

TERG

Discussion Paper No.306

**Tests for Parameter Instability
in Dynamic Factor Models**

Xu Han
Atsushi Inoue

First Version: January 2011
This Version: May 2013

TOHOKU ECONOMICS RESEARCH GROUP

GRADUATE SCHOOL OF ECONOMICS AND
MANAGEMENT TOHOKU UNIVERSITY
27-1 KAWAUCHI, AOBA-KU, SENDAI,
980-8576 JAPAN

Tests for Parameter Instability in Dynamic Factor Models *

Xu Han[†]

Atsushi Inoue[‡]

City University of Hong Kong North Carolina State University

Tohoku University

First Version: January 2011

This Version: May 2013

Abstract

We develop tests for structural breaks of factor loadings in dynamic factor models. We focus on the joint null hypothesis that all factor loadings are constant over time. Because the number of factor loading parameters goes to infinity as the sample size grows, conventional tests cannot be used. Based on the fact that the presence of a structural change in factor loadings yields a structural change in second moments of factors obtained from the full sample principal component estimation, we reduce the infinite-dimensional problem into a finite-dimensional one and our statistic compares the pre- and post-break subsample second moments of estimated factors. Our test is consistent under the alternative hypothesis in which a fraction of or all factor loadings have structural changes. The Monte Carlo results show that our test has good finite-sample size and power.

Keywords: factor models, structural break

JEL classification: C12, C33

*We thank Zhongjun Qu for helpful conversation, anonymous referees for insightful suggestions, and participants at the 2011 North American Summer Meeting of the Econometric Society, the fifth CIREQ Time Series Conference, and the 2011 NBER-NSF Time Series Conference for helpful comments. We thank Jesus Gonzalo for pointing out an error in an earlier version of the paper.

[†]Department of Economics and Finance, City University of Hong Kong, Kowloon, Kong Kong SAR. E-mail: xuhan25@cityu.edu.hk.

[‡]Department of Agricultural and Resource Economics, North Carolina State University, Raleigh, NC 27695-8109, and Faculty of Economics, Tohoku University, Sendai, Japan. E-mail: atsushi@ncsu.edu.

1 Introduction

Dynamic factor models have become popular in the recent macroeconometrics literature because a few factors can often explain a substantial amount of variations of many macroeconomic time series (Sargent and Sims, 1977). For example, they have been successfully used in forecasting (Stock and Watson, 2002a), factor augmented vector autoregressive (FAVAR) models (Bernanke, Boivin and Elias, 2005; Stock and Watson, 2005) and DSGE models (Boivin and Giannoni, 2006). While most of these applications implicitly assume that the factor loadings in dynamic factor models are time-invariant, there is strong evidence of structural instability in macroeconomic time series (Stock and Watson, 1996). If the common factors are driven by some structural shocks, it is possible that macroeconomic variables react to these structural shocks differently during different sample periods, resulting in time-varying factor loadings. For example, Eickmeier, Lemke and Marcellino (2011) consider time-varying FAVAR models to take into account changes in the monetary transmission mechanism. If parameter instability is ignored, the dynamic factor models may perform poorly or give misleading results. For example, Banerjee and Marcellino (2008) provide simulation evidence that the performance of forecasts based on dynamic factor models will be significantly worse off if the structural breaks in factor loadings are not taken into account.

While the estimated factors still consistently span the original factor space if the size of break is small enough (Stock and Watson, 2002b; Bates et al., 2012), such results do not hold when the size of breaks is large. Large break can augment the factor space, but simply introducing more factors cannot solve all the problems for two reasons. First, when structural break leads to an augmented factor space, the factor dynamics is changed as well. Stock and Watson (2009) argue that one should take into account such change in the forecasting regression if estimated factors are used as predictors. Second, the augmented factor space does not contain more information than the original factor space, so introducing more factors can reduce the efficiency and worsen the forecast performance. Hence, it is essential to know the existence of structural breaks in factor loadings and break dates for forecasting based on dynamic factor models.

In this paper, we consider testing the joint null hypothesis that factor loadings are constant

over time against the alternative that a non-negligible fraction of or all factor loadings are not. We are interested in the joint null hypothesis rather than the null hypothesis that a specific individual factor loading is constant over time because it is the joint null hypothesis under which one can estimate the factors consistently. Conventional tests of structural change, such as Andrews (1993), are designed to deal with finitely many parameters and cannot be used to test our null hypothesis that involves parameters whose number goes to infinity as the sample size grows. Directly extending the conventional test to our setup is challenging for two reasons. First, one needs to estimate an infinite dimensional covariance matrix and its inverse. This brings several technical difficulties: (1) the norm of the difference between the estimated and true covariance matrices can be very large even if each entry of the estimated matrix converges in probability; (2) taking the inverse of a high dimensional matrix will amplify the estimation error dramatically and lead to very poor results (Ledoit and Wolf, 2004); and (3) the number of time periods, T , can be even smaller than the dimension of the estimated covariance matrix, so the sample covariance matrix can be singular. Second, because the degree of freedom also goes to infinity, the limit distribution of such test statistics, even if it is well-defined, is likely to be nonstandard.

To the best of our knowledge there are three existing tests for structural instability of factor loadings.¹ One is proposed by Stock and Watson (2009) (henceforth SW), who regressed each variable on the estimated factors and implemented a Chow test for each of these regressions. Using a post-war quarterly data set for the United States, they found a substantial amount of instability in factor loadings: 41% (23%) of these Chow tests reject at the 5% (1%) significance level. This method cannot control the overall type I error for testing our joint null hypothesis and it may overstate the parameter instability in factor loadings. Another test is proposed by Breitung and Eickmeier (2011) (henceforth BE). They constructed a joint test that controls the overall type I error as well as tests for individual factor loadings allowing for an unknown break date. To test the joint null hypothesis they use the sample average of the Chow test statistics for each of the factor loadings. They require the idiosyncratic shocks to be cross-sectionally independent, however. This is more restrictive than Bai and Ng's (2002) setup for approximate factor models where idiosyncratic shocks are allowed to

¹Recently, Cheng et al. (2013) consider detecting structural changes of factor loadings using LASSO estimator.

have weak cross-sectional correlation. Also, their joint test is severely oversized in the presence of serial correlations in the idiosyncratic shocks when the HAC covariance matrix estimator is used.² More recently, Chen, Dolado and Gonzalo (2012) (CDG hereafter) develop Wald and LM tests of structural change in factor loadings using a regression of the first estimated factor on the remaining factors. Their tests are simple to implement in practice.

This paper proposes new joint tests and contributes in the following ways: First, we reduce the infinite-dimensional problem to a finite-dimensional one. Because the principal component analysis (PCA) implicitly imposes the restriction that the factor loadings are time-invariant, the estimated factors can demonstrate a higher dimension under the alternative hypothesis than under the null as pointed out by SW and BE. Based on this fact, our statistic compares the pre- and post-break subsample second moments of estimated factors. We allow for unknown break dates and our statistic has the same asymptotic distribution as the conventional supreme Wald test proposed by Andrews (1993). Second, we follow Bai and Ng's (2002) approximate factor setup in which serial, cross-sectional correlation and heteroskedasticity are allowed in the idiosyncratic shocks, and the knowledge about the form of such correlations and heteroskedasticity is not required to implement our test. This is more general than BE's framework which requires cross-sectional independence and AR(p) assumption on the idiosyncratic shocks. Third, we consider different types of structural breaks which have different impacts on the dimension of factor space. Besides the case pointed out by SW and BE where the number of factors is enlarged due to structural breaks in factor loadings, we also consider the case where structural change in the loading matrix does not change the number of factors. We establish conditions under which our tests are consistent against different types of breaks. Fourth, the number of factors are unknown and to be determined by the information criteria (IC) proposed by Bai and Ng (2002). We show that Bai and Ng's IC remain consistent for estimating

²In the working paper version of their paper, BE provide simulation evidence that the effective size of their test can be greater than 90% in some setups if HAC estimators are used. This may be explained by possible invalidity of the sequential asymptotics their test is based on. When the HAC estimator is used, each individual statistic converges to a chi-square distribution at a slower rate as the time series dimension diverges. When the cross-sectional average of these statistics are taken, the cross-sectional average of these errors may not vanish even asymptotically. BE also suggest fitting an AR(p) model for the residuals of factor models in the first stage and then compute the covariance using the filtered residuals, but this operation requires correct knowledge about the serial correlation structure of the residuals.

the dimension of an equivalent factor model with time-invariant loadings under the one-time break alternative. This helps our test statistic distinguish the null and alternative hypotheses. Finally, we establish the regularity conditions on N , T and bandwidth parameters for different kernel functions so that HAC estimators are applicable to our test statistics.

The test of CDG is also based on the fact that the dimension of factor space estimated by the information criterion is larger under the alternative hypothesis than under the null hypothesis. Testing structural change in their regression model is equivalent to testing structural change in a subset of the elements of the covariance matrix of the estimated factors. In contrast, our test statistics make use of all the elements of the covariance matrix. Also, our tests allow for the case where the break in factor loadings does not enlarge the number of factors, which is ruled out in CDG's framework. We compare the finite sample performance of their tests and ours in a Monte Carlo experiment in section 3.

In this paper, all limits are taken as both $N, T \rightarrow \infty$ simultaneously. $\|\cdot\|$ denotes the Euclidean norm of a vector or matrix, \xrightarrow{p} denotes convergence in probability, \xrightarrow{d} denotes convergence in distribution, \Rightarrow denotes weak convergence of stochastic processes, and $[\cdot]$ is the integer part operator. $\text{vech}(\cdot)$ is equal to the column-wise vectorization of a square matrix with the upper triangular excluded.

The remainder of this paper is organized as follows: Section 2 proposes a structural break test for factor loadings, and the asymptotic properties are established under the null and alternative hypotheses. Section 3 shows Monte Carlo results under various data generating processes (DGPs). Section 4 concludes. Proofs are relegated to appendices A and B.

2 A Structural Break Test in Factor Loadings

2.1 Factor Models and the Null Hypothesis of Interest

Let x_{it} denote the observation for the i th cross section at period t for $i = 1, 2, \dots, N$ and $t = 1, 2, \dots, T$. Let $[\pi^*T] + 1$ denote the break date of factor loadings and $\pi^* \in (0, 1)$. Suppose that x_{it} has r

common factors and follows the static factor representation:

$$x_{it} = \begin{cases} f'_{0,t}\lambda_{0,i} + f'_{1,t}\lambda_{1,i} + e_{it} & \text{if } 1 \leq t \leq \lfloor \pi^*T \rfloor \\ f'_{0,t}\lambda_{0,i} + f'_{1,t}\lambda_{2,i} + e_{it} & \text{if } \lfloor \pi^*T \rfloor + 1 \leq t \leq T \end{cases} \quad (2.1)$$

where $f_{0,t}$ is a $q_0 \times 1$ vector that denotes the factors with time-invariant factor loadings, $f_{1,t}$ is a $q_1 \times 1$ dimensional factor whose loadings have structural change at $\lfloor \pi^*T \rfloor + 1$, $\lambda_{0,i}$ is the factor loading on $f_{0,t}$, $\lambda_{1,i}$ and $\lambda_{2,i}$ are the pre- and post-break factor loadings on $f_{1,t}$, respectively, and e_{it} is the idiosyncratic shock for cross section i at period t . Let $f_t \equiv (f'_{0,t}, f'_{1,t})'$ denote the $r \times 1$ vector of common factors at period t , so we have $r = q_0 + q_1$. Define the vectors: $X_i \equiv (x_{i1}, x_{i2}, \dots, x_{iT})'$ and $e_i \equiv (e_{i1}, e_{i2}, \dots, e_{iT})'$. The matrix notation of the factor model is:

$$X = \begin{bmatrix} F_{0,1}\Lambda'_0 + F_{1,1}\Lambda'_1 \\ F_{0,2}\Lambda'_0 + F_{1,2}\Lambda'_2 \end{bmatrix} + e \quad (2.2)$$

where $X \equiv (X_1, X_2, \dots, X_N)$, $F_{0,1} \equiv (f_{0,1}, f_{0,2}, \dots, f_{0, \lfloor \pi^*T \rfloor})'$, $F_{0,2} \equiv (f_{0, \lfloor \pi^*T \rfloor + 1}, f_{0, \lfloor \pi^*T \rfloor + 2}, \dots, f_{0,T})'$, $F_{1,1} \equiv (f_{1,1}, f_{1,2}, \dots, f_{1, \lfloor \pi^*T \rfloor})'$, $F_{1,2} \equiv (f_{1, \lfloor \pi^*T \rfloor + 1}, f_{1, \lfloor \pi^*T \rfloor + 2}, \dots, f_{1,T})'$, $\Lambda_0 \equiv (\lambda_{0,1}, \lambda_{0,2}, \dots, \lambda_{0,N})'$, $\Lambda_1 \equiv (\lambda_{1,1}, \lambda_{1,2}, \dots, \lambda_{1,N})'$, $\Lambda_2 \equiv (\lambda_{2,1}, \lambda_{2,2}, \dots, \lambda_{2,N})'$, and $e \equiv (e_1, e_2, \dots, e_N)$.

We are interested in testing the null hypothesis of no break in factor loadings:

$$H_0 : \lambda_{1,i} = \lambda_{2,i} \quad \forall i \quad (2.3)$$

The test of Breitung and Eickmeier (2011) is designed to test the null hypothesis

$$H_{BE,0} : \lambda_{1,i} = \lambda_{2,i} \quad \text{for given } i \quad (2.4)$$

While their test is useful when one is interested in a specific factor loading, it is (2.3) under which factors are consistently estimated. If their test for (2.4) is applied to test (2.3), the null hypothesis (2.3) will be rejected with probability approaching one because the test is applied to factor loadings whose number goes to infinity. Breitung and Eickmeier (2011) also suggest a pooled LM test for

testing (2.3) but their pooled test requires that the idiosyncratic shocks e_{it} and e_{jt} are independent for all $i \neq j$, which is too restrictive compared with the approximate factor structure. Below we propose a test for the null hypothesis (2.3) that is valid under assumptions allowing cross-sectional correlated idiosyncratic errors.

2.2 The Test Statistic

We consider testing the null hypothesis H_0 : all factor loadings are constant over time against the alternative hypothesis H_1 : $\lfloor \alpha N \rfloor$ many variables have structural changes in factor loadings at a common break date, where $\alpha \in (0, 1]$. To motivate our test statistic, let us first consider the behavior of the model under the null hypothesis. Since $\Lambda_1 = \Lambda_2$, (2.2) can be rewritten in the following form:

$$\begin{aligned} X &= \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \Lambda' + e \\ &= F \Lambda' + e \end{aligned} \tag{2.5}$$

where $F_1 \equiv [F_{0,1}:F_{1,1}]$, $F_2 \equiv [F_{0,2}:F_{1,2}]$, $F \equiv [F_1':F_2']'$, and $\Lambda \equiv [\Lambda_0:\Lambda_1]$. If the fourth moment of f_t is time-invariant and some regularity conditions hold, a Wald statistic comparing the subsample means of $f_t f_t'$ should converge to a chi-square distributed random variable under the null hypothesis. Under the alternative hypothesis, (2.2) has time-varying factor loadings, but PCA implicitly imposes the restriction that factor loadings are constant over time. To see why the above Wald statistic has power, consider a simple example where all factors have time-varying loadings, i.e., Λ_1 and Λ_2 are both $N \times r$ and $q_0 = 0$. The PCA will estimate an equivalent factor model with time-invariant factor loadings:

$$X = \begin{bmatrix} F_{1,1} & 0_{\lfloor \pi^* T \rfloor \times r} \\ 0_{(T - \lfloor \pi^* T \rfloor) \times r} & F_{1,2} \end{bmatrix} \begin{bmatrix} \Lambda'_1 \\ \Lambda'_2 \end{bmatrix} + e$$

Let $G_1 \equiv [F_{1,1}; 0_{[\pi^*T] \times r}]$ and $G_2 \equiv [0_{(T-[\pi^*T]) \times r}; F_{1,2}]$, so it follows that

$$\frac{1}{\pi^*T} G_1' G_1 = \begin{bmatrix} \frac{1}{\pi^*T} F_{1,1}' F_{1,1} & 0_{r \times r} \\ 0_{r \times r} & 0_{r \times r} \end{bmatrix}$$

and

$$\frac{1}{(1-\pi^*)T} G_2' G_2 = \begin{bmatrix} 0_{r \times r} & 0_{r \times r} \\ 0_{r \times r} & \frac{1}{(1-\pi^*)T} F_{1,2}' F_{1,2} \end{bmatrix}$$

have different limits, so the Wald statistic that compares the second moments of G_1 and G_2 will reject the null hypothesis under the alternative. This example shows that the presence of a structural break in factor loadings implies structural change in the subsample second moments of factors. Using this fact, we can reduce the infinite-dimensional problem to a finite-dimensional one.

Therefore, we base the proposed test statistic on the pre- and post-break subsample means of $\hat{f}_t \hat{f}_t'$, where \hat{f}_t is the PCA estimate of the factors. Let

$$A(\pi, \hat{F}) \equiv \text{vech} \left(\sqrt{T} \left(\frac{1}{[\pi T]} \sum_{t=1}^{[\pi T]} \hat{f}_t \hat{f}_t' - \frac{1}{T - [\pi T]} \sum_{t=[\pi T]+1}^T \hat{f}_t \hat{f}_t' \right) \right).$$

$\hat{S}(\pi, \hat{F})$ and $\tilde{S}(\pi, \hat{F})$ denote unrestricted and restricted estimates of the long-run covariance matrix of $A(\pi, \hat{F})$, respectively, and will be defined more precisely in the next subsection. We define two test statistics by

$$\sup_{\pi \in [\pi_1, \pi_2]} W_T(\pi, \hat{F}) \equiv \sup_{\pi \in [\pi_1, \pi_2]} A(\pi, \hat{F})' \hat{S}(\pi, \hat{F})^{-1} A(\pi, \hat{F}), \quad (2.6)$$

$$\sup_{\pi \in [\pi_1, \pi_2]} LM_T(\pi, \hat{F}) \equiv \sup_{\pi \in [\pi_1, \pi_2]} A(\pi, \hat{F})' \tilde{S}(\pi, \hat{F})^{-1} A(\pi, \hat{F}), \quad (2.7)$$

where $W_T(\pi, \hat{F})$ and $LM_T(\pi, \hat{F})$ are Wald and LM-like statistics for testing whether the subsample means of $\hat{f}_t \hat{f}_t'$ are equal or not at a predetermined break date πT . The following subsections will discuss the assumptions and detailed properties of the proposed test statistics.

2.3 Assumptions

Recall that $\Lambda \equiv [\Lambda_0; \Lambda_1]$ in (2.5). Let λ_i denote the transpose of the i^{th} row of Λ . Let ι_{1t} and ι_{2t} denote two indicator functions: $\iota_{1t} \equiv \mathbf{1}\{t \leq \lfloor \pi^* T \rfloor\}$ and $\iota_{2t} \equiv \mathbf{1}\{t \geq \lfloor \pi^* T \rfloor + 1\}$.

Assumption 1: $E \|f_t\|^4 < \infty$, $E(f_t f_t') = \Sigma_F$ and $T^{-1} \sum_{t=1}^T f_t f_t' \xrightarrow{p} \Sigma_F$ as $T \rightarrow \infty$ for some positive definite matrix Σ_F .

Assumption 2: $\|\lambda_i\| \leq \bar{\lambda} < \infty$, $\|\Lambda' \Lambda / N - \Sigma_\Lambda\| \rightarrow 0$ for some $r \times r$ positive definite matrix Σ_Λ , and $\|\Lambda' \Lambda / N - \Sigma_\Lambda\| = O\left(\frac{1}{\sqrt{N}}\right)$.

Assumption 3: There exists a positive constant $M < \infty$ such that for all N and T ,

(a) $E(e_{it}) = 0$, $E|e_{it}|^8 \leq M$ for all i and t .

(b) $E(e'_s e_t / N) = E(N^{-1} \sum_{i=1}^N e_{is} e_{it}) = \gamma_N(s, t)$, $|\gamma_N(s, s)| \leq M$ for all s , and $T^{-1} \sum_{s=1}^T \sum_{t=1}^T |\gamma_N(s, t)| \leq M$.

(c) $E(e_{it} e_{jt}) = \tau_{ij,t}$ with $|\tau_{ij,t}| \leq |\tau_{ij}|$ for some τ_{ij} and for all t . In addition, $N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\tau_{ij}| \leq M$.

(d) $E(e_{it} e_{js}) = \tau_{ij,ts}$, and $(NT)^{-1} \sum_{i=1}^N \sum_{j=1}^N \sum_{s=1}^T \sum_{t=1}^T |\tau_{ij,ts}| \leq M$.

(e) for every (t, s) , $E \left| N^{-1/2} \sum_{i=1}^N [e_{is} e_{it} - E(e_{is} e_{it})] \right|^4 \leq M$.

Assumption 4: $E \left(\frac{1}{N} \sum_{i=1}^N \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T f_t e_{it} \cdot \iota_{mt} \right\|^2 \right) \leq M$ for $m = 1, 2$.

Assumption 5: There exists $M < \infty$ such that for all T and N , and for every $t \leq T$ and for every $i \leq N$:

(a) $\sum_{s=1}^T |\gamma_N(s, t)| \leq M$.

(b) $\sum_{k=1}^N |\tau_{ki}| \leq M$.

Assumption 6: There exists an $M < \infty$ such that for all N and T :

(a) for each t and $m = 1, 2$, $E \left\| \frac{1}{\sqrt{NT}} \sum_{s=1}^T \sum_{k=1}^N f_s [e_{ks} e_{kt} - E(e_{ks} e_{kt})] \cdot \iota_{ms} \right\|^2 \leq M$.

(b) $E \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^T \sum_{k=1}^N f_t \lambda'_{\ell,k} e_{kt} \cdot \iota_{mt} \right\|^2 \leq M$ for $m = 1, 2$ and $\ell = 0, 1, 2$.

(c) for each t and $\ell = 0, 1, 2$, $E \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_{\ell,i} e_{it} \right\|^4 \leq M$.

Assumption 7: The eigenvalues of $r \times r$ matrix $(\Sigma_\Lambda \Sigma_F)$ are distinct.

Assumption 8: For any constants π_1 and π_2 that satisfy $0 < \pi_1 \leq \pi^* \leq \pi_2 < 1$,

(a)

$$\begin{aligned} \sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{k=1}^N f_t \lambda'_{\ell,k} e_{kt} \cdot \iota_{mt} \right\|^2 &= O_p(1) \\ \sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{\sqrt{NT}} \sum_{t=\lfloor \pi T \rfloor + 1}^T \sum_{k=1}^N f_t \lambda'_{\ell,k} e_{kt} \cdot \iota_{mt} \right\|^2 &= O_p(1) \end{aligned}$$

for $m = 1, 2$ and $\ell = 0, 1, 2$.

(b) $\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{\sqrt{T}}{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} (f_t f'_t - \Sigma_F) \right\| = O_p(1)$ and $\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{\sqrt{T}}{T - \lfloor \pi T \rfloor} \sum_{t=\lfloor \pi T \rfloor + 1}^T (f_t f'_t - \Sigma_F) \right\| = O_p(1)$.

These assumptions are either from or slight modifications of those in Bai (2003) and enable us to conduct inference about subsample means of $\hat{f}_t \hat{f}'_t$. Assumption 1 is the same as Assumption A in Bai (2003) except that it requires time-invariant second moment of f_t . This assumption is made under both the null and alternative hypotheses. Note that factors and factor loadings are multiplicative and identified under some normalization, a factor model with a break in $E(f_t f'_t)$ and no break in factor loadings is observationally equivalent to a factor model with time-invariant $E(f_t f'_t)$ but a rotation in the post-break factor loading matrix. More details about the rotation in the loading matrix are discussed in Section 2.5 (See footnote 5). Given this identification issue, it is not restrictive to assume that factors have constant second moment.

The Assumption 2 is slightly different from Assumption B of Bai (2003) in that it specifies the convergence speed of $\Lambda' \Lambda / N$.³ Assumptions 3, 5 and 7 exactly follow from Bai's (2003) setup. Assumption 3 allows weak serial and cross-sectional dependence in the idiosyncratic shocks, and As-

³Since the factor loadings are assumed to be non-random, $\Lambda' \Lambda / N$ can converge to Σ_Λ at any rate. We assume that the rate is no slower than $1/\sqrt{N}$, which is not stringent compared to usual convergence rate in the Central Limit Theorem.

sumption 5 is a strengthened version of Assumption 3. Assumptions 3 and 5 also allow heterogeneity in time and cross section dimensions. Thus, this paper allows weaker assumptions on idiosyncratic shocks than BE who assume that the idiosyncratic shocks are independent in cross section dimension and follow AR(p) processes. Assumption 4 implies that $E(N^{-1} \sum_{i=1}^N \|T^{-1/2} \sum_{t=1}^T f_t e_{it}\|^2) \leq 4M$. Parts (a) and (b) of Assumption 6 imply that (a) for each t , $E\|(NT)^{-1/2} \sum_{s=1}^T \sum_{k=1}^N f_s [e_{ks} e_{kt} - E(e_{ks} e_{kt})]\|^2 \leq 4M$; and (b) $E\|(NT)^{-1/2} \sum_{t=1}^T \sum_{k=1}^N f_t \lambda'_k e_{kt}\|^2 \leq 8M$, which are Assumptions F1 and F2 of Bai (2003). The role of the indicators functions ι_{1t} and ι_{2t} will become clear under the alternative hypothesis.⁴ Assumption 6(c) is slightly stronger than Assumption F3 of Bai (2003) which only requires the existence of the second moment, but the asymptotic normal distribution of $\frac{1}{\sqrt{N}} \sum_{i=1}^N \lambda_i e_{it}$ in Bai (2003) is not necessary in this paper. Assumption 8 requires that the sample sizes before and after the hypothesized break date go to infinity. Assumption 8 also states that the terms in $\|\cdot\|$ are $O_p(1)$ uniformly in π . Hence, model (2.5) satisfies that $\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{k=1}^N f_t \lambda'_k e_{kt} \right\|^2 = O_p(1)$ and $\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{\sqrt{NT}} \sum_{t=\lfloor \pi T \rfloor + 1}^T \sum_{k=1}^N f_t \lambda'_k e_{kt} \right\|^2 = O_p(1)$ under the null hypothesis. Note that all summands have zero means, so Assumption 8 is an implication of the conventional functional central limit theorem.

2.4 Asymptotics under the Null Hypothesis

Before discussing the properties of our test statistic, it is useful to describe some useful notations and existing results. Let V_{NT} be the $r \times r$ diagonal matrix of the first r largest eigenvalues of $(1/TN)XX'$ in decreasing order. Lemma A3 of Bai (2003) shows that V_{NT} converges to V in probability, where V is the diagonal matrix consisting of the eigenvalues of $\Sigma_{\Lambda}^{\frac{1}{2}} \Sigma_F \Sigma_{\Lambda}^{\frac{1}{2}}$ in descending order. Let Υ denote $\Sigma_{\Lambda}^{\frac{1}{2}} \Sigma_F \Sigma_{\Lambda}^{\frac{1}{2}}$'s eigenvectors that corresponds to V such that $\Upsilon' \Upsilon = I_r$. Recall that the estimated factor matrix \hat{F} is \sqrt{T} times eigenvectors corresponding to the r largest eigenvalues of XX' . Let $H \equiv (\Lambda' \Lambda / N)(F' \hat{F} / T) V_{NT}^{-1}$ be an $r \times r$ matrix. Proposition 1 of Bai (2003) show that $F' \hat{F} / T$ converges to $\Sigma_{\Lambda}^{-\frac{1}{2}} \Upsilon V^{\frac{1}{2}}$. Thus, it follows that $H \xrightarrow{p} \Sigma_{\Lambda}^{\frac{1}{2}} \Upsilon V^{-\frac{1}{2}}$. Let $H_0 \equiv \text{plim}_{T, N \rightarrow \infty} H$, so it is obvious that

$$E(H'_0 f_t f'_t H_0) = H'_0 \Sigma_F H_0 = I_r \quad (2.8)$$

⁴See the proof of Lemma 10 in the appendix.

which is implied by the definition of H_0 and the fact that $V^{-\frac{1}{2}}\Upsilon'\Sigma_{\Lambda}^{\frac{1}{2}}\Sigma_F\Sigma_{\Lambda}^{\frac{1}{2}}\Upsilon V^{-\frac{1}{2}} = V^{-\frac{1}{2}}VV^{-\frac{1}{2}} = I_r$.

Equation (2.8) provides a bridge connecting the statistics using estimated factors and true factors. Let $A(\pi, FH_0) \equiv \text{vech}\left(\sqrt{T}\left(\frac{1}{\lfloor\pi T\rfloor}\sum_{t=1}^{\lfloor\pi T\rfloor}H_0'f_t f_t'H_0 - \frac{1}{T-\lfloor\pi T\rfloor}\sum_{t=\lfloor\pi T\rfloor+1}^T H_0'f_t f_t'H_0\right)\right)$. Under Assumption 1 that $E(f_t f_t') = \Sigma_F$, the central limit theorem implies that $A(\pi, FH_0)$ converges in distribution to some normally distributed random variable, and Wald statistics can be constructed based on $A(\pi, FH_0)$ and its sample variance. Although both F and H_0 are not observable, \hat{F} is a consistent estimate of FH (see Bai, 2003) and $H \xrightarrow{p} H_0$, so replacing $H_0'f_t$ by \hat{f}_t is a potential solution.

Theorem 1: Under Assumptions 1 - 8, if $\frac{\sqrt{T}}{N} \rightarrow 0$ as $N, T \rightarrow \infty$, then

$$\sup_{\pi \in [\pi_1, \pi_2]} \left\| A(\pi, \hat{F}) - A(\pi, FH_0) \right\| \xrightarrow{p} 0$$

Theorem 1 shows that the difference between $A(\pi, \hat{F})$ and $A(\pi, FH_0)$ is $o_p(1)$ uniformly in π , so $A(\pi, \hat{F})$ and $A(\pi, FH_0)$ will have the same asymptotic distribution. To construct a Wald statistic, we also need the sample variance of $A(\pi, FH_0)$. Let

$$\Omega \equiv \lim_{T \rightarrow \infty} \text{Var} \left(\text{vech} \left(\frac{1}{\sqrt{T}} \left(\sum_{t=1}^T H_0'f_t f_t'H_0 - I_r \right) \right) \right).$$

Let $\hat{\Omega}_1(\pi, FH_0)$ and $\hat{\Omega}_2(\pi, FH_0)$ be consistent estimates of Ω , where the subscripts “1” and “2” denote the pre- and post-break subsamples, respectively, π denotes the break date that splits the sample, and FH_0 means that the sample variance is computed using unobserved $f_t'H_0$. Since the common factors f_t are likely to be serially correlated, we consider the following estimates for the sample variances:

$$\begin{aligned} \hat{\Omega}_1(\pi, FH_0) &= \hat{\Gamma}_{1,0}(\pi, FH_0) + \sum_{j=1}^{\lfloor\pi T\rfloor-1} k \left(\frac{j}{S_{\lfloor\pi T\rfloor}} \right) (\hat{\Gamma}_{1,j}(\pi, FH_0) + \hat{\Gamma}_{1,j}(\pi, FH_0)') \\ \hat{\Omega}_2(\pi, FH_0) &= \hat{\Gamma}_{2,0}(\pi, FH_0) + \sum_{j=1}^{T-\lfloor\pi T\rfloor-1} k \left(\frac{j}{S_{T-\lfloor\pi T\rfloor}} \right) (\hat{\Gamma}_{2,j}(\pi, FH_0) + \hat{\Gamma}_{2,j}(\pi, FH_0)') \end{aligned} \quad (2.9)$$

where $k(\cdot)$ is a real-valued kernel

$$\begin{aligned}\hat{\Gamma}_{1,j}(\pi, FH_0) &= \frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \text{vech}(H'_0 f_t f'_t H_0 - I_r) \text{vech}(H'_0 f_{t-j} f'_{t-j} H_0 - I_r)' \\ \hat{\Gamma}_{2,j}(\pi, FH_0) &= \frac{1}{T - \lfloor \pi T \rfloor} \sum_{t=j+\lfloor \pi T \rfloor+1}^T \text{vech}(H'_0 f_t f'_t H_0 - I_r) \text{vech}(H'_0 f_{t-j} f'_{t-j} H_0 - I_r)'\end{aligned}\quad (2.10)$$

Alternatively, we can use all data to estimate $\hat{\Omega}_1(\pi, FH_0)$ and $\hat{\Omega}_2(\pi, FH_0)$,

$$\hat{\Omega}(FH_0) = \hat{\Gamma}_0(FH_0) + \sum_{j=1}^{T-1} k\left(\frac{j}{S_T}\right) (\hat{\Gamma}_j(FH_0) + \hat{\Gamma}_j(FH_0)') \quad (2.11)$$

where

$$\hat{\Gamma}_j(FH_0) = \frac{1}{T} \sum_{t=j+1}^T \text{vech}(H'_0 f_t f'_t H_0 - I_r) \text{vech}(H'_0 f_{t-j} f'_{t-j} H_0 - I_r)' \quad (2.12)$$

In this paper, we focus our analysis on three commonly used kernels that always give positive definite estimates: Bartlett, Parzen and Quadratic Spectral (henceforth QS). S is a band-width parameter, and its subscript denotes the size of the sample (or subsample) that is used to estimate the long-run variance. Let

$$\hat{S}(\pi, FH_0) \equiv \frac{1}{\pi} \hat{\Omega}_1(\pi, FH_0) + \frac{1}{1-\pi} \hat{\Omega}_2(\pi, FH_0),$$

so $\hat{S}(\pi, FH_0)$ is an estimate of the asymptotic variance of $A(\pi, FH_0)$. One can also construct the restricted estimator $\tilde{S}(\pi, FH_0)$ using $\hat{\Omega}(FH_0)$, i.e.

$$\tilde{S}(\pi, FH_0) \equiv \left(\frac{1}{\pi} + \frac{1}{1-\pi} \right) \hat{\Omega}(FH_0).$$

Note that all of $\hat{\Omega}_m(\pi, FH_0)$, $\hat{\Gamma}_{m,j}(\pi, FH_0)$, for $m = 1, 2$, $\hat{\Omega}(FH_0)$, $\hat{\Gamma}_j(FH_0)$, $\hat{S}(\pi, FH_0)$ and $\tilde{S}(\pi, FH_0)$ are computed using infeasible data $f'_t H_0$. We define $\hat{\Omega}_m(\pi, \hat{F})$, $\hat{\Gamma}_{m,j}(\pi, \hat{F})$, for $m = 1, 2$, $\hat{\Omega}(\hat{F})$, $\hat{\Gamma}_j(\hat{F})$, $\hat{S}(\pi, \hat{F})$ and $\tilde{S}(\pi, \hat{F})$ as the feasible analogs computed using the estimated regressors

\hat{F} .

Condition 1: (a) The Bartlett kernel is used to estimate $\hat{S}(\pi, FH_0)$, $\hat{S}(\pi, \hat{F})$, $\tilde{S}(\pi, FH_0)$ and $\tilde{S}(\pi, \hat{F})$, and there exists a constant $K > 0$ such that S_T , $S_{\lfloor \pi T \rfloor}$, and $S_{T - \lfloor \pi T \rfloor}$ are less than $KT^{\frac{1}{3}}$ for all $\pi \in [\pi_1, \pi_2] \subset (0, 1)$; and (b) $\frac{T^{\frac{2}{3}}}{N} \rightarrow 0$ as $N, T \rightarrow \infty$.

Condition 2: (a) The Parzen kernel is used to estimate $\hat{S}(\pi, FH_0)$, $\hat{S}(\pi, \hat{F})$, $\tilde{S}(\pi, FH_0)$ and $\tilde{S}(\pi, \hat{F})$, and there exists a constant $K > 0$ such that S_T , $S_{\lfloor \pi T \rfloor}$, and $S_{T - \lfloor \pi T \rfloor}$ are less than $KT^{\frac{1}{5}}$ for all $\pi \in [\pi_1, \pi_2] \subset (0, 1)$; or, the QS kernel is used to estimate $\hat{S}(\pi, FH_0)$, $\hat{S}(\pi, \hat{F})$, $\tilde{S}(\pi, FH_0)$ and $\tilde{S}(\pi, \hat{F})$, and there exist constants $K_1, K_2 > 0$ such that $K_1 T^{\frac{1}{5}} \leq S_T$, $S_{\lfloor \pi T \rfloor}$, $S_{T - \lfloor \pi T \rfloor} \leq K_2 T^{\frac{1}{5}}$ for all $\pi \in [\pi_1, \pi_2] \subset (0, 1)$; and (b) $\frac{T^{\frac{2}{5}}}{N} \rightarrow 0$ as $N, T \rightarrow \infty$.

Theorem 2: Under Assumptions 1 - 7, if Condition 1 or Condition 2 holds, then

$$\begin{aligned} \sup_{\pi \in [\pi_1, \pi_2]} \left\| \hat{S}(\pi, \hat{F}) - \hat{S}(\pi, FH_0) \right\| &\xrightarrow{p} 0 \\ \sup_{\pi \in [\pi_1, \pi_2]} \left\| \tilde{S}(\pi, \hat{F}) - \tilde{S}(\pi, FH_0) \right\| &\xrightarrow{p} 0. \end{aligned}$$

Theorem 2 shows that the infeasible sample variances can be replaced by the estimates computed using \hat{F} . Given this result, we can compute the Wald statistic and the LM-like statistics

$$\begin{aligned} W_T(\pi, \hat{F}) &\equiv A(\pi, \hat{F})' \hat{S}(\pi, \hat{F})^{-1} A(\pi, \hat{F}) \\ LM_T(\pi, \hat{F}) &\equiv A(\pi, \hat{F})' \tilde{S}(\pi, \hat{F})^{-1} A(\pi, \hat{F}) \end{aligned} \tag{2.13}$$

and sup-Wald and sup-LM defined in (2.6) and (2.7). To establish the asymptotic distributions of feasible statistics, we define their infeasible analogs as

$$\begin{aligned} W_T(\pi, FH_0) &\equiv A(\pi, FH_0)' \hat{S}(\pi, FH_0)^{-1} A(\pi, FH_0) \\ LM_T(\pi, FH_0) &\equiv A(\pi, FH_0)' \tilde{S}(\pi, FH_0)^{-1} A(\pi, FH_0) \end{aligned} \tag{2.14}$$

and we make the following assumption:

Assumption 9: (a) $\Omega \equiv \lim_{T \rightarrow \infty} \text{Var} \left(\text{vech} \left(\frac{1}{\sqrt{T}} \left(\sum_{t=1}^T H_0' f_t f_t' H_0 - I_r \right) \right) \right)$ is positive definite, and $\|\Omega\| < \infty$. $\hat{\Omega}_1(\pi, FH_0)$, $\hat{\Omega}_2(\pi, FH_0)$ and $\hat{\Omega}(FH_0)$ defined in (2.9) and (2.11) are consistent estimators of Ω satisfying that

$$\sup_{\pi \in [\pi_1, \pi_2]} \left\| \hat{\Omega}_m(\pi, FH_0) - \Omega \right\| = o_p(1) \text{ for } m = 1, 2$$

$$\left\| \hat{\Omega}(FH_0) - \Omega \right\| = o_p(1)$$

(b) $W_T(\pi, FH_0) \Rightarrow Q_p(\pi)$, $LM_T(\pi, FH_0) \Rightarrow Q_p(\pi)$, $\sup_{\pi \in [\pi_1, \pi_2]} W_T(\pi, FH_0) \xrightarrow{d} \sup_{\pi \in [\pi_1, \pi_2]} Q_p(\pi)$, and $\sup_{\pi \in [\pi_1, \pi_2]} LM_T(\pi, FH_0) \xrightarrow{d} \sup_{\pi \in [\pi_1, \pi_2]} Q_p(\pi)$, where $Q_p(\pi) = [B_p(\pi) - \pi B_p(1)]' [B_p(\pi) - \pi B_p(1)] / [\pi(1 - \pi)]$ and $B_p(\cdot)$ is a p -vector ($p = \frac{r(r+1)}{2}$) of independent Brownian motions on $[0, 1]$ restricted to $[\pi_1, \pi_2] \subset (0, 1)$.

Assumption 9(a) states that $\hat{\Omega}_1(\pi, FH_0)$, $\hat{\Omega}_2(\pi, FH_0)$ and $\hat{\Omega}(FH_0)$ converge to the population moment Ω uniformly in π . This is similar to Assumption 3 of Andrews (1993). Assumption 9(b) is just the main result of Theorem 3 in Andrews (1993): the sequences of Wald and LM statistics weakly converge to the stochastic process $Q_p(\pi)$ restricted to $[\pi_1, \pi_2] \subset (0, 1)$, and both $\sup_{\pi \in [\pi_1, \pi_2]} W_T(\pi, FH_0)$ and $\sup_{\pi \in [\pi_1, \pi_2]} LM_T(\pi, FH_0)$ converge to $\sup_{\pi \in [\pi_1, \pi_2]} Q_p(\pi)$ by the continuous mapping theorem. See CDG for more primitive assumptions under which high-level assumptions that are similar to Assumption 9(b) hold. Note that all the terms in Assumption 9 are computed using the infeasible data FH_0 , which means that if FH_0 were observable, one would be able to use the conventional supreme Wald test. The following theorem guarantees that one can use the estimated regressors, \hat{F} , to compute the supreme statistics, which have the same asymptotic distribution as those computed using FH_0 .

Theorem 3: Under Assumptions 1 – 9, if either Condition 1 or Condition 2 holds and $\frac{\sqrt{T}}{N} \rightarrow 0$ as $N, T \rightarrow \infty$, then

(i) $\sup_{\pi \in [\pi_1, \pi_2]} \left| W_T(\pi, \hat{F}) - W_T(\pi, FH_0) \right| = o_p(1)$ and $\sup_{\pi \in [\pi_1, \pi_2]} \left| LM_T(\pi, \hat{F}) - LM_T(\pi, FH_0) \right| = o_p(1)$.

(ii) $\sup_{\pi \in [\pi_1, \pi_2]} W_T(\pi, \hat{F}) \xrightarrow{d} \sup_{\pi \in [\pi_1, \pi_2]} Q_p(\pi)$ and $\sup_{\pi \in [\pi_1, \pi_2]} LM_T(\pi, \hat{F}) \xrightarrow{d} \sup_{\pi \in [\pi_1, \pi_2]} Q_p(\pi)$.

Theorem 3 shows that one can use the conventional critical values for the sup- W and sup- LM statistics computed using \hat{F} . The uniformity provided by part (i) of Theorem 3 also shows that $W_T(\pi, \hat{F}) \Rightarrow Q_p(\pi)$ and $LM_T(\pi, \hat{F}) \Rightarrow Q_p(\pi)$ by assumption 9(b). Thus, the continuous mapping theorem implies that the mean Wald statistic and the exponential Wald statistic proposed by Andrews and Ploberger (1994) can also be used to test structural breaks in factor loadings.

Corollary 1: Under Assumptions 1 – 9, if either Condition 1 or Condition 2 holds and $\frac{\sqrt{T}}{N} \rightarrow 0$ as $N, T \rightarrow \infty$, then

$$\int_{\pi_1}^{\pi_2} \exp\left(\frac{W_T(\pi, \hat{F})}{2}\right) d\pi - \int_{\pi_1}^{\pi_2} \exp\left(\frac{W_T(\pi, FH_0)}{2}\right) d\pi = o_p(1)$$

$$\int_{\pi_1}^{\pi_2} W_T(\pi, \hat{F}) d\pi - \int_{\pi_1}^{\pi_2} W_T(\pi, FH_0) d\pi = o_p(1)$$

Define $\exp\text{-}W(\hat{F}) \equiv \ln\left(\frac{1}{\pi_2 - \pi_1} \int_{\pi_1}^{\pi_2} \exp\left(\frac{W_T(\pi, \hat{F})}{2}\right) d\pi\right)$ and $\text{mean-}W(\hat{F}) \equiv \frac{1}{\pi_2 - \pi_1} \int_{\pi_1}^{\pi_2} W_T(\pi, \hat{F}) d\pi$.

Corollary 1 shows that critical values provided by Andrews and Ploberger (1994) can be applied to $\exp\text{-}W(\hat{F})$ and $\text{mean-}W(\hat{F})$ as if \hat{F} is observed rather than estimated. This result also holds for $\exp\text{-}LM(\hat{F})$ and $\text{mean-}LM(\hat{F})$ which can be defined in a similar way.

2.5 Asymptotics under the Alternative Hypothesis

We consider the alternative hypothesis that $[\alpha N]$ many cross sections have a single break at a common break date in their loadings. The model (2.2) can be rewritten as:

$$X = \begin{bmatrix} F_{0,1} & F_{1,1} & 0 \\ F_{0,2} & 0 & F_{1,2} \end{bmatrix} \begin{bmatrix} \Lambda'_0 \\ \Lambda'_1 \\ \Lambda'_2 \end{bmatrix} + e \quad (2.15)$$

Next, we partition the matrix $\Sigma_F \equiv E(f_t f_t')$ as

$$\Sigma_F \equiv \begin{bmatrix} \Sigma_F^{0,0} & \Sigma_F^{0,1} \\ \Sigma_F^{0,1'} & \Sigma_F^{1,1} \end{bmatrix}$$

where $\Sigma_F^{0,0} \equiv E(f_{0,t} f_{0,t}')$, $\Sigma_F^{0,1} \equiv E(f_{0,t} f_{1,t}')$, and $\Sigma_F^{1,1} \equiv E(f_{1,t} f_{1,t}')$.

Note that (2.15) is equivalent to a factor model that has time invariant factor loadings. Note that $[\Lambda_0 : \Lambda_1 : \Lambda_2]$ in (2.15) may not be of full column rank, so the representation of the equivalent model is not unique. However, it is only meaningful to deal with a representation with both factor and loading matrices of full column rank, so we reformulate the equivalent model as

$$X = G\Theta' + e. \quad (2.16)$$

Let g_t denote the transpose of the t^{th} row of G and θ_i denote the transpose of the i^{th} row of Θ , so $G = (g_1, g_2, \dots, g_T)'$ and $\Theta = (\theta_1, \theta_2, \dots, \theta_N)'$. If G has full column rank, the number of factors in (2.16) is determined by the rank of $\lim_{N \rightarrow \infty} \Theta' \Theta / N$. We use the rank of $\Theta' \Theta / N$ to define three different types of breaks in the factor loading matrix. The detailed expressions of G and Θ depend on the number of factors in (2.16) and are discussed below.

Type 1 Break: $\text{rank}(\Theta' \Theta / N) = r + q_1$.

Type 1 break requires the column rank of $[\Lambda_1 : \Lambda_2]$ to be $2q_1$. The break in λ_i should be so idiosyncratic across i 's that Λ_1 and Λ_2 are linearly independent. Under type 1 break, (2.15) is equivalent to a factor model that has $r + q_1$ factors with time-invariant factor loading matrix:

$$X = G_{(1)} \Theta'_{(1)} + e \quad (2.17)$$

where $G_{(1)} = \begin{bmatrix} F_{0,1} & F_{1,1} & 0 \\ F_{0,2} & 0 & F_{1,2} \end{bmatrix}_{T \times (r+q_1)}$ is the factor matrix and $\Theta_{(1)} = [\Lambda_0 : \Lambda_1 : \Lambda_2]_{N \times (r+q_1)}$ is

the factor loading matrix.

Type 2 Break: $\text{rank}(\Theta'\Theta/N) = r$.

Under type 2 break, the column rank of $[\Lambda_1:\Lambda_2]$ is q_1 , so there exists a $q_1 \times q_1$ matrix Z , either singular or nonsingular, such that $\Lambda_2 = \Lambda_1 Z$ (or $\Lambda_1 = \Lambda_2 Z$). Type 2 break means that all the loadings change in a very homogeneous way. For example, $\Lambda_2 = 2\Lambda_1$, which might be unlikely to happen in practice. A more empirically relevant example is that $\text{rank}(\Lambda_1) = q_1$ and $\Lambda_2 = 0$, so the q_1 -dimensional factors $f_{1,t}$ disappear after the break. Hence, type 2 break covers the case of emerging or disappearing factors. Without loss of generality, we assume that $\Lambda_2 = \Lambda_1 Z$, so the equivalent model has r transformed factors under type 2 break:

$$X = G_{(2)}\Theta'_{(2)} + e \quad (2.18)$$

where $G_{(2)} = \begin{bmatrix} F_{0,1} & F_{1,1} \\ F_{0,2} & F_{1,2}Z' \end{bmatrix}_{T \times r}$ is the factor matrix and $\Theta_{(2)} = [\Lambda_0:\Lambda_1]_{N \times r}$ is the factor loading

matrix. Note that (2.18) has the same loading matrix as under the null, i.e. $\Lambda \equiv [\Lambda_0:\Lambda_1]$. The change in Λ is transmitted to the factors, and $F_{1,2}$ is rotated by Z' .⁵

Type 3 Break: $\text{rank}(\Theta'\Theta/N) = r + \ell$, where $0 < \ell < q_1$.

Under type 3 break, the column rank of $[\Lambda_1:\Lambda_2]$ is greater than q_1 but less than $2q_1$, so some (but not all) columns of Λ_1 and Λ_2 are linearly dependent. Compare the equivalent models in (2.17) and (2.18): for type 1 break, the dimension of the factor space is augmented; for type 2 break, the factors are rotated with unchanged loadings. It is not difficult to see that type 3 break will lead to an equivalent model with characteristics of both types 1 and 2 breaks. The linearly dependent columns of Λ_1 and Λ_2 will transform the factors by rotation as under type 2 break, while the linearly independent columns of Λ_1 and Λ_2 will augment the factor space as under type 1 break. Without loss of the main insight, our analysis will focus on breaks of types 1 and 2 to avoid introducing more

⁵Note that we assume $E(f_t f_t')$ to be constant over time. If $E(f_t f_t')$ has a break, then the model is equivalent to a factor model with time-invariant $E(f_t f_t')$ but a type 2 break in the loading matrix. To see it, let Σ_F and Σ_F^* be the variances of f_t before and after the break, respectively. If both Σ_F and Σ_F^* are positive definite, then there exist a nonsingular matrix Ψ such that $\Psi \Sigma_F \Psi' = \Sigma_F^*$. It can be viewed as a factor model where factors have constant second moment Σ_F , the pre-break loading is Λ , and post-break loading is $\Lambda \Psi$.

tedious notations associated with type 3 break. The power properties of our statistics under type 3 break will be the combination of those under types 1 and 2.

Assumption 10: Conditions on the Break

Let $\Sigma_{G^{(j)}} \equiv \text{plim}_{T \rightarrow \infty} G^{(j)}G^{(j)}/T$ and $\Sigma_{\Theta^{(j)}} = \lim_{N \rightarrow \infty} \Theta^{(j)'}\Theta^{(j)}/N$ for $j = 1, 2$.

(a) Type 1 break: $\Sigma_{G^{(1)}}$ is positive definite. $\|\lambda_{2,i}\| \leq \bar{\lambda} < \infty$, $\|\Theta^{(1)'}\Theta^{(1)}/N - \Sigma_{\Theta^{(1)}}\| \rightarrow 0$ for some $(r + q_1) \times (r + q_1)$ positive definite matrix $\Sigma_{\Theta^{(1)}}$, and $\|\Theta^{(1)'}\Theta^{(1)}/N - \Sigma_{\Theta^{(1)}}\| \leq O\left(\frac{1}{\sqrt{N}}\right)$.

(b) Type 2 break: $\|Z\| < \infty$ and $Z\Sigma_F^{1,1}Z' \neq \Sigma_F^{1,1}$.

(c) For both types 1 and 2 breaks, The eigenvalues of $\Sigma_{G^{(j)}}\Sigma_{\Theta^{(j)}}$ are distinct.

Part (a) of Assumption 10 ensures that the number of factors in the equivalent model (2.17) is equal to $r + q_1$. The positive definiteness of $\Sigma_{G^{(1)}}$ requires the columns of $G^{(1)}$ to be linearly independent, which is not restrictive given the structure of $G^{(1)}$. The requirement on $\Theta^{(1)}$ is simply the analog of Assumption 2 on Λ . The restriction that $Z\Sigma_F^{1,1}Z' \neq \Sigma_F^{1,1}$ in part (b) of Assumption 10 is to ensure the consistency of our statistics. It rules out a very unlikely case where $Z = -1$, i.e., all the loadings switch their signs after the break. Part (c) plays the same role as Assumption 7 and ensures the convergence and nonsingularity of the rotation matrix J defined below.

Next, we define the analogs of V_{NT} , V , H , and H_0 under the alternative hypothesis. Recall that \hat{F} denotes the PCA estimate of factors and \hat{f}_t denotes the transpose of the t^{th} row of \hat{F} . Under the alternative hypothesis, however, \hat{F} will be an estimate of factors (up to a rotation) in the equivalent models (2.17) and (2.18). Since the equivalent models under types 1 and 2 are different, we will discuss the corresponding cases separately.

Type 1 break:

Let U_{NT} be the $(r + q_1) \times (r + q_1)$ diagonal matrix of the first $r + q_1$ largest eigenvalues of $(1/TN)XX'$ in descending order. Let U be the probability limit of U_{NT} , where U is the diagonal matrix consisting of the eigenvalues of $\Sigma_{\Theta^{(1)}}^{-\frac{1}{2}}\Sigma_{G^{(1)}}\Sigma_{\Theta^{(1)}}^{-\frac{1}{2}}$ in descending order (Lemma A3, Bai and Ng, 2003). Let $J = (\Theta^{(1)'}\Theta^{(1)}/N)(G^{(1)'}\hat{F}/T)U_{NT}^{-1}$ be an $(r + q_1) \times (r + q_1)$ matrix. Denote $\text{plim}_{T,N \rightarrow \infty} J$

as J_0 , which is a non-singular matrix (by Proposition 1 of Bai, 2003) ⁶. Let

$$D_1 \equiv \begin{bmatrix} \Sigma_F^{0,0} & \Sigma_F^{0,1} & 0 \\ \Sigma_F^{0,1'} & \Sigma_F^{1,1} & 0 \\ 0 & 0 & 0 \end{bmatrix}, D_2 \equiv \begin{bmatrix} \Sigma_F^{0,0} & 0 & \Sigma_F^{0,1} \\ 0 & 0 & 0 \\ \Sigma_F^{0,1'} & 0 & \Sigma_F^{1,1} \end{bmatrix}, \text{ and } C \equiv J_0'(D_1 - D_2)J_0. \quad (2.19)$$

Type 2 break:

Let U_{NT} be the $r \times r$ diagonal matrix of the first r largest eigenvalues of $(1/TN)XX'$ in descending order. Let U be the probability limit of U_{NT} , where U is the diagonal matrix consisting of the eigenvalues of $\Sigma_{\hat{\Theta}(2)}^{\frac{1}{2}} \Sigma_{G(2)} \Sigma_{\hat{\Theta}(2)}^{\frac{1}{2}}$ in descending order (Lemma A3, Bai and Ng, 2003). Let $J = (\Theta'_{(2)} \Theta_{(2)} / N)(G'_{(2)} \hat{F} / T) U_{NT}^{-1}$ be an $r \times r$ matrix. Denote $\text{plim}_{T,N \rightarrow \infty} J$ as J_0 , which is a non-singular matrix (by Proposition 1 of Bai, 2003). Let

$$D_1 \equiv \begin{bmatrix} \Sigma_F^{0,0} & \Sigma_F^{0,1} \\ \Sigma_F^{0,1'} & \Sigma_F^{1,1} \end{bmatrix}, D_2 \equiv \begin{bmatrix} \Sigma_F^{0,0} & \Sigma_F^{0,1} Z' \\ Z \Sigma_F^{0,1'} & Z \Sigma_F^{1,1} Z' \end{bmatrix}, \text{ and } C \equiv J_0'(D_1 - D_2)J_0. \quad (2.20)$$

To establish the consistency of the test under the alternative hypothesis, we need Assumption 11 that regulates the asymptotic property of the variance matrices in our statistics. Define $\hat{\Omega}_1(\pi^*, GJ_0)$, $\hat{\Omega}_2(\pi^*, GJ_0)$, $\hat{\Omega}(GJ_0)$ by replacing π , F , and H_0 in equations (2.9), (2.10), (2.11), and (2.12) with π^* , G , and J_0 , respectively. Let $\hat{S}(\pi^*, GJ_0) = \frac{1}{\pi^*} \hat{\Omega}_1(\pi^*, GJ_0) + \frac{1}{1-\pi^*} \hat{\Omega}_2(\pi^*, GJ_0)$ and $\tilde{S}(\pi^*, GJ_0) = \frac{1}{\pi^*} \hat{\Omega}(GJ_0) + \frac{1}{1-\pi^*} \hat{\Omega}(GJ_0)$.

Assumption 11: (a)

$$\text{plim}_{T \rightarrow \infty} \inf \left\{ \text{vech}(C)' \left[\max(S_{\lfloor \pi^* T \rfloor}, S_{T - \lfloor \pi^* T \rfloor}) \hat{S}(\pi^*, GJ_0)^{-1} \right] \text{vech}(C) \right\} > 0$$

$$\text{plim}_{T \rightarrow \infty} \inf \left\{ \text{vech}(C)' \left[S_T \tilde{S}(\pi^*, GJ_0)^{-1} \right] \text{vech}(C) \right\} > 0$$

where $S_{\lfloor \pi^* T \rfloor}$, $S_{T - \lfloor \pi^* T \rfloor}$, and S_T are the bandwidth parameters for $\hat{\Omega}_1(\pi^*, GJ_0)$, $\hat{\Omega}_2(\pi^*, GJ_0)$, and

⁶The subscripts (1) and (2) to distinguish two types of breaks in U_{NT} , U , J , J_0 and C are omitted to simplify the notations.

$\hat{\Omega}(GJ_0)$, respectively.

(b) Condition 1 or 2 holds for $\hat{\mathbb{S}}(\pi^*, GJ_0)$, $\hat{\mathbb{S}}(\pi^*, \hat{F})$, $\tilde{\mathbb{S}}(\pi^*, GJ_0)$, and $\tilde{\mathbb{S}}(\pi^*, \hat{F})$.

Remark 1: Assumption 11(a) ensures that $W_T(\pi^*, GJ_0)$ and $LM_T(\pi^*, GJ_0)$ diverge under the alternative as N and T go to infinity. Contrast to Equation (2.8), it is worth noting that $E(J_0'g_tg_t'J_0) \neq I_{r+q_1}$ under the alternative hypothesis, so the HAC estimators are not properly demeaned. Hall (2000) investigates the properties of HAC estimators that are not properly demeaned in the context of overidentifying restriction tests. He shows that if the HAC estimate is not correctly demeaned, then it will diverge at the rate of the bandwidth parameter b_T . He uses μ_* to denote the expectation of the invalid moment conditions and \mathfrak{S} to denote the HAC estimator, and shows that $b_T\mu_*'\mathfrak{S}^{-1}\mu_* \xrightarrow{P}$ a positive constant. (see Hall (2000), Proof of Theorem 2, p. 1525-1526). Although Hall's (2000) result is developed for the HAC estimator in the context of overidentifying restriction tests, it can be readily extended to our HAC estimators. In this paper, it turns out that $\text{vech}(C)$ is an analog of μ_* and our bandwidth parameters are analogs of b_T . Hence, Assumption 9'(a) is analogous to Hall's result that $b_T\mu_*'\mathfrak{S}^{-1}\mu_*$ is asymptotically bounded away from zero, and it can be proved under more primitive conditions in Hall (2000).

Theorem 4: Under Assumptions 1 – 8, and 11, if the break of factor loading matrix satisfies Assumption 10, then

(i) There exists some non-random matrix $C \neq 0$, such that $\frac{1}{[\pi^*T]} \sum_{t=1}^{[\pi^*T]} \hat{f}_t\hat{f}_t' - \frac{1}{T-[\pi^*T]} \sum_{t=[\pi^*T]+1}^T \hat{f}_t\hat{f}_t' \xrightarrow{P} C$.

(ii) For any constants π_1 and π_2 that satisfy $0 < \pi_1 \leq \pi^* \leq \pi_2 < 1$, $\sup_{\pi \in [\pi_1, \pi_2]} W_T(\pi, \hat{F})$ and $\sup_{\pi \in [\pi_1, \pi_2]} LM_T(\pi, \hat{F})$ are consistent under the alternative hypothesis that a fraction of N cross sections have structural breaks in their factor loadings at a common date $[\pi^*T]$.

Theorem 4(i) shows that pre- and post-break subsample means of $\hat{f}_t\hat{f}_t'$ converge to different limits under the alternative hypothesis. This explains why just using a Wald statistic computed using estimated factors can in fact detect the structural breaks in factor loadings. Note that the factors (in static form) are estimated by PCA which implicitly assumes that the factor loadings are time-

invariant. Hence, Theorem 4(a) shows that the PCA will transmit the structural breaks in the loading matrix to the subsample means of $\hat{f}_t \hat{f}_t'$ before and after $\lfloor \pi^* T \rfloor$.

Recall that the number of factors is increased under type 1 break. This indicates that the number of factors plays an important role in determining the asymptotics of the test statistics in this paper. In practice, the number of factors is commonly estimated using IC proposed by Bai and Ng (2002), so the asymptotics of IC under the null and alternative hypotheses will affect the performance of the structural break tests for factor loadings. It turns out that Bai and Ng's IC can consistently estimate the number of factors in the equivalent model and help our statistics distinguish the null and alternative hypotheses in large samples.

Proposition 1: Under Assumptions 1–4 and 10, Bai and Ng's information criteria consistently estimate the number of factors of the equivalent models (2.17) and (2.18).

IC is equivalent to determining the number of asymptotically non-zero eigenvalues of XX'/NT . Assumptions 1 – 4 play the role of Assumptions A – D of Bai and Ng (2002) to ensure IC's consistency under the null hypothesis. Under the alternative hypothesis, Proposition 1 shows that IC consistently estimates the number of factors in the equivalent models. Thus, the asymptotics of test statistics proposed in this paper will not be affected by implementing IC in the first stage as N and T tend to infinity, and the finite-sample effect of the first-stage IC will be investigated in Monte Carlo experiments in the next section.

When IC is used in the first stage to determine the number of factors estimated by PCA, tests for structural breaks should not be based on factor loadings. For example, the traditional Chow test for testing $\lambda_{1,i} = \lambda_{2,i}$ will not have power under the alternative hypothesis, because the factor loadings of the equivalent models are actually time-invariant.⁷ Therefore, in order to test structural breaks in factor loadings, one should focus on the estimated factors rather than the estimated factor loadings.

⁷Breitung and Eickmeier (2011) also point out that the Chow test statistic will also lack power if the number of factors is determined using IC based on full sample, so they suggest implementing IC to both pre- and post-break subsamples. This solution is subject to the knowledge of the break date, and the IC may become less accurate since the time dimension will be shortened once the sample is split.

3 Monte Carlo Simulations

In the Monte Carlo experiments we investigate the finite sample properties of our statistics for known and unknown break points. Section 3.1 compares the performance of our Wald test W and LM-like test LM with BE's pooled test S_{BE} and the Bonferroni test s_{Bon} , where W and LM abbreviate $W_T(\pi, \hat{F})$ and $LM_T(\pi, \hat{F})$, respectively. Recall that $S_{BE} = (\sum_{i=1}^N s_i - rN) / \sqrt{2rN}$, where s_i is BE's individual statistic⁸ for the i^{th} variable and r is estimated by Bai and Ng's IC. We use three different superscripts to denote the way to compute s_i : s_i^0 denotes BE's individual statistic assuming that the idiosyncratic shocks are conditionally homoskedastic and serially uncorrelated; s_i^{GLS} denotes the individual statistic computed using quasi-demeaned residuals which are based on AR models with lags selected by BIC; s_i^{HAC} denotes the individual statistic computed using HAC estimate. Let S_{BE}^0 , S_{BE}^{GLS} and S_{BE}^{HAC} denote the pool statistics computed using s_i^0 , s_i^{GLS} and s_i^{HAC} , respectively. Besides the pooled statistics, we also include the results based on Bonferroni critical values: $\mathcal{F}^{-1}(1 - 5\%/N)$, where \mathcal{F} is the chi-square CDF with degree of freedom \hat{r} . The reason for considering the Bonferroni test is that there are N BE individual statistics s_i , using the 5% significance level for each s_i will always result in a fraction of s_i 's rejecting the null hypothesis even if the factor loading matrix is constant over time. The Bonferroni method is a simple way to control the overall type I error of all s_i statistics. Let s_{Bon}^0 , s_{Bon}^{GLS} and s_{Bon}^{HAC} denote the Bonferroni statistics based on s_i^0 , s_i^{GLS} and s_i^{HAC} , respectively. In addition, our W and LM statistics are computed using three different estimates for the sample variances: W_0 and LM_0 are computed using White's (1980) conditional heteroskedasticity robust estimate; W_B and LM_B are computed using Newey and West's (1994) data dependent HAC estimate based on the Bartlett kernel; W_{QS} and LM_{QS} are computed using the same data dependent HAC estimate but based on the QS kernel.

Section 3.2 compares the performance of the sup- W , exp- W , mean- W , sup- LM , exp- LM , and mean- LM tests when the break date is unknown. Similar to Section 3.1, the subscripts "0", "B" and "QS" denote the statistics using the conditional heteroskedasticity robust estimate, HAC estimate based on the Bartlett kernel, and HAC estimate based on the QS kernel, respectively. Section 3.3

⁸We use BE's LM statistics to compute the pooled test statistic following Breitung and Eickmeier's (2011) suggestion based on their simulation results.

provides Monte Carlo experiments on the power comparison between our and CDG's tests.

In all Monte Carlo experiments, factor loadings are initially set randomly, and then fixed throughout 5000 replications for each DGP.

3.1 Testing Breaks with Known Break Date

Our first experiment focuses on the size of W , LM , s_{Bon} and S_{BE} when the break date is known. The model is $x_{it} = \sum_{k=1}^r \lambda_{ik} f_{kt} + \kappa e_{it}$, where $\lambda_{ik} \stackrel{iid}{\sim} N(\frac{b}{2}, 1)$, and f_{kt} and e_{it} are generated by the following DGPs:

$$\text{N1: } f_{kt}, e_{it} \stackrel{iid}{\sim} N(0, 1), \kappa = \sqrt{(1 + b^2/4)r}.$$

$$\text{N2: } f_{kt} \stackrel{iid}{\sim} N(0, 1), e_{it} = \sigma_i(\nu_{it} + \sum_{1 \leq |j| \leq P} \beta \nu_{i-j,t}), \sigma_i \stackrel{iid}{\sim} U(0.5, 1.5), \nu_{it} \stackrel{iid}{\sim} N(0, 1), \text{ and } \kappa = \sqrt{12(1 + b^2/4)r/13(1 + 2P\beta^2)}.$$

$$\text{N3: } f_{kt} = \rho_f f_{kt-1} + \mu_{it}, \mu_{it} \stackrel{iid}{\sim} N(0, 1 - \rho_f^2), e_{it} = \sigma_i \nu_{it}, \sigma_i \stackrel{iid}{\sim} U(0.5, 1.5), \nu_{it} = \rho_\nu \nu_{it-1} + \epsilon_{it} + \omega_\epsilon \epsilon_{it-1}, \epsilon_{it} \stackrel{iid}{\sim} N\left(0, \frac{1}{1 + (\rho_\epsilon + \omega_\epsilon)^2 / (1 - \rho_\epsilon^2)}\right), \text{ and } \kappa = \sqrt{12(1 + b^2/4)r/13}.$$

In N1 – N3, we set $b = 1$ and $r = 3$, and the value of κ is chosen so that $R^2 = \text{trace}(E(ee'))/\text{trace}(E(XX'))$ is 50%.⁹ N1 is the simplest DGP: both factors and idiosyncratic shocks are i.i.d, i.e. no correlation or heteroskedasticity is involved. Both N2 and N3 allow heteroskedasticity across i , and we follow Breitung and Eickmeier's (2011) setup: $\sigma_i \stackrel{iid}{\sim} U(0.5, 1.5)$. N2 also allows limited cross-sectional correlation in idiosyncratic shocks if $\beta \neq 0$ and $P \geq 1$. We let $\beta \in \{0, 0.1\}$ and $P \in \{6, 8\}$, and these values are similar to those of Onatski (2010). DGP N3 considers the case where both factors and idiosyncratic shocks are serially correlated. The factors are assumed to be AR(1) processes, and $\rho_f = 0.7$ which leads to mild persistency. ν_{it} follows an AR(1) process if ω_ϵ is zero, or an ARMA(1, 1) process otherwise. We set $\omega_\epsilon \in \{0, 0.5\}$ and $\rho_\nu = 0.5$.

Table 1 reports the size of the Bonferroni test, BE's pooled tests and our tests.¹⁰ The last column of Table 1 is averaged number of factors selected by IC_{p_1} of Bai and Ng (2002). It is remarkable that both BE's pooled statistic and Bonferroni statistic are valid under sequential asymptotics where T

⁹For the choice of κ , note that $E(\lambda_{ij} F_{tj})^2 = 1 + b^2/4$ and $E(\sigma_i^2) = 13/12$.

¹⁰For the s_{bon} and S_{BE} tests, we implement IC to the full sample. While BE suggest implementing IC to pre- and post-break subsamples to determine r , it is not very clear about which to use when the numbers of factors estimated from the pre- and post-break subsamples are different.

first goes to infinity while N is fixed and then N goes to infinity. In the context of factor models, however, it is assumed that N and T go to infinity simultaneously, so S_{BE} and s_{Bon} are theoretically invalid. Our experiments investigate the consequence of applying these invalid statistics in practice. It can be seen under DGP N3 that S_{BE}^{HAC} always rejects more than 75% of the times, which is similar to the results in the working paper version of Breitung and Eickmeier (2011). S_{BE}^{GLS} has a substantial size distortion when $\omega_\varepsilon = 0$ and $(N, T) = (500, 100)$. Under DGP N3, the effective size of s_{Bon}^{HAC} can be as high as 16.7%. While s_{Bon}^{GLS} does not exceed the 5% nominal size, it is theoretically invalid for the same reason as s_{Bon}^{HAC} and could have incorrect size under other DGPs not considered by our paper.

Also, under DGP N2 with cross-sectional correlations ($\beta \neq 0$ and $P > 0$), the pooled tests tend to over-reject the null hypothesis. For example, the effective size of S_{BE}^{GLS} is 16.2% when $P = 8$, $\beta = 0.1$, $N = 200$, and $T = 100$. In contrast, our tests do not require the independence of idiosyncratic shocks, so the size of our tests is robust to cross-sectional correlation in e_{it} . Moreover, IC tends to over-estimate the number of factors when the correlation is relatively strong. For instance, when $P = 8$, $\beta = 0.1$, $N = 100$, and $T = 200$, the average of estimated number of factors is 6.37, but our tests demonstrate robustness to the overestimation of r in simulations. (See also the cases where the averaged \hat{r} equal to 3.46 and 3.66). Finally, under DGP N3, the size of W and LM based on HAC estimates are close to 5%, and LM tends to have better size than W for small T .

The second experiment compares the powers of W , LM , s_{Bon} and S_{BE} when the break date is known. The break date is set to be $\frac{T}{2}$, and the data are generated by the following DGPs:

A1: $x_{it} = \sum_{k=1}^r \lambda_{ik} f_{kt} + \kappa e_{it}$ for $i = 1, 2, \dots, N$ and $t \leq T/2$, and $x_{it} = \sum_{k=1}^r (\lambda_{ik} - b) f_{kt} + \kappa e_{it}$ for $i = 1, 2, \dots, N$ and $t \geq T/2 + 1$, where f_{kt} , $e_{it} \stackrel{iid}{\sim} N(0, 1)$, $\kappa = \sqrt{(1 + b^2/4)r}$, and $\lambda_{ik} \stackrel{iid}{\sim} N(\frac{b}{2}, 1)$.

A2: $x_{it} = \sum_{k=1}^r \lambda_{ik} f_{kt} + \kappa e_{it}$ for $i = 1, 2, \dots, \alpha N$ and $t \leq T/2$, $x_{it} = \sum_{k=1}^r (\lambda_{ik} - b) f_{kt} + \kappa e_{it}$ for $i = 1, 2, \dots, \alpha N$ and $t \geq T/2 + 1$, and $x_{it} = \sum_{k=1}^r \lambda_{ik} f_{kt} + \kappa e_{it}$ for $i = \alpha N + 1, \dots, N$ and $t = 1, 2, \dots, T$, where f_{kt} , $e_{it} \stackrel{iid}{\sim} N(0, 1)$, $\kappa = \sqrt{(1 + b^2/4)r}$, $\lambda_{ik} \stackrel{iid}{\sim} N(\frac{b}{2}, 1)$, and $b = 1$.

A3: $x_{it} = \sum_{k=1}^r \lambda_{ik} f_{kt} + \kappa e_{it}$ for $i = 1, 2, \dots, N$, where f_{kt} , $e_{it} \stackrel{iid}{\sim} N(0, 1)$, $\kappa = \sqrt{(1 + b^2/4)r}$,

$\lambda_{ik} = \lambda_{ik}^0$ for $t \leq T/2$, $\lambda_{ik} = c \cdot \lambda_{ik}^0$ for $t \geq T/2 + 1$, $\lambda_{ik}^0 \stackrel{iid}{\sim} N(\frac{b}{2}, 1)$, and $b = 1$.

We set $r = 3$ for DGPs A1–A3. It is not difficult to verify that the equivalent factor model with time-invariant loadings has four factors under both DGPs A1 and A2. DGP A3 considers type 2 break discussed in section 2.5, so the number of factors under DGP A3 is unchanged.

DGP A1 focuses on how the power changes as the magnitude of break in factor loadings increases. We set $b \in \{1/3, 2/3, 1, 2\}$ in DGP A1 and the results are summarized in Table 2A. The pooled tests and our tests have very different patterns of power. When $b = 1/3$ and N and T are relatively small, the pooled tests are very powerful, while ours do not have good power. However, as N and T increase, our tests become powerful. When N and $T = 500$, our tests always reject the null, whereas the pooled tests reject less than 10%. Additionally, when b becomes larger, LM and W are powerful even for small N and T , while the power of the pooled tests is in fact close to the nominal size. Note that the equivalent model with time-invariant loadings has four factors under DGP A1, so XX'/NT has four nonzero eigenvalues asymptotically. When $b = 1/3$, and N and $T \leq 200$, IC only captures the first three nonzero eigenvalues, yielding three estimated factors. By Stock and Watson’s (2002b) result, these three factors consistently estimate the original factor space because the break is “small”. Under such a circumstance, PCA estimates the original factor models with time-varying loadings, so BE’s pooled statistics are powerful. However, as the break becomes larger, Stock and Watson’s (2002b) result based on “small” break does not hold any more. The factor space is augmented, and PCA in fact estimates the equivalent model with four factors (see the last column of Table 2A). It is clear that our tests are much more powerful for $b \geq 2/3$, whereas the rejection rates of the pooled statistics are close to 5% because the equivalent model does not have time-varying loadings.¹¹ Finally, since the Bonferroni tests are always more conservative than the pooled tests, it is not surprising that Bonferroni tests lack power when the sample size or b is large.

DGP A2 investigates the power when only a fraction of factor loadings have structural breaks. We set $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$. The results are shown in Table 2B. When $\alpha = 0.2$ and N and

¹¹The individual test s_i will also be lack of power due to the same reason. To conserve space, we do not report these results.

$T = 100$, the pooled and Bonferroni tests have high power, while our tests are less powerful. This is consistent with the recent result by Bates et al. (2012) that PCA estimator can consistently estimate the factor space when only a “small” (asymptotically negligible) fraction of factor loadings have breaks. As N and T increase to 200, however, our tests become more powerful than the pooled and Bonferroni tests. Also, when a larger fraction of factor loadings have structural breaks, LM and W have better power than the other two classes of tests.

DGP A3 considers the power against type 2 break. The post-break loadings are equal to pre-break loadings scaled by a parameter c . We set $c^2 \in \{0, 0.25, 0.5, 0.75\}$, so the factor structure becomes weaker after the break. When $c = 0$, there is no factor structure in the post-break subsample. The results are reported in Table 2C. It is remarkable that both Bonferroni and pooled statistics have almost no power under this type of break. The reason is that the factor space is not augmented under this type of break, and PCA always estimates a model with time-invariant loadings whether the break is small or big (see (2.18)). In contrast, the results in Table 2C show that the power of W and LM increases as N and T increase, confirming the consistency of our tests against type 2 break.

3.2 Testing Breaks with Unknown Break Date

In this subsection, we investigate the size and power of six statistics, $\text{sup-}W$, $\text{exp-}W$, $\text{mean-}W$, $\text{sup-}LM$, $\text{exp-}LM$, and $\text{mean-}LM$, without imposing the knowledge about the location of break date. Table 3 presents the simulation results under the null hypothesis. Under DGP N2, the size of our tests is not affected by the cross sectional correlation of the idiosyncratic shocks. Note that N2 does not allow serial correlation of factors or idiosyncratic shocks. It turns out that the statistics using HAC estimates become relatively conservative, but as the time dimension increases, their effective size becomes closer to 5%. Under DGP N3, we consider the case where both factors and idiosyncratic shocks are AR(1) processes. As expected, statistics without using HAC over-reject the null hypothesis more frequently than the nominal size, and those using HAC estimates have much better effective size. Since the LM-like test uses the full sample to estimate the sample variance, it is not surprising that the tests based on LM_B and LM_{QS} almost always have better size than those

based on W_B and W_{QS} . In fact, sup- LM has the best size: the size of both sup- LM_B and sup- LM_{QS} is clustered at around 5%; exp- LM also has decent size: the size of exp- LM_B and exp- LM_{QS} is clustered at 6%; mean- W test over-rejects the most frequently, but their size never exceeds 12%.

Tables 4A, 4B and 4C report the power of sup- W , exp- W , mean- W , sup- LM , exp- LM , and mean- LM under DGPs A1, A2 and A3, respectively. Under DGP A1, all of these tests do not have good power when the magnitude of the break is small, i.e., $b = 1/3$, for small N and T . This is similar to the results shown in Table 2A. As b increases, all of these tests become powerful, though the power is not monotonically increasing in b . Under DGP A2, all tests are not very powerful in small samples ($N = 100$ and $T = 200$) when a small fraction of ($\alpha = 0.2$) factor loadings have structural breaks. However, the power increases substantially and approaches one as the sample size increases, with $\alpha = 0.2$ unchanged. Additionally, as structural breaks become more prevalent, all tests become powerful. Under DGP A3, all tests have limited power when $c^2 = 0.75$ and $T \leq 200$. It is remarkable that mean tests are more powerful than supreme and exponential tests when $c^2 = 0.75$. This is consistent with the result by Andrews and Ploberger (1994) that the mean test is designed to be powerful against small breaks. As c^2 decreases, all tests become more powerful. When $T = 500$ and $c^2 \leq 0.5$, all tests have power equal to one.¹² Finally, note that our tests almost always detect reasonably big breaks when $N = T = 500$ under DGPs A1–A3. This confirms the consistency of our tests.

3.3 Power Comparison with the Chen, Dolado, and Gonzalo’s (2012) Test

Chen, Dolado, and Gonzalo (2012) propose a test for structural break in factor loadings by testing whether or not there is a break in the coefficients of the following regression:

$$\hat{f}_{1t} = \phi_2 \hat{f}_{2t} + \dots + \phi_{\hat{r}} \hat{f}_{\hat{r}t} + u_t = \phi' \hat{f}_{-1t} + u_t,$$

where \hat{r} is the number of factors determined by Bai and Ng’s (2002) IC, $\hat{f}_{1t}, \dots, \hat{f}_{\hat{r}t}$ are the factors associated with the 1st, ..., \hat{r}^{th} eigenvalues of XX' , respectively, and $\hat{f}_{-1t} \equiv [\hat{f}_{2t} \dots \hat{f}_{\hat{r}t}]'$ and $\phi \equiv$

¹²Note that DGPs A1–A3 do not allow serial correlation in the data, so the tests using HAC estimates are less powerful than those using heteroskedasticity robust estimates.

$[c_2 \dots c_{\hat{r}}]'$. They develop both Wald and LM versions for their test.

We perform power comparison between our tests and the CDG tests under DGPs A1 and A3 for unknown break date. Table 5A reports the results under DGP A1 (the superscript “CDG” denotes the CDG tests). Under DGP A1, neither our test nor CDG test uniformly dominates the other. When White’s (1980) conditional heteroskedasticity robust estimate is used, our $\sup -LM_0$ test tends to be more powerful than $\sup -LM_0^{CDG}$, while our $\sup -W_0$ has similar power to $\sup -W_0^{CDG}$. When the HAC estimator is used with the Bartlett kernel, the results are mixed. $\sup -LM_B$ is more powerful than $\sup -LM_B^{CDG}$ when $T = 500$. The power of W_B is very similar to that of W_B^{CDG} , while $\sup -W_B^{CDG}$ tends to be more powerful when $T \leq 200$.

Table 5B reports the results under DGP A3. It is remarkable that both $\sup -LM_0^{CDG}$ and $\sup -LM_B^{CDG}$ have very limited power even when the factor structure disappears after the break, i.e., $c = 0$, for $N = T = 500$. $\sup -W_0^{CDG}$ and $\sup -W_B^{CDG}$ also have little power except for $c = 0$. Note that this type of alternative is ruled out by CDG’s assumptions. In contrast, both our Wald and LM tests are powerful under DGP A3. Hence, our tests can detect more types of breaks in factor loadings.

Remark 2: Note that the power of our tests depends on the consistency of the Bai and Ng’s (2002) IC. It is well known that Bai and Ng’s (2002) IC tend to overestimate the number of factors if the idiosyncratic shocks are correlated (see simulation results of Onatski (2010)). It is remarkable that the power of our tests will not suffer too much because an \hat{F} with additional columns still has a break in its second moment. Hence, our test is still consistent when the number of factors is overestimated (namely, $\hat{r} > r + q_1$ under type 1 break or $\hat{r} > r$ under type 2 break). If the number of factors is underestimated, our test may lose some power in finite samples because there may be breaks in the ignored factors. This is also observed by Chen et al. (2012) in their simulations for the CDG tests.

4 Conclusions

This paper proposes new test statistics for structural breaks in factor loadings of dynamic factor models. We consider testing the null hypothesis that the factor loading matrix is constant over time against the alternative hypothesis that a fraction of or all factor loadings have a single break at a common date. Our contributions include the following: First, by reducing the infinite dimensional parameter problem into a finite dimensional problem, we are able to allow for the use of conventional critical values, unknown break dates, and serial and cross-sectional correlations in the idiosyncratic shocks. Second, we treat the number of factors to be determined rather than to be known, and our tests are more powerful than the existing tests based on factors selected by information criteria. Although we only considered the single-break alternative hypothesis, we expect that our tests have power against the alternative in which there are finitely many breaks in factor loadings.

Appendix

A Proofs of the Results in Section 2.4

First, recall that V_{NT} is the $r \times r$ diagonal matrix of the first r largest eigenvalues of $(1/TN)XX'$ in decreasing order, and the estimated factor matrix \hat{F} is \sqrt{T} times eigenvectors corresponding to the r largest eigenvalues of XX' . Therefore, we have $(1/NT)XX'\hat{F} = \hat{F}V_{NT}$ and $(1/NT)XX'\hat{F}V_{NT}^{-1} = \hat{F}$. Let $\delta_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$. Using $X = F\Lambda' + e$ gives:

$$\frac{1}{NT}(F\Lambda'\Lambda F' + F\Lambda'e' + e\Lambda F' + ee')\hat{F}V_{NT}^{-1} = \hat{F} \quad (\text{A.1})$$

Using the fact that $H = (\Lambda'\Lambda/N)(F'\hat{F}/T)V_{NT}^{-1}$ yields:

$$\hat{F} - FH = \frac{1}{NT} \left(F\Lambda'e'\hat{F} + e\Lambda F'\hat{F} + ee'\hat{F} \right) V_{NT}^{-1} \quad (\text{A.2})$$

$$\hat{f}_t - H'f_t = V_{NT}^{-1} \left(\frac{1}{T} \sum_{s=1}^T \hat{f}_s \gamma_N(s, t) + \frac{1}{T} \sum_{s=1}^T \hat{f}_s \zeta_{st} + \frac{1}{T} \sum_{s=1}^T \hat{f}_s \eta_{st} + \frac{1}{T} \sum_{s=1}^T \hat{f}_s \xi_{st} \right) \quad (\text{A.3})$$

where $\zeta_{st} = \frac{e'_s e_t}{N} - \gamma_N(s, t)$, $\eta_{st} = f'_s \Lambda' e_t / N$, and $\xi_{st} = f'_t \Lambda' e_s / N$. Before we prove Theorem 1 we present three lemmas, first two of which are due to Bai (2003) and are stated only for convenience:

Lemma 1 (Lemma A.1 of Bai, 2003): Under Assumptions 1 - 4,

$$\frac{1}{T} \sum_{t=1}^T \|\hat{f}_t - H'f_t\|^2 = O_p \left(\frac{1}{\delta_{NT}^2} \right)$$

Lemma 2 (Lemma B.2 of Bai, 2003): Under Assumptions 1 - 6,

$$\frac{1}{T} (\hat{F} - FH)'F = O_p \left(\frac{1}{\delta_{NT}^2} \right)$$

Lemma 3: Under Assumptions 1 - 6 and 8(a), for π_2 satisfies $0 < \pi_1 \leq \pi_2 < 1$,

$$\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{[\pi T]} \sum_{t=1}^{[\pi T]} (\hat{f}_t - H'f_t) f_t' \right\| = O_p \left(\frac{1}{\delta_{NT}^2} \right)$$

$$\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T - [\pi T]} \sum_{t=[\pi T]+1}^T (\hat{f}_t - H'f_t) f_t' \right\| = O_p \left(\frac{1}{\delta_{NT}^2} \right)$$

Proof of Lemma 3: It follows from (A.3) that

$$\begin{aligned}
& \sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T} \sum_{t=1}^{\lfloor \pi T \rfloor} (\hat{f}_t - H' f_t) f_t' \right\| \\
&= \sup_{\pi \in [\pi_1, \pi_2]} \left\| V_{NT}^{-1} \left[\frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T \hat{f}_s f_t' \gamma_N(s, t) + \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T \hat{f}_s f_t' \zeta_{st} + \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T \hat{f}_s f_t' \eta_{st} + \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T \hat{f}_s f_t' \xi_{st} \right] \right\| \\
&\leq \left(\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T \hat{f}_s f_t' \gamma_N(s, t) \right\| + \sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T \hat{f}_s f_t' \zeta_{st} \right\| + \right. \\
&\quad \left. \sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T \hat{f}_s f_t' \eta_{st} \right\| + \sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T \hat{f}_s f_t' \xi_{st} \right\| \right) \|V_{NT}^{-1}\| \\
&= I + II + III + IV
\end{aligned}$$

Term I can be expressed as

$$\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T (\hat{f}_s - H' f_s) f_t' \gamma_N(s, t) + \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T H' f_s f_t' \gamma_N(s, t) \right\| \quad (\text{A.4})$$

The first term in (A.4) is bounded by

$$\begin{aligned}
& \sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T (\hat{f}_s - H' f_s) f_t' \gamma_N(s, t) \right\| \\
&\leq \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s - H' f_s\|^2 \right)^{\frac{1}{2}} \sup_{\pi \in [\pi_1, \pi_2]} \left(\frac{1}{T} \sum_{s=1}^T \left\| \frac{1}{T} \sum_{t=1}^{\lfloor \pi T \rfloor} f_t' \gamma_N(s, t) \right\|^2 \right)^{\frac{1}{2}} \\
&\leq \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s - H' f_s\|^2 \right)^{\frac{1}{2}} \sup_{\pi \in [\pi_1, \pi_2]} \left(\frac{1}{T} \sum_{s=1}^T \frac{1}{T} \sum_{t=1}^{\lfloor \pi T \rfloor} \|f_t\|^2 \frac{1}{T} \sum_{t=1}^{\lfloor \pi T \rfloor} |\gamma_N(s, t)|^2 \right)^{\frac{1}{2}} \\
&= \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s - H' f_s\|^2 \right)^{\frac{1}{2}} \sup_{\pi \in [\pi_1, \pi_2]} \frac{1}{\sqrt{T}} \left[\left(\frac{1}{T} \sum_{t=1}^{\lfloor \pi T \rfloor} \|f_t\|^2 \right)^{\frac{1}{2}} \left(\frac{1}{T} \sum_{s=1}^T \sum_{t=1}^{\lfloor \pi T \rfloor} |\gamma_N(s, t)|^2 \right)^{\frac{1}{2}} \right] \\
&\leq \frac{1}{\sqrt{T}} \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s - H' f_s\|^2 \right)^{\frac{1}{2}} \left(\frac{1}{T} \sum_{t=1}^T \|f_t\|^2 \right)^{\frac{1}{2}} \left(\frac{1}{T} \sum_{s=1}^T \sum_{t=1}^T |\gamma_N(s, t)|^2 \right)^{\frac{1}{2}} \\
&= \frac{1}{\sqrt{T}} O_p \left(\frac{1}{\delta_{NT}} \right) O_p(1)
\end{aligned}$$

where the $O_p(1)$ follows from Assumption 1 that $E\|f_t\|^2 \leq M$ and the fact that $\frac{1}{T} \sum_{s=1}^T \sum_{t=1}^T |\gamma_N(s, t)|^2 \leq$

M by Lemma 1(i) of Bai and Ng (2002). For the second term in (A.4), we consider its expectation (excluding H),

$$\begin{aligned}
E \left(\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T f_s f_t' \gamma_N(s, t) \right\| \right) &\leq E \left(\sup_{\pi \in [\pi_1, \pi_2]} \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T \|f_s f_t'\| |\gamma_N(s, t)| \right) \\
&\leq E \left(\frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \|f_s f_t'\| |\gamma_N(s, t)| \right) \\
&\leq \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \left(E \|f_s\|^2 E \|f_t\|^2 \right)^{\frac{1}{2}} |\gamma_N(s, t)| \\
&\leq \frac{M}{T^2} \sum_{t=1}^T \sum_{s=1}^T |\gamma_N(s, t)| \leq \frac{M^2}{T}
\end{aligned}$$

where the last two inequalities use Assumptions 1 and 3(b). Note that H is $O_p(1)$ because it converges to a constant matrix H_0 . Therefore, term I is $O_p\left(\frac{1}{\sqrt{T\delta_{NT}}}\right)$.

Term II can be written as

$$\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T (\hat{f}_s - H' f_s) f_t' \zeta_{st} + \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T H' f_s f_t' \zeta_{st} \right\| \quad (\text{A.5})$$

The first term in (A.5) is bounded by

$$\begin{aligned}
\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T (\hat{f}_s - H' f_s) f_t' \zeta_{st} \right\| &\leq \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s - H' f_s\|^2 \right)^{\frac{1}{2}} \sup_{\pi \in [\pi_1, \pi_2]} \left(\frac{1}{T} \sum_{s=1}^T \left\| \frac{1}{T} \sum_{t=1}^{\lfloor \pi T \rfloor} f_t' \zeta_{st} \right\|^2 \right)^{\frac{1}{2}} \\
&\leq \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s - H' f_s\|^2 \right)^{\frac{1}{2}} \sup_{\pi \in [\pi_1, \pi_2]} \left(\frac{1}{T} \sum_{s=1}^T \frac{1}{T} \sum_{t=1}^{\lfloor \pi T \rfloor} \|f_t\|^2 \frac{1}{T} \sum_{t=1}^{\lfloor \pi T \rfloor} |\zeta_{st}|^2 \right)^{\frac{1}{2}} \\
&\leq \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s - H' f_s\|^2 \right)^{\frac{1}{2}} \sup_{\pi \in [\pi_1, \pi_2]} \left[\left(\frac{1}{T} \sum_{t=1}^{\lfloor \pi T \rfloor} \|f_t\|^2 \right)^{\frac{1}{2}} \left(\frac{1}{T^2} \sum_{s=1}^T \sum_{t=1}^{\lfloor \pi T \rfloor} |\zeta_{st}|^2 \right)^{\frac{1}{2}} \right] \\
&\leq \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s - H' f_s\|^2 \right)^{\frac{1}{2}} \left(\frac{1}{T} \sum_{t=1}^T \|f_t\|^2 \right)^{\frac{1}{2}} \left(\frac{1}{T^2} \sum_{s=1}^T \sum_{t=1}^T |\zeta_{st}|^2 \right)^{\frac{1}{2}} \\
&= O_p\left(\frac{1}{\delta_{NT}}\right) O_p(1) O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}$$

where $\left(\frac{1}{T^2} \sum_{s=1}^T \sum_{t=1}^T |\zeta_{st}|^2\right)^{\frac{1}{2}} = O_p\left(\frac{1}{\sqrt{N}}\right)$ follows from the fact that

$$\frac{1}{T^2} \sum_{s=1}^T \sum_{t=1}^T |\zeta_{st}|^2 = \frac{1}{NT^2} \sum_{s=1}^T \sum_{t=1}^T \left(\frac{1}{\sqrt{N}} \sum_{k=1}^N [e_{ks} e_{kt} - E(e_{ks} e_{kt})] \right)^2 = O_p\left(\frac{1}{N}\right)$$

by Assumption 3(e). For the second term in (A.5),

$$\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T H' f_s f_t' \zeta_{st} \right\| = \sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T H' f_s f_t' \frac{1}{N} \sum_{k=1}^N [e_{ks} e_{kt} - E(e_{ks} e_{kt})] \right\|$$

Let $\frac{1}{\sqrt{NT}} \sum_{s=1}^T \sum_{k=1}^N f_s [e_{ks} e_{kt} - E(e_{ks} e_{kt})] \equiv z_t$, so the above equation reduces to

$$\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{\sqrt{NT^3}} \sum_{t=1}^{\lfloor \pi T \rfloor} H' z_t f_t' \right\| \leq \frac{1}{\sqrt{NT}} \left(\frac{1}{T} \sum_{t=1}^T \|z_t\|^2 \right)^{\frac{1}{2}} \left(\frac{1}{T} \sum_{t=1}^T \|f_t'\|^2 \right)^{\frac{1}{2}} \|H\| = \frac{1}{\sqrt{NT}} O_p(1)$$

where $\frac{1}{T} \sum_{t=1}^T \|z_t\|^2 = O_p(1)$ follows from Assumption 6(a) $E\|z_t\| = E\left\| \frac{1}{\sqrt{NT}} \sum_{s=1}^T \sum_{k=1}^N f_s [e_{ks} e_{kt} - E(e_{ks} e_{kt})] \right\|^2 \leq M$ for all t . Thus term II is $O_p\left(\frac{1}{\delta_{NT}\sqrt{N}}\right)$.

Term III can be written as

$$\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T (\hat{f}_s - H' f_s) f_t' \eta_{st} + \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T H' f_s f_t' \eta_{st} \right\| \quad (\text{A.6})$$

The first term in (A.6) is bounded by

$$\begin{aligned} & \sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T (\hat{f}_s - H' f_s) f_t' \eta_{st} \right\| \\ & \leq \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s - H' f_s\|^2 \right)^{\frac{1}{2}} \sup_{\pi \in [\pi_1, \pi_2]} \left(\frac{1}{T} \sum_{s=1}^T \left\| \frac{1}{T} \sum_{t=1}^{\lfloor \pi T \rfloor} f_t' \eta_{st} \right\|^2 \right)^{\frac{1}{2}} \\ & \leq \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s - H' f_s\|^2 \right)^{\frac{1}{2}} \sup_{\pi \in [\pi_1, \pi_2]} \left(\frac{1}{T} \sum_{s=1}^T \left\| \frac{1}{T} \sum_{t=1}^{\lfloor \pi T \rfloor} f_t' \frac{1}{N} \sum_{k=1}^N f_s' \lambda_k e_{kt} \right\|^2 \right)^{\frac{1}{2}} \\ & \leq \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s - H' f_s\|^2 \right)^{\frac{1}{2}} \left[\frac{1}{T} \sum_{s=1}^T \left(\frac{1}{T} \sum_{t=1}^T \left\| f_t' \frac{1}{N} \sum_{k=1}^N f_s' \lambda_k e_{kt} \right\|^2 \right)^{27} \right]^{\frac{1}{2}} \\ & \leq \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s - H' f_s\|^2 \right)^{\frac{1}{2}} \left[\frac{1}{T} \sum_{s=1}^T \left(\frac{1}{T} \sum_{t=1}^T \|f_t\|^2 \frac{1}{T} \sum_{t=1}^T \left| \frac{1}{N} \sum_{k=1}^N f_s' \lambda_k e_{kt} \right|^2 \right) \right]^{\frac{1}{2}} \\ & \leq \frac{1}{\sqrt{N}} \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s - H' f_s\|^2 \right)^{\frac{1}{2}} \left[\frac{1}{T} \sum_{t=1}^T \|f_t\|^2 \frac{1}{T} \sum_{s=1}^T \|f_s\|^2 \frac{1}{T} \sum_{t=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{k=1}^N \lambda_k e_{kt} \right\|^2 \right]^{\frac{1}{2}} \\ & = \frac{1}{\sqrt{N}} O_p\left(\frac{1}{\delta_{NT}}\right) O_p(1) \end{aligned}$$

where the last equation uses Assumptions 1 and 6(c). The second term in (A.6) is bounded by

$$\begin{aligned}
\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T H' f_s f_t' \eta_{st} \right\| &= \sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T H' f_s \frac{1}{N} \sum_{k=1}^N f_s' \lambda_k e_{kt} f_t' \right\| \\
&= \sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T} \sum_{s=1}^T H' f_s f_s' \frac{1}{NT} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{k=1}^N \lambda_k e_{kt} f_t' \right\| \\
&\leq \|H\| \left\| \frac{1}{T} \sum_{s=1}^T f_s f_s' \right\| \sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{NT} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{k=1}^N \lambda_k f_t' e_{kt} \right\| \\
&= O_p \left(\frac{1}{\sqrt{NT}} \right)
\end{aligned}$$

where the last equation follows from Assumption 8(a). Thus, term III is $O_p \left(\frac{1}{\delta_{NT} \sqrt{N}} \right)$.

For term IV, we rewrite it as

$$\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T \hat{f}_s f_t' \xi_{st} \right\| \leq \sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T (\hat{f}_s - H' f_s) f_t' \xi_{st} + \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T H' f_s f_t' \xi_{st} \right\| \quad (\text{A.7})$$

The first term in (A.7) is bounded by

$$\begin{aligned}
&\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T (\hat{f}_s - H' f_s) f_t' \xi_{st} \right\| \\
&\leq \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s - H' f_s\|^2 \right)^{\frac{1}{2}} \sup_{\pi \in [\pi_1, \pi_2]} \left(\frac{1}{T} \sum_{s=1}^T \left\| \frac{1}{T} \sum_{t=1}^{\lfloor \pi T \rfloor} f_t' \xi_{st} \right\|^2 \right)^{\frac{1}{2}} \\
&\leq \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s - H' f_s\|^2 \right)^{\frac{1}{2}} \sup_{\pi \in [\pi_1, \pi_2]} \left(\frac{1}{T} \sum_{s=1}^T \left\| \frac{1}{T} \sum_{t=1}^{\lfloor \pi T \rfloor} f_t' \left(\frac{1}{N} \sum_{k=1}^N \lambda_k' f_t e_{ks} \right) \right\|^2 \right)^{\frac{1}{2}} \\
&\leq \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s - H' f_s\|^2 \right)^{\frac{1}{2}} \sup_{\pi \in [\pi_1, \pi_2]} \left(\frac{1}{T} \sum_{s=1}^T \left\| \left(\frac{1}{T} \sum_{t=1}^{\lfloor \pi T \rfloor} f_t' f_t \right) \left(\frac{1}{N} \sum_{k=1}^N \lambda_k' e_{ks} \right) \right\|^2 \right)^{\frac{1}{2}} \\
&\leq \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s - H' f_s\|^2 \right)^{\frac{1}{2}} \sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T} \sum_{t=1}^{\lfloor \pi T \rfloor} f_t' f_t \right\| \frac{1}{\sqrt{N}} \left(\frac{1}{T} \sum_{s=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{k=1}^N \lambda_k' e_{ks} \right\|^2 \right)^{\frac{1}{2}} \\
&\leq \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s - H' f_s\|^2 \right)^{\frac{1}{2}} \frac{1}{T} \sum_{t=1}^T \|f_t'\|^2 \frac{1}{\sqrt{N}} \left(\frac{1}{T} \sum_{s=1}^T \left\| \frac{1}{\sqrt{N}} \sum_{k=1}^N \lambda_k' e_{ks} \right\|^2 \right)^{\frac{1}{2}} = O_p \left(\frac{1}{\delta_{NT} \sqrt{N}} \right)
\end{aligned}$$

For the second term in (A.7), we have

$$\begin{aligned}
\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T H' f_s f_t' \xi_{st} \right\| &= \sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T^2} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{s=1}^T H' f_s f_t' \left(\frac{1}{N} \sum_{k=1}^N \lambda_k' f_t e_{ks} \right) \right\| \\
&= \sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{NT} \sum_{s=1}^T \sum_{k=1}^N H' f_s \lambda_k' e_{ks} \frac{1}{T} \sum_{t=1}^{\lfloor \pi T \rfloor} f_t' f_t \right\| \\
&\leq \left\| \frac{1}{NT} \sum_{s=1}^T \sum_{k=1}^N H' f_s \lambda_k' e_{ks} \right\| \left\| \frac{1}{T} \sum_{t=1}^T \|f_t'\|^2 \right\| = O_p \left(\frac{1}{\sqrt{NT}} \right)
\end{aligned}$$

where the last equation uses Assumption 6(b). Therefore, term IV is $O_p \left(\frac{1}{\delta_{NT} \sqrt{N}} \right)$.

Combining the above results the sum $I + II + III + IV$ is $O_p \left(\frac{1}{\delta_{NT}^2} \right)$. ■

Proof of Theorem 1: It is sufficient to prove that

$$V \equiv \sup_{\pi \in [\pi_1, \pi_2]} \sqrt{T} \left\| \text{vech} \left(\frac{\sum_{t=1}^{\lfloor \pi T \rfloor} \hat{f}_t \hat{f}_t'}{\lfloor \pi T \rfloor} - \frac{\sum_{t=\lfloor \pi T \rfloor+1}^T \hat{f}_t \hat{f}_t'}{T - \lfloor \pi T \rfloor} \right) - \text{vech} \left(\frac{\sum_{t=1}^{\lfloor \pi T \rfloor} H' f_t f_t' H}{\lfloor \pi T \rfloor} - \frac{\sum_{t=\lfloor \pi T \rfloor+1}^T H' f_t f_t' H}{T - \lfloor \pi T \rfloor} \right) \right\|$$

and

$$\begin{aligned}
VI \equiv & \sup_{\pi \in [\pi_1, \pi_2]} \sqrt{T} \left\| \text{vech} \left(\frac{\sum_{t=1}^{\lfloor \pi T \rfloor} H' f_t f_t' H}{\lfloor \pi T \rfloor} - \frac{\sum_{t=\lfloor \pi T \rfloor+1}^T H' f_t f_t' H}{T - \lfloor \pi T \rfloor} \right) \right. \\
& \left. - \text{vech} \left(\frac{\sum_{t=1}^{\lfloor \pi T \rfloor} H_0' f_t f_t' H_0}{\lfloor \pi T \rfloor} - \frac{\sum_{t=\lfloor \pi T \rfloor+1}^T H_0' f_t f_t' H_0}{T - \lfloor \pi T \rfloor} \right) \right\|
\end{aligned}$$

are both $o_p(1)$. Term V is bounded by

$$V \leq \sup_{\pi \in [\pi_1, \pi_2]} \sqrt{T} \left\| \text{vech} \left(\frac{\sum_{t=1}^{\lfloor \pi T \rfloor} \hat{f}_t \hat{f}_t' - H' f_t f_t' H}{\lfloor \pi T \rfloor} \right) \right\| + \sup_{\pi \in [\pi_1, \pi_2]} \sqrt{T} \left\| \text{vech} \left(\frac{\sum_{t=\lfloor \pi T \rfloor+1}^T \hat{f}_t \hat{f}_t' - H' f_t f_t' H}{T - \lfloor \pi T \rfloor} \right) \right\|$$

To save space, we will only prove that the first term in the above inequality is $o_p(1)$, because the negligibility of the second term can be proved in a similar way. The first term is bounded by

$$\begin{aligned}
& \sup_{\pi \in [\pi_1, \pi_2]} \sqrt{T} \left\| \text{vech} \left(\frac{\sum_{t=1}^{\lfloor \pi T \rfloor} \hat{f}_t (\hat{f}_t' - f_t' H) + (\hat{f}_t - H' f_t) f_t' H}{\lfloor \pi T \rfloor} \right) \right\| \\
&= \sup_{\pi \in [\pi_1, \pi_2]} \sqrt{T} \left\| \text{vech} \left(\frac{\sum_{t=1}^{\lfloor \pi T \rfloor} (\hat{f}_t - H' f_t) (\hat{f}_t' - f_t' H) + H' f_t (\hat{f}_t' - f_t' H) + (\hat{f}_t - H' f_t) f_t' H}{\lfloor \pi T \rfloor} \right) \right\| \\
&\leq \sup_{\pi \in [\pi_1, \pi_2]} \sqrt{T} \left(\left\| \frac{\sum_{t=1}^{\lfloor \pi T \rfloor} (\hat{f}_t - H' f_t) (\hat{f}_t' - f_t' H)}{\lfloor \pi T \rfloor} \right\| + 2 \left\| \frac{\sum_{t=1}^{\lfloor \pi T \rfloor} H' f_t (\hat{f}_t' - f_t' H)}{\lfloor \pi T \rfloor} \right\| \right)
\end{aligned}$$

Note that

$$\sup_{\pi \in [\pi_1, \pi_2]} \sqrt{T} \left\| \frac{\sum_{t=1}^{\lfloor \pi T \rfloor} (\hat{f}_t - H' f_t)(\hat{f}'_t - f'_t H)}{\lfloor \pi T \rfloor} \right\| \leq \sup_{\pi \in [\pi_1, \pi_2]} \frac{\sqrt{T} \sum_{t=1}^{\lfloor \pi T \rfloor} \|\hat{f}_t - H' f_t\|^2}{\lfloor \pi T \rfloor} \leq \frac{\sqrt{T} \sum_{t=1}^T \|\hat{f}_t - H' f_t\|^2}{\lfloor \pi_1 T \rfloor} = O_p \left(\frac{\sqrt{T}}{\delta_{NT}^2} \right)$$

where the last equation follows from Lemma 1. Also, Lemma 3 implies that

$$\sup_{\pi \in [\pi_1, \pi_2]} \sqrt{T} \left\| \frac{1}{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} H' f_t (\hat{f}'_t - f'_t H) \right\| = O_p \left(\frac{\sqrt{T}}{\delta_{NT}^2} \right)$$

Since $\frac{\sqrt{T}}{N} \rightarrow 0$ as $N, T \rightarrow \infty$, term V is $o_p(1)$. For term VI, we can bound it by

$$\begin{aligned} & \sup_{\pi \in [\pi_1, \pi_2]} \sqrt{T} \left\| \frac{\sum_{t=1}^{\lfloor \pi T \rfloor} H' f_t f'_t H - H'_0 f_t f'_t H_0}{\lfloor \pi T \rfloor} - \frac{\sum_{t=\lfloor \pi T \rfloor + 1}^T H' f_t f'_t H - H'_0 f_t f'_t H_0}{T - \lfloor \pi T \rfloor} \right\| \\ &= \sup_{\pi \in [\pi_1, \pi_2]} \sqrt{T} \left\| \frac{\sum_{t=1}^{\lfloor \pi T \rfloor} (H' - H'_0) f_t f'_t H + H'_0 f_t f'_t (H - H_0)}{\lfloor \pi T \rfloor} - \frac{\sum_{t=\lfloor \pi T \rfloor + 1}^T (H' - H'_0) f_t f'_t H + H'_0 f_t f'_t (H - H_0)}{T - \lfloor \pi T \rfloor} \right\| \\ &= \sup_{\pi \in [\pi_1, \pi_2]} \sqrt{T} \left\| (H' - H'_0) \left(\frac{\sum_{t=1}^{\lfloor \pi T \rfloor} f_t f'_t}{\lfloor \pi T \rfloor} - \frac{\sum_{t=\lfloor \pi T \rfloor + 1}^T f_t f'_t}{T - \lfloor \pi T \rfloor} \right) H + H_0 \left(\frac{\sum_{t=1}^{\lfloor \pi T \rfloor} f_t f'_t}{\lfloor \pi T \rfloor} - \frac{\sum_{t=\lfloor \pi T \rfloor + 1}^T f_t f'_t}{T - \lfloor \pi T \rfloor} \right) (H - H_0) \right\| \\ &\leq \sqrt{T} \|H - H_0\| \sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{\sum_{t=1}^{\lfloor \pi T \rfloor} f_t f'_t}{\lfloor \pi T \rfloor} - \frac{\sum_{t=\lfloor \pi T \rfloor + 1}^T f_t f'_t}{T - \lfloor \pi T \rfloor} \right\| (\|H\| + \|H_0\|) \\ &= \sqrt{T} o_p(1) O_p \left(\frac{1}{\sqrt{T}} \right) O_p(1) \end{aligned}$$

where the last equation uses Assumption 8(b) and the fact that $\|H - H_0\| \xrightarrow{p} 0$ implied by Assumptions 1 - 4 and 7 (See Bai, 2003) . Thus, term VI is $o_p(1)$. ■

Before proving Theorem 2 we present and prove Lemmas 4–8.

Lemma 4: Under Assumption 3(b), for all N and T , there exists a constant $M_1 < \infty$ such that $\frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \gamma_N^4(s, t) \leq M_1$. Under Assumption 5(b), for all N and T , there exists a constant $M_2 < \infty$ such that $\sum_{s=1}^T \gamma_N^2(s, t) \leq M_2$ for $t = 1, 2, \dots, T$.

Proof of Lemma 4: Let $\rho(s, t) \equiv \gamma_N(s, t) / \sqrt{\gamma_N(s, s) \gamma_N(t, t)}$, so $|\rho(s, t)| \leq 1$. Since $\gamma_N(s, s) \leq M$ for every s ,

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \gamma_N^4(s, t) &= \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \gamma_N^2(s, s) \gamma_N^2(t, t) \rho^4(s, t) \\ &\leq \frac{M^3}{T} \sum_{t=1}^T \sum_{s=1}^T |\gamma_N(s, s) \gamma_N(t, t)|^{\frac{1}{2}} |\rho(s, t)| \\ &= \frac{M^3}{T} \sum_{t=1}^T \sum_{s=1}^T |\gamma_N(s, t)| \leq M^4 \end{aligned}$$

where the last inequality follows from Assumption 3(b). Also, we note that for each t ,

$$\begin{aligned}\sum_{s=1}^T \gamma_N^2(s, t) &= \sum_{s=1}^T \gamma_N(s, s) \gamma_N(t, t) \rho^2(s, t) \\ &\leq M \sum_{s=1}^T |\gamma_N(s, s) \gamma_N(t, t)|^{\frac{1}{2}} |\rho(s, t)| \\ &= M \sum_{s=1}^T |\gamma_N(s, t)| \leq M^2\end{aligned}$$

where the last inequality follows from Assumption 5(a). ■

Lemma 5:

(i) Under Assumptions 1 - 4,

$$\frac{1}{T} \sum_{t=1}^T \|\hat{f}_t - H' f_t\|^4 = O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{N^2}\right)$$

(ii) Under Assumptions 1 - 5,

$$\frac{1}{T} \sum_{t=1}^T \|\hat{f}_t - H' f_t\|^4 = O_p\left(\frac{1}{\delta_{NT}^4}\right)$$

(iii) Under Assumptions 1 - 4.

$$\frac{1}{T} \sum_{t=1}^T \|\hat{f}_t\|^4 = O_p(1)$$

Proof of Lemma 5: Equation A.3 implies that

$$\frac{1}{T} \sum_{t=1}^T \|\hat{f}_t - H' f_t\|^4 \leq 64 \|V_{NT}^{-1}\|^4 \frac{1}{T} \sum_{t=1}^T (a_t + b_t + c_t + d_t)$$

where

$$a_t = \frac{1}{T^4} \left\| \sum_{s=1}^T \hat{f}_s \gamma_N(s, t) \right\|^4, \quad b_t = \frac{1}{T^4} \left\| \sum_{s=1}^T \hat{f}_s \zeta_{st} \right\|^4, \quad c_t = \frac{1}{T^4} \left\| \sum_{s=1}^T \hat{f}_s \eta_{st} \right\|^4, \quad d_t = \frac{1}{T^4} \left\| \sum_{s=1}^T \hat{f}_s \xi_{st} \right\|^4.$$

First,

$$\frac{1}{T} \sum_{t=1}^T a_t \leq \frac{1}{T} \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s\|^2 \right)^2 \sum_{t=1}^T \left(\frac{1}{T} \sum_{s=1}^T \gamma_N^2(s, t) \right)^2$$

Under Assumption 3(b), the above inequality is bounded by

$$\frac{1}{T} \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s\|^2 \right)^2 \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \gamma_N^A(s, t) = O_p \left(\frac{1}{T} \right)$$

because $\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s\|^2 = O_p(1)$ and $\frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \gamma_N^A(s, t) \leq M_1$ by Lemma 4. However, under Assumption 5(a), which is stronger than Assumption 3(b), we can derive a sharper bound

$$\frac{1}{T} \sum_{t=1}^T a_t \leq \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s\|^2 \right)^2 \frac{M_2^2}{T^2} = O_p \left(\frac{1}{T^2} \right)$$

The difference between (i) and (ii) of Lemma 5 is due to the different bounds of $\frac{1}{T} \sum_{t=1}^T a_t$. The rest of the proofs of parts (i) and (ii) are the same.

Second,

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T b_t &\leq \frac{1}{T^5} \sum_{t=1}^T \left\| \sum_{s=1}^T \hat{f}_s \zeta_{st} \right\|^4 \\ &= \frac{1}{T^5} \sum_{t=1}^T \left(\sum_{s=1}^T \sum_{u=1}^T \hat{f}_s' \hat{f}_u \zeta_{st} \zeta_{ut} \right)^2 \\ &\leq \frac{1}{T^5} \sum_{t=1}^T \left(\sum_{s=1}^T \sum_{u=1}^T (\hat{f}_s' \hat{f}_u)^2 \right) \left(\sum_{s=1}^T \sum_{u=1}^T \zeta_{st}^2 \zeta_{ut}^2 \right) \\ &\leq \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s\|^2 \right)^2 \frac{1}{T} \sum_{t=1}^T \left(\frac{1}{T^2} \sum_{s=1}^T \sum_{u=1}^T \zeta_{st}^2 \zeta_{ut}^2 \right) \end{aligned}$$

Since $E \zeta_{st}^2 \zeta_{ut}^2 \leq \max_{s,t} E |\zeta_{st}|^4$ and $E |\zeta_{st}|^4 = N^{-2} E \left| N^{-1/2} \sum_{i=1}^N [e_{is} e_{it} - E(e_{is} e_{it})] \right|^4 \leq N^{-2} M$, by Assumption 3(e), $\frac{1}{T} \sum_{t=1}^T b_t = O_p \left(\frac{1}{N^2} \right)$.

Third,

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T c_t &\leq \frac{1}{T^5} \sum_{t=1}^T \left\| \sum_{s=1}^T \hat{f}_s \eta_{st} \right\|^4 \\ &\leq \frac{1}{T^5} \sum_{t=1}^T \left\| \sum_{s=1}^T \hat{f}_s f_s' \Lambda' e_t / N \right\|^4 \\ &\leq \frac{1}{T} \sum_{t=1}^T \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s\|^2 \right)^2 \left(\frac{1}{T} \sum_{s=1}^T \|f_s\|^2 \right)^2 \left(\frac{1}{N^2} \left\| \frac{\Lambda' e_t}{\sqrt{N}} \right\|^4 \right) = O_p \left(\frac{1}{N^2} \right) \end{aligned}$$

by Assumptions 1 and 6(c).

Finally,

$$\begin{aligned}
\frac{1}{T} \sum_{t=1}^T d_t &\leq \frac{1}{T^5} \sum_{t=1}^T \left\| \sum_{s=1}^T \hat{f}_s \xi_{st} \right\|^4 \\
&= \frac{1}{T^5} \sum_{t=1}^T \left\| \sum_{s=1}^T \hat{f}_s e'_s \Lambda f_t / N \right\|^4 \\
&\leq \frac{1}{T^5} \left\| \sum_{s=1}^T \hat{f}_s e'_s \Lambda / N \right\|^4 \sum_{t=1}^T \|f_t\|^4 \\
&\leq \left(\frac{1}{T} \sum_{s=1}^T \|\hat{f}_s\|^2 \right)^2 \left(\frac{1}{TN} \sum_{s=1}^T \left\| \frac{e'_s \Lambda}{\sqrt{N}} \right\|^2 \right)^2 \frac{1}{T} \sum_{t=1}^T \|f_t\|^4 = O_p \left(\frac{1}{N^2} \right)
\end{aligned}$$

by Assumptions 1 and 6(c). Thus, under Assumptions 1 - 4, $\frac{1}{T} \sum_{t=1}^T a_t + b_t + c_t + d_t = O_p \left(\frac{1}{T} \right) + O_p \left(\frac{1}{N^2} \right)$, while under Assumptions 1 - 5, $\frac{1}{T} \sum_{t=1}^T a_t + b_t + c_t + d_t = O_p \left(\frac{1}{T^2} \right) + O_p \left(\frac{1}{N^2} \right)$. For part (iii), rewrite $\frac{1}{T} \sum_{t=1}^T \|\hat{f}_t\|^4$ as

$$\frac{1}{T} \sum_{t=1}^T \|\hat{f}_t - H' f_t + H' f_t\|^4 \leq \frac{8}{T} \left(\sum_{t=1}^T \|\hat{f}_t - H' f_t\|^4 + \sum_{t=1}^T \|H' f_t\|^4 \right) = o_p(1) + O_p(1)$$

by the result in part (i) and Assumption 1. ■

Lemma 6: Under Assumptions 1-7, $H - H_0 = O_p \left(\frac{1}{\delta_{NT}} \right)$.

Proof of Lemma 6: By the identity $(1/NT)XX'\hat{F} \equiv \hat{F}V_{NT}$, we have

$$\left(\frac{\Lambda' \Lambda}{N} \right)^{\frac{1}{2}} \frac{1}{T} F' \left(\frac{XX'}{NT} \right) \hat{F} = \left(\frac{\Lambda' \Lambda}{N} \right)^{\frac{1}{2}} \left(\frac{F' \hat{F}}{T} \right) V_{NT}$$

Note that $V_{NT} - V \xrightarrow{p} 0$ by Lemma A.3 of Bai 2003. We will prove that $V_{NT} - V = O_p \left(\frac{1}{\delta_{NT}} \right)$. Substituting $X = F\Lambda' + e$ into the above equation, we get

$$\left(\frac{\Lambda' \Lambda}{N} \right)^{\frac{1}{2}} \left(\frac{F' F}{T} \right) \left(\frac{\Lambda' \Lambda}{N} \right) \left(\frac{F' \hat{F}}{T} \right) + d_{NT} = \left(\frac{\Lambda' \Lambda}{N} \right)^{\frac{1}{2}} \left(\frac{F' \hat{F}}{T} \right) V_{NT}$$

where

$$\begin{aligned}
d_{NT} &= \left(\frac{\Lambda' \Lambda}{N} \right)^{\frac{1}{2}} \frac{1}{NT^2} F' (F\Lambda' e' + F' e \Lambda F' + F' e e') \hat{F} \\
&= \left(\frac{\Lambda' \Lambda}{N} \right)^{\frac{1}{2}} \frac{1}{T} F' (\hat{F} - FH) V_{NT} = O_p \left(\frac{1}{\delta_{NT}^2} \right)
\end{aligned}$$

by equation (A.2) and Lemma 2. Let

$$B_{NT} = \left(\frac{\Lambda'\Lambda}{N}\right)^{\frac{1}{2}} \left(\frac{F'F}{T}\right) \left(\frac{\Lambda'\Lambda}{N}\right)^{\frac{1}{2}}, \quad B = \Sigma_{\Lambda}^{\frac{1}{2}} \Sigma_F \Sigma_{\Lambda}^{\frac{1}{2}}, \quad R_{NT} = \left(\frac{\Lambda'\Lambda}{N}\right)^{\frac{1}{2}} \left(\frac{F'\hat{F}}{T}\right),$$

where $B_{NT} - B = O_p\left(\frac{1}{\delta_{NT}}\right)$ by Assumptions 1 and 2, and R_{NT} is $O_p(1)$ and invertible, because $\frac{F'\hat{F}}{T}$ is $O_p(1)$ and converges to some non-singular matrix by Proposition 1 of Bai (2003). Now we have

$$\left(B_{NT} + d_{NT}R_{NT}^{-1}\right)R_{NT} = R_{NT}V_{NT}$$

Let V_{NT}^* be a diagonal matrix consisting of the diagonal elements of $R'_{NT}R_{NT}$. Define $\Upsilon_{NT} \equiv R_{NT}V_{NT}^{*\frac{-1}{2}}$, so Υ_{NT} is $O_p(1)$ and contains eigenvectors of $B_{NT} + d_{NT}R_{NT}^{-1}$, i.e. $\left(B_{NT} + d_{NT}R_{NT}^{-1}\right)\Upsilon_{NT} = \Upsilon_{NT}V_{NT}$. Note that $d_{NT}R_{NT}^{-1} = O_p\left(\frac{1}{\delta_{NT}^2}\right)$ by the facts that $R_{NT} = O_p(1)$ and invertible and that $d_{NT} = O_p\left(\frac{1}{\delta_{NT}^2}\right)$, so $B_{NT} + d_{NT}R_{NT}^{-1} - B = O_p\left(\frac{1}{\delta_{NT}}\right)$. Assumption 7 implies that the eigenvalues of B are distinct, so $V_{NT} - V = O_p\left(\frac{1}{\delta_{NT}}\right)$ and $\Upsilon_{NT} - \Upsilon = O_p\left(\frac{1}{\delta_{NT}}\right)$. This is because both eigenvalues and eigenvectors are continuously differentiable functions for matrices with distinct eigenvalues. By Equations (A.1) and (A.2),

$$\begin{aligned} \frac{1}{NT^2}\hat{F}'F\Lambda'\Lambda F'\hat{F} + \frac{1}{T}\hat{F}'(\hat{F} - FH)V_{NT} &= V_{NT} \\ \frac{\hat{F}'F}{T} \frac{\Lambda'\Lambda}{N} \frac{F'\hat{F}}{T} - V_{NT} &= O_p\left(\frac{1}{\delta_{NT}^2}\right) \end{aligned}$$

where the last equation follows from Lemma 2. Based on the result that $V_{NT} - V = O_p\left(\frac{1}{\delta_{NT}}\right)$,

$$R'_{NT}R_{NT} - V = \frac{\hat{F}'F}{T} \frac{\Lambda'\Lambda}{N} \frac{F'\hat{F}}{T} - V = O_p\left(\frac{1}{\delta_{NT}}\right).$$

Recall that $\text{diag}(R'_{NT}R_{NT}) = V_{NT}^*$. Therefore,

$$V_{NT}^* - V = \text{diag}(R'_{NT}R_{NT} - V) = O_p\left(\frac{1}{\delta_{NT}}\right) \quad (\text{A.8})$$

Using $\Upsilon_{NT} = R_{NT}V_{NT}^{*\frac{-1}{2}}$ and the definition of R_{NT} , we have $\frac{F'\hat{F}}{T} = \left(\frac{\Lambda'\Lambda}{N}\right)^{-\frac{1}{2}} \Upsilon_{NT}V_{NT}^{*\frac{1}{2}}$. Comparing $\frac{F'\hat{F}}{T}$ and its probability limit gives

$$\begin{aligned} \frac{F'\hat{F}}{T} - \Sigma_{\Lambda}^{-\frac{1}{2}}\Upsilon V^{\frac{1}{2}} &= \left[\left(\frac{\Lambda'\Lambda}{N}\right)^{-\frac{1}{2}} - \Sigma_{\Lambda}^{-\frac{1}{2}}\right] \Upsilon_{NT}V_{NT}^{*\frac{1}{2}} + \Sigma_{\Lambda}^{-\frac{1}{2}}(\Upsilon_{NT} - \Upsilon)V_{NT}^{*\frac{1}{2}} + \Sigma_{\Lambda}^{-\frac{1}{2}}\Upsilon\left(V_{NT}^{*\frac{1}{2}} - V^{\frac{1}{2}}\right) \\ &= O_p\left(\frac{1}{\delta_{NT}}\right) \end{aligned}$$

by Assumption 2 and Equation (A.8). Therefore, $H - H_0 = (\Lambda' \Lambda / N)(F' \hat{F} / T) V_{NT}^{-1} - \Sigma_{\Lambda} \text{plim}(F' \hat{F} / T) V^{-1} = O_p\left(\frac{1}{\delta_{NT}}\right)$ by Assumption 2 and the result that $V_{NT} - V = O_p\left(\frac{1}{\delta_{NT}}\right)$. Note that V_{NT} and V are diagonal matrices consisting of positive eigenvalues, so they are invertible. ■

Lemma 7: (i) Under Assumptions 1 - 4,

$$\begin{aligned} \sup_{\pi \in [\pi_1, \pi_2]} \sup_{0 \leq j \leq \lfloor \pi T \rfloor - 1} \left\| \hat{\Gamma}_{1,j}(\pi, \hat{F}) - \hat{\Gamma}_{1,j}(\pi, FH) \right\| &= O_p\left(T^{-\frac{1}{4}}\right) + O_p\left(N^{-\frac{1}{2}}\right) \\ \sup_{\pi \in [\pi_1, \pi_2]} \sup_{0 \leq j \leq T - \lfloor \pi T \rfloor - 1} \left\| \hat{\Gamma}_{2,j}(\pi, \hat{F}) - \hat{\Gamma}_{2,j}(\pi, FH) \right\| &= O_p\left(T^{-\frac{1}{4}}\right) + O_p\left(N^{-\frac{1}{2}}\right) \\ \sup_{0 \leq j \leq T - 1} \left\| \hat{\Gamma}_j(\hat{F}) - \hat{\Gamma}_j(FH) \right\| &= O_p\left(T^{-\frac{1}{4}}\right) + O_p\left(N^{-\frac{1}{2}}\right) \end{aligned}$$

(ii) Under Assumptions 1 - 5,

$$\begin{aligned} \sup_{\pi \in [\pi_1, \pi_2]} \sup_{0 \leq j \leq \lfloor \pi T \rfloor - 1} \left\| \hat{\Gamma}_{1,j}(\pi, \hat{F}) - \hat{\Gamma}_{1,j}(\pi, FH) \right\| &= O_p\left(\frac{1}{\delta_{NT}}\right) \\ \sup_{\pi \in [\pi_1, \pi_2]} \sup_{0 \leq j \leq T - \lfloor \pi T \rfloor - 1} \left\| \hat{\Gamma}_{2,j}(\pi, \hat{F}) - \hat{\Gamma}_{2,j}(\pi, FH) \right\| &= O_p\left(\frac{1}{\delta_{NT}}\right) \\ \sup_{0 \leq j \leq T - 1} \left\| \hat{\Gamma}_j(\hat{F}) - \hat{\Gamma}_j(FH) \right\| &= O_p\left(\frac{1}{\delta_{NT}}\right) \end{aligned}$$

Proof: To save space, we will only prove the first equations of part (i) and (ii), because the rest can be proved using a similar argument.

$$\begin{aligned} & \sup_{\pi \in [\pi_1, \pi_2]} \sup_{0 \leq j \leq \lfloor \pi T \rfloor - 1} \left\| \hat{\Gamma}_{1,j}(\pi, \hat{F}) - \hat{\Gamma}_{1,j}(\pi, FH) \right\| \\ = & \sup_{\pi \in [\pi_1, \pi_2]} \sup_{0 \leq j \leq \lfloor \pi T \rfloor - 1} \left\| \frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \text{vech}(\hat{f}_t \hat{f}_t' - I_r) \text{vech}(\hat{f}_{t-j} \hat{f}_{t-j}' - I_r)' \right. \\ & \left. - \frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \text{vech}(H' f_t f_t' H - I_r) \text{vech}(H' f_{t-j} f_{t-j}' H - I_r)' \right\| \\ \leq & \sup_{\pi \in [\pi_1, \pi_2]} \sup_{0 \leq j \leq \lfloor \pi T \rfloor - 1} \left\{ \frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \left\| \text{vech}(\hat{f}_t \hat{f}_t' - I_r) \text{vech}(\hat{f}_{t-j} \hat{f}_{t-j}' - H' f_{t-j} f_{t-j}' H)' \right\| \right. \\ & \left. + \frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \left\| \text{vech}(\hat{f}_t \hat{f}_t' - H' f_t f_t' H) \text{vech}(H' f_{t-j} f_{t-j}' H - I_r)' \right\| \right\} \\ \leq & \sup_{\pi \in [\pi_1, \pi_2]} \sup_{0 \leq j \leq \lfloor \pi T \rfloor - 1} \left\{ \frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \left\| \hat{f}_t \hat{f}_t' \right\| \left\| \hat{f}_{t-j} \hat{f}_{t-j}' - H' f_{t-j} f_{t-j}' H \right\| + \frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} r \left\| \hat{f}_{t-j} \hat{f}_{t-j}' - H' f_{t-j} f_{t-j}' H \right\| \right. \\ & \left. + \frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \left\| \hat{f}_t \hat{f}_t' - H' f_t f_t' H \right\| \left\| H' f_{t-j} f_{t-j}' H \right\| + \frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} r \left\| \hat{f}_t \hat{f}_t' - H' f_t f_t' H \right\| \right\} \\ = & VII + VIII + IX + X \end{aligned}$$

Term VII is bounded by,

$$\begin{aligned}
& \sup_{\pi \in [\pi_1, \pi_2]} \sup_{0 \leq j \leq \lfloor \pi T \rfloor - 1} \left(\frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \|\hat{f}_t\|^4 \right)^{\frac{1}{2}} \left(\frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \left\| \hat{f}_{t-j}(\hat{f}'_{t-j} - f'_{t-j}H) + (\hat{f}_{t-j} - H'f_{t-j})f'_{t-j}H \right\|^2 \right)^{\frac{1}{2}} \\
& \leq \sup_{\pi \in [\pi_1, \pi_2]} \sup_{0 \leq j \leq \lfloor \pi T \rfloor - 1} \left(\frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \|\hat{f}_t\|^4 \right)^{\frac{1}{2}} \left(\frac{2}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \left\| \hat{f}_{t-j}(\hat{f}'_{t-j} - f'_{t-j}H) \right\|^2 \right. \\
& \quad \left. + \frac{2}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \left\| (\hat{f}_{t-j} - H'f_{t-j})f'_{t-j}H \right\|^2 \right)^{\frac{1}{2}} \\
& \leq \sup_{\pi \in [\pi_1, \pi_2]} \sup_{0 \leq j \leq \lfloor \pi T \rfloor - 1} \left(\frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \|\hat{f}_t\|^4 \right)^{\frac{1}{2}} \left[\left(\frac{2}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \|\hat{f}_{t-j}\|^4 \frac{2}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \|\hat{f}'_{t-j} - f'_{t-j}H\|^4 \right)^{\frac{1}{2}} \right. \\
& \quad \left. + \left(\frac{2}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \|\hat{f}_{t-j} - H'f_{t-j}\|^4 \frac{2}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \|f'_{t-j}H\|^4 \right)^{\frac{1}{2}} \right]^{\frac{1}{2}} \\
& \leq \left(\frac{1}{\lfloor \pi_1 T \rfloor} \sum_{t=1}^T \|\hat{f}_t\|^4 \right)^{\frac{1}{2}} \left[\left(\frac{2}{\lfloor \pi_1 T \rfloor} \sum_{t=1}^T \|\hat{f}_t\|^4 \frac{2}{\lfloor \pi_1 T \rfloor} \sum_{t=1}^T \|\hat{f}'_t - f'_t H\|^4 \right)^{\frac{1}{2}} \right. \\
& \quad \left. + \left(\frac{2}{\lfloor \pi_1 T \rfloor} \sum_{t=1}^T \|\hat{f}_t - H'f_t\|^4 \frac{2}{\lfloor \pi_1 T \rfloor} \sum_{t=1}^T \|f'_t H\|^4 \right)^{\frac{1}{2}} \right]^{\frac{1}{2}} \\
& = \begin{cases} O_p(T^{-\frac{1}{4}}) + O_p(N^{-\frac{1}{2}}) & \text{by Lemma 5(i) under Assumptions 1 - 4} \\ O_p\left(\frac{1}{\delta_{NT}}\right) & \text{by Lemma 5(ii) under Assumptions 1 - 5} \end{cases}
\end{aligned}$$

where $\frac{1}{T} \sum_{t=1}^T \|\hat{f}_t\|^4 = O_p(1)$ follows from Lemma 5(iii). Using a similar argument, one can prove that terms VIII, IX and X are $O_p(T^{-\frac{1}{4}}) + O_p(N^{-\frac{1}{2}})$ under Assumptions 1 - 4, and $O_p\left(\frac{1}{\delta_{NT}}\right)$ under Assumptions 1 - 5. ■

Lemma 8: Under Assumptions 1 - 7,

$$\begin{aligned}
& \sup_{\pi \in [\pi_1, \pi_2]} \sup_{0 \leq j \leq \lfloor \pi T \rfloor - 1} \left\| \hat{\Gamma}_{1,j}(\pi, FH) - \hat{\Gamma}_{1,j}(\pi, FH_0) \right\| = O_p\left(\frac{1}{\delta_{NT}}\right) \\
& \sup_{\pi \in [\pi_1, \pi_2]} \sup_{0 \leq j \leq T - \lfloor \pi T \rfloor - 1} \left\| \hat{\Gamma}_{2,j}(\pi, FH) - \hat{\Gamma}_{2,j}(\pi, FH_0) \right\| = O_p\left(\frac{1}{\delta_{NT}}\right) \\
& \sup_{0 \leq j \leq T-1} \left\| \hat{\Gamma}_j(FH) - \hat{\Gamma}_j(FH_0) \right\| = O_p\left(\frac{1}{\delta_{NT}}\right)
\end{aligned}$$

Proof: To save space, we will only prove the first equation of Lemma 8, because the proofs of the rests are analogous.

$$\begin{aligned}
& \sup_{\pi \in [\pi_1, \pi_2]} \sup_{0 \leq j \leq \lfloor \pi T \rfloor - 1} \left\| \hat{\Gamma}_{1,j}(\pi, FH) - \hat{\Gamma}_{1,j}(\pi, FH_0) \right\| \\
= & \sup_{\pi \in [\pi_1, \pi_2]} \sup_{0 \leq j \leq \lfloor \pi T \rfloor - 1} \left\| \frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \text{vech}(H' f_t f_t' H - I_r) \text{vech}(H' f_{t-j} f_{t-j}' H - I_r)' \right. \\
& \left. - \frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \text{vech}(H_0' f_t f_t' H_0 - I_r) \text{vech}(H_0' f_{t-j} f_{t-j}' H_0 - I_r)' \right\| \\
\leq & \sup_{\pi \in [\pi_1, \pi_2]} \sup_{0 \leq j \leq \lfloor \pi T \rfloor - 1} \left\{ \frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \left\| \text{vech}(H' f_t f_t' H - I_r) \text{vech}(H' f_{t-j} f_{t-j}' H - H_0' f_{t-j} f_{t-j}' H_0)' \right\| \right. \\
& \left. + \frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \left\| \text{vech}(H' f_t f_t' H - H_0' f_t f_t' H_0) \text{vech}(H_0' f_{t-j} f_{t-j}' H_0 - I_r)' \right\| \right\} \\
\leq & \sup_{\pi \in [\pi_1, \pi_2]} \sup_{0 \leq j \leq \lfloor \pi T \rfloor - 1} \left\{ \frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \left\| H' f_t f_t' H \right\| \left\| H' f_{t-j} f_{t-j}' H - H_0' f_{t-j} f_{t-j}' H_0 \right\| \right. \\
& + \frac{r}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \left\| H' f_{t-j} f_{t-j}' H - H_0' f_{t-j} f_{t-j}' H_0 \right\| + \frac{r}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \left\| H' f_t f_t' H - H_0' f_t f_t' H_0 \right\| \\
& \left. + \frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \left\| H' f_t f_t' H - H_0' f_t f_t' H_0 \right\| \left\| H_0' f_{t-j} f_{t-j}' H_0 \right\| \right\} \\
= & XI + XII + XIII + XIV
\end{aligned}$$

Term XI is bounded by

$$\begin{aligned}
XI & \leq \sup_{\pi \in [\pi_1, \pi_2]} \sup_{0 \leq j \leq \lfloor \pi T \rfloor - 1} \left(\frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \|f_t' H\|^4 \right)^{\frac{1}{2}} \left(\frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \|H' f_{t-j} f_{t-j}' (H - H_0) + (H - H_0)' f_{t-j} f_{t-j}' H_0\|^2 \right)^{\frac{1}{2}} \\
& \leq \sup_{\pi \in [\pi_1, \pi_2]} \sup_{0 \leq j \leq \lfloor \pi T \rfloor - 1} \left(\frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \|f_t' H\|^4 \right)^{\frac{1}{2}} \left[\frac{2}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} (\|H' f_{t-j} f_{t-j}'\|^2 + \|f_{t-j} f_{t-j}' H_0\|^2) \right]^{\frac{1}{2}} \|H - H_0\| \\
& \leq \sup_{\pi \in [\pi_1, \pi_2]} \left(\frac{1}{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} \|f_t' H\|^4 \right)^{\frac{1}{2}} \left[\frac{2}{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} (\|H' f_t f_t'\|^2 + \|f_t f_t' H_0\|^2) \right]^{\frac{1}{2}} \|H - H_0\| \\
& \leq \left(\frac{1}{\lfloor \pi_1 T \rfloor} \sum_{t=1}^T \|f_t' H\|^4 \right)^{\frac{1}{2}} \left[\frac{2}{\lfloor \pi_1 T \rfloor} (\|H'\|^2 + \|H_0\|^2) \sum_{t=1}^T \|f_t\|^4 \right]^{\frac{1}{2}} \|H - H_0\| = O_p(1) O_p\left(\frac{1}{\delta_{NT}}\right)
\end{aligned}$$

by Assumption 1 and Lemma 6. The proofs of terms XII, XIII and XIV are similar to that of term

XI. ■

Proof of Theorem 2: It is enough to show that

$$\sup_{\pi \in [\pi_1, \pi_2]} \left\| \hat{\Omega}_m(\pi, \hat{F}) - \hat{\Omega}_m(\pi, FH_0) \right\| \xrightarrow{P} 0 \quad \text{for } m = 1, 2$$

$$\left\| \hat{\Omega}(\hat{F}) - \hat{\Omega}(FH_0) \right\| \xrightarrow{P} 0$$

We will only prove the first equation when $m = 1$, and the rests can be proved using a similar argument. Note that

$$\begin{aligned} & \sup_{\pi \in [\pi_1, \pi_2]} \left\| \hat{\Omega}_1(\pi, \hat{F}) - \hat{\Omega}_1(\pi, FH_0) \right\| \\ & \leq \sup_{\pi \in [\pi_1, \pi_2]} \left\| \hat{\Omega}_1(\pi, \hat{F}) - \hat{\Omega}_1(\pi, FH) \right\| + \sup_{\pi \in [\pi_1, \pi_2]} \left\| \hat{\Omega}_1(\pi, FH) - \hat{\Omega}_1(\pi, FH_0) \right\| \\ & = XV + XVI \end{aligned}$$

For term XV,

$$\begin{aligned} & \sup_{\pi \in [\pi_1, \pi_2]} \left\| \hat{\Omega}_1(\pi, \hat{F}) - \hat{\Omega}_1(\pi, FH) \right\| \\ & \leq \sup_{\pi \in [\pi_1, \pi_2]} \left\| \hat{\Gamma}_{1,0}(\pi, \hat{F}) + \sum_{j=1}^{\lfloor \pi T \rfloor - 1} k \left(\frac{j}{S_{\lfloor \pi T \rfloor}} \right) \left(\hat{\Gamma}_{1,j}(\pi, \hat{F}) + \hat{\Gamma}_{1,j}(\pi, \hat{F})' \right) \right. \\ & \quad \left. - \hat{\Gamma}_{1,0}(\pi, FH) - \sum_{j=1}^{\lfloor \pi T \rfloor - 1} k \left(\frac{j}{S_{\lfloor \pi T \rfloor}} \right) \left(\hat{\Gamma}_{1,j}(\pi, FH) + \hat{\Gamma}_{1,j}(\pi, FH)' \right) \right\| \end{aligned}$$

Note that $\left| k \left(\frac{j}{S_{\lfloor \pi T \rfloor}} \right) \right| \leq 1$ and $k \left(\frac{j}{S_{\lfloor \pi T \rfloor}} \right) = 0$ if $j > S_{\lfloor \pi T \rfloor}$ for Bartlett and Parzen kernels. Thus,

$$XV \leq \sup_{\pi \in [\pi_1, \pi_2]} \left\| \hat{\Gamma}_{1,0}(\pi, \hat{F}) - \hat{\Gamma}_{1,0}(\pi, FH) \right\| + 2 \sup_{\pi \in [\pi_1, \pi_2]} \sum_{j=1}^{S_{\lfloor \pi T \rfloor}} \left\| \hat{\Gamma}_{1,j}(\pi, \hat{F}) - \hat{\Gamma}_{1,j}(\pi, FH) \right\| \quad (\text{A.9})$$

For Bartlett kernel, the RHS of (A.9) is $O_p \left(\frac{T^{\frac{1}{3}}}{\delta_{NT}} \right)$ by Lemma 7(ii) and Condition 1(a) that $S_{\lfloor \pi T \rfloor} \leq KT^{\frac{1}{3}}$ for any $\pi \in [\pi_1, \pi_2] \subset (0, 1)$, so term XV is $o_p(1)$ if $\frac{T^{\frac{2}{3}}}{N} \rightarrow 0$ as $N, T \rightarrow \infty$. For Parzen kernel, the RHS of (A.9) is $O_p \left(\frac{T^{\frac{1}{5}}}{T^{\frac{1}{4}}} \right) + O_p \left(\frac{T^{\frac{1}{5}}}{N^{\frac{1}{2}}} \right)$ by Lemma 7(i) and Condition 2(a) that $S_{\lfloor \pi T \rfloor} \leq KT^{\frac{1}{5}}$ for any $\pi \in [\pi_1, \pi_2] \subset (0, 1)$, so term XV is $o_p(1)$ if $\frac{T^{\frac{2}{5}}}{N} \rightarrow 0$ as $N, T \rightarrow \infty$. For the QS kernel, term

XV is bounded by

$$\begin{aligned}
& \sup_{\pi \in [\pi_1, \pi_2]} \left\{ \left\| \hat{\Gamma}_{1,0}(\pi, \hat{F}) - \hat{\Gamma}_{1,0}(\pi, FH) \right\| + 2 \sum_{j=1}^{S_{\lfloor \pi T \rfloor}} \left| k \left(\frac{j}{S_{\lfloor \pi T \rfloor}} \right) \right| \left\| \hat{\Gamma}_{1,j}(\pi, \hat{F}) - \hat{\Gamma}_{1,j}(\pi, FH) \right\| \right\} \\
& + \sup_{\pi \in [\pi_1, \pi_2]} \left\{ 2 \sum_{j=S_{\lfloor \pi T \rfloor}+1}^{\lfloor \pi T \rfloor - 1} \left| k \left(\frac{j}{S_{\lfloor \pi T \rfloor}} \right) \right| \left\| \hat{\Gamma}_{1,j}(\pi, \hat{F}) - \hat{\Gamma}_{1,j}(\pi, FH) \right\| \right\} \\
& = a + b
\end{aligned}$$

Term a is $O_p \left(\frac{T^{\frac{1}{5}}}{T^{\frac{1}{4}}} \right) + O_p \left(\frac{T^{\frac{1}{5}}}{N^{\frac{1}{2}}} \right)$ by Lemma 7(i) and Condition 2(a). For term b , note that

$$k(x) = \frac{25}{12\pi^2 x^2} \left(\frac{\sin \left(\frac{6\pi x}{5} \right)}{\frac{6\pi x}{5}} - \cos \left(\frac{6\pi x}{5} \right) \right)$$

It is obvious that for any $|x| > 1$, there exists $M > 0$, such that $\left| \frac{\sin \left(\frac{6\pi x}{5} \right)}{\frac{6\pi x}{5}} - \cos \left(\frac{6\pi x}{5} \right) \right| < M$. Therefore, term b is bounded by

$$\begin{aligned}
& \sup_{\pi \in [\pi_1, \pi_2]} \left\{ S_{\lfloor \pi T \rfloor}^2 \sum_{j=S_{\lfloor \pi T \rfloor}+1}^{\lfloor \pi T \rfloor - 1} \frac{25M}{6\pi^2 j^2} \left\| \hat{\Gamma}_{1,j}(\pi, \hat{F}) - \hat{\Gamma}_{1,j}(\pi, FH) \right\| \right\} \\
& \leq \left(\frac{25M \left(K_2 T^{\frac{1}{5}} \right)^2}{6\pi^2} \sum_{j=\lfloor K_1 T^{\frac{1}{5}} \rfloor}^T \frac{1}{j^2} \right) \sup_{\pi \in [\pi_1, \pi_2]} \left\| \hat{\Gamma}_{1,j}(\pi, \hat{F}) - \hat{\Gamma}_{1,j}(\pi, FH) \right\| \\
& \leq \left(\frac{25M \left(K_2 T^{\frac{1}{5}} \right)^2}{6\pi^2} \int_{\lfloor K_1 T^{\frac{1}{5}} \rfloor - 1}^T \frac{1}{j^2} \right) \sup_{\pi \in [\pi_1, \pi_2]} \left\| \hat{\Gamma}_{1,j}(\pi, \hat{F}) - \hat{\Gamma}_{1,j}(\pi, FH) \right\| \\
& \leq \frac{25M \left(K_2 T^{\frac{1}{5}} \right)^2}{6\pi^2} \left(\frac{1}{\lfloor K_1 T^{\frac{1}{5}} \rfloor - 1} - \frac{1}{T} \right) O_p \left(\frac{1}{\delta_{NT}} \right) \\
& = O \left(T^{\frac{2}{5}} \right) O \left(T^{-\frac{1}{5}} \right) O_p \left(\frac{1}{\delta_{NT}} \right) = O_p \left(\frac{T^{\frac{1}{5}}}{\delta_{NT}} \right)
\end{aligned}$$

Thus, for QS kernels, term XV is $o_p(1)$ if $\frac{T^{\frac{2}{5}}}{N} \rightarrow 0$ as $N, T \rightarrow \infty$. Using Lemma 8, one can show that term XVI is $o_p(1)$ in using a similar argument. ■

To prove Theorem 3 we present Lemma 9.

Lemma 9: Under Assumptions 1 - 9,

- (i) if $\sqrt{T}/N \rightarrow 0$, then $\sup_{\pi \in [\pi_1, \pi_2]} \|A(\pi, FH)\| = O_p(1)$ and $\sup_{\pi \in [\pi_1, \pi_2]} \|A(\pi, \hat{F})\| = O_p(1)$;
- (ii) $\sup_{\pi \in [\pi_1, \pi_2]} \|\hat{S}(\pi, FH_0)^{-1}\| = O_p(1)$ and $\sup_{\pi \in [\pi_1, \pi_2]} \|\hat{S}(\pi, \hat{F})^{-1}\| = O_p(1)$;

$$(iii) \sup_{\pi \in [\pi_1, \pi_2]} \left\| \hat{\mathbb{S}}(\pi, \hat{F})^{-1} - \hat{\mathbb{S}}(\pi, FH_0)^{-1} \right\| = o_p(1).$$

Proof of Lemma 9:

For part (i),

$$\begin{aligned} \sup_{\pi \in [\pi_1, \pi_2]} A(\pi, FH) &= \sup_{\pi \in [\pi_1, \pi_2]} \left\| \text{vech} \left(\frac{1}{[\pi T]} \sum_{t=1}^{[\pi T]} H' f_t f_t' H - \frac{1}{T - [\pi T]} \sum_{t=[\pi T]+1}^T H' f_t f_t' H \right) \right\| \\ &\leq \sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{[\pi T]} \sum_{t=1}^{[\pi T]} f_t f_t' - \frac{1}{T - [\pi T]} \sum_{t=[\pi T]+1}^T f_t f_t' \right\| \|H\|^2 \\ &= O_P(1) \end{aligned}$$

by Assumption 8(b).

$$\sup_{\pi \in [\pi_1, \pi_2]} \|A(\pi, \hat{F})\| \leq \sup_{\pi \in [\pi_1, \pi_2]} \|A(\pi, FH_0)\| + \sup_{\pi \in [\pi_1, \pi_2]} \|A(\pi, \hat{F}) - A(\pi, FH_0)\| = O_p(1) + O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right)$$

by Theorem 1. If $\sqrt{T}/N \rightarrow 0$, then the second term is $o_p(1)$.

Now, we consider part (ii). Assumption 9(a) and the fact that $0 < \pi_1 \leq \pi \leq \pi_2 < 1$ imply that $\sup_{\pi \in [\pi_1, \pi_2]} \|\hat{\mathbb{S}}(\pi, FH_0) - (\frac{1}{\pi} + \frac{1}{1-\pi})\Omega\| = o_p(1)$. Since Ω is positive definite, $\sup_{\pi \in [\pi_1, \pi_2]} |\rho_{min}(\hat{\mathbb{S}}(\pi, FH_0)) - \rho_{min}((\frac{1}{\pi} + \frac{1}{1-\pi})\Omega)| \leq \sup_{\pi \in [\pi_1, \pi_2]} \|\hat{\mathbb{S}}(\pi, FH_0) - (\frac{1}{\pi} + \frac{1}{1-\pi})\Omega\| = o_p(1)$,¹³ where $\rho_{min}(\cdot)$ denote the minimum eigenvalue of a symmetric matrix. This means that the eigenvalues of $\hat{\mathbb{S}}(\pi, FH_0)$ are bounded away from zero uniformly in π , so $\sup_{\pi \in [\pi_1, \pi_2]} \|\hat{\mathbb{S}}(\pi, FH_0)^{-1}\| = O_p(1)$. For the second part of (ii), we have $\sup_{\pi \in [\pi_1, \pi_2]} \|\hat{\mathbb{S}}(\pi, \hat{F}) - (\frac{1}{\pi} + \frac{1}{1-\pi})\Omega\| \leq \sup_{\pi \in [\pi_1, \pi_2]} \|\hat{\mathbb{S}}(\pi, \hat{F}) - \hat{\mathbb{S}}(\pi, FH_0)\| + \sup_{\pi \in [\pi_1, \pi_2]} \|\hat{\mathbb{S}}(\pi, FH_0) - (\frac{1}{\pi} + \frac{1}{1-\pi})\Omega\| = o_p(1)$ by Theorem 2 and Assumption 9(a), so

$$\sup_{\pi \in [\pi_1, \pi_2]} |\rho_{min}(\hat{\mathbb{S}}(\pi, \hat{F})) - \rho_{min}((\frac{1}{\pi} + \frac{1}{1-\pi})\Omega)| \leq \sup_{\pi \in [\pi_1, \pi_2]} \|\hat{\mathbb{S}}(\pi, \hat{F}) - (\frac{1}{\pi} + \frac{1}{1-\pi})\Omega\| = o_p(1).$$

This means that the eigenvalues of $\hat{\mathbb{S}}(\pi, \hat{F})$ are bounded away from zero uniformly in π , which implies $\sup_{\pi \in [\pi_1, \pi_2]} \|\hat{\mathbb{S}}(\pi, \hat{F})^{-1}\| = O_p(1)$.

For part (iii),

$$\begin{aligned} &\sup_{\pi \in [\pi_1, \pi_2]} \left\| \hat{\mathbb{S}}(\pi, \hat{F})^{-1} - \hat{\mathbb{S}}(\pi, FH_0)^{-1} \right\| \\ &= \sup_{\pi \in [\pi_1, \pi_2]} \left\| \hat{\mathbb{S}}(\pi, FH_0)^{-1} \left(\hat{\mathbb{S}}(\pi, FH_0) - \hat{\mathbb{S}}(\pi, \hat{F}) \right) \hat{\mathbb{S}}(\pi, \hat{F})^{-1} \right\| \\ &\leq \sup_{\pi \in [\pi_1, \pi_2]} \left\| \hat{\mathbb{S}}(\pi, FH_0)^{-1} \right\| \sup_{\pi \in [\pi_1, \pi_2]} \left\| \hat{\mathbb{S}}(\pi, FH_0) - \hat{\mathbb{S}}(\pi, \hat{F}) \right\| \sup_{\pi \in [\pi_1, \pi_2]} \left\| \hat{\mathbb{S}}(\pi, \hat{F})^{-1} \right\| \\ &= o_p(1) \end{aligned}$$

¹³This inequality follows from Golub and van Loan (1989, Corollary 8.1.3, p. 411).

using the result in part (ii) and Theorem 2. ■

Proof of Theorem 3: For part (i),

$$\begin{aligned}
& \sup_{\pi \in [\pi_1, \pi_2]} \left| W_T(\pi, \hat{F}) - W_T(\pi, FH_0) \right| \\
\leq & \sup_{\pi \in [\pi_1, \pi_2]} \left| A(\pi, \hat{F})' \left[\hat{S}(\pi, \hat{F})^{-1} - \hat{S}(\pi, FH_0)^{-1} \right] A(\pi, \hat{F}) \right| + \\
& \sup_{\pi \in [\pi_1, \pi_2]} \left| \left[A(\pi, \hat{F})' - A(\pi, FH_0)' \right] \hat{S}(\pi, FH_0)^{-1} A(\pi, \hat{F}) \right| + \sup_{\pi \in [\pi_1, \pi_2]} \left| A(\pi, FH_0)' \hat{S}(\pi, FH_0)^{-1} \left[A(\pi, \hat{F}) - A(\pi, FH_0) \right] \right| \\
= & o_p(1)
\end{aligned}$$

using results of Lemma 9 and Theorem 1. The result that $\sup_{\pi \in [\pi_1, \pi_2]} \left| LM_T(\pi, \hat{F}) - LM_T(\pi, FH_0) \right| = o_p(1)$ can be proved in a similar way.

For part (ii), since $\sup_{\pi \in [\pi_1, \pi_2]} W_T(\pi, FH_0) \xrightarrow{d} \sup_{\pi \in [\pi_1, \pi_2]} Q_p(\pi)$ by Assumption 9(b), it is sufficient to show that

$$\left| \sup_{\pi \in [\pi_1, \pi_2]} W_T(\pi, \hat{F}) - \sup_{\pi \in [\pi_1, \pi_2]} W_T(\pi, FH_0) \right| = o_p(1).$$

Note that $W_T(\pi, \hat{F}) = W_T(\pi, \hat{F}) - W_T(\pi, FH_0) + W_T(\pi, FH_0)$. Taking supreme on both sides gives

$$\sup_{\pi \in [\pi_1, \pi_2]} W_T(\pi, \hat{F}) \leq \sup_{\pi \in [\pi_1, \pi_2]} W_T(\pi, \hat{F}) - W_T(\pi, FH_0) + \sup_{\pi \in [\pi_1, \pi_2]} W_T(\pi, FH_0)$$

So $\sup_{\pi \in [\pi_1, \pi_2]} W_T(\pi, \hat{F}) - \sup_{\pi \in [\pi_1, \pi_2]} W_T(\pi, FH_0) \leq \sup_{\pi \in [\pi_1, \pi_2]} \left| W_T(\pi, \hat{F}) - W_T(\pi, FH_0) \right| = o_p(1)$ by the result in part (i). Similarly, one can also show that $\sup_{\pi \in [\pi_1, \pi_2]} W_T(\pi, FH_0) - \sup_{\pi \in [\pi_1, \pi_2]} W_T(\pi, \hat{F}) \leq \sup_{\pi \in [\pi_1, \pi_2]} \left| W_T(\pi, \hat{F}) - W_T(\pi, FH_0) \right| = o_p(1)$. Combining these two inequalities give the desired result. The result for the LM-like statistic can be proved in a similar way. ■

Proof of Corollary 1: It is sufficient to prove that $\sup_{\pi \in [\pi_1, \pi_2]} \left| \exp \left(\frac{W_T(\pi, \hat{F})}{2} \right) - \exp \left(\frac{W_T(\pi, FH_0)}{2} \right) \right| = o_p(1)$. By mean value theorem, there exists a sequence of $c_\pi \in [0, 1]$ such that

$$\begin{aligned}
& \sup_{\pi \in [\pi_1, \pi_2]} \left| \exp \left(\frac{W_T(\pi, \hat{F})}{2} \right) - \exp \left(\frac{W_T(\pi, FH_0)}{2} \right) \right| \\
= & \sup_{\pi \in [\pi_1, \pi_2]} \left| \frac{1}{2} \left[W_T(\pi, \hat{F}) - W_T(\pi, FH_0) \right] \exp \left(\frac{c_\pi W_T(\pi, \hat{F})}{2} + \frac{(1 - c_\pi) W_T(\pi, FH_0)}{2} \right) \right| \\
\leq & \sup_{\pi \in [\pi_1, \pi_2]} \left| \frac{1}{2} \left[W_T(\pi, \hat{F}) - W_T(\pi, FH_0) \right] \right| \sup_{\pi \in [\pi_1, \pi_2]} \left| \exp \left(\frac{W_T(\pi, \hat{F})}{2} + \frac{W_T(\pi, FH_0)}{2} \right) \right| \\
= & o_p(1)
\end{aligned}$$

where the last equality follows from the fact that both $\sup_{\pi \in [\pi_1, \pi_2]} W_T(\pi, \hat{F})$ and $\sup_{\pi \in [\pi_1, \pi_2]} W_T(\pi, FH_0)$ are $O_p(1)$ by Theorem 3 and Assumption 9(b). ■

B Proofs of the Results in Section 2.5

We present the proof of Proposition 1 and Lemma 10 and then prove Theorem 4.

Proof of Proposition 1: The proof is the same as that of Theorem 2 of Bai and Ng (2002). To show that Bai and Ng's information criteria are applicable to the equivalent models, it is sufficient to verify that (2.17) and (2.18) satisfy Assumptions A–D of Bai and Ng (2002). Our Assumption 3 is identical to Assumption C of Bai and Ng (2002). Thus, we only need to verify that (2.16) satisfies the following assumptions under types 1 and 2 breaks:

Assumption A (Bai and Ng, 2002): $E \|g_t\|^4 < \infty$, and $T^{-1} \sum_{t=1}^T g_t g_t' \xrightarrow{P} \Sigma_G$ as $T \rightarrow \infty$ for some positive definite matrix Σ_G .

Assumption B (Bai and Ng, 2002): $\|\theta_i\| \leq \bar{\theta} < \infty$, $\|\Theta' \Theta / N - \Sigma_\Theta\| \rightarrow 0$ for some $r \times r$ positive definite matrix Σ_Θ .

Assumption D (Bai and Ng, 2002): $E \left(\frac{1}{N} \sum_{i=1}^N \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T g_t e_{it} \right\|^2 \right) \leq M$.

Under type 1 break, $g_t \equiv (f'_{0,t}, f'_{1,t}, 0_{1 \times q_1})'$ if $1 \leq t \leq \lfloor \pi^* T \rfloor$, and $g_t \equiv (f'_{0,t}, 0_{1 \times q_1}, f'_{1,t})'$ if $\lfloor \pi^* T \rfloor + 1 \leq t \leq T$. It is straightforward that $E \|g_t\|^4 < \infty$ because of $E \|f_t\|^4 < \infty$ by Assumption 1. By Assumption 10(a), $\Sigma_{G(1)}$ is positive definite. To verify Assumption B, we have $\|\theta_i\| \leq \sqrt{2\bar{\lambda}} < \infty$ by Assumption 2 and 10(a), and the limit of $\Theta'_{(1)} \Theta_{(1)} / N$ is positive definite by Assumption 10(a). Also, the verification of Assumption D is straightforward based on Assumption 4 and structure of g_t .

Under type 2 break, $g_t \equiv (f'_{0,t}, f'_{1,t})'$ if $1 \leq t \leq \lfloor \pi^* T \rfloor$, and $g_t \equiv (f'_{0,t}, f'_{1,t} Z')'$ if $\lfloor \pi^* T \rfloor + 1 \leq t \leq T$. First note that $E \|g_t\|^4 < \infty$ by $\|Z\| < \infty$ and Assumption 1. Also, the limit of $T^{-1} \sum_{t=1}^T g_t g_t'$ is

$$\Sigma_{G(2)} = \pi^* \Sigma_F + (1 - \pi^*) \begin{bmatrix} \Sigma_F^{0,0} & \Sigma_F^{0,1} Z' \\ Z \Sigma_F^{0,1'} & Z \Sigma_F^{1,1} Z' \end{bmatrix}$$

by Assumption 1. $\Sigma_{G(2)}$ is positive definite because Σ_F is positive definite and $\begin{bmatrix} \Sigma_F^{0,0} & \Sigma_F^{0,1} Z' \\ Z \Sigma_F^{0,1'} & Z \Sigma_F^{1,1} Z' \end{bmatrix}$ is positive semi-definite. Assumption B automatically holds by Assumption 2 since $\Theta_{(2)} = \Lambda$ under type 2 break. Finally, the verification of Assumption D is straightforward based on Assumption 4 and structure of g_t . We have completed the verification of Assumptions A–D of Bai and Ng (2002). ■

Lemma 10: Under Assumptions 1–6, 8(a) and 10,

- (i) $T^{-1} \sum_{t=1}^T \|\hat{f}_t - J' g_t\|^2 = O_p(\delta_{NT}^{-2})$;
- (ii) For any π_1 and π_2 that satisfy $0 < \pi_1 < \pi_2 < 1$,

$$\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{\lfloor \pi T \rfloor} \sum_{t=1}^{\lfloor \pi T \rfloor} (\hat{f}_t - J' g_t) g_t' \right\| = O_p \left(\frac{1}{\delta_{NT}^2} \right)$$

$$\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{T - \lfloor \pi T \rfloor} \sum_{t=\lfloor \pi T \rfloor + 1}^T (\hat{f}_t - J' g_t) g_t' \right\| = O_p \left(\frac{1}{\delta_{NT}^2} \right)$$

Proof of Lemma 10: Parts (i) and (ii) are just analogs of Lemmas 1 and 3. Part (i) is simply Theorem 1 of Bai and Ng (2002) for the equivalent models. The proof only requires verification of Assumptions A–D in Bai and Ng (2002), which has been done in the proof of Proposition 1.

For part (ii), the proof is almost the same as that of Lemma 3. To use the argument in the proof of Lemma 3, we will need the following conditions:

Condition A. There exists an $M_1 < \infty$ such that for all N and T :

(a) for each t , $E \left\| \frac{1}{\sqrt{NT}} \sum_{s=1}^T \sum_{k=1}^N g_s [e_{ks} e_{kt} - E(e_{ks} e_{kt})] \right\|^2 \leq M_1$;

(b) $E \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^T \sum_{k=1}^N g_t \theta_k' e_{kt} \right\| \leq M_1$;

(c) for each t , $E \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N \theta_i e_{it} \right\|^4 \leq M_1$.

Condition B. For any constants π_1 and π_2 that satisfy $0 < \pi_1 \leq \pi^* \leq \pi_2 < 1$,

$$\sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^{\lfloor \pi T \rfloor} \sum_{k=1}^N g_t \theta_k' e_{kt} \right\|^2 = O_p(1) \text{ and } \sup_{\pi \in [\pi_1, \pi_2]} \left\| \frac{1}{\sqrt{NT}} \sum_{t=\lfloor \pi T \rfloor + 1}^T \sum_{k=1}^N g_t \theta_k' e_{kt} \right\|^2 = O_p(1).$$

These two conditions can be easily verified. Under type 1 break, $g_t \equiv (f'_{0,t}, f'_{1,t}, 0_{1 \times q_1})'$ if $1 \leq t \leq \lfloor \pi^* T \rfloor$, $g_t \equiv (f'_{0,t}, 0_{1 \times q_1}, f'_{1,t})'$ if $\lfloor \pi^* T \rfloor + 1 \leq t \leq T$, and $\theta_k \equiv (\lambda'_{0,k}, \lambda'_{1,k}, \lambda'_{2,k})'$. Hence, parts (a)–(c) of Condition A are implied by Assumptions 6(a)–6(c), respectively. Condition B is implied by Assumption 8(a). Under type 2 break, $g_t \equiv f_t$ if $1 \leq t \leq \lfloor \pi^* T \rfloor$, $g_t \equiv (f'_{0,t}, f'_{1,t} Z')$ if $\lfloor \pi^* T \rfloor + 1 \leq t \leq T$, and $\theta_k \equiv \lambda_k$. Since Z is bounded by Assumption 10(b), Conditions A and B hold by Assumptions 6 and 8(a). ■

Proof of Theorem 4: (i)

$$\begin{aligned} & \frac{1}{\lfloor \pi^* T \rfloor} \sum_{t=1}^{\lfloor \pi^* T \rfloor} \hat{f}_t \hat{f}_t' - \frac{1}{T - \lfloor \pi^* T \rfloor} \sum_{t=\lfloor \pi^* T \rfloor + 1}^T \hat{f}_t \hat{f}_t' \\ &= \left(\frac{1}{\lfloor \pi^* T \rfloor} \sum_{t=1}^{\lfloor \pi^* T \rfloor} J' g_t g_t' J - \frac{1}{T - \lfloor \pi^* T \rfloor} \sum_{t=\lfloor \pi^* T \rfloor + 1}^T J' g_t g_t' J \right) + \frac{1}{\lfloor \pi^* T \rfloor} \sum_{t=1}^{\lfloor \pi^* T \rfloor} (\hat{f}_t \hat{f}_t' - J' g_t g_t' J) \\ & \quad - \frac{1}{T - \lfloor \pi^* T \rfloor} \sum_{t=\lfloor \pi^* T \rfloor + 1}^T (\hat{f}_t \hat{f}_t' - J' g_t g_t' J) \end{aligned}$$

Note that

$$\begin{aligned} & \frac{1}{\lfloor \pi^* T \rfloor} \sum_{t=1}^{\lfloor \pi^* T \rfloor} (\hat{f}_t \hat{f}_t' - J' g_t g_t' J) \\ &= \frac{1}{\lfloor \pi^* T \rfloor} \sum_{t=1}^{\lfloor \pi^* T \rfloor} \left[(\hat{f}_t - J' g_t) g_t' J + (\hat{f}_t - J' g_t) (\hat{f}_t' - g_t' J) + J' g_t (\hat{f}_t' - g_t' J) \right] = O_p \left(\frac{1}{\delta_{NT}^2} \right) \end{aligned}$$

by Lemma 10. Similarly, $\frac{1}{[T-\pi^*T]} \sum_{t=[\pi^*T]+1}^T (\hat{f}_t \hat{f}_t' - J' g_t g_t' J) = O_p(\delta_{NT}^{-2})$.

Under type 1 break, recall D_1 , D_2 and C defined by (2.19), so we have

$$\frac{1}{[\pi^*T]} \sum_{t=1}^{[\pi^*T]} J' g_t g_t' J - \frac{1}{T - [\pi^*T]} \sum_{t=[\pi^*T]+1}^T J' g_t g_t' J \xrightarrow{p} J_0' \left(\begin{bmatrix} \Sigma_F^{0,0} & \Sigma_F^{0,1} & 0 \\ \Sigma_F^{0,1'} & \Sigma_F^{1,1} & 0 \\ 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} \Sigma_F^{0,0} & 0 & \Sigma_F^{0,1} \\ 0 & 0 & 0 \\ \Sigma_F^{0,1'} & 0 & \Sigma_F^{1,1} \end{bmatrix} \right) J_0 \equiv C$$

by Assumption 8(b) and the definitions of J and J_0 . Matrix C contains non-zero entries because $D_1 - D_2$ is not zero due to the positive definiteness of $\Sigma_F^{1,1}$ and the fact that J_0 is a non-singular matrix.

Under type 2 break, recall D_1 , D_2 and C defined by (2.20), so we have

$$J' \left(\frac{1}{[\pi^*T]} \sum_{t=1}^{[\pi^*T]} g_t g_t' - \frac{1}{T - [\pi^*T]} \sum_{t=[\pi^*T]+1}^T g_t g_t' \right) J \xrightarrow{p} J_0' \begin{bmatrix} 0_{q_0 \times q_0} & \Sigma_F^{0,1} (I_{q_1} - Z') \\ (I_{q_1} - Z) \Sigma_F^{0,1'} & \Sigma_F^{1,1} - Z \Sigma_F^{1,1} Z' \end{bmatrix} J_0 \equiv C$$

by Assumption 8(b) and the definitions of J and J_0 . Matrix C contains non-zero entries because $\Sigma_F^{1,1} - Z \Sigma_F^{1,1} Z'$ is not zero by Assumption 10(b) and the fact that J_0 is a non-singular matrix.

(ii) First, note that Assumption 10(c) is the analog of Assumption 7. Hence, Theorem 2 still holds for the equivalent models under the alternative and we have $\|\hat{\mathbb{S}}(\pi^*, GJ_0) - \hat{\mathbb{S}}(\pi^*, \hat{F})\| = o_p(1)$ and $\|\tilde{\mathbb{S}}(\pi^*, GJ_0) - \tilde{\mathbb{S}}(\pi^*, \hat{F})\| = o_p(1)$. Second, we will show that $\|\hat{\mathbb{S}}(\pi^*, GJ_0)^{-1} - \hat{\mathbb{S}}(\pi^*, \hat{F})^{-1}\| = o_p(1)$ and $\|\tilde{\mathbb{S}}(\pi^*, GJ_0)^{-1} - \tilde{\mathbb{S}}(\pi^*, \hat{F})^{-1}\| = o_p(1)$. Let $\rho_{max}(A)$ and $\rho_{min}(A)$ denote the largest and smallest eigenvalue of a matrix A , respectively. Hall (2000) shows that if the HAC estimates are not correctly demeaned, then the HAC estimator is asymptotically equivalent to the sum of two matrices: one of these matrices is positive definite and $O(1)$; the other is positive semi-definite and diverges at the rate of the bandwidth parameter. Using these results, we can see that $\text{plim}_{T \rightarrow \infty} (\rho_{min}[\hat{\mathbb{S}}(\pi^*, GJ_0)]) > 0$ and $\text{plim}_{T \rightarrow \infty} (\rho_{min}[\tilde{\mathbb{S}}(\pi^*, GJ_0)]) > 0$. This implies that both $\|\hat{\mathbb{S}}(\pi^*, GJ_0)^{-1}\|$ and $\|\tilde{\mathbb{S}}(\pi^*, GJ_0)^{-1}\|$ are $O_p(1)$, because $\|\hat{\mathbb{S}}(\pi^*, GJ_0)^{-1}\| \leq M \rho_{max}[\hat{\mathbb{S}}(\pi^*, GJ_0)^{-1}] = M (\rho_{min}[\hat{\mathbb{S}}(\pi^*, GJ_0)])^{-1} = O_p(1)$ for some positive constant $M < \infty$ (see Hall (2000), Eq. (16), p.1525). Since Theorem 2 implies that $\|\hat{\mathbb{S}}(\pi^*, GJ_0) - \hat{\mathbb{S}}(\pi^*, \hat{F})\| = o_p(1)$ and $\|\tilde{\mathbb{S}}(\pi^*, GJ_0) - \tilde{\mathbb{S}}(\pi^*, \hat{F})\| = o_p(1)$, it follows that $\text{plim}_{T \rightarrow \infty} (\rho_{min}[\hat{\mathbb{S}}(\pi^*, \hat{F})]) > 0$ and $\text{plim}_{T \rightarrow \infty} (\rho_{min}[\tilde{\mathbb{S}}(\pi^*, \hat{F})]) > 0$, which also implies that both $\|\hat{\mathbb{S}}(\pi^*, \hat{F})^{-1}\|$ and $\|\tilde{\mathbb{S}}(\pi^*, \hat{F})^{-1}\|$ are $O_p(1)$. Now, we have

$$\|\hat{\mathbb{S}}(\pi^*, GJ_0)^{-1} - \hat{\mathbb{S}}(\pi^*, \hat{F})^{-1}\| \leq \|\hat{\mathbb{S}}(\pi^*, GJ_0)^{-1}\| \|\hat{\mathbb{S}}(\pi^*, \hat{F}) - \hat{\mathbb{S}}(\pi^*, GJ_0)\| \|\hat{\mathbb{S}}(\pi^*, \hat{F})^{-1}\| = O_p(1) o_p(1) O_p(1). \quad (\text{B.1})$$

We can show $\|\tilde{\mathbb{S}}(\pi^*, GJ_0)^{-1} - \tilde{\mathbb{S}}(\pi^*, \hat{F})^{-1}\| = o_p(1)$ using a similar argument.

Finally,

$$\begin{aligned}
& W_T(\pi^*, \hat{F}) \\
= & A(\pi^*, \hat{F})' \hat{S}(\pi^*, \hat{F})^{-1} A(\pi^*, \hat{F}) \\
= & \frac{T}{\max(S_{\lfloor \pi^* T \rfloor}, S_{T - \lfloor \pi^* T \rfloor})} \left[\frac{1}{\sqrt{T}} A(\pi^*, \hat{F})' \right] \left[\max(S_{\lfloor \pi^* T \rfloor}, S_{T - \lfloor \pi^* T \rfloor}) \hat{S}(\pi^*, \hat{F})^{-1} \right] \left[\frac{1}{\sqrt{T}} A(\pi^*, \hat{F}) \right] \\
= & \frac{T}{\max(S_{\lfloor \pi^* T \rfloor}, S_{T - \lfloor \pi^* T \rfloor})} [\text{vech}(C)' + o_p(1)] \left\{ \max(S_{\lfloor \pi^* T \rfloor}, S_{T - \lfloor \pi^* T \rfloor}) \left[\hat{S}(\pi^*, GJ_0)^{-1} + o_p(1) \right] \right\} [\text{vech}(C) + o_p(1)] \\
\rightarrow & \infty
\end{aligned}$$

by Assumption 11, part (i) of Theorem 4, and (B.1). Also, it can be proved that $LM_T(\pi^*, \hat{F}) \rightarrow \infty$ using a similar argument. $\sup_{\pi \in [\pi_1, \pi_2]} W_T(\pi, \hat{F})$ and $\sup_{\pi \in [\pi_1, \pi_2]} LM_T(\pi, \hat{F})$ are also consistent tests due to the consistency of $W_T(\pi^*, \hat{F})$ and $LM_T(\pi^*, \hat{F})$. ■

References

- [1] Andrews, Donald W.K. (1993), “Tests for Parameter Instability and Structural Change with Unknown Change Point,” *Econometrica*, 61, 821-856.
- [2] Andrews, Donald W.K., and Werner Ploberger (1994), “Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative,” *Econometrica*, 62, 1383.
- [3] Bai, J., and S. Ng (2002), “Determining the Number of Factors in Approximate Factor Models”, *Econometrica* 70:191-221.
- [4] Bai, J. (2003), “Inferential Theory for Factor Models of Large Dimensions,” *Econometrica* 71, 135-172.
- [5] Bai, J., and S. Ng (2010), “Principal Components Estimation and Identification of the Factors,” Columbia University working paper.
- [6] Banerjee, A. and M. Marcellino (2008), “Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Change,” CEPR Working Paper 6706.
- [7] Bates, B., M. Plagborg-Moller, J. Stock, M. Watson (2012), “Consistent Factor Estimation in Dynamic Factor Models with Structural Instability,” *Journal of Econometrics*, forthcoming.
- [8] Bernanke, B.S., J. Boivin and P. Elias (2005), “Measuring the Effects of Monetary Policy: a Factor-augmented Vector Autoregressive (FAVAR) Approach”, *Quarterly Journal of Economics* 120, 387–422.
- [9] Boivin, J. and M. Giannoni (2006), “DSGE Models in a Data-Rich Environment,” NBER Working Paper No. 12772.
- [10] Breitung, Jörg. and Sandra Eickmeier (2011), “Testing For Structural Breaks in Dynamic Factor Models,” *Journal of Econometrics*, 163, 71–84.
- [11] Chen, L., Dolado, J., and Gonzalo, J. (2012), “Detecting Big Structural Breaks in Large Factor Models,” MPRA working paper.
- [12] Cheng, X., Z. Liao, and F. Schorfheide (2013), “Shrinkage Estimation of Dynamic Factor Models with Structural Instabilities”, Manuscript, University of Pennsylvania.
- [13] Eickmeier, Sandra, Wolfgang Lemke and Massimiliano Marcellino (2011), “Classical Time-Varying FAVAR Models – Estimation, Forecasting and Structural Analysis,” CEPR Discussion Paper DP8321.
- [14] Golub, G.H. and C.F. van Loan (1989), *Matrix Computations*, Second Edn. Baltimore, MD: Johns Hopkins Press.
- [15] Hall, Alastair R. (2000), “Covariance Matrix Estimation and the Power of the Overidentifying Restrictions Test,” *Econometrica*, 68, 1517-1527.

- [16] Ledoit, O. and M. Wolf (2004), “A Well-conditioned Estimator for Large-dimensional Covariance Matrices,” *Journal of Multivariate Analysis*, 88, 365-411.
- [17] Newey, Whitney K., and Kenneth D. West (1994), “Automatic Lag Selection in Covariance Matrix Estimation,” *The Review of Economic Studies*, 61(4), 631-653.
- [18] Onatski, A. (2010), “Determining the Number of Factors from Empirical Distribution of Eigenvalues,” *The Review of Economics and Statistics*, 92(4), 1004-1016.
- [19] Sargent, T.J. and C.A.Sims (1977), “Business Cycle Modelling without Pretending to Have Too Much a-priori Economic Theory.”, in: Sims et al., eds., *New Methods in Business Cycle Research* (Federal Reserve Bank of Minneapolis, Minneapolis).
- [20] Stock, James H., and Mark W. Watson (1996), “Evidence on Structural Instability in Macroeconomic Time Series Relations,” *Journal of Business and Economic Statistics*, 14, 11–30.
- [21] Stock, James H., and Mark W. Watson (2002a), “Macroeconomic forecasting using diffusion indexes”, *Journal of Business and Economic Statistics* 20, 147-162.
- [22] Stock, James H., and Mark W. Watson (2002b), “Forecasting Using Principal Components from a Large Number of Predictors,” *Journal of the American Statistical Association*, 97, 1167–1179.
- [23] Stock, James H., and Mark W. Watson (2005), “Implications of Dynamic Factor Models for VAR Analysis,” NBER Working Paper No. 11467.
- [24] Stock, James H., and Mark W. Watson (2009), “Forecasting in Dynamic Factor Models Subject to Structural Instability,” in *The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry*, Jennifer Castle and Neil Shephard (eds), Oxford: Oxford University Press.
- [25] White, Halbert (1980), “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity,” *Econometrica*, 48, 817-838.

Table 1: Size of Structural Break Tests with Known Break Date, $r = 3$

DGPs	β	P	ω_ϵ	N	T	Bonferroni Statistics			Pooled Statistics			W and LM					\hat{r}	
						s_{Bon}^0	s_{Bon}^{GLS}	s_{Bon}^{HAC}	S_{BE}^0	S_{BE}^{GLS}	S_{BE}^{HAC}	W_0	W_B	W_{QS}	LM_0	LM_B		LM_{QS}
N1				100	100	0.024	0.023	0.001	0.046	0.047	0.019	0.040	0.011	0.011	0.040	0.010	0.014	3.00
				100	200	0.038	0.039	0.006	0.049	0.049	0.032	0.039	0.020	0.021	0.039	0.023	0.023	3.00
				200	100	0.020	0.020	0.006	0.052	0.050	0.024	0.042	0.011	0.011	0.042	0.014	0.016	3.00
				200	200	0.034	0.034	0.003	0.050	0.052	0.041	0.048	0.024	0.026	0.048	0.027	0.028	3.00
N2	0			100	100	0.025	0.024	0.001	0.048	0.048	0.020	0.041	0.011	0.011	0.041	0.010	0.015	3.00
	0			100	200	0.039	0.040	0.001	0.048	0.048	0.034	0.040	0.019	0.021	0.040	0.022	0.023	3.00
	0			200	100	0.020	0.020	0.001	0.051	0.050	0.026	0.042	0.012	0.010	0.042	0.013	0.015	3.00
	0			200	200	0.033	0.033	0.003	0.051	0.053	0.041	0.049	0.024	0.026	0.049	0.029	0.027	3.00
N2	0.1	6		100	100	0.027	0.027	0.001	0.116	0.112	0.055	0.035	0.008	0.007	0.035	0.010	0.019	3.22
	0.1	6		100	200	0.039	0.038	0.004	0.105	0.102	0.075	0.042	0.015	0.014	0.042	0.016	0.016	3.66
	0.1	6		200	100	0.022	0.021	0.000	0.133	0.127	0.069	0.040	0.012	0.011	0.040	0.014	0.014	3.00
	0.1	6		200	200	0.034	0.035	0.005	0.126	0.121	0.094	0.044	0.021	0.024	0.044	0.022	0.024	3.00
	0.1	6		200	500	0.045	0.045	0.013	0.134	0.134	0.115	0.044	0.033	0.035	0.044	0.031	0.034	3.02
	0.1	8		100	200	0.030	0.030	0.001	0.104	0.103	0.057	0.031	0.002	0.001	0.031	0.003	0.037	6.37
	0.1	8		200	100	0.021	0.020	0.001	0.167	0.162	0.089	0.037	0.012	0.013	0.037	0.011	0.013	3.03
	0.1	8		200	200	0.037	0.036	0.004	0.157	0.156	0.111	0.043	0.015	0.015	0.043	0.018	0.017	3.46
N3			0	100	100	0.840	0.028	0.037	1.000	0.075	0.864	0.592	0.069	0.049	0.592	0.060	0.035	3.02
			0	100	200	0.923	0.039	0.060	1.000	0.057	0.759	0.608	0.087	0.061	0.608	0.057	0.050	3.00
			0	200	100	0.903	0.025	0.043	1.000	0.085	0.986	0.610	0.066	0.046	0.610	0.063	0.035	3.00
			0	200	200	0.968	0.033	0.063	1.000	0.061	0.964	0.620	0.082	0.059	0.620	0.058	0.054	3.00
			0	200	500	0.990	0.043	0.087	1.000	0.052	0.793	0.630	0.075	0.072	0.630	0.059	0.056	3.00
			0	500	100	0.954	0.020	0.037	1.000	0.147	1.000	0.596	0.069	0.045	0.596	0.059	0.042	3.00
			0	500	200	0.995	0.033	0.080	1.000	0.068	1.000	0.616	0.086	0.060	0.616	0.058	0.055	3.00
N3			0.5	100	150	0.995	0.031	0.109	1.000	0.054	0.926	0.658	0.076	0.063	0.658	0.053	0.041	3.15
			0.5	100	200	0.996	0.028	0.099	1.000	0.055	0.860	0.639	0.083	0.059	0.639	0.056	0.054	3.01
			0.5	200	150	1.000	0.026	0.141	1.000	0.063	0.997	0.658	0.078	0.067	0.658	0.062	0.053	3.11
			0.5	200	200	1.000	0.027	0.125	1.000	0.060	0.988	0.638	0.085	0.060	0.638	0.060	0.056	3.01
			0.5	200	500	1.000	0.032	0.115	1.000	0.054	0.839	0.630	0.079	0.075	0.630	0.059	0.061	3.00
			0.5	500	150	1.000	0.022	0.167	1.000	0.083	1.000	0.654	0.088	0.071	0.654	0.064	0.052	3.03
			0.5	500	200	1.000	0.023	0.161	1.000	0.083	1.000	0.642	0.083	0.064	0.642	0.057	0.052	3.00

Note: The nominal size is 5%. For Bonferroni statistics and pooled statistics, the superscript “0” denotes that the statistics assume conditional homoskedasticity and no serial correlation in the residuals; the superscript “GLS” stands for Breitung and Eickmeier’s (2011) estimates based on a quasi-demean transformation; the superscript “HAC” means that HAC estimates are used to compute the statistic. For LM and W , the subscript “0” means that the statistic uses White’s (1980) conditional heteroskedasticity robust estimate; subscripts “B” and “QS” denote statistics based on HAC estimates with Bartlett and QS kernels, respectively. All HAC estimates are based on Newey and West’s method (1994). \hat{r} is the number of factors estimated by IC_{p_1} of Bai and Ng (2002).

Table 2A: Power against a Break at $T/2$

DGP A1: the equivalent factor model with time-invariant loadings has 4 factors.

			Bonferroni Statistics			Pooled Statistics			W and LM						
b	N	T	s_{Bon}^0	s_{Bon}^{GLS}	s_{Bon}^{HAC}	S_{BE}^0	S_{BE}^{GLS}	S_{BE}^{HAC}	W_0	W_B	W_{QS}	LM_0	LM_B	LM_{QS}	\hat{r}
1/3	100	100	0.540	0.521	0.024	1.000	1.000	1.000	0.075	0.022	0.020	0.075	0.020	0.022	3.00
1/3	100	200	1.000	1.000	0.500	1.000	1.000	1.000	0.092	0.049	0.054	0.092	0.055	0.054	3.00
1/3	200	100	0.539	0.525	0.168	1.000	1.000	1.000	0.069	0.020	0.020	0.069	0.018	0.018	3.00
1/3	200	200	0.987	0.987	0.392	0.995	0.995	0.994	0.095	0.054	0.056	0.095	0.055	0.060	3.01
1/3	200	500	0.617	0.618	0.601	0.654	0.654	0.630	0.480	0.468	0.467	0.480	0.463	0.466	3.40
1/3	500	200	0.627	0.628	0.197	0.646	0.645	0.624	0.444	0.412	0.415	0.444	0.408	0.412	3.39
1/3	500	500	0.053	0.054	0.006	0.094	0.092	0.051	1.000	1.000	1.000	1.000	1.000	1.000	4.00
2/3	100	100	0.028	0.026	0.001	0.066	0.065	0.022	0.999	0.603	0.458	0.999	0.416	0.278	4.00
2/3	100	200	0.041	0.042	0.002	0.069	0.068	0.030	1.000	0.976	0.988	1.000	0.973	0.982	4.00
2/3	200	100	0.019	0.018	0.000	0.057	0.055	0.068	1.000	0.617	0.468	1.000	0.413	0.270	4.00
2/3	200	200	0.037	0.037	0.002	0.059	0.061	0.027	1.000	0.977	0.988	1.000	0.967	0.982	4.00
1	100	100	0.020	0.020	0.001	0.052	0.053	0.028	1.000	0.611	0.467	1.000	0.415	0.279	4.00
1	100	200	0.038	0.039	0.002	0.059	0.059	0.028	1.000	0.970	0.985	1.000	0.958	0.974	4.00
1	200	100	0.018	0.017	0.001	0.050	0.051	0.077	1.000	0.645	0.494	1.000	0.434	0.288	4.00
1	200	200	0.034	0.035	0.002	0.054	0.055	0.027	1.000	0.975	0.985	1.000	0.965	0.979	4.00
2	100	100	0.018	0.018	0.001	0.053	0.052	0.026	1.000	0.606	0.464	1.000	0.377	0.254	4.00
2	100	200	0.036	0.037	0.003	0.062	0.061	0.028	1.000	0.984	0.992	1.000	0.969	0.985	4.00
2	200	100	0.019	0.018	0.000	0.051	0.049	0.061	1.000	0.601	0.470	1.000	0.377	0.259	4.00
2	200	200	0.032	0.032	0.001	0.052	0.054	0.026	1.000	0.978	0.988	1.000	0.961	0.983	4.00

Note: The parameter b controls the size of the shift in factor loadings. For Bonferroni statistics and pooled statistics, the superscript “0” denotes that the statistic assumes conditional homoskedasticity and no serial correlation in the residuals; the superscript “GLS” stands for Breitung and Eickmeier’s (2011) estimates based on a quasi-demean transformation; the superscript “HAC” means that HAC estimates are used to compute the statistic. For LM and W , the subscript “0” means that the statistic uses White’s (1980) conditional heteroskedasticity robust estimate; subscripts “B” and “QS” denote statistics based on HAC estimates with Bartlett and QS kernels, respectively. All HAC estimates are based on Newey and West’s method (1994). \hat{r} is the number of factors estimated by IC_{p_1} of Bai and Ng (2002).

Table 2B: Power against a Break at $T/2$

DGP A2: the equivalent factor model with time-invariant loadings has 4 factors.

α	N	T	Bonferroni Statistics			Pooled Statistics			W and LM						\hat{r}
			s_{Bon}^0	s_{Bon}^{GLS}	s_{Bon}^{HAC}	S_{BE}^0	S_{BE}^{GLS}	S_{BE}^{HAC}	W_0	W_B	W_{QS}	LM_0	LM_B	LM_{QS}	
0.2	100	100	0.986	0.986	0.347	0.986	0.986	0.985	0.106	0.038	0.034	0.106	0.031	0.027	3.01
0.2	100	200	0.912	0.912	0.907	0.914	0.914	0.910	0.197	0.151	0.155	0.197	0.151	0.153	3.09
0.2	200	100	0.890	0.889	0.295	0.894	0.893	0.888	0.199	0.099	0.078	0.199	0.074	0.063	3.11
0.2	200	200	0.235	0.234	0.203	0.271	0.270	0.224	0.836	0.814	0.817	0.836	0.812	0.816	3.80
0.4	100	100	0.076	0.074	0.010	0.114	0.111	0.065	0.972	0.600	0.442	0.972	0.416	0.282	3.95
0.4	100	200	0.047	0.047	0.004	0.078	0.077	0.032	1.000	0.986	0.994	1.000	0.979	0.987	4.00
0.4	200	100	0.027	0.027	0.001	0.062	0.060	0.053	1.000	0.613	0.469	1.000	0.400	0.273	4.00
0.4	200	200	0.035	0.035	0.001	0.057	0.057	0.022	1.000	0.988	0.995	1.000	0.981	0.989	4.00
0.6	100	100	0.026	0.025	0.001	0.052	0.052	0.021	1.000	0.623	0.460	1.000	0.424	0.280	4.00
0.6	100	200	0.040	0.039	0.003	0.067	0.066	0.028	1.000	0.979	0.990	1.000	0.970	0.981	4.00
0.6	200	100	0.017	0.016	0.000	0.055	0.052	0.061	1.000	0.631	0.481	1.000	0.413	0.271	4.00
0.6	200	200	0.036	0.037	0.002	0.050	0.051	0.027	1.000	0.983	0.990	1.000	0.974	0.985	4.00
0.8	100	100	0.021	0.020	0.000	0.057	0.055	0.028	1.000	0.600	0.457	1.000	0.417	0.274	4.00
0.8	100	200	0.038	0.037	0.002	0.060	0.061	0.030	1.000	0.974	0.987	1.000	0.963	0.978	4.00
0.8	200	100	0.020	0.019	0.000	0.054	0.054	0.052	1.000	0.627	0.480	1.000	0.430	0.286	4.00
0.8	200	200	0.034	0.035	0.001	0.053	0.054	0.027	1.000	0.976	0.986	1.000	0.967	0.981	4.00

Note: The parameter α controls the percentage of factor loadings that have structural breaks. For Bonferroni statistics and pooled statistics, the superscript “0” denotes that the statistic assumes conditional homoskedasticity and no serial correlation in the residuals; the superscript “GLS” stands for Breitung and Eickmeier’s (2011) estimates based on a quasi-demean transformation; the superscript “HAC” means that HAC estimates are used to compute the statistic. For LM and W , the subscript “0” means that the statistic uses White’s (1980) conditional heteroskedasticity robust estimate; subscripts “B” and “QS” denote statistics based on HAC estimates with Bartlett and QS kernels, respectively. All HAC estimates are based on Newey and West’s method (1994). \hat{r} is the number of factors estimated by IC_{p_1} of Bai and Ng (2002).

Table 2C: Power against a Break at $T/2$

DGPA3: the equivalent factor model with time-invariant loadings has 3 factors.

c^2	N	T	Bonferroni Statistics			Pooled Statistics			W and LM					\hat{r}	
			s_{Bon}^0	s_{Bon}^{GLS}	s_{Bon}^{HAC}	S_{BE}^0	S_{BE}^{GLS}	S_{BE}^{HAC}	W_0	W_B	W_{QS}	LM_0	LM_B		LM_{QS}
3/4	100	100	0.025	0.024	0.001	0.047	0.046	0.019	0.134	0.048	0.044	0.134	0.040	0.038	3.00
3/4	100	200	0.040	0.040	0.005	0.050	0.050	0.033	0.335	0.204	0.199	0.335	0.207	0.203	3.00
3/4	200	100	0.022	0.022	0.000	0.050	0.048	0.022	0.140	0.050	0.054	0.140	0.050	0.046	3.00
3/4	200	200	0.034	0.035	0.003	0.051	0.052	0.038	0.349	0.219	0.211	0.349	0.223	0.214	3.00
3/4	200	500	0.045	0.049	0.010	0.055	0.055	0.047	0.831	0.792	0.784	0.831	0.777	0.785	3.00
3/4	500	200	0.029	0.029	0.002	0.047	0.046	0.040	0.351	0.226	0.215	0.351	0.225	0.218	3.00
3/4	500	500	0.046	0.047	0.010	0.052	0.053	0.047	0.825	0.782	0.779	0.825	0.772	0.779	3.00
1/2	100	100	0.025	0.025	0.001	0.047	0.047	0.016	0.761	0.430	0.394	0.761	0.344	0.287	3.00
1/2	100	200	0.039	0.040	0.005	0.050	0.051	0.029	0.993	0.966	0.957	0.993	0.954	0.947	3.00
1/2	200	100	0.024	0.023	0.001	0.049	0.050	0.019	0.797	0.467	0.431	0.797	0.380	0.307	3.00
1/2	200	200	0.033	0.033	0.003	0.051	0.053	0.032	0.996	0.969	0.963	0.996	0.961	0.956	3.00
1/4	100	100	0.026	0.026	0.001	0.044	0.048	0.026	1.000	0.956	0.941	1.000	0.749	0.691	3.00
1/4	100	200	0.040	0.040	0.003	0.053	0.054	0.028	1.000	1.000	1.000	1.000	1.000	1.000	3.00
1/4	200	100	0.023	0.023	0.000	0.048	0.047	0.059	1.000	0.968	0.960	1.000	0.754	0.704	3.00
1/4	200	200	0.031	0.031	0.002	0.051	0.052	0.031	1.000	1.000	1.000	1.000	1.000	1.000	3.00
0	100	100	0.031	0.031	0.000	0.070	0.074	0.235	1.000	1.000	0.999	1.000	0.779	0.806	3.00
0	100	200	0.043	0.043	0.001	0.074	0.075	0.041	1.000	1.000	1.000	1.000	1.000	1.000	3.00
0	200	100	0.028	0.029	0.000	0.057	0.061	0.770	1.000	1.000	0.999	1.000	0.777	0.814	3.00
0	200	200	0.043	0.044	0.000	0.064	0.066	0.142	1.000	1.000	1.000	1.000	1.000	1.000	3.00

Note: The parameter c controls the ratio of post- and pre-break factor loadings. For Bonferroni statistics and pooled statistics, the superscript “0” denotes that the statistic assumes conditional homoskedasticity and no serial correlation in the residuals; the superscript “GLS” stands for Breitung and Eickmeier’s (2011) estimates based on a quasi-demean transformation; the superscript “HAC” means that HAC estimates are used to compute the statistic. For LM and W , the subscript “0” means that the statistic uses White’s (1980) conditional heteroskedasticity robust estimate; subscripts “B” and “QS” denote statistics based on HAC estimates with Bartlett and QS kernels, respectively. All HAC estimates are based on Newey and West’s method (1994). \hat{r} is the number of factors estimated by IC_{p_1} of Bai and Ng (2002).

Table 3: Size of Structural Break Tests with Unknown Break Date, $r = 3$

						sup- W			exp- W			mean- W			
DGPs	β	P	ω_ϵ	N	T	sup- W_0	sup- W_B	sup- W_{QS}	exp- W_0	exp- W_B	exp- W_{QS}	mean- W_0	mean- W_B	mean- W_{QS}	\hat{r}
N2	0			100	200	0.049	0.011	0.007	0.067	0.011	0.006	0.064	0.015	0.013	3.00
				100	500	0.088	0.028	0.021	0.083	0.030	0.024	0.063	0.034	0.033	3.00
	0			200	200	0.055	0.014	0.009	0.073	0.015	0.011	0.073	0.018	0.017	3.00
				200	500	0.087	0.029	0.020	0.088	0.030	0.023	0.066	0.035	0.034	3.00
	0			500	200	0.060	0.013	0.008	0.079	0.013	0.009	0.073	0.019	0.016	3.00
				200	200	0.062	0.014	0.009	0.075	0.013	0.010	0.071	0.016	0.015	3.00
	0.1	6		200	500	0.090	0.029	0.021	0.085	0.031	0.022	0.063	0.034	0.033	3.02
N3				100	200	0.846	0.059	0.054	0.903	0.066	0.046	0.894	0.099	0.066	3.00
				100	500	0.919	0.083	0.095	0.931	0.091	0.082	0.896	0.117	0.102	3.00
				200	200	0.852	0.061	0.055	0.896	0.068	0.049	0.891	0.100	0.069	3.00
				200	500	0.920	0.092	0.101	0.929	0.100	0.088	0.903	0.119	0.104	3.00
				500	200	0.855	0.066	0.052	0.896	0.069	0.049	0.890	0.104	0.070	3.00
						sup- LM			exp- LM			mean- LM			
DGPs	β	P	ω_ϵ	N	T	sup- LM_0	sup- LM_B	sup- LM_{QS}	exp- LM_0	exp- LM_B	exp- LM_{QS}	mean- LM_0	mean- LM_B	mean- LM_{QS}	\hat{r}
N2	0			100	200	0.033	0.009	0.008	0.042	0.013	0.012	0.045	0.023	0.027	3.00
				100	500	0.043	0.024	0.028	0.051	0.028	0.035	0.054	0.039	0.040	3.00
	0			200	200	0.046	0.012	0.012	0.053	0.019	0.017	0.053	0.029	0.031	3.00
				200	500	0.041	0.019	0.023	0.046	0.024	0.030	0.052	0.038	0.042	3.00
	0			500	200	0.036	0.011	0.008	0.044	0.014	0.011	0.052	0.025	0.026	3.00
				200	200	0.031	0.009	0.007	0.043	0.012	0.011	0.050	0.026	0.025	3.00
	0.1	6		200	500	0.041	0.024	0.026	0.047	0.027	0.031	0.046	0.034	0.037	3.02
N3				100	200	0.826	0.048	0.046	0.867	0.059	0.059	0.856	0.089	0.078	3.00
				100	500	0.896	0.047	0.053	0.909	0.060	0.066	0.875	0.079	0.079	3.00
				200	200	0.840	0.050	0.045	0.875	0.065	0.060	0.856	0.091	0.083	3.00
				200	500	0.895	0.048	0.052	0.908	0.062	0.066	0.880	0.080	0.079	3.00
				500	200	0.847	0.052	0.046	0.879	0.068	0.060	0.861	0.101	0.087	3.00

Note: The nominal size is 5%. The subscript “0” means that the statistic uses White’s (1980) conditional heteroskedasticity robust estimate; subscripts “B” and “QS” denote statistics based on HAC estimates with Bartlett and QS kernels, respectively. All HAC estimates are based on Newey and West’s method (1994). \hat{r} is the number of factors estimated by IC_{p_1} of Bai and Ng (2002).

Table 4A: Power against Unknown Break Date

DGP A1: the equivalent factor model with time-invariant loadings has 4 factors.

b	N, T	sup- W_0	sup- W_B	exp- W_0	exp- W_B	mean- W_0	mean- W_B	\hat{r}
1/3	100,200	0.072	0.019	0.108	0.023	0.121	0.032	3.00
1/3	200,200	0.078	0.022	0.105	0.026	0.117	0.040	3.00
1/3	500,200	0.435	0.359	0.454	0.389	0.462	0.392	3.39
1/3	200,500	0.482	0.435	0.491	0.441	0.484	0.453	3.40
1/3	500,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
2/3	100,200	1.000	0.895	1.000	0.960	1.000	0.953	4.00
2/3	200,200	1.000	0.891	1.000	0.961	1.000	0.951	4.00
2/3	500,200	1.000	0.904	1.000	0.966	1.000	0.958	4.00
2/3	200,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
2/3	500,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
1	100,200	1.000	0.887	1.000	0.950	1.000	0.940	4.00
1	200,200	1.000	0.919	1.000	0.963	1.000	0.952	4.00
1	500,200	1.000	0.901	1.000	0.960	1.000	0.951	4.00
1	200,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
1	500,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
2	100,200	1.000	0.828	1.000	0.937	1.000	0.940	4.00
2	200,200	1.000	0.842	1.000	0.940	1.000	0.937	4.00
2	500,200	1.000	0.830	1.000	0.927	1.000	0.930	4.00
2	200,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
2	500,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
b	N	sup- LM_0	sup- LM_B	exp- LM_0	exp- LM_B	mean- LM_0	mean- LM_B	\hat{r}
1/3	100,200	0.058	0.017	0.076	0.029	0.088	0.048	3.00
1/3	200,200	0.065	0.023	0.081	0.033	0.090	0.050	3.00
1/3	500,200	0.418	0.216	0.432	0.283	0.462	0.392	3.39
1/3	200,500	0.453	0.431	0.464	0.443	0.470	0.453	3.40
1/3	500,500	1.000	0.999	1.000	1.000	1.000	1.000	4.00
2/3	100,200	1.000	0.571	1.000	0.702	1.000	0.847	4.00
2/3	200,200	1.000	0.561	1.000	0.693	1.000	0.833	4.00
2/3	500,200	1.000	0.561	1.000	0.700	1.000	0.848	4.00
2/3	200,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
2/3	500,500	1.000	0.999	1.000	1.000	1.000	1.000	4.00
1	100,200	1.000	0.565	1.000	0.692	1.000	0.822	4.00
1	200,200	1.000	0.615	1.000	0.743	1.000	0.856	4.00
1	500,200	1.000	0.579	1.000	0.703	1.000	0.836	4.00
1	200,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
1	500,500	1.000	0.998	1.000	0.999	1.000	0.999	4.00
2	100,200	1.000	0.466	1.000	0.603	1.000	0.782	4.00
2	200,200	1.000	0.493	1.000	0.626	1.000	0.784	4.00
2	500,200	1.000	0.473	1.000	0.598	1.000	0.765	4.00
2	200,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
2	500,500	1.000	0.997	1.000	0.999	1.000	1.000	4.00

Note: The parameter b controls the size of the shift in factor loadings. The subscript “0” means that the statistic uses White’s (1980) conditional heteroskedasticity robust estimate; subscript “B” denotes statistics based on HAC estimates with Bartlett kernel. All HAC estimates are based on Newey and West’s method (1994). \hat{r} is the number of factors estimated by IC_{p_1} of Bai and Ng (2002).

Table 4B: Power against Unknown Break Date

DGP A2: the equivalent factor model with time-invariant loadings has 4 factors.

α	N, T	sup- W_0	sup- W_B	exp- W_0	exp- W_B	mean- W_0	mean- W_B	\hat{r}
0.2	100,200	0.166	0.106	0.199	0.114	0.216	0.128	3.09
0.2	200,200	0.822	0.745	0.830	0.794	0.833	0.793	3.80
0.2	500,200	1.000	0.911	1.000	0.976	1.000	0.981	4.00
0.2	200,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
0.2	500,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
0.4	100,200	1.000	0.924	1.000	0.975	1.000	0.969	4.00
0.4	200,200	1.000	0.909	1.000	0.969	1.000	0.966	4.00
0.4	500,200	1.000	0.903	1.000	0.973	1.000	0.974	4.00
0.4	200,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
0.4	500,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
0.6	100,200	1.000	0.916	1.000	0.968	1.000	0.961	4.00
0.6	200,200	1.000	0.896	1.000	0.963	1.000	0.961	4.00
0.6	500,200	1.000	0.908	1.000	0.966	1.000	0.967	4.00
0.6	200,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
0.6	500,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
0.8	100,200	1.000	0.895	1.000	0.952	1.000	0.944	4.00
0.8	200,200	1.000	0.907	1.000	0.965	1.000	0.957	4.00
0.8	500,200	1.000	0.889	1.000	0.955	1.000	0.954	4.00
0.8	200,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
0.8	500,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
α	N, T	sup- LM_0	sup- LM_B	exp- LM_0	exp- LM_B	mean- LM_0	mean- LM_B	\hat{r}
0.2	100,200	0.150	0.070	0.172	0.093	0.184	0.128	3.09
0.2	200,200	0.820	0.455	0.826	0.580	0.828	0.708	3.80
0.2	500,200	1.000	0.519	1.000	0.678	1.000	0.852	4.00
0.2	200,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
0.2	500,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
0.4	100,200	1.000	0.600	1.000	0.731	1.000	0.869	4.00
0.4	200,200	1.000	0.551	1.000	0.698	1.000	0.855	4.00
0.4	500,200	1.000	0.520	1.000	0.674	1.000	0.842	4.00
0.4	200,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
0.4	500,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
0.6	100,200	1.000	0.591	1.000	0.721	1.000	0.850	4.00
0.6	200,200	1.000	0.551	1.000	0.695	1.000	0.843	4.00
0.6	500,200	1.000	0.547	1.000	0.689	1.000	0.846	4.00
0.6	200,500	1.000	0.999	1.000	0.999	1.000	1.000	4.00
0.6	500,500	1.000	0.999	1.000	1.000	1.000	1.000	4.00
0.8	100,200	1.000	0.578	1.000	0.699	1.000	0.944	4.00
0.8	200,200	1.000	0.570	1.000	0.706	1.000	0.843	4.00
0.8	500,200	1.000	0.551	1.000	0.685	1.000	0.829	4.00
0.8	200,500	1.000	0.998	1.000	1.000	1.000	1.000	4.00
0.8	500,500	1.000	0.999	1.000	0.999	1.000	0.999	4.00

Note: The parameter α controls the percentage of factor loadings that have structural breaks. The subscript “0” means that the statistic uses White’s (1980) conditional heteroskedasticity robust estimate; subscript “B” denotes statistics based on HAC estimates with Bartlett kernel. All HAC estimates are based on Newey and West’s method (1994). \hat{r} is the number of factors estimated by IC_{p1} of Bai and Ng (2002).

Table 4C: Power against Unknown Break Date

DGP A3: the equivalent factor model with time-invariant loadings has 3 factors.

c^2	N, T	sup- W_0	sup- W_B	exp- W_0	exp- W_B	mean- W_0	mean- W_B	\hat{r}
3/4	100,200	0.204	0.056	0.278	0.084	0.328	0.130	3.00
3/4	200,200	0.202	0.056	0.282	0.087	0.327	0.133	3.00
3/4	500,200	0.204	0.063	.0288	0.096	0.328	0.143	3.00
3/4	200,500	0.678	0.503	0.729	0.600	0.751	0.654	3.00
3/4	500,500	0.673	0.500	0.729	0.597	0.751	0.648	3.00
1/2	100,200	0.941	0.632	0.970	0.798	0.979	0.852	3.00
1/2	200,200	0.952	0.646	0.977	0.816	0.983	0.864	3.00
1/2	500,200	0.957	0.663	0.980	0.827	0.985	0.873	3.00
1/2	200,500	1.000	1.000	1.000	1.000	1.000	1.000	3.00
1/2	500,500	1.000	1.000	1.000	1.000	1.000	1.000	3.00
1/4	100,200	1.000	0.983	1.000	0.999	1.000	0.999	3.00
1/4	200,200	1.000	0.983	1.000	0.998	1.000	0.999	3.00
1/4	500,200	1.000	0.987	1.000	0.999	1.000	0.999	3.00
1/4	200,500	1.000	1.000	1.000	1.000	1.000	1.000	3.00
1/4	500,500	1.000	1.000	1.000	1.000	1.000	1.000	3.00
0	100,200	1.000	0.999	1.000	1.000	1.000	1.000	3.00
0	200,200	1.000	0.999	1.000	1.000	1.000	1.000	3.00
0	500,200	1.000	1.000	1.000	1.000	1.000	1.000	3.00
0	200,500	1.000	1.000	1.000	1.000	1.000	1.000	3.00
0	500,500	1.000	1.000	1.000	1.000	1.000	1.000	3.00
c^2	N, T	sup- LM_0	sup- LM_B	exp- LM_0	exp- LM_B	mean- LM_0	mean- LM_B	\hat{r}
3/4	100,200	0.195	0.071	0.249	0.102	0.282	0.161	3.00
3/4	200,200	0.194	0.061	0.251	0.100	0.285	0.163	3.00
3/4	500,200	0.188	0.063	0.248	0.104	0.285	0.169	3.00
3/4	200,500	0.646	0.482	0.706	0.579	0.723	0.637	3.00
3/4	500,500	0.646	0.478	0.700	0.570	0.724	0.630	3.00
1/2	100,200	0.948	0.533	0.968	0.715	0.969	0.811	3.00
1/2	200,200	0.955	0.534	0.974	0.728	0.974	0.821	3.00
1/2	500,200	0.961	0.538	0.978	0.732	0.976	0.827	3.00
1/2	200,500	1.000	1.000	1.000	1.000	1.000	1.000	3.00
1/2	500,500	1.000	1.000	1.000	1.000	1.000	1.000	3.00
1/4	100,200	1.000	0.773	1.000	0.965	1.000	0.983	3.00
1/4	200,200	1.000	0.759	1.000	0.966	1.000	0.982	3.00
1/4	500,200	1.000	0.746	1.000	0.963	1.000	0.979	3.00
1/4	200,500	1.000	1.000	1.000	1.000	1.000	1.000	3.00
1/4	500,500	1.000	1.000	1.000	1.000	1.000	1.000	3.00
0	100,200	1.000	0.671	1.000	0.991	1.000	0.996	3.00
0	200,200	1.000	0.644	1.000	0.987	1.000	0.994	3.00
0	500,200	1.000	0.633	1.000	0.987	1.000	0.991	3.00
0	200,500	1.000	1.000	1.000	1.000	1.000	1.000	3.00
0	500,500	1.000	1.000	1.000	1.000	1.000	1.000	3.00

Note: The parameter c controls the ratio of post- and pre-break factor loadings. The subscript “0” means that the statistic uses White’s (1980) conditional heteroskedasticity robust estimate; subscript “B” denotes statistics based on HAC estimates with Bartlett kernel. All HAC estimates are based on Newey and West’s method (1994). \hat{r} is the number of factors estimated by IC_{p_1} of Bai and Ng (2002).

Table 5A: Power Comparison with CDG tests for Unknown Break Date

DGP A1: the equivalent factor model with time-invariant loadings has 4 factors.

b	N	T	sup- LM_0	sup- LM_0^{CDG}	sup- LM_B	sup- LM_B^{CDG}	sup- W_0	sup- W_0^{CDG}	sup- W_B	sup- W_B^{CDG}	\hat{r}
0.33	100	200	0.058	0.052	0.017	0.033	0.072	0.071	0.019	0.061	3.00
0.33	200	200	0.065	0.048	0.023	0.029	0.078	0.070	0.022	0.059	3.01
0.33	500	200	0.418	0.389	0.216	0.371	0.435	0.428	0.359	0.419	3.39
0.33	200	500	0.453	0.353	0.431	0.338	0.482	0.445	0.435	0.440	3.40
0.33	500	500	1.000	0.937	0.999	0.931	1.000	1.000	1.000	1.000	4.00
0.66	100	200	1.000	0.902	0.571	0.866	1.000	1.000	0.895	1.000	4.00
0.66	200	200	1.000	0.668	0.561	0.574	1.000	1.000	0.891	1.000	4.00
0.66	500	200	1.000	0.864	0.561	0.817	1.000	1.000	0.904	1.000	4.00
0.66	200	500	1.000	0.746	0.999	0.716	1.000	1.000	1.000	1.000	4.00
0.66	500	500	1.000	0.946	0.999	0.938	1.000	1.000	1.000	1.000	4.00
1	100	200	1.000	0.940	0.565	0.892	1.000	1.000	0.887	1.000	4.00
1	200	200	1.000	0.778	0.614	0.663	1.000	1.000	0.919	1.000	4.00
1	500	200	1.000	0.914	0.579	0.854	1.000	1.000	0.901	1.000	4.00
1	200	500	1.000	0.861	0.997	0.839	1.000	1.000	1.000	1.000	4.00
1	500	500	1.000	0.970	0.998	0.964	1.000	1.000	1.000	1.000	4.00
2	100	200	1.000	1.000	0.465	0.992	1.000	1.000	0.827	0.997	4.00
2	200	200	1.000	1.000	0.494	0.995	1.000	1.000	0.842	0.997	4.00
2	500	200	1.000	1.000	0.473	0.995	1.000	1.000	0.830	0.996	4.00
2	200	500	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	4.00
2	500	500	1.000	1.000	0.997	1.000	1.000	1.000	1.000	1.000	4.00

Note: The parameter b controls the size of the shift in factor loadings. The superscript “CDG” denotes the CDG tests. The subscript “0” means that the statistic uses White’s (1980) conditional heteroskedasticity robust estimate; subscript “B” denotes statistics based on HAC estimates with Bartlett kernel. All HAC estimates are based on Newey and West’s method (1994). \hat{r} is the number of factors estimated by IC_{p_1} of Bai and Ng (2002).

Table 5B: Power Comparison with CDG tests for Unknown Break Date

DGP A3: the equivalent factor model with time-invariant loadings has 3 factors.

c^2	N	T	sup- LM_0	sup- LM_0^{CDG}	sup- LM_B	sup- LM_B^{CDG}	sup- W_0	sup- W_0^{CDG}	sup- W_B	sup- W_B^{CDG}	\hat{r}
3/4	100	200	0.195	0.039	0.071	0.024	0.204	0.049	0.056	0.042	3.00
3/4	200	200	0.194	0.039	0.061	0.028	0.202	0.048	0.056	0.044	3.00
3/4	500	200	0.188	0.038	0.063	0.024	0.204	0.054	0.063	0.043	3.00
3/4	200	500	0.646	0.046	0.482	0.038	0.678	0.045	0.503	0.036	3.00
3/4	500	500	0.646	0.050	0.478	0.043	0.673	0.046	0.500	0.042	3.00
1/2	100	200	0.948	0.055	0.533	0.035	0.941	0.036	0.632	0.035	3.00
1/2	200	200	0.955	0.049	0.534	0.035	0.952	0.040	0.646	0.041	3.00
1/2	500	200	0.961	0.054	0.538	0.034	0.957	0.042	0.663	0.041	3.00
1/2	200	500	1.000	0.068	1.000	0.056	1.000	0.031	1.000	0.030	3.00
1/2	500	500	1.000	0.069	1.000	0.058	1.000	0.032	1.000	0.034	3.00
1/4	100	200	1.000	0.068	0.773	0.049	1.000	0.017	0.983	0.028	3.00
1/4	200	200	1.000	0.056	0.759	0.043	1.000	0.023	0.983	0.035	3.00
1/4	500	200	1.000	0.066	0.746	0.048	1.000	0.024	0.987	0.036	3.00
1/4	200	500	1.000	0.081	1.000	0.070	1.000	0.017	1.000	0.018	3.00
1/4	500	500	1.000	0.084	1.000	0.069	1.000	0.021	1.000	0.023	3.00
0	100	200	1.000	0.072	0.671	0.052	1.000	0.138	0.999	0.210	3.00
0	200	200	1.000	0.062	0.644	0.051	1.000	0.400	0.999	0.487	3.00
0	500	200	1.000	0.067	0.633	0.052	1.000	0.665	1.000	0.728	3.00
0	200	500	1.000	0.082	1.000	0.068	1.000	0.519	1.000	0.563	3.00
0	500	500	1.000	0.082	1.000	0.067	1.000	0.804	1.000	0.828	3.00

Note: The parameter c controls the ratio of post- and pre-break factor loadings. The superscript “CDG” denotes the CDG tests. The subscript “0” means that the statistic uses White’s (1980) conditional heteroskedasticity robust estimate; subscript “B” denotes statistics based on HAC estimates with Bartlett kernel. All HAC estimates are based on Newey and West’s method (1994). \hat{r} is the number of factors estimated by IC_{p_1} of Bai and Ng (2002).