Econometric Theory, **31**, 2015, 1117–1152.
doi:10.1017/S0266466614000486

TESTS FOR PARAMETER INSTABILITY IN DYNAMIC FACTOR MODELS

XU HAN
City University of Hong Kong

ATSUSHI INOUE
Vanderbilt University and Tohoku University

In this paper, 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 postbreak 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.

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 Eliasz, 2005; Stock and Watson, 2005), dynamic stochastic general equilibrium (DSGE) models (Boivin and Giannoni, 2006), and handling a large number of instruments (Bai and Ng, 2010). 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

We would like to thank Zhongjun Qu for helpful conversation, Peter C.B. Phillips and 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 also thank Jesus Gonzalo for pointing out an error in an earlier version of the paper. Address correspondence to Xu Han, Department of Economics and Finance, City University of Hong Kong, Kowloon, Hong Kong SAR; e-mail: xuhan25@cityu.edu.hk or to Atsushi Inoue, Department of Economics, Vanderbilt University, Nashville, TN 37235, USA; email: atsushi.inoue@vanderbilt.edu.

© Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

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 breaks can augment the factor space, but simply introducing more factors cannot solve all the problems for two reasons. First, when a 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 nonnegligible 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

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

implemented a Chow test for each of these regressions. Using a postwar 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 have weak cross-sectional correlation. Also, their joint test is severely oversized in the presence of serial correlations in the idiosyncratic shocks when the Heteroskedasticity and Autocorrelation Consistent (henceforth 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 but the LM test is not consistent under some special cases.

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 that have different impacts on the dimension of factor space. Besides the case pointed out by SW and BE where the number of factors in the full sample 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 in the full sample. 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 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

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

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 $N, T \to \infty$ 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 $\lfloor\cdot\rfloor$ is the integer part operator. 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 provides conclusions. Proofs are relegated to Appendixes 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 ith cross-section at period t for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$. Let $\lfloor \pi^* T \rfloor + 1$ 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:

$$x_{it} = \begin{cases} f_{0,t}'\lambda_{0,i} + f_{1,t}'\lambda_{1,i} + e_{it} & \text{if } 1 \le t \le \lfloor \pi^* T \rfloor \\ f_{0,t}'\lambda_{0,i} + f_{1,t}'\lambda_{2,i} + e_{it} & \text{if } \lfloor \pi^* T \rfloor + 1 \le t \le T, \end{cases} \tag{2.1}$$

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 postbreak 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: $X_i \equiv (x_{i1}, x_{i2}, \ldots, x_{iT})'$ and $e_i \equiv (e_{i1}, e_{i2}, \ldots, e_{iT})'$. The matrix notation of the factor model is:

$$X = \begin{bmatrix} F_{0,1}\Lambda_0' + F_{1,1}\Lambda_1' \\ F_{0,2}\Lambda_0' + F_{1,2}\Lambda_2' \end{bmatrix} + e, \tag{2.2}$$

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

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

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

$$H_0 : \lambda_{1,i} = \lambda_{2,i} \quad \forall i. \tag{2.3}$$

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

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

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 $F_1 \equiv [F_{0,1} \vdots F_{1,1}]$, $F_2 \equiv [F_{0,2} \vdots F_{1,2}]$, $F \equiv [F_1' \vdots F_2']'$, and $\Lambda \equiv [\Lambda_0 \vdots \Lambda_1]$. 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:

$$X = \begin{bmatrix} F_{1,1} & 0_{\lfloor \pi^* T \rfloor \times r} \\ 0_{(T - \lfloor \pi^* T \rfloor) \times r} & F_{1,2} \end{bmatrix} \begin{bmatrix} \Lambda_1' \\ \Lambda_2' \end{bmatrix} + e.$$

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

Let $G_1 \equiv [F_{1,1} \vdots 0_{\lfloor \pi^* T \rfloor \times r}]$ and $G_2 \equiv [0_{(T-\lfloor \pi^* T \rfloor) \times r} \vdots F_{1,2}]$, thus 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 postbreak subsample means of $\hat{f}_t \hat{f}_t'$, where $\hat{f}_t$ is the PCA estimate of the factors. Let

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

$\hat{\mathbb{S}}(\pi, \hat{F})$ and $\tilde{\mathbb{S}}(\pi, \hat{F})$ 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

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

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

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 $\Lambda \equiv [\Lambda_0 \vdots \Lambda_1]$ 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 $T \to \infty$ for some positive definite matrix Σ_F.

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

Assumption 2. $\|\lambda_i\| \leq \bar{\lambda} < \infty$, $\|\Lambda'\Lambda/N - \Sigma_\Lambda\| \to 0$ 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)

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

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

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

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

(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 are discussed in Section 2.5 (see note 4). Given this identification issue, it is not restrictive to assume that factors have constant second moment.

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 Assumption 5 is a strengthened version of Assumption 3. Assumptions 3 and 5 also allow heterogeneity in time and cross-sectional dimensions. Thus, this paper allows weaker assumptions on idiosyncratic shocks than BE, who assume that the idiosyncratic shocks are independent in the cross-sectional dimension and follow AR(p) processes. Assumption 4 implies that $E(N^{-1}\sum_{i=1}^{N}\|T^{-1/2}\sum_{t=1}^{T}f_t e_{it}\|^2)\le 4M$. Parts (a) and (b) of Assumption 6 imply that (a) for each t, $E\|(NT)^{-1/2}\sum_{s=1}^{T}\sum_{k=1}^{N}f_s[e_{ks}e_{kt}-E(e_{ks}e_{kt})]\|^2\le 4M$; and (b) $E\|(NT)^{-1/2}\sum_{t=1}^{T}\sum_{k=1}^{N}f_t\lambda_k'e_{kt}\|^2\le 8M$. These correspond to Assumptions F1 and F2 of Bai (2003). The role of the indicator functions ι_{1t} and ι_{2t} will become clear under the alternative hypothesis. 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 $\|.\|$ 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 helpful to describe some useful notations and existing results. Let V_{NT} be the $r\times r$ diagonal matrix of

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

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 correspond 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) shows 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 $H_0 \equiv \text{plim}_{T,N\to\infty}H$, so it is obvious that

$$E(H_0'f_t f_t'H_0) = H_0'\Sigma_F H_0 = I_r, \tag{2.8}$$

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 $A(\pi, FH_0) \equiv \text{vech}\big(\sqrt{T}\big(\frac{1}{\lfloor \pi T\rfloor}\sum_{t=1}^{\lfloor \pi T\rfloor} H_0'f_t f_t'H_0 - \frac{1}{T-\lfloor \pi T\rfloor}\sum_{t=\lfloor \pi T\rfloor+1}^{T} H_0'f_t f_t'H_0\big)\big)$. 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* $\frac{\sqrt{T}}{N} \to 0$ *as* $N, T \to \infty$, *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

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

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 postbreak 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

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

likely to be serially correlated, we consider the following estimates for the sample variances:

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

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

$$\hat{\Gamma}_{1,j}(\pi, FH_0) = \frac{1}{\lfloor \pi T \rfloor} \sum_{t=j+1}^{\lfloor \pi T \rfloor} \mathrm{vech}(H_0' f_t f_t' H_0 - I_r)\mathrm{vech}(H_0' f_{t-j} f_{t-j}' H_0 - I_r)',$$

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

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

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

where

$$\hat{\Gamma}_j(FH_0) = \frac{1}{T} \sum_{t=j+1}^{T} \mathrm{vech}(H_0' f_t f_t' H_0 - I_r)\mathrm{vech}(H_0' f_{t-j} f_{t-j}' H_0 - I_r)'. \tag{2.12}$$

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 bandwidth parameter, and its subscript denotes the size of the sample (or subsample) that is used to estimate the long-run variance. Let

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

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

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

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

Note that all of $\hat{\Omega}_m(\pi, FH_0)$, $\hat{\Gamma}_{m,j}(\pi, FH_0)$, for $m = 1, 2$, $\hat{\Omega}(FH_0)$, $\hat{\Gamma}_j(FH_0)$, $\hat{\mathbb{S}}(\pi, FH_0)$, and $\tilde{\mathbb{S}}(\pi, FH_0)$ 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 $m = 1, 2$, $\hat{\Omega}(\hat{F})$, $\hat{\Gamma}_j(\hat{F})$, $\hat{\mathbb{S}}(\pi, \hat{F})$, and $\tilde{\mathbb{S}}(\pi, \hat{F})$ as the feasible analogs computed using the estimated regressors $\hat{F}$.

Condition 1. (a) The Bartlett kernel is used to estimate $\hat{\mathbb{S}}(\pi, FH_0)$, $\hat{\mathbb{S}}(\pi, \hat{F})$, $\tilde{\mathbb{S}}(\pi, FH_0)$, and $\tilde{\mathbb{S}}(\pi, \hat{F})$, 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) $\frac{T^{\frac{2}{3}}}{N} \to 0$ as N, $T \to \infty$.

Condition 2. (a) The Parzen kernel is used to estimate $\hat{\mathbb{S}}(\pi, FH_0)$, $\hat{\mathbb{S}}(\pi, \hat{F})$, $\tilde{\mathbb{S}}(\pi, FH_0)$, and $\tilde{\mathbb{S}}(\pi, \hat{F})$, 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 $\hat{\mathbb{S}}(\pi, FH_0)$, $\hat{\mathbb{S}}(\pi, \hat{F})$, $\tilde{\mathbb{S}}(\pi, FH_0)$, and $\tilde{\mathbb{S}}(\pi, \hat{F})$, 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) $\frac{T^{\frac{2}{5}}}{N} \to 0$ as N, $T \to \infty$.

THEOREM 2. *Under Assumptions 1–7, if Condition 1 or Condition 2 holds, then*

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

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

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

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

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

and we make the following assumption:

Assumption 9. (a) $\Omega \equiv \lim_{T \to \infty} \mathrm{Var}\left(\mathrm{vech}\left(\frac{1}{\sqrt{T}} \left(\sum_{t=1}^{T} H_0' f_t f_t' H_0 - I_r \right) \right) \right)$ 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

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

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

(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 $\left(p = \frac{r(r+1)}{2}\right)$ 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* $\frac{\sqrt{T}}{N} \to 0$ *as* $N, T \to \infty$, *then*

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

(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* $\frac{\sqrt{T}}{N} \to 0$ *as* $N, T \to \infty$, *then*

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

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

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

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* $mean\text{-}W(\hat{F}) \equiv \frac{1}{\pi_2-\pi_1}\int_{\pi_1}^{\pi_2} W_T(\pi,\hat{F})d\pi$.

Corollary 1 shows that critical values provided by Andrews and Ploberger (1994) can be applied to exp-$W(\hat{F})$ and mean-$W(\hat{F})$ as if $\hat{F}$ is observed rather than estimated. This result also holds for exp-$LM(\hat{F})$ and 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 $\lfloor \alpha N \rfloor$ many cross-sections have a single break at a common break date in their loadings. The model (2.2) can be rewritten as:

$$X = \begin{bmatrix} F_{0,1} & F_{1,1} & 0 \\ F_{0,2} & 0 & F_{1,2} \end{bmatrix} \begin{bmatrix} \Lambda_0' \\ \Lambda_1' \\ \Lambda_2' \end{bmatrix} + e. \tag{2.15}$$

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 $[\Lambda_0 \vdots \Lambda_1 \vdots \Lambda_2]$ 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

$$X = G\Theta' + e. \tag{2.16}$$

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 $G = (g_1, g_2, \ldots, g_T)'$ and $\Theta = (\theta_1, \theta_2, \ldots, \theta_N)'$. If G has full column rank, the number of factors in (2.16) is determined by the rank of $\lim_{N\to\infty} \Theta'\Theta/N$. 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.

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

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

The type 1 break requires the column rank of $[\Lambda_1 \vdots \Lambda_2]$ 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 the type 1 break, (2.15) is equivalent to a factor model that has $r + q_1$ factors with the time-invariant factor loading matrix:

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

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 $\Theta_{(1)} = [\Lambda_0 \vdots \Lambda_1 \vdots \Lambda_2]_{N\times(r+q_1)}$ is the factor loading matrix.

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

Under the type 2 break, the column rank of $[\Lambda_1 \vdots \Lambda_2]$ 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 $\Lambda_1 = \Lambda_2 Z$). The 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. This type of break is important because it is possible that the number of factors reduces after a major recession or crisis. Hence, the 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 the type 2 break:

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

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 $\Theta_{(2)} = [\Lambda_0 \vdots \Lambda_1]_{N\times r}$ is the factor loading matrix. Note that (2.18) has the same loading matrix as under the null, i.e., $\Lambda \equiv [\Lambda_0 \vdots \Lambda_1]$. The change in Λ is transmitted to the factors, and $F_{1,2}$ is rotated by Z'.[4]

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

Under the type 3 break, the column rank of $[\Lambda_1 \vdots \Lambda_2]$ 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 the type 1 break, the dimension of the factor space is augmented; for the type 2 break, the factors are rotated with unchanged loadings. It is not difficult to see that the type 3 break will lead to an equivalent model with characteristics of both type 1 and 2 breaks. The linearly dependent columns of Λ_1 and Λ_2 will transform the factors by rotation as under the type 2 break, while the linearly independent columns of Λ_1 and Λ_2 will augment the factor space as under the type 1 break. Without loss

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

of the main insight, our analysis will focus on breaks of type 1 and 2 to avoid introducing more tedious notations associated with the type 3 break. The power properties of our statistics under the type 3 break will be the combination of those under types 1 and 2.

Assumption 10 (Conditions on the Break)**.** Let $\Sigma_{G_{(j)}} \equiv \text{plim}_{T\to\infty} G_{(j)}G_{(j)}/T$ and $\Sigma_{\Theta_{(j)}} = \lim_{N\to\infty} \Theta'_{(j)}\Theta_{(j)}/N$ for $j = 1, 2$.

(a) Type 1 break: $\Sigma_{G_{(1)}}$ is positive definite. $\|\lambda_{2,i}\| \leq \bar{\lambda} < \infty$, $\|\Theta'_{(1)}\Theta_{(1)}/N - \Sigma_{\Theta_{(1)}}\| \to 0$ 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 type 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 (Bai, 2003, Lem. A3). 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 $\text{plim}_{T,N\to\infty} J$ as J_0, which is a nonsingular matrix (by Prop. 1 of Bai, 2003).[5] Let

$$D_1 \equiv \begin{bmatrix} \Sigma_F^{0,0} & \Sigma_F^{0,1} & 0 \\ \Sigma_F^{0,1'} & \Sigma_F^{1,1} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad D_2 \equiv \begin{bmatrix} \Sigma_F^{0,0} & 0 & \Sigma_F^{0,1} \\ 0 & 0 & 0 \\ \Sigma_F^{0,1'} & 0 & \Sigma_F^{1,1} \end{bmatrix}, \quad \text{and}$$

$$C \equiv J_0'(D_1 - D_2)J_0. \tag{2.19}$$

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

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 $\Sigma_{\Theta_{(2)}}^{\frac{1}{2}} \Sigma_{G_{(2)}} \Sigma_{\Theta_{(2)}}^{\frac{1}{2}}$ in descending order (Bai and Ng, 2003, Lem. A3). Let $J = (\Theta'_{(2)}\Theta_{(2)}/N)(G'_{(2)}\hat{F}/T)U_{NT}^{-1}$ be an $r \times r$ matrix. Denote $\text{plim}_{T,N\to\infty} J$ as J_0, which is a nonsingular matrix (by Prop. 1 of Bai, 2003). Let

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

To establish the consistency of the test under the alternative hypothesis, we need Assumption 11, which regulates the asymptotic property of the variance matrices in our statistics. Define $\hat{\Omega}_1(\pi^*, GJ_0)$, $\hat{\Omega}_2(\pi^*, GJ_0)$, and $\hat{\Omega}(GJ_0)$ 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 $\hat{\mathbb{S}}(\pi^*, GJ_0) = \frac{1}{\pi^*}\hat{\Omega}_1(\pi^*, GJ_0) + \frac{1}{1-\pi^*}\hat{\Omega}_2(\pi^*, GJ_0)$ and $\tilde{\mathbb{S}}(\pi^*, GJ_0) = \frac{1}{\pi^*}\hat{\Omega}(GJ_0) + \frac{1}{1-\pi^*}\hat{\Omega}(GJ_0)$.

Assumption 11. (a)

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

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 $\hat{\Omega}(GJ_0)$, 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. In contrast to equation (2.8), it is worth noting that $E(J_0' g_t g_t' J_0) \neq I_{r+q_1}$ 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 $b_T \mu_*' \mathfrak{S}^{-1} \mu_*$ converges in probability to a positive constant (see Hall, 2000, Proof of Theorem 2, pp. 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 vech(C) is an analog of μ_*, and our bandwidth parameters are analogs of b_T. Hence, Assumption 11(a) is analogous to Hall's

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

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 nonrandom matrix* $C \neq 0$, *such that* $\frac{1}{\lfloor \pi^* T \rfloor} \sum_{t=1}^{\lfloor \pi^* T \rfloor} \hat{f}_t \hat{f}_t' - \frac{1}{T - \lfloor \pi^* T \rfloor} \sum_{t=\lfloor \pi^* T \rfloor + 1}^{T} \hat{f}_t \hat{f}_t' \xrightarrow{p} C$.

(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* $\lfloor \pi^* T \rfloor$.

Theorem 4(i) shows that pre- and postbreak 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 the 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 the 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 (2002) 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 (2002) information criteria consistently estimate the number of factors of the equivalent models (2.17) and (2.18).*

IC are equivalent to determining the number of asymptotically nonzero eigenvalues of XX'/NT. Assumptions 1–4 play the role of Assumptions A–D in 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 are 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

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

the alternative hypothesis because the factor loadings of the equivalent models are actually time-invariant.[6] Therefore, in order to test structural breaks in factor loadings, one should focus on the estimated factors rather than the estimated factor loadings.

3. MONTE CARLO SIMULATIONS

In the Monte Carlo experiments, we investigate the finite sample properties of our statistics for unknown break points.[7] Section 3.1 compares the performance of the sup-W, exp-W, mean-W, sup-LM, exp-LM, and mean-LM tests, where W and LM abbreviate $W_T(\pi, \hat{F})$ and $LM_T(\pi, \hat{F})$, respectively. 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; and W_B and LM_B are computed using Newey and West's (1994) data-dependent HAC estimate based on the Bartlett kernel. Section 3.2 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 we then fixed throughout 5,000 replications for each DGP.

3.1. Testing Breaks with Unknown Break Date

We consider the model: $x_{it} = \sum_{k=1}^{r} \lambda_{ik} f_{kt} + \kappa e_{it}$, where $\lambda_{ik} \overset{iid}{\sim} N(\frac{b}{2}, 1)$, and f_{kt} and e_{it} are generated by the following DGPs:

N1: $f_{kt} \overset{iid}{\sim} N(0,1)$, $e_{it} = \sigma_i(v_{it} + \sum_{1\le|j|\le P} \beta v_{i-j,t})$, $\sigma_i \overset{iid}{\sim} U(0.5, 1.5)$, $v_{it} \overset{iid}{\sim} N(0,1)$, and $\kappa = \sqrt{12(1+b^2/4)r/13(1+2P\beta^2)}$.

N2: $f_{kt} = \rho_f f_{kt-1} + \mu_{it}$, $\mu_{it} \overset{iid}{\sim} N(0, 1-\rho_f^2)$, $e_{it} = \sigma_i v_{it}$, $\sigma_i \overset{iid}{\sim} U(0.5, 1.5)$, $v_{it} = \rho_v v_{it-1} + \epsilon_{it}$, $\epsilon_{it} \overset{iid}{\sim} N(0, 1-\rho_v^2)$, and $\kappa = \sqrt{12(1+b^2/4)r/13}$.

In N1–N2, 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%.[8] Both N1 and N2 allow heteroskedasticity across i, and we follow Breitung and Eickmeier's (2011) setup: $\sigma_i \overset{iid}{\sim} U(0.5, 1.5)$. N1 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 $P = 6$, and these values are similar to those of Onatski (2010). DGP N2 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. v_{it} follows an AR(1) process and we set $\rho_v = 0.5$.

Table 1 presents the simulation results under the null hypothesis. The last column of Table 1 shows averaged numbers of factors selected by IC_{p_1} of Bai and Ng (2002). Under DGP N1, the size of our tests is not affected by the cross-sectional correlation of the idiosyncratic shocks. Note that N1 does not allow

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

TABLE 1. Size of structural break tests with unknown break date, $r = 3$

					sup-W		exp-W		mean-W		
DGPs	β	P	N	T	sup-W_0	sup-W_B	exp-W_0	exp-W_B	mean-W_0	mean-W_B	$\hat{r}$
N1	0		200	200	0.055	0.014	0.073	0.015	0.073	0.018	3.00
	0		200	500	0.087	0.029	0.088	0.030	0.066	0.035	3.00
	0.1	6	200	200	0.062	0.014	0.075	0.013	0.071	0.016	3.00
	0.1	6	200	500	0.090	0.029	0.085	0.031	0.063	0.034	3.02
N2			200	200	0.852	0.061	0.896	0.068	0.891	0.100	3.00
			200	500	0.920	0.092	0.929	0.100	0.903	0.119	3.00
					sup-LM		exp-LM		mean-LM		
DGPs	β	P	N	T	sup-LM_0	sup-LM_B	exp-LM_0	exp-LM_B	mean-LM_0	mean-LM_B	$\hat{r}$
N1	0		200	200	0.046	0.012	0.053	0.019	0.053	0.029	3.00
	0		200	500	0.041	0.019	0.046	0.024	0.052	0.038	3.00
	0.1	6	200	200	0.031	0.009	0.043	0.012	0.050	0.026	3.00
	0.1	6	200	500	0.041	0.024	0.047	0.027	0.046	0.034	3.02
N2			200	200	0.840	0.050	0.875	0.065	0.856	0.091	3.00
			200	500	0.895	0.048	0.908	0.062	0.880	0.080	3.00

Note: The nominal size is 5%. The subscript "0" means that the statistic uses White's (1980) conditional heteroskedasticity robust estimate; subscript "B" denotes statistics based on HAC estimates with the 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).

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 N2, we consider the case where both factors and idiosyncratic shocks are AR(1) processes. As expected, the 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 almost always have better size than those based on W_B. Under DGP N2, sup-LM has the best size: the size of sup-LM_B is clustered at around 5%. The exp-LM test also has decent size: the size of exp-LM_B is clustered at 6%. The mean-W test over-rejects most frequently, but their size never exceeds 12%.

The second experiment compares the powers of sup-W, exp-W, mean-W, sup-LM, exp-LM, and mean-LM, when the break date is unknown. 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 $i = 1, 2, \ldots, N$ and $t \leq T/2$, and $x_{it} = \sum_{k=1}^{r} (\lambda_{ik} - b) f_{kt} + \kappa e_{it}$ for $i = 1, 2, \ldots, N$ and $t \geq T/2 + 1$, where f_{kt}, $e_{it} \overset{iid}{\sim} N(0, 1)$, $\kappa = \sqrt{(1 + b^2/4)r}$, and $\lambda_{ik} \overset{iid}{\sim} N(\frac{b}{2}, 1)$.

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

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

A3: $x_{it} = \sum_{k=1}^{r} \lambda_{ik} f_{kt} + \kappa e_{it}$ for $i = 1, 2, \ldots, N$, where f_{kt}, $e_{it} \overset{iid}{\sim} N(0,1)$, $\kappa = \sqrt{(1 + b^2/4)r}$, $\lambda_{ik} = \lambda_{ik}^0$ for $t \le T/2$, $\lambda_{ik} = c \cdot \lambda_{ik}^0$ for $t \ge T/2 + 1$, $\lambda_{ik}^0 \overset{iid}{\sim} N(\frac{b}{2}, 1)$, 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 the 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 the break in factor loadings increases. We set $b \in \{1/3,\ 2/3,\ 1,\ 2\}$ in DGP A1, and the results

TABLE 2A. 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	200,200	0.078	0.022	0.105	0.026	0.117	0.040	3.00
1/3	500,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
2/3	200,200	1.000	0.891	1.000	0.961	1.000	0.951	4.00
2/3	500,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
1	200,200	1.000	0.919	1.000	0.963	1.000	0.952	4.00
1	500,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
2	200,200	1.000	0.842	1.000	0.940	1.000	0.937	4.00
2	500,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
b	N, T	sup-LM_0	sup-LM_B	exp-LM_0	exp-LM_B	mean-LM_0	mean-LM_B	$\hat{r}$
1/3	200,200	0.065	0.023	0.081	0.033	0.090	0.050	3.00
1/3	500,500	1.000	0.999	1.000	1.000	1.000	1.000	4.00
2/3	200,200	1.000	0.561	1.000	0.693	1.000	0.833	4.00
2/3	500,500	1.000	0.999	1.000	1.000	1.000	1.000	4.00
1	200,200	1.000	0.615	1.000	0.743	1.000	0.856	4.00
1	500,500	1.000	0.998	1.000	0.999	1.000	0.999	4.00
2	200,200	1.000	0.493	1.000	0.626	1.000	0.784	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_{p1} of Bai and Ng (2002).

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

are summarized in Table 2A. Under DGP A1, none of these tests have good power when the magnitude of the break and sample size are small, i.e., $b = 1/3$ and N and $T = 200$. Note that IC only detect three factors when $b = 1/3$, and N and $T = 200$. By Stock and Watson's (2002b) result, these three factors consistently estimate the original factor space because the break is "small". Hence, the lack of power in such a circumstance is not a serious problem. As b or the sample size increases, however, all of these tests become powerful, though the power is not monotonically increasing in b.

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. All tests have good power for $N \geq 200$ even when $\alpha = 0.2$. For even smaller samples, the working paper version of this paper (Han and Inoue, 2014) provides simulation results that the tests become much less powerful when $N = 100$, $T = 200$, and $\alpha = 0.2$. This is expected due to the recent finding by Bates et al. (2012) that PCA estimators can consistently estimate the

TABLE 2B. 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	200,200	0.822	0.745	0.830	0.794	0.833	0.793	3.80
0.2	500,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
0.4	200,200	1.000	0.909	1.000	0.969	1.000	0.966	4.00
0.4	500,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
0.6	200,200	1.000	0.896	1.000	0.963	1.000	0.961	4.00
0.6	500,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
0.8	200,200	1.000	0.907	1.000	0.965	1.000	0.957	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	200,200	0.820	0.455	0.826	0.580	0.828	0.708	3.80
0.2	500,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
0.4	200,200	1.000	0.551	1.000	0.698	1.000	0.855	4.00
0.4	500,500	1.000	1.000	1.000	1.000	1.000	1.000	4.00
0.6	200,200	1.000	0.551	1.000	0.695	1.000	0.843	4.00
0.6	500,500	1.000	0.999	1.000	1.000	1.000	1.000	4.00
0.8	200,200	1.000	0.570	1.000	0.706	1.000	0.843	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).

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

TABLE 2C. 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	200,200	0.202	0.056	0.282	0.087	0.327	0.133	3.00
3/4	500,500	0.673	0.500	0.729	0.597	0.751	0.648	3.00
1/2	200,200	0.952	0.646	0.977	0.816	0.983	0.864	3.00
1/2	500,500	1.000	1.000	1.000	1.000	1.000	1.000	3.00
1/4	200,200	1.000	0.983	1.000	0.998	1.000	0.999	3.00
1/4	500,500	1.000	1.000	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	500,500	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	200,200	0.194	0.061	0.251	0.100	0.285	0.163	3.00
3/4	500,500	0.646	0.478	0.700	0.570	0.724	0.630	3.00
1/2	200,200	0.955	0.534	0.974	0.728	0.974	0.821	3.00
1/2	500,500	1.000	1.000	1.000	1.000	1.000	1.000	3.00
1/4	200,200	1.000	0.759	1.000	0.966	1.000	0.982	3.00
1/4	500,500	1.000	1.000	1.000	1.000	1.000	1.000	3.00
0	200,200	1.000	0.644	1.000	0.987	1.000	0.994	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 prebreak 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).

factor space when only a "small" (asymptotically negligible) fraction of factor loadings have breaks. However, the power increases substantially and approaches one as the sample size increases, with $\alpha = 0.2$ unchanged.

DGP A3 considers the power against the type 2 break. The post-break loadings are equal to prebreak 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. Under DGP A3, all tests have limited power when $c^2 = 0.75$ and $T = 200$. It is remarkable that mean tests are more powerful than supreme and exponential tests when $c^2 = 0.75$. This is consistent with the finding 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.[9] 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.

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

3.2. Power Comparison with Chenet al.'s (2012) Test

Chen et al. (2012) propose a test for structural breaks in factor loadings by testing whether or not there is a break in the coefficients of the following regression:

$$\hat{f}_{1t} = \phi_2 \hat{f}_{2t} + \cdots + \phi_{\hat{r}} \hat{f}_{\hat{r}t} + u_t = \phi' \hat{f}_{-1t} + u_t, \tag{3.1}$$

where $\hat{r}$ is the number of factors determined by Bai and Ng's (2002) IC, $\hat{f}_{1t}, \ldots, \hat{f}_{\hat{r}t}$ are the factors associated with the $1^{st}, \ldots, \hat{r}^{th}$ eigenvalues of XX', respectively, $\hat{f}_{-1t} \equiv [\hat{f}_{2t} \ldots \hat{f}_{\hat{r}t}]'$, and $\phi \equiv [c_2 \ldots c_{\hat{r}}]'$. They develop both Wald and LM versions for their test.

We perform power comparisons between our tests and the CDG tests under DGPs A1 and A3 for an unknown break date. Table 3A reports the results under DGP A1 (the superscript "cdg" denotes the CDG tests). Under DGP A1, neither our test nor the 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 $N = T = 500$, while $\sup - W_B^{cdg}$ tends to be more powerful when $N = T = 200$.

Table 3B reports the results under DGP A3. In general, both our supreme Wald and supreme LM tests are powerful under DGP A3. The case of $c = 0$ is especially interesting because the number of factors changes after the break. Note that this type of alternative is ruled out by CDG's assumptions. 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 $N = T = 500$. In addition, $\sup - W_0^{cdg}$ and $\sup - W_B^{cdg}$ do have some power but are less powerful than our

TABLE 3A. 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}	$\hat{r}$
1/3	200,200	0.065	0.048	0.023	0.029	0.078	0.070	0.022	0.059	3.01
1/3	500,500	1.000	0.937	0.999	0.931	1.000	1.000	1.000	1.000	4.00
2/3	200,200	1.000	0.668	0.561	0.574	1.000	1.000	0.891	1.000	4.00
2/3	500,500	1.000	0.946	0.999	0.938	1.000	1.000	1.000	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	500,500	1.000	0.970	0.998	0.964	1.000	1.000	1.000	1.000	4.00
2	200,200	1.000	1.000	0.494	0.995	1.000	1.000	0.842	0.997	4.00
2	500,500	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).

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

TABLE 3B. 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	200,200	0.194	0.039	0.061	0.028	0.202	0.048	0.056	0.044	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	200,200	0.955	0.049	0.534	0.035	0.952	0.040	0.646	0.041	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	200,200	1.000	0.056	0.759	0.043	1.000	0.023	0.983	0.035	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	200,200	1.000	0.062	0.644	0.051	1.000	0.400	0.999	0.487	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 prebreak 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 the 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).

tests when $c = 0$. Hence, our tests have some advantages in detecting type 2 breaks in factor loadings.

Remark 2. Note that the power of our tests depends on the consistency of 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 the type 1 break or $\hat{r} > r$ under the 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

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

power against the setup (e.g., Bai, 1997) in which there are finitely many breaks in factor loadings. In addition, Choi (2012) proposes a feasible generalized principal component estimator (FGPCE) for dynamic factors. The key insight of our test is unchanged in Choi's (2012) framework, so we expect that our test can be applied to factors estimated by FGPCE as well.

Finally, our tests are developed for linear factor models which do not take into account possible nonlinear structures in the high dimensional data. In practice, a rejection of the null hypothesis of no structural break may be due to the presence of a nonlinear factor structure. Developing a test for nonlinear factor models is beyond the scope of this paper and is left for future research.

NOTES

1. Recently, Cheng, Liao, and Schorfheide (2013) consider detecting structural changes of factor loadings using a LASSO estimator.

2. In 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 is 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 computing the covariance using the filtered residuals, but this operation requires correct knowledge about the serial correlation structure of the residuals.

3. Since the factor loadings are assumed to be nonrandom, $\Lambda'\Lambda/N$ 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.

4. Recall 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 exists 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 prebreak loading is Λ, and post-break loading is $\Lambda\Psi$. In such a scenario, the type 2 break in the loading matrix could also be viewed as a change in the volatility of factors, but one cannot tell whether the break comes from $E(f_t f_t')$ or the loading matrix because they are observationally equivalent. Hence, we assume $E(f_t f_t')$ to be constant for identification purposes and view the break in $E(f_t f_t')$ as a type 2 break in Λ.

5. The 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.

6. Breitung and Eickmeier (2011) also point out that the Chow test statistic will lack power if the number of factors is determined using IC based on the full sample, so they suggest implementing IC to both pre- and postbreak 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.

7. We also conduct simulations for the case of known break dates and other sample sizes in the working paper version of this paper (Han and Inoue, 2014).

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

9. Note 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.

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

REFERENCES

Andrews, D.W.K. (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821–856.

Andrews, D.W.K. & W. Ploberger (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383.

Bai, J. (1997) Estimating multiple breaks one at a time. *Econometric Theory* 13, 315–352.

Bai, J. (2003) Inferential theory for factor models of large dimensions. *Econometrica* 71, 135–172.

Bai, J. & S. Ng (2002) Determining the number of factors in approximate factor models. *Econometrica* 70, 191–221.

Bai, J. & S. Ng (2010) Instrumental variable estimation in a data rich environment. *Econometric Theory* 26, 1577–1606.

Banerjee, A. & M. Marcellino (2008) Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Change. CEPR Working paper 6706.

Bates, B., M. Plagborg-Moller, J. Stock, & M. Watson (2012) Consistent factor estimation in dynamic factor models with structural instability. *Journal of Econometrics*, forthcoming.

Bernanke, B.S., J. Boivin, & P. Eliasz (2005) Measuring the effects of monetary policy: A Factor-Augmented Vector Autoregressive (FAVAR) approach. *Quarterly Journal of Economics* 120, 387–422.

Boivin, J. & M. Giannoni (2006) DSGE Models in a Data-Rich Environment. NBER Working paper no. 12772.

Breitung, J. & S. Eickmeier (2011) Testing for structural breaks in dynamic factor models. *Journal of Econometrics* 163, 71–84.

Chen, L., J. Dolado, & J. Gonzalo (2012) Detecting Big Structural Breaks in Large Factor Models. MPRA Working paper.

Cheng, X., Z. Liao, & F. Schorfheide (2013) Shrinkage Estimation of Dynamic Factor Models with Structural Instabilities. Manuscript, University of Pennsylvania.

Choi, I. (2012) Efficient estimation of factor models. *Econometric Theory* 28, 274–308.

Eickmeier, S., W. Lemke, & M. Marcellino (2011) Classical Time-Varying FAVAR Models – Estimation, Forecasting and Structural Analysis. CEPR Discussion paper DP8321.

Hall, A.R. (2000) Covariance matrix estimation and the power of the overidentifying restrictions test. *Econometrica* 68, 1517–1527.

Han, X. & A. Inoue (2014) Tests for Parameter Instability in Dynamic Factor Models. Manuscript, City University of Hong Kong and Southern Methodist University. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2314338.

Ledoit, O. & M. Wolf (2004) A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* 88, 365–411.

Newey, W.K. & K.D. West (1994) Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies* 61(4), 631–653.

Onatski, A. (2010) Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics* 92(4), 1004–1016.

Sargent, T.J. & 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.

Stock, J.H. & M.W. Watson (1996) Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics* 14, 11–30.

Stock, J.H. & M.W. Watson (2002a) Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* 20, 147–162.

Stock, J.H. & M.W. Watson (2002b) Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97, 1167–1179.

Stock, J.H. & M.W. Watson (2005) Implications of Dynamic Factor Models for VAR Analysis. NBER Working paper no. 11467.

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

Stock, J.H. & M.W. Watson (2009) Forecasting in dynamic factor models subject to structural instability. In J. Castle & N. Shephard (eds.), *The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry*. Oxford University Press.

White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–838.

APPENDIX

To save space, we only present the statements of lemmas in this appendix. The complete proofs of the lemmas can be found in the working paper version (Han and Inoue, 2014).

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:

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

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

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

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

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.

LEMMA 1 (Lem. 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 (Lem. 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 $0 < \pi_1 \le \pi_2 < 1$,*

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

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

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

Proof of Theorem 1. It is sufficient to prove that

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

and

$$II \equiv \sup_{\pi\in[\pi_1,\pi_2]} \sqrt{T} \left\| \operatorname{vech}\left(\frac{\sum_{t=1}^{\lfloor \pi T\rfloor} H' f_t f_t' H}{\lfloor \pi T\rfloor} - \frac{\sum_{t=\lfloor \pi T\rfloor+1}^{T} H' f_t f_t' H}{T-\lfloor \pi T\rfloor} \right) - \operatorname{vech}\left(\frac{\sum_{t=1}^{\lfloor \pi T\rfloor} H_0' f_t f_t' H_0}{\lfloor \pi T\rfloor} - \frac{\sum_{t=\lfloor \pi T\rfloor+1}^{T} H_0' f_t f_t' H_0}{T-\lfloor \pi T\rfloor} \right) \right\|$$

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

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

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

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

Note that

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

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

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

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

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

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

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

Before proving Theorem 2, we present Lemmas 4–8.

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

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,*

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

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

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

LEMMA 7. (i) *Under Assumptions 1–4,*

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

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

$$\sup_{0\le j\le T-1}\left\|\hat{\Gamma}_j(\hat{F})-\hat{\Gamma}_j(FH)\right\| = O_p\left(T^{-\frac{1}{4}}\right)+O_p\left(N^{-\frac{1}{2}}\right).$$

(ii) *Under Assumptions 1–5,*

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

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

$$\sup_{0\le j\le T-1}\left\|\hat{\Gamma}_j(\hat{F})-\hat{\Gamma}_j(FH)\right\| = O_p\left(\frac{1}{\delta_{NT}}\right).$$

LEMMA 8. *Under Assumptions 1–7,*

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

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

$$\sup_{0\le j\le T-1}\left\|\hat{\Gamma}_j(FH)-\hat{\Gamma}_j(FH_0)\right\| = O_p\left(\frac{1}{\delta_{NT}}\right).$$

Proof of Theorem 2. It is enough to show that

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

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

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

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

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

For term *III*,

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

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 the Bartlett and Parzen kernels. Thus,

$$III\leq \sup_{\pi\in[\pi_1,\pi_2]}\left\|\hat{\Gamma}_{1,0}(\pi,\hat{F})-\hat{\Gamma}_{1,0}(\pi,FH)\right\|$$
$$+2\sup_{\pi\in[\pi_1,\pi_2]}\sum_{j=1}^{S_{\lfloor\pi T\rfloor}}\left\|\hat{\Gamma}_{1,j}(\pi,\hat{F})-\hat{\Gamma}_{1,j}(\pi,FH)\right\|. \tag{A.4}$$

For the Bartlett kernel, the RHS of (A.4) 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 III is $o_p(1)$ if $\frac{T^{\frac{2}{3}}}{N}\to 0$ as N, $T\to\infty$. For the Parzen kernel, the RHS of (A.4) 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 III is $o_p(1)$ if $\frac{T^{\frac{2}{5}}}{N}\to 0$ as $N,\ T\to\infty$. For the QS kernel, term III is bounded by

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

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

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

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

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

$$\sup_{\pi \in [\pi_1, \pi_2]} \left\{ S^2_{\lfloor \pi T \rfloor} \sum_{j=S_{\lfloor \pi T \rfloor}+1}^{\lfloor \pi T \rfloor - 1} \frac{25M}{6\pi^2 j^2} \left\| \hat{\Gamma}_{1,j}(\pi, \hat{F}) - \hat{\Gamma}_{1,j}(\pi, FH) \right\| \right\}$$

$$\leq \left(\frac{25M\left(K_2 T^{\frac{1}{5}}\right)^2}{6\pi^2} \sum_{j=\lfloor K_1 T^{\frac{1}{5}} \rfloor}^{T} \frac{1}{j^2} \right) \sup_{\pi \in [\pi_1, \pi_2]} \left\| \hat{\Gamma}_{1,j}(\pi, \hat{F}) - \hat{\Gamma}_{1,j}(\pi, FH) \right\|$$

$$\leq \left(\frac{25M\left(K_2 T^{\frac{1}{5}}\right)^2}{6\pi^2} \int_{\lfloor K_1 T^{\frac{1}{5}} \rfloor - 1}^{T} \frac{1}{j^2} \right) \sup_{\pi \in [\pi_1, \pi_2]} \left\| \hat{\Gamma}_{1,j}(\pi, \hat{F}) - \hat{\Gamma}_{1,j}(\pi, FH) \right\|$$

$$\leq \frac{25M\left(K_2 T^{\frac{1}{5}}\right)^2}{6\pi^2} \left(\frac{1}{\lfloor K_1 T^{\frac{1}{5}} \rfloor - 1} - \frac{1}{T} \right) O_p\left(\frac{1}{\delta_{NT}}\right)$$

$$= O\left(T^{\frac{2}{5}}\right) O\left(T^{-\frac{1}{5}}\right) O_p\left(\frac{1}{\delta_{NT}}\right) = O_p\left(\frac{T^{\frac{1}{5}}}{\delta_{NT}}\right).$$

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

To prove Theorem 3, we present Lemma 9.

LEMMA 9. *Under Assumptions 1–9,*

(i) *if* $\sqrt{T}/N \to 0$, *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) $\sup_{\pi \in [\pi_1, \pi_2]} \|\hat{\mathbb{S}}(\pi, FH_0)^{-1}\| = O_p(1)$ *and* $\sup_{\pi \in [\pi_1, \pi_2]} \|\hat{\mathbb{S}}(\pi, \hat{F})^{-1}\| = O_p(1)$.

(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 Theorem 3. For part (i),

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

$$\leq \sup_{\pi \in [\pi_1, \pi_2]} \left| A(\pi, \hat{F})' \left[\hat{\mathbb{S}}(\pi, \hat{F})^{-1} - \hat{\mathbb{S}}(\pi, FH_0)^{-1} \right] A(\pi, \hat{F}) \right|$$

$$+ \sup_{\pi \in [\pi_1, \pi_2]} \left| \left[A(\pi, \hat{F})' - A(\pi, FH_0)' \right] \hat{\mathbb{S}}(\pi, FH_0)^{-1} A(\pi, \hat{F}) \right|$$

$$+ \sup_{\pi \in [\pi_1, \pi_2]} \left| A(\pi, FH_0)' \hat{\mathbb{S}}(\pi, FH_0)^{-1} \left[A(\pi, \hat{F}) - A(\pi, FH_0) \right] \right|$$

$$= o_p(1)$$

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

by the 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

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

Thus, $\sup_{\pi\in[\pi_1,\pi_2]} W_T(\pi,\hat{F}) - \sup_{\pi\in[\pi_1,\pi_2]} W_T(\pi,FH_0) \le \sup_{\pi\in[\pi_1,\pi_2]}\left|W_T(\pi,\hat{F}) - W_T(\pi,FH_0)\right| = o_p(1)$ by the result in part (i). Similarly, one can also show that $\sup_{\pi\in[\pi_1,\pi_2]} W_T(\pi,FH_0) - \sup_{\pi\in[\pi_1,\pi_2]} W_T(\pi,\hat{F}) \le \sup_{\pi\in[\pi_1,\pi_2]}\left|W_T(\pi,\hat{F}) - W_T(\pi,FH_0)\right| = o_p(1)$. Combining these two inequalities gives 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 the mean value theorem, there exists a sequence of $c_\pi \in [0,1]$ such that

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

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 Lemma 10 and then prove Theorem 4. The proof of Proposition 1 is provided in the working paper version (Han and Inoue, 2014).

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

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 Theorem 4. (i) We have

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

Note that

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

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

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

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

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

by Assumption 8(b) and the definitions of J and J_0. Matrix C contains nonzero 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 nonsingular matrix.

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

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

by Assumption 8(b) and the definitions of J and J_0. Matrix C contains nonzero 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 nonsingular 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 semidefinite and diverges at the rate of the bandwidth parameter. Using these results, we can see that $\text{plim}_{T\to\infty}\left(\rho_{min}[\hat{\mathbb{S}}(\pi^*, GJ_0)]\right) > 0$ and $\text{plim}_{T\to\infty}\left(\rho_{min}[\tilde{\mathbb{S}}(\pi^*, GJ_0)]\right) > 0$. This implies that both $\|\hat{\mathbb{S}}(\pi^*, GJ_0)^{-1}\|$ and $\|\tilde{\mathbb{S}}(\pi^*, GJ_0)^{-1}\|$ are $O_p(1)$ because $\|\hat{\mathbb{S}}(\pi^*, GJ_0)^{-1}\| \le M\rho_{max}[\hat{\mathbb{S}}(\pi^*, GJ_0)^{-1}] = M\left(\rho_{min}[\hat{\mathbb{S}}(\pi^*, GJ_0)]\right)^{-1} = O_p(1)$ for some positive constant $M < \infty$ (see Hall, 2000, eqn. (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 $\text{plim}_{T\to\infty}\left(\rho_{min}[\hat{\mathbb{S}}(\pi^*, \hat{F})]\right) > 0$ and $\text{plim}_{T\to\infty}\left(\rho_{min}[\tilde{\mathbb{S}}(\pi^*, \hat{F})]\right) > 0$, 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

$$\begin{aligned}&\|\hat{\mathbb{S}}(\pi^*, GJ_0)^{-1} - \hat{\mathbb{S}}(\pi^*, \hat{F})^{-1}\| \\ &\quad\le \|\hat{\mathbb{S}}(\pi^*, GJ_0)^{-1}\|\|\hat{\mathbb{S}}(\pi^*, \hat{F}) - \hat{\mathbb{S}}(\pi^*, GJ_0)\|\|\hat{\mathbb{S}}(\pi^*, \hat{F})^{-1}\| \\ &\quad= O_p(1)o_p(1)O_p(1). \end{aligned} \tag{B.1}$$

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,

$$\begin{aligned}W_T(\pi^*, \hat{F}) = &\frac{T}{\max(S_{\lfloor \pi^* T\rfloor}, S_{T-\lfloor \pi^* T\rfloor})}\left[\frac{1}{\sqrt{T}}A(\pi^*, \hat{F})'\right] \\ &\times\left[\max(S_{\lfloor \pi^* T\rfloor}, S_{T-\lfloor \pi^* T\rfloor})\hat{\mathbb{S}}(\pi^*, \hat{F})^{-1}\right]\left[\frac{1}{\sqrt{T}}A(\pi^*, \hat{F})'\right]\end{aligned}$$

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press

$$
\begin{aligned}
&= \frac{T}{\max(S_{\lfloor \pi^* T \rfloor}, S_{T-\lfloor \pi^* T \rfloor})} \left[\operatorname{vech}(C)' + o_p(1)\right] \\
&\quad \times \left\{ \max(S_{\lfloor \pi^* T \rfloor}, S_{T-\lfloor \pi^* T \rfloor}) \left[\hat{\mathbb{S}}(\pi^*, GJ_0)^{-1} + o_p(1)\right] \right\} \left[\operatorname{vech}(C) + o_p(1)\right] \\
&\to \infty
\end{aligned}
$$

by Assumption 11, part (i) of Theorem 4, and (B.1). Also, it can be proved that $LM_T(\pi^*, \hat{F}) \to \infty$ 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 $LM_T(\pi^*, \hat{F})$. ■

https://doi.org/10.1017/S0266466614000486 Published online by Cambridge University Press