Detecting big structural breaks in large factor models

Chen, Liang and Dolado, Juan Jose and Gonzalo, Jesus

Universidad Carlos III de Madrid

8 June 2011

Detecting Big Structural Breaks in Large Factor Models*

Liang Chen† Juan J. Dolado‡ Jesús Gonzalo§

This Draft: June 2, 2011

Abstract

Constant factor loadings is a standard assumption in the analysis of large dimensional factor models. Yet, this assumption may be restrictive unless parameter shifts are mild. In this paper we develop a new testing procedure to detect *big* breaks in factor loadings at either known or unknown dates. It is based upon testing for structural breaks in a regression of the first of the $\bar{r}$ factors estimated by PC for the whole sample on the remaining $\bar{r}-1$ factors, where $\bar{r}$ is chosen using Bai and Ng´s (2002) information criteria. We argue that this test is more powerful than other tests available in the literature on this issue.

KEYWORDS: Structural break, large factor model, factor loadings, principal components.
JEL CODES: C12, C33.

*We are grateful to Soren Johansen, Hashem Pesaran and participants at the Conference in Honour of Sir David F. Hendry (St. Andrews) and the Workshop on High-Dimensional Econometric Modelling (Cass Business School). Financial support from the Spanish Ministerio de Ciencia e Innovación (grants SEJ2007-63098 and Consolider-2010) and Comunidad de Madrid (grant Excelecon) is gratefully acknowledged.

†Dept. of Economics, Universidad Carlos III de Madrid. lchen@eco.uc3m.es

‡Dept. of Economics, Universidad Carlos III de Madrid. dolado@eco.uc3m.es

§Dept. of Economics, Universidad Carlos III de Madrid. jesus.gonzalo@uc3m.es

1 Introduction

Despite the well-ackowledged fact that some parameters in economic relationships can become unstable due to important structural breaks (e.g., those related to technological change, globalization or strong policy reforms), a standard practice in the estimation of large factor models is to assume the constancy of the factor loadings. Possibly, one of the main reasons for this benign neglect of breaks is that the first attempt to address this issue, by means of time-varying factor loadings, focused on characterizing the properties of mild instabilities, under which the constructed factors using principal components (PC hereafter) remain consistently estimated (Stock and Watson, 2002).

Later on, however, a few studies have investigated the performance of factor-based forecasting subject not only to mild but also to large breaks in the factor model structures. Banerjee, Marcellino and Masten (2008) conclude that the instability of factor loadings is the most likely reason behind the worsening factor-based forecasts, particularly in small samples. Although their results are exclusively based on Monte Carlo simulations, they shed some light on the importance of detecting relevant structural breaks in the factor loadings. Two additional papers have contributed to this stream of research. The first one is by Stock and Watson (2009) who, extending their previous approach, propose several forms of mild structural instability in factor models to then use empirical evidence showing that the failure of factor-based forecasts is mainly due to the instability of forecast function, rather than of the factor loadings. As a result, they conclude that the estimated factors using PC are still consistent when instabilities are small in magnitude and independent, claiming therefore that forecasts can be improved by using full sample factor estimates and subsample forecasting equations. Yet, this focus on mild structural breaks, though very useful, has also been questioned by Giannone (2007) who argues that *"....to understand structural changes we should devote more effort in modelling the variables characterized by more severe instabilities..."*. In this paper, we follow this route by proving a precise characterization of the different conditions under which big and mild structural breaks in the factor loadings may occur, as well as develop a test to distinguish between them. We conclude that the influence of big breaks cannot be ignored since it may lead to misleading results in the usual econometric practices with factor models.

The second paper, which is the most closely related to ours, is by Breitung and Eickmeier (2010). Like us, these authors propose statistical tests for big structural breaks in the factor loadings. Their test relies on the argument that, under the null of no structural break plus some additional assumptions, the estimation error of the factors can be ignored and thus the estimated factors can be treated as the true factors. Consequently, a Chow-type test can be constructed by means of separate regressions for each variable in the dataset where the regressors are the estimated factors for the whole sample period and their truncated version from the date of the break onwards whre the coefficients on the latter are tested for statistical significance. However, in our view, the Breitung and Eickmeier's test suffers from two limitations: (i) it is based on comparing the empirical rejection frequency among the individual regressions to a nominal size of 5% under the null of no breaks despite the fact that the limiting distribution of this test statistic is not known; and (ii) it is claimed that the number of factors can be correctly estimated using subsamples before and after the known break date. However, if either the break date is not considered to be a priori known or the number of factors is not correctly specified, their test may exhibit poor power. For example, as explained below, a factor model with r common factors and 1 structural break in the factor loadings admits a standard factor representation with $r+1$ common factors without a break. Hence, if the number of factors is incorrectly specified as being $r+1$ instead of r, their test may not detect any break at all.[1]

Our contribution in this paper is to propose a simple testing procedure to detect structural breaks in the factor loadings which allows for different types of breaks and does not suffer from the previous shortcomings. In particular, we first derive some asymptotic results finding that, in contrast to small breaks where both the number of factors and the factor space are consistently estimated, the number of factors will be over-estimated when big breaks occur. We argue that ignoring those big breaks can have serious consequences on the forecasting performance of factors in some popular regression models. We then propose a simple two-step test procedure for testing big breaks. In the first step, the number of factors for the whole sample period is estimated as $\hat{r}$, and then the $\hat{r}$ factors are estimated using PC. In the second step, one of the estimated factors (e.g., the first one) is regressed on the remaining $\hat{r}-1$ factors, and the standard Chow Test or the Sup Type Test of Andrews (1993), depending on whether the date of the break is treated as known or unknown, is then used to test for a structural break in this regression. If the null of no structural breaks is rejected in the second-step regression, we conclude that there are big breaks and, otherwise, that either no breaks at all exist or that only small breaks occur. We also illustrate the finite sample performance of our test using simulations, as well as provide an empirical application of how to implement our testing approach.

The rest of the paper is organized as follows. In Section 2, we present the basic notation, assumptions and give precise definitions of two different types of structural breaks considered here: *big* and *small breaks*. In Section 3, we analyze the consequences of big breaks on the choice of the number of factors and their estimation, as well as the effects of those breaks on the factor augmented regressions. In Section 4, we derive the asymptotic results underlying our approach and discuss the advantages of our proposed test against Breitung and Eickmeier's (2010) test. Section 5 deals with the finite sample performance of our test procedure using Monte-Carlo simulations. Section 6 provides two empirical applications. Finally, Section 7 concludes.

2 Notation and Preliminaries

We consider factor models that can be rewritten in the static canonical form:

$$X_t = AF_t + e_t \tag{1}$$

[1]Even when the break date is known, the number of factors could still be incorrectly estimated due to finite-sample problems of the consistent information criteria used to choose the number of factors to be estimated.

where X_t is the $N \times 1$ vector of observed variables, $A = (\alpha_1, \dots, \alpha_N)'$ is the $N \times r$ matrix of factor loadings, r is the number of common factors, $F_t = (f_{t1}, \dots, f_{tr})'$ is the $r \times 1$ vector of common factors, and e_t is the $N \times 1$ vector of idiosyncratic errors. In the case of dynamic factor models, all the common factors f_t and their lags are stacked into F_t. Thus, a dynamic factor model with r dynamic factors and p lags of these factors can be written as a canonical static factor model with $(r+1) \times p$ static factors. Further, given the assumptions we make about the e_t error terms, the case analyzed by Breitung and Eickmeier (2010) where the e_{it} disturbances are generated by individual specific AR(p_i) processes is also considered. Notice, however, that our setup excludes the generalized dynamic factor models considered by Forni and Lippi (2001) when the polynomial distributed lag tends possibly to infinity.

We assume that there is a single structural break in the factor loadings of all factors at the same time τ:

$$X_t = AF_t + e_t \quad t = 1, 2 \dots, \tau \tag{2}$$

$$X_t = BF_t + e_t \quad t = \tau + 1, \dots, T \tag{3}$$

where $B = (\beta_1, \dots, \beta_N)'$ is the new factor loadings after the break. By defining the matrix $C = B - A$, which captures the size of the breaks, the factor model in (2) and (3) can be rewritten as:

$$X_t = AF_t + CG_t + e_t \tag{4}$$

where $G_t = 0$ for $t = 1, \dots, \tau$, and $G_t = F_t$ for $t = \tau + 1, \dots, T$.

As argued by Stock and Watson (2002), the effects of some mild instability in the factor loadings can be averaged out, so that estimation and inference based on PC remain valid. Our aim is to generalize their analysis by distinguishing between two types of break sizes: *big* and *small*. Whereas the latter correspond to those breaks characterized by Stock and Watson (2002, 2009) and therefore can be neglected, our goal is to analyze which are the effects of the former.We we will show that they cannot be ignored. Thus, to distinguish between both types of breaks, it is convenient to partition the matrix C as follows:

$$C = [\Lambda \quad H]$$

where Λ and H are $N \times k_1$ and $N \times k_2$ matrices that corresponds to the *big* and the *small* breaks, and $k_1 + k_2 = r$. In other words, we assume that, among the r factors, k_1 and k_2 factors are subject to *big* and *small* breaks in their loadings, respectively. Accordingly, we can also partition G_t into two parts, G_t^1 and G_t^2, such that (4) can be rewritten as:

$$X_t = AF_t + \Lambda G_t^1 + HG_t^2 + e_t \tag{5}$$

where $\Lambda = (\lambda_1, \dots, \lambda_N)'$ and $H = (\eta_1, \dots, \eta_N)'$.

Once the basic notation has been established, the next step is to provide precise definitions of the two types of breaks.

Assumption 1. ***Breaks***

a. $E||\lambda_i||^4 < \infty$. $N^{-1}\sum_{i=1}^{N}\lambda_i\lambda_i' \to \Sigma_\Lambda$ *as* $N \to \infty$ *for some positive definite matrix* Σ_Λ.

b. $\eta_i = O_p(\frac{1}{\sqrt{NT}})$ *for* $i = 1,2,\ldots,N$.

The matrices Λ and H are assumed to contain random elements. Assumption 1.a yields the definition of a big break which also includes the case where $\lambda_i = 0$ (no break) for a fixed proportion of variables as $N \to \infty$. Assumption 1.b, in turn, provides the definition of small breaks which can be ignored as N and T goes to infinity.

To investigate the influence of the breaks on the estimation of factors and the number of factors, some further assumptions need to be imposed. To achieve consistent notation with the previous literature in the discussion of these assumptions, we follow the presentation of Bai and Ng (2002) with a few slight modifications. Let $tr(\Sigma)$ and $||\Sigma|| = \sqrt{tr(\Sigma'\Sigma)}$denote the trace and the norm of a matrix Σ, respectively, while $[T\pi]$ denotes the integer part of $T \times \pi$ for $\pi \in (0,1)$. Then

Assumption 2. ***Factors****:* $E(F_t) = 0$*,* $E||F_t||^4 < \infty$*,* $T^{-1}\sum_{t=1}^{T}F_tF_t' \to \Sigma_F$ *and* $T^{-1}\sum_{t=1}^{\tau}F_tF_t' \to \pi^*\Sigma_F$ *as* $T \to \infty$ *for some positive definite matrix* Σ_F *where* $\pi^* = \lim_{t\to\infty}\frac{\tau}{T}$.

Assumption 3. ***Factor Loadings****:* $E||\alpha_i||^4 \leq M < \infty$*, and* $N^{-1}A'A \to \Sigma_A$*,* $N^{-1}\Gamma'\Gamma \to \Sigma_\Gamma$ *as* $N \to \infty$ *for some positive definite matrix* Σ_A *and* Σ_Γ*, where* $\Gamma = [A \quad \Lambda]$.

Assumption 4. ***Idiosyncratic Errors****: the error terms* e_t*, the factors* F_t *and the loadings* A_i *satisfy the Assumption A, B, C, E, F1 and F2 of Bai (2003).*

Assumption 5. ***Independence of Factors, Loadings, Breaks, and Idiosyncratic Errors****:* $[F_t]_{t=1}^{T}$*,* $[\alpha_i]_{i=1}^{N}$*,* $[\lambda_i]_{i=1}^{N}$*,* $[\eta_i]_{i=1}^{N}$ *and* $[e_t]_{t=1}^{T}$ *are mutually independent groups, and for all i*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_te_{it} = O_p(1).$$

While Assumptions 3 and 4 are standard in the literature on factor models allowing for weak cross-sectional and temporal correlations between the errors (see Bai and Ng, 2002), Assumption 2 is a new one. Since factors and factor loadings cannot be separately identified, we have to assume some stable properties for the factors in order to test the stability of the factor loadings. We also allow the different factors to be correlated at all leads and lags. Assumption 5 on the independence among the different groups is stronger than the usual assumptions made by Bai and Ng (2002). Notice, however, that we could have also assumed some dependence between these groups and then impose some restrictions on this dependence when necessary. Yet, this would complicate the proofs without essentially altering the intuition behind the main idea underlying our approach. Thus, for the sake of simplicity, we assume them to be independent.

3 The Effects of Structural Breaks

In this section, we study the effects of the structural breaks on the estimation of factors based on PC, and on the estimation of the number of factors based on the information criteria proposed by Bai and Ng (2002). Our main result is that the estimated factors using PC are not consistent and the number of factors tends to be overestimated when big breaks exist, in contrast to Stock and Watson's (2002, 2009) findings that the true factor space is still consistently estimated.

3.1 The estimation of factors

Let us rewrite model (5) with k_1 big breaks and k_2 small breaks in the more compact form:

$$X_t = AF_t + \Lambda G_t^1 + \varepsilon_t \tag{6}$$

where $\varepsilon_t = HG_t^2 + e_t$. The idea is to show that the new error terms ε_t still satisfy the necessary conditions for (6) being a standard factor model with new factors $F_t^* = [F_t' \quad G_t^{1'}]'$ and new factor loadings $[A \quad \Lambda]$.

Let $\bar{r}$ be the selected number of factors, either by the information criteria or by some prior knowledge. Note that $\bar{r}$ is not necessarily equal to r. Let $\tilde{F}$ be $\sqrt{T}$ times the $\bar{r}$ eigenvectors corresponding to the $\bar{r}$ largest eigenvalues of the matrix XX', and define

$$\hat{F} = \tilde{F}V_{N,T}$$

as the estimated factors, where the $T \times N$ matrix $X = [\bar{X}_1, \bar{X}_2 \ldots \bar{X}_T]'$, $\bar{X}_t = [X_{t1}, X_{t2}, \ldots, X_{tN}]'$, $\hat{F} = [\hat{F}_1, \hat{F}_2, \ldots, \hat{F}_T]'$, and $V_{N,T}$ is a diagonal matrix with the $\bar{r}$ largest eigenvalues of $(NT)^{-1}XX'$. Then we have

Proposition 1. *For any fixed $\bar{r} \geq 1$, under Assumptions 1 to 5, there exists a full rank $\bar{r} \times (r + k_1)$ matrix D and $\delta_{N,T} = \min\{\sqrt{N}, \sqrt{T}\}$ such that:*

$$\hat{F}_t = DF_t^* + O_p(1/\delta_{N,T}) \tag{7}$$

This result implies that $\hat{F}_t$ estimate consistently the space of the new factors, F_t^*, but not the space of the true factors, F_t.

Let us consider two cases. First, when $k_1 = 0$ (no big breaks), we have that $G_t^1 = 0$, and $F_t^* = F_t$, so that (7) becomes

$$\hat{F}_t = DF_t + O_p(1/\delta_{N,T}) \tag{8}$$

for a $\bar{r} \times r$ matrix D of full rank. This just trivially replicates the well-known consistency result of Bai and Ng (2002).

Secondly, in the more interesting case when $k_1 > 0$ (big breaks exist), we can rewrite (7) as

$$\hat{F}_t = [D_1 \quad D_2] \begin{pmatrix} F_t \\ G_t^1 \end{pmatrix} + o_p(1) = D_1F_t + D_2G_t^1 + o_p(1) \tag{9}$$

where the $\bar{r} \times (r+k_1)$ matrix D is partitioned into the $\bar{r} \times r$ matrix D_1 and the $\bar{r} \times k_1$ matrix D_2. Note that, by the definition of G_t, $G_t^1 = 0$ for $t = 1, 2, \ldots, \tau$, and $G_t^1 = F_t^1$ for $t = \tau+1, \ldots, T$, where F_t^1 is the $k_1 \times 1$ sub-vector of F_t that experiences big breaks in their loadings. Therefore (9) can be expressed as:

$$\hat{F}_t = D_1 F_t + o_p(1) \text{ for } t = 1, 2, \ldots, \tau \tag{10}$$

$$\hat{F}_t = D_2^* F_t + o_p(1) \text{ for } t = \tau+1, \ldots, T \tag{11}$$

where $D_2^* = D_1 + [D_2 \quad 0]$, 0 is a $\bar{r} \times (r-k_1)$ zero matrix, such that $D_2 \neq 0$ since D is a full-rank matrix. Hence, since $D_1 \neq D_2^*$, this result implies that, in contrast to small breaks, the estimated factors $\hat{F}$ are not consistent for the space of the true factors F under big breaks. Thus, in this case, the use of estimated factors as predictors or explanatory variables may lead to misleading results in the usual econometric practices with factor models.

To illustrate the consequences of having big breaks in the factor loadings, consider the following simple Factor Augmented Regression (FAR) model (see Bai and Ng, 2006):

$$y_t = a' F_t + b' W_t + u_t, \; t = 1, 2, .., T \tag{12}$$

where W_t is a small set of observable variables and the $r \times 1$ vector F_t contains the r common factors driving a large panel dataset x_{it} $(i = 1, 2, ...N; t = 1, 2, ...T)$ which excludes both y_t and W_t.The parameters of interest are the elements of vector b while F_t is included in (12) to control for potential endogeneity arising from omitted variables. Since we cannot identify F_t and a, only the product $a' F_t$ is relevant. Suppose there is a big break at date τ. From (10) and (11), we can rewrite (12) as:

$$y_t = (a' D_1^-)(D_1 F_t) + b' W_t + u_t \;\text{ for } t = 1, 2, \ldots, \tau$$
$$y_t = (a' D_2^{*-})(D_2^* F_t) + b' W_t + u_t \text{ for } t = \tau+1, \ldots, T$$

where $D_1^- D_1 = D_2^{*-} D_2 = I_r$, or equivalently

$$y_t = a_1' \hat{F}_t + b' W_t + \tilde{u}_t \;\text{ for } t = 1, 2, \ldots, \tau \tag{13}$$

$$y_t = a_2' \hat{F}_t + b' W_t + \tilde{u}_t \;\text{ for } t = \tau+1, \ldots, T \tag{14}$$

where $a_1' = a' D_1^-$ and $a_2' = a' D_2^{*-}$, and $\tilde{u}_t = u_t + o_p(1)$.

If the number of factors is assumed to be known a priori , $\bar{r} = r$, then $D_1^- = D_1^{-1}$, $D_2^{*-} = D_2^{*-1}$. Since $D_1 \neq D_2^*$, it follows that $D_1^{-1} \neq D_2^{*-1}$ and thus $a_1 \neq a_2$. Therefore, using the indicator function $\mathbb{I}(t > \tau\ )$, (13) and (14) can be rewritten as

$$y_t = a_1' \hat{F}_t + (a_2 - a_1)' \hat{F}_t \mathbb{I}(t > \tau) + b' W_t + \tilde{u}_t, \; t = 1, 2, .., T \tag{15}$$

The implication is that if we were to ignore the set of regressors $\hat{F}_t \mathbb{I}(t > \tau)$ in (15), the estimation of b will in general become inconsistent due to ommited variables bias. There are many examples in the literature where the number of factors is a priori imposed for theoretical reasons, e.g., to name a few, a single common factor representing a global effect is assumed in the well-known study by Bernanke, Boivin and Eliasz (2005) on measuring the effects of monetary policy in Factor Augmented VAR (FAVAR) models, or two factors are imposed by Rudebusch and Wu (2008) in their macro-finance model.

Alternatively, if the number of factors is not assumed to be apriori known and therefore needs to be estimated using some information criteria, we will show in Proposition 2 in the next section that the chosen number of factors will tend to $r+k_1$ as the sample size gets large. In this case, D_1 and D_2 are $(r+k_1)\times r$, and by the definitions of D_1 and D_2^*, it is easy to show that we can always find a $r\times(r+k_1)$ matrix $D^*=D_1^-=D_2^{*-}$ such that $D^*D_1=D^*D_2^*=I_r$. If we define

$$a^*=a'D^* \tag{16}$$

then $a_1'=a_2'=a^*$ so that (13) and (14) can be rewritten as

$$y_t=a^*\hat{F}_t+b'W_t+\tilde{u}_t,\ t=1,2,..,T \tag{17}$$

From above equation we can see that the estimation of (12) will not be affected by the estimated factors under big breaks if $\bar{r}=r+k_1$.

In sum, in the presence of big breaks, the use of estimated factors as the true factors when assuming that the number of factors is a priori known will lead to inconsistent estimates in a FAR. As a simple remedy, $\hat{F}_t\mathbb{I}(t>\tau)$ should be added as regressors when big breaks are detected and the break date is located. Alternatively, without pretending to know a priori the true number of factors, the estimation of FAR will be robust to the estimation of factors under big breaks if the number of factors is overestimated. Notice that a similar argument will render inconsistent the impulse response functions in FAVAR models where (12) becomes $y_{t+1}=(F_{t+1},W_{t+1})'$. As a result, in order to run regression (17), a formal test of whether big breaks exist is required.We will illustrate these points by using simulations in a typical forecasting exercise where the predictors are common factors estimated by PC.

3.2 The estimated number of factors

Breitung and Eickmeier (2010) have previouosly argued that the presence of structural breaks in the factor loadings may lead to the overestimation of the number of factors but they do not prove this result. In this part, we fill this gap by providing a rigorous proof.

Let $\hat{r}$ be the estimated number of factors in (6) using the information criteria of Bai and Ng (2002). Then the following result holds:

Proposition 2. *Under Assumptions 1 to 5, it holds that*

$$\lim_{N,T\to\infty}\mathbb{P}[\hat{r}=r+k_1]=1$$

When there is no big break ($k_1 = 0$), this result replicates Theorem 2 of Bai and Ng (2002). However, under big breaks ($k_1 > 0$), their information criteria will overestimate the number of factors by the number of big breaks (k_1). Actually, Bai and Ng (2002)'s criteria, that consistently estimate the number of true factors, will overestimate the number of original factors when there are big breaks because we have shown that a factor model with those breaks admits a representation without break but with more factors.

Finally, notice that, although the presence of structural breaks in the factor loadings may lead to wrong estimation of the factor space and the number of factors, the common part of a factor model (AF_t and BF_t) can still be consistently estimated if enough factors are extracted.

4 Testing for Structural Breaks

4.1 Hypotheses of interest and test statistics

From the previous discussion, we have found that the factor space and the number of factors are both consistently estimated only when mildl breaks exist. Therefore, our goal here is to develop a test for big breaks.

If we were to follow the usual approach in the literature to test of structural breaks, we should consider

$$H_0 : A = B$$
$$H_1 : A \neq B$$

However, if only small breaks occur, the alternative hypothesis may not be interesting since $C = A - B$ vanishes as $N \to \infty$ and $T \to \infty$. Thus, this kind of local alternatives for which the usual test should have no trivial power, is not relevant for the large factor models we consider here. Therefore, since our focus is on big breaks, we consider instead:

$$H_0 : k_1 = 0$$
$$H_1 : k_1 > 0$$

where the null and alternative hypotheses correspond to the cases where there are no big breaks (yet there may be small breaks) and there is at least one big break, respectively.

To test the above null hypothesis, we consider the following two-step procedure:

1. *In the first step, the number of factors to estimate, $\bar{r}$, is either determined by Bai and Ng ´s (2002) information criteria ($\bar{r} = \hat{r}$) or by prior knowledge, so that $\bar{r}$ common factors ($\hat{F}_t$) are estimated by PC.*

2. *In the second step, we consider the following linear regression of the first estimated factor on the remaining $\bar{r} - 1$ ones:*

$$\hat{F}_{1t} = c_2\hat{F}_{2t} + \cdots + c_{\bar{r}}\hat{F}_{\bar{r}t} + u_t = c'\hat{F}_{-1t} + u_t \tag{18}$$

where $\hat{F}_{-1t} = [\hat{F}_{2t} \cdots \hat{F}_{\bar{r}t}]'$ and $c = [c_2 \cdots c_{\bar{r}}]'$ are $(\bar{r}-1) \times 1$ vectors. Then we test for a structural break of c in the above regression. If a structural break is detected, then we reject $H_0 : k_1 = 0$; otherwise, we cannot reject the null stating that there are no big breaks.

Both steps can be easily implemented in practice. In the second step, although there are many methods of testing for structural breaks in a simple linear regression model, we consider the *Chow Test* when the possible break date is assumed to be known, and the *Sup-type Test* when no prior knowledge about the break date exists. Moreover, since the Wald, LR, and LM test statistics have the same asymptotic distribution under the null, we focus on the LM and Wald tests because they are simpler to compute.

Following Andrews (1993), the LM test statistic is defined as:

$$\mathscr{L}(\bar{\pi}) = \frac{T}{\bar{\pi}(1-\bar{\pi})}\Big(\frac{1}{T}\sum_{t=1}^{\tau}\hat{F}_{-1t}\hat{u}_t\Big)'\hat{S}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{\tau}\hat{F}_{-1t}\hat{u}_t\Big) \tag{19}$$

where $\bar{\pi} = \tau/T$, $\hat{u}_t$ is the residuals in the OLS regression of (18), $S = \lim_{T\to\infty} \mathrm{Var}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{F}_{-1t}u_t\Big)$, and $\hat{S}$ is a consistent estimate of S.

The Sup-LM statistic is defined as:

$$\mathscr{L}(\Pi) = \sup_{\pi\in\Pi}\frac{T}{\pi(1-\pi)}\Big(\frac{1}{T}\sum_{t=1}^{[T\pi]}\hat{F}_{-1t}\hat{u}_t\Big)'\hat{S}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{[T\pi]}\hat{F}_{-1t}\hat{u}_t\Big) \tag{20}$$

where Π is some pre-specified subset of $[0,1]$.

Similarly, the Wald and Sup-Wald test statistics can be constructed as:

$$\mathscr{L}^*(\bar{\pi}) = T\Big(\hat{c}_1(\bar{\pi}) - \hat{c}_2(\bar{\pi})\Big)'\hat{V}^{-1}\Big(\hat{c}_1(\bar{\pi}) - \hat{c}_2(\bar{\pi})\Big) \tag{21}$$

and

$$\mathscr{L}^*(\Pi) = \sup_{\pi\in\Pi} T\Big(\hat{c}_1(\pi) - \hat{c}_2(\pi)\Big)'\hat{V}^{-1}\Big(\hat{c}_1(\pi) - \hat{c}_2(\pi)\Big) \tag{22}$$

where $\hat{c}_1(\pi)$ and $\hat{c}_2(\pi)$ are OLS estimates of c using subsamples before and after the break point : $[T\pi]$. In addition, $\hat{V} = \hat{M}^{-1}\hat{S}\hat{M}^{-1}$, and $\hat{M} = T^{-1}\sum_{t=1}^{T}\hat{F}_{-1t}\hat{F}'_{-1t}$.

To illustrate why our two-step procedure is able to detect the big breaks, it is useful to consider a simple example where $r = 1, k_1 = 1$ (one common factor and one big break). Then (6) becomes:

$$X_t = Af_t + \Lambda g_t + \varepsilon_t$$

where $g_t = 0$ for $t = 1, \ldots, \tau$, and $g_t = f_t$ for $t = \tau+1, \ldots, T$. By Proposition 2, we will tend to get $\hat{r} = 2$ in this case. Suppose now that we estimate 2 factors ($\bar{r} = 2$). Then, by Proposition 1, we have:

$$\begin{pmatrix}\hat{f}_{t1}\\ \hat{f}_{t2}\end{pmatrix} = D\begin{pmatrix}f_t\\ g_t\end{pmatrix} + o_p(1)$$

where $D = \begin{pmatrix} d_1 & d_2 \\ d_3 & d_4 \end{pmatrix}$ is a non-singular matrix. By the definition of g_t we have:

$$\hat{f}_{t1} = d_1 f_t + o_p(1) \quad \hat{f}_{t2} = d_3 f_t + o_p(1) \quad \text{for} \quad t = 1, \ldots, \tau$$

$$\hat{f}_{t1} = (d_1 + d_2) f_t + o_p(1) \quad \hat{f}_{t2} = (d_3 + d_4) f_t + o_p(1) \quad \text{for} \quad t = \tau + 1, \ldots, T$$

which imply that:

$$\hat{f}_{t1} = \frac{d_1}{d_3} \hat{f}_{t2} + o_p(1) \quad \text{for} \quad t = 1, \ldots, \tau$$

$$\hat{f}_{t1} = \frac{d_1 + d_2}{d_3 + d_4} \hat{f}_{t2} + o_p(1) \quad \text{for} \quad t = \tau + 1, \ldots, T$$

Thus, we can observe that the two estimated factors are linearly related and that the coefficients $\frac{d_1}{d_3}$ and $\frac{d_1+d_2}{d_3+d_4}$ before and after the break date must be different due to the non-singularity of the matrix D. As a result, if we regress one of the estimated factors on the other and test for a structural break in this regression, we should reject the null of no big break. We choose the first estimated factor, $\hat{f}_{t1}$, as the regressand in the previous regressions because being the "main factor" in the PC analysis it is likely that $d_3 \neq 0$.[2]Likewise, if the break date τ is not a priori assumed to be known, the Sup-type Test will yield a natural estimate of τ at the date when the test reaches its maximum value. In what follows, we derive the asymptotic distribution of the test statistics (19) and (20) under the null hypothesis, as well as extend the intuition behind this simple example to the more general case in order to show that our test has power against relevant alternatives.

4.2 Limiting distributions under the null hypothesis

Since in most applications, the number of factors is estimated by means of the information criteria, and it converges to the true one under the null hypothesis of no big break, we start with the most interesting case where $\bar{r} = r$.

Note that use of PC implies that $\sum_{t=1}^{T} \hat{F}_{-1t} \hat{F}_{1t} = 0$ for any T by construction, so we have $\hat{c} = 0$ in (18) and $\hat{u}_t = \hat{F}_{1t}$ in (19). To derive the asymptotic distributions of the LM statistics, we impose the following additional assumptions:

Assumption 6. *$\sqrt{T}/N \to 0$ as $N \to \infty$ and $T \to \infty$. .*

Assumption 7. *$\{F_t\}$ is a stationary and ergodic sequence, and $\{F_{it}F_{jt} - E(F_{it}F_{jt}), \Omega_t\}$ is an adapted mixingale with γ_m of size -1 for $i, j = 1, 2, \ldots, r$, that is:*

$$\sqrt{E\left(E(Y_{ij,t} | \Omega_{t-m})^2\right)} \leq c_t \gamma_m$$

where $Y_{ij,t} = F_{it}F_{jt} - E(F_{it}F_{jt})$, Ω_t is a $\sigma-$ algebra generated by the information at time $t, t-1, \ldots$, $\{c_t\}$ and $\{\gamma_m\}$ are non-negative sequences and $\gamma_m = O(m^{-1-\delta})$ for some $\delta > 0$.

[2]Since D is non singular, even if $d_3 = 0$, d_1 cannot be equal to zero. If the regression for the first subsample yields an ill-defined (ie., very large) estimated slope, then we recommed using $\hat{f}_{t2}$ as the regressand and $\hat{f}_{t1}$ as the regressor.

Assumption 8. *For the subset* Π *of* $[0,1]$*:*

$$\sup_{\pi\in\Pi}\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T\pi}\sum_{i=1}^{N}\alpha_i F_t' e_{it}\right\|^2 = O_p(1)$$

Assumption 9. $\left\|\hat{S}-S\right\| = o_p(1)$*, and* S *is a* $(r-1)\times(r-1)$ *symmetric positive definite matrix.*

Assumption 6 and 8 are required to bound the estimation errors of $\hat{F}_t$, while Assumption 7 is necessary for deriving the weak convergence of the test statistics using the Functional Central Limit Theorem (FCLT).

Note that these assumptions are not restrictive. Assumption 6 allows T to be $O(N^{1+\delta})$ for $-1<\delta<1$. As for Assumption 7, it allows one to consider a quite general class of linear processes for the factors: $F_{it}=\sum_{k=1}^{\infty}\phi_{ik}v_{i,t-k}$, where $v_t=[v_{1t}\dots v_{rt}]'$ are i.i.d with zero means, and $Var(v_{it})=\sigma_i^2<\infty$. It can shown that in this case:

$$\sqrt{E\left(E(Y_{ij,t}|\Omega_{t-m})^2\right)} \leq \sigma_i\sigma_j\left(\sum_{k=m}^{\infty}|\phi_{ik}|\right)\left(\sum_{k=m}^{\infty}|\phi_{jk}|\right)$$

then it suffices that

$$\left(\sum_{k=m}^{\infty}|\phi_{ik}|\right) = O(m^{-1/2-\delta})$$

for some $\delta>0$, which is satisfied for a large class of ARMA processes. Assumption 8 is analogue to Assumption F.2 of Bai (2003), which involves zero- mean random variables. Finally, a consistent estimate of S can be calculated by a HAC estimator.

Let " $\xrightarrow{d}$ "denote *convergence in distribution* , and $\mathscr{W}_{r-1}(\cdot)$ denote a $r-1$ vector of standard Brownian Motions, then:

Theorem 1. *Under the null hypothesis* $H_0: k_1=0$ *and Assumptions 1 to 9:*

$$\mathscr{L}(\Pi) \xrightarrow{d} \sup_{\pi\in\Pi}\left(\mathscr{W}_{r-1}(\pi)-\pi\mathscr{W}_{r-1}(1)\right)'\left(\mathscr{W}_{r-1}(\pi)-\pi\mathscr{W}_{r-1}(1)\right)/[\pi(1-\pi)];$$

$$\mathscr{L}(\bar{\pi}) \xrightarrow{d} \chi^2(r-1).$$

The critical values for the Sup-type test are provided in Andrews (1993).

It is easy to show that Theorem 1 still holds when $\bar{r}<r$. Yet, when $\bar{r}>r$, the covariance matrix S is not full ranked, although $\hat{S}$ can be inverted in any given finite sample size. In the following section, we will show through simulations that Theorem 1 still provide good approximations for the test statistics even when $\bar{r}>r$.

4.3 Behavior of LM and Wald tests under the alternative hypothesis

We extend the idea of the simple example in section 4.1 to show that, under the alternative hypothesis ($k_1 > 0$), the linear relationship between the estimated factors changes at time τ, so that big breaks can be detected.

First, let us consider the case where $r < \bar{r} \le r + k_1$ so that D_1 and D_2^* in (10) and (11) become $\bar{r} \times r$ matrices with full column rank. Notice that, since $r < \bar{r}$, we can always find $\bar{r} \times 1$ vectors ρ_1 and ρ_2 which belong to the null spaces of D_1' and $D_2^{*'}$ separately, that is, $\rho_1' D_1 = 0$ and $\rho_2' D_2^* = 0$. Hence, premultiplying both sides of (10) and (11) by ρ_1' and ρ_2' leads to:

$$\rho_1' \hat{F}_t = o_p(1) \quad t = 1, 2 \ldots, \tau$$
$$\rho_2' \hat{F}_t = o_p(1) \quad t = \tau + 1, \ldots, T,$$

which, after normalizing the first elements of ρ_1 and ρ_2 to be 1, yields:

$$\hat{F}_{1t} = \hat{F}'_{-1t} \rho_1^* + o_p(1) \quad t = 1, 2 \ldots, \tau \tag{23}$$
$$\hat{F}_{1t} = \hat{F}'_{-1t} \rho_2^* + o_p(1) \quad t = \tau + 1, \ldots, T \tag{24}$$

Next, to show that $\rho_1^* \neq \rho_2^*$, we proceed as follows. Suppose that $\gamma \in \mathrm{Null}(D_1')$ and $\gamma \in \mathrm{Null}(D_2^{*'})$, then by the definition of D_1 and D_2^* and by the basic properties of full-rank matrices, it holds that $\gamma \in \mathrm{Null}(D')$. Since D is full rank $\bar{r} \times (r + k_1)$ matrix, then $\mathrm{Null}(D') = 0$ and thus $\gamma = 0$. Therefore, the only vector that belongs to the null space of D_1 and D_2^* is the trivial zero vector. Further, because the rank of the null space of D_1 and D_2^* is $\bar{r} - r > 0$, we can always find two non- zero-vectors such that $\rho_1 \neq \rho_2$.

Notice that when $\bar{r} \le r$, the rank of the null spaces of D_1 and D_2^* becomes zero. Hence, the preceding analysis does not apply in this case despite the existence of linear relationships among the estimated factors. If we regress one of the estimated factors on the others, with $\hat{\rho}_1$ and $\hat{\rho}_2$ denoting the OLS estimates of the coefficients using the subsamples before and after the break, it is easy to show that $\hat{\rho}_1 \to \theta_1$ and $\hat{\rho}_2 \to \theta_2$, but generally we cannot verify that $\theta_1 \neq \theta_2$.

In the case where $\bar{r} > r + k_1$, the rank of null space of D defined in Proposition 1 becomes $\bar{r} - (r + k_1)$. Applying similar arguments as above, we can find a non zero $\bar{r} \times 1$ vector ρ such that $\rho' D = 0$. Then, premultiplying both sides of (7) by ρ' and normalizing the first element of ρ to be 1, it follows that:

$$\hat{F}_{1t} = \hat{F}'_{-1t} \rho^* + o_p(1) \text{ for } t = 1, 2, \ldots, T$$

Hence, there is still a linear relationship between the estimated factors, but this relationship (ρ^*) is constant over time.

As a result, our test may fail to detect the breaks when $\bar{r} \le r$ or $\bar{r} > r + k_1$, which is confirmed by the simulation results shown in the following section. However, this may not be a problem due to two reasons. First, we usually equate the number of factors with the estimated ones, ($\bar{r} = \hat{r}$) and we have shown that $\mathbb{P}[\hat{r} = r + k_1] \to 1$. Secondly, instead of using a single value, we can try different values of $\bar{r}$. Then, under the null, we should not detect any break no matter which value of $\bar{r}$ we use while, under the alternative, we should detect breaks when $\bar{r}$ lies between r and $r + k_1$.

4.4 Comparison with other available tests

Although the issue of instability in factor models was initially raised by Stock and Watson (2002) in the context of small breaks, Breitung and Eickmeier (2010) (BE test, henceforth) is, to our knowledge, the only available paper that proposes a test for big breaks. Thus, it is natural to compare our testing procedure with theirs. In our view, the BE test suffers from three shortcomings which are worth mentioning before the comparison is made.

First, the BE test will lose power when the number of factors is overestimated. The BE test is equivalent to the Chow test in the regression $X_{it} = \alpha_i F_t + e_{it}$ where F_t is replaced by $\hat{F}_t$. However, as shown in equation (5), a factor model with big breaks in the factor loadings admits a new representation with more factors but no break. In other words, when the number of factors is overestimated, the PC estimators consistently estimate (up to a linear transformation) the new factors and loadings which are stable in the new representation. Thus the BE test may fail to detect breaks in this case. Although the authors are fully aware of this problem (see Remark B in their paper) and suggest to split the sample to estimate the correct number of factors, in principle this is not feasible when the break date is considered to be unknown. Using a Sup-Type Test, as BE propose, solves the problem of the unknown break date but, since the number of factors will tend to be overestimated, lack of power will still be a problem.

Secondly, their testing procedure is mainly heuiristic. Their null hypothesis is $A = B$, or $\alpha_i = \beta_i$ for all $i = 1, \ldots, N$, rather than $\alpha_j = \beta_j$ for a specific j [3]. They construct N test statistics (denoted by s_i $i = 1, \ldots, N$) for each of the N variables, but do not derive a single statistic for $H_0 : A = B$. One possibility that the authors mention is to combine the N individual statistics to obtain a pooled test, but this requires the errors e_{it} and e_{jt} to be independent if $i \neq j$, an assumption which is too restrictive. In their simulations and applications, the decisions are merely based on the *rejection frequencies*, i.e., the proportion of variables that are detected to have breaks using the individual statistics s_i. This rejection frequency, defined by $N^{-1} \sum_{i=1}^{N} \mathbb{I}(s_i > \alpha)$ where $\mathbb{I}\,(.)$ is an indicator function and α is some critical value, may converge to some predetermined nominal size (typically 5%), as shown by their simulations, but this is not a proper test insofar as its limiting distribution is not derived.

Finally, the individual tests for each of the variables may lead to incorrect conclusions about which individual variables are subject to breaks in their loadings of the factors, as

[3] The authors do not mention this, but it is implicitly assumed because they need the factors to be consistently estimated under the null, which will hold only if $\alpha_i = \beta_i$ for all $i = 1, \ldots, N$, or alternatively if the the break is small according to our definition..

BE seemingly do.[4] A key presumption for their individual test to work properly is that the estimated factors $\hat{F}_t$ can replace the true factors, even under the alternative hypothesis (given that the number of factors is correctly estimated). As we have shown before, the true factor space can only be consistently estimated under the null of no break or only small breaks. By contrast, when big breaks exist, the space of the true factors is not well estimated (see equations (10) and (11)). If we plug in the estimated factors in this case, some variables that have constant loadings may be detected to have breaks due to the poor estimation of the factors. For example, consider a factor model with big breaks in the factor loadings where we select the right number of factors $\bar{r} = r$, and there is one of the variables X_{it} that has constant loadings:[5]

$$X_{it} = \alpha_i' F_t + e_{it}.$$

Then, from (10) and (11), we can also write the above-mentioned equation as follows:

$$\begin{aligned} X_{it} &= (\alpha_i' D_1^{-1})(D_1 F_t) + e_{it} = (\alpha_i' D_1^{-1})\hat{F}_t + \tilde{e}_{it} \quad t = 1, 2 \ldots, \tau \\ X_{it} &= (\alpha_i' D_2^{*-1})(D_2^* F_t) + e_{it} = (\alpha_i' D_2^{*-1})\hat{F}_t + \tilde{e}_{it} \quad t = \tau + 1, \ldots, T \end{aligned}$$

where $\tilde{e}_{it} = e_{it} + o_p(1)$. Notice that $\alpha_i' D_1^{-1} \neq \alpha_i' D_2^{*-1}$ since $D_1 \neq D_2^*$. As a result, the factor loadings will exhibit a break when the true factors are replaced by the estimated factors. Hence if we apply the individual test to X_{it} using $\hat{F}_t$, we may wrongly conclude that there is a big break in that variable when there is none.

To analyze how serious this problem could be in practice, we design a very simple simulation. First, we generate a factor model with $N = T = 200$, $r = 1$, where the first 100 variables have constant factor loadings while the remaining 100 variables have big breaks in their loadings. Then we estimate the factors by PC and apply the individual tests for all the 200 variables.[6] Applying the BE test, we find that rejection frequency for all the 200 variables is 53.07%, close to the proportion of variables that have breaks. However, the rejection frequencies for the first and second 100 variables are 52.98% and 53.15% , respectively, which means that we falsely reject the null for more than half of the variables that are stable while we reject the null correctly for only half of the variables that have breaks. Further, if we increase the size of the breaks, the reject frequency can rise up to 90% while the true proportion is 50%.

Our LM and Wald tests cannot identify either which particular variables are subject to breaks in the factor loadings but avoid the other two problems. Regarding the first problem, we have derived its limiting distribution in Theorem 1 both for the cases of known and unknown breaking dates. As for the second one, contrary to the BE test, our

[4] For example, in BE (2010, Section 6, pg. 26), it is stated that "there seems to be a break in the loadings on the CPI and consumer expectations,..., but not in the loadings of commodity prices".

[5] Notice that this is possible because of Assumption 1.a.

[6] For simplicity, all the loadings, factors and errors are generated as standard normal variables, the mean of the factor loadings of the second 100 variables are shifted by 0.3 at time $\tau = 100$. The reported numbers are averages of 1000 replications

test needs more estimated factors than the true number $(r+k_1 \geq \bar{r} > r)$ to maintain the power. However, this overestimation it is still preferable to the BE test because in practice the number of factors to estimate is chosen by means of the information criteria $(\bar{r} = \hat{r})$, and we have proved in Proposition 2 that $\mathbb{P}[\hat{r} = r+k_1] \to 1$.

5 Simulations

In this section, we first study the finite sample properties of our proposed LM/Wald and Sup-LM/Wald tests. Then a comparison is made with the properties of the BE test by means of Monte-Carlo simulations. Since the only BE test with a known limiting distribution is their pooled statistic when the idiosyncratic components in the factor model are uncorrelated, we restrict the comparison to this specific case instead of using their rejection frequency approach whose asymptotic distribution remains unknown.

5.1 Size properties

We first simulate data from the following DGP:

$$X_{it} = \sum_{k=1}^{r} \alpha_{ik} F_{kt} + e_{it}$$

where $r = 3$, α_{ik} and e_{it} are generated as *i.i.d* standarised normal variables, and $\{F_{kt}\}$ are generated as:

$$F_{kt} = \phi_k F_{k,t-1} + v_{kt}$$

where $\phi_1 = 0.8$, $\phi_2 = 0.5$, $\phi_3 = 0.2$, and v_{kt} is another *i.i.d* standarized normal error term. The number of replications is 1000. We consider both the LM and Wald tests and their Sup-type versions defined in (19)-(20) and (21)-(22). The potential breaking date τ is considered to be a priori known and is set at $T/2$ for the LM/Wald tests while Π is chosen as $[0.15, 0.85]$ for the Sup-type versions of the tests. The covariance matrix S is estimated using the HAC estimator of Newey and West (1987).

Table 1 reports the empirical sizes (in percentages) for the LM/Wald tests and Sup-LM/Wald tests (in brackets) using 5% critical values for sample sizes (N and T) equal to 100, 150, 200, 250, 300 and 1000.[7]. We consider three cases regarding the choice of the number of factors to be estimated by PC: (i) the correct one ($\bar{r} = r = 3$), (ii) smaller than the true number of factors ($\bar{r} = 2 < r = 3$), and (iii) larger than the true number of factors ($\bar{r} = 4 > r = 3$).[8]

Broadly speaking, the LM and Wald tests are slightly undersized for $r = 2$ and 3 and more so when $r = 4$. Yet the effective sizes converge to the nominal size as N and T

[7] As mentioned earlier, the critical values of the Sup–type tests are taken from Andrews (1993).

[8] Notice that the choice of $r = 3$ allows us to analyze the consequences of performing our proposed test with the under-parameterised choice of $\bar{r} = 2$, where two factors are needed to perform the LM/Wald test in (18). Had we chosen $r = 2$ as the true number of factors, the test could not be performed for $\bar{r} = 1$.

increase This finite sample problem is more accurate with the Sup-LM test especially for small T, in line with the findings in other studies (see, Diebold and Chen, 1996) This is hardly surprising because, for instance, when $T = 100$ and $\Pi = [0.15, 0.85]$, we only have 15 observations in the first subsample. By contrast, the Sup-Wald test is too liberal for $T = 100$. Therefore, although we impose that $\sqrt{T}/N$ goes to zero, a large T is preferable when the Sup-LM test is used. Another conclusion to be drawn is that, despite some minor differences, the tests perform quite similarly in terms of size even when the selected number of factors is not correct.

5.2 Power properties

We next consider similar DGPs but this time with $r = 2$ and now subject to big breaks which are characterized as deterministic shifts in the means of the factor loadings [9]. The factors are simulated as AR(1) processes with coefficients of 0.8 for the first factor and 0.2 for the second. The shifts in the loadings are 0.2 and 0.4 at time $\tau = T/2$. Table 2 reports the empirical rejection rates of the LM/Wald and Sup-LM/Wald tests in percentage terms using again 1000 replications.

As expected, both tests are powerful to detect the breaks as long as $r = 2 < \bar{r} \leq r + k_1 = 4$, while the power is trivial when $\bar{r} = r = 2$. Moreover, the power is low for the Sup-LM test when $T \leq 150$, which is not surprising given that the Sup-LM test is undersized. This problem could be fixed by either using size-corrected critical values, or by the Sup-Wald test that is more powerful in finite samples. For safety, we recommend to use data sets with large T (at least around 200) in practice.

5.3 Comparison with BE test

To compare our test to the BE test, we need to construct a pooled statistic as suggested at the beginning of this section. The pooled BE test is constructed as follows:

$$\frac{\sum_{i=1}^{N} s_i - N\bar{r}}{\sqrt{2N\bar{r}}}$$

where s_i is the individual LM statistics in BE (2010). This test should converge to a standarised normal distribution as long as e_{it} and e_{jt} are independent, an assumption we also adopt here. For simplicity, we only report results for the case of known break dates.

We first generate factor models with $r = 2$, and compare the two tests under the null. The DGPs are similar to those used in the size study. The second column of Table 3 (no break) reports the 5% empirical sizes. In general, we find that the pooled BE and the LM tests exhibit similar sizes.

Then, we generate a break in the loadings of the first factor while the other parts of the models remain the same as in the DGP where we study the power properties. The break

[9]The results with other types of breaks such as random shifts are similar.

is generated as a shift of 0.1 in the mean of the loadings. We consider two cases: (i) the number of factors is correctly selected: $\bar{r} = r = 2$; and (ii) the selected number of factors is larger than the true one: $\bar{r} = 3 > r = 2$. The third and fourth columns in Table 3 report the empirical rejection rates of both tests. In agreement with our previous discussion, the differences in power are striking: when $\bar{r} = 2$, the pooled BE test is much more powerful while the opposite occurs when $\bar{r} = 3$. However, according to our result in Proposition 2, the use of Bai and Ng's (2002) selection criteria will yield the choice of $\bar{r} = 3$ as a much more likely outcome as N and T increase.

5.4 The effect of big breaks on forecasting

Finally, in this section we consider the effect of having big breaks in a typical forecasting exercise where the predictors are estimated common factors. First, we have a large panel of data X_t driven by the factors F_t which are subject to a break in the factor loadings:

$$X_t = AF_t\mathbb{I}(t \leq \tau) + BF_t\mathbb{I}(t > \tau) + e_t$$

Secondly, the variable we wish to forecast y_t , which is excluded from to X_t, is assumed to be related to F_t as follows:

$$y_{t+1} = a'F_t + v_{t+1}$$

We consider a DGP where $N = 100, T = 200, \tau = 100, r = 2, a' = [1 \quad 1]$, F_t are generated as two AR(1) processes with coefficients 0.8 and 0.2, respectively, e_t and v_t are i.i.d normal variables, and the break size is characterized by a mean shift between loadings A and B occuring at half of the time sample size.

The following forecasting methods are compared in our simulation:

Bechmark Forecasting: The factors F_t are treated as observed and are used directly as predictors. The one-step-ahead forecast of y_t is defined as $y_t(1) = \hat{a}'F_t$, where $\hat{a}$ is the OLS estimate of a using y_{t+1} and F_t.

Forecasting 1: We first estimated 2 factors $\hat{F}_t$ from X_t by PCs, which are then used as predictors. $y_t(1) = \hat{a}'\hat{F}_t$, where $\hat{a}$ is the OLS estimate of a using y_{t+1} and $\hat{F}_t$.

Forecasting 2: We first estimated 2 factors $\hat{F}_t$ from X_t by PC, then use $\hat{F}_t$ and $\hat{F}_t\mathbb{I}(t > \tau)$ as predictors. $y_t(1) = \hat{a}'[\hat{F}_t \quad \hat{F}_t\mathbb{I}(t > \tau)]$, where $\hat{a}$ is the OLS estimate of a in the regression of y_{t+1} on $\hat{F}_t$ and $\hat{F}_t\mathbb{I}(t > \tau)]$.

Forecasting 3: We first estimated 4 factors $\hat{F}_t$ from X_t by PC, then use them as predictors. $y_t(1) = \hat{a}'\hat{F}_t$, where $\hat{a}$ is the OLS estimate of a using y_{t+1} and $\hat{F}_t$.

The above forecastings are implemented recursively, e.g., at each time t, the data $X_t, X_{t-1}, \ldots, X_1$ and $y_t, y_{t-1}, \ldots, y_1$ are treated as known to forecast y_{t+1}. This process starts from $t = 149$ to $t = 199$, and the mean square errors (MSEs) are calculated by

$$MSE = \sum_{t=149}^{199} \frac{(y_{t+1} - y_t(1))^2}{51}$$

To facilitate the comparisons, the MSE of the Benchmark Forecasting is standardized to 1.

The results of 1000 replications are reported in Figure 1 with different break sizes ranging from 0 to 1. It is clear that the MSE of the Forecasting 1 method increases drastically with the size of the breaks. The Forecasting 1 and 2 procedures perform equally well and their MSEs remain constant as the break size increases, although they can not outperform the benchmark forecasting due to the estimation errors of the factors.

6 Empirical Applications

To provide a few empirical applications of our test, we first use the dataset of Stock and Watson (2009). This data consist of 144 quarterly time series for the US ranging from 1959:Q1 to 2006:Q4, concerning nominal and real variables. Since not all variables are available for the whole period, we end up using their suggested balanced panel of standarized variables with $T = 190, N = 109$ which more or less corresond to the case where T=200, N=100 in Table 2, where no severe size distortions are found. We refer to Stock and Watson (2009) for the details of the the data description and the standardization procedure to achieve stationarity.

Since the estimated numbers of factors using various Bai and Ng's (2002) information criteria range from 3 to 6, we implement the test for $\bar{r} = 3,4,5$ and 6. For the Sup- LM and Wald tests, the trimming $\Pi = [0.3, 0.7]$ is used. It corresponds to the time period ranging from 1973Q3 to 1992Q3 which includes several relevant events like, e.g., the second oil price shock (1979) and the beginning of great moderation (1984). The graphs displayed in Figure 1 are the series of LM and Wald tests for different values of $\bar{r}$, with the horizontal lines representing the 5% critical values of the Sup-type test.

As can be observed, the LM and the Wald tests reject the null of no big breaks (i.e., exceeds the lower horizontal line) for $\bar{r} = 4,5,6$ when the break date is assumed to be known at 1984 in agreement with the results in BE (2010). Stock and Watson (2009) get similar conclusions about the existence of breaks around the early 1980s. However, one important implication of our results is that the breaks should be interpreted as being big and thus cannot be neglected.

As for the case when the break date is not assumed to be a priori known, we find that, while the Sup-LM test cannot reject the null for all values of $\bar{r}$, the Sup-Wald test rejects the null when $\bar{r} = 5,6$.(i.e., exceeds the upper horizontal line) The estimate of the break date provided by the last test is around 1979 (second oil price shock), rather than 1984 which, as mentioned before, is the only date considered by BE (2010) as a potential break date in their empirical application using the same dataset.

A second empirical application relies on another dataset of Stock and Watson (2003). The data we use consists of 240 monthly marco series from 11 European countries from 1982M1 to 1997M8. This data set is standardized to a panel with $T = 188$ and $N = 240$ (see the original paper for the details). We use the trimming $\Pi = [0.15, 0.85]$ which spans the period from 1984M12 to 1995M6, during which the Maastricht Treaty was signed and the European Union was created. The results of the LM and Wald tests are shown in Figure 2 with the 5% critical values of the Sup-type test for $\bar{r} = 3$ to 6.

It is clear that, under the assumption of a known break date, the comparison of the test values to the 5% critical values of a χ^2 distribution implies that we can easily reject the null of no big break around 1994. However, in contrast to the Stock and Watson's (2009) dataset, if we compare the maximum of the LM and Wald tests to the critical values of the Sup-type test, no big break is detected during the sample period.

7 Conclusions

In this paper, we propose a simple two- step procedure to test for big structural breaks in the factor loadings of large dimensional factor models.that overcomes some limitations in other available tests, like Breitung and Eickmeier (2010). In particular, we derive the limiting distributions of our test, while theirs remains unknown, and show that it has much better power than their test when the choice of the number of factors is based upon Bai and Ng's (2002) consistent information criteria Our method can be easily implemented in practice either when the break date is considered to be known or unknown, and can be adapted to a sequential procedure when the number of factors might not be correctly chosen in finite samples. Lastly, and foremost, our testing procedure is useful to avoid serious forecasting/estimation problems in standard econometric practices with factors, like FAR and FAVAR models, especially if the number of factor is a priori determined and the factor loadings are subject to big breaks.

In the second step of our testing approach, a Sup-type test is used to detect a break of the parameters in that regression when the break date is assumed to be unknown. As the simulations show, this test does not perform very well especially when T is not too large $(T < 200)$. As other studies on the size of sup-type tests suggest, bootstrap can improve the finite sample performance of the test.compared to the tabulated asymptotic critical values of Andrews (1993). It is high in our research agenda to explore this possibility.

Further research is also needed if we were to allow for multiple breaks. As Breitung and Eickmeier point out, sequential estimation, as in Bai and Perron (1998), or an efficient search procedure, as in Bai and Perron (2003), for finding the candidate break dates may be employed.

Finally, while we only consider the case where structural breaks affect the factor loadings, it has been noted by Stock and Watson (2009) that there could be other sources of parameter instability stemming from breaks in the factor dynamics and/or in the idiosyncratic errors. Given the instability of the whole model, how to identify the instability of each of these components is an issue that also requires further investigation.

Appendix

A.1: Proof of Propositions 1 and 2

The proof procedes by showing that the errors, factors and loadings in model (6) satisfy Assumptions A to D of Bai and Ng (2002). Then, once this is shown, Propositions 1 and 2 just follow immediately from Theorems 1 and 2 of Bai and Ng (2002). Define $F_t^* = [F_t' \quad G_t^{1'}]'$, $\varepsilon_t = HG_2^t + e_t$, $\Gamma = [A \quad \Lambda]$.

Lemma 1. *$E||F_t^*||^4 < \infty$ and $T^{-1}\sum_{t=1}^T F_t^* F_t^{*'} \to \Sigma_F^*$ as $T \to \infty$ for some positive matrix Σ_F^*.*

Proof. $E||F_t^*||^4 < \infty$ follows from $E||F_t||^4 < \infty$ by Assumption 2 and the definition of G_t^1.

To prove the second part, we partition the matrix $\Sigma_F (= \lim_{T\to\infty} T^{-1}\sum_{t=1}^T F_t F_t')$ into:

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}$$

where $\Sigma_{11} = \lim_{T\to\infty} T^{-1}\sum_{t=1}^T F_t^1 F_t^{1'}$, $\Sigma_{22} = \lim_{T\to\infty} T^{-1}\sum_{t=1}^T F_t^2 F_t^{2'}$, $\Sigma_{12} = \lim_{T\to\infty} T^{-1}\sum_{t=1}^T F_t^1 F_t^{2'}$, and F_t^1 is the $k_1 \times 1$ subvector of F_t that has big breaks in their loadings, F_t^2 is the $k_2 \times 1$ subvector of F_t that doesn't have big breaks in their loadings. By the definition of F_t^* and G_t^1 we have:

$$T^{-1}\sum_{t=1}^T F_t^* F_t^{*'} = \begin{pmatrix} T^{-1}\sum_{t=1}^T F_t^1 F_t^{1'} & T^{-1}\sum_{t=1}^T F_t^1 F_t^{2'} & T^{-1}\sum_{t=\tau+1}^T F_t^1 F_t^{1'} \\ T^{-1}\sum_{t=1}^T F_t^2 F_t^{1'} & T^{-1}\sum_{t=1}^T F_t^2 F_t^{2'} & T^{-1}\sum_{t=\tau+1}^T F_t^2 F_t^{1'} \\ T^{-1}\sum_{t=\tau+1}^T F_t^1 F_t^{1'} & T^{-1}\sum_{t=\tau+1}^T F_t^1 F_t^{2'} & T^{-1}\sum_{t=\tau+1}^T F_t^1 F_t^{1'} \end{pmatrix}$$

By Assumption 2, the above matrix converges to

$$\Sigma_F^* = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & (1-\pi^*)\Sigma_{11} \\ \Sigma_{12}' & \Sigma_{22} & (1-\pi^*)\Sigma_{12}' \\ (1-\pi^*)\Sigma_{11} & (1-\pi^*)\Sigma_{12} & (1-\pi^*)\Sigma_{11} \end{pmatrix}$$

Moreover,

$$\det(\Sigma_F^*) = \det\begin{pmatrix} \Sigma_{11} & \Sigma_{12} & 0 \\ \Sigma_{12}' & \Sigma_{22} & (1-\pi^*)\Sigma_{12}' \\ 0 & 0 & \pi^*(1-\pi^*)\Sigma_{11} \end{pmatrix} = \det(\Sigma_F)\det(\pi^*(1-\pi^*)\Sigma_{11}) > 0$$

because Σ_F is positive definite by assumption. This completes the proof. □

Lemma 2. *$E||\Gamma_i||^4 < \infty$, and $N^{-1}\Gamma'\Gamma \to \Sigma_\Gamma$ as $N \to \infty$ for some positive definite matrix Σ_Γ.*

Proof. This follows directly from Assumptions 1.a and 3. □

The following lemmas involve the new errors ε_t. Let M and M^* denote some positive constants.

Lemma 3. *$E(\varepsilon_{it}) = 0$, $E|\varepsilon_{it}|^8 \leq M^*$*

Proof. This follows easily from $E|e_{it}|^8 \leq M$ (Assumption 4), $E(F_t) = 0$, $E||F_t||^4 < \infty$ (Assumption 2), and $\eta_i = o_p(1)$ (Assumption 1.b). □

Lemma 4. *$E(\varepsilon_s'\varepsilon_t/N) = E(N^{-1}\sum_{i=1}^N \varepsilon_{is}\varepsilon_{it}) = \gamma_N^*(s,t)$, $|\gamma_N^*(s,s)| \leq M^*$ for all s, and $\sum_{s=1}^T \gamma_N^*(s,t)^2 \leq M^*$ for all t and T.*

Proof.

$$\begin{aligned}
\gamma_N^*(s,t) &= N^{-1}\sum_{i=1}^N E(\varepsilon_{is}\varepsilon_{it}) \\
&= N^{-1}\sum_{i=1}^N E(e_{is}+\eta_i'G_s^2)E(e_{it}+\eta_i'G_t^2) \\
&= N^{-1}\sum_{i=1}^N \left[E(e_{is}e_{it}) + E(\eta_i'G_s^2\eta_i'G_t^2)\right] \\
&\leq N^{-1}\sum_{i=1}^N E(e_{is}e_{it}) + N^{-1}\sum_{i=1}^N \sqrt{E\left(\eta_i'G_s^2\right)^2 E\left(\eta_i'G_t^2\right)^2}
\end{aligned}$$

Since $N^{-1}\sum_{i=1}^N E(e_{is}e_{it}) = \gamma_N(s,t)$ by Assumption C of Bai and Ng (2002), and $E\left(\eta_i'G_t^2\right)^2 = O(\frac{1}{NT})$ for all t by Assumptions 1.b and 2, we have $\gamma_N^*(s,t) \leq \gamma_N(s,t) + O(\frac{1}{NT})$. Then

$$|\gamma_N^*(s,s)| \leq |\gamma_N(s,s)| + O(\frac{1}{NT}) \leq M^*$$

by Assumption C of Bai and Ng (2002). Moreover,

$$\begin{aligned}
\sum_{s=1}^T \gamma_N^*(s,t)^2 &\leq \sum_{s=1}^T \left(\gamma_N(s,t) + O(\frac{1}{NT})\right)^2 \\
&= \sum_{s=1}^T \gamma_N(s,t)^2 + O(\frac{1}{N}) \\
&\leq M + O(\frac{1}{N}) \leq M^*
\end{aligned}$$

by Assumption E.1 of Bai (2003). The proof is complete. □

Lemma 5. *$E(\varepsilon_{it}\varepsilon_{jt}) = \tau_{ij,t}^*$ with $|\tau_{ij,t}^*| \leq |\tau_{ij}^*|$ for some τ_{ij}^* and for all t; and $N^{-1}\sum_{i=1}^N\sum_{j=1}^N |\tau_{ij}^*| \leq M^*$.*

Proof. By Assumption C.3 of Bai and Ng (2002), $|\tau_{ij,t}| \leq |\tau_{ij}|$ for some τ_{ij} and all t, where $\tau_{ij,t} = E(e_{it}e_{jt})$. Then:

$$\begin{aligned}
|\hat{\tau}_{ij,t}| &= |E(\varepsilon_{it}\varepsilon_{jt})| \\
&= |E(e_{it}+\eta_i'G_t^2)(e_{jt}+\eta_j'G_t^2)| \\
&\leq |E(e_{it}e_{jt})| + \sqrt{E\left(\eta_i'G_s^2\right)^2 E\left(\eta_i'G_t^2\right)^2} \\
&\leq |\tau_{ij}| + O(\frac{1}{NT})
\end{aligned}$$

for all t. Therefore

$$\begin{aligned}
N^{-1}\sum_{i=1}^N\sum_{j=1}^N |\tau_{ij}^*| &\leq N^{-1}\sum_{i=1}^N\sum_{j=1}^N \left(|\tau_{ij}| + O(\frac{1}{NT})\right) \\
&\leq M + O(\frac{1}{T}) \\
&\leq M^*
\end{aligned}$$

by Assumption C.3 of Bai and Ng (2002). $\square$

Lemma 6. *$E(\varepsilon_{it}\varepsilon_{js}) = \tau^*_{ij,ts}$ and $(NT)^{-1}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\sum_{s=1}^T|\tau^*_{ij,ts}| \leq M^*$.*

Proof. By Assumption C.4 of Bai and Ng (2002), $(NT)^{-1}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\sum_{s=1}^T|\tau_{ij,ts}| \leq M$, where $E(e_{it}e_{js}) = \tau_{ij,ts}$. Then:

$$E(\varepsilon_{it}\varepsilon_{js}) = \tau^*_{ij,ts} = E(e_{it}e_{js}) + E(\eta_i' G_t^2 \eta_j' G_s^2) = \tau_{ij,ts} + E(\eta_i' G_t^2 \eta_j' G_s^2)$$

and we have

$$\begin{aligned}(NT)^{-1}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\sum_{t=1}^T|\tau^*_{ij,ts}| &\leq (NT)^{-1}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\sum_{t=1}^T|\tau_{ij,ts}| + (NT)^{-1}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\sum_{t=1}^T|E(\eta_i' G_t^2 \eta_j' G_s^2)| \\ &\leq M + O(1) \\ &\leq M^*\end{aligned}$$

following the same arguments as above.

$\square$

Lemma 7. *For every (t,s), $E|N^{-1/2}\sum_{i=1}^N[\varepsilon_{is}\varepsilon_{it} - E(\varepsilon_{is}\varepsilon_{it})]|^4 \leq M^*$.*

Proof. Since $\varepsilon_{it} = e_{it} + \eta_i' G_t^2 = e_{it} + O_p(\frac{1}{\sqrt{NT}})$, we have:

$$\begin{aligned}E|N^{-1/2}\sum_{i=1}^N[\varepsilon_{it}\varepsilon_{is} - E(\varepsilon_{it}\varepsilon_{is})]|^4 &= E|N^{-1/2}\sum_{i=1}^N[e_{it}e_{is} - E(e_{it}e_{is}) + Op(\frac{1}{\sqrt{NT}}) + O(\frac{1}{\sqrt{NT}})]|^4 \\ &= E|N^{-1/2}\sum_{i=1}^N[e_{it}e_{is} - E(e_{it}e_{is})] + Op(\frac{1}{\sqrt{T}}) + O(\frac{1}{\sqrt{T}})|^4 \\ &\leq M + O(\frac{1}{\sqrt{T}}) \\ &\leq M^*\end{aligned}$$

$\square$

Lemma 8. *$E\left(\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T F_t^*\varepsilon_{it}\right\|^2\right) \leq M^*$.*

Proof. By the definition of ε_{it} we have:

$$E\Big(\frac{1}{N}\sum_{i=1}^N\Big\|\frac{1}{\sqrt{T}}\sum_{t=1}^T F_t^*\varepsilon_{it}\Big\|^2\Big) \leq E\Big(\frac{1}{N}\sum_{i=1}^N\Big\|\frac{1}{\sqrt{T}}\sum_{t=1}^T F_t^* e_{it}\Big\|^2\Big) + E\Big(\frac{1}{N}\sum_{i=1}^N\Big\|\frac{1}{\sqrt{T}}\sum_{t=1}^T F_t^*\eta_i' G_t^2\Big\|^2\Big)$$

then by the definition of F_t^* and G_t^2,

$$\Big\|\frac{1}{\sqrt{T}}\sum_{t=1}^T F_t^* e_{it}\Big\|^2 = \Big\|\frac{1}{\sqrt{T}}\sum_{t=1}^T F_t e_{it}\Big\|^2 + \Big\|\frac{1}{\sqrt{T}}\sum_{t=\tau+1}^T F_t^1 e_{it}\Big\|^2$$

$$\Big\|\frac{1}{\sqrt{T}}\sum_{t=1}^T F_t^*\eta_i' G_t^2\Big\|^2 = \Big\|\frac{1}{\sqrt{T}}\sum_{t=1}^T F_t\eta_i' F_t^2\Big\|^2 + \Big\|\frac{1}{\sqrt{T}}\sum_{t=1}^T F_t^1\eta_i' F_t^2\Big\|^2$$

Therefore, by Assumptions 1.b, 2 and 5, it follows easily that the first part of the right hand side of the last inequality is $O(1)$ and the second part is $O(\frac{1}{N})$. Thus the proof is complete. $\square$

Once we have proved that the new factors: F_t^*, the new loadings: Γ and the new errors: ε_t all satisfy the necessary conditions of Bai and Ng (2002), Propositions 1 and 2 just follow directly from theirTheorems 1 and 2, with r replaced by $r + k_1$ and F_t replaced by F_t^*.

A.2: Proof of Theorem 1

Let $\hat{F}_t$ define the $r \times 1$ vector of estimated factors. Under the null: $k_1 = 0$, when $\bar{r} = r$ we have

$$\hat{F}_t = DF_t + o_p(1)$$

let $D_{(i\cdot)}$ denote the ith row of D, and $D_{(\cdot j)}$ denote the jth column of D. Define $\hat{\mathscr{F}}_t = DF_t$, and $\hat{\mathscr{F}}_{kt} = D_{(k\cdot)} \times F_t$ as the kth element of $\hat{\mathscr{F}}_t$. Let $\hat{F}_{1t}$ be the first element of $\hat{F}_t$, and $\hat{F}_{-1t} = [\hat{F}_{2t} \cdots \hat{F}_{rt}]$. $\hat{\mathscr{F}}_{1t}$ and $\hat{\mathscr{F}}_{-1t}$ can be defined in the same way. Note that $\hat{\mathscr{F}}_t$ depends on N and T because $D = (\hat{F}F/T)(A'A/N)$ (see Bai and Ng (2002)).

Lemma 9.

$$\sup_{\pi \in \Pi} \left\| \frac{1}{T} \sum_{t=1}^{T\pi} (\hat{F}_t - \hat{\mathscr{F}}_t) F_t' \right\| = O_p(\delta_{N,T}^{-2})$$

Proof. Following Bai (2003) we have:

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T\pi}(\hat{F}_t - \hat{\mathscr{F}}_t)F_t' &= T^{-2}\sum_{t=1}^{T\pi}\sum_{s=1}^{T}\hat{F}_s F_t' \gamma_N(s,t) + T^{-2}\sum_{t=1}^{T\pi}\sum_{s=1}^{T}\hat{F}_s F_t' \zeta_{st} + T^{-2}\sum_{t=1}^{T\pi}\sum_{s=1}^{T}\hat{F}_s F_t' \kappa_{st} + T^{-2}\sum_{t=1}^{T\pi}\sum_{s=1}^{T}\hat{F}_s F_t' \xi_{st} \\
&= I + II + III + IV
\end{aligned}$$

where

$$\begin{aligned}
\zeta_{st} &= \frac{e_s' e_t}{N} - \gamma_N(s,t). \\
\kappa_{st} &= F_s' A' e_t / N. \\
\xi_{st} &= F_t' A' e_s / N.
\end{aligned}$$

First, note that:

$$I = T^{-2}\sum_{t=1}^{T\pi}\sum_{s=1}^{T}(\hat{F}_s - DF_s)F_t'\gamma_N(s,t) + T^{-2}D\sum_{t=1}^{T\pi}\sum_{s=1}^{T}F_s F_t' \gamma_N(s,t)$$

Consider the first part of the right hand side, we have

$$\begin{aligned}
& \left\| T^{-2}\sum_{t=1}^{T\pi}\sum_{s=1}^{T}(\hat{F}_s - DF_s)F_t'\gamma_N(s,t) \right\| \\
= & \left\| T^{-2}\sum_{s=1}^{T}\left((\hat{F}_s - DF_s)\sum_{t=1}^{T\pi}F_t'\gamma_N(s,t)\right) \right\| \\
\leq & T^{-1/2}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s - DF_s\right\|^2}\sqrt{\frac{1}{T}\sum_{t=1}^{T\pi}\left\|F_t\right\|^2}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T\pi}\gamma_N(s,t)^2} \\
= & T^{-1/2}O_p(\delta_{N,T}^{-1})O_p(1)
\end{aligned}$$

because $\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s - DF_s\right\|^2$ is $O_p(\delta_{N,T}^{-2})$ by Theorem 1 of Bai and Ng (2002), $\sum_{t=1}^{T\pi}\left\|F_t\right\|^2$ is $O_p(1)$ by Assumption 2, and $\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T\pi}\gamma_N(s,t)^2 \leq \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma_N(s,t)^2 = O_p(1)$ by Lemma 1(i) of Bai and Ng (2002).

For the second part, note that:

$$
\begin{aligned}
& \left\| T^{-2} D \sum_{t=1}^{T\pi} \sum_{s=1}^{T} F_s F_t' \gamma_N(s,t) \right\| \\
\leq \quad & T^{-2} \|D\| \sum_{t=1}^{T} \sum_{s=1}^{T} \left\| F_s F_t' \right\| |\gamma_N(s,t)| \\
\leq \quad & O_p(1) T^{-1} \Big(\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} |\gamma_N(s,t)| \Big) \\
= \quad & O_p(T^{-1})
\end{aligned}
$$

because $\|D\|$, $\left\|F_s F_t'\right\|$ and $\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}|\gamma_N(s,t)|$ are all $O_p(1)$ from Bai and Ng (2002) and Assumptions 2 and 4. Therefore, we have

$$
\sup_{\pi \in \Pi} \|I\| = O_p\Big(\frac{1}{\delta_{N,T}\sqrt{T}} \Big). \tag{A.1}
$$

Next, II can be written as:

$$
T^{-2} \sum_{t=1}^{T\pi} \sum_{s=1}^{T} (\hat{F}_s - DF_s) F_t' \zeta_{st} + T^{-2} D \sum_{t=1}^{T\pi} \sum_{s=1}^{T} F_s F_t' \zeta_{st}
$$

Similarly, we have

$$
\begin{aligned}
& \left\| T^{-2} \sum_{t=1}^{T\pi} \sum_{s=1}^{T} (\hat{F}_s - DF_s) F_t' \zeta_{st} \right\| \\
\leq \quad & \sqrt{\frac{1}{T} \sum_{s=1}^{T} \left\| \hat{F}_s - DF_s \right\|^2} \sqrt{\frac{1}{T} \sum_{t=1}^{T\pi} \left\| F_t \right\|^2} \sqrt{\frac{1}{T^2} \sum_{s=1}^{T} \sum_{t=1}^{T\pi} \zeta_{st}^2} \\
\leq \quad & \sqrt{\frac{1}{T} \sum_{s=1}^{T} \left\| \hat{F}_s - DF_s \right\|^2} \sqrt{\frac{1}{T} \sum_{t=1}^{T} \left\| F_t \right\|^2} \sqrt{\frac{1}{T^2} \sum_{s=1}^{T} \sum_{t=1}^{T} \zeta_{st}^2} \\
= \quad & O_p\Big(\frac{1}{\delta_{N,T}\sqrt{N}} \Big)
\end{aligned}
$$

because $\zeta_{st} = N^{-1}\sum_{i=1}^{N}[e_{it}e_{is} - E(e_{it}e_{is})]$ is $O_p(1/\sqrt{N})$ by Assumption C.5 of Bai (2003). For the second term, we can write:

$$
T^{-2} D \sum_{t=1}^{T\pi} \sum_{s=1}^{T} F_s F_t' \zeta_{st} = \frac{1}{\sqrt{NT}} \frac{1}{T} D \sum_{t=1}^{T\pi} q_t F_t'
$$

where

$$
q_t = \frac{1}{\sqrt{NT}} \sum_{s=1}^{T} \sum_{i=1}^{N} [e_{it}e_{is} - E(e_{it}e_{is})] F_s
$$

Since $E\|q_t\|^2 < M$ by Assumption F.1 of Bai (2003), we have

$$\begin{aligned}
&\sup_{\pi\in\Pi}\Big\|T^{-2}D\sum_{t=1}^{T\pi}\sum_{s=1}^{T}F_sF_t'\zeta_{st}\Big\| \\
= \quad & \frac{1}{\sqrt{NT}}\sup_{\pi\in\Pi}\Big\|T^{-1}D\sum_{t=1}^{T\pi}q_tF_t'\Big\| \\
\leq \quad & \frac{1}{\sqrt{NT}}\|D\|\sup_{\pi\in\Pi}\Big\|\sqrt{\frac{1}{T}\sum_{t=1}^{T\pi}\|q_t\|^2}\sqrt{\frac{1}{T}\sum_{t=1}^{T\pi}\|F_t\|^2} \\
\leq \quad & O_p(1)\frac{1}{\sqrt{NT}}\Big\|\sqrt{\frac{1}{T}\sum_{t=1}^{T}\|q_t\|^2}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\|F_t\|^2} \\
= \quad & O_p\Big(\frac{1}{\sqrt{NT}}\Big)
\end{aligned}$$

Then it follows that

$$\sup_{\pi\in\Pi}\|II\| = O_p\Big(\frac{1}{\delta_{N,T}\sqrt{N}}\Big). \tag{A.2}$$

III can be written as:

$$III = T^{-2}\sum_{t=1}^{T\pi}\sum_{s=1}^{T}(\hat{F}_s - DF_s)F_t'\kappa_{st} + T^{-2}D\sum_{t=1}^{T\pi}\sum_{s=1}^{T}F_sF_t'\kappa_{st}$$

and the second part on the right hand side can be written as

$$D\Big(\frac{1}{T}\sum_{s=1}^{T}F_sF_s'\Big)\frac{1}{NT}\sum_{t=1}^{T\pi}\sum_{i=1}^{N}\alpha_iF_t'e_{it}$$

therefore:

$$\begin{aligned}
&\sup_{\pi\in\Pi}\Big\|T^{-2}D\sum_{t=1}^{T\pi}\sum_{s=1}^{T}F_sF_t'\kappa_{st}\Big\| \\
\leq \quad & \frac{1}{\sqrt{NT}}\|D\|\Big\|\frac{1}{T}\sum_{s=1}^{T}F_sF_s'\Big\|\sup_{\pi\in\Pi}\Big\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T\pi}\sum_{i=1}^{N}\alpha_iF_t'e_{it}\Big\| \\
= \quad & O_p\Big(\frac{1}{\sqrt{NT}}\Big)
\end{aligned}$$

by Assumption 8.

For the first part on the right hand side, we have

$$
\begin{aligned}
&\sup_{\pi\in\Pi}\Big\|T^{-2}\sum_{t=1}^{T\pi}\sum_{s=1}^{T}(\hat{F}_s-DF_s)F_t'\kappa_{st}\Big\| \\
\leq\ & \sqrt{\frac{1}{T}\sum_{s=1}^{T}\big\|\hat{F}_s-DF_s\big\|^2}\sup_{\pi\in\Pi}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\Big\|\frac{1}{T}\sum_{t=1}^{T\pi}F_t'\kappa_{st}\Big\|^2} \\
=\ & O_p(\delta_{N,T}^{-1})\frac{1}{\sqrt{NT}}\sup_{\pi\in\Pi}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\Big\|F_s'\frac{1}{\sqrt{NT}}\sum_{t=1}^{T\pi}\sum_{i=1}^{N}\alpha_iF_t'e_{it}\Big\|^2} \\
\leq\ & O_p(\delta_{N,T}^{-1})\frac{1}{\sqrt{NT}}\sqrt{\frac{1}{T}\sum_{s=1}^{T}\Big\|F_s\Big\|^2}\sup_{\pi\in\Pi}\sqrt{\Big\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T\pi}\sum_{i=1}^{N}\alpha_iF_t'e_{it}\Big\|^2} \\
=\ & O_p\Big(\frac{1}{\delta_{N,T}}\frac{1}{\sqrt{NT}}\Big)
\end{aligned}
$$

by Assumption 8. Thus,

$$\sup_{\pi\in\Pi}\|III\| = O_p\Big(\frac{1}{\sqrt{NT}}\Big). \tag{A.3}$$

It can also be proved in the similar way that

$$\sup_{\pi\in\Pi}\|IV\| = O_p\Big(\frac{1}{\sqrt{NT}}\Big). \tag{A.4}$$

Finally we have:

$$
\begin{aligned}
&\sup_{\pi\in\Pi}\Big\|\frac{1}{T}\sum_{t=1}^{T\pi}(\hat{F}_t-\hat{\mathscr{F}}_t)F_t'\Big\| \leq \sup_{\pi\in\Pi}\|I\|+\sup_{\pi\in\Pi}\|II\|+\sup_{\pi\in\Pi}\|III\|+\sup_{\pi\in\Pi}\|IV\| \\
=\ & O_p\Big(\frac{1}{\sqrt{T}\delta_{N,T}}\Big)+O_p\Big(\frac{1}{\sqrt{N}\delta_{N,T}}\Big)+O_p\Big(\frac{1}{\sqrt{NT}}\Big)=O_p\Big(\frac{1}{\delta_{N,T}^2}\Big)
\end{aligned}
$$

□

Lemma 10.

$$\sup_{\pi\in\Pi}\Big\|\frac{1}{T}\sum_{t=1}^{T\pi}\hat{F}_t\hat{F}_t'-\frac{1}{T}\sum_{t=1}^{T\pi}\hat{\mathscr{F}}_t\hat{\mathscr{F}}_t'\Big\| = O_p(\delta_{N,T}^{-2})$$

Proof. Note that:

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T\pi}\hat{F}_t\hat{F}_t'-\frac{1}{T}\sum_{t=1}^{T\pi}\hat{\mathscr{F}}_t\hat{\mathscr{F}}_t' \\
=\ & \frac{1}{T}\sum_{t=1}^{T\pi}\hat{F}_t\hat{F}_t'-\frac{1}{T}\sum_{t=1}^{T\pi}(DF_t)(F_t'D') \\
=\ & \frac{1}{T}\sum_{t=1}^{T\pi}\hat{F}_t(\hat{F}_t'-F_t'D')+\frac{1}{T}\sum_{t=1}^{T\pi}(\hat{F}_t-DF_t)(F_t'D') \\
=\ & \frac{1}{T}\sum_{t=1}^{T\pi}(\hat{F}_t-DF_t)(\hat{F}_t-DF_t)'+\frac{1}{T}D\sum_{t=1}^{T\pi}F_t(\hat{F}_t-DF_t)'+\frac{1}{T}\sum_{t=1}^{T\pi}(\hat{F}_t-DF_t)(F_t'D')
\end{aligned}
$$

Thus:

$$
\begin{aligned}
& \sup_{\pi\in\Pi}\left\|\frac{1}{T}\sum_{t=1}^{T\pi}\hat{F}_t\hat{F}_t' - \frac{1}{T}\sum_{t=1}^{T\pi}\hat{\mathscr{F}}_t\hat{\mathscr{F}}_t'\right\| \\
\leq\ & \sup_{\pi\in\Pi}\left\|\frac{1}{T}\sum_{t=1}^{T\pi}(\hat{F}_t - DF_t)(\hat{F}_t - DF_t)'\right\| + 2\|D\|\sup_{\pi\in\Pi}\left\|\frac{1}{T}\sum_{t=1}^{T\pi}(\hat{F}_t - DF_t)F_t'\right\| \\
\leq\ & \frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_t - DF_t\right\|^2 + 2\|D\|\sup_{\pi\in\Pi}\left\|\frac{1}{T}\sum_{t=1}^{T\pi}(\hat{F}_t - DF_t)F_t'\right\|
\end{aligned}
$$

since $\left\|\hat{F}_t - DF_t\right\| = O_p(\delta_{N,T}^{-1})$ and $\sup_{\pi\in\Pi}\left\|\frac{1}{T}\sum_{t=1}^{T\pi}(\hat{F}_t - DF_t)F_t'\right\|$ is $O_p(\delta_{N,T}^{-2})$ by Lemma 9, the proof is complete.

□

The following two lemmas follow from Lemma 10 and Assumption 6:

Lemma 11.

$$\sup_{\pi\in\Pi}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\hat{F}_{-1t}\hat{F}_{1t} - \frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\hat{\mathscr{F}}_{-1t}\hat{\mathscr{F}}_{1t}\right\| = o_p(1)$$

Proof. See Lemma 10 and Assumption 6. □

Lemma 12.

$$\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{\mathscr{F}}_{-1t}\hat{\mathscr{F}}_{1t}'\right\| = o_p(1)$$

Proof. By construction we have $\frac{1}{T}\sum_{t=1}^{T}\hat{F}_{-1t}\hat{F}_{1t}' = 0$, then the result follows from Lemma 11. □

Let $\Rightarrow$ denote *weak convergence*, $D^* = Q\Sigma_A$, where $Q = \lim\frac{\hat{F}'F}{T}$(See proposition 1 of Bai (2003)), $\Sigma_A = \lim\frac{A'A}{N}$. And define $\mathscr{F}_t = D^*F_t$, $S = \lim\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T}\mathscr{F}\mathscr{F}'\right)$. Then:

Lemma 13.

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\left(\mathscr{F}_{-1t}\mathscr{F}_{1t} - E(\mathscr{F}_{-1t}\mathscr{F}_{1t})\right) \Rightarrow S^{1/2}\mathscr{W}_{r-1}(\pi)$$

for $\pi\in[0,1]$.

Proof. $\mathscr{F}_{-1t}\mathscr{F}_{1t}$ is stationary and ergodic because F_t is stationary and ergodic by Assumption 7. First, we show that $\{\mathscr{F}_{kt}\mathscr{F}_{1t} - E(\mathscr{F}_{kt}\mathscr{F}_{1t}), \Omega_t\}$ is an adapted mixingale of size -1 for $k = 2,\ldots,r$. By definition, we have $\mathscr{F}_{kt}\mathscr{F}_{1t} = (D^*_{(k\cdot)}F_t)(D^*_{(1\cdot)}F_t) = \left(\sum_{p=1}^{r}D^*_{kp}F_{pt}\right)\left(\sum_{p=1}^{r}D^*_{1p}F_{pt}\right) = \sum_{h=1}^{r}\sum_{p=1}^{r}D^*_{kp}D^*_{1h}F_{pt}F_{ht}$, and $\mathscr{F}_{kt}\mathscr{F}_{1t} - E(\mathscr{F}_{kt}\mathscr{F}_{1t}) = \sum_{h=1}^{r}\sum_{p=1}^{r}D^*_{kp}D^*_{1h}(F_{pt}F_{ht} - E(F_{pt}F_{ht})) =$

$\sum_{h=1}^{r}\sum_{p=1}^{r} D_{kp}^{*} D_{1h}^{*} Y_{hp,t}$. Thus:

$$
\begin{aligned}
& \sqrt{E\Big(E\big(\mathscr{F}_{kt}\mathscr{F}_{1t}-E(\mathscr{F}_{kt}\mathscr{F}_{1t})|\Omega_{t-m}\big)\Big)^2} \\
= \ & \sqrt{E\Big(\sum_{h=1}^{r}\sum_{p=1}^{r} D_{kp}^{*} D_{1h}^{*} E(Y_{hp,t}|\Omega_{t-m})\Big)^2} \\
\leq \ & \sum_{h=1}^{r}\sum_{p=1}^{r} |D_{kp}^{*} D_{1h}^{*}| \sqrt{E\big(E(Y_{hp,t}|\Omega_{t-m})\big)^2} \\
\leq \ & \Delta \sum_{h=1}^{r}\sum_{p=1}^{r} c_t^{hp} \gamma_m^{hp} \\
\leq \ & \Delta r^2 \max(c_t^{hp}) \max(\gamma_m^{hp})
\end{aligned}
$$

since $\max(\gamma_m^{hp})$ is $O(m^{-1-\delta})$ for some $\delta > 0$ by Assumption 7, we conclude that $\{\mathscr{F}_{kt}\mathscr{F}_{1t} - E(\mathscr{F}_{kt}\mathscr{F}_{1t}), \Omega_t\}$ is an adapted mixingale of size -1 for $k = 2,\dots,r$.

Next, we proof the weak convergence using the Crame-Rao device. Define

$$
z_t = a' S^{-1/2}\big(\mathscr{F}_{-1t}\mathscr{F}_{1t} - E(\mathscr{F}_{-1t}\mathscr{F}_{1t})\big)
$$

where $a \in \mathbb{R}^{r-1}$, and $a'a = 1$. Note that

$$
z_t = \sum_{k=2}^{r} \tilde{a}_k [\mathscr{F}_{kt}\mathscr{F}_{1t} - E(\mathscr{F}_{kt}\mathscr{F}_{1t})]
$$

where $\tilde{a}_k$ is the $k-1$th element of $a'S^{-1/2}$.

$$
\begin{aligned}
E(z_t^2) \ & \leq \ \left(\sum_{k=2}^{r}\sqrt{E\Big(\tilde{a}_k[\mathscr{F}_{kt}\mathscr{F}_{1t} - E(\mathscr{F}_{kt}\mathscr{F}_{1t})]\Big)^2}\right)^2 \\
& \leq \ \Delta\left(\sum_{k=2}^{r}\sqrt{E\Big(\mathscr{F}_{kt}\mathscr{F}_{1t}\Big)^2 - \Big(E(\mathscr{F}_{kt}\mathscr{F}_{1t})\Big)^2}\right)^2 \leq M
\end{aligned}
$$

because $E\|F_t\|^4 < \infty$ and $\mathscr{F}_{kt} = D_{k\cdot}^{*} F_t$. Moreover, z_t is stationary and ergodic, and $\{z_t, \Omega_t\}$ is an adapted mixingale sequence of size -1, because:

$$
\begin{aligned}
\sqrt{E\Big(E(z_t|\Omega_{t-m})\Big)^2} \ & = \ \sqrt{E\left(\sum_{k=2}^{r}\tilde{a}_k E\Big(\mathscr{F}_{kt}\mathscr{F}_{1t} - E(\mathscr{F}_{kt}\mathscr{F}_{1t})|\Omega_{t-m}\Big)\right)^2} \\
& \leq \ \sum_{k=2}^{r}|\tilde{a}_k|\sqrt{E\left(E\Big(\mathscr{F}_{kt}\mathscr{F}_{1t} - E(\mathscr{F}_{kt}\mathscr{F}_{1t})|\Omega_{t-m}\Big)^2\right)} \\
& \leq \ \max(|\tilde{a}_k|)\sum_{k=2}^{r}\tilde{c}_t^k \tilde{\gamma}_m^k
\end{aligned}
$$

by the results above we known $\tilde{\gamma}_m^k$ is $O(m-1-\delta)$ for $k = 2,\dots,r$, it follows that $\{z_t, \Omega_t\}$ is an adapted mixingale sequence of size -1. Then it follows from Theorem 7.17 of White (2001) that:

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi} z_t = a'S^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\big(\mathscr{F}_{-1t}\mathscr{F}_{1t} - E(\mathscr{F}_{-1t}\mathscr{F}_{1t})\big) \Rightarrow \mathscr{W}(\pi)
$$

Moreover, it can be proved that:

$$a_1'\frac{1}{\sqrt{T}}\sum_{t=T\pi_1}^{T\pi_2}\left(\mathscr{F}_{-1t}\mathscr{F}_{1t}-E(\mathscr{F}_{-1t}\mathscr{F}_{1t})\right)+a_2'\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi_0}\left(\mathscr{F}_{-1t}\mathscr{F}_{1t}-E(\mathscr{F}_{-1t}\mathscr{F}_{1t})\right)\rightsquigarrow\mathscr{N}(0,(\pi_2-\pi_1)a_1'Sa_1+\pi_0a_2'Sa_2)$$

by using Corrollary 3.1 of Woodridge and White (1988). The proof is complete by using Lemma A.4 of Andrews (1993).

□

Lemma 14.

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\hat{F}_{-1t}\hat{F}_{1t}\Rightarrow S^{1/2}\mathscr{B}^0_{r-1}(\pi)$$

for $\pi\in\Pi$, where $\mathscr{B}^0_{r-1}(\pi)=\mathscr{W}_{r-1}(\pi)-\pi\mathscr{W}_{r-1}(1)$ is a vector of Brownian Bridge.

Proof. If we show that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\left[\mathscr{F}_{-1t}\mathscr{F}_{1t}-T^{-1}\sum_{s=1}^{s}\mathscr{F}_{-1s}\mathscr{F}_{1s}\right]\Rightarrow S^{1/2}\mathscr{B}^0_{r-1}(\pi)\tag{A.5}$$

and

$$\sup_{\pi\in\Pi}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\hat{\mathscr{F}}_{-1t}\hat{\mathscr{F}}_{1t}-\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\left[\mathscr{F}_{-1t}\mathscr{F}_{1t}-T^{-1}\sum_{s=1}^{s}\mathscr{F}_{-1s}\mathscr{F}_{1s}\right]\right\|=o_p(1)\tag{A.6}$$

then the result follows from Lemma 11.

First note that

$$\begin{aligned}&\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\left[\mathscr{F}_{-1t}\mathscr{F}_{1t}-T^{-1}\sum_{s=1}^{s}\mathscr{F}_{-1s}\mathscr{F}_{1s}\right]\\ =\;&\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\left(\mathscr{F}_{-1t}\mathscr{F}_{1t}-E(\mathscr{F}_{-1t}\mathscr{F}_{1t})\right)+\frac{1}{T}\sum_{t=1}^{T\pi}\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\left(\mathscr{F}_{-1s}\mathscr{F}_{1s}-E(\mathscr{F}_{-1s}\mathscr{F}_{1s})\right)\right)\end{aligned}$$

hence A.5 can be verified by applying Lemma 13.

To prove A.6, we first define D_{-1} as the second to last rows of D, and D_1 as the first row of D. Then we have

$$\hat{\mathscr{F}}_{-1t}\hat{\mathscr{F}}_{1t}=D_{-1}F_tF_t'D_1'$$

and

$$\mathscr{F}_{-1t}\mathscr{F}_{1t}=D^*_{-1}F_tF_t'D_1^{*'}$$

it follows that:

$$\begin{aligned}&\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\left(\hat{\mathscr{F}}_{-1t}\hat{\mathscr{F}}_{1t}-\mathscr{F}_{-1t}\mathscr{F}_{1t}\right)\\ =\;&\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\left(D_{-1}F_tF_t'D_1'-D_{-1}F_tF_t'D_1^*+D_{-1}F_tF_t'D_1^*-D^*_{-1}F_tF_t'D_1^*\right)\\ =\;&D_{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}F_tF_t'\right)\left(D_1'-D_1^{*'}\right)+\left(D_{-1}-D^*_{-1}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}F_tF_t'\right)D_1^{*'}\end{aligned}$$

Next, define $\overline{\mathscr{F}_{-1t}\mathscr{F}_{1t}} = \frac{1}{T}\sum_{s=1}^{T}\mathscr{F}_{-1s}\mathscr{F}_{1s}$, and $\overline{F_{-1t}F_{1t}} = \frac{1}{T}\sum_{s=1}^{T}F_{-1s}F_{1s}$, then:

$$
\begin{aligned}
& \frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\left(T^{-1}\sum_{s=1}^{s}\mathscr{F}_{-1s}\mathscr{F}_{1s}\right) \\
= \; & D_{-1}^{*}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\overline{F_{-1t}F_{1t}}\right)D_{1}^{*'} \\
= \; & D_{-1}^{*}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\overline{F_{-1t}F_{1t}}\right)D_{1}^{*'} - D_{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\overline{F_{-1t}F_{1t}}\right)D_{1}^{*'} + D_{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\overline{F_{-1t}F_{1t}}\right)D_{1}^{*'} \\
& - D_{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\overline{F_{-1t}F_{1t}}\right)D_{1}' + D_{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\overline{F_{-1t}F_{1t}}\right)D_{1}' \\
= \; & \left(D_{-1}^{*} - D_{-1}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\overline{F_{-1t}F_{1t}}\right)D_{1}^{*'} + D_{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\overline{F_{-1t}F_{1t}}\right)\left(D_{1}^{*'} - D_{1}'\right) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\left(\frac{1}{T}\sum_{s=1}^{T}\hat{\mathscr{F}}_{-1s}\hat{\mathscr{F}}_{1s}\right)
\end{aligned}
$$

Combining the above results gives:

$$
\begin{aligned}
& \frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\hat{\mathscr{F}}_{-1t}\hat{\mathscr{F}}_{1t} - \frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\left[\mathscr{F}_{-1t}\mathscr{F}_{1t} - T^{-1}\sum_{s=1}^{s}\mathscr{F}_{-1s}\mathscr{F}_{1s}\right] \\
= \; & \frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\left(\hat{\mathscr{F}}_{-1t}\hat{\mathscr{F}}_{1t} - \mathscr{F}_{-1t}\mathscr{F}_{1t}\right) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\left(T^{-1}\sum_{s=1}^{s}\mathscr{F}_{-1s}\mathscr{F}_{1s}\right) \\
= \; & D_{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\left(F_tF_t' - \overline{F_{-1t}F_{1t}}\right)\right)\left(D_1' - D_1^{*'}\right) + \left(D_{-1} - D_{-1}^{*}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\left(F_tF_t' - \overline{F_{-1t}F_{1t}}\right)\right)D_1^{*'} \\
& + \frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\left(\frac{1}{T}\sum_{s=1}^{T}\hat{\mathscr{F}}_{-1s}\hat{\mathscr{F}}_{1s}\right)
\end{aligned}
$$

Following the similar arguments of Lemma 13, we can prove that

$$
\sup_{\pi\in\Pi}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T\pi}\left(F_tF_t' - \overline{F_{-1t}F_{1t}}\right)\right\| = O_p(1).
$$

Moreover, it is easy to see that $\|D\| = O_p(1)$ and $\|D - D^{*}\| = o_p(1)$. Finally, $\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\hat{\mathscr{F}}_{-1s}\hat{\mathscr{F}}_{1s}\right\|$ is $o_p(1)$ by Lemma 12. Then A.6 follows easily and the proof is complete. □

Theorem 1:

Proof. First note that $\lim\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T}\hat{F}_{-1t}\hat{F}_{1t}\right) = S$ because $\frac{1}{T}\sum_{t=1}^{T}\hat{F}_{-1t}\hat{F}_{1t} - \frac{1}{T}\sum_{t=1}^{T}\mathscr{F}_{-1t}\mathscr{F}_{1t} = o_p(1)$ and $E\|F_t\|^4 < \infty$. Then Theorem 1 follows from Assumption 9, Lemma 14, and Continuous Mapping Theorem. □

References

Andrews, D. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*(4), 821–856.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bai, J. and S. Ng (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica 74*(4), 1133–1150.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*(1), 47–78.

Bai, J. and P. Perron (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics 18*(1), 1–22.

Banerjee, A., M. Marcellino, and I. Masten (2008). *Forecasting macroeconomic variables using diffusion indexes in short samples with structural change*. CEPR DP. 6706.

Bernanke, B., J. Boivin, and P. Eliasz (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics 120*(1), 387.

Breitung, J. and S. Eickmeier (2010). Testing for structural breaks in dynamic factor models. *Journal of Econometrics, forthcoming*.

Diebold, F. and C. Chen (1996). Testing structural stability with endogenous breakpoint a size comparison of analytic and bootstrap procedures. *Journal of Econometrics 70*(1), 221–241.

Forni, M. and M. Lippi (2001). The generalized dynamic factor model: representation theory. *Econometric Theory 17*(06), 1113–1141.

Giannone, D. (2007). Discussion: "forecasting in dynamic factor models subject to structural instability" by stock and watson. In *New Developments in Dynamic Factor Modelling*. Bank of England, Centre for Central Banking Studies.

Newey, W. and K. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Rudebusch, G. and T. Wu (2008). A macro-finance model of the term structure, monetary policy and the economy. *The Economic Journal 118*(530), 906–926.

Stock, J. and M. Watson (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*(460), 1167–1179.

Stock, J. and M. Watson (2009). Forecasting in dynamic factor models subject to structural instability. *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry*, 173–205.

White, H. (2001). *Asymptotic theory for econometricians*. Academic Press Orlando, Florida.

Wooldridge, J. and H. White (1988). Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory 4*(2), 210–230.

Table 1: Size study, 3 factors

N	T	$\hat{\alpha}_{0.05}\|\bar{r}=2$				$\hat{\alpha}_{0.05}\|\bar{r}=3$				$\hat{\alpha}_{0.05}\|\bar{r}=4$			
		LM	Sup LM	Wald	Sup Wald	LM	Sup LM	Wald	Sup Wald	LM	Sup LM	Wald	Sup Wald
100	100	5.0	1.0	5.9	4.8	2.3	0.2	4.2	6.7	0.5	0.2	1.3	11.6
100	150	5.0	1.9	4.9	3.1	3.5	0.7	3.7	4.8	1.1	0.3	1.9	7.0
100	200	5.7	2.7	5.0	4.0	4.9	1.8	4.0	3.5	3.0	0.5	2.9	3.9
100	250	5.3	3.2	5.3	3.9	4.4	1.8	4.7	3.2	2.3	0.9	3.4	3.1
100	300	6.2	4.5	6.7	4.0	5.3	2.0	5.1	3.4	3.8	1.1	4.7	3.9
150	100	5.3	1.2	5.9	5.1	2.6	0.2	4.0	7.9	0.8	0.2	2.3	12.9
150	150	5.9	1.8	5.2	4.0	2.9	0.5	3.4	4.0	1.3	0.3	2.7	6.1
150	200	5.5	2.6	6.2	4.5	3.5	1.2	5.1	3.4	2.3	0.9	3.0	4.3
150	250	6.0	2.9	6.9	3.8	3.5	1.6	5.7	3.1	3.2	0.5	3.6	4.7
150	300	5.8	3.7	6.3	4.4	3.9	2.5	5.1	4.0	3.5	1.3	4.0	3.7
200	100	4.6	1.1	5.4	5.0	2.3	0.1	3.0	8.6	0.4	0.4	1.5	15.6
200	150	4.7	2.3	5.6	3.2	2.8	0.2	3.7	4.3	1.2	0.1	2.7	5.6
200	200	5.4	3.0	5.1	2.9	4.0	1.6	3.4	2.5	2.6	1.3	3.2	3.5
200	250	6.2	3.7	7.0	4.0	3.8	2.0	6.8	4.1	2.4	1.1	4.1	5.2
200	300	5.3	3.1	5.5	4.6	3.2	1.5	3.5	4.0	3.4	1.3	2.6	4.5
250	100	5.2	0.8	7.4	5.1	2.1	0.4	4.5	7.0	0.6	0.2	3.5	12.9
250	150	4.1	2.5	5.7	3.6	2.9	0.5	3.9	4.2	1.6	0.0	2.4	6.4
250	200	5.3	2.6	6.5	4.9	3.5	0.8	4.6	5.0	2.9	0.3	3.4	5.2
250	250	5.3	3.1	6.2	4.3	4.7	1.8	5.6	3.1	4.0	0.7	3.5	3.6
250	300	5.5	4.0	5.1	3.7	4.3	1.5	4.0	3.3	3.4	1.4	2.9	3.7
300	100	4.7	0.6	5.2	5.4	1.5	0.2	3.4	8.5	0.3	0.3	2.9	14.0
300	150	4.6	1.8	6.4	5.4	2.9	0.8	4.8	4.7	1.7	0.5	2.8	7.0
300	200	3.7	2.6	7.0	4.0	3.2	0.8	6.5	4.1	1.7	0.5	4.2	5.5
300	250	5.9	3.5	6.3	4.1	4.8	1.7	5.2	3.4	2.7	1.0	3.3	3.5
300	300	5.7	4.2	4.2	4.1	6.2	3.2	4.4	3.4	3.9	1.4	2.8	3.2
1000	1000	5.7	6.1	7.1	5.9	5.8	4.2	6.2	4.9	6.5	4.7	5.8	3.5

Table 2: Power study, 2 factors

N	T	$\hat{\alpha}_{0.05}\|\bar{r}=2$				$\hat{\alpha}_{0.05}\|\bar{r}=3$				$\hat{\alpha}_{0.05}\|\bar{r}=4$			
		LM	Sup LM	Wald	Sup Wald	LM	Sup LM	Wald	Sup Wald	LM	Sup LM	Wald	Sup Wald
100	100	6.3	1.8	8.1	5.4	77.9	1.8	100	98.3	41.7	0.5	100	97.3
100	150	8.9	2.5	10.0	4.8	95.8	24.0	100	100	88.8	2.8	100	99.9
100	200	8.9	4.1	9.3	5.4	97.6	72.9	92.0	92.0	95.5	39.6	91.8	92.5
100	250	12.0	5.3	12.4	6.5	99.1	98.0	97.4	97.4	99.0	77.9	97.4	97.4
100	300	13.0	6.5	11.6	6.0	99.6	98.0	83.6	83.6	99.4	94.1	83.5	83.7
150	100	6.1	2.2	7.8	5.9	77.9	1.4	99.7	99.5	41.6	0.6	99.8	99.0
150	150	7.5	2.2	8.3	5.0	95.4	24.5	100	100	88.5	2.2	100	100
150	200	8.8	4.1	9.8	5.4	98.8	76.5	100	100	97.7	40.2	100	100
150	250	9.7	4.8	10.3	6.0	99.4	94.4	99.0	99.1	98.5	79.1	99.0	99.1
150	300	11.4	6.3	10.8	7.1	99.7	98.6	90.5	91.1	99.7	94.5	90.7	91.1
200	100	6.4	1.5	7.6	4.6	79.4	2.3	100	97.7	42.9	0.7	100	99.2
200	150	8.5	3.4	9.5	6.3	97.0	24.1	100	100	89.0	3.0	100	100
200	200	8.6	3.5	9.3	4.5	99.0	77.6	100	100	98.0	38.8	100	100
200	250	11.5	4.5	12.3	5.7	100	96.8	100	100	100	82.7	100	100
200	300	11.2	5.4	12.6	6.4	99.8	98.8	99.9	99.9	99.7	95.1	99.9	99.9
250	100	5.1	1.4	6.7	4.5	80.4	1.8	100	99.7	45.2	1.0	100	99.2
250	150	6.7	2.4	7.8	5.0	97.0	24.5	99.9	100	90.7	3.2	100	100
250	200	7.2	3.4	7.8	5.0	99.2	78.9	100	100	98.4	40.9	100	100
250	250	10.5	5.5	11.3	5.8	99.8	95.6	100	100	99.7	82.4	100	100
250	300	11.5	5.7	12.0	7.6	99.9	99.2	100	100	99.9	95.1	100	100
300	100	6.0	1.6	7.0	6.7	80.1	1.2	100	99.1	45.4	0.3	100	98.9
300	150	8.6	2.1	9.9	4.7	97.3	24.9	100	100	91.5	3.4	100	100
300	200	8.6	4.3	9.2	6.8	99.3	79.0	100	100	98.4	43.3	100	100
300	250	11.4	4.4	11.9	5.8	99.8	94.3	100	100	99.5	82.6	100	100
300	300	11.3	5.9	12.1	7.7	99.8	99.0	100	100	99.8	96.3	100	100

Table 3: Comparison of LM test, 2 factors: known break date

N	T	no break, $\bar{r}=2$		1 break, $\bar{r}=2$		1 break, $\bar{r}=3$	
		BE	LM	BE	LM	BE	LM
100	100	6.0	3.9	100	5.6	21.9	96.8
100	150	5.9	5.2	100	7.2	18.2	100
100	200	5.2	4.3	100	6.2	26.0	89.8
100	250	5.3	4.8	100	8.7	17.9	97.7
100	300	5.7	4.3	100	7.4	30.2	83.9
150	100	6.4	4.3	100	5.8	18.3	94.6
150	150	5.9	5.7	100	6.6	16.2	100
150	200	5.6	4.3	100	6.2	12.5	100
150	250	5.5	4.5	100	5.7	14.9	98.3
150	300	4.9	4.0	100	5.6	20.6	89.7
200	100	5.5	4.1	100	4.1	20.0	95.8
200	150	5.4	4.8	100	6.6	15.8	100
200	200	7.0	4.5	100	6.3	14.0	100
200	250	6.5	4.7	100	7.5	12.6	100
200	300	5.0	4.7	100	7.8	12.0	99.7
250	100	6.8	3.9	100	4.2	18.8	97.0
250	150	5.4	5.3	100	5.9	14.9	100
250	200	4.5	4.6	100	6.1	11.3	100
250	250	5.1	4.2	100	6.6	10.9	100
250	300	6.6	4.9	100	8.3	7.9	100
300	100	7.3	4.7	100	5.4	19.7	96.3
300	150	7.0	3.6	100	6.1	14.4	100
300	200	5.9	3.4	100	6.0	13.6	100
300	250	5.9	5.4	100	6.7	12.0	100
300	300	5.7	6.1	100	7.0	10.0	100

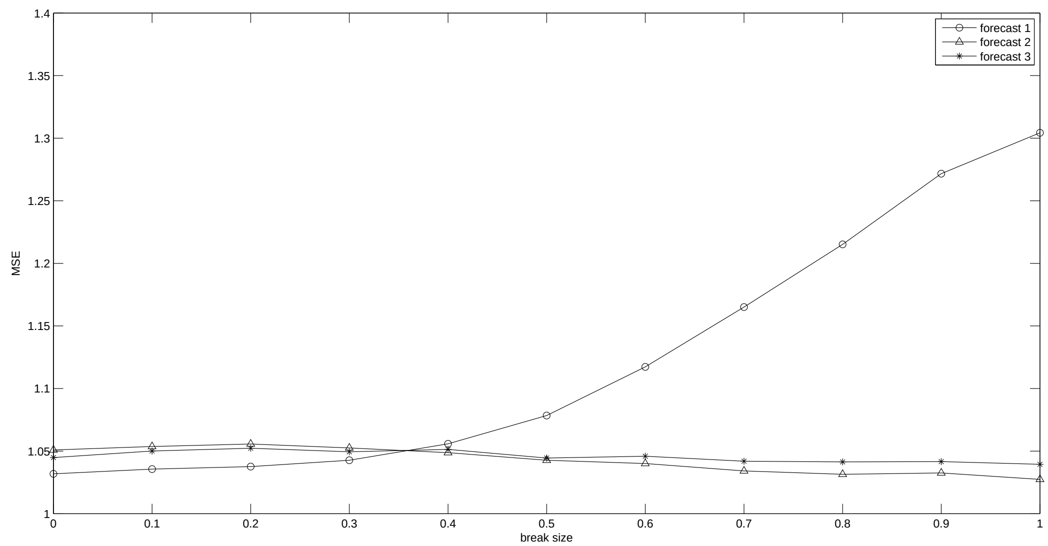

Figure 1: The MSEs of different forecasting methods.

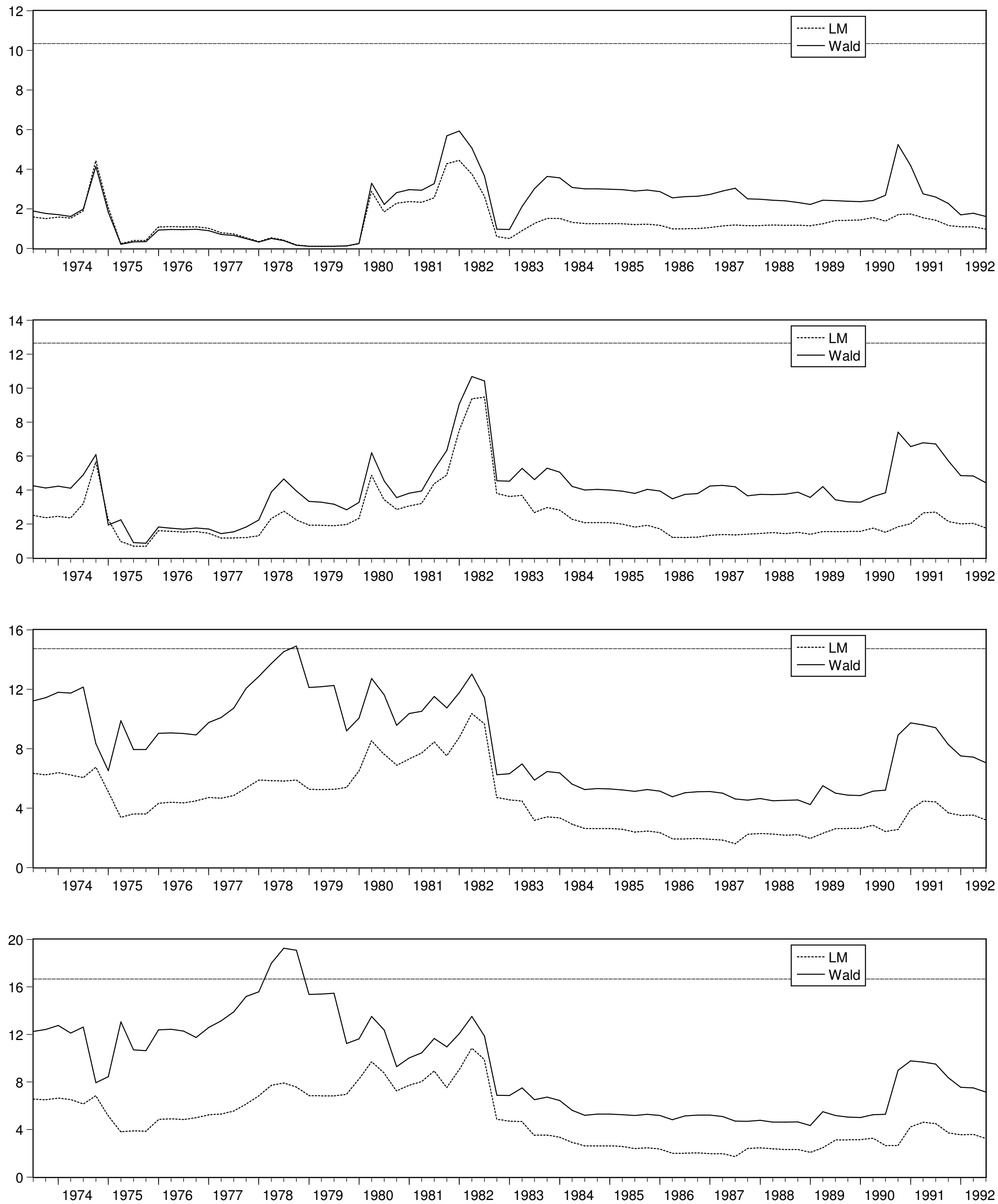

Figure 2: US data set. The LM test (dotted) and Wald test (solid) using the trimming $\Pi = [0.3, 0.7]$, for $\bar{r} = 3$ to 6 (from top to bottom), and the corresponding critical values (horizontal dotted lines) for the Sup Test.

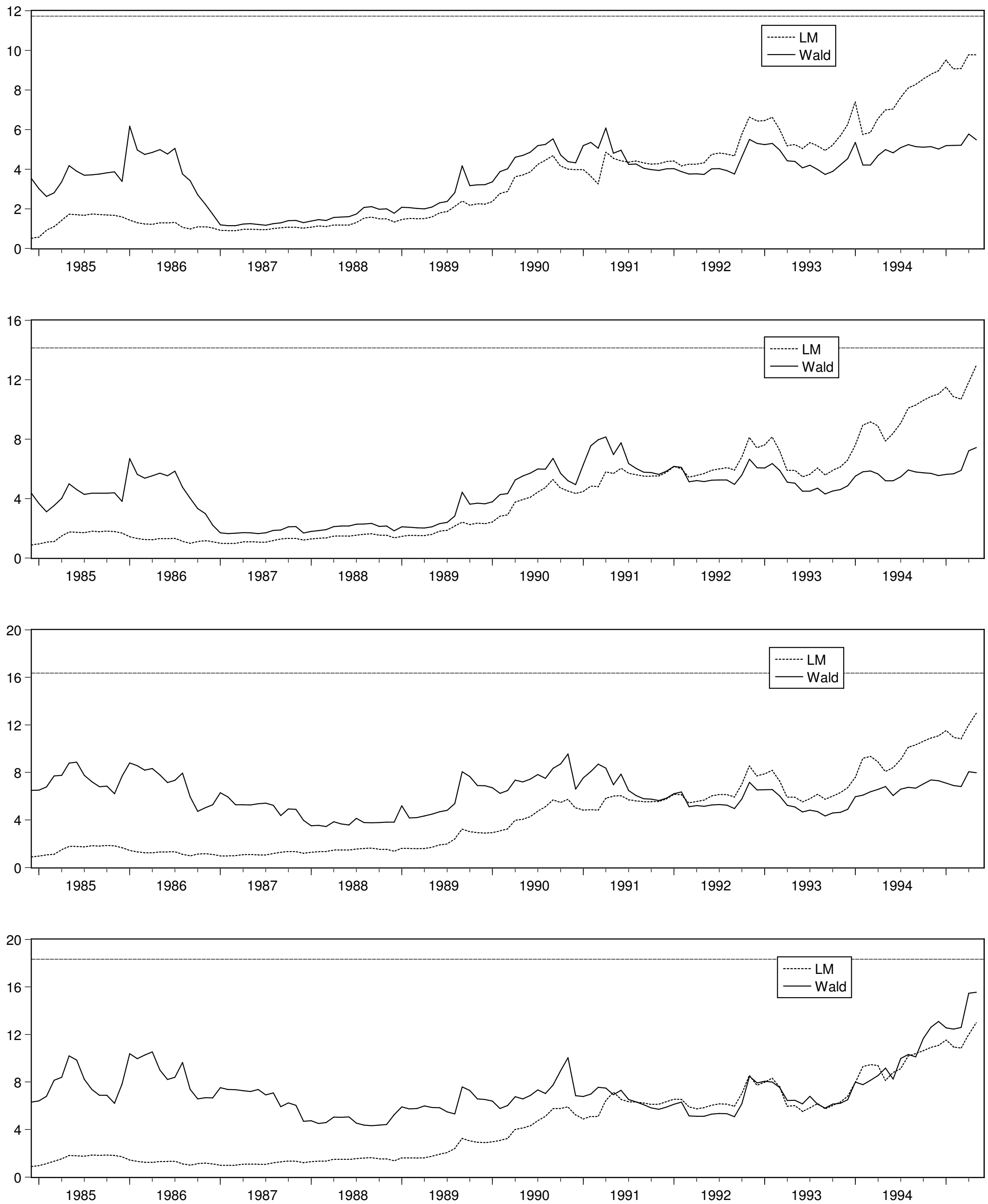

Figure 3: EU data set. The LM test (dotted) and Wald test (solid) using the trimming $\Pi = [0.15, 0.85]$, for $\bar{r} = 3$ to 6 (from top to bottom), and the corresponding critical values (horizontal dotted lines) for the Sup Test.