left(1\right)=o_{P}\left(1\right),
\end{aligned}$$

where we also use the fact that $E\left(b_{4NT}\right)=O\left(1\right)$ by using Lemma B.4 and arguments as used in the above study of b_{1NT}. Similarly, we can show that $M_{1,1}^{(3,2)}=o_{P}\left(1\right).$ Thus $M_{1,1}^{(3)}=o_{P}\left(1\right).$ By the same token, we can show that $M_{1,1}^{(4)}=o_{P}\left(1\right).$ Consequently, we have shown that $M_{1,1}-\mathbb{B}_{NT}\overset{d}{\to}N\left(0,\mathbb{V}_{0}\right).$

Next, we show (i2) $M_{1,2}=\Pi_{1}+o_{P}\left(1\right).$ We make the following decomposition

$$\begin{aligned}
M_{1,2} &= N^{-1/2}h^{1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}d_{1it,2}^{2}\\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[F_{t}^{\prime}H^{(t)}H^{(t)\prime}\frac{1}{T}\sum_{s=1}^{T}F_{s}g_{is}^{\dagger}+\lambda_{i0}^{\prime}\left(H^{-1}\right)^{\prime}V_{NT}^{-1}\left(\frac{1}{T}\tilde{F}^{\prime}F\right)\Lambda_{0}^{\prime}g_{t}^{\dagger}/N\right]^{2}\\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[F_{t}^{\prime}H^{(t)}H^{(t)\prime}\frac{1}{T}\sum_{s=1}^{T}F_{s}F_{s}^{\prime}g_{is}\right]^{2}+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\lambda_{i0}^{\prime}\left(H^{-1}\right)^{\prime}V_{NT}^{-1}\left(\frac{1}{T}\tilde{F}^{\prime}F\right)\Lambda_{0}^{\prime}g_{t}^{\dagger}/N\right]^{2}\\
&\quad+2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[F_{t}^{\prime}H^{(t)}H^{(t)\prime}\frac{1}{T}\sum_{s=1}^{T}F_{s}F_{s}^{\prime}g_{is}\right]\left[\lambda_{i0}^{\prime}\left(H^{-1}\right)^{\prime}V_{NT}^{-1}\left(\frac{1}{T}\tilde{F}^{\prime}F\right)\Lambda_{0}^{\prime}g_{t}^{\dagger}/N\right]\\
&\equiv M_{1,2}^{(1)}+M_{1,2}^{(2)}+2M_{1,2}^{(3)},\text{ say.}
\end{aligned}$$

In view of the fact that $\frac{1}{T}\sum_{s=1}^{T}F_{s}F_{s}^{\prime}g_{is}=\Sigma_{F}\frac{1}{T}\sum_{s=1}^{T}g_{is}+\frac{1}{T}\sum_{s=1}^{T}\left(F_{s}F_{s}^{\prime}-\Sigma_{F}\right)g_{is}=\Sigma_{F}[\int_{0}^{1}g_{i}\left(u\right)du+T^{-1}]+o_{P}\left(1\right)=o_{P}\left(1\right)$ uniformly in i, we can readily show that $M_{1,2}^{(1)}=o_{P}\left(1\right).$ Noting that $\left(H^{-1}\right)^{\prime}V_{NT}^{-1}$

$\left(\frac{1}{T}\tilde{F}'F\right) = \left(H^{-1}\right)' V_{NT}^{-1}\{\frac{1}{T}\tilde{F}'[\tilde{F}H^{-1} + (F - \tilde{F}H^{-1})]\} = \left(H^{-1}\right)' V_{NT}^{-1}H^{-1} + o_P(1) = \left(H_0^{-1}\right)' V_0^{-1}H_0^{-1} + o_P(1)$ by Lemmas B.2(i), (iii), (v) and (vi) and the fact that $\frac{1}{T}\tilde{F}'\tilde{F} = \mathbb{I}_R$, we have

$$\begin{aligned}
M_{1,2}^{(2)} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\lambda_{i0}'\left(H_0^{-1}\right)' V_0^{-1}H_0^{-1}\Lambda_0' g_t^{\dagger}/N\right]^2 + o_P(1) \\
&= \frac{1}{T}\sum_{t=1}^{T}\mathrm{tr}\left[\left(H_0^{-1}\right)' V_0^{-1}H_0^{-1}\left(N^{-1}\Lambda_0' g_t^{\dagger}\right)\left(N^{-1}g_t^{\dagger\prime}\Lambda_0\right)\left(H_0^{-1}\right)' V_0^{-1}H_0^{-1}\left(N^{-1}\Lambda_0'\Lambda_0\right)\right] + o_P(1) \\
&= \frac{1}{T}\sum_{t=1}^{T}\mathrm{tr}\left[\left(H_0^{-1}\right)' V_0^{-1}H_0^{-1}\left[\left(N^{-1}\Lambda_0' g_t^{\dagger}\right)\left(N^{-1}g_t^{\dagger\prime}\Lambda_0\right)\right]\left(H_0^{-1}\right)' V_0^{-1}H_0^{-1}\Sigma_{\Lambda_0}\right] + o_P(1) \\
&= \Pi_1 + o_P(1).
\end{aligned}$$

In addition, $M_{1,2}^{(3)} \leq \{M_{1,2}^{(1)}M_{1,2}^{(2)}\}^{1/2} = o_P(1)$ by CS inequality. It follows that $M_{1,2} = \Pi_1 + o_P(1)$.

Now, we show (i3) $M_{1,l} = o_P(1)$ **for** $l = 3, 4, ..., 10$. Let $d_{1it,3}^{(1)} = \lambda_{i0}'[(H^{(t)-1})'\hat{S}_{\lambda,t}^{-1}H^{(t)-1} - (H^{-1})'V_{NT}^{-1}(\frac{1}{T}\tilde{F}'F)]\Lambda_0' e_t/N$, $d_{1it,3}^{(2)} = a_{NT}\lambda_{i0}'(H^{(t)-1})'\hat{S}_{\lambda,t}^{-1}H^{(t)-1}\frac{1}{N}\sum_{i=1}^{N} g_{it}e_{it}$, and $d_{1it,3}^{(3)} = a_{NT}g_{it}'(H^{(t)-1})'\hat{S}_{\lambda,t}^{-1}H^{(t)-1}\frac{1}{N}\sum_{i=1}^{N}\lambda_{it}e_{it}$. Then $d_{1it,3} = d_{1it,3}^{(1)} + d_{1it,3}^{(2)} + d_{1it,3}^{(3)}$ and

$$\begin{aligned}
M_{1,3} &= N^{-1/2}h^{1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}d_{1it,3}^2 \leq 3N^{-1/2}h^{1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\left(d_{1it,3}^{(1)}\right)^2 + \left(d_{1it,3}^{(2)}\right)^2 + \left(d_{1it,3}^{(3)}\right)^2\right] \\
&\equiv 3(M_{1,3}^{(1)} + M_{1,3}^{(2)} + M_{1,3}^{(3)}), \text{ say.}
\end{aligned}$$

For $M_{1,3}^{(1)}$, we apply Lemma B.2(viii) to obtain

$$\begin{aligned}
M_{1,3}^{(1)} &\leq N^{-1/2}Th^{1/2}\max_t\left\|\left(H^{(t)-1}\right)'\hat{S}_{\lambda,t}^{-1}H^{(t)-1} - \left(H^{-1}\right)' V_{NT}^{-1}\left(\frac{1}{T}\tilde{F}'F^0\right)\right\|^2\frac{1}{N}\sum_{i=1}^{N}\|\lambda_{i0}\|^2\frac{1}{TN}\sum_{t=1}^{T}\|\Lambda_0' e_t\|^2 \\
&= N^{-1/2}Th^{1/2}O_P((Th\ln T)^{-1} + N^{-1})O_P(1) = o_P(1).
\end{aligned}$$

For $M_{1,3}^{(2)}$, we have by Lemmas B.1(ii) and (vi)

$$M_{1,3}^{(2)} \leq \max_t\left\|\left(H^{(t)-1}\right)'\hat{S}_{\lambda,t}^{-1}H^{(t)-1}\right\|^2\frac{1}{N}\sum_{i=1}^{N}\|\lambda_{i0}\|^2\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{j=1}^{N}g_{jt}e_{jt}\right\|^2 = O_P(1)\,O(1)\,o_P(1) = o_P(1).$$

Similarly, we can show that $M_{1,3}^{(3)} = o_P(1)$. Thus $M_{1,3} = o_P(1)$. By CS inequality and Lemmas B.3(i)-(iv), $M_{1,4} = N^{-1/2}h^{1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}d_{1it,4}^2 \leq 4N^{-1/2}h^{1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\{\left\|F_t' H^{(t)}R_{\Lambda}(i,t)\right\|^2 + \left\|F_t' H R_{\Lambda}^0(i)\right\|^2 + ||\lambda_{it}'(H^{(t)-1})'R_F(t)||^2 + ||\lambda_{i0}'(H^{-1})'R_F^0(t)||^2\} = N^{1/2}Th^{1/2}O_P\left(T^{-2}h^{-2} + N^{-2}(\ln T)^2\right) = o_P(1)$. By CS inequality,

$$\begin{aligned}
|M_{1,8}| &= \left|N^{-1/2}h^{1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}d_{1it,2}d_{1it,3}\right| \leq \{M_{1,2}M_{1,3}\}^{1/2} = o_P(1), \\
|M_{1,9}| &= \left|N^{-1/2}h^{1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}d_{1it,2}d_{1it,4}\right| \leq \{M_{1,2}M_{1,4}\}^{1/2} = o_P(1), \\
|M_{1,10}| &= \left|N^{-1/2}h^{1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}d_{1it,3}d_{1it,4}\right| \leq \{M_{1,3}M_{1,4}\}^{1/2} = o_P(1).
\end{aligned}$$

We are left to show that $M_{1,l} = o_P(1)$ for $l = 5, 6, 7$. To conserve the space, we prove these claims in Lemma B.5(i)-(iii), below. This completes the proof of the theorem. ■

Lemma B.5 *Suppose that Assumptions A.1, A.3(i) and (ii*) and A.6-A.7 hold. Suppose that* $\mathbb{H}_1(a_{NT})$ *holds true. Then*

(i) $M_{1,5} = o_P(1)$,

(ii) $M_{1,6} = o_P(1)$,

(iv) $M_{1,7} = o_P(1)$.

Lemma B.6 *Suppose that Assumptions A.1, A.3(i) and (ii*) and A.6-A.7 hold. Suppose that* $\mathbb{H}_1(a_{NT})$ *holds true. Then*

(i) $M_2 = \Pi_2 + o_P(1)$,

(ii) $M_3 = o_P(1)$,

(iii) $M_4 = o_P(1)$,

(iv) $M_5 = o_P(1)$,

(v) $M_6 = o_P(1)$.

To prove Theorem 4.2, we need another lemma.

Lemma B.7 *Suppose that Assumptions A.1, A.3(i) and (ii*) and A.6-A.7 hold. Then under* $\mathbb{H}_1(a_{NT})$,

(i) $\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_t\right\|^l = O_P(1)$ *for* $l = 4, 6, 8$,

(ii) $\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{L}_{st}^2\left\|\hat{F}_s\right\|^l = O_P\left(h^{-1}\right)$ *for* $l = 1, 2$,

(iii) $\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{L}_{st}^2\left\|\hat{F}_s - H^{(s)\prime}F_s\right\|^2 = o_P\left(N^{-1/2}h^{-1/2}\right)$,

(iv) $\frac{1}{T^3}\sum_{s=1}^{T}\left\|\hat{F}_s\right\|^l\left[\sum_{t=1}^{T}\hat{L}_{st}^2\right]^2 = O_P\left(h^{-2}\right)$ *for* $l = 0, 2$,

(v) $\frac{h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}k_{h,st}^2\left[F_t'H^{(t)}(\hat{F}_s - H^{(s)\prime}F_s)\right]F_s'H^{(t)}H^{(t)\prime}F_te_{is}^2 = o_P(1)$,

(vi) $\frac{h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}k_{h,st}^2\left[F_s'H^{(s)}(\hat{F}_t - H^{(t)\prime}F_t)\right]F_s'H^{(t)}H^{(t)\prime}F_te_{is}^2 = o_P(1)$.

Proof of Theorem 4.2. Given Theorem 4.1, it suffices to prove the first two parts of the theorem. In fact, we prove the first two parts of the theorem under $\mathbb{H}_1(a_{NT})$ so that they are still applicable for Theorem 4.3 below.

Step 1. We prove (i) $\hat{\mathbb{B}}_{NT} - \mathbb{B}_{NT} = o_P(1)$ **under** $\mathbb{H}_1(a_{NT})$. Let $\bar{L}_{st} = F_s'L_{st}F_t = F_s'[k_{h,st}H^{(t)}H^{(t)\prime} - HH']F_t$ and $\hat{L}_{st} = k_{h,st}\hat{F}_s'\hat{F}_t - \tilde{F}_s'\tilde{F}_t$. Using $\hat{e}_{is}^2 - e_{is}^2 = (\hat{e}_{is} - e_{is})^2 + 2(\hat{e}_{is} - e_{is})e_{is}$, we have

$$
\begin{aligned}
\hat{\mathbb{B}}_{NT} - \mathbb{B}_{NT} &= \frac{h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\hat{L}_{st}^2\hat{e}_{is}^2 - \bar{L}_{st}^2e_{is}^2\right) \\
&= \frac{h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\hat{L}_{st}^2(\hat{e}_{is} - e_{is})^2 + 2\hat{L}_{st}^2(\hat{e}_{is} - e_{is})e_{is} + \left(\hat{L}_{st}^2 - \bar{L}_{st}^2\right)e_{is}^2\right] \\
&\equiv B_1 + 2B_2 + B_3, \text{ say.}
\end{aligned}
$$

It suffices to show that (i1) $B_1 = o_P(1)$, (i2) $B_2 = o_P(1)$, and (i3) $B_3 = o_P(1)$. To show (i1), we make the following decomposition:

$$e_{is} - \hat{e}_{is} = \hat{\lambda}_{is}'\hat{F}_s - \lambda_{is}'F_s = (\hat{\lambda}_{is} - H^{(s)^{-1}}\lambda_{is})'\hat{F}_s + \lambda_{is}'H^{(s)\prime^{-1}}(\hat{F}_s - H^{(s)\prime}F_s) \equiv d_{e1,is} + d_{e2,is}, \text{ say.} \quad \text{(B.5)}$$

By CS inequality, $B_1 \leq \frac{2h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T \hat{L}_{st}^2\left(d_{e1,is}^2 + d_{e2,is}^2\right) \equiv 2B_{1,1} + 2B_{1,2}$. By Lemmas B.1(vii) and B.7(ii)

$$\begin{aligned} B_{1,1} &= \frac{h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T \hat{L}_{st}^2\hat{F}_s'\left(\hat{\lambda}_{is} - H^{(s)^{-1}}\lambda_{is}\right)\left(\hat{\lambda}_{is} - H^{(s)^{-1}}\lambda_{is}\right)'\hat{F}_s \\ &\leq N^{1/2}h^{1/2}\left\{\max_s \frac{1}{N}\sum_{i=1}^N\left\|\hat{\lambda}_{is} - H^{(s)^{-1}}\lambda_{is}\right\|^2\right\}\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T\hat{L}_{st}^2\left\|\hat{F}_s\right\|^2 \\ &= N^{1/2}h^{1/2}O_P\left(C_{NT}^{-2}\ln T\right)O_P\left(h^{-1}\right) = o_P(1). \end{aligned}$$

Similarly, by Lemmas B.1(ii) and B.7(iii)

$$\begin{aligned} B_{1,2} &= \frac{h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T \hat{L}_{st}^2\lambda_{is}'H^{(s)\prime^{-1}}\left(\hat{F}_s - H^{(s)\prime}F_s\right)\left(\hat{F}_s - H^{(s)\prime}F_s\right)'H^{(s)^{-1}}\lambda_{is} \\ &\leq \bar{c}_\lambda^2\max_s\left\|H^{(s)^{-1}}\right\|^2\frac{N^{1/2}h^{1/2}}{T^2}\sum_{t=1}^T\sum_{s=1}^T\hat{L}_{st}^2\left\|\hat{F}_s - H^{(s)\prime}F_s\right\|^2 = o_P(1). \end{aligned}$$

Next, we show (i2). Using (B.5), we decompose B_2 as follows

$$B_2 = \frac{h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^N\sum_{t=1}^T\hat{L}_{st}^2\left(\hat{e}_{is} - e_{is}\right)e_{is} = \frac{-h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T\hat{L}_{st}^2\left(d_{e1,is} + d_{e2,is}\right)e_{is} \equiv -B_{2,1} - B_{2,2}, \text{ say.}$$

By (B.2), we further decompose $B_{2,1}$:

$$\begin{aligned} B_{2,1} &= \frac{h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T\hat{L}_{st}^2\hat{F}_s'\left(\hat{\lambda}_{is} - H^{(s)^{-1}}\lambda_{is}\right)e_{is} \\ &= \frac{h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T\hat{L}_{st}^2\hat{F}_s'\left[D_1(i,s) + D_2(i,s) - D_3(i,s)\right]e_{is} \equiv B_{2,1}^{(1)} + B_{2,1}^{(2)} - B_{2,1}^{(3)}, \text{ say.} \end{aligned}$$

For $B_{2,1}^{(1)}$, we have by Lemma B.7(iv),

$$\begin{aligned} B_{2,1}^{(1)} &= \frac{h^{1/2}}{T^2}\sum_{t=1}^T\sum_{s=1}^T\hat{L}_{st}^2\hat{F}_s'H^{(s)\prime}\left(\frac{1}{TN^{1/2}}\sum_{i=1}^N\sum_{r=1}^T F_r^{(s)}e_{ir}^{(s)}e_{is}\right) \\ &\leq h^{1/2}\max_s\left\|H^{(s)}\right\|\left\{\frac{1}{T^3}\sum_{s=1}^T\left\|\hat{F}_s\right\|^2\left(\sum_{t=1}^T\hat{L}_{st}^2\right)^2\right\}^{1/2}\left\{\frac{1}{NT}\sum_{s=1}^T\left\|\frac{1}{T}\sum_{i=1}^N\sum_{r=1}^T k_{h,rs}F_re_{ir}e_{is}\right\|^2\right\}^{1/2} \\ &= h^{1/2}O_P(1)\,O_P\left(h^{-1}\right)O_P\left((Th)^{-1/2} + N^{1/2}T^{-1}h^{-1/2}\right) = o_P(1), \end{aligned}$$

where we use the fact that $\frac{1}{NT}\sum_{s=1}^T E\left\|\frac{1}{T}\sum_{i=1}^N\sum_{r=1}^T k_{h,rs}F_re_{ir}e_{is}\right\|^2 = O\left((Th)^{-1} + NT^{-2}h^{-1}\right) = o(1)$ by moment calculations.

For $B_{2,1}^{(2)}$, we have by Lemmas B.1(v) and B.7(ii),

$$\begin{aligned}
B_{2,1}^{(2)} &= \frac{h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{L}_{st}^2\hat{F}_s' D_2(i,s)e_{is} \\
&= \frac{h^{1/2}}{T^2N^{1/2}}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{L}_{st}^2\hat{F}_s'\sum_{i=1}^{N}\frac{1}{T}\sum_{r=1}^{T}\left[\hat{F}_r^{(s)}-H^{(s)\prime}F_r^{(s)}\right]e_{ir}^{(s)}e_{is} \\
&\leq N^{1/2}h^{1/2}\max_{i,s}\left\|\frac{1}{T}\sum_{r=1}^{T}\left[\hat{F}_r^{(s)}-H^{(s)\prime}F_r^{(s)}\right]e_{ir}^{(s)}\right\|\left\{\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{L}_{st}^2\left\|\hat{F}_s\right\|\right\}\frac{1}{N}\sum_{i=1}^{N}|e_{is}| \\
&= N^{1/2}h^{1/2}O_P\left(T^{-1}h^{-1}+N^{-1}\ln(NT)\right)O_P\left(h^{-1}\right)O_P\left(1\right)=o_P\left(1\right).
\end{aligned}$$

For $B_{2,1}^{(3)}$, by Lemma B.1(iv)-(v) and B.7(ii),

$$\begin{aligned}
B_{2,1}^{(3)} &= \frac{h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{L}_{st}^2\hat{F}_s' D_3(i,s)e_{is} \\
&= \frac{h^{1/2}}{T^2N^{1/2}}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{L}_{st}^2\hat{F}_s'\frac{1}{T}\hat{F}^{(s)\prime}(\hat{F}^{(s)}H^{(s)-1}-F^{(s)})\sum_{i=1}^{N}\lambda_{is}e_{is} \\
&\leq N^{1/2}h^{-1/2}\max_{s}\frac{1}{T}\left\|\hat{F}^{(s)\prime}(\hat{F}^{(s)}H^{(s)-1}-F^{(s)})\right\|\max_{s}\left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_{is}e_{is}\right\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{L}_{st}^2\left\|\hat{F}_s\right\| \\
&= O_P\left(N^{1/2}h^{-1/2}\right)\left(O_P\left(C_{NT}^{-2}\right)+o_P\left(a_{NT}\right)\right)O_P\left(N^{-1/2}\ln T\right)O_P\left(h^{-1}\right)=o_P\left(1\right).
\end{aligned}$$

Thus $B_{2,1}=o_P\left(1\right)$. In addition, by Lemma B.7(iii),

$$\begin{aligned}
\left|B_{2,2}\right| &= \frac{h^{1/2}}{T^2N^{1/2}}\left|\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{L}_{st}^2\lambda_{is}'H^{(s)\prime^{-1}}\left(\hat{F}_s-H^{(s)\prime}F_s\right)e_{is}\right| \\
&\leq \frac{h^{1/2}}{TN^{1/2}}\left|\sum_{s=1}^{T}\left[\frac{1}{T}\sum_{t=1}^{T}\hat{L}_{st}^2\right]\left[\sum_{i=1}^{N}\lambda_{is}'H^{(s)\prime^{-1}}\left(\hat{F}_s-H^{(s)\prime}F_s\right)e_{is}\right]\right| \\
&\leq h^{1/2}\left\{\frac{1}{T}\sum_{s=1}^{T}\left(\frac{1}{T}\sum_{t=1}^{T}\hat{L}_{st}^2\right)^2\right\}^{1/2}\left\{\frac{1}{NT}\sum_{s=1}^{T}\left\|\sum_{i=1}^{N}\lambda_{is}'H^{(s)\prime^{-1}}\left(\hat{F}_s-H^{(s)\prime}F_s\right)e_{is}\right\|^2\right\}^{1/2} \\
&= h^{1/2}O_P\left(h^{-1}\right)O_P\left(C_{NT}^{-1}\right)=o_P\left(1\right),
\end{aligned}$$

where we use the fact that $\frac{1}{NT}\sum_{s=1}^{T}\left\|\sum_{i=1}^{N}\lambda_{is}'H^{(s)\prime^{-1}}\left(\hat{F}_s-H^{(s)\prime}F_s\right)e_{is}\right\|^2=O_P\left(C_{NT}^{-2}\right)$ by arguments as used in the proof of Lemma B.1(viii). Thus $B_2=o_P\left(1\right)$.

Now, we show (i3). For B_3, using $\bar{L}_{st}=F_s'L_{st}F_t=F_s'[k_{h,st}H^{(t)}H^{(t)\prime}-HH']F_t$ and $\hat{L}_{st}=k_{h,st}\hat{F}_s'\hat{F}_t-$

$\tilde{F}_s'\tilde{F}_t$, we make the following decomposition:

$$
\begin{aligned}
B_3 &= \frac{h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{L}_{st}^2-\bar{L}_{st}^2\right)e_{is}^2 \\
&= \frac{h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}k_{h,st}^2\left[\left(\hat{F}_s'\hat{F}_t\right)^2-\left(F_s'H^{(t)}H^{(t)\prime}F_t\right)^2\right]e_{is}^2 \\
&\quad -\frac{2h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}k_{h,st}\left[\hat{F}_s'\hat{F}_t\tilde{F}_s'\tilde{F}_t-\left(F_s'H^{(t)}H^{(t)\prime}F_t\right)F_s'HH'F_t\right]e_{is}^2 \\
&\quad +\frac{h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\left(\tilde{F}_s'\tilde{F}_t\right)^2-\left(F_s'HH'F_t\right)^2\right]e_{is}^2\equiv B_{3,1}+B_{3,2}+B_{3,3},\text{ say.}
\end{aligned}
$$

Using $a^2-b^2=(a-b)^2+2(a-b)b$ and $\hat{F}_s'\hat{F}_t-H^{(s)\prime}F_sF_t'H^{(t)}=(\hat{F}_s-H^{(s)\prime}F_s)'(\hat{F}_t-H^{(t)\prime}F_t)+(\hat{F}_s-H^{(s)\prime}F_s)'H^{(t)\prime}F_t+F_s'H^{(s)}(\hat{F}_t-H^{(t)\prime}F_t)$, we can bound $|B_{3,1}|$ as follows

$$
\begin{aligned}
|B_{3,1}| &\le \frac{h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}k_{h,st}^2\Big\{3\left[(\hat{F}_s-H^{(s)\prime}F_s)'(\hat{F}_t-H^{(t)\prime}F_t)\right]^2+3\left[(\hat{F}_s-H^{(s)\prime}F_s)'H^{(t)\prime}F_t\right]^2 \\
&\quad +3\left[F_s'H^{(s)}(\hat{F}_t-H^{(t)\prime}F_t)\right]^2+2\left[(\hat{F}_t-H^{(t)\prime}F_t)'(\hat{F}_s-H^{(s)\prime}F_s)\right]F_s'H^{(t)}H^{(t)\prime}F_t \\
&\quad +2\left[F_t'H^{(t)}(\hat{F}_s-H^{(s)\prime}F_s)\right]F_s'H^{(t)}H^{(t)\prime}F_t+2\left[F_s'H^{(s)}(\hat{F}_t-H^{(t)\prime}F_t)\right]F_s'H^{(t)}H^{(t)\prime}F_t\Big\}e_{is}^2 \\
&\equiv 3B_{3,1}^{(1)}+3B_{3,1}^{(2)}+3B_{3,1}^{(3)}+2B_{3,1}^{(4)}+2B_{3,1}^{(5)}+2B_{3,1}^{(6)},\text{ say.}
\end{aligned}
$$

By Lemma A.2(i) and the fact that $\max_i \frac{1}{N}\sum_{i=1}^{N}e_{is}^2=O_P(1)$, we have

$$
\begin{aligned}
B_{3,1}^{(1)} &= \frac{h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}k_{h,st}^2\left[(\hat{F}_s-H^{(s)\prime}F_s)'(\hat{F}_t-H^{(t)\prime}F_t)\right]^2e_{is}^2 \\
&\le N^{1/2}h^{1/2}\left(\max_{s,t}k_{h,st}^2\right)\max_i\left(\frac{1}{N}\sum_{i=1}^{N}e_{is}^2\right)\left\{\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s-H^{(s)\prime}F_s\right\|^2\right\}^2 \\
&= N^{1/2}h^{1/2}O\left(h^{-2}\right)O_P(1)\,O_P\left(C_{NT}^{-4}\right)=o_P(1).
\end{aligned}
$$

By Lemmas B.1(x), we can readily show that

$$
\begin{aligned}
B_{3,1}^{(2)} &= \frac{h^{1/2}}{T^2N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}k_{h,st}^2\left[(\hat{F}_s-H^{(s)\prime}F_s)'H^{(t)\prime}F_t\right]^2e_{is}^2 \\
&\le N^{1/2}h^{1/2}\left\{\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s-H^{(s)\prime}F_s\right\|^2\right\}\max_t\left\|H^{(t)}\right\|^2\max_s\left(\frac{1}{N}\sum_{i=1}^{N}e_{is}^2\right)\max_s\left(\frac{1}{T}\sum_{t=1}^{T}k_{h,st}^2\left\|F_t\right\|^2\right) \\
&= N^{1/2}h^{1/2}O_P\left(C_{NT}^{-2}\right)O_P(1)\,O_P(1)\,O_P\left(h^{-1}\right)=o_P(1).
\end{aligned}
$$

Similarly, we can show that $B_{3,1}^{(l)}=o_P(1)$ for $l=3,4$ by using Lemma B.1(x). By Lemmas B.7(v)-(vi), $B_{3,1}^{(l)}=o_P(1)$ for $l=5,6$. It follows that $B_{3,1}=o_P(1)$. Similarly, we have $B_{3,l}=o_P(1)$ for $l=2,3$. Then $B_3=o_P(1)$. This completes the proof of part (i).

Step 2. We show (ii) $\hat{\mathbb{V}}_{NT}=\mathbb{V}_{NT}+o_P(1)$ **under** $\mathbb{H}_1(a_{NT})$. Let $\bar{k}_{sr}=\bar{K}\left(\frac{s-r}{Th}\right)$. Let $\bar{\Sigma}_F=H_0'\Sigma_F H_0$ and $\bar{\mathbb{V}}_{NT}=2T^{-2}N^{-1}h^{-1}\sum_{1\leq s\neq r\leq T}\bar{k}_{sr}^2\left(F_s'H_0\bar{\Sigma}_F H_0'F_r\right)^2(e_r'e_s)^2$. we make the following decomposition

$$
\begin{aligned}
\hat{\mathbb{V}}_{NT}-\mathbb{V}_{NT} &= 2T^{-2}N^{-1}h^{-1}\sum_{1\leq s\neq r\leq T}\bar{k}_{sr}^2\left(\hat{F}_s'\hat{\Sigma}_F\hat{F}_r\right)^2\left[(\hat{e}_r'\hat{e}_s)^2-(e_r'e_s)^2\right] \\
&\quad +2T^{-2}N^{-1}h^{-1}\sum_{1\leq s\neq r\leq T}\bar{k}_{sr}^2\left[\left(\hat{F}_s'\hat{\Sigma}_F\hat{F}_r\right)^2-\left(F_s'H_0\bar{\Sigma}_F H_0'F_r\right)^2\right](e_r'e_s)^2 \\
&\quad +2T^{-2}N^{-1}h^{-1}\sum_{1\leq s\neq r\leq T}\bar{k}_{sr}^2\left[\phi_{sr}^2-E\left(\phi_{sr}^2\right)\right] \\
&\equiv 2\mathbb{V}_{1NT}+2\mathbb{V}_{2NT}+2\mathbb{V}_{3NT},\ \text{say,}
\end{aligned}
$$

where recall $\phi_{sr}=F_s'H_0\bar{\Sigma}_F H_0'F_r e_r'e_s$. It suffices to show (ii1) $\mathbb{V}_{1NT}=o_P(1)$, (ii1) $\mathbb{V}_{2NT}=o_P(1)$, and (ii3) $\mathbb{V}_{3NT}=o_P(1)$. We prove (ii1)-(ii2) in Lemma B.8 below. For $\mathbb{V}_{3NT}$, observe that $E(\mathbb{V}_{3NT})=0$ and

$$
\begin{aligned}
\mathrm{Var}(2\mathbb{V}_{3NT}) &= \mathrm{Var}\left(4T^{-2}N^{-1}h^{-1}\sum_{s=2}^{T}\sum_{r=1}^{s-1}\bar{k}_{sr}^2\phi_{sr}^2\right)=\frac{16}{T^4N^2h^2}E\left(\sum_{s=2}^{T}\sum_{r=1}^{s-1}\bar{k}_{sr}^2\phi_{sr}^2\right)^2-\mathbb{V}_{NT}^2 \\
&= b_{1NT}-\mathbb{V}_{NT}^2=o(1)
\end{aligned}
$$

where b_{1NT} is defined in the proof of Theorem 4.1. Then $\mathbb{V}_{3NT}=o_P(1)$ by Chebyshev inequality. This completes the proof of the theorem ■

Lemma B.8 *Suppose that Assumptions A.1, A.3(i) and (ii*) and A.6-A.7 hold. Then under* $\mathbb{H}_1(a_{NT})$,

(i) $\mathbb{V}_{1NT}=T^{-2}N^{-1}h^{-1}\sum_{1\leq s\neq r\leq T}\bar{k}_{sr}^2\left(\hat{F}_s'\hat{\Sigma}_F\hat{F}_r\right)^2\left[(\hat{e}_r'\hat{e}_s)^2-(e_r'e_s)^2\right]=o_P(1)$,

(ii) $\mathbb{V}_{2NT}=T^{-2}N^{-1}h^{-1}\sum_{1\leq s\neq r\leq T}\bar{k}_{sr}^2\left[\left(\hat{F}_s'\hat{\Sigma}_F\hat{F}_r\right)^2-\left(F_s'H_0\bar{\Sigma}_F H_0'F_r\right)^2\right](e_r'e_s)^2=o_P(1)$.

Proof of Theorem 4.3. By the proof of Theorem 4.1, $J_{NT}\equiv\mathbb{V}_{NT}^{-1/2}\left(TN^{1/2}h^{1/2}\hat{M}-\mathbb{B}_{NT}\right)\xrightarrow{d}N(\pi_0,1)$ under $\mathbb{H}_1(a_{NT})$. By the proof of Theorem 4.2, $\hat{\mathbb{B}}_{NT}=\mathbb{B}_{NT}+o_P(1)$ and $\hat{\mathbb{V}}_{NT}=\mathbb{V}_{NT}+o_P(1)$ under under $\mathbb{H}_1(a_{NT})$. It follows that $\hat{J}_{NT}\equiv\hat{\mathbb{V}}_{NT}^{-1/2}\left(TN^{1/2}h^{1/2}\hat{M}-\hat{\mathbb{B}}_{NT}\right)\xrightarrow{d}N(\pi_0,1)$ under $\mathbb{H}_1(a_{NT})$. ■

Proof of Theorem 4.4. Under the global alternative $\mathbb{H}_1$, we have by (4.4)

$$
\begin{aligned}
\hat{M} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\left(\hat{\lambda}_{it}'\hat{F}_t-\lambda_{it}'F_t\right)+\left(\lambda_{it}'F_t-\tilde{\lambda}_{i0}'\tilde{F}_t\right)\right]^2 \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{\lambda}_{it}'\hat{F}_t-\lambda_{it}'F_t\right)^2+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\lambda_{it}'F_t-\tilde{\lambda}_{i0}'\tilde{F}_t\right)^2 \\
&\quad +\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{\lambda}_{it}'\hat{F}_t-\lambda_{it}'F_t\right)\left(\lambda_{it}'F_t-\tilde{\lambda}_{i0}'\tilde{F}_t\right) \\
&\equiv \hat{M}_1+\hat{M}_2+2\hat{M}_3,\ \text{say.}
\end{aligned}
$$

Using $\hat{\lambda}_{it}'\hat{F}_t - \lambda_{it}'F_t = (\hat{\lambda}_{it} - H^{(t)-1}\lambda_{it})'(\hat{F}_t - H^{(t)\prime}F_t) + (\hat{\lambda}_{it} - H^{(t)-1}\lambda_{it})'H^{(t)\prime}F_t + \left(H^{(t)-1}\lambda_{it}\right)'(\hat{F}_t - H^{(t)\prime}F_t)$, we can readily show that $\hat{M}_1 = o_P(1)$ by Lemmas B.1(viii) and (xi). By Assumption A.7, we have that for sufficiently large N and T,

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\lambda_{it}'F_t - \tilde{\lambda}_{i0}'\tilde{F}_t\right)^2 \geq \inf_{(\breve{\Lambda},\breve{F})\in\mathbf{\Lambda}_N\times\mathbf{F}_T}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\lambda_{it}'F_t - \breve{\lambda}_i'\breve{F}_t\right)^2 \geq \underline{c}_{\Lambda F}/2 > 0.$$

We can easily show that the left hand side object is $O_P(1)$ under $\mathbb{H}_1$. Then by CS inequality, $\hat{M}_3 \leq \left\{\hat{M}_1\hat{M}_2\right\}^{1/2} = o_P(1)$. Consequently, we have $P(\hat{M} \geq \underline{c}_{\Lambda F}/2) \to 1$.

In addition, we can show that $\hat{\mathbb{V}}_{NT}$ also converges to a positive number (say $\mathbb{V}_0$) and $\hat{\mathbb{B}}_{NT} = O_P\left(N^{1/2}h^{-1/2}\right) = o_P\left(TN^{1/2}h^{1/2}\right)$ under $\mathbb{H}_1$. It follows that

$$\frac{\hat{J}_{NT}}{TN^{1/2}h^{1/2}} = \hat{\mathbb{V}}_{NT}^{-1/2}\left(\hat{M} - \frac{\hat{\mathbb{B}}_{NT}}{TN^{1/2}h^{1/2}}\right) \geq \mathbb{V}_0^{-1/2}\underline{c}_{\Lambda F}/4$$

with probability approaching 1. Consequently $P\left(\hat{J}_{NT} \geq c_{NT}\right) \to 1$ as $(N,T) \to \infty$ for any $c_{NT} = o\left(TN^{1/2}h^{1/2}\right)$. ■

Proof of Theorem 4.5. Let P^* denote the probability measure induced by the wild bootstrap conditional on X . Let E^* and Var^* denote the expectation and variance under P^* and $O_{P^*}(\cdot)$ and $o_{P^*}(\cdot)$ the probability order under P^*. In view of the fact that (1) the null hypothesis is satisfied in the bootstrap world, (2) e_t^*'s are independent over t conditional on X, and (3) both $\tilde{\lambda}_{i0}$ and $\tilde{F}_t$ are fixed given X, the proof is similar to but simpler than that of 4.1 and 4.2.

Let $\hat{M}^*$, J_{NT}^*, $\mathbb{B}_{NT}^*$, $\mathbb{V}_{NT}^*$, $\hat{J}_{NT}^*$, $\hat{\mathbb{B}}_{NT}^*$, and $\hat{\mathbb{V}}_{NT}^*$ denote the bootstrap analogue of $\hat{M}$, J_{NT}, $\mathbb{B}_{NT}$, $\mathbb{V}_{NT}$, $\hat{J}_{NT}$, $\hat{\mathbb{B}}_{NT}$, and $\hat{\mathbb{V}}_{NT}$, respectively. Then $J_{NT}^* \equiv (TN^{1/2}h^{1/2}\hat{M}^* - \mathbb{B}_{NT}^*)/\sqrt{\mathbb{V}_{NT}^*}$ and $\hat{J}_{NT}^* \equiv \left(N^{-1/2}\hat{M}^* - \hat{\mathbb{B}}_{NT}^*\right)/\sqrt{\hat{\mathbb{V}}_{NT}^*}$. Following the proof of Theorem 4.1, we can show that

$$TN^{1/2}h^{1/2}\hat{M}^* - \mathbb{B}_{NT}^* = \sum_{s=2}^{T} Z_{NT,s}^* + o_{P^*}(1)$$

where $Z_{NT,s}^* = 2T^{-1}N^{-1/2}h^{-1/2}\sum_{r=1}^{s-1}\bar{k}_{sr}\tilde{F}_s'H\tilde{\Sigma}_{\tilde{F}}H'\tilde{F}_r e_s^{*\prime}e_r^*$, $e_s^* = (e_{N1}^*, ..., e_{Ns}^*)'$, and $\tilde{\Sigma}_{\tilde{F}} = T^{-1}\sum_{t=1}^{T}\tilde{F}_s\tilde{F}_s'$. [c.f. $Z_{NT,s} = 2T^{-1}N^{-1/2}h^{-1/2}\sum_{r=1}^{s-1}\bar{k}_{sr}F_s'H_0\Sigma_F H_0'F_r e_s'e_r$] Then we can prove the theorem by showing that: (i) $\sum_{s=2}^{T} Z_{NT,s}^*/\sqrt{\mathbb{V}_{NT}^*} \xrightarrow{D^*} N(0,1)$, (ii) $\hat{\mathbb{B}}_{NT}^* = \mathbb{B}_{NT}^* + o_{P^*}(1)$, and (iii) $\hat{\mathbb{V}}_{NT}^* = \mathbb{V}_{NT}^* + o_{P^*}(1)$.

We only outline the proof of (i) as those of other parts are analogous to the corresponding parts in the proof of Theorem 4.2. Noting that $\{Z_{NT,t}^*, \mathcal{F}_{NT,t}^*\}$ is an m.d.s., we can continue to apply the martingale CLT by showing that

$$\mathcal{Z}^* \equiv \sum_{t=2}^{T} E_{\mathcal{F}_{NT,t-1}^*}^*\left|Z_{NT,t}^*\right|^4 = o_{P^*}(1), \text{ and } \sum_{t=2}^{T} Z_{NT,t}^{*2} - \mathbb{V}_{NT}^* = o_{P^*}(1). \quad \text{(B.6)}$$

As in the proof of Theorem 4.1,

$$
\begin{aligned}
&E^*\left(\mathcal{Z}^*\right) \\
&= \frac{16}{T^4N^2h^2}\sum_{s=2}^{T}E^*\left[\sum_{r=1}^{s-1}\bar{k}_{sr}^4\phi_{sr}^{*4}+2\sum_{1\le r_1<r_2\le s-1}\bar{k}_{sr_1}^2\bar{k}_{sr_2}^2\phi_{sr_1}^{*2}\phi_{sr_2}^{*2}\right. \\
&\quad\left.+4\sum_{t=1}^{s-1}\sum_{1\le r_1<r_2\le s-1}\bar{k}_{st}^2\bar{k}_{sr_1}\bar{k}_{sr_2}\phi_{st}^{*2}\phi_{sr_1}^*\phi_{sr_2}^*+4\sum_{1\le r_1<r_2\le s-1,1\le t_1<t_2\le s-1}\bar{k}_{r_1s}\bar{k}_{r_2s}\bar{k}_{t_1s}\bar{k}_{t_2s}\phi_{sr_1}^*\phi_{sr_2}^*\phi_{st_1}^*\phi_{st_2}^*\right] \\
&\equiv \mathcal{Z}_1^*+\mathcal{Z}_2^*+\mathcal{Z}_3^*+\mathcal{Z}_4^*,\ \text{say.}
\end{aligned}
$$

where $\phi_{sr}^*=\tilde{F}_s'H\tilde{\Sigma}_{\tilde{F}}H'\tilde{F}_re_s^{*\prime}e_r^*$. Using the IID property of ς_{it} and the conditions in Theorem 4.5, we can readily verify that $\mathcal{Z}_l^*=o_P(1)$ for $l=1,2,3,4$. For example, noting that $E\left[\varsigma_{i_1s}\varsigma_{i_2s}\varsigma_{i_3s}\varsigma_{i_4s}\right]=3$ if $i_1=i_2=i_3=i_4$, $=1$ if $i_1=i_2\neq i_3=i_4$, $i_1=i_3\neq i_2=i_4$, or $i_1=i_4\neq i_2=i_4$, and zero otherwise, we have for any $s\neq r$,

$$
\begin{aligned}
E^*\left(e_s^{*\prime}e_r^*\right)^4 &= E^*\left(\varsigma_s'\tilde{\Sigma}\varsigma_r\right)^4=\sum_{i_1,\ldots,i_4,j_1,\ldots,j_4}\tilde{\sigma}_{i_1j_1}\tilde{\sigma}_{i_2j_2}\tilde{\sigma}_{i_3j_3}\tilde{\sigma}_{i_4j_4}E\left[\varsigma_{i_1s}\varsigma_{i_2s}\varsigma_{i_3s}\varsigma_{i_4s}\right]E\left[\varsigma_{j_1r}\varsigma_{j_2r}\varsigma_{j_3r}\varsigma_{j_4r}\right] \\
&= 9\sum_{i,j}\tilde{\sigma}_{ij}^4+9\sum_{i}\sum_{j_1\neq j_2}\tilde{\sigma}_{ij_1}^2\tilde{\sigma}_{ij_2}^2+9\sum_{i_1\neq i_2}\sum_{j}\tilde{\sigma}_{i_1j}^2\tilde{\sigma}_{i_2j}^2 \\
&\quad+\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}\left[\tilde{\sigma}_{i_1j_1}\tilde{\sigma}_{i_1j_1}\tilde{\sigma}_{i_2j_2}\tilde{\sigma}_{i_2j_2}+\tilde{\sigma}_{i_1j_1}\tilde{\sigma}_{i_1j_2}\tilde{\sigma}_{i_2j_1}\tilde{\sigma}_{i_2j_2}+\tilde{\sigma}_{i_1j_1}\tilde{\sigma}_{i_1j_2}\tilde{\sigma}_{i_2j_2}\tilde{\sigma}_{i_2j_1}\right. \\
&\quad+\tilde{\sigma}_{i_1j_1}\tilde{\sigma}_{i_2j_1}\tilde{\sigma}_{i_1j_2}\tilde{\sigma}_{i_2j_2}+\tilde{\sigma}_{i_1j_1}\tilde{\sigma}_{i_2j_2}\tilde{\sigma}_{i_1j_1}\tilde{\sigma}_{i_2j_2}+\tilde{\sigma}_{i_1j_1}\tilde{\sigma}_{i_2j_2}\tilde{\sigma}_{i_1j_2}\tilde{\sigma}_{i_2j_1} \\
&\quad\left.+\tilde{\sigma}_{i_1j_1}\tilde{\sigma}_{i_2j_1}\tilde{\sigma}_{i_2j_2}\tilde{\sigma}_{i_1j_2}+\tilde{\sigma}_{i_1j_1}\tilde{\sigma}_{i_2j_2}\tilde{\sigma}_{i_2j_1}\tilde{\sigma}_{i_1j_2}+\tilde{\sigma}_{i_1j_1}\tilde{\sigma}_{i_2j_2}\tilde{\sigma}_{i_2j_2}\tilde{\sigma}_{i_1j_1}\right] \\
&= 9\sum_{i,j}\tilde{\sigma}_{ij}^4+18\sum_{i}\sum_{j_1\neq j_2}\tilde{\sigma}_{ij_1}^2\tilde{\sigma}_{ij_2}^2+3\sum_{i_1\neq i_2}\sum_{j_1\neq j_2}\left[\tilde{\sigma}_{i_1j_1}^2\tilde{\sigma}_{i_2j_2}^2+2\tilde{\sigma}_{i_1j_1}\tilde{\sigma}_{i_1j_2}\tilde{\sigma}_{i_2j_1}\tilde{\sigma}_{i_2j_2}\right] \\
&= O_P\left(\xi_{NT}^3N+N\xi_{NT}^2+N^2\xi_{NT}^2\right)=O_P\left(N^2\xi_{NT}^2\right).
\end{aligned}
$$

Then

$$
\begin{aligned}
\mathcal{Z}_1^* &= \frac{16}{T^4N^2h^2}\sum_{s=2}^{T}\sum_{r=1}^{s-1}\bar{k}_{sr}^4\left(\tilde{F}_s'H\tilde{\Sigma}_{\tilde{F}}H'\tilde{F}_r\right)^4E^*\left(e_s^{*\prime}e_r^*\right)^4 \\
&= O_P\left(N^2\xi_{NT}^2\right)\frac{16}{T^4N^2h^2}\sum_{s=2}^{T}\sum_{r=1}^{s-1}\bar{k}_{sr}^4\left(\tilde{F}_s'H\tilde{\Sigma}_{\tilde{F}}H'\tilde{F}_r\right)^4=O_P\left(\xi_{NT}^2T^{-2}h^{-1}\right)=o_P(1),
\end{aligned}
$$

where we use the fact that $\frac{1}{T^2h}\sum_{s=2}^{T}\sum_{r=1}^{s-1}\bar{k}_{sr}^4\left\|\tilde{F}_s\right\|^8=O_P(1)$ under Assumption A.3 and the extra conditions in the theorem. Similarly, noting that for any $r_1 < r_2 < s$,

$$
\begin{aligned}
E^*\left[\left(e_s^{*\prime}e_{r_1}^*\right)^2\left(e_s^{*\prime}e_{r_2}^*\right)^2\right] &= E^*\left[\left(\varsigma_s'\tilde{\Sigma}\varsigma_{r_1}\varsigma_{r_1}'\tilde{\Sigma}\varsigma_s\right)\left(\varsigma_s'\tilde{\Sigma}\varsigma_{r_2}\varsigma_{r_1}'\tilde{\Sigma}\varsigma_s\right)\right] = E^*\left[\varsigma_s'\tilde{\Sigma}\tilde{\Sigma}\varsigma_s\varsigma_s'\tilde{\Sigma}\tilde{\Sigma}\varsigma_s\right] \\
&= \sum_{i_1,\ldots,i_4,j_1,j_2}\tilde{\sigma}_{i_1j_1}\tilde{\sigma}_{j_1i_2}\tilde{\sigma}_{i_3j_2}\tilde{\sigma}_{j_2i_4}E\left[\varsigma_{i_1s}\varsigma_{i_2s}\varsigma_{i_3s}\varsigma_{i_4s}\right] \\
&= 3\sum_{i,j_1,j_2}\tilde{\sigma}_{ij_1}^2\tilde{\sigma}_{ij_2}^2 + \sum_{i_1,i_2,j_1,j_2}\left[\tilde{\sigma}_{i_1j_1}^2\tilde{\sigma}_{i_2j_2}^2 + 2\tilde{\sigma}_{i_1j_1}\tilde{\sigma}_{j_1i_2}\tilde{\sigma}_{i_1j_2}\tilde{\sigma}_{j_2i_2}\right] \\
&= O_P\left(N\xi_{NT}^3\right) + O_P\left(N^2\xi_{NT}^2\right) = O_P\left(N^2\xi_{NT}^2\right),
\end{aligned}
$$

where we use the fact that $\tilde{\sigma}_{ij} = \tilde{\sigma}_{ji}$ and $\xi_{NT} = o\left(T^{1/2}\right) = o\left(N\right)$, we have

$$
\begin{aligned}
\mathcal{Z}_4^* &= \frac{64}{T^4N^2h^2}\sum_{s=2}^{T}\sum_{1\le r_1<r_2\le s-1,1\le t_1<t_2\le s-1}\bar{k}_{r_1s}\bar{k}_{r_2s}\bar{k}_{t_1s}\bar{k}_{t_2s}\tilde{F}_s'H\tilde{\Sigma}_{\tilde{F}}H'\tilde{F}_{r_1}\tilde{F}_s'H\tilde{\Sigma}_{\tilde{F}}H'\tilde{F}_{r_2}\tilde{F}_s'H\tilde{\Sigma}_{\tilde{F}}H'\tilde{F}_{t_1}\tilde{F}_s' \\
&\quad \times H\tilde{\Sigma}_{\tilde{F}}H'\tilde{F}_{t_2}E^*\left[\left(e_s^{*\prime}e_{r_1}^*\right)\left(e_s^{*\prime}e_{r_2}^*\right)\left(e_s^{*\prime}e_{t_1}^*\right)\left(e_s^{*\prime}e_{t_2}^*\right)\right] \\
&= \frac{64}{T^4N^2h^2}\sum_{s=2}^{T}\sum_{1\le r_1<r_2\le s-1}\bar{k}_{r_1s}^2\bar{k}_{r_2s}^2\left(\tilde{F}_s'H\tilde{\Sigma}_{\tilde{F}}H'\tilde{F}_{r_1}\right)^2\left(\tilde{F}_s'H\tilde{\Sigma}_{\tilde{F}}H'\tilde{F}_{r_2}\right)^2E^*[\left(e_s^{*\prime}e_{r_1}^*\right)^2\left(e_s^{*\prime}e_{r_2}^*\right)^2] \\
&= \frac{64}{T^4N^2h^2}\sum_{s=2}^{T}\sum_{1\le r_1<r_2\le s-1}\bar{k}_{r_1s}^2\bar{k}_{r_2s}^2\left(\tilde{F}_s'H\tilde{\Sigma}_{\tilde{F}}H'\tilde{F}_{r_1}\right)^2\left(\tilde{F}_s'H\tilde{\Sigma}_{\tilde{F}}H'\tilde{F}_{r_2}\right)^2O_P\left(N^2\xi_{NT}^2\right) \\
&= O_P\left(\xi_{NT}^2T^{-1}\right) = o_P\left(1\right).
\end{aligned}
$$

Then $\mathcal{Z}^* = o_{P^*}\left(1\right)$ by the conditional Markov inequality. Now $\sum_{t=2}^{T}E^*(Z_{NT,t}^{*2}) = 4T^{-2}N^{-1}h^{-1}E^*[\sum_{r=1}^{s-1}\bar{k}_{sr}\tilde{F}_s'H\tilde{\Sigma}_{\tilde{F}}H'\tilde{F}_{r_1}e_s^{*\prime}e_r^*]^2 = \mathbb{V}_{NT}^*$. Straightforward moment calculations yield that $E^*(\sum_{t=2}^{T}Z_{NT,t}^{*2})^2 = \mathbb{V}_{NT}^{*2} + o_P\left(1\right)$. Thus $\text{Var}^*(\sum_{t=2}^{T}Z_{NT,t}^{*2}) = o_P\left(1\right)$ and $\sum_{t=2}^{T}Z_{NT,t}^{*2} - \mathbb{V}_{NT}^* = o_{P^*}(1)$. This completes the proof of (i). ■

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.

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., 2014. Shrinkage estimation of high-dimensional factor models with structural instabilities. Working Paper, University of Pennsylvania.

Han, X., Inoue A., 2014. Tests for parameter instability in dynamic factor models. Econometric Theory, forthcoming.

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.

Moon, H., Weidner, M., 2014. Linear regression for panel with unknown number of factors as interactive fixed effects. Working paper, University of Southern California.

Onatski, A., 2009. Testing hypotheses about the number of factors in large factor models. Econometrica 77, 1447-1479.

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.

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, Oxford University Press.

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., 2015. 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.