

Graduate School of Economics, Hitotsubashi University  
Discussion Paper Series No. 2013-17

**Testing for Factor Loading Structural Change under  
Common Breaks**

Yohei Yamamoto      Shinya Tanaka

December 2013

# Testing for Factor Loading Structural Change under Common Breaks\*

Yohei Yamamoto<sup>†</sup>

Shinya Tanaka<sup>‡</sup>

Hitotsubashi University

JSPS Research Fellow

December 13, 2013

## Abstract

This paper proposes a new test for factor loading structural change in dynamic factor models. The proposed test is robust to the nonmonotonic power problem that occurs if the factor loadings exhibit structural changes at common dates over cross-sections. To illustrate the usefulness of our test, we first show that the leading test proposed by Breitung and Eickmeier (2011) exhibits nonmonotonic power, essentially because the breaks are considered as spurious factors with stable factor loadings. We use both local and non-local asymptotic frameworks to investigate the power of their test. The new test eliminates the effects of the spurious factors by maximizing the test statistic over possible numbers of the original factors. This approach is effective because the original factors are not identified under the alternative hypothesis. Monte Carlo simulations and an empirical example using U.S. Treasury yield curve data clearly illustrate the validity of the asymptotic power analysis and usefulness of the proposed test.

**JEL Classification Number:** C12, C38

**Keywords:** factor model, principal components, common breaks, spurious factors, local alternative asymptotics, fixed alternative asymptotics, nonmonotonic power, yield curve

---

\*This is a revised version of the paper previously circulated under the title “Power Analysis for Factor Loading Structural Change Tests under Common Breaks.” Yamamoto acknowledges financial support for this work from the MEXT Grants-in-Aid for Scientific Research Nos. 25870235 and 25285067. Tanaka acknowledges financial support from Grant-in-Aid for JSPS Fellows 259598. We appreciate Eiji Kurozumi, Pierre Perron, and Zhongjun Qu for useful suggestions.

<sup>†</sup>Hitotsubashi University, Department of Economics, 2-1 Naka, Kunitachi, Tokyo, Japan 186-8601 (yohei.yamamoto@econ.hit-u.ac.jp).

<sup>‡</sup>Takasaki City University of Economics, Department of Economics, 1300 Kaminamie-machi, Takasaki, Gunma, Japan 370-0801 (stanaka@wd5.so-net.ne.jp).

## 1 Introduction

One of the most important problems in classical hypothesis testing is the fact that a certain test may have power when the model involves parameter values that are close to the null hypothesis, but loses power as they get distant from the null region. This is called the nonmonotonic power problem and is found to be widespread in structural change tests.<sup>1</sup> As far as the authors know, Perron (1991) is the first paper that points out this problem in structural change tests. Later, Vogelsang (1999) documents that various structural change tests applied to a model with lagged dependent variables lose power as the magnitude of the break increases. Crainiceanu and Vogelsang (2007) point out that the same problem occurs even in a model without lagged dependent variables, but the variances are estimated using a standard nonparametric heteroskedasticity and autocorrelation consistent (HAC) estimator. Subsequent literature carries out asymptotic analyses of the power problem for the case of CUSUM and CUSUM squared tests (Deng and Perron, 2006), the tests proposed by Andrews (1993) and Andrews and Ploberger (1994) (Kim and Perron, 2009), and Elliott and Müller's (2006)  $\widehat{qLL}$  test (Perron and Yamamoto, 2013). In these cases, one of the important sources of nonmonotonic power is found to be the problem of identifying the persistence parameter and structural change in the conditional mean of the model. As documented by Perron (1990), unaccounted-for structural change in the conditional mean of the model generates a bias of the persistence parameter estimate toward one. This occurs either in the model parameter or the variance estimates of the tests, and results in the structural change tests becoming uninformative. In other words, the persistence parameter is not correctly identified under the alternative hypothesis of structural change.<sup>2</sup>

This paper casts new light on this nonmonotonic power problem in considering the factor loading structural change tests recently proposed by Breitung and Eickmeier (2011) (hereafter, BE tests) in the context of dynamic factor models.<sup>3</sup> Our focus is on the case in which the factor loadings exhibit structural changes at common dates over cross-sections. The

---

<sup>1</sup>See Perron (2006) for a comprehensive review. In the context outside of structural change tests, Nelson and Savin (1990) uncover the case of the Wald test in a nonlinear logit model. Hall (2000) documents the case of the overidentification restrictions test.

<sup>2</sup>Incorrectly identified persistence parameter is not sufficient for the nonmonotonic power problem to occur. Indeed, Perron and Yamamoto (2013) show that the SupWald test does not suffer from this problem when the coefficients have multiple breaks or when the coefficients follow random walks, even if the persistence parameter is not correctly identified in these cases.

<sup>3</sup>Recently, two other factor loading structural change tests have been proposed by Chen et al. (2012) and Han and Inoue (2013). However, the scope of both tests is different from ours, since they need to test factor loadings of many response variables jointly and cannot be applied to factor loadings of a particular variable.

BE tests are based on the standard tests of Andrews (1993) and Andrews and Ploberger (1994) applied to the regression of one response variable on the principal components factors. Furthermore, given that the principal components consistently estimate the original factors (up to rotation), the tests have standard limit distributions. A crucial assumption is that the model must have no or very small structural changes in the factor loadings so that the original factors are consistently estimated prior to testing. Therefore, as long as the common breaks in the factor loadings are small, the original factors are indeed identified and consistently estimated, resulting in the tests having power. However, when the breaks are larger, these breaks generate extra common factors (we call them spurious factors) that can have factor loadings with no structural change. If so, the tests applied to the model with these loadings only have trivial power. Thus, the source of power problem is identification problem between common factors and factor loadings.

This motivates two goals of this paper. First, we carry out a power analysis of the BE tests by investigating their asymptotic properties under the alternative hypothesis of common breaks. Following the literature, we use local and non-local asymptotic frameworks. The local framework assumes that the magnitudes of structural changes shrink at a fast rate as the sample size increases,<sup>4</sup> whereas the non-local framework assumes that magnitudes of breaks are fixed. We find that in the former approach, the BE tests can be increasing functions of the break magnitudes and the asymptotic result approximates finite sample behavior of the tests under small breaks. On the other hand, in the latter approach, the test statistics are neither consistent nor increasing in break magnitudes. This result is valid when the breaks are moderate or large. The novel thing in this paper is that nonmonotonic power occurs even if the model includes no lagged dependent variables or the serial correlation in the errors is unaccounted for.

The second goal of this paper is to propose a new test for factor loading structural change, which is robust to this nonmonotonic power problem. The idea of the test is to eliminate the effects of the spurious factors that appear only under the alternative hypothesis. This is carried out by maximizing the test statistic over possible numbers of factors. This approach is effective because the original factors are not identified under the alternative hypothesis of common breaks. The proposed test is shown to have a simple asymptotic distribution under the null hypothesis and is consistent under the alternative hypothesis. These facts are confirmed via Monte Carlo simulations under various settings.

Finally, we present an empirical example using U.S. Treasury yield curve data by changing

---

<sup>4</sup>In the same context, this framework is used by Bates et al. (2013).

the sample period of 10-year rolling windows starting from November 1985. The new test clearly finds structural changes in factor loadings of most maturities when the sample includes either of the recessions in 1990-91, 2001, and 2008-09. In other periods, the factor structure is stable overall. However, if we use the SupWald version of the BE test, the test gives fewer rejections that are very similar throughout different sample periods. If the SupLM version of the BE test is used, the test never shows a rejection and is completely uninformative. This result highlights the validity of the asymptotic power analysis and usefulness of the new test.

This paper is structured as follows. In Section 2, we present the models and assumptions. In Section 3, we introduce two asymptotic frameworks and carry out asymptotic power analysis of the BE tests. In Section 4, we propose a modified test that is robust to the nonmonotonic power problem. In Section 5, we use Monte Carlo simulations to investigate the finite sample validity of the power analysis and finite sample properties of the new tests. In Section 6, we provide an empirical example based on the U.S. Treasury yield curve data. In Section 7, we present our conclusions. Throughout the paper, the following notations are used. The Euclidean norm of vector  $x$  is denoted by  $\|x\|$ . For matrices, the vector-induced norm is used. The symbols  $O(\cdot)$  and  $o(\cdot)$  denote the standard asymptotic orders of sequences. The symbol  $\xrightarrow{P}$  represents convergence in probability under the probability measure  $P$  and the symbol  $\Rightarrow$  denotes convergence in distribution.  $O_p(\cdot)$  and  $o_p(\cdot)$  are the orders of convergence in probability under  $P$ .

## 2 Models and test statistics

Consider a factor model

$$X = F\Lambda' + e, \quad (1)$$

where  $X$  is a  $T \times N$  matrix of response variables,  $F$  is a  $T \times r$  matrix of unobservable common factors,  $\Lambda$  is an  $N \times r$  matrix of factor loadings, and  $e$  is a  $T \times N$  matrix of the idiosyncratic errors. Denote the  $i$ th row of  $\Lambda$  by  $\lambda'_i$  and the  $t$ th row of  $X$ ,  $F$  and  $e$  by  $X'_t$ ,  $F'_t$ , and  $e'_t$ , respectively. Further, let  $x_{it}$  and  $e_{it}$  be the  $(i, t)$  th elements of  $X'$  and  $e'$ , respectively. The following assumptions hold for this model:

### Assumption 1:

- (a)  $E\|F_t\|^4 \leq M < \infty$  and  $F'F/T \xrightarrow{P} \Sigma_F$ , a positive definite matrix.
- (b)  $E\|\lambda_i\|^2 \leq \lambda < \infty$  and  $\Lambda'\Lambda/N \xrightarrow{P} \Sigma_\Lambda$ , a positive definite matrix.
- (c)  $E(e_{it}) = 0$  and  $\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T e_{it} \xrightarrow{P} \sigma_e^2$ , a positive constant.
- (d)  $E\text{trace} \left[ (ee')^j \right] = O(NT \times \max\{N, T\}^{j-1})$  for  $j = 1, 2$ , and  $3$ .

- (e)  $E \sum_{t=1}^T \sum_{s=1}^T \left( \sum_{i=1}^N \lambda'_i F_t e_{is} \right)^2 = O(NT^2)$ .  
(f)  $E \sum_{t=1}^T \sum_{i=1}^N \lambda'_i \lambda_i e_{it}^2 = O(NT)$ .  
(g)  $E \sum_{i=1}^N \left\| \sum_{t=1}^T F_t e_{it} \right\|^2 = O(NT)$ .

**Assumption 2:**  $F_t$ ,  $\lambda_i$ , and  $e_{is}$  are independent for all  $i, t$ , and  $s$ .

Assumption 1 is equivalent to the set of assumptions discussed by Bai and Ng (2002) and adopted by Amengual and Watson (2006). It enables us to consistently estimate the number of factors  $r$  by using the standard information criteria suggested by Bai and Ng (2002). It also corresponds to the assumptions of Bai (2003) that enable Gaussian asymptotic inference on the common factor space and the associated factor loadings by using the principal components estimators. Parts (a), (b), and (c) are standard assumptions. Part (d) is a key assumption to consistently estimate the number of common factors by imposing a restriction on the cross-sectional correlations in the idiosyncratic error terms. Parts (e), (f), and (g) restrict dependency among the components  $\lambda_i$ ,  $F_t$ , and  $e_{it}$ , respectively. Assumption 2 imposes a more stringent condition than Assumption 1 (e), (f), and (g). However, it greatly simplifies the derivation of our results without losing much substance. Note that Assumption 2 is also employed in Breitung and Eickmeier (2011), who derived the limiting distributions of structural change tests that this paper focuses on.

We are interested in assessing if the factor loadings of the  $i$ th response variable  $\lambda_i$  have structural changes.<sup>5</sup> To this end, we take a leading example of the tests proposed by Breitung and Eickmeier (2011). Suppose that there are two regimes and we denote the loadings in each regime by  $\lambda_i^{(1)}$  and  $\lambda_i^{(2)}$ . Then, the null hypothesis is  $\lambda_i^{(1)} = \lambda_i^{(2)}$  and the alternative hypothesis is  $\lambda_i^{(1)} \neq \lambda_i^{(2)}$ . If the change date is unknown, the tests are simply constructed as the SupWald, SupLM, or SupLR test for the coefficients in the time-series regression of  $X_i$  the  $i$ th column of  $X$  on the principal components factor estimates  $\hat{F}$  suggested by literature so that

$$BE_{NT,i} = \sup_{\epsilon \leq \tau \leq 1-\epsilon} S_{NT,i}(\tau),$$

with  $S_{NT,i}(\tau) = T(SSR_{r,i} - SSR(\tau)_i)/SSR(\tau)_i$  for the Wald test,  $S_{NT,i}(\tau) = T(SSR_{r,i} - SSR(\tau)_i)/SSR_{r,i}$  for the LM test, and  $S_{NT,i}(\tau) = T(\log(SSR_{r,i}) - \log(SSR(\tau)_i))$  for the LR

<sup>5</sup>Following Bai and Ng (2007), the model (1) is also interpreted as a dynamic factor model.

$$x_{it} = \lambda'_{i0} f_t + \lambda'_{i1} f_{t-1} + \cdots + \lambda'_{iq} f_{t-q} + e_{it},$$

by defining  $\lambda_i = [\lambda'_{i0} \lambda'_{i1} \cdots \lambda'_{iq}]'$  and  $F_t = [f'_t f'_{t-1} \cdots f'_{t-q}]'$  so that we can simply work out structural changes in  $\lambda_i$ .

test.  $[\epsilon, 1 - \epsilon]$  specifies the permissible break fraction with a trimming parameter  $\epsilon$ . The sum of squared residuals are

$$\begin{aligned} SSR_{r,i} &= \sum_{t=1}^T (X_{it} - \hat{\lambda}'_i \hat{F}_t)^2, \\ SSR(\tau)_i &= \sum_{t=1}^{[T\tau]} (X_{it} - \hat{\lambda}_i^{(1)'} \hat{F}_t)^2 + \sum_{t=[T\tau]+1}^T (X_{it} - \hat{\lambda}_i^{(2)'} \hat{F}_t)^2, \end{aligned}$$

with  $\hat{\lambda}_i = \left( \sum_{t=1}^T \hat{F}_t \hat{F}_t' \right)^{-1} \left( \sum_{t=1}^T \hat{F}_t X_{it} \right)$ ,  $\hat{\lambda}_i^{(1)} = \left( \sum_{t=1}^{[T\tau]} \hat{F}_t \hat{F}_t' \right)^{-1} \left( \sum_{t=1}^{[T\tau]} \hat{F}_t X_{it} \right)$  and  $\hat{\lambda}_i^{(2)} = \left( \sum_{t=[T\tau]+1}^T \hat{F}_t \hat{F}_t' \right)^{-1} \left( \sum_{t=[T\tau]+1}^T \hat{F}_t X_{it} \right)$ . The tests are modified accordingly when heteroskedasticity and serial correlations are suspected in the errors by using standard heteroskedasticity and autocorrelation covariance matrix.<sup>6</sup>

When the tests are constructed, the unknown number of factors must be estimated by, for example, the standard information criteria of Bai and Ng (2002) such that  $\hat{r} = \arg \min_{0 \leq q \leq \bar{r}} IC(q)$  where

$$IC(q) = \ln \left( V(q, \hat{F}^q) \right) + q \cdot g(N, T),$$

$V(q, \hat{F}^q) = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \hat{e}_{it}^2$  and  $\hat{e}_{it}$  is the idiosyncratic residual from the regression of  $X_{it}$  on the first  $q$  principal component factors. The penalty satisfies  $g(N, T) \rightarrow 0$  and  $c_{NT}^2 \cdot g(N, T) \rightarrow \infty$  where  $c_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$ .<sup>7</sup>  $\bar{r}$  is the prespecified maximum number of factors.

Specifically, we consider a model with  $m$  time structural changes in the factor loadings of all or many response variables at common dates  $\{T_1, \dots, T_m\}$ . Let  $\Lambda^{(j)}$  be an  $N \times r$  matrix of factor loadings at regime  $j$  ( $j = 1, \dots, m+1$ ) so that the model is now

$$X_t = \Lambda^{(j)} F_t + e_t, \tag{2}$$

where

$$\Lambda^{(j)} = \begin{cases} \Lambda^{(1)} & \text{for } t \leq T_1 \\ \Lambda^{(2)} & \text{for } T_1 + 1 \leq t \leq T_2 \\ \vdots & \\ \Lambda^{(m+1)} & \text{for } T_m + 1 \leq t \leq T. \end{cases} \tag{3}$$

<sup>6</sup>There are two remarks. First, if the change date is known, one naturally uses the Chow test. Second, these tests must apparently have power against multiple breaks, although explicitly accounting for them; say, employing Bai and Perron's (1998) double maximum tests must result in power gain.

<sup>7</sup>Bai and Ng (2002) suggest three types of  $g(N, T)$ :  $q \left( \frac{N+T}{NT} \right) \ln \left( \frac{NT}{N+T} \right)$ ,  $q \left( \frac{N+T}{NT} \right) \ln(c_{NT}^2)$ , and  $q \left( \frac{\ln c_{NT}^2}{c_{NT}^2} \right)$ . They define the information criteria with these penalties as  $IC_{p1}$ ,  $IC_{p2}$ , and  $IC_{p3}$ , respectively.

Then, the magnitudes of the  $j$ th break are  $\Delta^{(j)} = \Lambda^{(j+1)} - \Lambda^{(j)}$ , where  $\Delta^{(j)}$  denotes either fixed or random quantities. For later use, we denote the  $L^2$ -norm of the matrix  $\|\Delta^{(j)}\| = \delta^{(j)}$  and  $\delta = \max_j \delta^{(j)}$ . This model includes the case where only a subset of response variables has the stated common breaks. In this case, the rows of  $\Delta^{(j)}$  corresponding to the stable loadings are simply zeros. It is not hard to find empirical examples of such common breaks in factor loadings. For example, Stock and Watson (2009) document evidence of a break in many factor loadings around 1984 in the U.S. macroeconomic time series. Bai (2010) discusses several examples of common breaks in panel data models.

When the factor loadings are modeled as (3), model (2) can be rewritten as

$$X = F\Lambda^{(1)'} + S\Gamma' + e, \quad (4)$$

where  $S$  is a  $T \times mr$  matrix of the so-called spurious factors

$$S = \begin{bmatrix} 0 & 0 & 0 \\ F_{T_1+1:T_2} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & F_{T_{m+1}:T} \end{bmatrix},$$

with  $F_{T_j+1:T_{j+1}} = [F_{T_j+1} \ F_{T_j+2} \ \cdots \ F_{T_{j+1}}]'$ , and  $\Gamma$  is an  $N \times mr$  matrix

$$\Gamma = \begin{bmatrix} \Gamma^{(2)} & \cdots & \Gamma^{(m+1)} \end{bmatrix},$$

where  $\Gamma^{(j)} = \Lambda^{(j)} - \Lambda^{(1)}$  and let  $\Gamma^{(1)} = \Lambda^{(1)}$ . Further, we let  $\gamma'_i$  be the  $i$ th row of  $\Gamma$  with dimension  $mr$  and  $\gamma_i^{(j)'} be the  $i$ th row of  $\Gamma^{(j)}$  with dimension  $r$ . We call (4) spurious factor representations of the original factor model (2) with common breaks in the factor loadings (3). Under this representation, Assumptions 1 (a) and (b) are extended as follows.$

**Assumption 1'**: If the  $m$  common breaks exist, with  $F^* = [F \ S]$  and  $\Lambda^* = [\Lambda_1 \ \Gamma]$ ,

(a)  $F^{*'}F^*/T \xrightarrow{p} \Sigma_F^*$ , a positive definite matrix.

(b)  $E \left\| \gamma_i^{(j)'} \right\|^2 \leq \bar{\gamma} < \infty$  and  $\Lambda^{*'}\Lambda^*/N \xrightarrow{p} \Sigma_\Lambda^*$ , a positive definite matrix.

Note that, under the null hypothesis of no breaks, models (1) and (2) coincide by notating  $\Lambda^{(1)} = \Lambda$ . Hence, Assumptions 1' (a) and (b) nest Assumptions 1 (a) and (b) under the null hypothesis. Thus, we replace them and rename Assumptions 1 (a) and (b) in the following analysis.

### 3 Asymptotic power analysis

The goal of this section is to obtain theoretical results that can reasonably approximate the exact behavior of the BE test statistics. To this end, we derive asymptotic results based on two different frameworks. The first method is a local alternative asymptotic framework and is popular in literature. In a related work, Bates et al. (2013) derive conditions in which the factors are consistently estimated using local asymptotic frameworks under various factor loading processes. The second approach is non-local and designs the break magnitudes to be fixed as the sample size increases. This approach is called the fixed alternative asymptotic framework.<sup>8</sup> The following assumptions elaborate these methodologies.

**Assumption 3a: (Local alternative asymptotic framework)** The break magnitudes shrink so that  $\Delta_{NT}^{(j)} = \Delta^{(j)}/k_{NT}$ , where  $k_{NT} = \max\{\sqrt{N}, \sqrt{T}\}$ .

**Assumption 3b: (Fixed alternative asymptotic framework)** The break magnitudes  $\Delta^{(j)}$  comprise an  $N \times r$  matrix of constants or a random matrix such that  $\|\Delta^{(j)}\| = O_p(1)$ .

Assumption 3a introduces a scaling factor  $k_{NT}$  as an increasing function of  $N$  and/or  $T$ . The specified rate  $k_{NT} = \max\{\sqrt{N}, \sqrt{T}\}$  is actually derived in Proposition 1 so that the original factors are identified. As we will see below, this device enables us to derive the standard power function that is increasing in the break magnitudes. On the other hand, Assumption 3b simply assumes that the break magnitudes are fixed as the sample size grows.

Because the factor number identification substantially affects the testing results, we take the standard approach by using information criteria suggested by Bai and Ng (2002).<sup>9</sup> The following proposition is obtained.

**Proposition 1** *Let  $\hat{r}$  be the number of factors estimated by Bai and Ng's (2002) method and Assumptions 1-3 hold. (i) Under the local alternative asymptotic framework with  $k_{N,T} = \max\{\sqrt{N}, \sqrt{T}\}$ ,  $\hat{r} \xrightarrow{P} r$  as  $N, T \rightarrow \infty$ ; and (ii) under the fixed alternative asymptotic framework,  $\hat{r} \xrightarrow{P} r(m+1)$  as  $N, T \rightarrow \infty$ .*

---

<sup>8</sup>See Perron and Yamamoto (2013), for example.

<sup>9</sup>Recently, other methods to estimate the number of factors have been proposed, for example, Onatski (2010) and Ahn and Horenstein (2013). Extending our theoretical results to these method is beyond the scope of this paper, however, we conducted a simulation analysis and discuss it in the Monte Carlo section.

The first part pertains to the fact that the number of original factors  $r$  can still be identified in the presence of common breaks that shrink. In this case, time variations of the factor loadings are so small that they are considered to be a part of the idiosyncratic errors, rather than variations in the common part. The second part suggests that when the common breaks are medium or large, the methods identify the original factors as well as all the spurious factors through the principal components.

We move on to the main asymptotic properties of the test statistics when they are constructed based on  $\hat{r}$  factors. The following theorem provides the results.

**Theorem 1** *Let  $BE_{NT,i}$  be one of the SupLR, SupLM, and SupWald structural change tests proposed by Breitung and Eickmeier (2011), assuming i.i.d. errors for the null hypothesis of  $\lambda_i$  being stable. Suppose that the number of factors is estimated by Bai and Ng's (2002) method and Assumptions 1-3 hold. i) Under the local alternative asymptotic framework with  $k_{NT} = \max\{\sqrt{N}, \sqrt{T}\}$ ,*

$$BE_{NT,i} = \begin{cases} O_p(\delta^2) & \text{if } \liminf N/T = 0 \\ O_p(1) & \text{if } \liminf T/N = 0 \end{cases},$$

*as  $N, T \rightarrow \infty$ . ii) Under the fixed alternative asymptotic framework,*

$$BE_{NT,i} = \begin{cases} O_p(1) & \text{if } \liminf \sqrt{T}/N = 0 \\ O_p(TN^{-2}) & \text{if } \liminf \sqrt{T}/N > 0 \end{cases},$$

*as  $N, T \rightarrow \infty$ .*

This theorem has useful implications for the power of the BE tests. First, part i) says that under the local approach, the tests are increasing in the break magnitude, especially when  $T$  is large as compared to  $N$ . This result captures the standard upward behavior of the power function as long as the breaks are small. On the other hand, when  $N$  grows faster, the test statistics are not increasing in magnitude, and therefore, upward behavior is not expected. Second, part ii) says that if the breaks are not very small, the statistics are not increasing functions of the magnitudes. Furthermore, in the usual situation where  $N$  is not so small as compared to  $T$  so that  $\sqrt{T}/N \rightarrow 0$  is relevant, the test statistics are simply  $O_p(1)$  and are inconsistent. When  $N$  is very small such that  $\sqrt{T}/N \rightarrow 0$  would be irrelevant, the tests are consistent but only at the rate  $TN^{-2}$ .

Based on the above investigation, we also provide the following theorem to reveal another aspect of the test statistics. This property can help understand testing results that are often too sensitive to the choice of the number of factors used in empirical analysis.

**Theorem 2** *Let  $BE_{NT,i}$  be one of the SupLR, SupLM, and SupWald structural change tests proposed by Breitung and Eickmeier (2011), assuming i.i.d. errors for the null hypothesis of  $\lambda_i$  being stable. Suppose that the number of factors is assumed to be  $r^*$ , where  $r \leq r^* < r(m+1)$  and Assumptions 1-3 hold. Under the fixed alternative asymptotic framework,*

$$BE_{NT,i} = O_p(T),$$

*as  $N, T \rightarrow \infty$ . In the special case of  $m = 1$  and the SupWald test being used,  $BE_{NT,i} = O_p(T\delta^2)$ .*

A proof of this theorem is provided in the Appendix and it goes as follows. Suppose that the factor model has common breaks such that the model can also be written as the spurious factor representation with no break in the factor loadings. However, if we select only a subset of the total factors, then the representation is incomplete and the components pertaining to the non-selected factors will generate structural changes in the factor loadings of the selected factors. This leads to the structural change tests that have standard power properties.<sup>10</sup>

#### 4 A modified test

This section proposes a modified Breitung and Eickmeier (2011) test (hereafter, MBE test) for factor loading structural change of one response variable. Our test is robust to the non-monotonic power problem under the common breaks described in the previous section. We have found that the source of power loss is the failure to identify factors and factor loadings, that is, the fact that original factors are not identified under the alternative hypothesis. The modified test takes the maximum of the SupWald structural change test statistics that are applied to the regressions of the response variable on each individual principal component estimate up to the  $\hat{r}$ th one. It is formed as follows:

$$MBE_{NT,i} = \max_{1 \leq j \leq \hat{r}} \sup_{\mu \in \Lambda_\varepsilon} W_{NT,i}(\mu; j),$$

where  $\Lambda_\varepsilon$  specifies the permissible break dates and  $W_{NT,i}(\mu; j)$  is the Wald test statistic for the null hypothesis of  $H_0 : \lambda_{ji}^{(1)} = \lambda_{ji}^{(2)}$  in the regression of  $x_{it}$  on the  $j$ th principal component  $\hat{f}_{jt}$  estimated using all the variables. That is,

$$x_{it} = \lambda_{ji} \hat{f}_{jt} + error,$$

---

<sup>10</sup>The results are consistent with the results of these test statistics using the standard model derived by Kim and Perron (2009) and Perron and Yamamoto (2013).

where  $\lambda_{ji}$  possibly takes different values before and after  $t = \lfloor \mu T \rfloor$  so that

$$\lambda_{ji} = \begin{cases} \lambda_{ji}^{(1)} & \text{for } t \leq \lfloor \mu T \rfloor \\ \lambda_{ji}^{(2)} & \text{for } t \geq \lfloor \mu T \rfloor + 1 \end{cases}.$$

To ease notation, we redefine the model (1) by using  $FK$  and  $\Lambda K^{-1}$ , where  $K = p \lim(\Lambda' \Lambda / N)(F' \hat{F} / T) V^{-1}$  and  $V$  is an  $r \times r$  diagonal matrix of the  $r$  largest eigenvalues of  $XX' / (NT)$ . We denote  $FK$  by  $F$  and  $\Lambda K^{-1}$  by  $\Lambda$  in this subsection. Let  $f_{jt}$  be the  $j$ th element of  $F_t$  and  $\lambda_{ji}$  be the  $j$ th element of  $\lambda_i$  and rewrite the model (1) in scalar form as follows:

$$\begin{aligned} x_{it} &= \lambda_{ji} f_{jt} + \sum_{l \neq j} \lambda_{li} f_{lt} + e_{it}, \\ &= \lambda_{ji} f_{jt} + v_{jit}, \end{aligned}$$

where  $v_{jit} \equiv \sum_{l \neq j} \lambda_{li} f_{lt} + e_{it}$ . We make the following assumptions:

**Assumption 4:** Under the null hypothesis,  $\lambda_{ji}$  ( $j = 1, \dots, r$ ) are independent and  $E(\lambda_{ji}) = 0$ .

**Assumption 5:** Under the null hypothesis of no structural change in the factor loadings in the original representation, the following hold:

- a)  $T^{-1} \sum_{t=1}^{\lfloor T\mu \rfloor} F_t F_t' \xrightarrow{P} \mu I_r$ , uniformly in  $\mu \in [0, 1]$ , and  $E(F_t F_t') = I_r$ .
- b)  $\frac{1}{T} \sum_{t=1}^{\lfloor T\mu \rfloor} \sum_{s=1}^{\lfloor T\mu \rfloor} E(f_{jt} f_{js} v_{jit} v_{jis}) \xrightarrow{P} \mu \sigma_{vji}^2$ , uniformly in  $\mu \in [0, 1]$ .
- c)  $\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\mu \rfloor} f_{jt} v_{jit} \Rightarrow \sigma_{vji} W(\mu)$  with  $W(\mu)$  being a univariate Wiener process defined on  $\mu \in [0, 1]$ .

Assumption 4 is somewhat more restrictive than the standard assumption. However, the robustness of finite sample size of the test when this assumption is not satisfied is examined via Monte Carlo simulations. The first part of Assumption 5a) is implied by the original representation (1) with rotation adjustment. We also have an additional identification assumption of the constant second moment of factors.<sup>11</sup> Parts b) and c) of Assumption 5 are standard in the literature to derive the limit distribution under the i.i.d. assumption; however, potential serial correlations and conditional heteroskedasticity can be accounted for using the standard heteroskedasticity and autocorrelation consistent variance estimates.

---

<sup>11</sup>Han and Inoue (2013) use the same assumption of constant second moment of factors.

**Theorem 3** Let  $W_j(\cdot)$  ( $j = 1, \dots, r$ ) be the  $j$ th component of  $r$  independent Wiener processes  $W$  defined on  $[0, 1]$ . *i)* Suppose Assumptions 1, 2, 4, and 5 hold. Under the null hypothesis of no breaks

$$MBE_{NT,i} \Rightarrow \max_{1 \leq j \leq r} \sup_{\mu \in \Lambda_c} \frac{[\mu W_j(1) - W_j(\mu)]^2}{\mu(1 - \mu)},$$

*ii)* Suppose Assumptions 1-3 hold. Let  $\lambda_{j_i}$  have the largest break among  $j = 1, \dots, r$  with the magnitude  $\delta^*$  and the rest of the loadings have breaks with magnitudes  $\delta$ . Under the fixed alternative framework,  $MBE_{NT,i} = O_p(T\delta^{*2}\delta^{-2})$ .

Part i) of Theorem 3 provides the asymptotic distribution of the MBE test. The critical values are tabulated using 10,000 replications of 10,000 discrete step Wiener processes in Table 1. Part ii) guarantees the consistency of the test and upward behavior of power functions.

**Remark 1** It is easily understood that the MBE test is consistent under any alternative hypothesis under which the BE test is consistent. This is because  $BE_{NT,i} \leq MBE_{NT,i}$  as long as the maximum number of factors used for  $MBE_{NT,i}$  is equal to (or possibly larger than) the number of factors used for  $BE_{NT,i}$ . Hence, the MBE test is expected to have power under the alternative hypothesis that is not necessarily common breaks.

**Remark 2** We alternatively consider the test that maximizes the SupWald statistics for the regressions of  $x_i$  on the first  $j$  factors over  $j$  up to  $\hat{r}$ . However, various Monte Carlo simulation results find that this test is less powerful than the proposed test. Hence, this method is not considered in this paper.

## 5 Finite sample evidence

In this section, we use Monte Carlo simulations to provide finite sample evidence of the theoretical results derived in the previous sections. Throughout the section, we consider the following  $r$  factor data generating process:

$$x_{it} = \lambda_i^{(j)'} f_t + e_{it}, \quad \text{for } i = 1, \dots, N, t = 1, \dots, T, \text{ and } j = 1, \dots, m, \quad (5)$$

where  $f_t \sim i.i.d.N(0, I_r)$  and  $e_{it} \sim i.i.d.N(0, 1)$ , unless otherwise specified. The factor loadings exhibit  $m$  time common structural changes at  $\{T_1, \dots, T_m\}$  for all  $i$ . Specifically, for the initial value, each component of  $\lambda_i^{(1)}$  follows  $i.i.d.N(0, 1)$  for all  $i$  and the loadings follow

$$\lambda_i^{(j+1)} = \lambda_i^{(j)} + \delta_i^{(j)} \quad \text{for } i = 1, \dots, N \text{ and } j = 1, \dots, m, \quad (6)$$

where the break magnitude  $\delta_i^{(j)}$  is generated by  $i.i.d.N(0, \sigma I_r)$ . The  $m$  break dates are located with equal intervals, that is, at  $T_j = j \times \lfloor T/(m+1) \rfloor$  for  $j = 1, \dots, m$ . The number of replications is 3,000 throughout the experiments.

## 5.1 Number of factors

We are first interested in the effect of the common breaks on the number of factors estimated by Bai and Ng's (2002) information criteria. To this end, we compute the mean of the estimated numbers under cases of  $r = 2, 3$ , and 4. When  $r = 2$ , the number of the common breaks is either  $m = 1, 2, 3$ , and 4, so that the total (true plus spurious) factors are  $r(m+1) = 4, 6, 8$ , and 10, respectively. When  $r = 3$ ,  $m = 1$  and 2, so that  $r(m+1) = 6$  and 9 respectively are considered. When  $r = 4$ , a case of  $m = 1$  such that  $r(m+1) = 8$  is considered.<sup>12</sup> The sample sizes are combinations of  $T = 100, 200, 500$  and  $N = 100, 200, 500$ . The results based on the three information criteria  $IC_p1$ ,  $IC_p2$ , and  $IC_p3$  are reported in Table 2. To see how the local approach works, we set the magnitude of the break  $\sigma = 5/\max\{\sqrt{N}, \sqrt{T}\}$ , the rate derived in Proposition 1. In most cases, the number of factors is close to the original number  $r$  and also gets closer as the sample size increases. Table 3 shows the results when the breaks are relatively large and fixed ( $\sigma = 5$ ) over the sample sizes. In this case, we expect the number estimates to approach  $r(m+1)$ , not  $r$ . Indeed, the results are very close to this expectation in most cases. Thus, finite sample evidence of Proposition 1 is obtained.

## 5.2 The BE test

We move on to finite sample power of the BE structural change tests. To construct the tests, the number of factors is estimated using the  $IC_p2$  criterion as  $\hat{r}$  and the structural change tests are applied to the  $\hat{r}$  loadings of the first response variable.<sup>13</sup> The experiment pertains to the power in relation to the sample size. The data are generated by models (5) and (6) with  $r = 2$  and  $m = 1$ . We consider  $\sigma = 5/\max\{\sqrt{N}, \sqrt{T}\}$  and  $10/\max\{\sqrt{N}, \sqrt{T}\}$  to investigate the local approach and  $\sigma = 5$  and 10 for the fixed approach. The rejection frequencies with various sample sizes are reported in Table 4. The results capture the theoretical results very well. First, under local breaks and when  $T$  is larger than  $N$ , the power tends to go up as the magnitudes increase. This is prominent when  $(T, N) = (500, 20)$  and  $(500, 50)$ , where the rejection frequencies go up by more than 0.2 percentage points. However, when  $N$  is

<sup>12</sup>We set the maximum number of factors at 12 throughout the simulations.

<sup>13</sup>We also computed power functions of the BE tests using the ER and GR statistics of Ahn and Horenstein (2013) and got similar results. The results based on Onatski (2010) are less sensitive to common breaks.

relatively large, the increase of power is smaller. Second, the results for fixed breaks are not sensitive to the break magnitudes and provide very similar power for  $\sigma = 5$  and  $\sigma = 10$ . This result supports the fact that the statistic is  $O_p(1)$  in most cases under the fixed alternative asymptotic framework.

We proceed to the computation of power functions with respect to the break magnitudes. The models are (5) and (6), the number of factors  $r = 2$ , and cases of  $m = 1, 2, 3$ , and 5 are considered. The power functions of the 5% nominal level in terms of the magnitudes of  $\sigma$  from 0 through 5, are reported in Figure 1. We first set the sample size  $(N, T) = (100, 100)$ . For the SupWald test (left panel), the power functions show similar patterns over the values of  $m$ . The power initially goes up as  $\sigma$  increases. This reflects the local asymptotic results where the tests are  $O_p(\delta)$ . However, the powers quickly reach their peaks and start to decline. After some points of  $\sigma$ , they become flat functions. Thus, the SupWald test exhibits clear evidence of nonmonotonic power. The right panel of Figure 1 shows the power functions of the SupLM test; these are very similar to those of the SupWald test. Both results imply that the SupLR test also provides similar patterns in its power function. What is impressive is that either of the SupWald, SupLM, and SupLR tests exhibit nonmonotonic power functions.

Theorem 1 suggests that the asymptotic behavior of the tests depends on the relative rate between  $N$  and  $T$ . Based on this result, we also compute power functions with different sample sizes. Figure 2 illustrates the power functions under the same setting, but the sample size is now  $T = 50$  and  $N = 200$ . When  $T$  is small relative to  $N$ , Theorem 1 shows that the power under small breaks is not a function of  $\sigma$ ; hence, the functions are more flat than in the case of  $T = 100$  and  $N = 100$ . On the other hand, Figure 3 shows power functions with  $T = 200$  and  $N = 50$ , corresponding to a case with a larger  $T$  relative to  $N$ . In this case, we observe a larger increase of the power for small breaks, because it is more explicitly an increasing function in magnitude. However, the powers quickly peak out and eventually become flat. Hence, our theoretical results on nonmonotonic power are confirmed by finite sample experiments.

Figure 4 illustrates power functions when the number of factors is arbitrarily chosen. The data are generated with  $r = 2$  and  $m = 3$ , so that  $r(m + 1) = 8$ . The sample size is  $T = 100$  and  $N = 100$ . We draw power functions when the number of factors is set at  $r^* = 3, 5, 8$ , and 10. The figure also includes the power when the number is selected by the  $IC_p2$  criterion (labeled IC). Of particular interest is the standard shape of power functions when we set  $r^*$  to be smaller than 8, so that Theorem 2 is valid. It shows that when  $r^*$  is 3 or 5, we see standard power functions; however, the tests lose power when  $r^*$  is equal to or larger than 8

or when the  $IC_p2$  criterion is used.

The experiments so far are all based on tests assuming i.i.d. errors. An interesting question is what happens when the errors are corrected by the standard method of heteroskedasticity and autocorrelation consistent (HAC) estimates. The literature suggests that when LM-based structural change tests are used instead of Wald-based tests, they are a source of nonmonotonic power (see Kim and Perron, 2009). To this end, we compute power functions under the same setting as Figure 1 but with the HAC standard errors suggested by Andrews (1991), using Bartlett kernel with AR1 approximation in Figure 5. The left panel shows that the SupWald test exhibits nonmonotonic power, although the results are quite similar to those obtained in Figure 1. It is remarkable that the SupLM test loses all the power in every case of  $m$  (right panel). Hence, the problem worsens if the SupLM test is used. This is consistent with the findings in the literature.

Finally, we proceed to an extension that is not fully covered by the theory presented in this paper, but is popular in both theoretical and empirical studies, that is, the parameters follow random walks<sup>14</sup>. Of particular interest is that the  $N$  dimensional random walk innovations only have a few ( $q$ ) common sources. We consider the factor loadings  $\lambda_{it}$  to be varying at every time period, instead of  $\lambda_{ji}$  in (6) varying only  $m$  times, so that the  $N \times r$  matrix  $\Lambda_t = (\lambda_{it})$  follows the following process

$$\Lambda_t = \Phi B_t, \tag{7}$$

where  $\Phi$  is an  $N \times q$  matrix and  $B_t$  is a  $q \times r$  independent random walk process where  $vec(B_t) = vec(B_{t-1}) + u_t$ . Let every  $qr$  component in  $u_t$  follows the i.i.d. standard normal distribution. The components in  $\Phi$  are generated from i.i.d.  $U[0, 1]$  and are fixed over time. Other specifications are the same as the model (5). Power functions of the tests with i.i.d. standard errors, various  $q$  values ( $q = 2, 4, 6,$  and  $10$ ), and sample size  $T = 100$  and  $N = 100$  are presented in Figure 6. The results are largely the same as those of the models with common breaks, with the number of breaks  $m$  corresponding to the rank  $q$ . Hence, we show that the power loss of the structural change tests could be prevalent in empirical models that assume the factor loadings to be random walks with a few common sources of innovations.

### 5.3 The modified BE test

This section investigates the finite sample size and power of the MBE test. To see the size, the data is generated by (5) with the factor loadings  $\lambda_i^{(j)} = \lambda_i$  for all  $j$  drawn from

---

<sup>14</sup>The cases Bates et al. (2013) consider include this example.

$i.i.d.N(0, I_r)$  and fixed within a replication. The true number of factors  $r$  considered is 1 through 8; the number is estimated by  $IC_p2$  criteria and used as the upper bound  $\hat{r}$ .<sup>15</sup> The first case generates factors by  $i.i.d.N(0, I_r)$ . We also consider the case of serially correlated factors by the following AR1 processes:

$$\begin{aligned} f_{jt} &= \rho_j f_{jt-1} + v_{jt}, \quad v_{jt} \sim i.i.d.N(0, 1) \\ \rho_j &= 0.4 - (j - 1) \times 0.05 \quad \text{for } j = 1, \dots, r. \end{aligned}$$

Thus, each factor has a different persistency parameter. With the i.i.d. factors, the non-robust standard errors are used to construct the test. With the serially correlated factors, the correlation is accounted for by using the nonparametric HAC variance with Bartlett kernel and bandwidth selected using an AR1 approximation proposed by Andrews (1991). Table 5-1 reports the 5% nominal size with the sample size  $(N, T) = (100, 100)$ ,  $(50, 200)$ , and  $(200, 50)$ . The truncation parameter  $\varepsilon$  specifying the permissible break dates  $\Lambda_\varepsilon = [\varepsilon, 1 - \varepsilon]$  is 0.05, 0.1, and 0.15, respectively. The results show that the non-robust test has very good size properties for various specifications and confirms Theorem 3 (i). When we use the HAC variance, we see some size-distortions. However, this is within the finite sample size distortions of the SupWald structural change test documented in the literature.<sup>16</sup> We next check the robustness of the size properties when  $E(\lambda_{ji}) = 0$  in Assumption 4 is not satisfied, by generating  $\lambda_{ji}$  from i.i.d.  $U[0, 1]$ . Table 5-2 shows that the size of the MBE test is not affected by this change.

We next investigate the power of the MBE test. The data is generated from the models (5) and (6) with  $r = 2$  and the sample size  $(N, T) = (100, 100)$ . For the process of factor loadings, we first consider multiple common structural changes with the number of breaks  $m = 2, 3, 5$ , and 8. Figure 7 reports power functions of the MBE test with non-robust variance estimates in the left panel and power functions of the test with the HAC variance in the right panel. The power with sample sizes  $(N, T) = (50, 200)$  and  $(200, 50)$  are also computed; however, they are on the same lines and are not reported. They clearly show that the MBE test has standard power functions and confirm Theorem 3 (ii) with either the non-robust or HAC robust test. Figure 8 shows power functions of the MBE test when the factor loadings follow reduced rank random walk (7). We see that the power is much better than the BE test that was illustrated in Figure 6. Overall, the Monte Carlo simulations show good power properties of the new test.

<sup>15</sup>The size and power of the MBE test using the methods of Ahn and Horenstein (2013) and Onatski (2010) to estimate numbers are also computed. The results are basically the same.

<sup>16</sup>See Kejriwal (2009) etc.

## 6 Empirical illustration

In this section, we provide an empirical example to illustrate how useful the results obtained in this paper are in practice. We use the U.S. Treasury yield curve data of Gürkaynak, Sack, and Wright (2007) and updated by these authors. The data span from June 1961 through June 2011. Since maturities as long as thirty years are available only after November 1985, data only after that month are used and the maturity years are numbered from 1 through 30 ( $N = 30$ ). To reduce the computational burden, the original daily data are converted into monthly data by picking the daily observation on the last day of each month. Figure 9 illustrates the yields of the selected maturities: 1, 5, 10, and 30 years. We conjecture that there may be common breaks in the factor loadings for a few time-periods, for example, around the recessions in 1990-91, 2001, and 2008-09.

Since the level data are likely to have stochastic trends in the factors, the data are first differenced. It is of interest to see whether the results are affected by a possible inclusion of recessions; so, we use a 10-year rolling window that starts from the sample of November 1985 through October 1995 ( $T = 120$ ) and ends at the sample of July 2001 through June 2011. In every window, the data are demeaned in time and the number of factors is estimated using the  $IC_p2$  criterion. The estimated number is  $\hat{r} = 11$  up to the sample ending at June 1997 and  $\hat{r} = 10$  thereafter. We also obtain similar results when  $IC_p1$  or  $IC_p3$  is used.

We carry out the BE tests of the SupWald and SupLM versions as well as the MBE test, all with the HAC variance estimates, for factor loadings of every maturity in a window. The results regarding each maturity for which the null of no break is rejected are stored. We then move the window by one month and repeat the same exercise. This is continued until the end of the window reaches June 2011. Figure 10 provides detailed information on which maturity is rejected or not rejected in each 10-year sample period. The horizontal axis corresponds to the end of the 10-year window and the vertical axis shows maturity. The points in black, dark gray and light gray show significance at the 1%, 5%, and 10% levels, respectively. If we use the MBE test, structural changes are found in factor loadings of most maturities in the periods that include either the recessions of 2001 or 2008-09,<sup>17</sup> suggesting that there are likely common breaks in the factor structure of the U.S. Treasury yield curve in these periods. Rejections are also seen in the late 1990s. On the other hand, there is a stable period between the two recessions in which most points are in white. The SupWald

---

<sup>17</sup>We set the permissible break dates from  $0.15T$  to  $0.85T$ , so that the rejections will appear 1.5 years after the breaks.

version of the BE test is not very informative in the sense that the results are very similar in any sample period. If the SupLM version of the BE test is used, the test almost never rejects at any window so that the colored points do not appear in the figure. This is consistent with the Monte Carlo study where the SupLM version of the BE test has no power when the HAC variance is used. Hence, these figures clearly illustrate the usefulness of the modified test in investigating the stability of the factor structure of the U.S. Treasury yield curve.

## 7 Conclusion

This paper proposes a new test for factor loading structural change in the context of dynamic factor models. To illustrate the usefulness of the proposed test, we carry out an asymptotic power analysis for the leading test proposed by Breitung and Eickmeier (2011) under common breaks. We use two complementary asymptotic frameworks to approximate the finite sample behavior of the BE test. Both theoretical and simulation results indicate that the tests have nonmonotonic power under common breaks. The new test maximizes the test statistic over possible number of factors and is robust to this nonmonotonic power problem. This approach is effective, because the original factors are not identified under the alternative hypothesis of common breaks. An empirical analysis using the U.S. Treasury yield curve data clearly illustrates the usefulness of the proposed test.

## Appendix : Technical derivations

In the Appendix, we denote the model under the alternative hypothesis (4) as

$$X = F^* \Lambda^{*'} + e, \quad (\text{A.1})$$

where  $F^* = [F \ S]$  is a  $T \times r(m+1)$  matrix and  $\Lambda^* = [\Lambda \ \Gamma]$  is an  $N \times r(m+1)$  matrix. We also notate  $k_{NT}$  by  $k$ .

**Lemma A1.** Let  $s'_t$  be the  $t$ th row of  $S$ . Under Assumption 1, (i)  $E |\gamma'_i s_t|^2 = O(1)$ ; (ii)  $E |\gamma'_i \gamma_i| = O(m)$ ; (iii)  $E |s'_t s_t| = O(1)$ ; for all  $i$  and  $t$ .

**Proof of Lemma A1:** For (i),  $E |\gamma'_i s_t|^2 = E \left| \gamma_i^{(j)'} F_t \right|^2 = O(1)$  since  $\gamma_i^{(j)'} F_t$  includes  $r^2$  terms that are all bounded. This is because  $E \|F_t\|^2 < \infty$  by Assumption 1(a) and  $E \left\| \gamma_i^{(j)} \right\|^2 < \infty$  by Assumption 1'(b). For (ii),  $E |\gamma'_i \gamma_i| \leq \sum_{j=1}^m E \left\| \gamma_i^{(j)} \right\|^2 = O(m)$ , since  $E \left\| \gamma_i^{(j)} \right\|^2 < \infty$  by Assumption 1'(b). For (iii),  $E |s'_t s_t| \leq E |F'_t F_t| = O(1)$ , by Assumption 1(a).

**Proof of Proposition 1:** For part i), we first show that under the local alternative asymptotic framework with  $k = \max\{\sqrt{N}, \sqrt{T}\}$ , the original factor space  $F$  is consistently estimated even in the presence of  $S\Gamma$ . To this end, we define the new idiosyncratic error term  $u_{it} + e_{it}$ , where  $u_{it} = \gamma'_i s_t / k$ . Since  $u_{it}$  is a linear combination of  $F_t$  and is independent of  $e_{it}$  and the latter satisfies Assumption 1, we consider whether or not the new errors  $u_{it}$  satisfy Assumption 1. Parts (a) and (b) are unaffected by this change of the errors from  $e_{it}$  to  $u_{it} + e_{it}$ . For part (c),

$$(NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T u_{it}^2 = k^{-2} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T (\gamma'_i s_t)^2 = O_p(k^{-2}), \quad (\text{A.2})$$

by Lemma A1(i). Hence,  $(NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T u_{it}^2$  vanishes as  $k \rightarrow \infty$ . For part (d),  $E \text{trace} [(uu')^j]$  is a sum of  $N^j T^j$  terms that are all bounded by  $E |\gamma'_i s_t|^{2j}$ . Hence,

$$E \text{trace} [(uu')^j] \leq O(N^j T^j k^{-2j}).$$

It is required that

$$N^j T^j k^{-2j} \leq NT \times \max\{N, T\},$$

for  $j = 1, 2, 3$ .

For part (e),

$$\begin{aligned} E \sum_{t=1}^T \sum_{s=1}^T \left( \sum_{i=1}^N \lambda'_i F_t u_{is} \right)^2 &= k^{-2} E \sum_{t=1}^T \sum_{s=1}^T \left( \sum_{i=1}^N \lambda'_i F_t \gamma'_i s_s \right)^2 \\ &\leq k^{-2} \sum_{t=1}^T \sum_{s=1}^T \left( \sum_{i=1}^N E |\lambda'_i F_t|^2 \right) \left( \sum_{i=1}^N E |\gamma'_i s_s|^2 \right) \\ &= O(N^2 T^2 k^{-2}). \end{aligned}$$

This follows from the Cauchy-Schwarz inequality and Lemma A1(i). In order to satisfy the whole term  $O_p(NT^2)$ , we require  $N \leq k^2$ .

For part (f),

$$E \sum_{t=1}^T \sum_{i=1}^N \lambda'_i \lambda_i u_{is}^2 = k^{-2} E \sum_{i=1}^N \sum_{t=1}^T [(\lambda'_i \lambda_i) (\gamma'_i s_t)^2] = O(NTk^{-2}),$$

since  $E |\lambda'_i \lambda_i| = O(1)$  by Assumption 1(b) and so is  $E |\gamma'_i s_t|^2$  by Lemma A1(i). This is smaller than  $O(NT)$  because  $k \rightarrow \infty$ .

For part (g),

$$\begin{aligned} E \sum_{i=1}^N \left\| \sum_{t=1}^T F_t u_{it} \right\|^2 &= k^{-2} E \sum_{i=1}^N \left\| \sum_{t=1}^T F_t (\gamma'_i s_t) \right\|^2 \\ &\leq k^{-2} \sum_{i=1}^N E \left( \sum_{t=1}^T \|F_t\|^2 \right) E \left( \sum_{t=1}^T \|\gamma'_i s_t\|^2 \right) = O(NT^2 k^{-2}) \end{aligned}$$

by Cauchy-Schwarz inequality, Assumption 1(a) and Lemma A(i). Then, it holds  $O(NT^2 k^{-2}) \leq O(NT)$  if  $k^2 \geq T$ . Finally, we confirm that the condition in Lemma 2 in Amengual and Watson (2006) is shown straightforwardly by (A.2) if  $k^2 \geq \min\{N, T\}$ . Therefore, the set of Assumption 1 holds for the model (1) if the following conditions are satisfied.

$$\begin{aligned} k^{-2} &\leq \max\{N, T\}, \\ NTk^{-4} &\leq \max\{N, T\}, \\ N^2 T^2 k^{-6} &\leq \max\{N, T\}, \\ N &\leq k^2, \\ T &\leq k^2, \\ k^2 &\geq \min\{N, T\}. \end{aligned}$$

The sufficient condition that satisfies the above six is  $k = \max\{\sqrt{N}, \sqrt{T}\}$ .

For part (ii), we show that the set of Assumption 1 holds for the model (A.1). Parts (a) and (b) are directly shown by Assumption 1'. Part (e) holds because  $E(e_{it}) = 0$  and  $s_t$  and  $e_{it}$  are independent by Assumption 2. Part (g) holds because  $\|s_t e_{is}\| = \|F_t e_{is}\|$  so that  $\|F_t^* e_{is}\| = 2\|F_t e_{is}\|$ . Multiplication with a constant does not affect the order of  $\sum_{i=1}^N \left\| \sum_{t=1}^T F_t e_{is} \right\|^2$ .

**Proof of Theorem 1 (i):** We consider the regression model

$$x_{it} = \lambda'_{it} F_t + e_{it}.$$

The residuals from the regression of  $x_{it}$  on  $\hat{F}_t$  assuming constant factor loadings are

$$\hat{e}_{it} = e_{it} + \lambda'_{it} F_t - \hat{\lambda}'_i \hat{F}_t,$$

where  $\hat{\lambda}_i = (\hat{F}' \hat{F})^{-1} (\hat{F}' X_i) = T^{-1} \hat{F}' X_i$  by using the property of principal components

$\hat{F}'\hat{F}/T = I$ . Also,

$$\begin{aligned}\hat{\lambda}_i &= T^{-1} \sum_{t=1}^T \hat{F}_t F_t' \lambda_{it} + T^{-1} \sum_{t=1}^T \hat{F}_t e_{it} \\ &= T^{-1} \sum_{j=1}^m \left( \sum_{t=T_{j-1}+1}^{T_j} \hat{F}_t F_t' \right) \gamma_i^{(j)} + T^{-1} \sum_{t=1}^T \hat{F}_t e_{it} \\ &= k^{-1} T^{-1} \underbrace{\sum_{j=1}^m \left[ \left( \sum_{t=T_{j-1}+1}^{T_j} \hat{F}_t F_t' \right) \sum_{l=1}^j \delta_i^{(l)} \right]}_{=O_p(T\delta)} + \underbrace{T^{-1} \sum_{t=1}^T \hat{F}_t e_{it}}_{=O_p(T^{-1/2})},\end{aligned}$$

where  $\delta_i^{(l)}$  is the  $i$ th row of  $\Delta^{(l)}$ . Note that we denote the magnitude of breaks by  $\delta = \max_l \left\| \delta_i^{(l)} \right\|$ . Therefore, the order of  $\hat{\lambda}_i$  is the maximum of  $O_p(k^{-1}\delta)$  where  $k = \max\{\sqrt{N}, \sqrt{T}\}$  and  $O_p(T^{-1/2})$ . In other words,

$$\hat{\lambda}_i = \begin{cases} O_p(T^{-1/2}\delta) & \text{if } \liminf N/T = 0 \\ O_p(T^{-1/2}) & \text{if } \liminf T/N = 0 \end{cases}. \quad (\text{A.3})$$

Let us now consider the restricted SSR.

$$\begin{aligned}SSR_r &= \sum_{t=1}^T \left( e_{it} + \lambda'_{it} F_t - \hat{\lambda}'_i \hat{F}_t \right)^2, \\ &= \sum_{t=1}^T e_{it}^2 + k^{-2} \underbrace{\sum_{j=1}^m \left[ \sum_{l=1}^j \delta_i^{(l)'} \left( \sum_{t=T_{j-1}+1}^{T_j} F_t F_t' \right) \sum_{l=1}^j \delta_i^{(l)} \right]}_{=O_p(T\delta^2)} \\ &\quad + 2k^{-1} \underbrace{\sum_{j=1}^m \left[ \sum_{l=1}^j \delta_i^{(l)'} \left( \sum_{t=T_{j-1}+1}^{T_j} F_t e_{it} \right) \right]}_{=O_p(\sqrt{T}\delta)} + \underbrace{\hat{\lambda}'_i}_{=(\text{A.3})} \underbrace{\left( \sum_{t=1}^T \hat{F}_t \hat{F}_t' \right)}_{=O_p(T)} \underbrace{\hat{\lambda}_i}_{=(\text{A.3})} \\ &\quad - 2 \underbrace{\hat{\lambda}'_i}_{=(\text{A.3})} \underbrace{\left( \sum_{t=1}^T \hat{F}_t e_{it} \right)}_{=O_p(T^{1/2})} - 2k^{-1} \underbrace{\hat{\lambda}'_i}_{=(\text{A.3})} \underbrace{\sum_{j=1}^m \left[ \left( \sum_{t=T_{j-1}+1}^{T_j} \hat{F}_t F_t' \right) \sum_{l=1}^j \delta_i^{(l)} \right]}_{=O_p(T\delta)} \\ &= I + II + III + IV + V + VI.\end{aligned}$$

Hence,  $I = O_p(T)$ ,  $II = O_p(Tk^{-2}\delta^2)$ , and  $III = O_p(Tk^{-1}\delta)$ . The terms  $IV$ ,  $V$ , and  $VI$  depend on the rate of  $\hat{\lambda}_i$ . If  $\liminf N/T = 0$ , then  $k = \sqrt{T}$  and  $\hat{\lambda}_i = O_p(T^{-1/2}\delta)$ . In this case,  $I = O_p(T)$ ,  $II = O_p(\delta^2)$ ,  $III = O_p(T^{1/2}\delta)$ ,  $IV = O_p(\delta^2)$ ,  $V = O_p(\delta)$ , and  $VI = O_p(\delta^2)$ . Term  $I$  dominates and  $SSR_r$  is shown to be  $O_p(T)$ . Second, if  $\liminf T/N = 0$ , then  $k = \sqrt{N}$  and  $\hat{\lambda}_i = O_p(T^{-1/2})$ . In this case,  $I = O_p(T)$ ,  $II = O_p(TN^{-1}\delta^2)$ ,  $III = O_p(TN^{-1/2}\delta)$ ,  $IV = O_p(1)$ ,  $V = O_p(1)$ , and  $VI = O_p(T^{1/2}N^{-1/2}\delta^2)$ . Term  $I$  dominates and the whole term is  $O_p(T)$ .

We next consider unrestricted residuals assuming one break in  $\lambda_i$  at  $t = \lfloor \mu T \rfloor$ . Let  $\mu$  be a constant with  $0 < \mu < 1$  and  $\lfloor \cdot \rfloor$  be an operator returning the integer part. Further, define  $\hat{F}_t^{(1)} = \hat{F}_t \times I(t \leq \lfloor \mu T \rfloor)$  and  $\hat{F}_t^{(2)} = \hat{F}_t \times I(t > \lfloor \mu T \rfloor)$  for  $t = 1, \dots, T$  where  $I(\cdot)$  is the indicator function. The unrestricted residuals are

$$\hat{e}_{it}(\mu) = e_{it} + \lambda'_{it} F_t - \hat{\lambda}_i^{(1)'} \hat{F}_t^{(1)} - \hat{\lambda}_i^{(2)'} \hat{F}_t^{(2)}.$$

First, we determine the order of  $\hat{\lambda}_i^{(1)}$  and  $\hat{\lambda}_i^{(2)}$ . They coincide with the order of  $T^{-1} \sum_{t=1}^{\lfloor \mu T \rfloor} \hat{F}_t x_{it}$  and  $T^{-1} \sum_{t=\lfloor \mu T \rfloor + 1}^T \hat{F}_t x_{it}$ . Now,

$$T^{-1} \sum_{t=1}^{\lfloor \mu T \rfloor} \hat{F}_t x_{it} = k^{-1} T^{-1} \underbrace{\sum_{j=1}^m \left[ \left( \sum_{t=T_{j-1}+1}^{T_j} \hat{F}_t^{(1)} F_t \right) \sum_{l=1}^j \delta_i^{(l)'} \right]}_{=O_p(T\delta)} + \underbrace{T^{-1} \sum_{t=1}^T \hat{F}_t^{(1)} e_{it}}_{=O_p(T^{-1/2})},$$

so that it has the same order as  $\hat{\lambda}_i$ . Similarly, it is shown that  $\hat{\lambda}_i^{(2)}$  has the same order as  $\hat{\lambda}_i$ . The unrestricted SSR, assuming one break at  $t = \lfloor \mu T \rfloor$  is

$$\begin{aligned} SSR(\mu) &= \sum_{t=1}^T \left( e_{it} + \lambda'_{it} F_t - \hat{\lambda}_i^{(1)'} \hat{F}_t^{(1)} - \hat{\lambda}_i^{(2)'} \hat{F}_t^{(2)} \right)^2, \\ &= \sum_{t=1}^T e_{it}^2 + k^{-2} \underbrace{\sum_{j=1}^m \left[ \sum_{l=1}^j \delta_i^{(l)'} \left( \sum_{t=T_{j-1}+1}^{T_j} F_t F_t' \right) \sum_{l=1}^j \delta_i^{(l)} \right]}_{=O_p(T\delta^2)} \\ &\quad + 2k^{-1} \underbrace{\sum_{j=1}^m \left[ \sum_{l=1}^j \delta_i^{(l)'} \left( \sum_{t=T_{j-1}+1}^{T_j} F_t e_{it} \right) \right]}_{=O_p(T^{1/2}\delta)} \\ &\quad + \underbrace{\hat{\lambda}_i^{(1)'} \left( \sum_{t=1}^{\lfloor \mu T \rfloor} \hat{F}_t \hat{F}_t' \right)}_{=(A.3)} \underbrace{\hat{\lambda}_i^{(1)}}_{=(A.3)} + \hat{\lambda}_i^{(2)'} \left( \sum_{t=\lfloor \mu T \rfloor + 1}^T \hat{F}_t \hat{F}_t' \right) \hat{\lambda}_i^{(2)} \\ &\quad - 2 \underbrace{\hat{\lambda}_i^{(1)'} \left( \sum_{t=1}^{\lfloor \mu T \rfloor} \hat{F}_t e_t \right)}_{=(A.3)} \underbrace{- 2 \hat{\lambda}_i^{(2)'} \left( \sum_{t=\lfloor \mu T \rfloor + 1}^T \hat{F}_t e_{it} \right)}_{=O_p(T^{1/2})} \\ &\quad - 2k^{-1} \underbrace{\hat{\lambda}_i^{(1)'} \sum_{j=1}^m \left[ \left( \sum_{t=T_{j-1}+1}^{T_j} \hat{F}_t^{(1)} F_t' \right) \sum_{l=1}^j \delta_i^{(l)} \right]}_{=(A.3)} \underbrace{\quad}_{=O_p(T\delta)} \\ &\quad - 2k^{-1} \hat{\lambda}_i^{(2)'} \sum_{j=1}^m \left[ \left( \sum_{t=T_{j-1}+1}^{T_j} \hat{F}_t^{(2)} F_t' \right) \sum_{l=1}^j \delta_i^{(l)} \right], \\ &= I + II + III + IV(1) + IV(2) + V(1) + V(2) + VI(1) + VI(1). \end{aligned}$$

Similar to the restricted SSR, if  $\liminf N/T = 0$ , then  $I = O_p(T)$ ,  $II = O_p(\delta^2)$ ,  $III = O_p(T^{1/2}\delta)$ ,  $IV = O_p(\delta^2)$ ,  $V = O_p(\delta)$ , and  $VI = O_p(\delta^2)$  so that the whole term is  $O_p(T)$ . If  $\liminf T/N = 0$ , then  $I = O_p(T)$ ,  $II = O_p(TN^{-1}\delta^2)$ ,  $III = O_p(TN^{-1/2}\delta)$ ,  $IV = O_p(1)$ ,

$V = O_p(1)$ , and  $VI = O_p(T^{1/2}N^{-1/2}\delta^2)$  and the whole term is  $O_p(T)$ . Finally,

$$\begin{aligned}
SSR_r - SSR(\mu) &= \hat{\lambda}'_i \left( \sum_{t=1}^T \hat{F}_t \hat{F}'_t \right) \hat{\lambda}_i - \hat{\lambda}_i^{(1)'} \left( \sum_{t=1}^{\lfloor \mu T \rfloor} \hat{F}_t \hat{F}'_t \right) \hat{\lambda}_i^{(1)} - \hat{\lambda}_i^{(2)'} \left( \sum_{t=\lfloor \mu T \rfloor + 1}^T \hat{F}_t \hat{F}'_t \right) \hat{\lambda}_i^{(2)} \\
&\quad - 2\hat{\lambda}'_i \left( \sum_{t=1}^T \hat{F}_t e_{it} \right) + 2\hat{\lambda}_i^{(1)'} \left( \sum_{t=1}^{\lfloor \mu T \rfloor} \hat{F}_t e_{it} \right) + 2\hat{\lambda}_i^{(2)'} \left( \sum_{t=\lfloor \mu T \rfloor + 1}^T \hat{F}_t e_{it} \right) \\
&\quad - 2k^{-1} \hat{\lambda}'_i \sum_{j=1}^m \left[ \left( \sum_{t=T_{j-1}+1}^{T_j} \hat{F}_t F'_t \right) \sum_{l=1}^j \delta_i^{(l)} \right] \\
&\quad + 2k^{-1} \hat{\lambda}_i^{(1)'} \sum_{j=1}^m \left[ \left( \sum_{t=T_{j-1}+1}^{T_j} \hat{F}_t^{(1)} F'_t \right) \sum_{l=1}^j \delta_i^{(l)} \right] \\
&\quad + 2k^{-1} \hat{\lambda}_i^{(2)'} \sum_{j=1}^m \left[ \left( \sum_{t=T_{j-1}+1}^{T_j} \hat{F}_t^{(2)} F'_t \right) \sum_{l=1}^j \delta_i^{(l)} \right] \\
&= IV - IV(1) - IV(2) + V - V(1) - V(2) + VI - VI(1) - VI(2).
\end{aligned}$$

If  $\liminf N/T = 0$ , then  $IV = O_p(\delta^2)$ ,  $V = O_p(\delta)$ , and  $VI = O_p(\delta^2)$ . In this case, the terms  $IV$ s and  $VI$ s dominate, and this is  $O_p(\delta^2)$ . If  $\liminf T/N = 0$ , then  $IV = O_p(1)$ ,  $V = O_p(1)$ , and  $VI = O_p(T^{1/2}N^{-1/2}\delta^2)$ . Since  $T/N \rightarrow 0$ , the whole term is  $O_p(1)$ . Therefore,

$$\begin{aligned}
Wald(\mu) &= \frac{SSR_r - SSR(\mu)}{SSR(\mu)/T} = \begin{cases} O_p(\delta^2) & \text{if } \liminf N/T = 0 \\ O_p(1) & \text{if } \liminf T/N = 0 \end{cases}, \\
LM(\mu) &= \frac{SSR_r - SSR(\mu)}{SSR_r/T} = \begin{cases} O_p(\delta^2) & \text{if } \liminf N/T = 0 \\ O_p(1) & \text{if } \liminf T/N = 0 \end{cases},
\end{aligned}$$

uniformly in  $\mu$ . The relationship  $Wald > LR > LM$  implies that  $LR(\mu)$  has the same order uniformly in  $\mu$  and Theorem 1(i) follows.

**Proof of Theorem 1 (ii):** We consider the regression

$$x_{it} = \lambda_i^{*'} K'^{-1} \hat{F}_t^* + \lambda_i^{*'} K'^{-1} \left( K' F_t^* - \hat{F}_t^* \right) + e_{it},$$

where  $\lambda_i^* = [\lambda_i' \gamma_i']'$ ,  $F_t^* = [F_t' s_t']'$ , and  $K = (\Lambda^* \Lambda^* / N) \left( F^{*'} \hat{F}^* / T \right) V^{-1}$  with  $V$  an  $r(m+1) \times r(m+1)$  diagonal matrix of the  $r(m+1)$  largest eigenvalues of  $XX'/(NT)$ . The Proof of Proposition 1 shows that the set of assumptions holds for the model (A.1). Hence, using the results of Bai (2003) for the PC estimator  $\hat{F}_t^*$ , it holds that

$$T^{-1} \sum_{t=1}^T \left\| \hat{F}_t^* - K' F_t^* \right\|^2 = O_p(c^{-2}), \quad (\text{A.4})$$

$$T^{-1} (\hat{F}^* - F^* K)' e_i = O_p(c^{-2}), \quad (\text{A.5})$$

$$T^{-1} (\hat{F}^* - F^* K)' \hat{F}^* = O_p(c^{-2}), \quad (\text{A.6})$$

where  $c = \min\{\sqrt{T}, \sqrt{N}\}$ .

The restricted residuals from the regression of  $x_{it}$  on  $\hat{F}_t^*$  under no changes in the factor loadings are

$$\hat{e}_{it} = e_{it} + \lambda_i^{*'} K'^{-1} (K' F_t^* - \hat{F}_t^*) + (\lambda_i^{*'} K'^{-1} - \hat{\lambda}_i^{*'}) \hat{F}_t^*.$$

We also define  $\hat{F}_t^{*(1)} = \hat{F}_t^* \times I(t \leq \lfloor \mu T \rfloor)$  and  $\hat{F}_t^{*(2)} = \hat{F}_t^* \times I(t > \lfloor \mu T \rfloor)$  for  $t = 1, \dots, T$ . Then, the unrestricted residuals, assuming one break in the loadings at  $t = \lfloor \mu T \rfloor$  are

$$\hat{e}_{it}(\mu) = e_{it} + \lambda_i^{*'} K'^{-1} (K' F_t^* - \hat{F}_t^*) + (\lambda_i^{*'} K'^{-1} - \hat{\lambda}_i^{*(1)'}) \hat{F}_t^{*(1)} + (\lambda_i^{*'} K'^{-1} - \hat{\lambda}_i^{*(2)'}) \hat{F}_t^{*(2)},$$

where  $\hat{\lambda}_i^{*(1)} = (\sum_{t=1}^{\lfloor \mu T \rfloor} \hat{F}_t^* \hat{F}_t^{*'})^{-1} (\sum_{t=1}^{\lfloor \mu T \rfloor} \hat{F}_t^* x_{it})$  and  $\hat{\lambda}_i^{*(2)} = (\sum_{t=\lfloor \mu T \rfloor+1}^T \hat{F}_t^* \hat{F}_t^{*'})^{-1} (\sum_{t=\lfloor \mu T \rfloor+1}^T \hat{F}_t^* x_{it})$ . The restricted SSR is

$$\begin{aligned} SSR_r &= e_i' e_i + \lambda_i^{*'} K'^{-1} \underbrace{(F^* K - \hat{F}^*)' (F^* K - \hat{F}^*)}_{=O_p(Tc^{-2}) \text{ by (A.4)}} K^{-1} \lambda_i^* + 2 \underbrace{e_i' (F^* K - \hat{F}^*)}_{=O_p(Tc^{-2}) \text{ by (A.5)}} K^{-1} \lambda_i^* \\ &+ \underbrace{(\lambda_i^{*'} K'^{-1} - \hat{\lambda}_i^{*'})}_{=O_p(s)} \underbrace{\hat{F}^* \hat{F}^*'}_{=O_p(T)} \underbrace{(K'^{-1} \lambda_i^* - \hat{\lambda}_i^*)}_{=O_p(s)} + 2 \underbrace{e_i' \hat{F}^*}_{=O_p(T^{1/2})} \underbrace{(K^{-1} \lambda_i^* - \hat{\lambda}_i^*)}_{=O_p(s)} \\ &+ 2 \lambda_i^{*'} K'^{-1} \underbrace{(F^* K - \hat{F}^*)' \hat{F}^*}_{=O_p(Tc^{-2}) \text{ by (A.6)}} \underbrace{(K^{-1} \lambda_i^* - \hat{\lambda}_i^*)}_{=O_p(s)}, \\ &= I + II + III + IV + V + VI. \end{aligned}$$

However,  $I = O_p(T)$ ,  $II = O_p(Tc^{-2})$ ,  $III = O_p(Tc^{-2})$ ,  $IV = O_p(Ts^2)$ ,  $V = O_p(T^{1/2}s)$ , and  $VI = O_p(Tsc^{-2})$ , where  $s$  represents the order of  $\|\hat{\lambda}_i^* - K^{-1} \lambda_i^*\|$ . Under the standard assumption as ours, Bai (2003) shows

$$s = \begin{cases} T^{-1/2} & \text{if } \liminf \sqrt{T}/N = 0 \\ N^{-1} & \text{if } \liminf \sqrt{T}/N > 0 \end{cases}.$$

Let us first suppose that  $\liminf T/N = 0$ . This case gives  $s = T^{-1/2}$  and  $c = \sqrt{T}$ . Hence,  $I = O_p(T)$ ,  $II = O_p(1)$ ,  $III = O_p(1)$ ,  $IV = O_p(1)$ ,  $V = O_p(1)$ , and  $VI = O_p(T^{-1/2})$ . The term  $I$  dominates and  $SSR_r = O_p(T)$ . Second, suppose that  $\liminf T/N > 0$  and  $\liminf \sqrt{T}/N = 0$ . This case gives  $s = T^{-1/2}$  and  $c = \sqrt{N}$  so that  $I = O_p(T)$ ,  $II = O_p(TN^{-1})$ ,  $III = O_p(TN^{-1})$ ,  $IV = O_p(1)$ ,  $V = O_p(1)$ , and  $VI = O_p(T^{1/2}N^{-1})$ . Again the term  $I$  dominates and  $SSR_r = O_p(T)$ . Finally suppose  $\liminf \sqrt{T}/N > 0$ . This case gives  $s = N^{-1}$  and  $c = \sqrt{N}$ . Hence,  $I = O_p(T)$ ,  $II = O_p(TN^{-1})$ ,  $III = O_p(TN^{-1})$ ,  $IV = O_p(TN^{-2})$ ,  $V = O_p(T^{1/2}N^{-1})$ , and  $VI = O_p(TN^{-2})$  so that the term  $I$  dominates and  $SSR_r = O_p(T)$ . Therefore,  $SSR_r = O_p(T)$  for every relative rate of  $N$  and  $T$ . The

unrestricted SSR is

$$\begin{aligned}
SSR(\mu) &= e_i' e_i + \lambda_i^{*'} K'^{-1} \underbrace{(F^* K - \hat{F}^*)' (F^* K - \hat{F}^*)}_{=O_p(Tc^{-2}) \text{ by (A.4)}} K^{-1} \lambda_i \\
&+ 2 \underbrace{e_i' (F^* K - \hat{F}^*)}_{=O_p(Tc^{-2}) \text{ by (A.5)}} K^{-1} \lambda_i^* \\
&+ \underbrace{(\lambda_i^{*'} K'^{-1} - \hat{\lambda}_i^{*(1)'})}_{=O_p(s)} \underbrace{\left( \sum_{t=1}^{\lfloor \mu T \rfloor} \hat{F}_t^* \hat{F}_t^{*'} \right)}_{=O_p(T)} \underbrace{(K^{-1} \lambda_i^* - \hat{\lambda}_i^{*(1)})}_{=O_p(s)} \\
&+ (\lambda_i^{*'} K'^{-1} - \hat{\lambda}_i^{*(2)'}) \left( \sum_{t=\lfloor \mu T \rfloor + 1}^T \hat{F}_t^* \hat{F}_t^{*'} \right) (K^{-1} \lambda_i^* - \hat{\lambda}_i^{*(2)}) \\
&+ 2 \underbrace{\left( \sum_{t=1}^{\lfloor \mu T \rfloor} e_{it} \hat{F}_t^{*'} \right)}_{=O_p(T^{1/2})} \underbrace{(K^{-1} \lambda_i^* - \hat{\lambda}_i^{*(1)})}_{=O_p(s)} + 2 \left( \sum_{t=\lfloor \mu T \rfloor + 1}^T e_{it} \hat{F}_t^{*'} \right) (K^{-1} \lambda_i^* - \hat{\lambda}_i^{*(2)}) \\
&+ 2 \lambda_i^{*'} K'^{-1} \underbrace{\left[ \sum_{t=1}^{\lfloor \mu T \rfloor} (K' F_t^* - \hat{F}_t^*) \hat{F}_t^{*'} \right]}_{=O_p(Tc^{-2}) \text{ by (A.6)}} \underbrace{(K^{-1} \lambda_i^* - \hat{\lambda}_i^{*(1)})}_{=O_p(s)} \\
&+ 2 \lambda_i^{*'} K'^{-1} \left[ \sum_{t=\lfloor \mu T \rfloor + 1}^T (K' F_t^* - \hat{F}_t^*) \hat{F}_t^{*'} \right] (K^{-1} \lambda_i^* - \hat{\lambda}_i^{*(2)}) \\
&= I + II + III + IV(1) + IV(2) + V(1) + V(2) + VI(1) + VI(2).
\end{aligned}$$

For all the terms  $I$ ,  $II$ ,  $III$ ,  $IV$ s,  $V$ s, and  $VI$ s, the same discussion as for the restricted SSR goes through to evaluate the orders of convergence. Therefore, similarly we can show that

$SSR(\mu)$  is  $O_p(T)$  uniformly in  $\mu$ . Finally,

$$\begin{aligned}
SSR_r - SSR(\mu) &= \underbrace{(\lambda_i^{*'} K'^{-1} - \hat{\lambda}_i^{*'})}_{=O_p(s)} \underbrace{\hat{F}^{*'} \hat{F}^*}_{=O_p(T)} \underbrace{(K^{-1} \lambda_i^* - \hat{\lambda}_i^*)}_{=O_p(s)} \\
&- \underbrace{(\lambda_i^{*'} K'^{-1} - \hat{\lambda}_i^{*(1)'})}_{=O_p(s)} \underbrace{\left( \sum_{t=1}^{\lfloor \mu T \rfloor} \hat{F}_t^* \hat{F}_t^{*'} \right)}_{=O_p(T)} \underbrace{(K^{-1} \lambda_i^* - \hat{\lambda}_i^{*(1)})}_{=O_p(s)} \\
&- (\lambda_i^{*'} K'^{-1} - \hat{\lambda}_i^{*(2)'}) \left( \sum_{t=\lfloor \mu T \rfloor + 1}^T \hat{F}_t^* \hat{F}_t^{*'} \right) (K^{-1} \lambda_i^* - \hat{\lambda}_i^{*(2)}) \\
&+ 2 \underbrace{e_i' \hat{F}^*}_{=O_p(T^{1/2})} \underbrace{(K^{-1} \lambda_i^* - \hat{\lambda}_i^*)}_{=O_p(s)} \\
&- 2 \underbrace{\left( \sum_{t=1}^{\lfloor \mu T \rfloor} e_{it} \hat{F}_t^{*'} \right)}_{=O_p(T^{1/2})} \underbrace{(K^{-1} \lambda_i^* - \hat{\lambda}_i^{*(1)})}_{=O_p(s)} \\
&- 2 \left( \sum_{t=\lfloor \mu T \rfloor + 1}^T e_{it} \hat{F}_t^{*'} \right) (K^{-1} \lambda_i^* - \hat{\lambda}_i^{*(2)}) \\
&+ 2 \lambda_i^{*'} K'^{-1} \underbrace{(F^* K - \hat{F}^*)' \hat{F}^*}_{=O_p(Tc^{-2})} \underbrace{(K^{-1} \lambda_i^* - \hat{\lambda}_i^*)}_{=O_p(s)} \\
&- 2 \lambda_i^{*'} K'^{-1} \underbrace{\left[ \sum_{t=1}^{\lfloor \mu T \rfloor} (K' F_t^* - \hat{F}_t^*) \hat{F}_t^{*'} \right]}_{=O_p(Tc^{-2})} \underbrace{(K^{-1} \lambda_i^* - \hat{\lambda}_i^{*(1)})}_{=O_p(s)} \\
&- 2 \lambda_i^{*'} K'^{-1} \left[ \sum_{t=\lfloor \mu T \rfloor + 1}^T (K' F_t^* - \hat{F}_t^*) \hat{F}_t^{*'} \right] (K^{-1} \lambda_i^* - \hat{\lambda}_i^{*(2)}), \\
&= IV - IV(1) - IV(2) + V - V(1) - V(2) + VI - VI(1) - VI(2).
\end{aligned}$$

First, suppose that  $\liminf T/N = 0$ . This case gives  $s = T^{-1/2}$  and  $c = \sqrt{T}$ . Hence,  $IV = O_p(1)$ ,  $V = O_p(1)$ , and  $VI = O_p(T^{-1/2})$  so that  $SSR_r - SSR(\mu) = O_p(1)$ . Second, suppose that  $\liminf T/N > 0$  and  $\liminf \sqrt{T}/N = 0$ . This case gives  $s = T^{-1/2}$  and  $c = \sqrt{N}$  so that  $IV = O_p(1)$ ,  $V = O_p(1)$ , and  $VI = O_p(T^{1/2}N^{-1})$ . Since  $\sqrt{T}/N \rightarrow 0$ ,  $IV$  and  $V$  dominate, and  $SSR_r - SSR(\mu) = O_p(1)$ . Finally, suppose that  $\liminf \sqrt{T}/N > 0$ . This case gives  $s = N^{-1}$  and  $c = \sqrt{N}$ . Hence,  $IV = O_p(TN^{-2})$ ,  $V = O_p(T^{1/2}N^{-1})$ , and  $VI = O_p(TN^{-2})$  so that  $SSR_r - SSR(\mu) = O_p(TN^{-2})$ . Hence,

$$\begin{aligned}
Wald &= \frac{SSR_r - SSR(\mu)}{SSR(\mu)/T} = \begin{cases} O_p(1) & \text{if } \liminf \sqrt{T}/N = 0 \\ O_p(TN^{-2}) & \text{if } \liminf \sqrt{T}/N > 0 \end{cases} \\
LM &= \frac{SSR_r - SSR(\mu)}{SSR_r/T} = \begin{cases} O_p(1) & \text{if } \liminf \sqrt{T}/N = 0 \\ O_p(TN^{-2}) & \text{if } \liminf \sqrt{T}/N > 0 \end{cases}
\end{aligned}$$

uniformly in  $\mu$ . The relationship  $Wald > LR > LM$  in linear models uniformly in  $\mu$  implies that the same results apply to  $LR$ .

**Proof of Theorem 2:** Let  $\sum_1$  and  $\sum_2$  denote  $\sum_{t=1}^{\lfloor T\mu_0 \rfloor}$  and  $\sum_{t=\lfloor T\mu_0 \rfloor+1}^T$ , respectively. We

primarily consider the case of  $m = 1$  but extend the proof to the cases of  $m > 1$  if necessary. The model is

$$x_{it} = \lambda'_i F_t + \gamma'_i F_t \cdot I(t > \lfloor T\mu_0 \rfloor) + e_{it}. \quad (\text{A.7})$$

Suppose we identify  $r^*$  factors that are  $r^*$  linear combinations of the total  $r(m+1)$  factors  $F_t$  and  $F_t \cdot I(t > \lfloor T\mu_0 \rfloor)$  so that  $Q_{11}F_t + Q_{12}F_t \cdot I(t > \lfloor T\mu_0 \rfloor)$ . Note that  $Q_{11}$  and  $Q_{12}$  are any  $r^* \times r$  and  $r^* \times r$  matrices with full column ranks specified later.<sup>18</sup> Let  $Q = [Q_{11} \ Q_{12}]$  be a  $r^* \times r(m+1)$  matrix. By doing so, model (A.7) is equivalent to

$$x_{it} = \bar{\lambda}'_i [Q_{11}F_t + Q_{12}F_t \cdot I(t > \lfloor T\mu_0 \rfloor)] + \bar{\gamma}'_i [Q_{21}F_t + Q_{22}F_t \cdot I(t > \lfloor T\mu_0 \rfloor)] + e_{it}, \quad (\text{A.8})$$

where  $Q_{21}$  and  $Q_{22}$  are  $(r(m+1) - r^*) \times r$  and  $(r(m+1) - r^*) \times m$  matrices and  $\bar{\lambda}_i$  and  $\bar{\gamma}_i$  are appropriately defined time invariant matrices. The key fact is that using only the factors  $Q_{11}F_t + Q_{12}F_t \cdot I(t > \lfloor T\mu_0 \rfloor)$  but not  $Q_{21}F_t + Q_{22}F_t \cdot I(t > \lfloor T\mu_0 \rfloor)$  yields the model (A.8) equivalent to

$$x_{it} = \lambda_{it}^* [Q_{11}F_t + Q_{12}F_t \cdot I(t > \lfloor T\mu_0 \rfloor)] + e_{it}, \quad (\text{A.9})$$

with coefficients with break

$$\lambda_{it}^* = \begin{cases} \lambda_i^{*(1)} & \text{for } t \leq \lfloor T\mu_0 \rfloor, \\ \lambda_i^{*(2)} & \text{for } t \geq \lfloor T\mu_0 \rfloor + 1. \end{cases}$$

The exact forms of the regime-specific coefficients are  $\lambda_i^{*(1)'} = (\bar{\lambda}'_i Q_{11} + \bar{\gamma}'_i Q_{21}) Q_{11}^-$  and  $\lambda_i^{*(2)'} = (\bar{\lambda}'_i (Q_{11} + Q_{12}) + \bar{\gamma}'_i (Q_{21} + Q_{22})) (Q_{11} + Q_{12})^-$ . Note that  $Q_{11}^-$  and  $(Q_{11} + Q_{12})^-$  are the Moore-Penrose inverses of  $Q_{11}$  and  $Q_{11} + Q_{12}$  that exist as long as  $r^* \geq r$ . Since  $Q = [Q_{11} \ Q_{12}]$  has not been specified, we regard it as  $Q = BK$  with  $B = [I_{r^*} \ 0_{r^* \times [r(m+1) - r^*]}]$  a  $r^* \times r(m+1)$  selector matrix where  $K$  is defined in the Proof of Theorem 1 (ii) if the first  $r^*$  principal components are used.<sup>19</sup> Therefore, the model (A.9) is simply

$$\begin{aligned} x_{it} &= \lambda_{it}^{*'} B K F_t^* + e_{it}, \\ &= \lambda_{it}^{*'} B \hat{F}_t^* + \lambda_{it}^{*'} B (K F_t^* - \hat{F}_t^*) + e_{it}, \end{aligned} \quad (\text{A.10})$$

where  $F^*$  is defined as in (A.1) and  $\hat{F}^*$  are the  $r(m+1)$  principal components of  $XX'/(NT)$ . Note that  $B\hat{F}_t^*$  are the factor estimates when one assumes that there are only  $r^*$  factors.

Based on the model (A.10), the restricted residuals from the regression of  $x_{it}$  on  $B\hat{F}_t^*$  are

$$\hat{e}_{it} = e_{it} + \lambda_{it}^{*'} B \xi_t + \lambda_{it}^{*'} B \hat{F}_t^* - \hat{\lambda}_i^{*'} B \hat{F}_t^*,$$

<sup>18</sup>If  $m > 1$ , then additional matrices  $Q_{13}, \dots, Q_{1m}$  appear, but the following discussion essentially holds.

<sup>19</sup>Otherwise,  $B$  is changed according to the choice of the factors as regressors.

where  $\xi_t = KF_t^* - \hat{F}_t^*$  and

$$\begin{aligned}\hat{\lambda}_i^* &= \left( \begin{array}{c} B \underbrace{\hat{F}_t^* \hat{F}_t^{*'}}_{T \times I_r(m+1)} B' \end{array} \right)^{-1} \left[ \sum B \hat{F}_t^* \hat{F}_t^{*'} B' \lambda_{it}^* + \sum B \hat{F}_t^* \xi_t' B' \lambda_{it}^* + \sum B \hat{F}_t^* e_{it} \right], \\ &= (BB')^{-1} BT^{-1} \underbrace{\sum \hat{F}_t^* \hat{F}_t^{*'} B' \lambda_{it}^*}_{=I} + (BB')^{-1} BT^{-1} \underbrace{\sum \hat{F}_t^* \xi_t' B' \lambda_{it}^*}_{=II} \\ &\quad + (BB')^{-1} B \underbrace{T^{-1} \sum \hat{F}_t^* e_{it}}_{=O_p(c^{-2}) + O_p(T^{-1/2})}.\end{aligned}$$

However,

$$\begin{aligned}I &= \underbrace{\left( T^{-1} \sum \hat{F}_t^* \hat{F}_t^{*'} \right) B' \lambda_i^{*(1)}}_{=I_r(m+1)} + \underbrace{\left( T^{-1} \sum_2 \hat{F}_t^* \hat{F}_t^{*'} \right) B' \left( \lambda_i^{*(2)} - \lambda_i^{*(1)} \right)}_{=O_p(1)} = O_p(\delta), \\ II &= \underbrace{\left( T^{-1} \sum \hat{F}_t^* \xi_t' \right) B' \lambda_i^{*(1)}}_{=O_p(c^{-2})} + \underbrace{\left( T^{-1} \sum_2 \hat{F}_t^* \xi_t' \right) B' \left( \lambda_i^{*(2)} - \lambda_i^{*(1)} \right)}_{=O_p(c^{-2})} = O_p(c^{-2}),\end{aligned}$$

so that  $I$  dominates, and  $\hat{\lambda}_i^* = O_p(\delta)$ . Now, the restricted SSR is

$$\begin{aligned}SSR_r &= \sum e_{it}^2 + \sum \lambda_{it}^{*'} B \xi_t \xi_t' B' \lambda_{it}^* + \sum \lambda_{it}^{*'} B \hat{F}_t^* \hat{F}_t^{*'} B' \lambda_{it}^* \\ &\quad + \hat{\lambda}_i^{*'} B \left( \sum \xi_t \xi_t' \right) B' \hat{\lambda}_i^* \\ &\quad + 2 \sum \lambda_{it}^{*'} B \xi_t e_{it} + 2 \sum \lambda_{it}^{*'} B \hat{F}_t^* e_{it} - 2 \hat{\lambda}_i^{*'} \left( \sum B \hat{F}_t^* e_{it} \right) \\ &\quad + 2 \sum \lambda_{it}^{*'} \xi_t \hat{F}_t^{*'} B' \lambda_{it}^* - 2 \left( \sum \lambda_{it}^{*'} \xi_t \hat{F}_t^{*'} B' \right) \hat{\lambda}_i^* - 2 \left( \sum \lambda_{it}^{*'} B \hat{F}_t^* \hat{F}_t^{*'} B' \right) \hat{\lambda}_i^*, \\ &= I + II + III + IV + V + VI + VII + VIII + XI + X.\end{aligned}$$

Then,  $I = O_p(T)$  and the rest of the terms are shown as follows:

$$\begin{aligned}II &= \lambda_i^{*(1)'} B \underbrace{\left( \sum \xi_t \xi_t' \right) B' \lambda_i^{*(1)}}_{=O_p(Tc^{-2})} + 2 \left( \lambda_i^{*(2)'} - \lambda_i^{*(1)'} \right) B \underbrace{\left( \sum_2 \xi_t \xi_t' \right) B' \lambda_i^{*(1)}}_{=O_p(Tc^{-2})} \\ &\quad + \left( \lambda_i^{*(2)'} - \lambda_i^{*(1)'} \right) B \underbrace{\left( \sum_2 \xi_t \xi_t' \right) B' \left( \lambda_i^{*(2)} - \lambda_i^{*(1)} \right)}_{=O_p(Tc^{-2})}, \\ &= O_p(Tc^{-2}\delta^2), \\ III &= \lambda_i^{*(1)'} B \underbrace{\left( \sum \hat{F}_t^* \hat{F}_t^{*'} \right) B' \lambda_i^{*(1)}}_{=O_p(T)} + 2 \left( \lambda_i^{*(2)'} - \lambda_i^{*(1)'} \right) B \underbrace{\left( \sum_2 \hat{F}_t^* \hat{F}_t^{*'} \right) B' \lambda_i^{*(1)}}_{=O_p(T)} \\ &\quad + \left( \lambda_i^{*(2)'} - \lambda_i^{*(1)'} \right) B \underbrace{\left( \sum_2 \hat{F}_t^* \hat{F}_t^{*'} \right) B' \left( \lambda_i^{*(2)} - \lambda_i^{*(1)} \right)}_{=O_p(T)}, \\ &= O_p(T\delta^2),\end{aligned}$$

$$\begin{aligned}
IV &= \hat{\lambda}_i^{*'} B \underbrace{\left( \sum \xi_t \xi_t' \right)}_{=O_p(Tc^{-2})} B' \hat{\lambda}_i^* = O_p(Tc^{-2}\delta^2), \\
V &= \lambda_i^{*(1)'} B \underbrace{\sum \xi_t e_{it}}_{=O_p(Tc^{-2})} + 2 \left( \lambda_i^{*(2)'} - \lambda_i^{*(1)'} \right) B \underbrace{\sum_2 \xi_t e_{it}}_{=O_p(Tc^{-2})} = O_p(Tc^{-2}\delta), \\
VI &= \lambda_i^{*(1)'} B \underbrace{\sum \hat{F}_t^* e_{it}}_{=O_p(T^{1/2})+O_p(Tc^{-2})} + 2 \left( \lambda_i^{*(2)'} - \lambda_i^{*(1)'} \right) B \underbrace{\sum_2 \hat{F}_t^* e_{it}}_{=O_p(T^{1/2})+O_p(Tc^{-2})} = O_p(T^{1/2}\delta) + O_p(Tc^{-2}\delta), \\
VII &= -2\hat{\lambda}_i^{*'} B \left( \sum \hat{F}_t^* e_{it} \right) = O_p(T^{1/2}\delta) + O_p(Tc^{-2}\delta), \\
VIII &= \lambda_i^{*(1)'} \underbrace{\left( \sum \xi_t \hat{F}_t^{*'} \right)}_{=O_p(Tc^{-2})} B' \lambda_i^{*(1)} + 2 \left( \lambda_i^{*(2)'} - \lambda_i^{*(1)'} \right) \underbrace{\left( \sum_2 \xi_t \hat{F}_t^{*'} \right)}_{=O_p(Tc^{-2})} B' \lambda_i^{*(1)} \\
&\quad + \left( \lambda_i^{*(2)'} - \lambda_i^{*(1)'} \right) \underbrace{\left( \sum_2 \xi_t \hat{F}_t^{*'} \right)}_{=O_p(Tc^{-2})} B' \left( \lambda_i^{*(2)} - \lambda_i^{*(1)} \right) = O_p(Tc^{-2}\delta^2), \\
XI &= \lambda_i^{*(1)'} \underbrace{\left( \sum \xi_t \hat{F}_t^{*'} \right)}_{=O_p(Tc^{-2})} B' \hat{\lambda}_i^* + \left( \lambda_i^{*(2)'} - \lambda_i^{*(1)'} \right) \underbrace{\left( \sum_2 \xi_t \hat{F}_t^{*'} \right)}_{=O_p(Tc^{-2})} B' \hat{\lambda}_i^* = O_p(Tc^{-2}\delta^2), \\
X &= \lambda_i^{*(1)'} B \underbrace{\left( \sum \hat{F}_t^* \hat{F}_t^{*'} \right)}_{=T \times I_{r(m+1)}} B' \hat{\lambda}_i^* + \left( \lambda_i^{*(2)'} - \lambda_i^{*(1)'} \right) B \underbrace{\left( \sum_2 \hat{F}_t^* \hat{F}_t^{*'} \right)}_{=O_p(T)} B' \hat{\lambda}_i^* = O_p(T\delta^2).
\end{aligned}$$

Hence, the terms *III* and *X* dominate and  $SSR_r = O_p(T\delta^2)$ .

Consider unrestricted residuals accounting for one break at  $t = \lfloor T\mu \rfloor$ . Without loss of generality, we let  $\mu > \mu_0$ .

$$\hat{e}_{it}(\mu) = e_{it} + \lambda_{it}^{*'} B \xi_t + \lambda_{it}^{*'} B \hat{F}_t^* - \hat{\lambda}_i^{*(1)'} B \hat{F}_t^* \cdot I(t \leq \lfloor T\mu \rfloor) - \hat{\lambda}_i^{*(2)'} B \hat{F}_t^* \cdot I(t \geq \lfloor T\mu \rfloor + 1),$$

where

$$\begin{aligned}
\hat{\lambda}_i^{*(1)} &= \left( \sum_1 B \hat{F}_t^* \hat{F}_t^{*'} B' \right)^{-1} \left[ \sum_1 B \hat{F}_t^* \hat{F}_t^{*'} B' \lambda_{it}^* + \sum_1 B \hat{F}_t^* \xi_t' B' \lambda_{it}^* + B \sum_1 \hat{F}_t^* e_i \right], \\
\hat{\lambda}_i^{*(2)} &= \left( \sum_2 B \hat{F}_t^* \hat{F}_t^{*'} B' \right)^{-1} \left[ \sum_2 B \hat{F}_t^* \hat{F}_t^{*'} B' \lambda_{it}^* + \sum_2 B \hat{F}_t^* \xi_t' B' \lambda_{it}^* + B \sum_2 \hat{F}_t^* e_i \right].
\end{aligned}$$

Since  $\sum_1 B \hat{F}_t^* \hat{F}_t^{*'} B' \lambda_{it}^*$  and  $\sum_1 B \hat{F}_t^* \xi_t' B' \lambda_{it}^*$  include a break,  $\hat{\lambda}_i^{*(1)} = O_p(\delta)$  and  $\hat{\lambda}_i^{*(2)} = O_p(1)$ . The result is the opposite when  $\mu < \mu_0$ . It is also shown that if we have multiple breaks  $m \geq 2$ , then  $\hat{\lambda}_i^{*(1)} = O_p(\delta)$  and  $\hat{\lambda}_i^{*(2)} = O_p(1)$  when all the breaks are before  $\lfloor T\mu \rfloor$ ,  $\hat{\lambda}_i^{*(1)} = O_p(1)$

and  $\hat{\lambda}_i^{*(2)} = O_p(\delta)$  when all the breaks are after  $\lfloor T\mu \rfloor$ , and  $\hat{\lambda}_i^{*(1)}, \hat{\lambda}_i^{*(2)} = O_p(\delta)$  otherwise. The unrestricted SSR is

$$\begin{aligned}
SSR(\mu) &= \sum e_{it}^2 + \sum \lambda_{it}^{*'} B \xi_t \xi_t' B' \lambda_{it}^* + \sum \lambda_{it}^{*'} B \hat{F}_t^* \hat{F}_t^{*'} B' \lambda_{it}^* \\
&\quad + \hat{\lambda}_i^{*(1)'} \left( \sum_1 B \xi_t \xi_t' B' \right) \hat{\lambda}_i^{*(1)} + \hat{\lambda}_i^{*(2)'} \left( \sum_2 B \xi_t \xi_t' B' \right) \hat{\lambda}_i^{*(2)} \\
&\quad + 2 \sum \lambda_{it}^{*'} B \xi_t e_{it} + 2 \sum \lambda_{it}^{*'} B \hat{F}_t^* e_{it} \\
&\quad - 2 \hat{\lambda}_i^{*(1)'} \left( \sum_1 B \hat{F}_t^* e_{it} \right) - 2 \hat{\lambda}_i^{*(2)'} \left( \sum_2 B \hat{F}_t^* e_{it} \right) \\
&\quad + 2 \sum \lambda_{it}^{*'} \xi_t \hat{F}_t^{*'} B' \lambda_{it}^* \\
&\quad - 2 \left( \sum_1 \lambda_{it}^{*'} \xi_t \hat{F}_t^{*'} B' \right) \hat{\lambda}_i^{*(1)} - 2 \left( \sum_2 \lambda_{it}^{*'} \xi_t \hat{F}_t^{*'} B' \right) \hat{\lambda}_i^{*(2)} \\
&\quad - 2 \left( \sum_1 \lambda_{it}^{*'} B \hat{F}_t^* \hat{F}_t^{*'} B' \right) \hat{\lambda}_i^{*(1)} - 2 \left( \sum_2 \lambda_{it}^{*'} B \hat{F}_t^* \hat{F}_t^{*'} B' \right) \hat{\lambda}_i^{*(2)}, \\
&= I + II + III + IV(1) + IV(2) + V + VI + VII(1) + VII(2) \\
&\quad + VIII + XI(1) + XI(2) + X(1) + X(2).
\end{aligned}$$

However,  $I, II, III, V, VI,$  and  $VIII$  are the same as the ones in the restricted SSR. We can also follow a similar discussion as in the restricted SSR to derive

$$\begin{aligned}
IV(1), IV(2) &= O_p(Tc^{-2}\delta^2), O_p(Tc^{-2}), \\
VII(1), VII(2) &= O_p(Tc^{-2}\delta) + O_p(T^{1/2}\delta), O_p(Tc^{-2}) + O_p(T^{1/2}), \\
XI(1), XI(2) &= O_p(Tc^{-2}\delta^2), O_p(Tc^{-2}), \\
X(1), X(2) &= O_p(T\delta^2), O_p(T),
\end{aligned}$$

so that  $SSR(\mu) = O_p(T\delta^2)$  uniformly in  $\mu$  except for the case of  $\mu = \mu_0$  and  $m = 1$ . Finally,

$$\begin{aligned}
&SSR_r - SSR(\mu) \\
&= \hat{\lambda}_i^{*'} B \left( \sum \xi_t \xi_t' \right) B' \hat{\lambda}_i^* - \hat{\lambda}_i^{*(1)'} B \left( \sum_1 \xi_t \xi_t' \right) B' \hat{\lambda}_i^{*(1)} - \hat{\lambda}_i^{*(2)'} B \left( \sum_2 \xi_t \xi_t' \right) B' \hat{\lambda}_i^{*(2)} \\
&\quad - 2 \hat{\lambda}_i^{*'} B \left( \sum \hat{F}_t^* e_{it} \right) + 2 \hat{\lambda}_i^{*(1)'} B \left( \sum_1 \hat{F}_t^* e_{it} \right) + 2 \hat{\lambda}_i^{*(2)'} B \left( \sum_2 \hat{F}_t^* e_{it} \right) \\
&\quad - 2 \left( \sum \lambda_{it}^{*'} \xi_t \hat{F}_t^{*'} \right) B' \hat{\lambda}_i^* + 2 \left( \sum_1 \lambda_{it}^{*'} \xi_t \hat{F}_t^{*'} \right) B' \hat{\lambda}_i^{*(1)} + 2 \left( \sum_2 \lambda_{it}^{*'} \xi_t \hat{F}_t^{*'} \right) B' \hat{\lambda}_i^{*(2)} \\
&\quad - 2 \left( \sum \lambda_{it}^{*'} B \hat{F}_t^* \hat{F}_t^{*'} B' \right) \hat{\lambda}_i^* + 2 \left( \sum_1 \lambda_{it}^{*'} B \hat{F}_t^* \hat{F}_t^{*'} B' \right) \hat{\lambda}_i^{*(1)} + 2 \left( \sum_2 \lambda_{it}^{*'} B \hat{F}_t^* \hat{F}_t^{*'} B' \right) \hat{\lambda}_i^{*(2)}, \\
&= IV - IV(1) - IV(2) + VII - VII(1) - VII(2) \\
&\quad + XI - XI(1) - XI(2) + X - X(1) - X(2).
\end{aligned}$$

However, we know that

$$\begin{aligned}
IV, IV(1) &= O_p(Tc^{-2}\delta^2) \text{ and } IV(2) = O_p(Tc^{-2}), \\
VII, VII(1) &= O_p(Tc^{-2}\delta) + O_p(T^{1/2}\delta) \text{ and } VII(2) = O_p(Tc^{-2}) + O_p(T^{1/2}), \\
XI, XI(1) &= O_p(Tc^{-2}\delta^2) \text{ and } XI(2) = O_p(Tc^{-2}), \\
X, X(1) &= O_p(T\delta^2) \text{ and } X(2) = O_p(T),
\end{aligned}$$

so that the term  $X$  dominates, and  $SSR_r - SSR(\mu) = O_p(T\delta^2)$  uniformly in  $\mu$ . Hence,

$$\begin{aligned} Wald &= \frac{SSR_r - SSR(\mu)}{SSR(\mu)/T} = O_p(T), \\ LM &= \frac{SSR_r - SSR(\mu)}{SSR_r/T} = O_p(T), \end{aligned}$$

uniformly in  $\mu$  except for the case of  $\mu = \mu_0$  and  $m = 1$ . The relationship  $Wald > LR > LM$  in linear models uniformly in  $\mu$  implies that the same results apply to  $LR$ . When  $m = 1$ ,  $SSR(\mu_0) = O_p(T)$ , with which the following holds:

$$Wald = \frac{SSR_r - SSR(\mu_0)}{SSR(\mu_0)/T} = O_p(T\delta^2).$$

**Proof of Theorem 3 (i):** Consider the regression of  $x_i$  on  $\hat{f}_j$  where  $\hat{f}_j$  is the  $j$ th principal component estimate for the model

$$x_{it} = \lambda_{ji}f_{jt} + v_{jit}.$$

The restricted residual is

$$\begin{aligned} \hat{v}_{jit} &= v_{jit} + \lambda_{ji}f_{jt} - \hat{\lambda}_{ji}\hat{f}_{jt} \\ &= v_{jit} - \left(\hat{\lambda}_{ji} - \lambda_{ji}\right) f_{jt} - \hat{\lambda}_{ji} \left(\hat{f}_{jt} - f_{jt}\right), \end{aligned}$$

where  $\hat{\lambda}_{ji} = \left(T^{-1} \sum_{t=1}^T \hat{f}_{jt}^2\right)^{-1} T^{-1} \sum_{t=1}^T \hat{f}_{jt}x_{it} = T^{-1} \sum_{t=1}^T \hat{f}_{jt}x_{it}$ .

Then, the restricted SSR is

$$\begin{aligned} SSR_r &= \sum_{t=1}^T v_{jit}^2 + (\hat{\lambda}_{ji} - \lambda_{ji})^2 \sum_{t=1}^T f_{jt}^2 \\ &\quad + \hat{\lambda}_{ji}^2 \sum_{t=1}^T (\hat{f}_{jt} - f_{jt})^2 - 2\hat{\lambda}_{ji} \sum_{t=1}^T (\hat{f}_{jt} - f_{jt})v_{jit} \\ &\quad - 2(\hat{\lambda}_{ji} - \lambda_{ji}) \sum_{t=1}^T f_{jt}v_{jit} + 2\hat{\lambda}_{ji}(\hat{\lambda}_{ji} - \lambda_{ji}) \sum_{t=1}^T f_{jt}(\hat{f}_{jt} - f_{jt}). \end{aligned}$$

Similarly, the unrestricted residual considering a break at  $t = \lfloor \mu T \rfloor$  is

$$\hat{v}_{jit}(\mu) = v_{jit} + \lambda_{ji}f_{jt} - \hat{\lambda}_{ji}^{(1)}\hat{f}_{jt}^{(1)} - \hat{\lambda}_{ji}^{(2)}\hat{f}_{jt}^{(2)},$$

where  $\hat{\lambda}_{ji}^{(s)} = \left(T^{-1} \sum_s \hat{f}_{jt}^2\right)^{-1} T^{-1} \sum_s \hat{f}_{jt}x_{it}$  for  $s = 1, 2$ .

The unrestricted SSR is

$$\begin{aligned} SSR(\mu) &= \sum_{t=1}^T v_{jit}^2 + (\hat{\lambda}_{ji}^{(1)} - \lambda_{ji})^2 \sum_1 f_{jt}^2 + (\hat{\lambda}_{ji}^{(2)} - \lambda_{ji})^2 \sum_2 f_{jt}^2 \\ &\quad + \hat{\lambda}_{ji}^{(1)2} \sum_1 (\hat{f}_{jt} - f_{jt})^2 + \hat{\lambda}_{ji}^{(2)2} \sum_2 (\hat{f}_{jt} - f_{jt})^2 \\ &\quad - 2\hat{\lambda}_{ji}^{(1)} \sum_1 (\hat{f}_{jt} - f_{jt})v_{jit} - 2\hat{\lambda}_{ji}^{(2)} \sum_2 (\hat{f}_{jt} - f_{jt})v_{jit} \\ &\quad - 2(\hat{\lambda}_{ji}^{(1)} - \lambda_{ji}) \sum_1 f_{jt}v_{jit} - 2(\hat{\lambda}_{ji}^{(2)} - \lambda_{ji}) \sum_2 f_{jt}v_{jit} \\ &\quad + 2\hat{\lambda}_{ji}^{(1)}(\hat{\lambda}_{ji}^{(1)} - \lambda_{ji}) \sum_1 f_{jt}(\hat{f}_{jt} - f_{jt}) + 2\hat{\lambda}_{ji}^{(2)}(\hat{\lambda}_{ji}^{(2)} - \lambda_{ji}) \sum_2 f_{jt}(\hat{f}_{jt} - f_{jt}). \end{aligned}$$

Note that

$$\begin{aligned}\hat{\lambda}_{ji} - \lambda_{ji} &= T^{-1} \sum_{t=1}^T f_{jt} v_{jit} + O_p(c^{-2}), \\ \hat{\lambda}_{ji}^{(s)} - \lambda_{ji} &= m_1^{-1} T^{-1} \sum_s f_{jt} v_{jit} + O_p(c^{-2}),\end{aligned}$$

where  $m_s = T^{-1} \sum_s \hat{f}_{jt}^2$  for  $s = 1, 2$ . Then, we have

$$\begin{aligned}SSR_r - SSR(\mu) &= (\hat{\lambda}_{ji} - \lambda_{ji})^2 \sum_{t=1}^T f_{jt}^2 - 2(\hat{\lambda}_{ji} - \lambda_{ji}) \sum_{t=1}^T f_{jt} v_{jit} \\ &\quad - (\hat{\lambda}_{ji}^{(1)} - \lambda_{ji})^2 \sum_1 f_{jt}^2 + 2(\hat{\lambda}_{ji}^{(1)} - \lambda_{ji}) \sum_1 f_{jt} v_{jit} \\ &\quad - (\hat{\lambda}_{ji}^{(2)} - \lambda_{ji})^2 \sum_2 f_{jt}^2 + 2(\hat{\lambda}_{ji}^{(2)} - \lambda_{ji}) \sum_2 f_{jt} v_{jit} + o_p(1) \\ &= A_1 - 2A_2 - B_1 + 2B_2 - C_1 + 2C_2 + o_p(1).\end{aligned}$$

Here, we see

$$\begin{aligned}A_1 &= \left( T^{-1/2} \sum_{t=1}^T f_{jt} v_{jit} \right)^2 T^{-1} \sum_{t=1}^T f_{jt}^2 + O_p(T^{1/2} c^{-2}), \\ &\Rightarrow \sigma_{vji}^2 W^2(1),\end{aligned}$$

$$\begin{aligned}A_2 &= \left( T^{-1/2} \sum_{t=1}^T f_{jt} v_{jit} \right)^2 + O_p(T^{1/2} c^{-2}) \\ &\Rightarrow \sigma_{vji}^2 W^2(1),\end{aligned}$$

$$\begin{aligned}B_1 &= m_1^{-2} \left( T^{-1/2} \sum_1 f_{jt} v_{jit} \right)^2 T^{-1} \sum_1 f_{jt}^2 + O_p(T^{1/2} c^{-2}), \\ &\Rightarrow \frac{1}{\mu} \sigma_{vji}^2 W^2(\mu),\end{aligned}$$

$$\begin{aligned}B_2 &= m_1^{-1} \left( T^{-1/2} \sum_1 f_{jt} v_{jit} \right)^2 + O_p(T^{1/2} c^{-2}), \\ &\Rightarrow \frac{1}{\mu} \sigma_{vji}^2 W^2(\mu),\end{aligned}$$

$$\begin{aligned}C_1 &= m_2^{-2} \left( T^{-1/2} \sum_2 f_{jt} v_{jit} \right)^2 T^{-1} \sum_2 f_{jt}^2 + O_p(T^{1/2} c^{-2}), \\ &\Rightarrow \frac{1}{1-\mu} \sigma_{vji}^2 \{W(1) - W(\mu)\}^2,\end{aligned}$$

$$\begin{aligned}C_2 &= m_2^{-1} \left( T^{-1/2} \sum_2 f_{jt} v_{jit} \right)^2 + O_p(T^{1/2} c^{-2}), \\ &\Rightarrow \frac{1}{1-\mu} \sigma_{vji}^2 \{W(1) - W(\mu)\}^2,\end{aligned}$$

by Assumption 5c) and the facts that  $m_1 \xrightarrow{p} \mu$  and  $m_2 \xrightarrow{p} 1 - \mu$ . It can be deduced that

$$\begin{aligned}SSR_r - SSR(\mu) &\Rightarrow \sigma_{vji}^2 \left[ -W(1)^2 + \frac{1}{\mu} W(\mu)^2 + \frac{1}{1-\mu} \{W(1) - W(\mu)\}^2 \right], \\ &= \sigma_{vji}^2 \frac{\{\mu W(1) - W(\mu)\}^2}{\mu(1-\mu)}.\end{aligned}$$

On the other hand, the denominator is

$$T^{-1}SSR(\mu) = T^{-1} \sum_{t=1}^T v_{jit}^2 + o_p(1).$$

It is straightforward that  $T^{-1} \sum_{t=1}^T v_{jit}^2 \xrightarrow{p} \sigma_{vji}^2$  by Assumption 5b). Therefore,

$$W_T(\mu; j) \Rightarrow \frac{\{\mu W(1) - W(\mu)\}^2}{\mu(1 - \mu)}.$$

Second,

$$\begin{aligned} E(f_{jt}f_{lt}v_{jit}v_{lit}) &= E \left[ f_{jt}f_{lt}(e_{it} + \sum_{k \neq j} \lambda_{ki}f_{kt})(e_{it} + \sum_{h \neq l} \lambda_{hi}f_{ht}) \right], \\ &= E(f_{jt}f_{lt}e_{it}^2) + E \left( f_{jt}f_{lt}e_{it}(\sum_{h \neq l} \lambda_{hi}f_{ht}) \right) \\ &\quad + E \left( f_{jt}f_{lt}e_{it}(\sum_{k \neq j} \lambda_{ki}f_{kt}) \right) \\ &\quad + E \left( f_{jt}f_{lt}(\sum_{h \neq l} \lambda_{hi}f_{ht})(\sum_{k \neq j} \lambda_{ki}f_{kt}) \right), \\ &= E(f_{jt}f_{lt})E(e_{it}^2) + E \left( f_{jt}f_{lt}(\sum_{h \neq l} \lambda_{hi}f_{ht}) \right) E(e_{it}) \\ &\quad + E \left( f_{jt}f_{lt}(\sum_{k \neq j} \lambda_{ki}f_{kt}) \right) E(e_{it}) \\ &\quad + E \left( f_{jt}f_{lt}(\sum_{h \neq l} \lambda_{hi}f_{ht})(\sum_{k \neq j} \lambda_{ki}f_{kt}) \right), \\ &= 0 + 0 + 0 + E(\lambda_j)E(\lambda_l)E(f_{jt}^2f_{lt}^2), \\ &= 0, \end{aligned}$$

for  $j \neq l$ . The third and fourth equalities follow Assumption 2 (independency of  $F_t$ ,  $\lambda_i$ , and  $e_{it}$ ) and the last equality is from Assumption 4 ( $E(\lambda_i) = 0$ ). Hence,  $W_T(\mu_j; j)$  and  $W_T(\mu_l; l)$  are uncorrelated for  $j \neq l$ . This implies that the limiting random variables  $B_j$  and  $B_l$  for  $j \neq l$  are independent as they are Gaussian processes. Therefore, we finally obtain

$$MBE_T \Rightarrow \max_{1 \leq j \leq r} \sup_{\mu \in \Lambda_\varepsilon} \frac{\{\mu W_j(1) - W_j(\mu)\}^2}{\mu(1 - \mu)}.$$

**Proof of Theorem 3 (ii):** Without loss of generality, we consider the case of  $m = 1$  but can extend the result for multiple breaks. The model under the alternative hypothesis is

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

where  $\lambda_{it} = \lambda_i^{(1)}$  for  $t \leq \lfloor \mu_0 T \rfloor$  and  $\lambda_{it} = \lambda_i^{(2)}$  otherwise. Consider the regression of  $x_{it}$  on the  $j$ th principal component. Without loss of generality, choose  $j$  such that

$$j = \arg \max_{1 \leq l \leq r} \left| \lambda_{li}^{(2)} - \lambda_{li}^{(1)} \right|.$$

Under the alternative, we know that there is at least one  $j$  that satisfies  $\lambda_{ji}^{(2)} - \lambda_{ji}^{(1)} \neq 0$ . To derive additional implications, we also denote the break magnitude of  $\lambda_{ji}$  by  $\delta^*$  and the magnitude of the rest of the loadings by  $\delta$ .

Now the model is rewritten as

$$x_{it} = \lambda_{jit}f_{jt} + \sum_{l \neq j} \lambda_{lit}f_{lt} + e_{it}.$$

Let  $h_{jit} = \sum_{l \neq j} \lambda_{lit}f_{lt}$  and  $\xi_{jt} = f_{jt} - \hat{f}_{jt}$ . Then using (A.4) and (A.6)

$$\begin{aligned} \sum_{t=1}^T \xi_{jt}h_{jit} &= \sum_{l \neq j} \sum_{t=1}^T \lambda_{lit}\xi_{jt}\hat{f}_{lt} + \sum_{l \neq j} \sum_{t=1}^T \lambda_{lit}\xi_{jt}\xi_{lt} \\ &\leq \sum_{l \neq j} \sum_{t=1}^T \lambda_{lit}\xi_{jt}\hat{f}_{lt} + \sum_{l \neq j} \left( \lambda_{lit} \sum_{t=1}^T \xi_{jt}^2 \right)^{1/2} \left( \lambda_{lit} \sum_{t=1}^T \xi_{lt}^2 \right)^{1/2} \\ &= O_p(Tc^{-2}\delta). \end{aligned}$$

We also have

$$\begin{aligned} \sum_{t=1}^T \hat{f}_{jt}h_{jit} &= \sum_{l \neq j} \sum_{t=1}^T \lambda_{lit}\hat{f}_{jt}f_{lt} \\ &= \sum_{l \neq j} \sum_{t=1}^T \lambda_{lit}\hat{f}_{jt}\hat{f}_{lt} + \sum_{l \neq j} \sum_{t=1}^T \lambda_{lit}\hat{f}_{jt}\xi_{lt} \\ &= O_p(Tc^{-2}\delta). \end{aligned}$$

Finally,

$$\sum_{t=1}^T e_{it}h_{jit} = \sum_{l \neq j} \sum_{t=1}^T \lambda_{lit}e_{it}f_{lt} = O_p(T^{1/2}\delta).$$

When the summation is over a subsample, we can similarly show that  $\sum_s \xi_{jt}h_{jit} = O_p(Tc^{-2})$ ,  $\sum_s \hat{f}_{jt}h_{jit} = O_p(Tc^{-2})$  and  $\sum_s e_{it}h_{jit} = O_p(T^{1/2})$  for  $s = 1, 2$ .

The restricted residual is

$$\hat{e}_{it} = e_{it} + \lambda_{jit}\xi_{jt} + \lambda_{jit}\hat{f}_{jt} - \hat{\lambda}_{ji}\hat{f}_{jt} + h_{jit},$$

where

$$\begin{aligned} \hat{\lambda}_{ji} &= \frac{T^{-1} \sum_{t=1}^T \hat{f}_{jt}x_{it}}{T^{-1} \sum_{t=1}^T \hat{f}_{jt}^2} = T^{-1} \sum_{t=1}^T \hat{f}_{jt}x_{it} \\ &= T^{-1} \sum_{t=1}^T \hat{f}_{jt}f_{jt}\lambda_{jit} + T^{-1} \sum_{t=1}^T \hat{f}_{jt}h_{jit} + T^{-1} \sum_{t=1}^T \hat{f}_{jt}e_{it} \\ &= \underbrace{T^{-1} \sum_{t=1}^T \hat{f}_{jt}^2 \lambda_{jit}}_{=O_p(\delta^*)} + \underbrace{T^{-1} \sum_{t=1}^T \hat{f}_{jt}\xi_{jt}\lambda_{jit}}_{=O_p(c^{-2}\delta^*)} + \underbrace{T^{-1} \sum_{t=1}^T \hat{f}_{jt}h_{jit}}_{=O_p(c^{-2}\delta)} + \underbrace{T^{-1} \sum_{t=1}^T \hat{f}_{jt}e_{it}}_{=O_p(T^{-1/2})} \\ &= O_p(\delta^*). \end{aligned}$$

The restricted SSR is

$$\begin{aligned}
SSR_r &= \underbrace{\sum_{t=1}^T e_{it}^2}_{=O_p(T)} + \underbrace{\sum_{t=1}^T \lambda_{jit}^2 \xi_{jt}^2}_{=O_p(Tc^{-2}\delta^{*2})} + \underbrace{\sum_{t=1}^T \lambda_{jit}^2 \hat{f}_{jt}^2}_{=O_p(T\delta^{*2})} + \underbrace{\hat{\lambda}_{ji}^2 \sum_{t=1}^T \hat{f}_{jt}^2}_{=O_p(T\delta^{*2})} + \underbrace{\sum_{t=1}^T h_{jit}^2}_{=O_p(T\delta^2)} \\
&+ 2 \underbrace{\sum_{t=1}^T \lambda_{jit} \xi_{jt} e_{it}}_{=O_p(Tc^{-2}\delta^*)} + 2 \underbrace{\sum_{t=1}^T \lambda_{jit} \hat{f}_{jt} e_{it}}_{=O_p(T^{1/2}\delta^*)} - 2 \underbrace{\hat{\lambda}_{ji} \sum_{t=1}^T \hat{f}_{jt} e_{it}}_{=O_p(T^{1/2}\delta^*)} \\
&+ 2 \underbrace{\sum_{t=1}^T e_{it} h_{jit}}_{=O_p(T^{1/2}\delta)} + 2 \underbrace{\sum_{t=1}^T \lambda_{jit}^2 \xi_{jt} \hat{f}_{jt}}_{=O_p(Tc^{-2}\delta^{*2})} + 2 \underbrace{\sum_{t=1}^T \lambda_{jit} \xi_{jt} h_{jit}}_{=O_p(Tc^{-2}\delta^{*2})} + 2 \underbrace{\sum_{t=1}^T \lambda_{jit} \hat{f}_{jt} h_{jit}}_{=O_p(T\delta^{*2})} \\
&- 2 \underbrace{\hat{\lambda}_{ji} \sum_{t=1}^T \lambda_{jit} \xi_{jt} \hat{f}_{jt}}_{=O_p(Tc^{-2}\delta^{*2})} - 2 \underbrace{\hat{\lambda}_{ji} \sum_{t=1}^T \lambda_{jit} \hat{f}_{jt}^2}_{=O_p(T\delta^{*2})} - 2 \underbrace{\hat{\lambda}_{ji} \sum_{t=1}^T \hat{f}_{jt} h_{jit}}_{=O_p(\sqrt{T}\delta^2)} \\
&= O_p(T\delta^{*2}).
\end{aligned}$$

To consider the unrestricted residuals, since the convergence results are all uniform in  $\mu$  as we saw in Proof of Theorem 1(ii), we only consider  $\mu = \mu_0$ .

$$\hat{e}(\mu_0)_{it} = e_{it} + \lambda_{jit} \xi_{jt} + \lambda_{jit} \hat{f}_{jt} - \hat{\lambda}_{ji}^{(1)} \hat{f}_{jt}^{(1)} - \hat{\lambda}_{ji}^{(2)} \hat{f}_{jt}^{(2)} + h_{jit},$$

where  $\hat{\lambda}_{ji}^{(s)} = \left( T^{-1} \sum_s \hat{f}_{jt}^2 \right)^{-1} T^{-1} \sum_s \hat{f}_{jt} x_{it}$  for  $s = 1, 2$ . Note that  $\hat{\lambda}_{ij}^{(1)} - \lambda_{ji}^{(1)} = O_p(T^{-1/2})$  since

$$\begin{aligned}
T^{-1} \sum_1 \hat{f}_{jt} x_{it} &= \left( T^{-1} \sum_s \hat{f}_{jt}^2 \right) \lambda_{ji}^{(1)} + \underbrace{T^{-1} \lambda_{ji}^{(1)} \sum_1 \hat{f}_{jt} \xi_{jt}}_{=O_p(c^{-2})} \\
&+ \underbrace{T^{-1} \sum_1 \hat{f}_{jt} h_{jit}}_{=O_p(c^{-2})} + \underbrace{T^{-1} \sum_1 \hat{f}_{jt} e_{it}}_{=O_p(T^{-1/2})} \\
&= O_p(1)
\end{aligned}$$

and  $0 < T^{-1} \sum_1 \hat{f}_{jt}^2 < T^{-1} \sum_{t=1}^T \hat{f}_{jt}^2 = 1$ . Similarly,  $\hat{\lambda}_{ji}^{(2)} - \lambda_{ji}^{(2)} = O_p(T^{-1/2})$ .

The unrestricted SSR evaluated at  $\mu_0$  is

$$\begin{aligned}
SSR(\mu_0) &= \underbrace{\sum_{t=1}^T e_{it}^2}_{=O_p(T)} + \underbrace{\sum_{t=1}^T \lambda_{jit}^2 \xi_{jt}^2}_{=O_p(Tc^{-2}\delta^{*2})} + \underbrace{(\lambda_{ji}^{(1)} - \hat{\lambda}_{ji}^{(1)})^2 \sum_1 \hat{f}_{jt}^2}_{=O_p(1)} + \underbrace{(\lambda_{ji}^{(2)} - \hat{\lambda}_{ji}^{(2)})^2 \sum_2 \hat{f}_{jt}^2}_{=O_p(1)} \\
&+ \underbrace{\sum_{t=1}^T h_{jit}^2}_{=O_p(T\delta^2)} + \underbrace{2\sum_{t=1}^T \lambda_{jit} \xi_{jt} e_{it}}_{=O_p(Tc^{-2}\delta^*)} + \underbrace{2\sum_{t=1}^T e_{it} h_{jit}}_{=O_p(T^{1/2}\delta)} + \underbrace{2\sum_{t=1}^T \lambda_{jit} \xi_{jt} h_{jit}}_{=O_p(Tc^{-2}\delta^*\delta)} \\
&+ 2\underbrace{(\hat{\lambda}_{ji}^{(1)} - \lambda_{ji}^{(1)}) \sum_1 \hat{f}_{jt} e_{it}}_{=O_p(1)} + 2\underbrace{(\hat{\lambda}_{ji}^{(2)} - \lambda_{ji}^{(2)}) \sum_2 \hat{f}_{jt} e_{it}}_{=O_p(1)} \\
&+ 2\underbrace{\lambda_{ji}^{(1)} (\hat{\lambda}_{ji}^{(1)} - \lambda_{ji}^{(1)}) \sum_1 \xi_{jt} \hat{f}_{jt}}_{=O_p(T^{1/2}c^{-2})} + 2\underbrace{\lambda_{ji}^{(2)} (\hat{\lambda}_{ji}^{(2)} - \lambda_{ji}^{(2)}) \sum_2 \xi_{jt} \hat{f}_{jt}}_{=O_p(T^{1/2}c^{-2})} \\
&+ 2\underbrace{(\hat{\lambda}_{ji}^{(1)} - \lambda_{ji}^{(1)}) \sum_1 \hat{f}_{jt} h_{jit}}_{=O_p(1)} + 2\underbrace{(\hat{\lambda}_{ji}^{(2)} - \lambda_{ji}^{(2)}) \sum_2 \hat{f}_{jt} h_{jit}}_{=O_p(1)}.
\end{aligned}$$

Hence,  $SSR(\mu_0) = O_p(T\delta^2)$ . Also,

$$\begin{aligned}
SSR_r - SSR(\mu_0) &= \underbrace{\hat{\lambda}_{ji}^2 \sum_{t=1}^T \hat{f}_{jt}^2}_{=O_p(T\delta^{*2})} - \underbrace{\hat{\lambda}_{ji}^{(1)2} \sum_1 \hat{f}_{jt}^2}_{=O_p(1)} - \hat{\lambda}_{ji}^{(2)2} \sum_2 \hat{f}_{jt}^2 \\
&+ 2\underbrace{\lambda_{ji}^{(1)} \hat{\lambda}_{ji}^{(1)} \sum_1 \hat{f}_{jt}^2}_{=O_p(T)} + 2\underbrace{\lambda_{ji}^{(2)} \hat{\lambda}_{ji}^{(2)} \sum_2 \hat{f}_{jt}^2}_{=O_p(T)} \\
&- 2\underbrace{\hat{\lambda}_{ji} \sum_{t=1}^T \hat{f}_{jt} e_{it}}_{=O_p(T^{1/2}\delta^*)} + 2\underbrace{\hat{\lambda}_{ij}^{(1)} \sum_1 \hat{f}_{jt} e_{it}}_{=O_p(T^{1/2})} + 2\underbrace{\hat{\lambda}_{ji}^{(2)} \sum_2 \hat{f}_{jt} e_{it}}_{=O_p(T^{1/2})} \\
&- 2\underbrace{\hat{\lambda}_{ji} \sum_{t=1}^T \lambda_{jit} \xi_{jt} \hat{f}_{jt}}_{=O_p(Tc^{-2}\delta^{*2})} + 2\underbrace{\hat{\lambda}_{ji}^{(1)} \sum_1 \lambda_{jit} \xi_{jt} \hat{f}_{jt}}_{=O_p(Tc^{-2}\delta^*)} + 2\underbrace{\hat{\lambda}_{ji}^{(2)} \sum_2 \lambda_{jit} \xi_{jt} \hat{f}_{jt}}_{=O_p(Tc^{-2}\delta^*)} \\
&- 2\underbrace{\hat{\lambda}_{ji} \sum_{t=1}^T \lambda_{jit} \hat{f}_{jt}^2}_{=O_p(T\delta^{*2})} + 2\underbrace{\hat{\lambda}_{ji}^{(1)} \sum_1 \lambda_{jit} \hat{f}_{jt}^2}_{=O_p(T)} + 2\underbrace{\hat{\lambda}_{ji}^{(2)} \sum_2 \lambda_{jit} \hat{f}_{jt}^2}_{=O_p(T)} \\
&- 2\underbrace{\hat{\lambda}_{ji} \sum_{t=1}^T \hat{f}_{jt} h_{jit}}_{=O_p(Tc^{-2}\delta^*\delta)} + 2\underbrace{\hat{\lambda}_{ji}^{(1)} \sum_1 \hat{f}_{jt} h_{jit}}_{=O_p(T\delta)} + 2\underbrace{\hat{\lambda}_{ji}^{(2)} \sum_2 \hat{f}_{jt} h_{jit}}_{=O_p(T\delta)}.
\end{aligned}$$

Hence,  $SSR_r - SSR(\mu) = O_p(T\delta^{*2})$  and  $MBE = O_p(T\delta^{*2}\delta^{-2})$ .

## References

- Ahn, S.C., and A.R.Horenstein, 2013, Eigenvalue ratio test for the number of factors, *Econometrica* 81, 1203-1227.
- Amengual, D. and M.W.Watson, 2006, Consistent estimation of the number of dynamic factors in a large  $N$  and  $T$  panel, *Journal of Business and Economic Statistics* 25, 91-96.
- Andrews, D.W.K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, 817-858.
- Andrews, D.W.K., 1993, Tests for parameter instability and structural change with unknown change point, *Econometrica* 61, 821-856.
- Andrews, D.W.K. and W. Ploberger, 1994, Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica* 62, 1383-1414.
- Bai, J., 2003, Inferential theory for factor models of large dimensions, *Econometrica* 71, 135-171.
- Bai, J. 2010, Common breaks in means and variances for panel data, *Journal of Econometrics* 157, 78-92.
- Bai, J. and S. Ng, 2002, Determining the number of factors in approximate factor models, *Econometrica* 70, 191-221.
- Bai, J. and S. Ng, 2007, Determining the number of primitive shocks in factor models, *Journal of Business and Economic Statistics* 25, 52-60.
- Bai, J. and P. Perron, 1998, Estimating and testing linear models with multiple structural changes, *Econometrica* 66, 47-78.
- Bates, B.J., M. Plagborg-Moller, J.H. Stock, and M.W.Watson, 2013, Consistent factor estimation in dynamic factor models with structural instability, forthcoming in *Journal of Econometrics*.
- Breitung, J. and S. Eickmeier, 2011, Testing for structural breaks in dynamic factor models, *Journal of Econometrics* 163, 71-84.
- Chen, L., J.J.Dolado. and J. Gonzalo, 2012, Detecting big structural breaks in large factor models, Unpublished manuscript.
- Crainiceanu, C.M. and T.J. Vogelsang, 2007, Nonmonotonic power for tests of mean shift in a time series, *Journal of Statistical Computation and Simulation* 77, 457-476.

- Deng, A. and P. Perron, 2006, A non-local perspective on the power properties of the CUSUM and CUSUM of squares tests for structural change, *Journal of Econometrics* 142, 212-240.
- Elliott, G. and U.K. Müller, 2006, Efficient tests for general persistent time variation in regression coefficients, *Review of Economic Studies* 73, 907-940.
- Gürkaynak, R. S., S. Brian, J.H. Wright, 2007, The U.S. treasury yield curve: 1961 to the present, *Journal of Monetary Economics* 54, 2291-2304.
- Hall, A.R., 2000, Covariance matrix estimation and the power of the overidentifying restriction test, *Econometrica* 68, 1517-1527.
- Han, X. and A. Inoue, 2013, Tests for parameter instability in dynamic factor models, Unpublished Manuscript, North Carolina State University.
- Kejriwal, M., 2009, Tests for a mean shift with good size and monotonic power, *Economic Letters* 102, 78-82.
- Kim, D. and P. Perron, 2009, Assessing the relative power of structural break tests using a framework based on the approximate Bahadur slope, *Journal of Econometrics* 149, 26-51.
- Nelson, F.D. and N.E. Savin, 1990, The danger of extrapolating asymptotic local power, *Econometrica* 58, 977-981.
- Onatski, A., 2010, Determining the number of factors from empirical distribution of eigenvalues, *Review of Economics and Statistics* 92, 1004-1016.
- Perron, P., 1990, Testing for a unit root in a time series with a changing mean, *Journal of Business and Economic Statistics* 8, 153-162.
- Perron, P., 1991, A test for changes in a polynomial trend function for a dynamic time series, Econometric Research Program Research Memorandum No. 363, Princeton University.
- Perron, P., 2006, Dealing with structural breaks, Patterson, K. Ed. Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory. Palgrave Macmillan pp. 278-352.
- Perron P. and Y. Yamamoto, 2013, On the usefulness or lack thereof of optimality criteria in structural change tests, forthcoming in *Econometric Reviews*.
- Stock, J.H. and M.W. Watson, 2009, Forecasting in dynamic factor models subject to structural instability, In *The methodology and practice of econometrics: Festschrift in honor of D. F. Hendry*, (Ed.) N. Shephard and J. Castle, Oxford: Oxford University Press.
- Vogelsang, T. J., 1999, Sources of nonmonotonic power when testing for a shift in the mean of a dynamic time series, *Journal of Econometrics* 88, 283-299.

Table 1: Critical values of the MBE test

eps=0.05											
	maxr	1	2	3	4	5	6	7	8	9	10
quantile	0.9	8.27	9.82	10.64	11.31	11.77	12.22	12.34	12.83	12.92	13.24
	0.95	9.91	11.48	12.27	12.80	13.31	13.75	13.84	14.26	14.45	14.83
	0.975	11.57	12.96	13.70	14.36	14.85	15.17	15.36	15.76	16.08	16.25
	0.99	13.69	14.97	15.96	16.42	16.83	16.80	17.56	17.66	17.82	18.09
	maxr	11	12	13	14	15	16	17	18	19	20
quantile	0.9	13.43	13.53	13.77	13.99	14.03	14.22	14.32	14.45	14.48	14.75
	0.95	15.01	15.04	15.21	15.48	15.59	15.77	15.71	15.88	16.16	16.32
	0.975	16.51	16.56	16.72	17.01	17.10	17.17	17.34	17.50	17.83	17.86
	0.99	18.34	18.56	18.65	18.81	19.06	18.97	19.27	19.46	19.91	19.84

eps=0.1											
	maxr	1	2	3	4	5	6	7	8	9	10
quantile	0.9	7.68	9.25	10.11	10.70	11.06	11.57	11.84	12.25	12.36	12.62
	0.95	9.24	10.94	11.79	12.34	12.67	13.21	13.30	13.81	13.75	14.20
	0.975	10.94	12.51	13.39	13.76	14.35	14.59	14.81	15.17	15.36	15.63
	0.99	13.23	14.35	15.41	15.76	16.26	16.48	16.66	17.06	17.35	17.59
	maxr	11	12	13	14	15	16	17	18	19	20
quantile	0.9	12.85	12.90	13.24	13.46	13.49	13.77	13.84	13.90	13.92	14.17
	0.95	14.43	14.47	14.66	15.02	15.02	15.31	15.32	15.44	15.60	15.74
	0.975	16.03	15.90	16.15	16.56	16.74	16.65	16.68	16.98	17.40	17.23
	0.99	17.81	18.14	18.05	18.50	18.53	18.44	18.78	18.71	19.35	19.24

eps=0.15											
	maxr	1	2	3	4	5	6	7	8	9	10
quantile	0.9	7.28	8.84	9.67	10.25	10.66	11.29	11.46	11.70	11.93	12.25
	0.95	8.78	10.44	11.28	11.88	12.11	12.88	12.96	13.23	13.45	13.82
	0.975	10.44	12.13	12.74	13.18	13.77	14.38	14.58	14.67	14.87	15.23
	0.99	12.81	13.91	15.01	15.23	15.67	16.47	16.74	16.44	16.90	17.11
	maxr	11	12	13	14	15	16	17	18	19	20
quantile	0.9	12.56	12.60	12.87	12.87	13.10	13.30	13.29	13.45	13.57	13.61
	0.95	14.15	14.11	14.28	14.46	14.71	14.87	14.76	15.02	15.16	15.07
	0.975	15.70	15.54	15.73	16.04	16.11	16.46	16.21	16.57	16.53	16.71
	0.99	17.86	17.31	17.80	17.95	18.07	18.33	18.26	18.33	18.54	18.41

Table 2: Number of common factors estimated by Bai and Ng's (2002) criteria under local breaks (mean of simulations)

a)  $IC_p1$

		2				3		4
		1	2	3	4	1	2	1
		4	6	8	10	6	9	8
$r$	$m$							
$r(m+1)$	$r(m+1)$							
T=100	N=100	2.34	2.63	2.94	3.21	3.38	3.67	4.38
200	200	2.08	2.48	2.89	3.10	3.08	3.52	4.09
500	500	2.00	2.09	2.80	3.16	3.00	3.10	4.00
200	100	2.00	2.06	2.29	2.58	3.00	3.07	4.00
500	100	2.00	2.00	2.00	2.00	3.00	3.00	4.00
100	200	2.00	2.07	2.30	2.61	3.00	3.09	4.00
100	500	2.00	2.00	2.00	2.00	3.00	3.00	4.00

b)  $IC_p2$

		2				3		4
		1	2	3	4	1	2	1
		4	6	8	10	6	9	8
$r$	$m$							
$r(m+1)$	$r(m+1)$							
T=100	N=100	2.06	2.24	2.49	2.73	3.07	3.23	4.05
200	200	2.01	2.11	2.49	2.80	3.00	3.12	4.00
500	500	2.00	2.00	2.26	2.87	3.00	3.00	4.00
200	100	2.00	2.01	2.11	2.31	3.00	3.01	4.00
500	100	2.00	2.00	2.00	2.00	3.00	3.00	4.00
100	200	2.00	2.02	2.12	2.33	3.00	3.02	4.00
100	500	2.00	2.00	2.00	2.00	3.00	3.00	4.00

c)  $IC_p3$

		2				3		4
		1	2	3	4	1	2	1
		4	6	8	10	6	9	8
$r$	$m$							
$r(m+1)$	$r(m+1)$							
T=100	N=100	3.84	5.04	5.74	6.26	5.66	7.45	7.49
200	200	3.95	4.43	4.69	5.03	5.85	6.55	7.68
500	500	3.99	4.00	4.01	4.20	5.97	6.00	7.93
200	100	2.51	2.91	3.23	3.55	3.61	4.04	4.71
500	100	2.00	2.00	2.00	2.01	3.00	3.00	4.00
100	200	2.51	2.88	3.22	3.52	3.61	4.03	4.70
100	500	2.00	2.00	2.00	2.02	3.00	3.00	4.00

Table 3: Number of common factors estimated by Bai and Ng's (2002) criteria under fixed breaks (mean of simulations)

a)  $IC_p1$

		2				3		4
		1	2	3	4	1	2	1
		4	6	8	10	6	9	8
$r$	$m$							
$r(m+1)$	$r(m+1)$							
T=100	N=100	3.98	6.45	7.89	9.86	5.20	8.80	6.39
200	200	4.64	5.08	7.88	9.88	6.38	7.25	8.04
500	500	4.84	6.00	8.66	10.13	6.95	9.00	8.99
200	100	4.02	5.01	7.82	9.83	5.42	7.02	6.72
500	100	4.16	5.01	7.66	9.66	5.73	7.03	7.25
100	200	4.15	6.50	7.91	9.91	5.52	8.89	6.87
100	500	4.40	6.59	7.94	9.92	5.93	8.99	7.42

b)  $IC_p2$

		2				3		4
		1	2	3	4	1	2	1
		4	6	8	10	6	9	8
$r$	$m$							
$r(m+1)$	$r(m+1)$							
T=100	N=100	3.87	6.39	7.85	9.76	4.98	8.74	6.04
200	200	4.20	5.01	7.86	9.85	5.67	7.03	7.09
500	500	4.81	5.98	8.26	9.87	6.94	8.94	8.98
200	100	3.86	5.00	7.80	9.80	5.11	7.00	6.28
500	100	4.01	5.00	7.64	9.65	5.48	7.01	6.92
100	200	4.01	6.45	7.90	9.88	5.23	8.83	6.44
100	500	4.28	6.56	7.93	9.92	5.77	8.94	7.19

c)  $IC_p3$

		2				3		4
		1	2	3	4	1	2	1
		4	6	8	10	6	9	8
$r$	$m$							
$r(m+1)$	$r(m+1)$							
T=100	N=100	4.96	7.52	8.60	10.53	6.98	10.71	8.97
200	200	4.95	6.00	8.86	10.63	6.99	8.99	9.00
500	500	4.92	6.00	8.93	10.93	6.98	9.00	9.00
200	100	4.82	5.39	8.00	9.92	6.79	7.85	8.65
500	100	4.59	5.13	7.71	9.72	6.53	7.32	8.32
100	200	4.85	6.99	8.09	9.99	6.74	9.81	8.51
100	500	4.73	6.78	7.99	9.95	6.51	9.39	8.18

Table 4: Rejection frequencies of the BE tests as functions of sample size

( $\delta = 5$ )

		local breaks			fixed breaks		
		SW	SLR	SLM	SW	SLR	SLM
T=50	N=50	0.298	0.228	0.170	0.175	0.092	0.034
100	100	0.387	0.358	0.319	0.164	0.123	0.077
300	300	0.542	0.530	0.519	0.090	0.076	0.055
500	500	0.556	0.553	0.550	0.096	0.086	0.079
20	500	0.140	0.076	0.018	0.521	0.205	0.004
50	500	0.123	0.082	0.044	0.185	0.095	0.033
100	500	0.180	0.151	0.134	0.126	0.084	0.046
500	20	0.425	0.421	0.417	0.319	0.311	0.304
500	50	0.515	0.512	0.503	0.295	0.291	0.281
500	100	0.537	0.527	0.520	0.214	0.205	0.197

( $\delta = 10$ )

		local breaks			fixed breaks		
		SW	SLR	SLM	SW	SLR	SLM
T=50	N=50	0.198	0.140	0.076	0.181	0.091	0.032
100	100	0.195	0.158	0.116	0.162	0.123	0.067
300	300	0.177	0.164	0.148	0.083	0.064	0.052
500	500	0.088	0.078	0.071	0.088	0.082	0.071
20	500	0.234	0.119	0.022	0.518	0.191	0.005
50	500	0.260	0.197	0.137	0.181	0.087	0.030
100	500	0.269	0.235	0.203	0.128	0.082	0.037
500	20	0.713	0.711	0.708	0.271	0.263	0.251
500	50	0.739	0.737	0.736	0.283	0.271	0.260
500	100	0.550	0.544	0.540	0.160	0.146	0.139

Figure 1: Power functions of the BE tests under common breaks  
 ( $T = 100$  and  $N = 100$ )

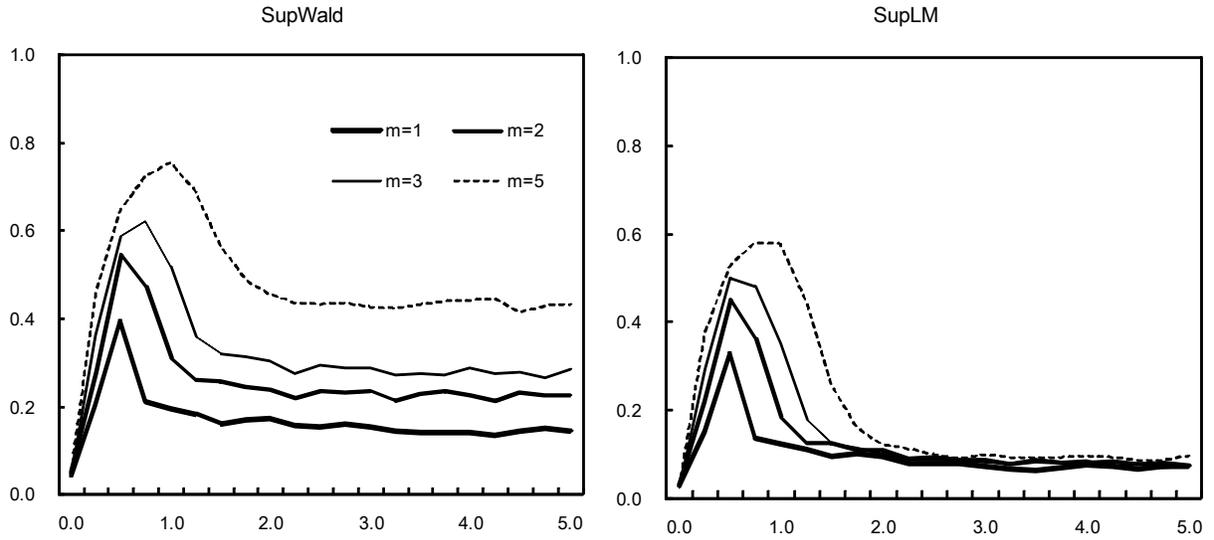


Figure 2: Power functions of the BE tests under common breaks  
 ( $T = 50$  and  $N = 200$ )

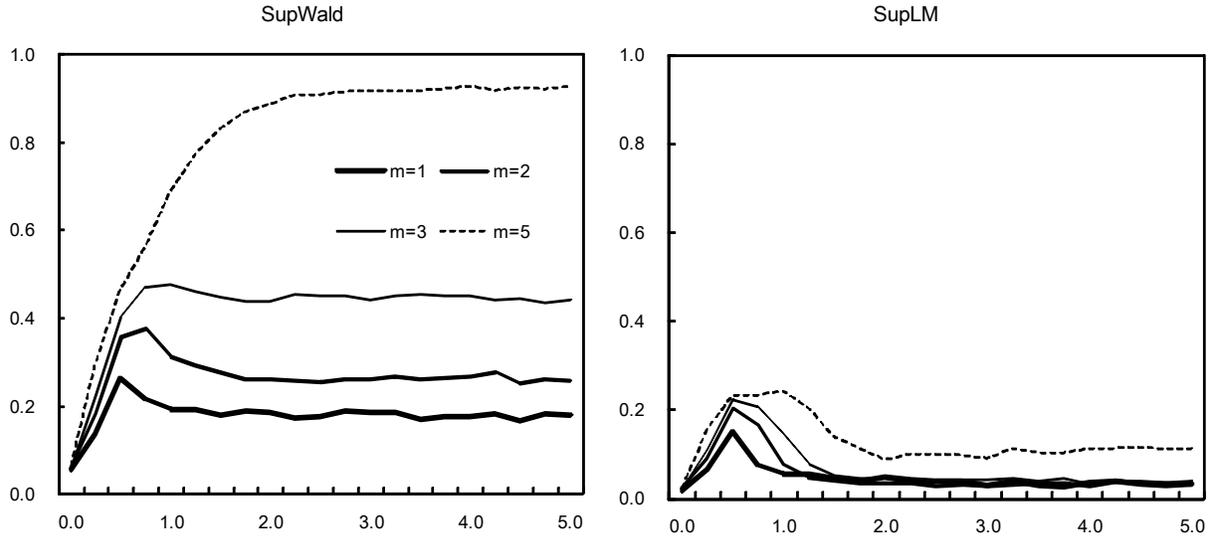


Figure 3: Power functions of the BE tests under common breaks  
 ( $T = 200$  and  $N = 50$ )

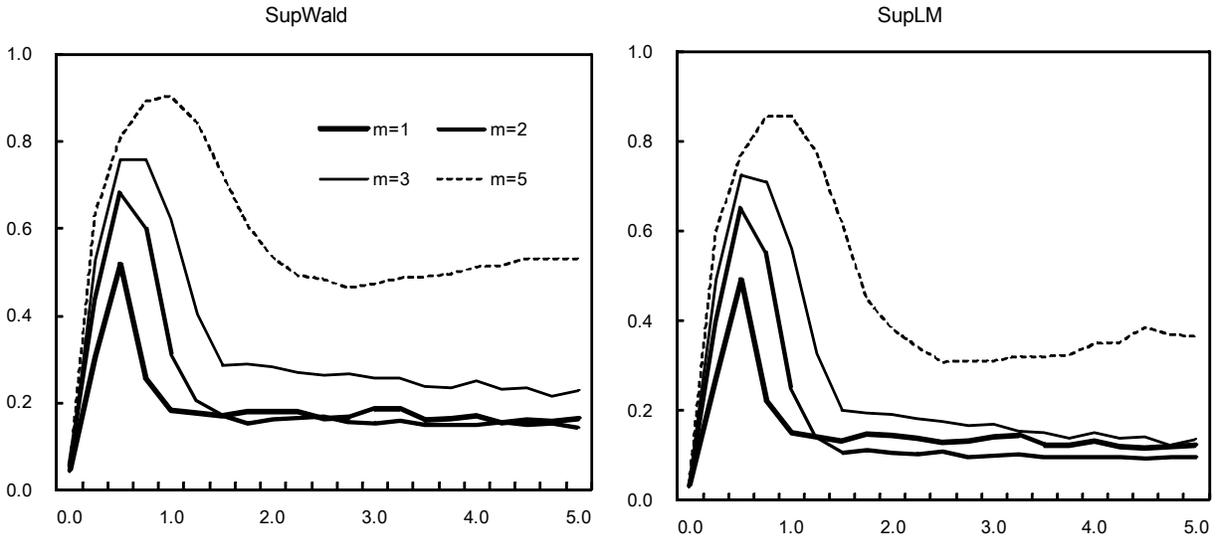


Figure 4: Power functions of the BE tests under common breaks  
 using a fixed number of factors ( $T = 100$  and  $N = 100$ )

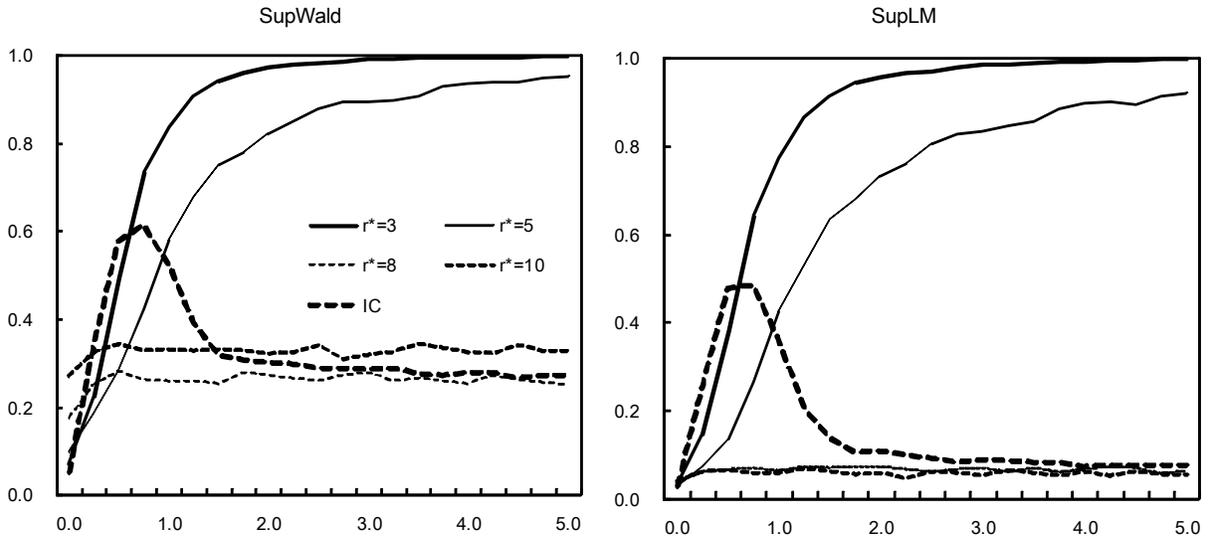


Figure 5: Power functions of the BE tests under common breaks  
 using HAC standard errors  
 ( $T = 100$  and  $N = 100$ )

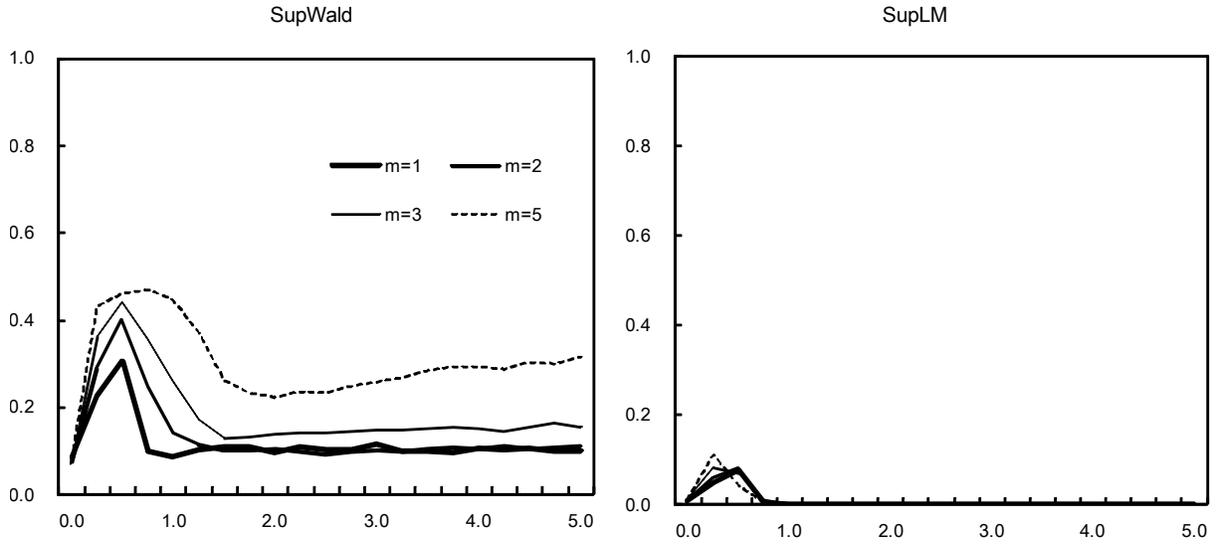


Figure 6: Power functions of the BE tests  
 under reduced rank random walk parameters  
 ( $T = 100$  and  $N = 100$ )

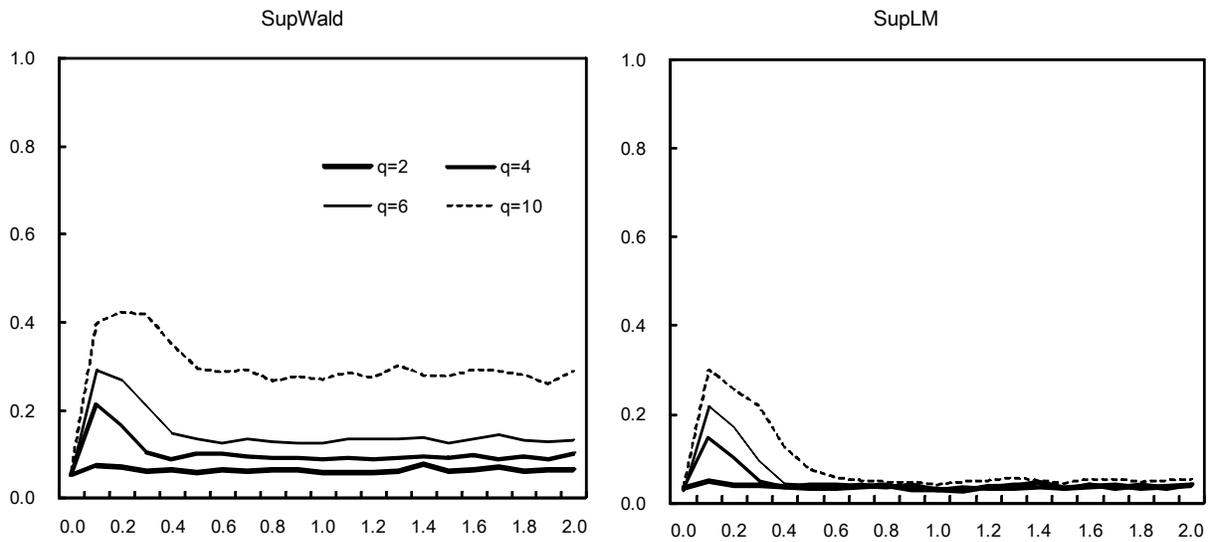


Table 5-1: Size of the MBE test ( $\lambda_{ji} \sim i.i.d.N(0, 1)$ )

		eps	r							
			1	2	3	4	5	6	7	8
i.i.d.	T=100 N=100	0.05	0.038	0.036	0.035	0.043	0.037	0.047	0.035	0.048
		0.10	0.047	0.048	0.043	0.051	0.048	0.039	0.050	0.043
		0.15	0.030	0.037	0.049	0.042	0.050	0.052	0.052	0.040
	T=50 N=200	0.05	0.037	0.042	0.045	0.053	0.049	0.049	0.047	0.051
		0.10	0.048	0.051	0.034	0.058	0.056	0.052	0.046	0.062
		0.15	0.040	0.043	0.051	0.053	0.047	0.044	0.051	0.055
	T=200 N=50	0.05	0.038	0.039	0.050	0.045	0.041	0.048	0.040	0.046
		0.10	0.051	0.050	0.036	0.048	0.040	0.034	0.047	0.047
		0.15	0.042	0.042	0.054	0.040	0.054	0.050	0.051	0.049
HAC	T=100 N=100	0.05	0.107	0.125	0.159	0.163	0.187	0.180	0.210	0.203
		0.10	0.084	0.137	0.131	0.195	0.175	0.158	0.159	0.194
		0.15	0.080	0.112	0.134	0.135	0.152	0.171	0.159	0.187
	T=50 N=200	0.05	0.105	0.161	0.208	0.243	0.265	0.294	0.306	0.341
		0.10	0.123	0.167	0.214	0.231	0.293	0.282	0.287	0.320
		0.15	0.120	0.151	0.200	0.228	0.238	0.255	0.268	0.290
	T=200 N=50	0.05	0.071	0.101	0.124	0.130	0.125	0.141	0.131	0.159
		0.10	0.068	0.080	0.106	0.124	0.123	0.113	0.106	0.131
		0.15	0.064	0.076	0.108	0.107	0.103	0.109	0.097	0.127

Table 5-2: Size of the MBE test ( $\lambda_{ji} \sim i.i.d.U[0, 1]$ )

		eps	r							
			1	2	3	4	5	6	7	8
i.i.d.	T=100 N=100	0.05	0.044	0.042	0.045	0.044	0.037	0.041	0.038	0.037
		0.10	0.036	0.04	0.052	0.045	0.049	0.041	0.043	0.043
		0.15	0.043	0.044	0.051	0.041	0.05	0.054	0.049	0.049
	T=50 N=200	0.05	0.039	0.049	0.044	0.035	0.037	0.031	0.041	0.038
		0.10	0.050	0.056	0.049	0.040	0.053	0.036	0.053	0.045
		0.15	0.040	0.045	0.063	0.051	0.046	0.041	0.053	0.047
	T=200 N=50	0.05	0.040	0.049	0.045	0.047	0.034	0.045	0.037	0.036
		0.10	0.053	0.055	0.041	0.042	0.042	0.046	0.041	0.039
		0.15	0.045	0.039	0.047	0.048	0.053	0.052	0.055	0.047
HAC	T=100 N=100	0.05	0.109	0.118	0.134	0.147	0.170	0.169	0.179	0.176
		0.10	0.083	0.113	0.123	0.119	0.154	0.164	0.153	0.160
		0.15	0.075	0.087	0.118	0.100	0.132	0.141	0.133	0.158
	T=50 N=200	0.05	0.132	0.189	0.214	0.211	0.233	0.243	0.264	0.259
		0.10	0.124	0.173	0.192	0.200	0.210	0.256	0.233	0.246
		0.15	0.098	0.151	0.171	0.200	0.191	0.221	0.198	0.232
	T=200 N=50	0.05	0.074	0.085	0.107	0.099	0.129	0.104	0.107	0.113
		0.10	0.060	0.082	0.087	0.092	0.110	0.100	0.090	0.107
		0.15	0.054	0.078	0.085	0.089	0.106	0.088	0.082	0.088

Figure 7: Power functions of the MBE test under common breaks

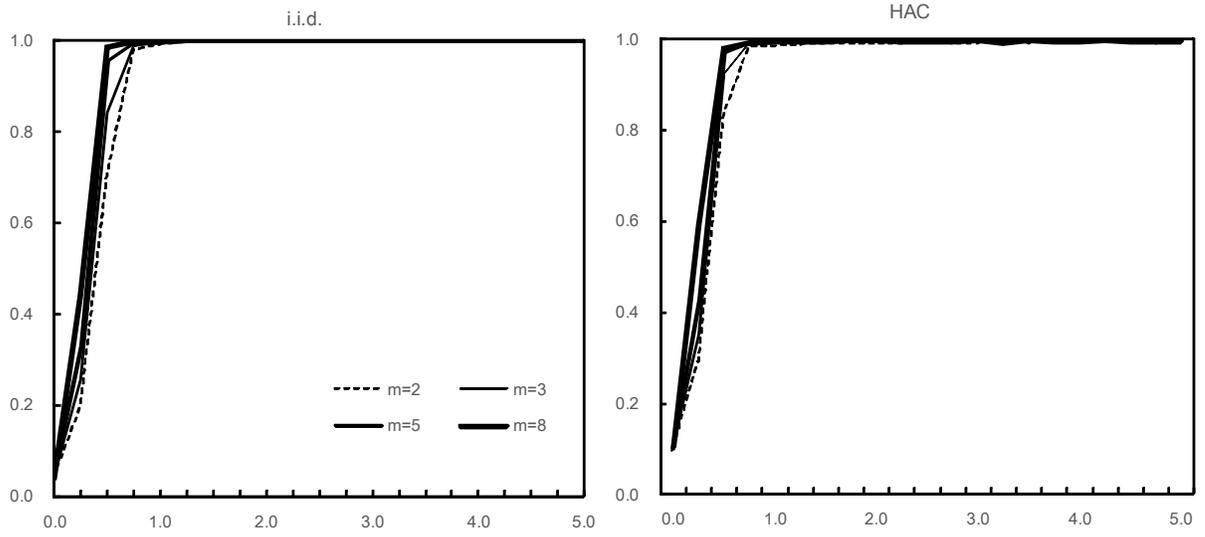


Figure 8: Power functions of the MBE test under reduced rank random walk parameters

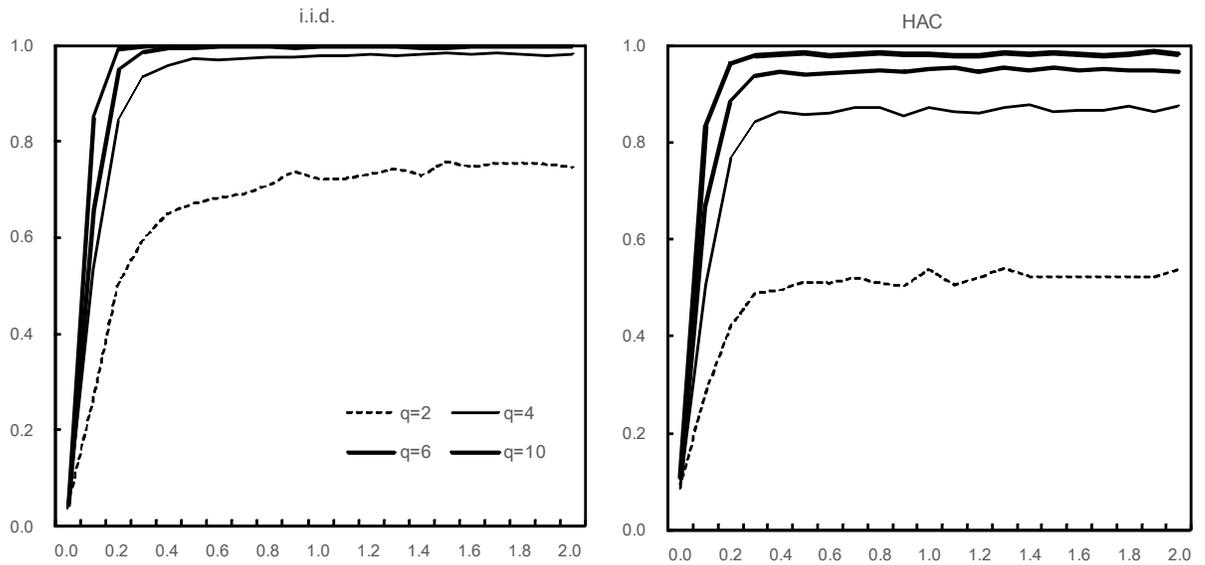


Figure 9: U.S. Treasury yields

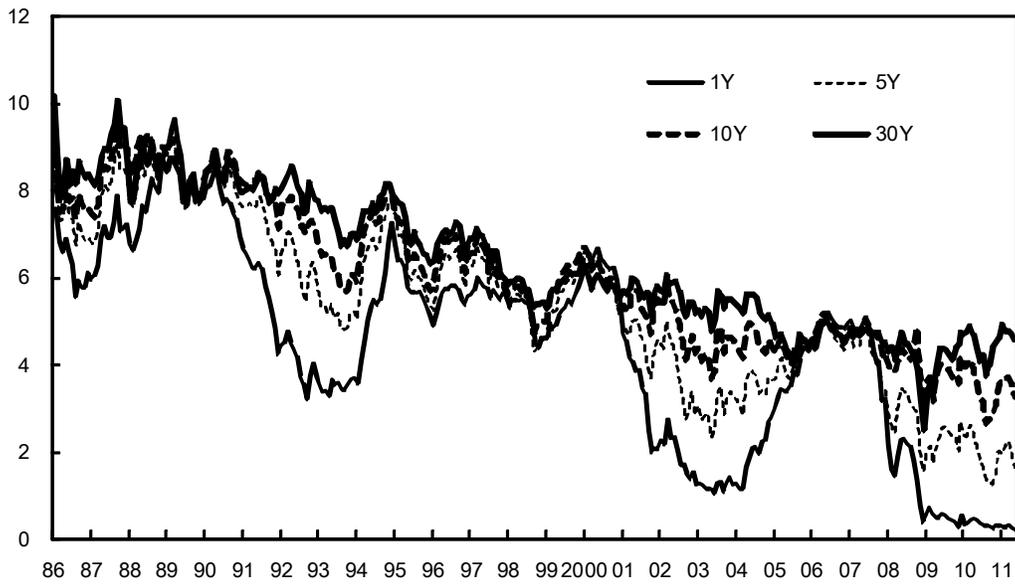
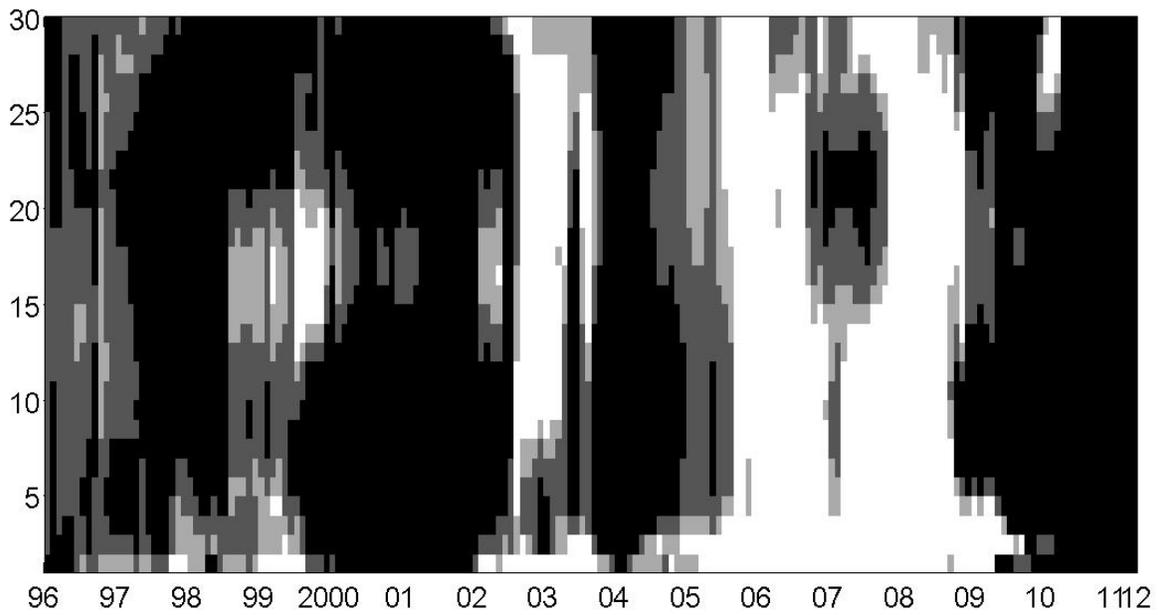
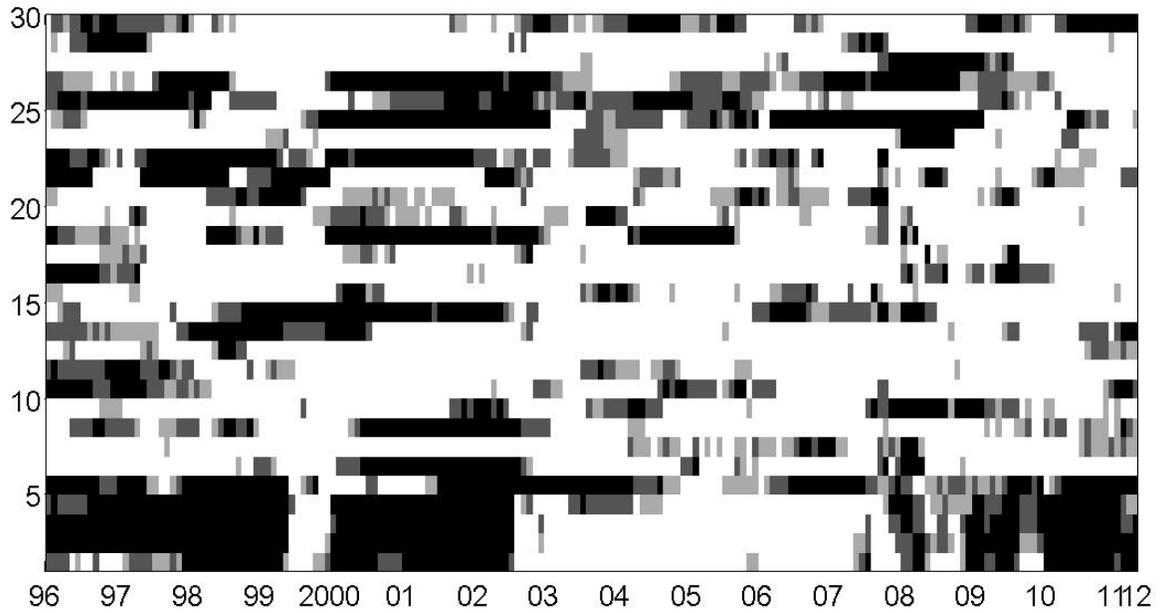


Figure 10: Results for the factor loading structural change tests in the U.S. Treasury yield data

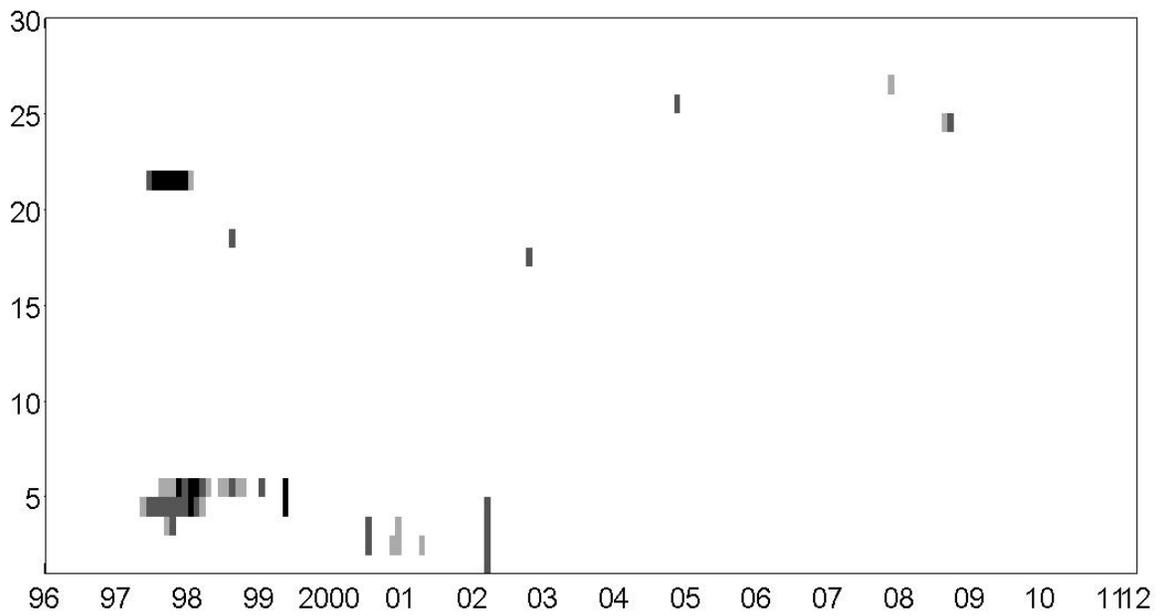
(MBE test)



(BE test : SupWald)



(BE test : SupLM)



- Notes: 1. The vertical and horizontal axes correspond to the maturity and the end of 10-year window, respectively.  
2. The maturities in black, dark gray, and light gray show significance at the 1%, 5%, and 10% levels, respectively.  
The maturities in white show insignificance even at the 10% level.