UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 11-41
Economic Series
December, 2011

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 916249875

# Detecting Big Structural Breaks in Large Factor Models*

Liang Chen† Juan J. Dolado‡ Jesús Gonzalo§

First Draft: May 2010
This draft: December 2011

**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 Honor 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 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, in 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 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.

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 factors. 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 version from the date of the break onwards. Focusing on the 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) their test is mostly heuristic since its limiting distribution remains unknown except in some restrictive cases; and (ii) it lacks 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 factor model 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 the factor loadings which allows for different types of breaks and does not suffer from the previous shortcomings. In particular, we first derive some asymptotic results finding that, in contrast to the consistency result for the original factor space under 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. To avoid this problem, we then propose a simple regression-based testing procedure which, as sketched earlier, is based on the idea that that a model with big breaks in the loadings can be reparameterized as a model 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 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 the standard Chow Test or the Sup-type Test of Andrews (1993), depending on whether the date of the break is treated as known or unknown, is then 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, we also provide some guidance, based on the rank properties of the covariance matrix of the estimated factors in different subsamples, 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. Finally, we illustrate the finite sample performance of our tests using simulations, as well as provide a few empirical applications of how to implement our testing approach in practice.

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 independent research, adopt a similar approach to ours in testing for big breaks. The two approaches, however, differ. In effect, rather than using a simple regression-based approach to avoid the infinite-dimensionality problem, as we do here, HI (2011) propose to 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 a very specific and highly unrealistic data generating process (DGP), where one of our proposed tests based on the LM principle might be inconsistent. Yet, even in this special

case, an alternative regression-based test based on the Wald principle has high power. We also show that, in more general and realistic cases, both our LM and Wald test behave well, with the Wald test being even more powerful than the corresponding HI's test for small sample sizes such as $N = T = 50$. An additional advantage of considering our simple linear-regression setup is that it permits to use many other existing methods for testing structural breaks (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 precise definitions of two types of instabilities considered here: *big* and *small breaks*. In Section 3, we analyze the consequences of big breaks on the choice of the number of factors and their estimation, as well as the effects of those breaks on 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 two empirical applications. Finally, Section 7 concludes. Three appendices contain detailed proofs of the main results.

# 2 Notation and Preliminaries

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

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

where $X_t$ is the $N \times 1$ vector of observed variables, $A = (\alpha_1, \ldots, \alpha_N)'$ is the $N \times r$ matrix of factor loadings, $r$ is the number of common factors which is finite, $F_t = (f_{t1}, \ldots, f_{tr})'$ is the $r \times 1$ vector of common factors, and $e_t$ is the $N \times 1$ vector of idiosyncratic errors. In the case of dynamic factor models, all the common factors $f_t$ and their lags are stacked into $F_t$. Thus, a dynamic factor model with $r$ dynamic factors and $p$ lags of these factors can be written as a static factor model 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 factor models 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 $\tau$:

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

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

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

rewritten as:

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

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

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 $\Lambda$ 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:

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

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

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

**Assumption 1.** ***Breaks***

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

*b.* $\eta_i = \frac{\kappa_i}{\sqrt{NT}}$ *for* $i = 1, 2, \ldots, N$ *and* $E\|\kappa_i\|^8 < \infty$.

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

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 $\Sigma$, respectively, while $[T\pi]$ denotes the integer part of $T \times \pi$ for $\pi \in [0, 1]$. Then

**Assumption 2.** ***Factors****:* $E(F_t) = 0$, $E||F_t||^8 < \infty$, $T^{-1}\sum_{t=1}^{T} F_t F_t' \to \Sigma_F$ *and* $T^{-1}\sum_{t=1}^{\tau} F_t F_t' \to \pi^* \Sigma_F$ *as* $T \to \infty$ *for some positive definite matrix* $\Sigma_F$ *where* $\pi^* = \lim_{T\to\infty} \frac{\tau}{T}$.

**Assumption 3.** ***Factor Loadings****:* $E||\alpha_i||^4 \leq M < \infty$, *and* $N^{-1}A'A \to \Sigma_A$, $N^{-1}\Gamma'\Gamma \to \Sigma_\Gamma$ *as* $N \to \infty$ *for some positive definite matrix* $\Sigma_A$ *and* $\Sigma_\Gamma$, *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 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 factor models allowing for weak cross-sectional and temporal correlations between the errors ( see Bai and Ng, 2002). Notice that Assumption 3 excludes in our specific setup the case where a new (old) factor appears (disappears) after the break since this event would imply that $\Sigma_\Gamma$ 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 some stable properties for the factors in order to test the stability of the loadings. We also allow the different factors to be correlated at all leads and lags. Assumption 5 on the independence among the different groups is stronger than the usual assumptions made by Bai and Ng (2002). Notice, however, that we could have also assumed some dependence between these groups and then impose some restrictions on this dependence when necessary. Yet, this would complicate the proofs without essentially altering the 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 the structural breaks on the estimation of factors based on PCA, and on the estimation of the number of factors based on the information criteria proposed by Bai and Ng (2002). 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 estimated by PCA is inconsistent for the true factor space, and 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:

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

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

Let $\bar{r}$ be the selected number of factors, either by 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 $X = [\bar{X}_1, \bar{X}_2 \ldots \bar{X}_T]'$, $\bar{X}_t = [X_{t1}, X_{t2}, \ldots, X_{tN}]'$, $\hat{F} = [\hat{F}_1, \hat{F}_2, \ldots, \hat{F}_T]'$. 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:*

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

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

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

for a $\bar{r} \times r$ matrix $D$ of full rank. This just trivially replicates the well-known consistency result of Bai and Ng (2002)[1].

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

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

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

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

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

where $D_2^* = D_1 + [D_2 \quad 0]$, 0 is a $\bar{r} \times (r - k_1)$ zero matrix, 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 $D_1$ 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):

[1] Notice 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 principal component of $X_t$, $\bar{r}$ can be larger than $r$.

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

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

$$\begin{aligned} y_t &= (a'D_1^-)(D_1F_t) + b'W_t + u_t \ \text{ for } t = 1,2,\ldots,\tau, \\ y_t &= (a'D_2^{*-})(D_2^*F_t) + b'W_t + u_t \ \text{for } t = \tau+1,\ldots,T \end{aligned}$$

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

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

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

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

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

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

A straightforward implication is that if we were to impose on a priori grounds the number of factors, ignoring therefore the set of regressors $\hat{F}_t\mathbf{1}(t > \tau)$ in (15), the estimation of $b$ will, in general, become inconsistent due to omitted variables. 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 Eliasz (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.

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

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

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

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

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

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 as a priori the true number of factors, the estimation of FAR will be robust to the estimation of factors under big breaks if the number of factors is overestimated. Notice that a similar argument will render inconsistent the impulse response functions in FAVAR models where the regressand in (12) becomes $y_{t+1} = (F_{t+1}, W_{t+1})'$. As a result, in order to run regression (17), our advice is not to impose a priori the number of factors unless a formal test of big breaks exist 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 information criteria proposed by Bai and Ng (2002). Then the following result holds:

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

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

Again, when there are no big breaks ($k_1 = 0$), this result trivially replicates Theorem 2 of Bai and Ng (2002). However, under big breaks ($k_1 > 0$), their information criteria will overestimate the number of original factors by the number of big breaks ($0 < k_1 \leq r$) because, as shown above, a factor model with this type of break admits a representation without 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

From the previous discussion, we have found that the factor space and the number of factors are both consistently estimated only when mild breaks occur. Therefore, our goal

here is to develop a test for big breaks.

If we were to follow the usual approach in the literature to test 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 $N \to \infty$ and $T \to \infty$. In other words, this kind of local alternatives, for which the usual test should have no trivial power, is not relevant for the large factor models we consider here. 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 $\widehat{F}$ 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 $\widehat{r}$ 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**:**

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

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) information criteria ($\bar{r} = \hat{r}$) applied to the whole sample (t=1,...,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:*

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

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

It is important to notice 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 is obviously not verified 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 versions* when no prior knowledge about the break date exists. Moreover, since the Wald, LR, and LM test statistics have the same asymptotic distribution under the null, we focus on the LM and Wald tests because they are simpler to compute.

Define $D^* = V^{-1/2}\Gamma'\Sigma_A^{1/2}$ as the limit of the matrix $D$ in equation (8), where $V = diag(v_1, v_2, \ldots, v_r)$, $v_1 > v_2 > \ldots > v_r$ are the eigenvalues of $\Sigma_A^{1/2}\Sigma_F\Sigma_A^{1/2}$, and $\Gamma$ is the corresponding eigenvector matrix (See Bai, 2003). Define $\mathscr{F}_{1t} = D_1^* F_t$ and $\mathscr{F}_{-1t} = D_{-1}^* F_t$, 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 $S = \lim \mathrm{Var}\big(\frac{1}{T}\sum_{t=1}^{T}\mathscr{F}_{-1t}\mathscr{F}_{1t}\big)$.

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

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

where $\bar{\pi} = \tau/T$, $\tau$ is a pre-assumed date for the potential break, $\hat{u}_t$ is the residuals in the OLS regression of (18)[2], and $\hat{S}$ is a consistent estimator of $S$.[3]

The Sup-LM statistic is defined as:

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

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

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

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

and

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

where $\hat{c}_1(\pi)$ and $\hat{c}_2(\pi)$ are OLS estimates of $c$ using subsamples before and after the break point : $[T\pi]$[4].

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

[3] See Appendix A.3 for discussions on the estimation of $S$.

[4] We can also construct the Wald test as $T\Big(\hat{c}_1(\bar{\pi}) - \hat{c}_2(\bar{\pi})\Big)'\Big(\frac{\hat{S}_1}{\pi} + \frac{\hat{S}_2}{(1-\pi)}\Big)^{-1}\Big(\hat{c}_1(\bar{\pi}) - \hat{c}_2(\bar{\pi})\Big)$ 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 using the full sample estimation of $S$, as in (21) and (22).

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

$$X_t = A f_{1t} + B g_t + e_t$$

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

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

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

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

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

which imply that:

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

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

Thus, we can observe that the two estimated factors are linearly related and that the coefficients $\frac{d_1}{d_3}$ and $\frac{d_1+d_2}{d_3+d_4}$ before and after the break date must be different due to the non-singularity of the $D$ 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 an special alternative where our LM test may become inconsistent while our Wald test still exhibits very high power.

Likewise, if the break date $\tau$ is not a priori assumed to be known, the Sup-type test will yield a natural estimate of $\tau$ 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 the information criteria, and it converges to the true one under the null hypothesis of no big break, we start with the most interesting case where $\bar{r} = r$.

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

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

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

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

**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.** *$\left\|\hat{S} - S\right\| = o_p(1)$, 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 $v_t = [v_{1t} \ldots v_{rt}]'$ 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 which is the one used in our simulations below.[5]

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

[5]Though 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 $N, T \to \infty$ , we have that both the LM and Wald tests verify*

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

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

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

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

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

It is easy to show that Theorem 1 still holds when $\bar{r} < r$. However, when $\bar{r} > r$, the matrix $D^*$ may not be asymptotically full ranked. Yet, as simulations in section 5 show, in such an instance Theorem 1 still provides a reasonably good approximation of the distribution of test statistics in finite samples. Moreover, the case where $\bar{r} > r$ can be avoided if, instead of relying on a priori choice of $\bar{r}$, practitioners follow Bai and Ng's (2002) information criteria, 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 (with the whole sample) changes at time $\tau$, so that big breaks can be detected.

First, let us consider the case where $r < \bar{r} \leq r + k_1$ so that $D_1$ and $D_2^*$ in (10) and (11) become $\bar{r} \times r$ matrices with full column rank. Notice that, since $r < \bar{r}$, we can always find $\bar{r} \times 1$ vectors $\rho_1$ and $\rho_2$ 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 $\rho_1'$ and $\rho_2'$ leads to:

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

which, after normalizing the first elements of $\rho_1$ and $\rho_2$ to be 1, implies that:

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

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

Next, to show that $\rho_1^* \neq \rho_2^*$, we proceed as follows. Suppose that $\gamma \in \text{Null}(D_1')$ 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 $\bar{r} \times (r + k_1)$ matrix, 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 $\bar{r}-r>0$, we can always find two non-zero vectors such that $\rho_1 \neq \rho_2$ .

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 $\hat{\rho}_1 \to \theta_1$ and $\hat{\rho}_2 \to \theta_2$, but generally we cannot verify that $\theta_1 \neq \theta_2$.

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

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

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

Consequently, as will be confirmed by our simulations below, our test may fail to detect the breaks when a value of $\bar{r}$ is a priori imposed such that $\bar{r} \leq r$ or $\bar{r} > r+k_1$. However, this is unlikely to happen due to two reasons. First, we usually equate the number of factors with the estimated ones, ( $\bar{r} = \hat{r}$) and we have shown that $\mathbb{P}[\hat{r} = r+k_1] \to 1$. Secondly, instead of using a single value, we can try different values of $\bar{r}$. Then, under the null, we should not detect any break no matter which value of $\bar{r}$ is used while, under the alternative, we should detect breaks when $\bar{r}$ lies between $r$ and $r+k_1$.

## 4.4 Disentangling breaks in loadings from breaks in factors

A potential critique of our LM and Wald tests is that they cannot differentiate between breaks in factor loadings and breaks in the covariance matrix of factors. For illustrative purposes, consider a factor model 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 $t = 1,\ldots,T/2$, and $E(F_t F_t') = \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$ for $t = T/2+1,\ldots,T$, with $\rho \neq 0$ and 1. If we further assume $\Sigma_A = \lim_{N\to\infty} A'A/N$ 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 $2 \times 1$ vector. In this case, regressing $\hat{F}_{1t}$ on $\hat{F}_{2t}$ will yield estimated coefficients close to $\rho$ and $-\rho$ before and after the break. 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, in practice, it is plausible that the factor dynamics are also subject to structural breaks. For instance, if the factors are interpreted as common shocks to the economy, then it is likely that the 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. It is noteworthy that this identification problem not only affects our test but also the competing ones in the literature. For example, as discussed in the Introduction, the HI tests, being based on the comparison of the covariance matrices of the estimated factors before and after the break, will also exhibit power against breaks in the factors as long as the estimated factors are consistent.[6] 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 when 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 $\tau$. 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 $\tau$ 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 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, in the first DGP, the covariance matrices of the estimated factors before and after the break converge to matrices with reduced rank ( i.e, equal to 1 in this example) given by

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

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, the corresponding covariance matrices will converge

[6] In view of Bai and Ng (2002), 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 matrix.

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

$$\begin{aligned} \widehat{\Sigma}_1 &\Rightarrow D\Sigma_1 D' \quad for\ t = 1,..,\tau, \\ \widehat{\Sigma}_2 &\Rightarrow D\Sigma_2 D' \quad for\ t = \tau+1,..,T. \end{aligned}$$

Therefore, when our tests detect a break in the factor model, 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) consistent information criteria 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 information criteria will choose this time 2 factors in each subsample.[7]

## 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 comparing our testing procedure with theirs. In our view, the BE test suffers from three shortcomings which are worth mentioning before the comparison is made.

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

Secondly, their testing procedure is mainly heuristic. Their null hypothesis is $A = B$, or $\alpha_i = \beta_i$ for all $i = 1,\ldots,N$, rather than $\alpha_j = \beta_j$ for a specific $j$.[8] They construct $N$ test statistics (denoted by $s_i$, $i = 1,\ldots,N$) for each of the $N$ variables, but do not derive

[7] In Chen, Dolado and Gonzalo (2011, 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.

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

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 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 $\alpha$ 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.[9] A key presumption for their individual test to work properly is that the estimated factors $\hat{F}_t$ can replace the true factors, even under the alternative hypothesis (given that the number of factors is correctly estimated). As we have shown before, the true factor space can only be consistently estimated under the null of no break at all or only small breaks. By contrast, when big breaks exist, the space of the true factors is not well estimated (see equations (10) and (11)). If we plug in the estimated factors in this case, some variables that have constant loadings may be detected to have breaks due to the poor estimation of the factors. To illustrate this caveat, let us consider a factor model with big breaks in the factor loadings where we select the right number of factors $\bar{r} = r$, and there is one of the variables $X_{it}$ that has constant loadings:[10]

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

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

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

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

To analyze how serious this problem could be in practice, we design a simple simulation. First, we generate a factor model with $N = T = 200$, $r = 1$, where the first 100 variables have constant factor loadings while the remaining 100 variables have big breaks in their loadings. Then we estimate the factors by PCA and apply the individual tests for all the 200 variables.[11] Applying the BE test, we find that rejection frequency for all the

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

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

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

200 variables is 53.07%, close to the proportion of variables that have breaks. However, the rejection frequencies for the first and second 100 variables are 52.98% and 53.15%, respectively, which means that we falsely reject the null for more than half of the variables that are stable while we reject the null correctly for only half of the variables that have breaks. Further, 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 one, contrary to the BE test, our test needs more estimated factors than the true number ($r+k_1 \geq \bar{r} > r$) to maintain power. However, this overestimation it is still preferable to using the BE test because in practice the number of factors to estimate is chosen by means of the information criteria ($\bar{r} = \hat{r}$), and we have shown in Proposition 2 that $\mathbb{P}[\hat{r} = r+k_1] \to 1$.

### 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 $\hat{\Sigma}_F \to \Sigma_F$ and $D_1$ and $D_2^*$ converge to different limits as $N$ and $T$ go to infinity. Specifically, their test is defined as:

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

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 look 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, such as Eviews, etc., while theirs is computationally more burdensome. Yet, our claim in what follows is that, despite exploring less information, our tests are still consistent and exhibit very similar power to theirs, with the only exception of the LM test (not the Wald

test) that may be less powerful 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 factor model 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, $\tau$ 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 $t = 1, \ldots, \tau$, 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 $O_p(T^{-1/5}\delta_{N,T}^2)$,[12] 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 $N < T$, is $O_p(T^{4/5}/N)$ which might lead (albeit not necessarily so) to inconsistency given that our Assumption 6 only assumes $\sqrt{T}/N \to 0$ as both $N$ and $T$ increase.

Nonetheless, our Wald and Sup-Wald tests are always consistent (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 $\frac{1}{\tau}\sum_{t=1}^{\tau} \hat{f}_{1t}\hat{f}_{2t}$ is a small term converging to zero, the fact that $\frac{1}{\tau}\sum_{t=1}^{\tau} \hat{f}_{2t}^2$ also converges to zero at the same rate implies that the difference in estimated slopes is $O_p(1)$.[13] 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 quite 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

[12] Notice 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 $\hat{S}^{-1} = O_p(T^{-1/5}\delta_{N,T}^2)$.

[13] Note that, although $\frac{1}{\tau}\sum_{t=1}^{\tau} \hat{f}_{2t}^2$ converges to zero in probability in the first subsample, 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.

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 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, both our LM and Wald tests would be 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 our Wald test has even better power properties than theirs in small samples ($T$ and $N \leq 100$).

Finally, 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 as 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 a single test.

Summing up, compared to the HI tests, our tests use less information but are much easier to implement. 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 are still 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 $r = 3$, $\alpha_{ik}$ 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 $\tau$ is considered to be a priori known and is set at $T/2$ for the LM/Wald tests while $\Pi$ is chosen as $[0.15, 0.85]$ for the Sup-type versions of the tests. The covariance matrix $S$ is estimated using the HAC estimator of Newey and West (1987).

Table 1 reports the empirical sizes (in percentages) for the LM/Wald tests and Sup-LM/Wald tests using 5% critical values for sample sizes ($N$ and $T$) equal to 100, 150, 200, 250, 300 and 1000.[14] 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 ($\bar{r} = 4 > r = 3$).[15]

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, the Sup-Wald test is too liberal for $T = 100$. Therefore, although we impose that $\sqrt{T}/N$ goes to zero, a large $T$ is preferable when the Sup-LM test is used. Another conclusion to be drawn is that, despite some minor differences, the tests perform quite similarly in terms of size even when the selected number of factors is not correct.

## 5.2 Power properties

We next consider similar DGPs but this time with $r = 2$ and now subject to big breaks which are characterized as deterministic shifts in the means of the factor loadings.[16] The factors are simulated as AR(1) processes with coefficients of 0.8 for the first factor and 0.2

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

[15] Notice 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 $\bar{r} = 1$.

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

for the second. The shifts in the loadings are 0.2 and 0.4 at time $\tau = T/2$. Table 2 reports the empirical rejection rates of the LM/Wald and Sup-LM/Wald tests in percentage terms using again 1000 replications.

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

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

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 two tests under the null. The DGPs are similar to those used in the size study. The second column of Table 3 (no break) reports the 5% empirical sizes. In general, we find that the pooled BE and the LM tests exhibit similar sizes.

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

## 5.4 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 $t = 1, \ldots, T/2$, and $X_{it} = A_{2i} f_t + e_{it}$ for $t = T/2 + 1, \ldots, T$, 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 cases. 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 $t = 1, \ldots, T/2$, and $X_{it} = (A_{1i} + b) f_t + e_{it}$ for $t = T/2+1, \ldots, T$, $A_{it}, u_t, e_{it} \sim i.i.d\ N(0,1)$, $\rho = 0.8$, and $b$ is the break size which ranges from 0 to 0.5. As can be observed, our Wald test has better power properties that 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 to $X_t$, is assumed to be related to $F_t$ as follows:

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

We consider a DGP where $N = 100$, $T = 200$, $\tau = 100$, $r = 2$, $a' = [1 \quad 1]$, $F_t$ are generated as two AR(1) processes with coefficients 0.8 and 0.2, respectively, $e_t$ and $v_t$ are i.i.d 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 $X_t, X_{t-1}, \ldots, X_1$ and $y_t, y_{t-1}, \ldots, y_1$ are treated as known to forecast $y_{t+1}$. This process starts from $t = 149$ to $t = 199$, and the mean square errors (MSEs) are calculated by

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

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

# 6 Empirical Applications

To provide a few empirical applications of our test, we first use the data set of Stock and Watson (2009). This data set consist 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 numbers of factors using various Bai and Ng's (2002) information criteria range from 3 to 6, we implement the test for $\bar{r} = 3, 4, 5$ and 6. For the Sup- LM and Wald tests, though other trimmings have been used, we report result 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 LM and Wald tests for different values of $\bar{r}$, with the horizontal lines representing the 5% critical values of the Sup-type test.

When compared to the critical values of the $\chi^2$ distributions, the LM and the Wald tests reject the null of no big breaks (i.e., exceeds the lower horizontal line) for $\bar{r} = 4, 5, 6$ when the break date is assumed to be known at 1984 in agreement with the results in BE (2010). Stock and Watson (2009) get similar conclusions about the existence of breaks around the early 1980s. However, one important implication of our results is that the breaks should be interpreted as being big and thus 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, while the Sup-LM test cannot reject the null for all values of $\bar{r}$, the Sup-Wald test rejects the null when $\bar{r} = 5, 6$ (i.e., exceeds the upper horizontal line). The estimate of the break date provided by the last test is around 1979 (second oil price shock), rather than 1984 which, as mentioned before, is the only date considered by BE (2010) as a potential break date in their empirical application using the same dataset. Using Bai and Ng's (2002) $IC_{p2}$ criterion (the one that chooses $\bar{r} = 6$) separately for each the two subsamples pre and post-1979 we cannot find evidence that the chosen number of factors is smaller that 6. Thus, following our results in subsection 4.4, we interpret the break as stemming from the factor loadings rather than from an abrupt change in the variance of the factors.

A second empirical application relies on another standardized dataset use by Stock and Watson (2003). It consists of 240 monthly macro series ($N = 240$) from 11 European countries from 1982M1 to 1997M8 ($T = 188$) and again we refer to the original paper for the details of the standardization of this panel. Since $N$ is quite large in this case, we report results for the trimming $\Pi = [0.15, 0.85]$ which spans the period from 1984M12 to 1995M6, during which the Maastricht Treaty was signed and the European Union was created. The results of the LM and Wald tests are shown in Figure 4 with the 5% critical values of the Sup-type test for $\bar{r} = 3$ to 6.

It is clear that, under the assumption of a known break date, comparison of the test values to the 5% critical values of a $\chi^2$ distribution implies that we can easily reject the null of no big break around 1994. However, in contrast to previous results for the U.S data set, a similar comparison of the maxima of the LM and Wald tests to the critical values of the Sup-type tests, indicates that no big breaks are present in this European data set during the sample period at hand.

# 7 Conclusions

In this paper, we propose a simple two-step procedure to test for big structural breaks in the factor loadings of large factor models 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 information criteria. 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 claim that our regression-based tests are simpler than theirs and that, though we use less information in exchange for simplicity, the one used in our approach is combined in a more efficient way so that in general the power of both testing approaches is similar. There is however an rather implausible DGP where our LM test may be inconsistent. Yet, for this very particular case, the power our Wald test and HI's tests is similar. We also show that for general cases our Wald test has better powers in very 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 does not perform very well especially when $T$ is not too large ($T < 200$). As other studies on the size of sup-type tests suggest, bootstrap can improve the finite sample performance of the test compared to the tabulated asymptotic critical values of Andrews (1993). It is high in our research agenda to explore this possibility.

Further research is also needed if we were to allow for multiple breaks. As BE (2010) point out, sequential estimation, like in Bai and Perron (1998), or an efficient search procedure, like in Bai and Perron (2003), for locating the candidate break dates may be employed, an issue which remains high in our research agenda.

Finally, though we have outlined 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 using 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.

## 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 $F_t^* = [F_t' \quad G_t^{1'}]'$, $\varepsilon_t = HG_2^t + e_t$, 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 $T^{-1}\sum_{t=1}^T F_t^* F_t^{*'} \to \Sigma_F^*$ as $T \to \infty$ for some positive matrix $\Sigma_F^*$.*

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

The following lemmae involve the new errors $\varepsilon_t$. Let $M$ and $M^*$ denote some positive constants.

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

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

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

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

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

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

As mentioned before, once it has been shown that the new factors: $F_t^*$, the new loadings: $\Gamma$ and the new errors: $\varepsilon_t$ 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 $D_{(i\cdot)}$ denote the $i$th row of $D$, and $D_{(\cdot j)}$ denote the $j$th column of $D$. Define $\hat{\mathscr{F}}_t = DF_t$, and $\hat{\mathscr{F}}_{kt} = D_{(k\cdot)} \times F_t$ as the $k$th element of $\hat{\mathscr{F}}_t$. Let $\hat{F}_{1t}$ be the first element of $\hat{F}_t$, and $\hat{F}_{-1t} = [\hat{F}_{2t} \cdots \hat{F}_{rt}]$, while $\hat{\mathscr{F}}_{1t}$ and $\hat{\mathscr{F}}_{-1t}$ can be defined in the same way. Note that $\hat{\mathscr{F}}_t$ depends on $N$ and $T$ because $D = (\hat{F}F/T)(A'A/N)$ (see Bai and Ng, 2002). 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 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 lemmae (9 to 13) required to prove Lemma 14 which is the key one in the proof of Theorem 1.

**Lemma 9.**

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

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

**Lemma 10.**

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

*Proof.* Note that:

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

Thus,

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

since $\frac{1}{T} \sum_{t=1}^{T} \left\| \hat{F}_t - DF_t \right\|^2 = O_p(\delta_{N,T}^{-2})$ 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.

□

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

**Lemma 11.**

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

*Proof.* See Lemma 10 and Assumption 6. □

**Lemma 12.**

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

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

Let $\Rightarrow$ denote *weak convergence*, then:

**Lemma 13.**

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

*for $\pi\in[0,1]$, where $\mathscr{W}_{r-1}(\cdot)$ 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. □

**Lemma 14.**

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

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

*Proof.* If we show that

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

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

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

then the result follows from Lemma 11.

First note that

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

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

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

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

and

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

It follows that:

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

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

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

Combining the above results gives:

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

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

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

Moreover, it is easy to check that $\|D\| = O_p(1)$ and $\|D - D^*\| = o_p(1)$ (See Bai 2003). Finally, $\left\| \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \hat{\mathscr{F}}_{-1s} \hat{\mathscr{F}}_{1s} \right\|$ is $o_p(1)$ by Lemma 12. Then A.6 has been shown and the proof is complete. □

**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:

$$\sqrt{T}\big(\hat{c}_1(\pi) - \hat{c}_2(\pi)\big) = \big(1/T \sum_{t=1}^{\tau} \hat{F}_{-1t}\hat{F}'_{-1t}\big)^{-1}\big(1/\sqrt{T}\sum_{t=1}^{\tau} \hat{F}_{-1t}\hat{F}_{1t}\big) - \big(1/T \sum_{t=\tau+1}^{T} \hat{F}_{-1t}\hat{F}'_{-1t}\big)^{-1}\big(1/\sqrt{T}\sum_{t=\tau+1}^{T} \hat{F}_{-1t}\hat{F}_{1t}\big)$$

$$= \left[\big(1/T \sum_{t=1}^{\tau} \hat{F}_{-1t}\hat{F}'_{-1t}\big)^{-1} + \big(I - 1/T \sum_{t=1}^{\tau} \hat{F}_{-1t}\hat{F}'_{-1t}\big)^{-1}\right]\big(1/\sqrt{T}\sum_{t=1}^{\tau} \hat{F}_{-1t}\hat{F}_{1t}\big).$$

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

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

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

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

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

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

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

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

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

### 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 $S = \lim \mathrm{Var}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \mathscr{F}_{-1t}\mathscr{F}_{it}\Big)$.

Notice that $E(\mathscr{F}_{-1t}\mathscr{F}_{1t}) = 0$, because $\frac{1}{T}\sum_{t=1}^{T} \mathscr{F}_{-1t}\mathscr{F}_{1t} = \frac{1}{T}\sum_{t=1}^{T} \hat{F}_{-1t}\hat{F}_{1t} + o_p(1)$, and $\frac{1}{T}\sum_{t=1}^{T} \mathscr{F}_{-1t}\mathscr{F}_{1t} \to E(\mathscr{F}_{-1t}\mathscr{F}_{1t})$ 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 $S$:

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

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

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

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

**Assumption 11.** $\left\|\hat{S}(\mathscr{F}) - S\right\| = o_p(1)$.

Next we consider a feasible estimator of $S$ where $\mathscr{F}_t$ 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.** $\left\|D^* - D\right\| = o_p(T^{-\frac{1}{5}})$.

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*

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

*Proof.* Given Assumption 10, it suffices to show that

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

It is easy to see that:

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

and

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

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

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

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

Finally we have

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

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

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. Eliasz (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.

Chen, B. and Y. Hong (2011). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, forthcoming.

Chen, L., J. Dolado, and J. Gonzalo (2011). Indentifying the source of structural breaks in large factor models. *In progress*.

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.

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.

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

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

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

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

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

| $N$ | $T$ | $\hat{\alpha}_{0.05}\|\overline{r}=2$ | | | | $\hat{\alpha}_{0.05}\|\overline{r}=3$ | | | | $\hat{\alpha}_{0.05}\|\overline{r}=4$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 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 $F_{kt} = \phi_k F_{k,t-1} + v_{kt}$; $\alpha_{ik}, e_{it}$ 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 $\Pi$ 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 ($\overline{r}$) in a Factor Model with $r = 2$ and Shifts in the Factor Loadings.

| $N$ | $T$ | $\hat{\alpha}_{0.05}\|\overline{r}=2$ | | | | $\hat{\alpha}_{0.05}\|\overline{r}=3$ | | | | $\hat{\alpha}_{0.05}\|\overline{r}=4$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 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 $F_{kt} = \phi_k F_{k,t-1} + v_{kt}$; $\alpha_{ik}, e_{it}$, 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 $\overline{r}$ in a Factor Model with $r = 2$ and No Shift or One Shift in the Factor Loadings.

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

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, $f_{1t} = f_t * \mathbf{1}(t < \tau)$, $f_{2t} = f_t * \mathbf{1}(t \geq \tau)$, $f_t = 0.8 f_{t-1} + u_t$, $u_t$and $e_{it} \sim$ i.i.d N(0,1). $\overline{r} = 2$.

| $N$ | $T$ | LM | Sup LM | Wald | Sup Wald | $\log\left(\frac{Wald}{Wald_{HI}}\right)$ |
|---|---|---|---|---|---|---|
| 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 |

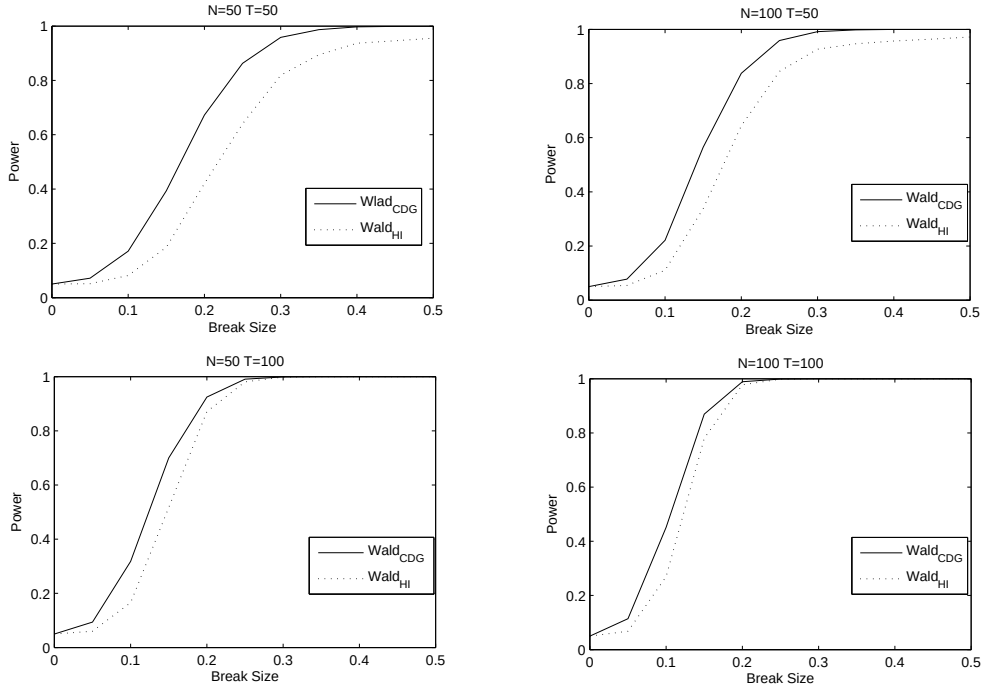


Figure 1: Comparisons of Wald tests.

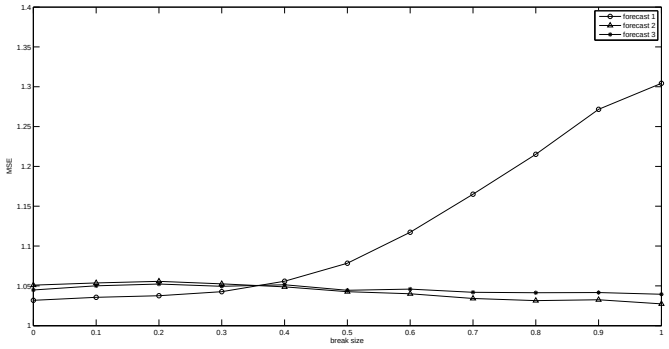


Figure 2: The MSEs of different forecasting methods (see Section 5.4).

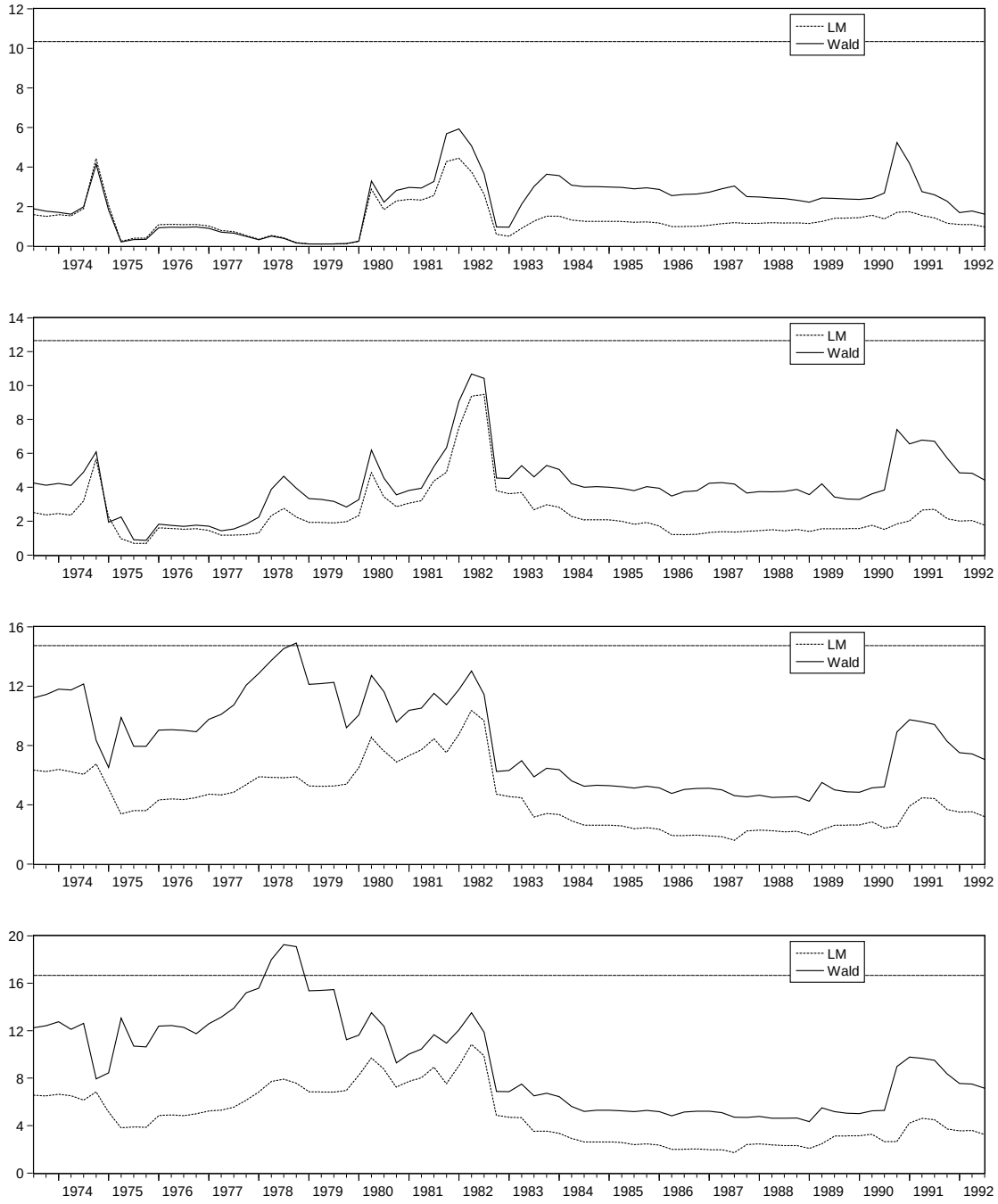


Figure 3: US data set. The LM test (dotted) and Wald test (solid) using the trimming $\Pi = [0.3, 0.7]$, for $\bar{r} = 3$ to 6 (from top to bottom), and the corresponding critical values (horizontal dotted lines) for the Sup-type Test.

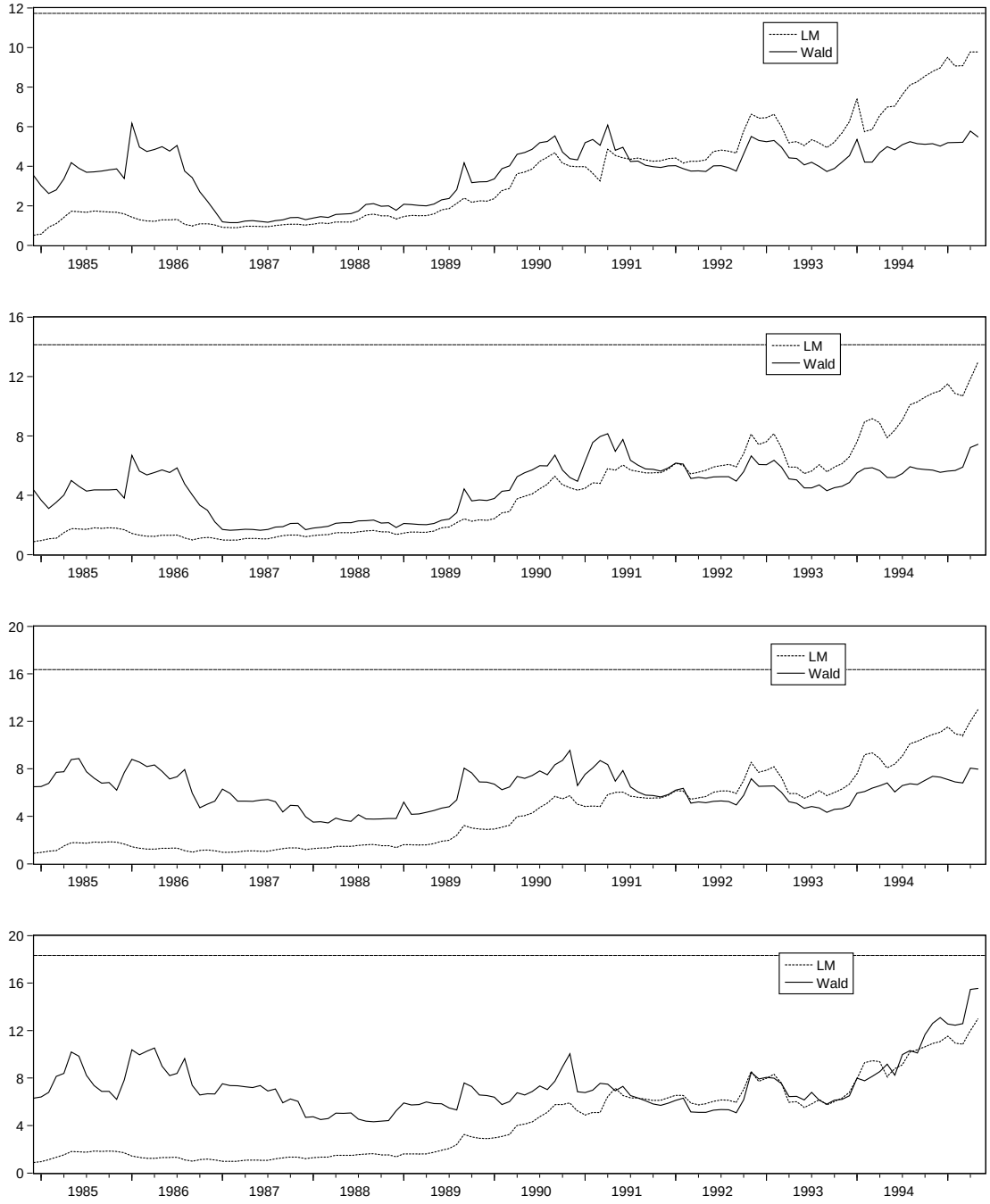


Figure 4: EU data set. The LM test (dotted) and Wald test (solid) using the trimming $\Pi = [0.15, 0.85]$, for $\overline{r} = 3$ to 6 (from top to bottom), and the corresponding critical values (horizontal dotted lines) for the Sup-type Test.