

Detecting Big Structural Breaks in Large Factor Models*

Liang Chen[†] Juan J. Dolado[‡] Jesús Gonzalo[§]

First Draft: May 2010

This draft: April 2012

Abstract

Time invariance of factor loadings is a standard assumption in the analysis of large factor models. Yet, this assumption may be restrictive unless parameter shifts are mild (i.e., local to zero). In this paper we develop a new testing procedure to detect *big* breaks in these loadings at either known or unknown dates. It relies upon testing for parameter breaks in a regression of the first of the \bar{r} factors estimated by PCA on the remaining $\bar{r} - 1$ factors, where \bar{r} is chosen according to Bai and Ng's (2002) information criteria. The test fares well in terms of power relative to other recently proposed tests on this issue, and can be easily implemented to avoid forecasting failures in standard factor-augmented (FAR, FAVAR) models where the number of factors is a priori imposed on the basis of theoretical considerations.

KEYWORDS: Structural break, large factor model, loadings, principal components.

JEL CODES: C12, C33.

*We are grateful to Dante Amengual, Manuel Arellano, Stephane Bonhomme and Soren Johansen for insightful comments, and to Atsushi Inoue for discussing with us about his related but independent research. We also thank seminar participants at CEMFI, Conference in Honour of Sir David F. Hendry (St. Andrews), ESEM 2011 (Oslo), and Workshops on High-Dimensional Econometric Modelling (Cass Business School) and Time Series Econometrics (Zaragoza) for helpful discussions. Financial support from the Spanish Ministerio de Ciencia e Innovación (grants ECO2010-19357 and Consolider-2010), Comunidad de Madrid (grant Excelecon) and Bank of Spain (grant ER program), 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 being well acknowledged 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 (FM, hereafter) is the assumption of time-invariant factor loadings. Possibly, one of the main reasons for this benign neglect of breaks stems from the important results obtained by Stock and Watson (2002, 2009) regarding the consistency of the estimated factors by principal components analysis (PCA hereafter) when the loadings are subject to small (i.e., local-to-zero) instabilities. Accordingly, these authors conclude that the failure of factor-based forecasts is mainly due to the instability of the forecast function, rather than of the different components, and hence, their advice is to use full-sample factor estimates and subsample forecasting equations to improve forecasts.

However, the main emphasis placed on local-to-zero breaks has been subsequently questioned. For example, by means of a Monte Carlo study, Banerjee, Marcellino and Masten (2008) conclude that, in contrast to Stock and Watson's diagnosis, the instability of factor loadings when big (i.e., not local-to-zero) breaks occur is the most likely reason behind the worsening factor-based forecasts, particularly in small samples. Likewise, when discussing Stock and Watson's research on this topic, Giannone (2007) argues that *"...to understand structural changes we should devote more effort in modelling the variables characterized by more severe instabilities..."*. In this paper, we pursue this goal by providing a precise characterization of the different conditions under which big and small breaks in the factor loadings may occur, as well as develop a simple test to distinguish between them. We conclude that, in contrast to small breaks, big breaks should not be ignored since they may lead to misleading results in standard econometric practices using factor models and in the potential interpretation of the factor themselves.

A forerunner of our paper is Breitung and Eickmeier (2010, BE henceforth) who are the first to propose a proper testing procedure to detect big breaks in the factor loadings. Their test relies on the idea 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 ones. Consequently, a Chow-type test can be constructed by regressing each variable in the data set on both the estimated factors using the whole sample period and their truncated versions from the date of the break onwards. Focusing on the joint statistical significance of the estimated coefficients for the truncated factors, their test compares the empirical rejection frequency among the individual regressions to a nominal size of 5% under the null of no breaks. In our view, this approach suffers from two limitations: (i) the BE test is mostly heuristic since its limiting distribution remains unknown except in some highly restrictive cases; and (ii) it lacks non-trivial power when the number of factors is not correctly specified. This loss of power can be very serious. For example, as explained further below, one of our main results is that a FM with r original factors where the loadings of one of them exhibit a big structural break at the same date admits a standard factor representation with $r + 1$ factors without a break. Hence, if the number of factors is chosen as $r + 1$, instead of r , their testing approach may not detect

any break at all when in fact there is one.

Our contribution here is to propose a simple testing procedure to detect breaks in FMs stemming from different sources which does not suffer from the previous shortcomings. In particular, we first derive some asymptotic results finding that, in contrast to small breaks, the number of factors will be over-estimated under big breaks. We argue that neglecting these breaks can have serious consequences on the forecasting performance of some popular regression-based models using factors, whose number is a priori imposed without testing for big breaks. Likewise, it may give rise to difficulties when trying to provide a structural interpretation of the estimated factors chosen according to some consistent information criteria (see Bai and Ng, 2006, and Chen, 2012). To avoid these problems, we propose a very simple regression-based testing procedure. As sketched earlier, it is based on the idea that a FM with big breaks in the loadings can be re-parameterized as another FM with constant loadings but a larger set of factors, where the number and the space spanned by the latter can be consistently estimated by PCA under fairly standard assumptions. Hence, rather than directly testing for whether all the elements of the loadings matrix are stable, which will suffer from an infinite-dimensionality problem as the number of variables in the panel data set grows, one can instead test whether the relationships among the larger finite-dimensional set of estimated factors are stable.

Specifically, our procedure consists of two steps. First, the number of factors for the whole sample period is estimated as \hat{r} using Bai and Ng's (2002) information criteria, and then the \hat{r} factors are estimated by PCA. Next, one of the estimated factors (e.g., the one associated to the largest eigenvalue) is regressed on the remaining $\hat{r} - 1$ factors, and then 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 implemented to test for a structural break in this regression. If the null of no breaks is rejected in the second-step regression, we conclude that there are big breaks and, otherwise, that either no breaks exist at all or that only small breaks occur. Further, on the basis of the rank properties of the covariance matrix of the estimated factors in different subsamples, we also provide some guidance on how to distinguish between breaks stemming from the loadings or from the DGP of the factors. This difference is important since the latter may lead to reject the null of constant loadings when it is true, leading to a misleading interpretation of the source of the break.

After completing a first draft of this paper, we became aware of a closely related (unpublished) paper by Han and Inoue (2011, HI hereafter) who, in an independent research, adopt a similar approach to ours in testing for big breaks in FM. The two approaches, however, differ in several respects. In effect, rather than using a simple regression-based approach to avoid the infinite-dimensionality problem, as we do here, HI (2011) test directly for differences before and after the break in all the elements of the covariance matrix of the estimated factors. We will argue below that, despite the fact that their tests use more information than ours, the potential lack of power entailed by our much simpler approach only applies to one of our tests (i.e., an LM type-test which becomes inconsistent) under a very specific and highly unrealistic data generating process (DGP). Yet, even in this special case, an alternative regression-based test based on the Wald principle, which in

general behaves much better than LM tests for detecting structural breaks,¹ does not suffer from any loss in power relative to HI's. In more general and realistic cases, both of our tests behave well, with the Wald test being even more powerful than the corresponding HI's test for small sample sizes, such as $N = T = 50$. One additional advantage of our simple linear-regression setup is that it permits to use many other existing methods for testing breaks, including multiple ones (see Perron, 2006, for an extensive review of these tests).

The rest of the paper is organized as follows. In Section 2, we present the basic notation, assumptions and the definitions of *small* and *big* 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 their effects on standard econometric practices with factor-augmented regressions. In Section 4, we first derive the asymptotic distributions of our tests and next discuss, when a big break is detected, how one can identify whether it stems from the loadings or from the process driving the factors; lastly we provide some analytical results on how our test fares, in terms of power, relative to BE's and HI's tests. Section 5 deals with the finite sample performance of our test relative to the competing tests using Monte-Carlo simulations. Section 6 provides an empirical application. Finally, Section 7 concludes. Three appendices contain detailed proofs of the main results.

2 Notation and Preliminaries

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

$$X_t = AF_t + e_t \quad (1)$$

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 which is finite, $F_t = (f_{1t}, \dots, f_{rt})'$ 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 static FM with $r \times (p + 1)$ static factors. Further, given the assumptions we make about the e_t error terms, the case analyzed by BE (2010) where the e_{it} disturbances are generated by individual specific AR processes is also considered. Notice, however, that our setup excludes the generalized dynamic FM considered by Forni and Lippi (2001) when the polynomial distributed lag possibly tends to infinity.

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

$$X_t = AF_t + e_t \quad t = 1, 2, \dots, \tau, \quad (2)$$

$$X_t = BF_t + e_t \quad t = \tau + 1, \dots, T \quad (3)$$

¹See Kim and Perron (2009).

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 FM in (2) and (3) can be rewritten as:

$$X_t = AF_t + CG_t + e_t \quad (4)$$

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, 2009), the effects of some mild (local to zero) instability in the factor loadings can be averaged out, so that estimation and inference based on PCA remain valid. We generalize their analysis by allowing for two types of break sizes: *small* and *big*. In contrast to the former, we will show that the latter cannot be neglected. To distinguish between them, it is convenient to partition the C matrix as follows:

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

where Λ and H are $N \times k_1$ and $N \times k_2$ matrices that correspond 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 the G_t matrix into G_t^1 and G_t^2 , such that (4) becomes:

$$X_t = AF_t + \Lambda G_t^1 + H G_t^2 + e_t \quad (5)$$

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. $E\|\lambda_i\|^4 < \infty$. $N^{-1} \sum_{i=1}^N \lambda_i \lambda_i' \rightarrow \Sigma_\Lambda$ as $N \rightarrow \infty$ for some positive definite matrix Σ_Λ .
- b. $\eta_i = \frac{\kappa_i}{\sqrt{NT}}$ for $i = 1, 2, \dots, N$ and $E\|\kappa_i\|^8 < \infty$.

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 $N \rightarrow \infty$. Assumption 1.b, in turn, provides the definition of small breaks which, besides having some bounded moments, are characterized as being of order $1/\sqrt{NT}$ so that they can be neglected as N and T go to infinity.

REMARK 1. It could be argued that a natural way to define the relevant hypotheses would be $H_0 : A = B$ and $H_1 : A \neq B$, as in BE (2010) and HI (2011). However, in the context of FMs with large N , the size of the breaks $C = B - A$ has important implications. For example, when $C = O_p(1/\sqrt{NT})$, we can invoke Stock and Watson's (2002) results showing the usual inferences, including estimation and testing using PC, are still valid because the effects of these small breaks will be averaged out as N and T go to infinity, and therefore they can be safely ignored. Our definitions of big and small breaks correspond to fixed and local alternatives respectively where the the order of small breaks give the sufficient conditions for ignoring these breaks. In other words, when $C = O_p(1)$, we have

a FM with $r + k_1$ new factors while, when $C = O_p(1/\sqrt{NT})$, we have a FM with the r original factors. By contrast, when $O_p(1/\sqrt{NT}) < C < O_p(1)$, we do not know how many new factors will be detected and the arguments we use to show the consistency of our test do not work in these cases. Hence, for reasons to be spelt out below, we do not lose anything by focusing exclusively on big breaks ($H_0 : k_1 = 0$ vs. $H_1 : k_1 \neq 0$). Notice also that we do not examine here the behavior of our tests under local alternatives of the form $\frac{1}{N}C'C = \gamma_{N,T}\Sigma_C$, where $\gamma_{N,T}$ is $o_p(1)$, because, according to Assumption 1a, when $\frac{1}{N}C'C = o_p(1)$, the FM will not have a new representation with $r + k_1$ factors.

To investigate the influence of the breaks on the number and estimation 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: $E(F_t) = 0$, $E\|F_t\|^8 < \infty$, $T^{-1}\sum_{t=1}^T F_t F_t' \rightarrow \Sigma_F$ and $T^{-1}\sum_{t=1}^{\tau} F_t F_t' \rightarrow \pi^* \Sigma_F$ as $T \rightarrow \infty$ for some positive definite matrix Σ_F where $\pi^* = \lim_{T \rightarrow \infty} \frac{\tau}{T}$.

Assumption 3. Factor Loadings: $E\|\alpha_i\|^4 \leq M < \infty$, and $N^{-1}A'A \rightarrow \Sigma_A$, $N^{-1}\Gamma'\Gamma \rightarrow \Sigma_\Gamma$ as $N \rightarrow \infty$ 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: $[F_t]_{t=1}^T$, $[\alpha_i]_{i=1}^N$, $[\lambda_i]_{i=1}^N$, $[\kappa_i]_{i=1}^N$ and $[e_t]_{t=1}^T$ are mutually independent groups, and uniformly for all i

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

Assumptions 3 and 4 are standard in the literature on FMs allowing for weak cross-sectional and temporal correlations between the errors (see Bai and Ng, 2002). Notice that, in our specific setup, Assumption 3 excludes the case where a new (old) factor appears (disappears) after the break since this event would imply that Σ_Γ becomes singular. However, this is not restrictive since we could always envisage any potential factor as having non-zero, albeit small, loadings in either of the relevant subsamples. Assumption 2, in turn, is a new one. Since factors and factor loadings cannot be separately identified, we have to assume that DGPs with breaks in the loadings, which can be reabsorbed by transformations of the factors, should not be included in the alternative. 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 insight underlying our approach. Thus, for the sake of simplicity, we assume them to be independent in the sequel.

3 The Effects of Structural Breaks

In this section, we study the effects of structural breaks on the estimation of both the number of factors based on the information criteria (IC, henceforth) proposed by Bai and Ng (2002) and the factors themselves through PCA. Our main finding is that, in contrast to Stock and Watson's (2002, 2009) consistency result for the true factor space under small breaks, the factor space estimated by PCA is inconsistent, and that the number of factors tends to be overestimated under big breaks.

3.1 The estimation of factors

Let us rewrite model (5) with k_1 big breaks and k_2 small breaks, so that $r = k_1 + k_2$, in the more compact form:

$$X_t = AF_t + \Lambda G_t^1 + \varepsilon_t \quad (6)$$

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 FM with new factors $F_t^* = [F_t' \ G_t^{1'}]'$ and new factor loadings $[A \ \Lambda]$.

Let \bar{r} be the selected number of factors, either by some prior knowledge or using the information criteria, where notice that \bar{r} is not necessarily equal to r . Let \hat{F} be \sqrt{T} times the \bar{r} eigenvectors corresponding to the \bar{r} largest eigenvalues of the matrix XX' , where the $T \times N$ matrix $X = [\bar{X}_1, \bar{X}_2, \dots, \bar{X}_T]'$, $\bar{X}_t = [X_{t1}, X_{t2}, \dots, X_{tN}]'$, $\hat{F} = [\hat{F}_1, \hat{F}_2, \dots, \hat{F}_T]'$. Then we have:

Proposition 1. *For any fixed $\bar{r}(1 \leq \bar{r} \leq r + k_1)$, 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:*

$$\hat{F}_t = DF_t^* + O_p(\delta_{N,T}^{-1}) \text{ for } t = 1, 2, \dots, T. \quad (7)$$

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

Let us next 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

$$\hat{F}_t = DF_t + O_p(\delta_{N,T}^{-1}) \quad (8)$$

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).²

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

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

²Notice that for the estimators \hat{F} defined here, \bar{r} has to be smaller or equal to r for (8) to hold. If \hat{F}_t is defined as the PC of X_t , \bar{r} can be larger than r .

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 $t = 1, 2, \dots, \tau$, and $G_t^1 = F_t^1$ for $t = \tau + 1, \dots, T$, where F_t^1 is the $k_1 \times 1$ sub-vector of F_t that is subject to big breaks in their loadings. Therefore (9) can be expressed as:

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

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

where $D_2^* = D_1 + [D_2 \ 0]$, 0 is a $\bar{r} \times (r - k_1)$ zero matrix, and in general $D_2 \neq 0$. Hence, since $D_1 \neq D_2^*$, this implies that, in contrast to small breaks where D_2 tends to 0 due to the local-to-zero properties of the elements of H in (5) (see Assumption 1.b), under big breaks the estimated factors \hat{F} will not be consistent for the space of the true factors F when these breaks are neglected. Accordingly, as will be explained below, imposing a priori the number of estimated factors to be used as predictors or explanatory variables in standard factor-augmented models may lead to misleading results.

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):

$$y_t = a' F_t + b' W_t + u_t, \quad t = 1, 2, \dots, T \quad (12)$$

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} ($i = 1, 2, \dots, N; t = 1, 2, \dots, T$) 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 that there is a big break at date τ . From (10) and (11), we can rewrite (12) as:

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

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

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

$$y_t = a_1' \hat{F}_t + b' W_t + \tilde{u}_t \text{ for } t = 1, 2, \dots, \tau, \quad (13)$$

$$y_t = a_2' \hat{F}_t + b' W_t + \tilde{u}_t \text{ for } t = \tau + 1, \dots, T \quad (14)$$

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 $a_1 \neq a_2$. Thus, using the indicator function $\mathbf{1}(t > \tau)$, (13) and (14) can be rewritten as

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

REMARK 2. A straightforward implication of the previous result is that if we were to impose the number of factors, on a priori grounds, therefore ignoring the set of regressors

$\hat{F}_t \mathbf{1}(t > \tau)$ in (15), in general the estimation of b will become inconsistent due to omitted-variables bias. Interestingly, there are many examples in the literature where the number of factors is a priori imposed for theoretical or practical reasons. For example, to name a few, a single common factor representing a global effect is assumed in the well-known study by Bernanke, Boivin and Elias (2005) on measuring the effects of monetary policy in Factor Augmented VAR (FAVAR) models, as well as in the risk analysis in portfolios of corporate debt by Gourieroux and Gagliardini (2011) where a single factor is supposed to capture a latent macro-variable. Likewise, two factors are a priori imposed by Rudebusch and Wu (2008) in their macro-finance model. Notice that a similar argument will render inconsistent the impulse response functions in FAVAR models where the regressand in (12) becomes $y_{t+1} = (F_{t+1}, W_{t+1})'$.

REMARK 3. Alternatively, if the number of factors is not imposed a priori and instead is estimated by means of some consistent IC, we will show in the next section (Proposition 2) that the chosen number of factors will tend to $r + k_1$ as the sample size grows. In this case, D_1 and D_2^* are $(r + k_1) \times r$ matrices, 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

$$a^* = a' D^*, \quad (16)$$

then $a'_1 = a'_2 = a^*$ so that (13) and (14) can be rewritten as

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

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

REMARK 4. Yet, even in this case, the factors themselves may be direct objects of interest and thus interpreting their meaning can have important implications for structural analysis. For example, a large body of empirical research in financial economics is concerned with identifying the factors that determine the prices of the securities.³ Likewise, the predictions of inflation rates could be more accurate if the factors associated with monetary policy shocks are given more weight than other factors identified as productivity changes. However, the existence of big breaks may hamper this identification procedure. For instance, if $\hat{r} = 2$, a relevant question would be: are there two genuine factors or one factor and one break? Our testing procedure provides a useful tool to disentangle these two cases. In effect, in case of rejection of the null hypothesis of no big breaks, our procedure allows to estimate consistently the break date. Therefore, one should split the sample and estimate the chosen number of factors according the IC in the different subsamples and merge them to become the new set of factors for the whole sample. In this fashion, one could test if the factors are more or less volatile after the break date or whether there is a change in their correlation with some observed variables.

Another area where our testing approach can be useful is in applications where, once the factors are estimated, a VAR model is specified in order to identify the structural shocks

³Chen et al (1986) and Shanken and Weinstein (2006) are good illustrations of attempts to interpret the underlying forces in the stock market developments in terms of some observed macro variables.

which drive them (see e.g. Charnavoki and Dolado, 2012). This identification becomes more difficult as the number of factors increases. For this reason it is very important to determine whether the selection of a large number of factors is due to having several genuine factors or to a break in some of them. The insight is that instead of having to identify $r + k_1$ shocks we would only have to identify r .

In sum, 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 under big breaks. As a simple remedy, $\hat{F}_t \mathbf{1}(t > \tau)$ 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 chosen according to some consistent IC, although it may hinder the correct interpretation of the estimated factors in terms of observables. As a result, in order to run regression (17), our advice is to avoid imposing the number of factors a priori, unless a formal test of big breaks is implemented. We will illustrate these points in Section 5 by means of simulations in a typical forecasting exercise where the predictors are common factors estimated by PCA.

3.2 The estimated number of factors

Breitung and Eickmeier (2010) have previously argued that the presence of structural breaks in the factor loadings may lead to the overestimation of the number of factors. Yet, since they do not have a formal proof of this result, we proceed to fill this gap by providing a rigorous proof.

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

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

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

Again, in the absence of big breaks ($k_1 = 0$), this result trivially replicates Theorem 2 of Bai and Ng (2002). However, under big breaks ($k_1 > 0$), their IC will overestimate the number of original factors by the number of big breaks ($0 < k_1 \leq r$) because, as shown above, a FM with this type of break admits a representation without a break but with more factors.

Finally, it is important to stress 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

Our goal here is to develop a test for big breaks. As mentioned above, if we were to follow the usual approach in the literature to test for structural breaks, we would consider the following null and alternative hypotheses in (2) and (3): $H_0 : A = B$ vs. $H_1 : A \neq B$. However, this standard formulation faces two problems. First, if only small breaks occur, the alternative hypothesis may not be interesting since $C = A - B$ vanishes as $N \rightarrow \infty$ and $T \rightarrow \infty$. Secondly, and foremost, since A and B are $N \times r$ matrices, we would face an infinite-dimensional parameter problem as N grows if we were to consider differences in all their individual elements.

Relying upon the discussion in (10) and (11) about the inconsistency of \hat{F} for the space of the true factors F when big breaks occur, our strategy to circumvent this problem is to focus instead on how the dependence properties of the \hat{r} estimated factors (using the whole sample) change before and after the potential break date. Since, in line with the standard assumption in Bai and Ng (2002) and Stock and Watson (2002), the (unknown) number of true factors, r , is considered to be invariant to the sample size, our previous result in Proposition 2 ensures that $r + k_1$, with $k_1 \leq r$, is finite-dimensional. Hence, to detect big breaks, we should rather consider:

$$\begin{aligned} H_0 : k_1 &= 0, \\ H_1 : k_1 &> 0, \end{aligned}$$

where the new 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 determined by Bai and Ng's (2002) IC ($\bar{r} = \hat{r}$) applied to the whole sample ($t=1, \dots, T$), and \bar{r} common factors (\hat{F}_t) are estimated by PCA.*
2. *In the second step, we consider the following linear regression of the first estimated factor on the remaining $\bar{r} - 1$ ones:*

$$\hat{F}_{1t} = c_2 \hat{F}_{2t} + \dots + c_{\bar{r}} \hat{F}_{\bar{r}t} + u_t = c' \hat{F}_{-1t} + u_t \quad (18)$$

where $\hat{F}_{-1t} = [\hat{F}_{2t} \dots \hat{F}_{\bar{r}t}]'$ and $c = [c_2 \dots c_{\bar{r}}]'$ 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 of no big breaks.

REMARK 5. It is noteworthy that, though the \bar{r} estimated factors by PCA are orthogonal by construction for the whole sample, and hence the c vector of coefficients in (18) is the zero vector when the regression is run from 1 to T , this property does not obviously

hold when the sample is split into two subsamples, before and after the potential break date. Hence regression (18) is always meaningful and both steps in the testing procedure can be very easily implemented in practice. In the second step, although there are many methods of testing for breaks in a simple linear regression model, we follow Andrews (1993) to define the LM and Wald tests when the possible break date is assumed to be known, and their *Sup-type* versions when there is no prior knowledge about the break date. Moreover, since the LM, Wald and LR test statistics have the same asymptotic distribution under the null, we focus on the first two because they are computationally simpler.

Define $D^* = V^{-1/2}\Gamma'\Sigma_A^{1/2}$ as the limit of the matrix D in equation (8), where $V = \text{diag}(v_1, v_2, \dots, v_r)$, $v_1 > v_2 > \dots > v_r$ are the eigenvalues of $\Sigma_A^{1/2}\Sigma_F\Sigma_A^{1/2}$, and Γ is the corresponding eigenvector matrix (see Bai, 2003). Define $\mathcal{F}_{1t} = D_1^*F_t$ and $\mathcal{F}_{-1t} = D_{-1}^*F_t$, where D_1^* is the first row of D^* and D_{-1}^* is the matrix containing the second to last rows of D^* . Finally, let the $(r-1) \times (r-1)$ matrix $S = \lim \text{Var}\left(\frac{1}{T}\sum_{t=1}^T \mathcal{F}_{-1t}\mathcal{F}_{1t}\right)$.

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

$$\mathcal{L}(\bar{\pi}) = \frac{T}{\bar{\pi}(1-\bar{\pi})} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{\tau} \hat{F}_{-1t} \hat{u}_t \right)' \hat{S}^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{\tau} \hat{F}_{-1t} \hat{u}_t \right) \quad (19)$$

where $\bar{\pi} = \tau/T$, τ is a pre-assumed date for the potential break, \hat{u}_t is the residuals in the OLS regression of (18)⁴, and \hat{S} is a consistent estimator of S .⁵

The corresponding Sup-LM statistic is defined as:

$$\mathcal{L}(\Pi) = \sup_{\pi \in \Pi} \frac{T}{\pi(1-\pi)} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{[T\pi]} \hat{F}_{-1t} \hat{u}_t \right)' \hat{S}^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{[T\pi]} \hat{F}_{-1t} \hat{u}_t \right) \quad (20)$$

where Π is any set with closure lies in $(0, 1)$.

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

$$\mathcal{L}^*(\bar{\pi}) = \pi(1-\pi) \cdot T \left(\hat{c}_1(\bar{\pi}) - \hat{c}_2(\bar{\pi}) \right)' \hat{S}^{-1} \left(\hat{c}_1(\bar{\pi}) - \hat{c}_2(\bar{\pi}) \right) \quad (21)$$

and

$$\mathcal{L}^*(\Pi) = \sup_{\pi \in \Pi} \pi(1-\pi) \cdot T \left(\hat{c}_1(\pi) - \hat{c}_2(\pi) \right)' \hat{S}^{-1} \left(\hat{c}_1(\pi) - \hat{c}_2(\pi) \right) \quad (22)$$

where $\hat{c}_1(\pi)$ and $\hat{c}_2(\pi)$ are OLS estimates of c using subsamples before and after the break point : $[T\pi]$.⁶

⁴By construction, $\hat{u}_t = \hat{F}_{1t}$.

⁵See Appendix A.3 for discussions on the estimation of S .

⁶We can also construct the Wald test as $T \left(\hat{c}_1(\bar{\pi}) - \hat{c}_2(\bar{\pi}) \right)' \left(\frac{\hat{S}_1}{\bar{\pi}} + \frac{\hat{S}_2}{(1-\bar{\pi})} \right)^{-1} \left(\hat{c}_1(\bar{\pi}) - \hat{c}_2(\bar{\pi}) \right)$ and the Sup-Wald test similarly, where \hat{S}_1 and \hat{S}_2 are estimates of S using subsamples. Yet, in all our simulations, the results based on these two methods are very similar. Therefore, for brevity, we focus on the ones obtained using the full sample estimation of S , as in (21) and (22).

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

$$X_t = Af_{1t} + Bg_t + e_t$$

where $g_t = 0$ for $t = 1, \dots, \tau$, and $g_t = f_t$ for $t = \tau + 1, \dots, T$. 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:

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

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

which imply that:

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

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

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 D matrix. 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. The case where $d_3 = 0$ (although excluded by our Assumptions) may render the above regression for the first subsample invalid. In section 4.5.2, we will consider such a special alternative where our LM test may become inconsistent while our Wald test still exhibits very high power.

Likewise, if the break date τ is not a priori assumed to be known, the Sup-type tests 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) to (22) under the null hypothesis, as well as extend the intuition behind this simple example to the more general case 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 Bai and Ng's (2002) IC, and it converges to the true one under the null hypothesis of no big break, we start with the most interesting case where $\bar{r} = r$.

As discussed earlier, under the null, the use of PCA 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 and Wald statistics, we adopt the following additional assumptions:

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

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 $i, j = 1, 2, \dots, r$, 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 σ -algebra generated by the information at time $t, t-1, \dots$, $\{c_t\}$ and $\{\gamma_m\}$ are non-negative sequences and $\gamma_m = O(m^{-1-\delta})$ for some $\delta > 0$.

Assumption 8. $\sup_{\pi \in [0,1]} \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^T \pi \sum_{i=1}^N \alpha_i F_t' e_{it} \right\|^2 = O_p(1)$.

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

Assumption 10. The eigenvalues of the $r \times r$ matrix $\Sigma_A \Sigma_F$ are distinct.

Assumption 6 and 8 are required to bound the estimation errors of \hat{F}_t , while Assumption 7 is needed to derive the weak convergence of the test statistics using the Functional Central Limit Theorem (FCLT). Assumption 10 is the Assumption G of Bai (2003), which is necessary for $D \xrightarrow{p} D^*$.

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 $Var(v_{it}) = \sigma_i^2 < \infty$. In this case, it can be shown that:

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

for which 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 similar 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, yielding positive definite estimates of the covariance matrix like, e.g., the popular Newey and West's (1987) estimator with a Barlett kernel which is the one used in our simulations below.⁷

Let " \xrightarrow{d} " denote *convergence in distribution*, then:

⁷Though not reported, other estimators like those based on Parzen kernels, yield similar results in our simulations about the size and power properties of the LM and Wald tests.

Theorem 1. Under the null hypothesis $H_0 : k_1 = 0$ and Assumptions 1 to 10, as $N, T \rightarrow \infty$, we have that both the LM and Wald tests verify

$$\mathcal{L}(\bar{\pi}), \mathcal{L}^*(\bar{\pi}) \xrightarrow{d} \chi^2(r-1)$$

where $\bar{\pi} = \tau/T$ for a given τ ; and

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

for any set Π with closure lies in $(0, 1)$, where $\mathcal{W}_{r-1}(\cdot)$ is a $r-1$ vector of independent Brownian Motions on $[0, 1]$ restricted to Π .

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

REMARK 6. It is easy to show that Theorem 1 still holds when $\bar{r} < r$. However, when $\bar{r} > r$, the matrix D^* may not be asymptotically full ranked. Yet, as simulations in section 5 show, Theorem 1 still provides a reasonably good approximation of the distribution of test statistics in finite samples in such an instance. Moreover, the case where $\bar{r} > r$ can be avoided if, instead of relying on a priori choice of \bar{r} , practitioners use Bai and Ng's (2002) IC, in line with the result in Proposition 2.

4.3 Behavior of LM and Wald tests under the alternative hypothesis

We now extend the insight of the simple example considered in section 4.1 to show that, under the alternative hypothesis ($k_1 > 0$), the linear relationship between the estimated factors (using the whole sample) changes at time τ , so that the proposed test is consistent and therefore that big breaks can be detected.

First, let us consider the case where $\bar{r} = r + 1$ so that D_1 and D_2^* in (10) and (11) become $(r+1) \times r$ matrices with full column rank. Notice that we can always find $(r+1) \times 1$ 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:

$$\begin{aligned} \rho_1' \hat{F}_t &= o_p(1) \quad t = 1, 2, \dots, \tau, \\ \rho_2' \hat{F}_t &= o_p(1) \quad t = \tau + 1, \dots, T \end{aligned}$$

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

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

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

where $\hat{F}'_{-1t} = [\hat{f}_{2t}, \dots, \hat{f}_{r+1,t}]$ and ρ_1^*, ρ_2^* are both $r \times 1$ vectors. Next, to show that $\rho_1^* \neq \rho_2^*$, we proceed as follows. Suppose that $\gamma \in \text{Null}(D_1')$ and $\gamma \in \text{Null}(D_2^{*})$, then by the definitions of D_1 and D_2^* and by the basic properties of full-rank matrices, it holds that $\gamma \in \text{Null}(D')$. Since D is full rank $(r+1) \times (r+k_1)$ matrix and $k_1 \geq 1$, then $\text{Null}(D') = 0$

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 1, we can always find two non-zero vectors such that $\rho_1 \neq b\rho_2$ for any constant $b \neq 0$. More generally, when $\bar{r} > r + 1$, the equations (23) and (24) still hold by giving zero coefficients to the extra estimated factors:

$$\begin{aligned}\hat{F}_{1t} &= \hat{F}'_{-1t}\rho_1^* + 0\hat{f}_{r+2,t} + \cdots + 0\hat{f}_{\bar{r}t} + o_p(1) & t = 1, 2, \dots, \tau, \\ \hat{F}_{1t} &= \hat{F}'_{-1t}\rho_2^* + 0\hat{f}_{r+2,t} + \cdots + 0\hat{f}_{\bar{r}t} + o_p(1) & t = \tau + 1, \dots, T.\end{aligned}$$

Notice that when $\bar{r} \leq r$, 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 two subsamples before and after the break, it is easy to show that $\hat{\rho}_1 \rightarrow \theta_1$ and $\hat{\rho}_2 \rightarrow \theta_2$, but generally we cannot verify that $\theta_1 \neq \theta_2$.

REMARK 7. Consequently, as will be confirmed by our simulations below, our test is consistent when $\bar{r} > r$.⁸ Hence, by equating \bar{r} to possibly different values of \hat{r} generated by Bai and Ng's (2002) IC in finite samples we would get more robust results if the test rejects the null for some of these values rather than for only one. In other words, running the test for different \bar{r} s can provide useful information. For example, if the null is rejected for $\bar{r} = 3, 4$ but not for $\bar{r} = 2$, this could be an indication that $r = 2$, since our test begins to have power when $\bar{r} > r$.

4.4 Disentangling breaks in loadings from breaks in factors

A potential critique for all available tests for big breaks in FM is that they cannot differentiate between breaks in factor loadings and breaks in the covariance matrix of factors. For illustrative purposes, consider a FM where the factor loadings are constant but the covariance matrix of the factors breaks: $r = 2$, $E(F_t F_t') = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ for $t = 1, \dots, T/2$, and $E(F_t F_t') = \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$ for $t = T/2 + 1, \dots, T$, with $\rho \neq 0$ and 1. If we further assume $\Sigma_A = \lim_{N \rightarrow \infty} A'A/N$ is a diagonal matrix, then, in view of Bai (2003), we have that $\hat{F}_t = F_t + o_p(1)$, where \hat{F}_t is a 2×1 vector. In this case, regressing \hat{F}_{1t} on \hat{F}_{2t} will yield estimated coefficients close to ρ and $-\rho$ before and after the break, respectively. As a result, our tests will detect a big break in the factor loadings, while the true DGP has a break in the factors and not in the constant loadings.

Although the above example has been excluded by our Assumptions 2 and 7, it is plausible that the factor dynamics may be in practice subject to structural breaks. For instance, if the factors are interpreted as common shocks to the economy, then it is likely that the

⁸We will consider a rather special example in the next section where our LM test may fail to achieve consistency but our Wald test is still consistent.

volatilities of these shocks have decreased since the beginning of 1980s as evidenced by recent studies on the effect of the Great Moderation (see, e.g., Gali and Gambetti, 2009). Hence, for interpretational purposes, it becomes relevant to identify whether breaks stem from changes in the factors themselves or in their loadings.

The main difficulty in doing so lies in the fact that, since they are multiplicative, factors and loadings can not be separately identified using PCA, so that breaks in the former can be easily expressed as breaks in the latter, and conversely. As mentioned earlier, this identification problem not only affects our test but also the competing ones in the literature. For example, as discussed in the Introduction, given that the HI test is based on the comparison of the covariance matrices of the estimated factors before and after the break, it will also exhibit power against breaks in the factors as long as the estimated factors are consistent.⁹ As regards the BE test, though in principle seems to be robust to breaks in the true factors because it is conditional on the estimated factors, Hansen's (2000) results imply that the asymptotic distribution is bound to differ insofar as the marginal distributions of the regressors (i.e., the estimated factors) change.

Yet, in spite of these shortcomings, we briefly outline here a procedure that allows to differentiate the source of breaks by examining the rank of the covariance matrices of the estimated factors. To illustrate this idea, consider first the simple example in subsection 4.1, where the DGP has 1 factor and 1 break in the loadings at date τ . As shown in Proposition 2, the estimated number of factors will converge to 2, and our test will detect the break. Next, consider an alternative DGP, this time with 2 factors $F_t = [F_{1t}, F_{2t}]'$, constant factor loadings, and a break in the covariance matrix of the factors at date τ such that $E(F_t F_t') = \Sigma_1$ in the first subsample and $E(F_t F_t') = \Sigma_2$ in the second subsample. Under this DGP, as long as $1/T \sum_{t=1}^T F_t F_t'$ converges to some positive-definite matrix, the estimated number of factors will also converge to 2, and both will be consistently estimated. Hence, as discussed above, our tests may also detect a big break in the factor loadings when it actually happens in the covariance matrix of the factors.

To differentiate between these two cases, notice that, the covariance matrices of the estimated factors before and after the break in the first DGP (break in loadings) converge to matrices with reduced rank (i.e, equal to 1 in this example) given by

$$\begin{aligned}\widehat{\Sigma}_1 &\Rightarrow E(f_t^2) D_1 D_1' \text{ for } t = 1, \dots, \tau, \\ \widehat{\Sigma}_2 &\Rightarrow E(f_t^2) (D_1 + D_2) (D_1 + D_2)' \text{ for } t = \tau + 1, \dots, T.\end{aligned}$$

where D_1 and D_2 denote the first and second 2x1 column vectors, respectively, of the 2x2 full-ranked matrix given by $D = \begin{pmatrix} d_1 & d_2 \\ d_3 & d_4 \end{pmatrix}$ in this DGP.

By contrast, in the second DGP (breaks in factors), the corresponding covariance ma-

⁹In view of Bai and Ng's (2002) results, the estimated number of factors and the factors themselves are still consistent as long as $1/T \sum_{t=1}^T F_t F_t'$ converges to some positive-definite matrix.

trices will converge to matrices with full rank (i.e., equal to 2 in this example) given by

$$\begin{aligned}\widehat{\Sigma}_1 &\Rightarrow D\Sigma_1D' \text{ for } t = 1, \dots, \tau, \\ \widehat{\Sigma}_2 &\Rightarrow D\Sigma_2D' \text{ for } t = \tau + 1, \dots, T.\end{aligned}$$

Therefore, when our tests detect a break in the FM, a rank test could be applied to the covariance matrices of the estimated factors in the subsamples before and after the break date. If the break happens in the loadings, the matrices should have reduced rank and, as result, separate application of Bai and Ng's (2002) IC to the respective data sets will choose a single factor in each of the two subsamples. Conversely, if the break happens in the factors, the matrices should have full rank and, as a result, the IC will choose this time 2 factors in each of the subsamples.¹⁰

4.5 Comparison with other available tests

4.5.1 Comparison with the BE test

BE (2010) is, to our knowledge, the first paper that proposes a test for big breaks. Thus, it seems natural to start by discussing how our testing procedure compares to theirs. In our view, the BE test suffers from three shortcomings which are worth mentioning before the comparison is carried out.

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 we have shown, a FM with big breaks in the factor loadings admits a new representation with more factors but no break. 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 heuristic. Their null hypothesis is $A = B$, or $\alpha_i = \beta_i$ for all $i = 1, \dots, N$, rather than $\alpha_j = \beta_j$ for a specific j .¹¹ They construct N test statistics (denoted by s_i , $i = 1, \dots, N$) 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 which is asymptotically normal, but this requires the errors e_{it} and e_{jt} to be independent if $i \neq j$, an assumption which is

¹⁰In Chen, Dolado and Gonzalo (2012, in process) we are currently working on the asymptotic distribution of the rank test based on the eigenvalues of the sample covariance matrices of \hat{F}_t in each of the subsamples, as an alternative approach to the direct use of Bai and Ng's (2002) criteria. Interestingly, notice that this rank test could also account for the possibility of new factors arising after the break date (excluded by Assumption 3), since in this case the rank will be larger in the second than in the first subsample.

¹¹The 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 $i = 1, \dots, N$, or alternatively if the break is small according to our previous definition.

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 . As shown by their simulation, this rejection frequency, defined by $N^{-1} \sum_{i=1}^N \mathbf{1}(s_i > \alpha)$ where α is some critical value, may converge to some predetermined nominal size (typically 5%). Yet, in principle, it cannot be considered as 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 BE seemingly do.¹² 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, 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. To illustrate this caveat, let us consider a FM 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:¹³

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

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

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

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 in fact there is none.

To analyze how serious this problem could be in practice, we design a simple simulation exercise. First, we generate a FM 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 PCA and apply the individual tests for all the 200 variables.¹⁴ 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 one half of the variables that have

¹²For 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".

¹³Notice that this is possible because of Assumption 1.a.

¹⁴For 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.

breaks. Further, we have checked that, by increasing the size of the breaks, the rejection frequency can rise up to 90% while the true fraction is 50%.

Our LM and Wald tests cannot either identify which particular variables are subject to breaks in the factor loadings but can circumvent 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 problem, contrary to the BE test, our test needs more estimated factors than the true number ($\bar{r} > r$) to maintain power. However, as shown in Proposition 2, use of consistent IC ($\bar{r} = \hat{r}$) implies that $\mathbb{P}[\hat{r} = r + k_1] \rightarrow 1$.

4.5.2 Comparison with the HI test

The HI (2011) test is based on the comparison of the covariance matrices of the estimated factors before and after the break. In view of our results in equation (10) and (11), $\frac{1}{\tau} \sum_1^\tau \hat{F}_t \hat{F}_t' = D_1 \hat{\Sigma}_F D_1' + o_p(1)$, and $\frac{1}{T-\tau} \sum_{\tau+1}^T \hat{F}_t \hat{F}_t' = D_2^* \hat{\Sigma}_F D_2^{*'} + o_p(1)$. Therefore, the HI test is able to detect breaks if $\hat{\Sigma}_F \rightarrow \Sigma_F$ and D_1 and D_2^* converge to different limits as N and T go to infinity. Specifically, their test is defined as:

$$T (A(\pi))' \hat{V}^{-1} A(\pi)$$

where

$$A(\pi) = \text{vech} \left[\frac{1}{\tau} \sum_1^\tau \hat{F}_t \hat{F}_t' - \frac{1}{T-\tau} \sum_{\tau+1}^T \hat{F}_t \hat{F}_t' \right].$$

and \hat{V} is a HAC estimator of the covariance matrix of $A(\pi)$ which is either constructed using the whole sample (LM version of the test) or using subsamples before and after the break (Wald version).

Basically, the HI test exploits the same insight as our tests in converting an infinite-dimensionality problem to a finite one, except that it relies on a different use of the estimated factors. Compared to ours, the HI test uses more information since our LM test only uses the first row (except the first element) of $\frac{1}{\tau} \sum_{t=1}^\tau \hat{F}_t \hat{F}_t'$, while our Wald test uses all the elements of the matrix except the first one ($\frac{1}{\tau} \sum_1^\tau \hat{F}_{1t}^2$).

In principle, it may seem that the use of less information is the price one has to pay to render our testing procedure much simpler than theirs. After all, both steps in our approach can be easily implemented in any conventional statistical software used by practitioners, while HI's test is computationally more burdensome. Yet, our claim in what follows is that, despite exploring less information, our Wald tests are always consistent and exhibit very similar power to theirs. Yet, the LM test may be less powerful (even inconsistent) than the HI test for a rather specific type of alternatives suggested by HI (2011), to be discussed in the sequel.

Consider the simple case where $r = 1$ and $k_1 = 1$. Then we can write the FM as:

$$X_t = A_1 f_{1t} + A_2 f_{2t} + e_t$$

where A_1 and A_2 are $N \times 1$ vectors of factor loadings before and after the break, τ is the break date, $f_{1t} = f_t \mathbf{1}(t \leq \tau)$ and $f_{2t} = f_t \mathbf{1}(t > \tau)$. In this case, $\hat{\Sigma}_F = \frac{1}{T} \sum_{t=1}^T F_t F_t'$ is diagonal by construction. If it is further assumed that $\frac{1}{N} A' A$ is diagonal for given N , where $A = [A_1, A_2]$. Then, it can be shown that $\hat{f}_{1t} = O_p(1)$ and $\hat{f}_{2t} = O_p(\delta_{N,T}^{-1})$ for $t = 1, \dots, \tau$, so that $\frac{1}{T} \sum_{t=1}^{\tau} \hat{f}_{2t}^2 = O_p(\delta_{N,T}^{-2})$, and $\frac{1}{T} \sum_{t=1}^{\tau} \hat{f}_{1t} \hat{f}_{2t} = O_p(\delta_{N,T}^{-2})$ (see Bai and Ng, 2011). Although \hat{S}^{-1} diverges in this case at a rate $O_p(T^{-1/5} \delta_{N,T}^2)$,¹⁵ our LM/ Sup-LM tests may fail to achieve consistency since they rely exclusively on the (squared) covariance term weighted by \hat{S}^{-1} . A heuristic explanation of this problem is that, combining the previous probabilistic orders of magnitude, the LM/Sup-LM tests behave in general as $O_p(T^{4/5} \delta_{N,T}^{-2})$ which, for $T < N$, becomes $O_p(T^{-1/5})$ rendering the test inconsistent while, for $N < T$, it is $O_p(T^{4/5}/N)$ which might lead (albeit not necessarily so) to inconsistency given that our Assumption 6 only assumes $\sqrt{T}/N \rightarrow 0$ as both N and T increase.

Nonetheless, our Wald and Sup-Wald tests always verify consistency (including for this DGP), as will be confirmed by the simulations in section 5. The insight is that, in contrast to the LM tests which only involve the comparison of the covariance terms of \hat{f}_{1t} and \hat{f}_{2t} in each of the two subsamples, the Wald tests rely on a similar comparison of the ratios between such a covariance and the variance of \hat{f}_{2t} (i.e., the OLS-estimated slopes in regression (18) for the two subsamples). Thus, though under the previous DGP $\frac{1}{T} \sum_{t=1}^{\tau} \hat{f}_{1t} \hat{f}_{2t}$ is a small term converging to zero, the fact that $\frac{1}{T} \sum_{t=1}^{\tau} \hat{f}_{2t}^2$ also converges to zero at the same rate implies that the difference in estimated slopes is $O_p(1)$.¹⁶ Yet, there could be some odd draws in the simulations (see the discussion in subsection 5.4 below) where this $O_p(1)$ term might not be bounded away from zero because the numerator in these peculiar realizations turns out to be very small in absolute value for finite samples. Except in these cases, the previous reasoning implies that $\sqrt{T}(\hat{c}_1(\pi) - \hat{c}_2(\pi))$ is $O_p(\sqrt{T})$. However, even when $(\hat{c}_1(\pi) - \hat{c}_2(\pi))$ is very small in this special DGP, the fact that \hat{S}^{-1} diverges at a rate $O_p(T^{-1/5} \delta_{N,T}^2)$, makes our Wald test rather powerful. Indeed, except for the above-mentioned draws, our Wald tests diverge under this DGP at a much faster rate than the corresponding HI tests. To see this, notice that, since the elements in $A(\pi)$ are always $O_p(1)$ and bounded away from zero so that $\sqrt{T}A(\pi)$ is $O_p(\sqrt{T})$, HI's (2011, Assumption 9'a) computation of \hat{V} implies that their test becomes $O_p(T/B_T) = O_p(T^{4/5})$, where $B_T = O_p(T^{1/5})$ is the Bartlett bandwidth parameter in the HAC estimator. In turn, our Wald test with such a bandwidth is $O_p(T \delta_{N,T}^2 / B_T) = O_p(T^{4/5} \delta_{N,T}^2) > O_p(T^{4/5})$. Hence, its rate of divergence under this specific departure from the null will be in general (i.e., except for the odd simulation draws) larger than the rate of HI's tests.

It is also noteworthy that the assumption made by HI (2011) about $A'A/N$ being exactly a diagonal matrix for a given value of N is extremely restrictive, making the special

¹⁵Notice that, as shown in Appendix 3, since \hat{S} is a HAC estimator using a Bartlett kernel, this matrix is always positive definite and therefore $\hat{S} = O_p(T^{1/5} \delta_{N,T}^{-2})$ implies that $\hat{S}^{-1} = O_p(T^{-1/5} \delta_{N,T}^2)$.

¹⁶Note that, although $\frac{1}{T} \sum_{t=1}^{\tau} \hat{f}_{2t}^2$ converges to zero in probability in the first subsample, it is always positive by construction. Therefore $\hat{c}_1(\pi)$ is $O_p(1)$, as is obviously the case of $\hat{c}_2(\pi)$ in the second subsample, where $\frac{1}{1-\tau} \sum_{t=1}^{\tau} \hat{f}_{2t}^2$ converges to a positive constant.

DGP considered above of very little practical relevance. In the slightly more plausible case where, in the limit, $\Sigma_A = \lim \hat{\Sigma}_A$ is diagonal (although this case is also excluded by our Assumptions 1 and 5) but the elements in the sequence $\{A'A/N\}$ do not satisfy exactly the previous assumption, our LM and Wald tests would be both consistent. Indeed, our simulations reported in section 5 below show that ours and HI's tests perform very similarly in terms of power under the type of breaks allowed by our Assumptions 1 and 5, and that the Wald test has even better power properties in small samples (T and $N \leq 100$) than theirs.

REMARK 8. Since our tests are based on a linear regression model, many other available methods in the literature can also be applied to our second-step procedure. For instance, when the break date is not known a priori, one can use the CUSUM type-test first proposed by Brown, Durbin and Evans (1975), and also Chen and Hong's (2011) test via nonparametric regression. Thus, this flexibility allows practitioners to draw conclusions about breaks based on broader evidence than just that provided by a single test.

REMARK 9. Finally, it is important to notice that, although most tests are designed for single break, our tests also have power against multiple breaks, since the second step involves a well-established test for a linear relationship. To see this, consider the simple example of a FM with one factor and two big breaks:

$$\begin{aligned} X_t &= Af_t \cdot 1(t \leq \tau_1) + Bf_t \cdot 1(\tau_1 < t < \tau_2) + Df_t \cdot 1(t \geq \tau_2) + e_t \\ &= Ag_t + Bh_t + Ds_t + e_t \end{aligned}$$

where $g_t = f_t \cdot 1(t \leq \tau_1)$, $h_t = f_t \cdot 1(\tau_1 < t < \tau_2)$, and $s_t = f_t \cdot 1(t \geq \tau_2)$. In view of Proposition 2, Bai and Ng's (2002) IC will lead to the choice of 3 factors which, when estimated by PCA, imply the following result:

$$\begin{pmatrix} \hat{f}_{1t} \\ \hat{f}_{2t} \\ \hat{f}_{3t} \end{pmatrix} = \begin{pmatrix} d_1 & d_4 & d_7 \\ d_2 & d_5 & d_8 \\ d_3 & d_6 & d_9 \end{pmatrix} \begin{pmatrix} g_t \\ h_t \\ s_t \end{pmatrix} + o_p(1).$$

Then, by the definition of g_t, h_t and s_t we have:

$$\begin{aligned} \begin{pmatrix} \hat{f}_{1t} \\ \hat{f}_{2t} \\ \hat{f}_{3t} \end{pmatrix} &= \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} f_t + o_p(1) \text{ for } t = 1, \dots, \tau_1 \\ \begin{pmatrix} \hat{f}_{1t} \\ \hat{f}_{2t} \\ \hat{f}_{3t} \end{pmatrix} &= \begin{pmatrix} d_4 \\ d_5 \\ d_6 \end{pmatrix} f_t + o_p(1) \text{ for } t = \tau_1, \dots, \tau_2 \\ \begin{pmatrix} \hat{f}_{1t} \\ \hat{f}_{2t} \\ \hat{f}_{3t} \end{pmatrix} &= \begin{pmatrix} d_7 \\ d_8 \\ d_9 \end{pmatrix} f_t + o_p(1) \text{ for } t = \tau_2, \dots, T \end{aligned}$$

Hence, we can find one vector $[p_1, p_2, p_3]'$ which is orthogonal to $[d_1, d_2, d_3]'$ and $[d_4, d_5, d_6]'$, plus another vector $[p_4, p_5, p_6]'$ which is orthogonal to $[d_7, d_8, d_9]'$. It is easy to see that $[p_1, p_2, p_3] \neq [p_4, p_5, p_6]$ (otherwise the D matrix will be singular), and thus we can find a breaking relationship between the estimated factors and even use Bai and Perron's (1998, 2003) to detect a second break.

Summing up, compared to the HI tests, our tests use less information but are much easier to implement, and any available test for the existence of breaks in a linear regression model can be used for (18). The price we pay for using less information in our simpler and more flexible procedure is that our LM and Sup-LM tests (but not the Wald tests) may be inconsistent for a special class of DGPs that have very little practical relevance. For less restrictive DGPs, both the LM and Wald tests are consistent, and the latter turns out to be even more powerful than HI's test in small samples. Simulation results supporting these statements are provided in the next section.

5 Simulations

In this section, we first study the finite sample properties of our proposed LM/Wald and Sup-LM/Wald tests. Then comparisons are made with BE test and HI test by means of simulations. Since the only BE test with a known limiting distribution is their pooled statistic when the idiosyncratic components in the factor model are independent, we restrict the comparison to this specific case instead of using their rejection frequency approach whose asymptotic distribution remains unknown. For HI test, we first show the case where our LM and Sup-LM tests lose powers due to the use of less information, and where our Wald and Sup-Wald test remain consistent. Then we compare our Wald test and HI test for small samples. Finally, we use a simple factor-based forecasting model to illustrate the consequences of ignoring big breaks as discussed in Section 3.1.

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 standard 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 standard 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 Newey and West's (1987) HAC estimator.

Table 1 reports the empirical sizes (in percentages) for the LM/Wald tests and Sup-LM/Wald tests using 5% asymptotic critical values for sample sizes (N and T) equal to 100, 150, 200, 250, 300 and 1000.¹⁷ 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 ($\bar{r} = 4 > r = 3$).¹⁸ Although for $N, T \geq 150$, Bai and Ng's (2002) IC tests very often select 3 factors, there are some cases where 2 and 4 are also selected.

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 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, although the Sup-Wald test is too liberal for $T = 100$, in general behaves better than the LM test (see Kim and Perron, 2009). 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.¹⁹ 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 Bai and Ng's (2002) IC with 1000 replications.

As expected, both tests are powerful to detect the breaks as long as $\bar{r} > r = 2$, while the power is trivial when $\bar{r} = r = 2$. However, the Wald tests exhibit in general more power than the LM tests and hence, in line with Kim and Perron's (2009) results, we recommend using them in empirical applications. For safety, we recommend to use data sets with large T (at least around 200), with $N > 100$, in order to run these tests reliably in practice.

5.3 Power comparison with the 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:

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

¹⁷As mentioned earlier, the critical values of the Sup-type tests are taken from Andrews (1993).

¹⁸Notice that the choice of $r = 3$ allows us to analyze the consequences of performing our proposed test with the under-parameterized 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 $\bar{r} = 1$.

¹⁹The results with other types of breaks such as random shifts are similar.

where s_i is the individual LM statistics in BE (2010). This test should converge to a standard 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 performances of the pooled BE test with our Wald test 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 both 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 is generated as a shift of size 0.1 in the mean of the loadings. As before, 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 $\bar{r} = 3$. Notice that, according to Remarks 6 and 7, for $\bar{r} = 2$, our test, will not be able to detect the break, whereas for $\bar{r} = 3$, the pooled BE test will be powerless. However, according to our result in Proposition 2, the use of Bai and Ng's (2002) IC will yield the choice of $\bar{r} = 3$ as a much more likely outcome as N and T increase. For example, for this simulation, on average the IC chooses $\hat{r} = 3$ in 94.6% of the cases.

5.4 Power comparison with the HI test

We first consider the following DGP borrowed from HI(2011): $f_t = \rho f_{t-1} + u_t$, $A_1 = \sqrt{0.8}[L'_1, L'_1]'$, $A_2 = \sqrt{1.2}[L'_2, -L'_2]'$, $X_{it} = A_{1i}f_t + e_{it}$ for $t = 1, \dots, T/2$, and $X_{it} = A_{2i}f_t + e_{it}$ for $t = T/2 + 1, \dots, T$, where L_1 and L_2 are $\frac{1}{2}N \times 1$ vectors with standard normal i.i.d elements, $u_t, e_{it} \sim i.i.d N(0, 1)$, and $\rho = 0.8$.

This is the very special DGP discussed in subsection 4.5.2 where $A'A$ is an exact diagonal matrix, a feature which may lead to the inconsistency of our LM and Sup-LM due to the use of less information. The results presented in Table 4 provide evidence for this theoretical result by showing that, even for $T = N = 500$ the rejection rates of these tests do not exceed 70%. On the contrary, the power of the Wald and Sup-Wald tests is very close or even equal to 100% in all instances. Moreover, in line with our previous argument, it can also be observed that the latter diverge at much faster rate than HI's tests. This can be verified in the last column of Table 4 which displays is log of the average ratios of $Wald/Wald_{HI}$, where $Wald$ and $Wald_{HI}$ denote are Wald test in (21) and the Wald test of HI respectively for a given break date. Yet, despite this much higher divergence rate on average, the power of our Wald test sometimes does not reach 100% because, as discussed above, there are a few ($\leq 0.5\%$) simulation draws where the numerator in (21) is very close to zero. Nonetheless, power is always very close to 100%.

For more general alternatives, where $A'A$ is not an exact diagonal matrix, both our LM and Wald tests and HI test perform equally well for N and T larger than 100. These

results, omitted to save space, are available upon request. However, for smaller sample sizes between 50 and 100, our test outperforms HI test. To illustrate this feature, the (size-adjusted) power curves of ours (solid line) and HI's Wald tests (thin lines), using the Bartlett kernel, are plotted in Figure 1 for the following DGP: $f_t = \rho f_{t-1} + u_t$, $X_{it} = A_{1i}f_t + e_{it}$ for $t = 1, \dots, T/2$, and $X_{it} = (A_{1i} + b)f_t + e_{it}$ for $t = T/2 + 1, \dots, T$, $A_{it}, u_t, e_{it} \sim i.i.d N(0, 1)$, $\rho = 0.8$, and b is the break size which ranges from 0 to 0.5. As can be observed, our Wald test has better power properties than HI's test in all these cases.

5.5 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:

$$X_t = AF_t\mathbf{1}(t \leq \tau) + BF_t\mathbf{1}(t > \tau) + e_t.$$

Secondly, the variable we wish to forecast y_t , which is excluded from 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 standard normal variables, and the break size is characterized by a range of mean shifts between 0 and 1 occurring at half of the time sample size.

The following forecasting procedures are compared in our simulation:

Benchmark 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 in the regression of y_{t+1} on F_t .

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

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

Forecasting 3: We first estimate 4 factors (replicating $r + k_1 = 4$) \hat{F}_t from X_t by PCA, which are then used as predictors in $y_t(1) = \hat{a}'\hat{F}_t$, where \hat{a} is the OLS estimate of a in the regression of y_{t+1} on \hat{F}_t .

The above forecasting exercises are implemented recursively, e.g., at each time t , the data X_t, X_{t-1}, \dots, X_1 and y_t, y_{t-1}, \dots, 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 be 1.

The results obtained from 1000 replications are reported in Figure 2 with the 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, in line with our discussion in Section 3. By contrast, the Forecasting 2 and 3 procedures perform equally well and their MSEs remain constant as the break size increases, although they cannot outperform the benchmark forecasting due to the estimation errors of the factors for the chosen sizes of N and T . In line with our previous analysis, the lesson to be drawn from this exercise is that, in case of a big break, imposing the number of factors a priori can significantly worsen forecasts.

6 An Empirical Application

To provide an empirical applications of our test, we use Stock and Watson's (2009) data set consisting of 144 quarterly time series of nominal and real variables for the US ranging from 1959:Q1 to 2006:Q4. Since not all variables are available for the whole period, we end up using their suggested balanced panel of standardized variables with $T = 190$, $N = 109$ which more or less correspond 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 number of factors using various Bai and Ng's (2002) IC range from 3 to 6, we implement the test for $\bar{r} = 3, 4, 5$ and 6. In view of their better performance, we use the Wald tests. Although other trimmings have been used, we report results using $\Pi = [0.3, 0.7]$ since, being N not too large, it is preferable to use longer subsamples to compute the maxima. It corresponds to the time period ranging from 1973Q3 to 1992Q3 which includes several relevant events like, e.g., the oil price shocks (1973, 1979) and the beginning of great moderation period (1984). The graphs displayed in Figure 3 are the series of Wald tests for different values of \bar{r} , with the horizontal lines representing the 10% critical values of the Wald and Sup-Wald tests.

When compared to the critical values of the χ^2 distributions, the Wald tests reject the null of no big breaks for $\bar{r} = 3, 4, 5, 6$ when the break date is assumed to be known at 1982 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 rather than mild and therefore should not be neglected in subsequent forecasting practices with these factors.

As for the case when the break date is not assumed to be a priori known, we find that the Sup-Wald test rejects the null when $\bar{r} = 4, 5, 6$ (i.e., exceeds the upper horizontal line). Given that our Wald test is consistent when $\bar{r} > r$ (Remark 7), we may conclude that $r = 3$ and that there is a big break. The estimate of the break date is around 1979 (second oil price shock), rather than 1984 which, as mentioned before, is the only candidate considered by BE (2010) as a potential break date in their empirical application with the same data set. Using Bai and Ng's (2002) IC_{p1} criterion we find 3 factors for each the two subsamples pre and post-1979 and 4 factors for the whole sample. Thus, following our results in

subsection 4.4, our preliminary interpretation of that break is as stemming from an abrupt change in the loadings of one factor ($r = 3$ and $k_1 = 1$) rather than from the variance of the factors.

7 Conclusions

In this paper, we propose a simple two-step procedure to test for big structural breaks in the factor loadings of large FMs that circumvents some limitations affecting other available tests, like BE (2010). In particular, we derive the limiting distributions of our test, while theirs remains unknown, and show that it has much better power properties when the choice of the number of factors is based upon Bai and Ng's (2002) consistent IC. We also compare our tests to those developed in independent work by HI (2011, unpublished) which are closely related to ours but differ in their implementation. We show that our regression-based tests are simpler than theirs and that, though we use less information in exchange for simplicity, the information used in our approach is combined in a more efficient way. Further, our Wald test has better power than the HI test when dealing with rather small samples ($N < 100, T < 100$).

Our method can be easily implemented in practice by means of PCA when either 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, when the factor loadings are subject to big breaks and the number of factors is a priori determined (as is conventional in several macroeconomic and financial applications).

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 performs very well especially when $T > 200$. For smaller samples, as it happens with many other Sup-type, bootstrap can improve the finite-sample significance level performance of the test compared to the tabulated asymptotic critical values of Andrews (1993)²⁰. It is high in our research agenda to explore this possibility.

Moreover, as discussed earlier (see Remark 9), our testing approach can allow for multiple big breaks, either through sequential estimation, like in Bai and Perron (1998), or by using an efficient search procedure, like in Bai and Perron (2003), for locating the candidate break dates. Exploring further this issue remains also high in our research agenda.

Finally, though we have outlined in section 4.4 a simple testing procedure to identify whether breaks stem from loadings or the volatility of the factors, we plan to derive other alternative tests based on the rank of the covariance matrix of the estimated factors in different subsamples which can also be extended to test for other sources of parameter instability, like e.g., the appearance of new factors or abrupt changes in the DGP of the idiosyncratic errors.

²⁰See, e.g., Hansen (2000).

Appendix

A.1: Proof of Propositions 1 and 2

The proof proceeds by showing that the errors, factors and loadings in model (6) satisfy Assumptions A to D of Bai and Ng (2002). Then, once these results are proven, Propositions 1 and 2 just follow immediately from application of Theorems 1 and 2 of Bai and Ng (2002). Define $F_t^* = [F_t' \quad G_t^1]'$, $\varepsilon_t = HG_t^2 + e_t$, and $\Gamma = [A \quad \Lambda]$. The proofs of Lemma 1 to Lemma 8 are similar to those in Bai and Ng (2002). To save space and avoid repetition, we put them in our online appendix (<http://www.eco.uc3m.es/~jgonzalo/WP1.html>).

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

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

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

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

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 .

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 $N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\tau_{ij}^*| \leq M^*$.

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

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

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

As mentioned before, once it has been shown 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 their Theorems 1 and 2, with r replaced by $r + k_1$ and F_t replaced by F_t^* .

A.2: Proof of Theorem 1

We only derive the limiting distributions for the two versions of the LM test, since the proofs for the Wald tests are very similar. 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 $D_{(i)}$ denote the i th row of D , and $D_{(\cdot, j)}$ denote the j th column of D . Define $\hat{\mathcal{F}}_t = DF_t$, and $\hat{\mathcal{F}}_{kt} = D_{(k, \cdot)} \times F_t$ as the k th element of $\hat{\mathcal{F}}_t$. Let \hat{F}_{1t} be the first element of \hat{F}_t , and $\hat{F}_{-1t} = [\hat{F}_{2t} \cdots \hat{F}_{rt}]$, while $\hat{\mathcal{F}}_{1t}$ and $\hat{\mathcal{F}}_{-1t}$ can be defined in the same way. Note that $\hat{\mathcal{F}}_t$ depends on N and T because $D = (\hat{F}F'/T)(A'A/N)$ (see Bai and Ng, 2002). For simplicity, let $T\pi$ denote $[T\pi]$.

Note that under H_0 , we allow for the existence of small breaks, so that the model can be written as $X_{it} = A_i F_t + e_{it} + \eta_i G_t^2$. However, since $\eta_i G_t^2$ is $O_p(1/\sqrt{NT})$ by Assumption 1, we can use similar methods as in Appendix A.1 to show that an error term of this order can be ignored and that the asymptotic properties of \hat{F}_t will not be affected (See Remark 5 of Bai, 2009). Therefore, for the sake of simplicity in the presentation below, we eliminate the last term and consider instead the model $X_{it} = A_i F_t + e_{it}$ in the following lemmata (9 to 13) required to prove Lemma 14 which is the key one in the proof of Theorem 1.

Lemma 9.

$$\sup_{\pi \in [0,1]} \left\| \frac{1}{T} \sum_{t=1}^{T\pi} (\hat{F}_t - \hat{\mathcal{F}}_t) F_t' \right\| = O_p(\delta_{N,T}^{-2}).$$

Proof. The proof is similar to Lemma B.2 of Bai (2003). For details see our online appendix. \square

Lemma 10.

$$\sup_{\pi \in [0,1]} \left\| \frac{1}{T} \sum_{t=1}^{T\pi} \hat{F}_t \hat{F}_t' - \frac{1}{T} \sum_{t=1}^{T\pi} \hat{\mathcal{F}}_t \hat{\mathcal{F}}_t' \right\| = O_p(\delta_{N,T}^{-2}).$$

Proof. Note that:

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

Thus,

$$\begin{aligned} & \sup_{\pi \in [0,1]} \left\| \frac{1}{T} \sum_{t=1}^{T\pi} \hat{F}_t \hat{F}_t' - \frac{1}{T} \sum_{t=1}^{T\pi} \hat{\mathcal{F}}_t \hat{\mathcal{F}}_t' \right\| \\ \leq & \sup_{\pi \in [0,1]} \left\| \frac{1}{T} \sum_{t=1}^{T\pi} (\hat{F}_t - DF_t)(\hat{F}_t - DF_t)' \right\| + 2\|D\| \sup_{\pi \in [0,1]} \left\| \frac{1}{T} \sum_{t=1}^{T\pi} (\hat{F}_t - DF_t) F_t' \right\| \\ \leq & \frac{1}{T} \sum_{t=1}^T \|\hat{F}_t - DF_t\|^2 + 2\|D\| \sup_{\pi \in [0,1]} \left\| \frac{1}{T} \sum_{t=1}^{T\pi} (\hat{F}_t - DF_t) F_t' \right\| \end{aligned}$$

since $\frac{1}{T} \sum_{t=1}^T \|\hat{F}_t - DF_t\|^2 = O_p(\delta_{N,T}^{-2})$ and $\sup_{\pi \in [0,1]} \left\| \frac{1}{T} \sum_{t=1}^{T\pi} (\hat{F}_t - DF_t) F_t' \right\|$ is $O_p(\delta_{N,T}^{-2})$ by Lemma 9, the proof is complete. \square

The next two lemmata follow from Lemma 10 and Assumption 6:

Lemma 11.

$$\sup_{\pi \in [0,1]} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T\pi} \hat{F}_{-1t} \hat{F}_{1t} - \frac{1}{\sqrt{T}} \sum_{t=1}^{T\pi} \hat{\mathcal{F}}_{-1t} \hat{\mathcal{F}}_{1t} \right\| = o_p(1).$$

Proof. See Lemma 10 and Assumption 6. \square

Lemma 12.

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

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

Let \Rightarrow denote *weak convergence*, then:

Lemma 13.

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

for $\boldsymbol{\pi} \in [0, 1]$, where $\mathcal{W}_{r-1}(\cdot)$ is a $r-1$ vector of independent Brownian motions on $[0, 1]$.

Proof. The proof is a standard application of Functional CLT. For details see our online appendix. \square

Lemma 14.

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

for $\boldsymbol{\pi} \in [0, 1]$, where the process $\mathcal{B}_{r-1}^0(\boldsymbol{\pi}) = \mathcal{W}_{r-1}(\boldsymbol{\pi}) - \boldsymbol{\pi} \mathcal{W}_{r-1}(1)$ indexed by $\boldsymbol{\pi}$ is a vector of Brownian Bridge on $[0, 1]$.

Proof. If we show that

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

for $\boldsymbol{\pi} \in [0, 1]$ and

$$\sup_{\boldsymbol{\pi} \in [0,1]} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T\pi} \hat{\mathcal{F}}_{-1t} \hat{\mathcal{F}}_{1t} - \frac{1}{\sqrt{T}} \sum_{t=1}^{T\pi} \left[\mathcal{F}_{-1t} \mathcal{F}_{1t} - T^{-1} \sum_{s=1}^T \mathcal{F}_{-1s} \mathcal{F}_{1s} \right] \right\| = o_p(1), \quad (\text{A.6})$$

then the result follows from Lemma 11.

First note that

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

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

$$\hat{\mathcal{F}}_{-1t} \hat{\mathcal{F}}_{1t} = D_{-1} F_t F_t' D_1'$$

and

$$\mathcal{F}_{-1t} \mathcal{F}_{1t} = D_{-1}^* F_t F_t' D_1^{*'}.$$

It follows that:

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

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

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

Combining the above results gives:

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

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

$$\sup_{\pi \in [0,1]} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T\pi} (F_t F_t' - \overline{F_{-1t} F_{1t}}) \right\| = O_p(1).$$

Moreover, it is easy to check that $\|D\| = O_p(1)$ and $\|D - D^*\| = o_p(1)$ (See Bai 2003). Finally, $\left\| \frac{1}{\sqrt{T}} \sum_{s=1}^T \hat{\mathcal{F}}_{-1s} \hat{\mathcal{F}}_{1s}' \right\|$ is $o_p(1)$ by Lemma 12. Then A.6 has been shown and the proof is complete. \square

Theorem 1:

Proof. The results for LM and Sup-LM tests follow from Assumption 9, Lemma 14, and Continuous Mapping Theorem.

For the Wald and Sup-Wald tests, notice that:

$$\begin{aligned} \sqrt{T}(\hat{c}_1(\pi) - \hat{c}_2(\pi)) &= \left(\frac{1}{T} \sum_{t=1}^{\tau} \hat{F}_{-1t} \hat{F}'_{-1t} \right)^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{\tau} \hat{F}_{-1t} \hat{F}_{1t}' \right) - \left(\frac{1}{T} \sum_{t=\tau+1}^T \hat{F}_{-1t} \hat{F}'_{-1t} \right)^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=\tau+1}^T \hat{F}_{-1t} \hat{F}_{1t}' \right) \\ &= \left[\left(\frac{1}{T} \sum_{t=1}^{\tau} \hat{F}_{-1t} \hat{F}'_{-1t} \right)^{-1} + \left(I - \frac{1}{T} \sum_{t=1}^{\tau} \hat{F}_{-1t} \hat{F}'_{-1t} \right)^{-1} \right] \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{\tau} \hat{F}_{-1t} \hat{F}_{1t}' \right). \end{aligned}$$

By Lemma 10 and that $D - D^* = o_p(1)$, we have:

$$\frac{1}{T} \sum_{t=1}^{\tau} \hat{F}_{-1t} \hat{F}'_{-1t} = \pi \frac{1}{\tau} \sum_{t=1}^{\tau} \hat{F}_{-1t} \hat{F}'_{-1t} = \pi \frac{1}{\tau} \sum_{t=1}^{\tau} \mathcal{F}_{-1t} \mathcal{F}'_{-1t} + o_p(1). \quad (\text{A.7})$$

when $\tau = T$ ($\pi = \tau/T = 1$), this implies

$$I_{r-1} = \frac{1}{T} \sum_{t=1}^T \hat{F}_{-1t} \hat{F}'_{-1t} = \frac{1}{T} \sum_{t=1}^T \mathcal{F}_{-1t} \mathcal{F}'_{-1t} + o_p(1).$$

By LLN we have $E(\mathcal{F}_{-1t} \mathcal{F}'_{-1t}) = I_{r-1}$. Applying LLN again to A.7 gives:

$$\frac{1}{T} \sum_{t=1}^{\tau} \hat{F}_{-1t} \hat{F}'_{-1t} \xrightarrow{p} \pi I_{r-1}$$

as N and T go to infinity. It then follows from Lemma 14 that:

$$\sqrt{T}(\hat{c}_1(\pi) - \hat{c}_2(\pi)) \Rightarrow \frac{S^{1/2} \mathcal{B}_{r-1}^0(\pi)}{\pi(1-\pi)}$$

and the limit distributions of the Sup-Wald and Wald tests follow easily. \square

A.3: Consistent Estimator of S

We now discuss the consistent estimator of S using the HAC estimator of Newey and West (1987).

Recall that $S = \lim \text{Var} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \mathcal{F}_{-1t} \mathcal{F}_{1t}' \right)$.

Notice that $E(\mathcal{F}_{-1t} \mathcal{F}_{1t}') = 0$, because $\frac{1}{T} \sum_{t=1}^T \mathcal{F}_{-1t} \mathcal{F}_{1t}' = \frac{1}{T} \sum_{t=1}^T \hat{F}_{-1t} \hat{F}_{1t}' + o_p(1)$, and $\frac{1}{T} \sum_{t=1}^T \mathcal{F}_{-1t} \mathcal{F}_{1t}' \rightarrow E(\mathcal{F}_{-1t} \mathcal{F}_{1t}')$ by Assumption 7 and $\frac{1}{T} \sum_{t=1}^T \hat{F}_{-1t} \hat{F}_{1t}' = 0$.

First, we define the infeasible estimator of S :

$$\hat{S}(\mathcal{F}) = \hat{\Gamma}_0(\mathcal{F}) + \sum_{j=1}^m w(j, m) [\hat{\Gamma}_j(\mathcal{F}) + \hat{\Gamma}_j(\mathcal{F})']$$

where $m = O_p(T^{\frac{1}{5}})$, $w(j, m) = 1 - \frac{j}{m+1}$ is the Bartlett kernel, and

$$\hat{\Gamma}_j(\mathcal{F}) = \frac{1}{T} \sum_{t=j+1}^T \mathcal{F}_{-1t} \mathcal{F}_{1t} \mathcal{F}_{1t-j} \mathcal{F}'_{-1t-j}.$$

Since the above estimator is a HAC, it is natural to make the following assumption:

Assumption 11. $\|\hat{S}(\mathcal{F}) - S\| = o_p(1)$.

Next we consider a feasible estimator of S where \mathcal{F}_t is replaced by \hat{F}_t :

$$\hat{S}(\hat{F}) = \hat{\Gamma}_0(\hat{F}) + \sum_{j=1}^m w(j, m) [\hat{\Gamma}_j(\hat{F}) + \hat{\Gamma}_j(\hat{F})']$$

where

$$\hat{\Gamma}_j(\hat{F}) = \frac{1}{T} \sum_{t=j+1}^T \hat{F}_{-1t} \hat{F}_{1t} \hat{F}_{1t-j} \hat{F}'_{-1t-j}.$$

If we further assume:

Assumption 12. $\|D^* - D\| = o_p(T^{-\frac{1}{5}})$.

then we have the following results:

Proposition 3. *Assume that Assumptions 1 to 12 hold, under the null $H_0 : k_1 = 0$, we have*

$$\|\hat{S}(\hat{F}) - S\| = o_p(1).$$

Proof. Given Assumption 10, it suffices to show that

$$\|\hat{S}(\hat{F}) - \hat{S}(\mathcal{F})\| = o_p(1).$$

It is easy to see that:

$$\|\hat{S}(\hat{F}) - \hat{S}(\mathcal{F})\| \leq 2 \sum_{j=0}^m \|\hat{\Gamma}_j(\hat{F}) - \hat{\Gamma}_j(\mathcal{F})\|,$$

and

$$\begin{aligned} & \|\hat{\Gamma}_j(\hat{F}) - \hat{\Gamma}_j(\mathcal{F})\| \\ & \leq \frac{1}{T} \sum_{t=1}^T \left\| \hat{F}_{-1t} \hat{F}_{1t} \hat{F}_{1t-j} \hat{F}'_{-1t-j} - \mathcal{F}_{-1t} \mathcal{F}_{1t} \mathcal{F}_{1t-j} \mathcal{F}'_{-1t-j} \right\| \\ & = O_p(\|\hat{F}_t - \mathcal{F}_t\|) \\ & = O_p(\delta_{N,T}^{-1}) + O_p(\|D^* - D\|). \end{aligned}$$

Notice that the result in the last line follows from the fact that

$$\|\hat{F}_t - \mathcal{F}_t\| \leq \|\hat{F}_t - DF_t\| + \|DF_t - D^*F_t\|$$

and $\|\hat{F}_t - DF_t\| = O_p(\delta_{N,T}^{-1})$.

Finally we have

$$\|\hat{S}(\hat{F}) - \hat{S}(\mathcal{F})\| = O_p(T^{\frac{1}{5}} \delta_{N,T}^{-1}) + O_p(\|D^* - D\|) O_p(T^{\frac{1}{5}}) = o_p(1)$$

by Assumptions 6 and 11. □

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. (2009). Panel data models with interactive fixed effects. *Econometrica* 77(4), 1229–1279.
- 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 (2006a). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica* 74(4), 1133–1150.
- Bai, J. and S. Ng (2006b). Evaluating latent and observed factors in macroeconomics and finance. *Journal of Econometrics* 131(1-2), 507–537.
- Bai, J. and S. Ng (2011). Principal components estimation and identification of the factors. *Columbia University, Working paper*.
- 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. Elias (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics* 120(1), 387–422.
- Breitung, J. and S. Eickmeier (2010). Testing for structural breaks in dynamic factor models. *Journal of Econometrics*, forthcoming.
- Brown, R., J. Durbin, and J. Evans (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, 149–192.
- Charnavoki, V. and J. Dolado (2012). The effects of global shocks on small commodity-exporting economies: New evidence from Canada. *CEPR. DP. 8825*.
- Chen, B. and Y. Hong (2011). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, forthcoming.

- Chen, L. (2012). Identifying observed factors in approximate factor models: estimation and hypothesis testing. *Universidad Carlos III de Madrid, Working paper*.
- Chen, L., J. Dolado, and J. Gonzalo (2011). Identifying the source of structural breaks in large factor models. *In progress*.
- Chen, N., R. Roll, and S. Ross (1986). Economic forces and the stock market. *Journal of Business*, 383–403.
- 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.
- Gagliardini, P. and C. Gouriéroux (2011). Efficiency in large dynamic panel models with common factor. *U. of Toronto (mimeo)*.
- Gali, J. and L. Gambetti (2009). On the sources of the great moderation. *American Economic Journal. Macroeconomics* 1(1), 26–57.
- Giannone, D. (2007). Discussion on: "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.
- Han, X. and A. Inoue (2011). Tests for parameter instability in dynamic factor models. *North Carolina State University, Working paper*.
- Hansen, B. (2000). Testing for structural change in conditional models. *Journal of Econometrics* 97(1), 93–115.
- Kim, D. and P. Perron (2009). Unit root tests allowing for a break in the trend function at an unknown time under both the null and alternative hypotheses. *Journal of Econometrics* 148(1), 1–13.
- Newey, W. and K. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3), 703–708.
- Perron, P. (2006). Dealing with structural breaks. *Palgrave handbook of econometrics* 1, 278–352.
- 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.
- Shanken, J. and M. Weinstein (2006). Economic forces and the stock market revisited. *Journal of Empirical Finance* 13(2), 129–144.

- 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.

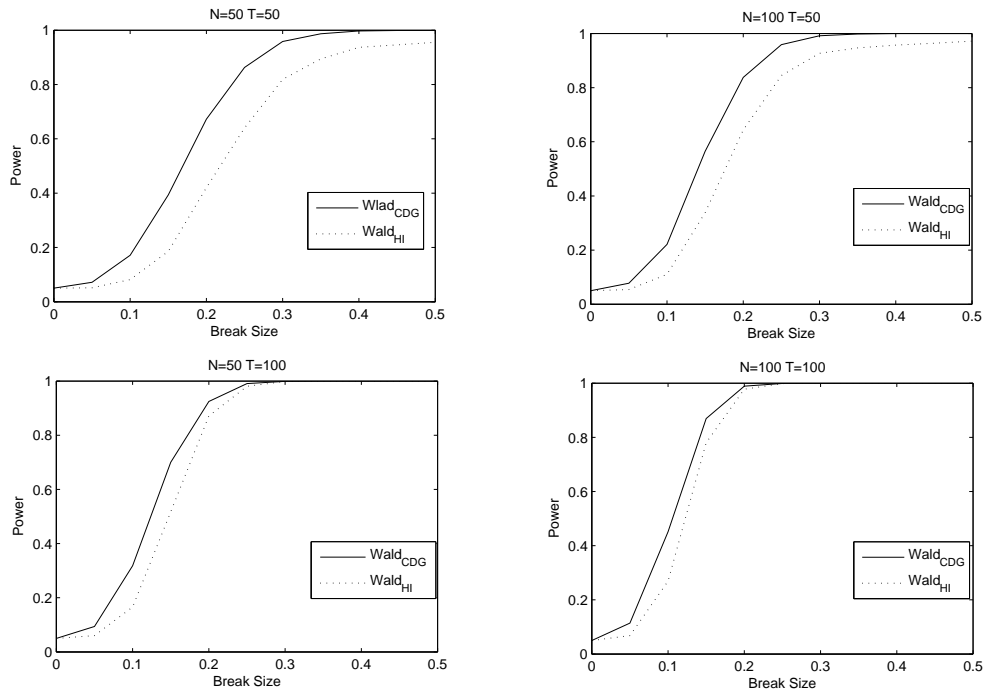


Figure 1: Comparisons of Wald tests.

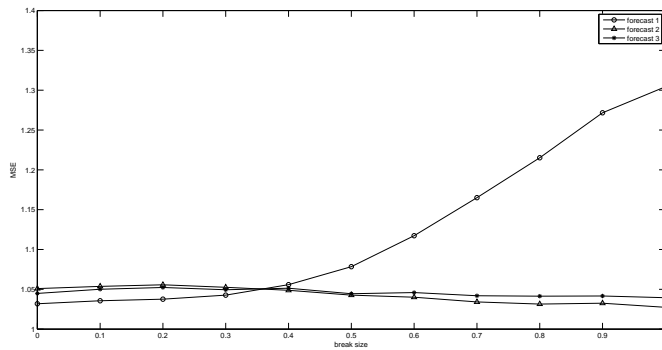


Figure 2: The MSEs of different forecasting methods (see Section 5.4).

Table 1: Actual Size of the Nominal 5% Size Tests for Different Sample Sizes and Different Choices of the Number of Factors (\bar{r}) in a Factor Model with $r = 3$.

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

Notes: The DGP is $X_{it} = \sum_{k=1}^3 \alpha_{ik} F_{kt} + e_{it}$ where $F_{kt} = \phi_k F_{k,t-1} + v_{kt}$; α_{ik}, e_{it} and v_{kt} are i.i.d standard normal variables, and $\phi_1 = 0.8, \phi_2 = 0.5, \phi_3 = 0.2$. The number of replications is 1000. The reported sizes correspond to the LM and Wald tests and their Sup-type versions. The potential breaking date $\tau = T/2$ is considered to be a priori known for the LM/Wald tests while Π is chosen as $[0.15, 0.85]$ for the Sup-type versions. The covariance matrix S is estimated using the HAC estimator of Newey and West (1987).

Table 2: Power of Nominal 5% Size Tests for Different Sample Sizes and Different Choices of the Number of Factors (\bar{r}) in a Factor Model with $r = 2$ and Shifts in the Factor Loadings.

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

Notes: The DGP is $X_{it} = \sum_{k=1}^2 \alpha_{ik} F_{kt} + e_{it}$ where $F_{kt} = \phi_k F_{k,t-1} + v_{kt}$; α_{ik} , e_{it} , and v_{kt} are i.i.d standard normal variables, and $\phi_1 = 0.8$, $\phi_2 = 0.2$. The number of replications is 1000. The shifts in the means of the factor loadings are 0.4 and 0.2 at $\tau = T/2$. The other characteristics of the Monte Carlo experiment are as in Table 1.

Table 3: Size and Power Comparison of BE and LM Tests at Nominal 5% Size for Different Sample Sizes and Different Choices of \bar{r} in a Factor Model with $r = 2$ and No Shift or One Shift in the Factor Loadings.

N	T	no break, $\bar{r} = 2$		1 break, $\bar{r} = 2$		1 break, $\bar{r} = 3$	
		BE	Wald	BE	Wald	BE	Wald
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

Notes: The DGP is as in Table 2. The shift in the mean of the factor loadings is either zero (no break) or 0.1 (break).

Table 4: DGP: $X_t = \lambda_1 f_{1t} + \lambda_2 f_{2t} + e_t$, $\lambda_1 = \sqrt{0.8}[L'_1, L'_1]'$, $\Lambda_2 = \sqrt{1.2}[L'_2, -L'_2]'$, where L_1 and L_2 are $N/2 \times 1$ matrices with i.i.d normal elements, $f_{1t} = f_t * \mathbf{1}(t < \tau)$, $f_{2t} = f_t * \mathbf{1}(t \geq \tau)$, $f_t = 0.8f_{t-1} + u_t$, u_t and $e_{it} \sim$ i.i.d $N(0,1)$. $\bar{r} = 2$.

N	T	LM	Sup LM	Wald	Sup Wald	$\log\left(\frac{Wald}{Wald_{lit}}\right)$
100	100	54.0	39.3	99.5	100	7.20
100	150	58.0	48.7	99.5	100	7.14
100	200	61.3	52.2	99.4	100	7.07
100	250	61.0	51.1	99.6	100	6.97
100	300	63.6	55.1	99.4	100	6.91
150	100	58.0	43.4	99.9	100	8.11
150	150	61.1	50.5	99.6	100	7.96
150	200	61.0	52.5	100	100	7.79
150	250	64.7	57.6	99.7	100	7.77
150	300	62.7	53.9	99.9	100	7.75
200	100	56.5	42.0	99.9	100	8.64
200	150	59.5	49.9	100	100	8.50
200	200	64.0	57.5	99.9	100	8.42
200	250	63.6	54.7	100	100	8.30
200	300	62.8	55.2	100	100	8.28
250	100	56.2	41.9	99.6	100	9.04
250	150	59.7	49.4	99.8	100	8.94
250	200	63.3	53.9	99.7	100	8.82
250	250	64.4	57.0	99.9	100	8.81
250	300	65.0	59.1	99.8	100	8.77
300	100	55.9	42.1	99.8	100	9.46
300	150	57.7	48.0	99.8	100	9.27
300	200	60.6	50.2	99.5	100	9.23
300	250	64.4	56.2	99.7	100	9.19
300	300	64.6	57.0	99.9	100	9.17
500	500	67.4	61.8	100	100	10.03

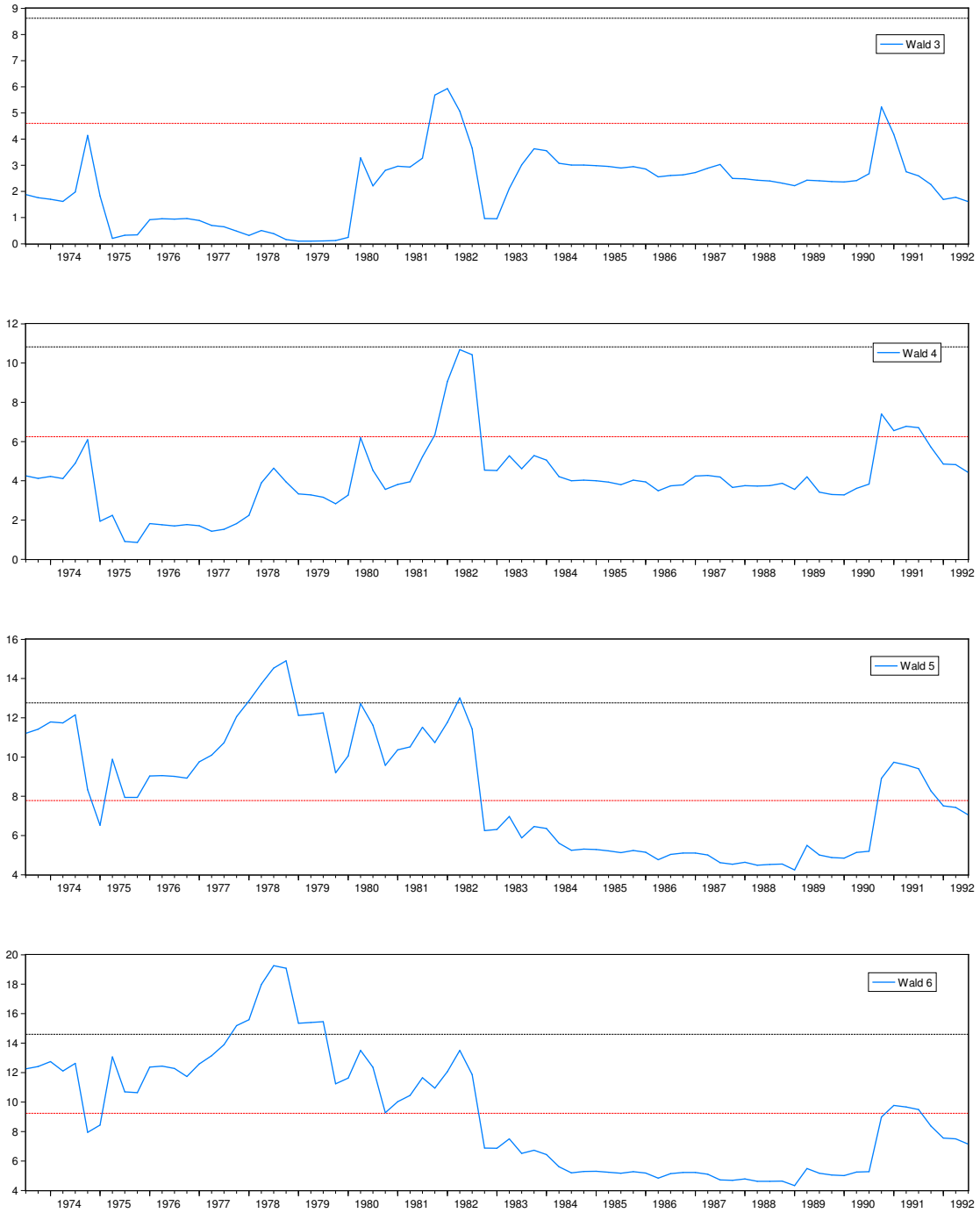


Figure 3: US data set. The Wald tests using the trimming $\Pi = [0.3, 0.7]$, for $\bar{r} = 3$ to 6 (from top to bottom), and the corresponding 10% asymptotic critical values (horizontal dotted lines) for the Wald (red) and Sup-Wald Test (black).