

Accepted Manuscript

Detecting big structural breaks in large factor models

Liang Chen, Juan J. Dolado, Jesús Gonzalo

PII: S0304-4076(14)00018-9

DOI: <http://dx.doi.org/10.1016/j.jeconom.2014.01.006>

Reference: ECONOM 3865

To appear in: *Journal of Econometrics*

Received date: 9 May 2012

Revised date: 3 October 2013

Accepted date: 21 January 2014



Please cite this article as: Chen, L., Dolado, J.J., Gonzalo, J., Detecting big structural breaks in large factor models. *Journal of Econometrics* (2014), <http://dx.doi.org/10.1016/j.jeconom.2014.01.006>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Detecting Big Structural Breaks in Large Factor Models*

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

First Draft: May 2010

This draft: September 2013

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 one of the factors estimated by Principal Components analysis on the remaining estimated factors, where the number of factors 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, factor loadings, principal components.

JEL CODES: C12, C33.

*We are grateful to a co-editor, two anonymous referees, Dante Amengual, Manuel Arellano, Stephane Bonhomme and Soren Johansen for insightful comments, and to Atsushi Inoue for helpful discussions on his related but independent research. We also thank seminar participants at Aarhus, CEMFI, Conference in Honour of Sir David F. Hendry (St. Andrews), ESEM 2011 (Oslo), and Workshops on High-Dimensional Econometric Modelling (Cass Business School) and Time Series Econometrics (Zaragoza) for suggestions. 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, University of Oxford. liang.chen@economics.ox.ac.uk

[‡]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 be subject to important structural breaks (e.g., those related to technological change, globalization or strong policy reforms), a standard practice in the estimation of large factor models (FM, hereafter) has been to assume 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. 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 of the FM. As a result, their advice is to use full-sample factor estimates and subsample forecasting equations to improve forecasts.

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

A forerunner of our paper is Breitung and Eickmeier (2011, BE henceforth) who were the first to propose a proper testing procedure to detect big breaks in the factor loadings. Their test relies on the idea that, under the null of no structural break (plus some additional assumptions), the estimation error of the factors can be ignored and thus the estimated factors can be treated as the true ones. Consequently, a Chow-type test can be implemented by regressing each variable in the data set on both the estimated factors using the whole sample and their truncated versions from the date of the break onwards. Focusing on the joint statistical significance of the estimated coefficients on the truncated factors, their test compares the empirical rejection frequency among the individual regressions to a nominal size of 5% under the null. In our view, this approach suffers from two limitations: (i) the overall limiting distribution of their test remains unknown when testing for the equality of all the elements of the loading matrix in both subsamples;¹ and (ii) it lacks non-trivial

¹With the notation used below in (1)-(2), the limiting distribution of the rejection frequencies for the joint hypothesis $A = B$ is not known, although the individual tests for the hypothesis $\alpha_i = \beta_i$ have known limiting distributions.

power when the number of factors is chosen according to some consistent estimator of r . This last problem can be serious. For example, as explained further below, a FM with r original factors where the loadings of one of them exhibit a big structural break at the same date admits a standard factor representation with $r + 1$ factors without a break. Hence, if the number of factors were to be chosen as $r + 1$, instead of r , their testing approach would fail to detect any break at all when in fact there is one.

Our contribution here is to propose a simple testing procedure to detect big breaks in FMs stemming from a single source which does not suffer from the previous shortcomings. We focus on breaks in the loadings though we also provide a procedure to detect whether the breaks originate from the loadings or from factors themselves. In particular, we first derive some asymptotic results finding that, in contrast to small breaks, the number of factors is overestimated under big breaks, a result which was also used by BE (2011). We argue that neglecting these breaks can have serious consequences on the forecasting performance of some popular regression-based models using factors, where their number is a priori imposed. Likewise, under big breaks, it may be difficult to provide a structural interpretation of the estimated factors when they are chosen according to some consistent information criteria (see Bai and Ng, 2006b, and Chen, 2012). Our proposal relies upon a very simple regression-based testing procedure. As sketched earlier, the insight is that a FM with big breaks in the loadings can be re-parameterized as another FM with constant loadings but a larger set of factors, where the number and the space spanned by the latter can be consistently estimated by PCA under fairly standard assumptions. Hence, rather than directly testing for whether all the elements of the loadings matrix are stable, which will suffer from an infinite-dimensionality problem as the number of variables in the panel data set grows, one can test if 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 chosen as \bar{r} according to Bai and Ng's (2002; BN henceforth) information criteria, and then \bar{r} factors are estimated by PCA. Next, one of the estimated factors (e.g., the first one) is regressed on the remaining $\bar{r} - 1$ factors, to next apply the standard Chow Test or the Sup-type Test of Andrews (1993) to this regression, depending on whether the date of the break is treated as known or unknown. If the null of no breaks is rejected in the second-step regression, we conclude that there are big breaks and, otherwise, that either no breaks exist at all or that only small breaks occur. Further, on the basis of the rank properties of the covariance matrix of the estimated factors in different subsamples, we also provide some guidance on how to distinguish between breaks stemming either from the loadings or from the data generating process (DGP hereafter) of the factors. This difference is important since the latter may lead to reject the null of constant loadings when it is true, implying a misleading interpretation of the source of the break.

After completing the first draft of this paper, we became aware of a closely related paper by Han and Inoue (2012, HI hereafter) who, in an independent research, adopt a similar approach to

ours in testing for big breaks in FM. The two approaches, however, differ in some relevant respects. In effect, rather than using a simple regression-based approach to avoid the infinite-dimensionality problem as we do here, HI (2012) 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 the HI tests use more information than ours, both tests generally exhibit similar power. Indeed, our regression-based test based on the Wald principle, which behaves much better in general than the Lagrange multiplier (LM hereafter) tests for detecting structural breaks, is even more powerful than the corresponding HI's test for small sample sizes, such as $N = T = 50$. One additional advantage of our simple linear-regression setup is that it is amenable to use many other existing methods for testing breaks, including multiple ones (see, e.g., Perron, 2006, for an extensive review of these tests).

The rest of the paper is organized as follows. In Section 2, we present the basic notation, assumptions and the definitions of *small* and *big* breaks. In Section 3, we analyze the consequences of big breaks on the choice of the number of factors and their estimation, as well as their effects on standard econometric practices with factor-augmented regressions. In Section 4, we first derive the asymptotic distributions of our tests and next discuss, when a big break is detected, how to identify its sources. Section 5 deals with the finite sample performance of our test relative to the competing tests using Monte-Carlo simulations. Section 6 provides an empirical application. Finally, Section 7 concludes. An Appendix contains detailed proofs of the main results.

2 Notation and Preliminaries

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

$$X_t = AF_t + e_t$$

where X_t is the $N \times 1$ vector of observed variables, $A = (\alpha_1, \dots, \alpha_N)'$ is the $N \times r$ matrix of factor loadings, r is the number of common factors which is finite, $F_t = (F_{1t}, \dots, F_{rt})'$ is the $r \times 1$ vector of common factors, and e_t is the $N \times 1$ vector of idiosyncratic errors. In the case of dynamic FMs, all the common factors f_t and their lags are stacked into F_t . Thus, a dynamic FM with r dynamic factors and p lags of these factors can be written as a static FM with $r \times (p + 1)$ static factors. Further, given the assumptions we make about e_t , the case analyzed by BE (2011) where the e_{it} disturbances are generated by individual specific autoregressive (AR hereafter) processes is also considered.²

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

²Notice, however, that our setup excludes the generalized dynamic FM considered by Forni and Lippi (2001), where the polynomial distributed lag possibly tends to infinity.

same date τ :

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

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

where $B = (\beta_1, \dots, \beta_N)'$ is the new factor loadings after the break. By defining the matrix $C = B - A$, which captures the size of the breaks, the FM in (1) and (2) can be rewritten as:

$$X_t = AF_t + CG_t + e_t \quad (3)$$

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

As argued by Stock and Watson (2002, 2009), the effects of some mild (local to zero) instability in the factor loadings can be averaged out, so that estimation and inference based on PCA remain valid. We generalize their analysis by allowing for two types of break sizes: *small* and *big*. In contrast to the former, we will show that the latter cannot be neglected. To distinguish between them, it is convenient to partition the C matrix as follows:

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

where Λ and H are $N \times k_1$ and $N \times k_2$ matrices that correspond to the *big* and the *small* breaks, and $k_1 + k_2 = r$. Accordingly, we can also partition the G_t matrix into G_t^1 and G_t^2 , such that (3) becomes:

$$X_t = AF_t + \Lambda G_t^1 + H G_t^2 + e_t \quad (4)$$

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

Throughout the paper, $tr(\Sigma)$ and $\|\Sigma\| = \sqrt{tr(\Sigma'\Sigma)}$ will denote the trace and the norm of a matrix Σ , respectively. For a finite dimensional vector v , we write $v = O_p(1)$ ($v = o_p(1)$) when $\|v\| = O_p(1)$ ($\|v\| = o_p(1)$). $[T\pi]$ denotes the integer part of $T \times \pi$ for $\pi \in [0, 1]$. Once the basic notation has been established, we next provide precise definitions of the two types of breaks.

Assumption 1. Breaks: (a) $\|\lambda_i\| \leq \bar{\lambda} < \infty$ for all i . $N^{-1}\Lambda'\Lambda \rightarrow \Sigma_\Lambda$ as $N \rightarrow \infty$ for some positive definite matrix Σ_Λ . (b) $\eta_i = (NT)^{-1/2}\kappa_i$ and $\|\kappa_i\| \leq \bar{\kappa} < \infty$ for all i .

The matrices Λ and H are assumed to contain non-random elements. Assumption 1.a yields the definition of a big break. It also includes the case where $\lambda_i = 0$ (no break) for a fixed proportion of variables as $N \rightarrow \infty$. As will be shown, this type of breaks will lead to inconsistency of the estimated factors and the overestimation of r . Assumption 1.b, in turn, provides the definition of small breaks which are characterized as being of order $1/\sqrt{NT}$, so the true factor space and r can be both consistently estimated under such breaks. If we only focus on the consistency of the estimated factors, the definition of small breaks can be relaxed to $\|\eta_i\| = O(N^{\delta_1}T^{\delta_2})$ with

$\delta_1, \delta_2 \leq 0$ and $\delta_1 + \delta_2 \neq 0$. As shown by Theorem 1 of Bates et al. (2013), the estimated factors using PCA are still consistent for breaks with such sizes, although at a slower convergence rate than $\min(\sqrt{N}, \sqrt{T})$. However, as argued in the Introduction, we are concerned with the number of factors as well, and such breaks will possibly lead to overestimated number of factors. Therefore, our assumption for small breaks is more stringent.

Remark 1. Bates et al. (2013) considers instabilities in the factor loadings such that: $A_t = A + h_{NT} * u_t$, where h_{NT} is a scalar which depends on N and T , and u_t is a vector of possibly random disturbances. They propose conditions about h_{NT} and u_t under which the PCA estimator of the factors are still consistent. In the case of structural breaks, they assume $h_{NT} = 1$ whereas u_t is vector of $O(1)$ elements (say $u_t = \Delta \mathbf{1}(t > \tau)$ where $\Delta = (\Delta_1, \dots, \Delta_N)'$) that do not depend on N or T after the break date τ . Unlike our definitions, they characterize the *size* of breaks as the number of nonzero elements in Δ , i.e., the number of variables having breaks in their factor loadings. For the consistency of the estimated factors, their conditions allow at most N^δ ($\delta < 1$) variables to have breaks ($\sum_{i=1}^N \|\Delta_i\| = O(N^\delta)$). By contrast, for the consistency of the estimated *number* of factors using BN's (2002) method, only a fixed number of variables are allowed to have breaks ($\sum_{i=1}^N \|\Delta_i\| = O(1)$), when N/T converges to some nonzero constant.

To compare these conditions with our definitions, notice that when N^δ variables are allowed to have breaks, we have $\|N^{-1} \sum_{i=1}^N \Delta_i \Delta_i'\| \leq N^{-1} \sum_{i=1}^N \|\Delta_i\|^2 = O(N^{\delta-1}) = o(1)$ when $\delta < 1$. On the other hand, our Assumption 1.a for big breaks implies that $\|\frac{1}{N} \sum_{i=1}^N \lambda_i \lambda_i'\|$ converges to a positive constant. In this sense, the big breaks defined in our paper have larger sizes than those considered in Bates et al. (2013), under which the estimated factors are proved to be consistent. As for the small breaks, our Assumption 1.b implies that $\sum_{i=1}^N \|\eta_i\| = O(1)$ when N and T have the same order, similar to the conditions of Bates et al. (2003) which also allows the number of factors to be consistently estimated. \square

To investigate the influence of 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 BN (2002) and Bai (2003) with a few slight modifications.

Assumption 2. Factors: $E(F_t) = 0$, $E\|F_t\|^4 < \infty$, $(NT)^{-4} E\|F_t\|^8 < \infty$, $T^{-1} \sum_{t=1}^T F_t F_t' \xrightarrow{P} \Sigma_F$ and $T^{-1} \sum_{t=1}^T F_t F_t' \xrightarrow{P} \pi^* \Sigma_F$ as $T \rightarrow \infty$ for some positive definite matrix Σ_F where $\pi^* = \lim_{T \rightarrow \infty} \frac{\tau}{T}$.

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

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

Assumption 5. Independence of Factors and Idiosyncratic Errors: $[F_t]_{t=1}^T$ and $[e_t]_{t=1}^T$ are two mutually independent groups, and $1/\sqrt{T} \sum_{t=1}^T F_t e_{it} = O_p(1)$ for each $i = 1, \dots, N$.

Assumptions 3 and 4 are standard in the literature on FM allowing for weak cross-sectional and temporal correlations between the errors. In our specific setup, Assumption 3 excludes the case where a new (old) factor appears (disappears) after the break since this event would imply that Σ_Γ becomes singular. However, this is not restrictive since we could always envisage any potential factor as having non-zero, albeit small, loadings in either of the relevant subsamples. Assumption 2, in turn, is a new one. Since factors and factor loadings cannot be separately identified, we have to assume that DGPs with breaks in the loadings, which can be reabsorbed by transformations of the factors, should not be included in the alternative. In Section 4.4, we will discuss how to differentiate between breaks in the factor loadings and breaks in the dynamics of the factors. Different factors are allowed 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 BN (2002). Notice, however, that we could have also assumed some dependence between these groups and then impose some restrictions on this dependence when necessary. Yet, this would complicate the proofs without essentially altering the insight underlying our approach. Thus, for the sake of simplicity, we assume them to be independent in the sequel.

3 The Effects of Structural Breaks

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

3.1 The estimation of factors

Let us rewrite model (4) with k_1 big breaks and k_2 small breaks in the more compact form:

$$X_t = AF_t + \Lambda G_t^1 + \epsilon_t \quad (5)$$

where $\epsilon_t = HG_t^2 + e_t$. The idea is to show that the new error terms ϵ_t still satisfy the necessary conditions for (5) being a standard FM with new factors $F_t^* = [F_t' \ G_t^{1'}]'$ and new factor loadings $[A \ \Lambda]$.

Let \bar{r} be the selected number of factors, either chosen by some prior knowledge or by using some consistent estimator such as the IC of BN (2002). Notice that \bar{r} is not necessarily equal to r . Let \hat{F} be \sqrt{T} times the \bar{r} eigenvectors corresponding to the \bar{r} largest eigenvalues of the matrix XX' , where the $T \times N$ matrix $X = [\bar{X}_1, \bar{X}_2, \dots, \bar{X}_T]'$, $\bar{X}_t = [X_{t1}, X_{t2}, \dots, X_{tN}]'$, $\hat{F} = [\hat{F}_1, \hat{F}_2, \dots, \hat{F}_T]'$. Then we have:

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

$$\hat{F}_t = DF_t^* + O_p(\delta_{N,T}^{-1}) \text{ for } t = 1, 2, \dots, T. \quad (6)$$

This result implies that \hat{F}_t estimates consistently the space of the new factors, F_t^* , but not the space of the 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 (6) becomes

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

for a $\bar{r} \times r$ matrix D of full rank. This just trivially replicates the well-known consistency result of BN (2002).³

Secondly, in the more interesting case where $k_1 > 0$ (big breaks exist), we can rewrite (6) as

$$\hat{F}_t = [D_1 \quad D_2] \begin{pmatrix} F_t \\ G_t^1 \end{pmatrix} + o_p(1) = D_1 F_t + D_2 G_t^1 + o_p(1) \quad (8)$$

where the $\bar{r} \times (r + k_1)$ matrix D is partitioned into the $\bar{r} \times r$ matrix D_1 and the $\bar{r} \times k_1$ matrix D_2 . Note that, by the definition of G_t , $G_t^1 = 0$ for $t = 1, 2, \dots, \tau$, and $G_t^1 = F_t^1$ for $t = \tau + 1, \dots, T$, where F_t^1 is the $k_1 \times 1$ sub-vector of F_t that is subject to big breaks in their loadings. Therefore (8) can be expressed as:

$$\hat{F}_t = D_1 F_t + o_p(1) \text{ for } t = 1, 2, \dots, \tau, \quad (9)$$

$$\hat{F}_t = D_2^\dagger F_t + o_p(1) \text{ for } t = \tau + 1, \dots, T \quad (10)$$

where $D_2^\dagger = D_1 + [D_2 \quad 0]$, 0 is a $\bar{r} \times (r - k_1)$ zero matrix, and $D_2 \neq 0$ when $\bar{r} > r$. Hence, since $D_1 \neq D_2^\dagger$, this implies that, in contrast to small breaks where D_2 tends to 0 due to the local-to-zero properties of the elements of H (see Assumption 1.b), under big breaks the estimated factors \hat{F} will not be consistent for the space of the true factors F . Accordingly, as will be discussed 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 BN, 2006a):

$$y_t = a' F_t + b' W_t + u_t, \quad t = 1, 2, \dots, T \quad (11)$$

where W_t is a small set of observable variables and the $r \times 1$ vector F_t contains the r common

³Notice that, for the estimator \hat{F} defined here, \bar{r} has to be smaller or equal to r for (7) to hold.

factors driving a large panel dataset X_{it} ($i = 1, 2, \dots, N$; $t = 1, 2, \dots, T$) which excludes both y_t and W_t . The parameters of interest are the elements of vector b while F_t is included in (11) to control for potential endogeneity arising from omitted variables. Since we cannot identify F_t and a , only the product $a'F_t$ is relevant. Suppose that there is a big break at date τ . From (9) and (10), we can rewrite (11) as:

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

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

$$y_t = a'_1\hat{F}_t + b'W_t + \tilde{u}_t \quad \text{for } t = 1, 2, \dots, \tau, \quad (12)$$

$$y_t = a'_2\hat{F}_t + b'W_t + \tilde{u}_t \quad \text{for } t = \tau + 1, \dots, T \quad (13)$$

where $a'_1 = a'D_1^-$ and $a'_2 = a'D_2^{\dagger-}$, and $\tilde{u}_t = u_t + o_p(1)$.

If the number of factors is a priori known to be $\bar{r} = r$, then $D_1^- = D_1^{-1}$, $D_2^{\dagger-} = D_2^{\dagger-1}$. If we assume $D_1 \neq D_2^{\dagger}$, it follows that $D_1^{-1} \neq D_2^{\dagger-1}$ and thus $a_1 \neq a_2$. Therefore, using the indicator function $\mathbf{1}(t > \tau)$, (12) and (13) can be rewritten as

$$y_t = a'_1\hat{F}_t + (a_2 - a_1)'\hat{F}_t\mathbf{1}(t > \tau) + b'W_t + \tilde{u}_t, \quad t = 1, 2, \dots, T. \quad (14)$$

A straightforward implication of the previous result is that if we were to impose the number of factors, on a priori ground, therefore ignoring the set of regressors $\hat{F}_t\mathbf{1}(t > \tau)$ in (14), in general the estimation of b will become inconsistent due to omitted variables.

Remark 2. Interestingly, there are many examples in the literature where, for theoretical or practical reasons, the number of factors is imposed as a prior. For example, to name a few, a single common factor representing a global effect is assumed in the well-known study by Bernanke, Boivin and Elias (2005) on measuring the effects of monetary policy in Factor Augmented VAR (FAVAR) models, as well as in the risk analysis in portfolios of corporate debt by Gourioux and Gagliardini (2011) where a single factor is supposed to capture a latent macro-variable. Likewise, two factors are a priori imposed by Rudebusch and Wu (2008) in their macro-finance model. Notice that a similar argument will render inconsistent the impulse response functions in FAVAR models where the regressand in (11) becomes $y_{t+1} = (F_{t+1}, W_{t+1})'$. \square

Remark 3. Alternatively, if the number of factors is not a priori imposed and it is estimated by the IC of BN (2002), we will show in the next section (Proposition 2) that the estimated number of factors will tend to $r + k_1$ as the sample size grows. In this case, D_1 and D_2^{\dagger} are $(r + k_1) \times r$ matrices, and by the definitions of D_1 and D_2^{\dagger} , it is easy to show that we can always find a $r \times (r + k_1)$ matrix

$D^- = D_1^- = D_2^{\dagger-}$ such that $D^- D_1 = D^- D_2^{\dagger} = I_r$. If we define

$$\bar{a}' = a' D^-, \quad (15)$$

then $a'_1 = a'_2 = \bar{a}'$ so that (12) and (13) can be rewritten as

$$y_t = \bar{a}' \hat{F}_t + b' W_t + \tilde{u}_t, \quad t = 1, 2, \dots, T, \quad (16)$$

so that the estimation of (11) will not be affected by the estimated factors under big breaks if $\bar{r} = r + k_1$. \square

Remark 4. Yet, even in this case, the factors themselves may be the direct subject of interest and thus their interpretation can have important implications for structural analysis. For example, a large body of empirical research in financial economics is concerned with identifying the factors that determine asset returns.⁴ However, the existence of big breaks may hamper this identification procedure. For instance, if $\hat{r} = 2$, a relevant question would be: are there two genuine factors or one factor and one break? Our testing procedure provides a useful tool to disentangle these two cases (see Section 4.4 below).

Another area where our testing approach could be useful is in applications where the estimated factors are modeled in a VAR in order to identify the structural shocks driving them (see e.g. Charnavoki and Dolado, 2012). This identification becomes more difficult as the number of factors increases. For this reason, it is very important to determine whether the selection of a large number of factors is due to having several genuine factors or to a break affecting some of them. The insight is that instead of having to identify $r + k_1$ shocks we would only have to identify r . \square

Summing up, the use of estimated factors as the true factors when assuming that the number of factors is a priori known will lead to inconsistent estimates in a FAR under big breaks. As a simple remedy, $\hat{F}_t \mathbf{1}(t > \tau)$ should be added as regressors when big breaks are detected and the break date is located. Alternatively, without pretending to know a priori the true number of factors, the estimation of FAR will be robust to the estimation of factors under big breaks if the number of factors is chosen according to some consistent estimator. Yet, this may hinder the correct interpretation of the estimated factors in terms of observables. As a result, in order to run regression (16), our advice is to avoid imposing the number of factors a priori, unless a formal test of big breaks is implemented. We will illustrate these points in Section 5 by means of simulations in a typical forecasting exercise where the predictors are common factors estimated by PCA.

⁴Chen et al (1986) and Shanken and Weinstein (2006) are good illustrations of attempts to interpret the underlying forces in the stock market developments in terms of some observed macro variables.

3.2 The estimated number of factors

BE (2011) 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 provide a formal proof of this result, we proceed to fill this gap.

Let \hat{r} be the estimated number of factors in (5) using the IC proposed by BN (2002), and ϑ_{NT}^1 be the largest eigenvalue of $(NT)^{-1} \sum_{t=1}^T e_t e_t'$. Then, the following result holds:

Proposition 2. *Suppose $\vartheta_{NT}^1 = O_p(\delta_{NT}^{-2})$ and Assumptions 1 to 5 hold, then:*

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

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

In sum, when we use PCA to estimate the factor space and the IC of BN (2002) to estimate the number of factors, the small breaks can be safely ignored, while the big breaks will lead to the inconsistencies of \hat{F}_t and \hat{r} .

4 Testing for Structural Breaks

4.1 Hypotheses of interest and test statistics

Our goal here is to develop a test for big breaks. As mentioned above, if we were to follow the usual approach in the literature to test for structural breaks, we would consider the following null and alternative hypotheses in (1) and (2): $H_0 : A = B$ vs. $H_1 : A \neq B$. However, this standard formulation faces two problems. First, if only small breaks occur, the estimation and inference based on PAC are not affected. Thus, we can ignore these breaks. 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. To solve the first problem, we focus only on big breaks and consider $H_0 : k_1 = 0$ vs. $H_1 : k_1 > 0$, 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.

Relying upon the discussion in Section 3.1 about the inconsistency of \hat{F} for the space of the true factors F when big breaks occur, our strategy to circumvent the second problem is to focus on how the dependence properties of the \bar{r} estimated factors (using the whole sample) change before and after the potential break date. Since, in line with the standard assumption in FM, the 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.

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 and \bar{r} common factors (\hat{F}_t) are estimated by PCA.
2. In the second step, we consider the following linear regression of one of the estimated factors on the remaining $\bar{r} - 1$ ones. For example, using the first factor as the regressand, this leads to the regression:

$$\hat{F}_{1t} = c_2 \hat{F}_{2t} + \cdots + c_{\bar{r}} \hat{F}_{\bar{r}t} + u_t = c' \hat{F}_{-1t} + u_t \quad (17)$$

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

Remark 5. In the first stage, we have recommended to choose \bar{r} as some consistent estimator of r to obtain the best size and power properties. Although our Proposition 2 is based on the \hat{r} estimated by the IC of BN (2002), one can also use other procedures to consistently estimate r . For example, Onatski (2009, 2010), Ahn and Horenstein (2013) show that their methods have better finite sample properties than BN (2002) especially when the errors are cross sectionally correlated. That being said, as will be discussed below, one important feature of our tests is that they do not rely on the correct estimation of r . In the second step, although there are many methods of testing for breaks in a simple linear regression model, we follow Andrews (1993) to define the LM and Wald tests when the possible break date is assumed to be known, and their *Sup-type* versions when there is no prior knowledge about it. Moreover, since the LM, Wald and LR test statistics have the same asymptotic distribution under the null, we focus on the first two because they are computationally simpler. \square

Define $D^* = V^{-1/2} \Upsilon' \Sigma_A^{1/2}$ as the limit of the matrix D in equation (7), where $V = \text{diag}(v_1, v_2, \dots, v_r)$, $v_1 > v_2 > \dots > v_r$ are the eigenvalues of $\Sigma_A^{1/2} \Sigma_F \Sigma_A^{1/2}$, and Υ is the corresponding eigenvector matrix (see Bai, 2003). Define $\mathcal{F}_{1t} = D_1^* F_t$ and $\mathcal{F}_{-1t} = D_{-1}^* F_t$, where D_1^* is the first row of D^* and D_{-1}^* is the matrix containing the second to last rows of D^* . Finally, let the $(r - 1) \times (r - 1)$ matrix $S = \lim \text{Var}(\frac{1}{T} \sum_{t=1}^T \mathcal{F}_{-1t} \mathcal{F}_{1t})$.

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

$$\mathcal{L}(\bar{\pi}) = \frac{1}{\bar{\pi}(1 - \bar{\pi})} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{\tau} \hat{F}_{-1t} \hat{u}_t \right)' \hat{S}^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{\tau} \hat{F}_{-1t} \hat{u}_t \right) \quad (18)$$

where $\bar{\pi} = \tau/T$, τ is a pre-assumed date for the potential break, \hat{u}_t is the residual in the OLS

regression (17), which by construction equals \hat{F}_{1t} , and \hat{S} is a consistent estimator of S .⁵

The corresponding Sup-LM statistic is defined as:

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

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

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

$$\mathcal{L}^*(\bar{\pi}) = \bar{\pi}(1-\bar{\pi}) \cdot T \left(\hat{c}_1(\bar{\pi}) - \hat{c}_2(\bar{\pi}) \right)' \hat{S}^{-1} \left(\hat{c}_1(\bar{\pi}) - \hat{c}_2(\bar{\pi}) \right) \quad (20)$$

and

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

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

To illustrate why our two-step testing 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 (5) becomes:

$$X_t = A f_t + B g_t + e_t$$

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

$$\begin{pmatrix} \hat{f}_{1t} \\ \hat{f}_{2t} \end{pmatrix} = D \begin{pmatrix} f_t \\ g_t \end{pmatrix} + o_p(1)$$

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:

$$\hat{f}_{1t} = d_1 f_t + o_p(1) \quad \hat{f}_{2t} = d_3 f_t + o_p(1) \quad \text{for } t = 1, \dots, \tau,$$

$$\hat{f}_{1t} = (d_1 + d_2) f_t + o_p(1) \quad \hat{f}_{2t} = (d_3 + d_4) f_t + o_p(1) \quad \text{for } t = \tau + 1, \dots, T,$$

which imply that:

$$\hat{f}_{1t} = \frac{d_1}{d_3} \hat{f}_{2t} + o_p(1) \quad \text{for } t = 1, \dots, \tau,$$

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

⁶We can also construct the Wald test as $T \left(\hat{c}_1(\bar{\pi}) - \hat{c}_2(\bar{\pi}) \right)' \left(\frac{\hat{S}_1}{\bar{\pi}} + \frac{\hat{S}_2}{(1-\bar{\pi})} \right)^{-1} \left(\hat{c}_1(\bar{\pi}) - \hat{c}_2(\bar{\pi}) \right)$ and the Sup-Wald test similarly, where \hat{S}_1 and \hat{S}_2 are estimates of S using subsamples. Yet, in all our simulations, the results based on these two methods are very similar. Therefore, for brevity, we focus on the ones obtained using the full sample estimation of S , as in (20) and (21).

$$\hat{f}_{1t} = \frac{d_1 + d_2}{d_3 + d_4} \hat{f}_{2t} + o_p(1) \quad \text{for } t = \tau + 1, \dots, T.$$

Thus, we can observe that the two estimated factors are linearly related and that the coefficients $\frac{d_1}{d_3}$ and $\frac{d_1+d_2}{d_3+d_4}$ before and after the break date must be different due to the non-singularity of the D matrix. As a result, regressing one of the estimated factors on the other and testing for a structural break in this regression, we should reject the null of no big break. In the case where $d_3 = 0$, the above argument fails. But since d_1 and d_4 are non-zeros (otherwise D will have reduced-rank), the estimated slope in the first subsample will diverge while it will converge to some bounded number in the second subsample. Therefore our test also has power in this case.⁷

Likewise, if the break date τ is not a priori assumed to be known, the Sup-type tests will yield a natural estimate of τ at the date when the test reaches its maximum value. In what follows, we derive the asymptotic distribution of the test statistics (18) to (21) under the null hypothesis, as well as extend the intuition behind this simple example to the more general case to show that our tests have 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 BN's (2002) IC, and it converges to the true one under the null hypothesis of no big break, we start with the most interesting case where $\bar{r} = r$. To derive the asymptotic distributions of the LM and Wald statistics, we adopt the following additional assumptions:

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

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

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

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

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

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

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

Assumption 6 and 8 are required to bound the estimation errors of \hat{F}_t , while Assumption 7 is needed to derive the weak convergence of the test statistics using the Functional Central Limit

⁷This is the case for the Wald test but may not be true for the LM test, because our Wald test is directly built upon the difference between the estimated coefficients.

Theorem (FCLT). Assumption 10 corresponds to Assumption G of Bai (2003), which is required 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 us to consider a quite general class of linear processes for the factors: $F_{it} = \sum_{k=0}^{\infty} \theta_{ik} v_{i,t-k}$, where $v_t = [v_{1t} \dots v_{rt}]'$ are i.i.d with zero means, and $Var(v_{it}) = \sigma_i^2 < \infty$. In this case, it can be shown that:

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

for which it suffices that

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

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 such as Newey and West's (1987) estimator with a Barlett kernel, which is the one used in our simulations below.⁸

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

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

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

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

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

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

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

Remark 6. It is easy to show that Theorem 1 still holds when $\bar{r} < r$. However, when $\bar{r} > r$, the properties of $\hat{F}_{\bar{r}t}$ are unknown since the \bar{r} th eigenvectors of XX' may be related to the properties of e_t . Thus, the asymptotic distribution cannot be derived in a similar way. Yet, as the simulations in Section 5 show, in such an instance Theorem 1 still provides a reasonably good approximation for the distributions of our test statistics in finite samples. Moreover, the case where $\bar{r} > r$ can

⁸Though 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.

be avoided if, instead of relying on a priori choice of \bar{r} , practitioners use BN's (2002) IC or other consistent estimators of r , in line with Proposition 2. \square

4.3 Performance of the 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 changes at time τ , so that the proposed tests are able to detect big breaks.

Assuming that $r < \bar{r} \leq r + k_1$, then the matrix D_1 and D_2^\dagger in (9) and (10) become $\bar{r} \times r$ matrices. Notice that since $\bar{r} > r$ we can always find $\bar{r} \times 1$ vectors ρ_1 and ρ_2 which belong to the null spaces of D_1' and $D_2^{\dagger'}$ separately, that is, $\rho_1' D_1 = 0$ and $\rho_2' D_2^\dagger = 0$. Hence, premultiplying both sides of (9) and (10) by ρ_1' and ρ_2' leads to:

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

which, after normalizing one of the elements of ρ_1 and ρ_2 (e.g., the first one) to be 1, implies that:

$$\hat{F}_{1t} = \hat{F}'_{-1t} \rho_1^* + o_p(1) \quad t = 1, 2, \dots, \tau, \quad (22)$$

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

where $\hat{F}'_{-1t} = [\hat{F}_{2t}, \dots, \hat{F}_{\bar{r}t}]$ and ρ_1^*, ρ_2^* are both $(\bar{r} - 1) \times 1$ vectors. Next, to show that $\rho_1^* \neq \rho_2^*$, we proceed as follows. Suppose that $\gamma \in \text{Null}(D_1')$ and $\gamma \in \text{Null}(D_2^{\dagger'})$, then by the definitions of D_1 and D_2^\dagger and by the basic properties of full-rank matrices, it holds that $\gamma \in \text{Null}(D')$. Since D is a full rank $\bar{r} \times (r + k_1)$ matrix and $\bar{r} \leq r + k_1$, then $\text{Null}(D') = 0$ and thus $\gamma = 0$. Therefore, the only vector that belongs to the null space of D_1 and D_2^\dagger is the trivial zero vector. Further, because the rank of the null space of D_1 and D_2^\dagger is larger than 1, we can always find two non-zero vectors such that $\rho_1 \neq b\rho_2$ for any constant $b \neq 0$.

Notice that when $\bar{r} \leq r$, the rank of the null spaces of D_1 and D_2^\dagger will possibly become 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, in general we cannot verify that $\text{plim} \hat{\rho}_1 \neq \text{plim} \hat{\rho}_2$.

Remark 7. One underlying assumption in the above argument is that one of the elements of ρ_1 and ρ_2 (e.g., the first ones) are different from zero. This assumption is hard to verify since the D matrix depends on Γ and F^* in a highly nonlinear way,⁹ and it is not difficult to find DGPs where

⁹By the result of Bai (2003), when $\bar{r} = r + k_1$, the $(r + k_1) \times (r + k_1)$ matrix $D = V_{NT}^{-1}(\hat{F}^{*'} F^*/T)(\Gamma' \Gamma/N)$, where V_{NT} is the diagonal matrix with the first $r + k_1$ eigenvalues of $X'X/NT$.

this assumption does not hold. This normalization issue makes it really hard to come up with a formal result on the consistency of our test. Instead, we have run a large number of simulations to study the actual power properties of our tests for various DGPs, including the ones where the first elements of ρ_1 and ρ_2 are both zeros. The general finding is that our Wald and Sup-Wald tests are very powerful for all the DGPs we have considered, while the LM and Sup-LM tests may lose power when the normalization issue arises.¹⁰ For this reason, we recommend to use the Wald test in practice. We present some of the representative simulation results in Section 5 but more results are available upon request. \square

Remark 8. Since our tests are based on a linear regression model, many other available methods in the literature can also be applied in our second-stage procedure. For instance, when the break date is not known a priori, one can use the CUSUM type-test first proposed by Brown, Durbin and Evans (1975), and also Chen and Hong's (2012) test via nonparametric regression. Thus, this flexibility allows practitioners to draw conclusions about breaks based on broader evidence than that just provided by a single test. \square

Remark 9. Although our tests have been designed for a single common break at date τ , they should also exhibit powers against other interesting alternatives such as multiple breaks and a change in the number of factors.¹¹ \square

4.4 Disentangling breaks in loadings from breaks in factors

A potential critique for all available tests of big breaks in FM is that they cannot differentiate between breaks in factor loadings and breaks in the covariance matrix of factors. For illustrative purposes, let us consider a FM with $r = 2$, where the factor loadings are constant but the covariance matrix of the factors breaks, such that: $E(F_t F_t') = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ for $t = 1, \dots, T/2$, and $E(F_t F_t') = \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$ for $t = T/2 + 1, \dots, T$, with $\rho \notin \{0, 1\}$. If we further assume $\Sigma_A = \lim_{N \rightarrow \infty} A' A / N$ is a diagonal matrix, then, in view of Bai (2003), we have that $\hat{F}_t = F_t + o_p(1)$, where \hat{F}_t is a 2×1 vector. In this case, regressing \hat{F}_{1t} on \hat{F}_{2t} will yield estimated coefficients close to ρ and $-\rho$ before and after the break, respectively. As a result, our tests will reject the null of no big break in the loadings while the true DGP has a break in the factors.

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

¹⁰Since the LM test is the one more affected by the normalization problem, an interesting issue raised by a referee is whether this problem could be avoided by using $\max_{j=1, \dots, \bar{r}} LM(\hat{F}_{jt})$ where $LM(\hat{F}_{jt})$ is the LM test using \hat{F}_{jt} as the regressand. However, because its limiting distribution is unknown under the null and likely dependent on nuisance parameters, we leave this analysis for our further research agenda.

¹¹See our online appendix for details.

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 which is the source of breaks.

Assuming that there is only one source of instability and that one break in the FM has been detected at some date τ , one can differentiate between the break in the loadings and the break in the factor dynamics by comparing the number of factors obtained with the whole sample to those in each of the subsamples split by τ . To see this, notice that, absent a big break in the factor loadings, the number of factors will be consistently estimated for the whole sample and each sub samples, as long as the factors satisfy:

$$\frac{1}{\tau} \sum_{t=1}^{\tau} F_t F_t' \xrightarrow{p} \Sigma_F^1 > 0, \text{ and } \frac{1}{T-\tau} \sum_{t=\tau+1}^T F_t F_t' \xrightarrow{p} \Sigma_F^2 > 0.$$

One important observation is that, when the factors are stationary and the rejection of the null is due to big breaks in the loadings, the true number of factors, r , can be consistently estimated for each of the two subsamples while, for the whole sample, it will be overestimated in light of Proposition 2. Therefore, for a given data set, if $\hat{r} = 2$ in the whole sample and our test rejects the null with $\bar{r} = 2$, it could be that there is one factor and one big break, or two factors with changing correlation, as shown above. In the first case, the estimated numbers of factors in each of the two subsamples converge to 1, while for the latter case it converges to 2. Thus, these different rank properties could become the basis of our identification strategy for the source of the break.

When both breaks in the loadings and factors exist, our strategy of considering a finite dimensional linear regression still works, but the asymptotic distribution of the proposed test statistics may be different. The reason is that the estimated factors, which are consistent estimators of the true factor space under the null, may also experience breaks in their dynamics. As a result, Hansen's (2000) results imply that the asymptotic distributions of the Wald and Sup Wald tests may change if the dynamics of regressors (the estimated factors in our case) have breaks. We leave this interesting question for future research.

5 Simulations

In this section, we first use a simple factor-based forecasting model to illustrate the consequences of ignoring big breaks, as discussed in Section 3. Next, we study the finite sample properties of our proposed LM/Wald and Sup-LM/Wald tests. We pay special attention to the sizes and the powers when $\bar{r} > r$ since, as discussed previously, obtaining formal results for this case is very difficult. A comparison with the tests of BE (2011) and HI (2012) is also provided to illustrate the advantages of our tests in term of power. Throughout this section, the potential breaking date is considered to located at half of the sample ($\tau = T/2$) and is taken to be a priori known for the

LM/Wald tests while Π is chosen as $[0.15, 0.85]$ for the Sup-type versions. Finally, the covariance matrix S is estimated using the HAC estimator of Newey and West (1987).

5.1 The effect of big breaks on forecasting

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 big break in the factor loadings:

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

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

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

We consider a DGP where $N = 100$, $T = 200$, $\tau = 100$, $r = 2$, $a' = [1 \quad 1]$, F_t are generated as two AR(1) processes with coefficients 0.8 and 0.2, respectively, e_t and v_t are i.i.d standard normal variables, and the break size is characterized by a range of mean shifts between 0 and 1.

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 $\hat{y}_{t+1|t} = \hat{a}'F_t$, 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 $\hat{y}_{t+1|t} = \hat{a}'\hat{F}_t$, 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. $\hat{y}_{t+1|t} = \hat{a}'[\hat{F}_t' \quad \hat{F}_t'\mathbf{1}(t > \tau)']'$, 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 $\hat{y}_{t+1|t} = \hat{a}'\hat{F}_t$, where \hat{a} is the OLS estimate of a in the regression of y_{t+1} on \hat{F}_t .

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

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

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

The results obtained from 1000 replications are reported in Figure 1, plotting MSEs against the different break sizes in the above-mentioned range. It is clear that the MSEs 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. Notice, however, that they cannot outperform the benchmark forecasting due to the estimation errors of the factors for the chosen sizes of N and T . In line with our previous analysis, the lesson to be drawn from this exercise is that, in case of a big break, imposing the number of factors a priori can significantly worsen forecasts.

5.2 Size properties

We first simulate data from the following DGP:

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

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

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

where $[\phi_1, \phi_2, \phi_3] = [0.8, 0.5, 0.2]$, 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.

Table 1 reports the empirical sizes (in percentages) for the LM/Wald tests and Sup-LM/Wald tests using 5% asymptotic critical values for sample sizes (N and T) equal to 100, 150, 200, 250, 300 and 1000.¹² 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 ($\bar{r} = 4 > r = 3$).¹³ Although, for $N, T \geq 100$, BN's (2002) IC very often select 3 factors, there are some cases where 2 and 4 are also selected.

Broadly speaking, the LM and Wald tests are slightly undersized for $r = 2$ and 3, and especially so when $r = 4$. Yet, the empirical sizes converge to the nominal size as N and T increase. This finite sample problem is more accurate with the Sup-LM test especially for small T , in line with the findings in other studies (see, Diebold and Chen, 1996). This is hardly surprising because, for instance, when $T = 100$ and $\Pi = [0.15, 0.85]$, we only have 15 observations in the first subsample. By contrast, although the Sup-Wald test is too liberal for $T = 100$, in general it behaves better than the Sup-LM test (see Kim and Perron, 2009). Theoretically, the asymptotic distribution in Theorem 1 applies only for $\bar{r} = 2, 3$. Yet, the results in Table 1 show that this distribution also

¹²As mentioned earlier, the critical values of the Sup-type tests are taken from Andrews (1993).

¹³Notice 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 tests. Had we chosen $r = 2$ as the true number of factors, then the test could not be implemented for $\bar{r} = 1$.

provides reasonably good approximations for the case $\bar{r} = 4$, the only exception being the Sup-Wald test, which is oversized for $T = 100$.

To further study how the size of our Wald tests is affected by the selection of \bar{r} in a more general setup, we repeat the above simulations by generating the idiosyncratic errors in the following way as in BN (2002):

$$e_{it} = \rho e_{i,t-1} + u_{it} + \beta \sum_{h=i-J}^{i-1} u_{ht} + \beta \sum_{h=i+1}^{i+J} u_{ht}$$

where $u_{it} \sim i.i.d N(0, 1)$, β and J control the cross sectional correlation of e_{it} , and ρ controls the serial correlation of e_{it} . The other parts of the DGP remain the same as above, and we consider there different types of e_t : (1) only serially correlated ($\rho = 0.2, 0.5, 0.7$), (2) only cross sectionally correlated ($\beta = 0.2, 0.5, J = 5, 10$), (3) both cross sectionally and serially correlated. To save space, we fix $N = 100$ (similar to the data set in our application) and focus on the role of T .¹⁴

The results are reported in Table 2 for the Wald and Sup-Wald tests, where two conclusions can be drawn. First, the serial correlations of e_t generally have a larger impact on the sizes than the cross-sectional correlations. Second, the size distortions of the Wald tests disappear as the sample size grows. Overall, the Wald tests have correct sizes for $\bar{r} = 4$ when $T = 200$.¹⁵

5.3 Power properties

We next consider similar DGPs as in Table 1 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 for the second. The shifts in the loadings are 0.2 and 0.4 at time $\tau = T/2$. The other parts of the DGP are the same as in Table 1. Table 3 reports the empirical powers of the LM/Wald and Sup-LM/Wald tests in percentage terms with 1000 replications. As expected, both tests are powerful to detect the breaks as long as $\bar{r} > r = 2$, while the power is trivial when $\bar{r} = r = 2$.

Next, we study the powers of our tests when the argument in Section 4.3 fails, i.e., the first element of ρ_1 and ρ_2 are both zero. The DGPs are constructed as follows:

$$\begin{aligned} X_{it} &= A_1 F_{1t} + A_2 F_{2t} + e_{1t} \text{ for } t = 1, \dots, T/2 \\ X_{it} &= B_1 F_{1t} + A_2 F_{2t} + e_{1t} \text{ for } t = T/2 + 1, \dots, T, \end{aligned}$$

where F_{1t} and F_{2t} are two AR(1) process defined as above, $A_1 \sim N(0, 1)$, $A_2 \sim 0.9 \cdot N(0, 1)$, $B_1 \sim 0.8 \cdot N(0, 1)$, and $e_{it} \sim N(0, 1)$. Define $G_t^1 = F_{1t} \mathbf{1}(t \geq \tau)$, $\Gamma = [A_1 \ A_2 \ (B_1 - A_1)]$, and

¹⁴More simulation results with different N and T are available upon request.

¹⁵The only exception is when the errors are all strongly correlated ($\rho = 0.7$ implies that the errors are even more persistent than two of the factors). In practice, this can be tested since e_t can be consistently estimated.

¹⁶The results with other types of breaks, such as random shifts, are similar and available upon request.

$F_t^* = (F_{1t} \ F_{2t} \ G_t^1)'$. Then, if $\bar{r} = 3$, using results of Bai (2003) it is easy to show that¹⁷:

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

where

$$D^* = \begin{pmatrix} 0 & 0.6 & 0 \\ -1.38 & 0 & 1.38 \\ 0 & 0 & -1.38 \end{pmatrix}.$$

It then follows from the definition of F_t^* that ρ_1 and ρ_2 should be vectors taking the form $(0, 0, a)$ and $(0, b, 0)$, respectively, with $a, b \neq 0$. Our tests are applied to such DGPs for $\bar{r} = 2, 3, 4$, first using \hat{F}_{1t} as the regressand and then \hat{F}_{2t} . The results with 1000 replications are reported in Tables 4 and 5.

First, it is clear that, even for this special DGP, our Wald and Sup-Wald tests exhibit good power in finite samples when \hat{F}_{1t} is used as the regressand. Second, although our LM and Sup-LM tests lose power when \hat{F}_{1t} is the regressand, this is not the case when \hat{F}_{2t} is used as the regressand. Finally, the Wald and Sup-Wald tests strongly dominates the LM and Sup-LM tests in term of power. Many other simulations have been run in which the power of our Wald and Sup-Wald tests are found to be strong and robust. As will be discussed below, the differences between the power of our LM and Wald tests for this DGP can be explained by the fact that the former uses much less information than the latter. Therefore, we recommend the use of Wald and Sup-Wald on the basis of their good size and power properties.

5.4 Comparison with the BE test

As discussed earlier, the BE test relies on the consistent estimation of the original factors. These authors construct N test statistics s_i for each of the hypothesis $\alpha_i = \beta_i$, but not for the joint hypothesis $A = B$. Their method have two limitations. First, as we have shown, the big breaks lead to a new FM representation, as in (5), in which the new factor loadings are constant. Thus, the BE test will lose power when the number of factors is chosen to be $r + k_1$, which is quite possible in light of our Proposition 2. On the contrary, our tests will not suffer from this problem when $\bar{r} > r$. To compare the performance of our test againts the BE test for the joint hypothesis ($A = B$), we need to construct the following pooled statistic as suggested by Remark A of BE (2011):

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

where s_i is the individual LM statistics in BE (2011). This test should converge to a standard normal distribution as long as e_{it} and e_{jt} are independent, a restrictive assumption that we also

¹⁷The matrix D^* is defined as in Section 4.1 with Σ_A and Σ_F replaced by Σ_Γ and Σ_{F^*} .

adopt here for comparative purpose. For simplicity, we only report results for the case of known break dates.

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

Then, we generate a break in the loadings of the first factor while the other elements of the DGPs remain the same as in Table 3 where we study the power properties. The break is generated as a shift of size 0.1 in the mean of the loadings. As before, we consider two cases: (i) the number of factors is correctly selected: $\bar{r} = r = 2$; and (ii) the selected number of factors is larger than the true one: $\bar{r} = 3 > r = 2$. The third and fourth columns in Table 5 report the empirical rejection rates of both tests. In agreement with our previous discussion, the differences in power are quite striking: when $\bar{r} = 2$, the pooled BE test is much more powerful while the opposite occurs when $\bar{r} = 3$. Notice that, as discussed above, for $\bar{r} = 2$, our test will not be able to detect the break whereas, for $\bar{r} = 3$, the pooled BE test will be powerless. However, according to our Proposition 2, the use of BN's (2002) IC will yield the choice of $\bar{r} = 3$ as a much more likely outcome as N and T increase. For example, for this simulation, on average the PC_{p1} of BN (2002) chooses $\hat{r} = 3$ in 94.6% of the cases.

The second problem of the BE approach is that, even when r is assumed to be known, their method may falsely reject the null hypothesis $\alpha_i = \beta_i$ for some i . To see this, we can use the same argument as in (11) to (13) to show that, in the presence of big breaks, for some variables with constant loadings¹⁸, replacing the true factors by the estimated factors will result in a "break" in the factor loadings.

5.5 Comparison with the HI test

The HI (2012) 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 (9) and (10), $\tau^{-1} \sum_{t=1}^{\tau} \hat{F}_t \hat{F}_t' = D_1 \hat{\Sigma}_F^1 D_1' + o_p(1)$, and $(T - \tau)^{-1} \sum_{t=\tau+1}^T \hat{F}_t \hat{F}_t' = D_2^\dagger \hat{\Sigma}_F^2 D_2^{\dagger'} + o_p(1)$, where $\hat{\Sigma}_F^1 = \tau^{-1} \sum_{t=1}^{\tau} F_t F_t'$ and $\hat{\Sigma}_F^2 = (T - \tau)^{-1} \sum_{t=\tau+1}^T F_t F_t'$. Therefore, the HI test is able to detect breaks if $\hat{\Sigma}_F^1, \hat{\Sigma}_F^2 \rightarrow \Sigma_F$ and D_1 and D_2^\dagger converge to different limits as N and T go to infinity. Specifically, their test is defined as:

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

where

$$\mathcal{C}(\pi) = \text{vech} \left[\frac{1}{\tau} \sum_{t=1}^{\tau} \hat{F}_t \hat{F}_t' - \frac{1}{T - \tau} \sum_{t=\tau+1}^T \hat{F}_t \hat{F}_t' \right],$$

¹⁸This is possible since our definition for big breaks allows a fixed proportion of variables to have no breaks.

and \hat{V} is a HAC estimator of the covariance matrix of $\mathcal{C}(\pi)$ 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-dimensional 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 ($\frac{1}{\tau} \sum_{t=1}^{\tau} \hat{F}_{1t}^2$).

In principle, it may seem that the use of less information is the price one has to pay to render our testing procedure much simpler than theirs. After all, both steps in our approach can be easily implemented in any conventional statistical software used by practitioners, while HI's test is computationally more burdensome. Yet, our Wald test exhibits very similar power to theirs in all the simulations we have run. HI (2012) reports some simulation results for the power comparisons with our Wald tests under very general DGPs. Therefore, to avoid repetitions, we focus only on small samples ($N, T \leq 100$), and compare the (size-adjusted) power curves of HI's and our Wald tests (using the Bartlett kernel) for the following DGP: $X_{it} = A_{1i}F_t + e_{it}$ for $t = 1, \dots, T/2$, and $X_{it} = (A_{1i} + b)F_t + e_{it}$ for $t = T/2 + 1, \dots, T$, where $F_t = \rho F_{t-1} + u_t$, $A_{1i}, u_t, e_{it} \sim i.i.d N(0, 1)$, $\rho = 0.8$, and b is the break size which ranges from 0 to 0.5. As can be observed in Table 7, our Wald test has better power properties than HI's test in all these cases. However, not surprisingly, as N and T get large, both tests perform very similarly in term of power.¹⁹

6 An Empirical Application

To provide an empirical application of our tests, we use Stock and Watson's (2009) data set consisting of 144 quarterly time series of nominal and real variables for the US ranging from 1959:Q1 to 2006:Q4. Since not all the variables are available for the whole period, we end up using their suggested balanced panel of standardized variables with $T = 190$, $N = 109$. This more or less corresponds to the case where $T = 200$, $N = 100$ in Tables 1 and 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.

Using various BN's (2002) IC (IC_{p1} , IC_{p2} , PC_{p1} and PC_{p2}) the estimated number of factors ranges from 2 to 6, therefore we implement our test for $\bar{r} = 2$ to 6. The Sup-Wald test is applied since no priori break date is assumed to be known. In order to have enough observations in both subamples, we use the trimming $\Pi = [0.3, 0.7]$. It corresponds to the time period ranging from 1973:Q3 to 1992:Q3 which includes several relevant events like, e.g., the oil price shocks (1973,

¹⁹This can be seen from Tables 5A and 5B of HI (2012). We also have similar unreported simulation results that are available upon request.

1979) and the beginning of great moderation period in the early 1980s. The graphs displayed in Figure 2 are the series of Wald tests for different values of \bar{r} , with the horizontal lines representing the 5% asymptotic critical values of the Sup-Wald tests.

We find that the Sup-Wald test rejects the null when $\bar{r} = 5, 6$. The estimated break date is around 1979-1980 (second oil price shock), rather than 1984, which is the only candidate considered by Stock and Watson (2009) as a potential break date in their empirical application with the same data set. One possible explanation for this break date could be the Iranian revolution at the beginning of 1979 and its subsequent impact on monetary policy in the US (see Fernandez-Villaverde et al., 2010).

7 Conclusions

In this paper, we propose a simple two-step procedure to test for big structural breaks in the factor loadings of large FM that circumvents some limitations affecting other available tests. In particular, after choosing the number of factors in the whole sample according to BN's (2002) IC and estimating them by PCA, our test relies on a regression of one of the estimated factors on the remaining ones, allowing for a break in the parameters at known or unknown date. LM and Wald tests for the null of parameter stability are applied. We show that our test may have better power than the test of BE (2011) under the alternative of big breaks and, and that it is simpler than the test of HI (2012). Despite using less information than the latter, we show that it is used in a more efficient way, and that our Wald test has better power properties than the HI test when dealing with small samples.

Our testing approach can be easily implemented in any statistical package and it is useful to avoid serious forecasting/estimation problems in standard econometric practices with factors. This could be the case of FAR and FAVAR models, when the factor loadings are subject to big breaks and the number of factors is a priori determined (as is conventional in several macroeconomic and financial applications).

In the second step of our testing approach, a Sup-type test is used to detect a break of the parameters in that regression when the break date is assumed to be unknown. As the simulations show, this test performs very well especially when $T \geq 100$. For smaller samples, as it happens with many other Sup-type tests, bootstrap can improve the finite-sample performance of the test compared to the tabulated asymptotic critical values of Andrews (1993), as suggested by Diebold and Chen (1996). It is high in our research agenda to explore this possibility.

Moreover, as discussed earlier in Remarks 8 and 9, many other existing methods for testing structural changes in linear regressions can also be applied in our second-stage procedure. Further, our testing approach can allow for multiple big breaks through sequential estimation, like in Bai and Perron (1998, 2003), for locating the candidate break dates. Exploring further this issue remains

also high in our research agenda.

Finally, though a simple testing procedure has been outlined in section 4.4 to identify whether breaks stem from loadings or from the volatility of the factors, we plan to derive other alternative tests based on the rank of the covariance matrix of the estimated factors in different subsamples which can also be extended to test for other sources of parameter instability.

Appendix

A.1: Proof of Propositions 1 and 2

The proof proceeds by showing that the errors, factors and loadings in model (5) satisfy Assumptions A to D of BN (2002). Then, once these results are proven, Propositions 1 and 2 just follow immediately from application of Theorems 1 and 2 of BN (2002). Define $F_t^* = [F_t' \ G_t^{1'}]'$, $\epsilon_t = HG_t^2 + e_t$, and $\Gamma = [A \ \Lambda]$. The proofs of Lemma 1 to Lemma 8 are similar to those in BN (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 $T^{-1} \sum_{t=1}^T F_t^* F_t^{*'} \xrightarrow{p} \Sigma_F^*$ as $T \rightarrow \infty$ for some positive matrix Σ_F^* .

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

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

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

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

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

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

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

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

Finally, it is easy to verify that $\sum_{i=1}^N \sum_{t=1}^T \|\eta_i\|^2 E\|F_t\|^2 = O(1)$ and thus the new idiosyncratic errors ϵ_t satisfy the necessary condition for the consistency of \hat{r} (See Observation 1 of Bates et al. 2013).

Once it has been shown that the new factors: F_t^* , the new loadings: Γ and the new errors: ϵ_t all satisfy the conditions of BN (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

Under the null: $k_1 = 0$, when $\bar{r} = r$ we have

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

Let $D_{(i)}$ denote the i th row of D , and $D_{(\cdot, j)}$ denote the j th column of D . Define $\hat{F}_t = DF_t$, and $\hat{F}_{kt} = D_{(k, \cdot)} \times F_t$ as the k th element of \hat{F}_t . Let \hat{F}_{1t} be the first element of \hat{F}_t , and $\hat{F}_{-1t} = [\hat{F}_{2t}, \dots, \hat{F}_{rt}]'$, while \hat{F}_{1t} and \hat{F}_{-1t} can be defined in the same way. Note that \hat{F}_t depends on N and T . 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 $X_{it} = \alpha_i F_t + e_{it} + \eta_i G_t^2$. However, since $\eta_i G_t^2$ is $O_p(1/\sqrt{NT})$ by Assumption 1, we can use similar methods as in Appendix A.1 to show that an error term of this order can be ignored and that the asymptotic properties of \hat{F}_t will not be affected (See Remark 5 of Bai, 2009). Therefore, for the sake of simplicity in the presentation below, we eliminate the last term and consider instead the model $X_{it} = \alpha_i F_t + e_{it}$ in the following lemmatae (9 to 13) required to prove Lemma 14, which is the key result in the proof of Theorem 1.

Lemma 9.

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

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

Lemma 10.

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

Proof. Note that:

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

Thus,

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

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

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

Lemma 11.

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

Proof. See Lemma 10 and Assumption 6. \square

Lemma 12.

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

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

Let \Rightarrow denote *weak convergence*. \mathcal{F}_{1t} , \mathcal{F}_{-1t} , D^* and S are as defined in the paper (Page 12). Then:

Lemma 13.

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

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

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

Lemma 14.

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

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

Proof. If we show that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T\pi} \left[\mathcal{F}_{-1t} \mathcal{F}_{1t} - T^{-1} \sum_{s=1}^T \mathcal{F}_{-1s} \mathcal{F}_{1s} \right] \Rightarrow S^{1/2} \mathcal{B}_{r-1}^0(\pi) \quad (24)$$

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

$$\sup_{\pi \in [0,1]} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T\pi} \hat{\mathcal{F}}_{-1t} \hat{\mathcal{F}}_{1t} - \frac{1}{\sqrt{T}} \sum_{t=1}^{T\pi} \left[\mathcal{F}_{-1t} \mathcal{F}_{1t} - T^{-1} \sum_{s=1}^T \mathcal{F}_{-1s} \mathcal{F}_{1s} \right] \right\| = o_p(1), \quad (25)$$

then the result follows from Lemma 11.

First note that

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

hence (24) can be verified by applying Lemma 13.

To prove (25), we first define D_{-1} as the second to last rows of D , and D_1 as the first row of D . D_{-1}^* and D_1^* are defined in the same manner. Then we have

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

and

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

It follows that:

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

Next, define $\overline{FF'} = \frac{1}{T} \sum_{s=1}^T F_sF_s'$, then:

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

Combining the above results gives:

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

Following similar arguments as in Lemma 13, it can be proved that

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

Moreover, it is easy to check that $\|D\| = O_p(1)$ and $\|D - D^*\| = o_p(1)$ (See Bai 2003). Finally, $\left\| \frac{1}{\sqrt{T}} \sum_{s=1}^T \hat{F}_{-1s} \hat{F}_{1s} \right\|$ is $o_p(1)$ by Lemma 12. Then (25) holds and we obtain the desired conclusion. \square

Theorem 1:

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

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

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

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

$$1/T \sum_{t=1}^{\tau} \hat{F}_{-1t} \hat{F}'_{-1t} = \pi \frac{1}{\tau} \sum_{t=1}^{\tau} \hat{F}_{-1t} \hat{F}'_{-1t} = \pi \frac{1}{\tau} \sum_{t=1}^{\tau} \mathcal{F}_{-1t} \mathcal{F}'_{-1t} + o_p(1). \quad (26)$$

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

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

Notice that $E(\mathcal{F}_{-1t} \mathcal{F}'_{-1t}) = I_{r-1}$, because $E(\mathcal{F}_t \mathcal{F}'_t) = D^* \Sigma_F D^* = V^{-1/2} \Gamma' \Sigma_{\Lambda}^{-1/2} \Sigma_F \Sigma_{\Lambda}^{-1/2} \Gamma V^{-1/2} = I_r$. Applying law of large numbers to (26) 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. Then it follows from Lemma 14 that:

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

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

A.3: Consistent Estimator of S

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

that $S = \lim \text{Var} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \mathcal{F}_{-1t} \mathcal{F}_{1t} \right)$. Notice that $E(\mathcal{F}_{-1t} \mathcal{F}_{1t}) = 0$ as shown above.

First, we define the infeasible estimator of S :

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

where $m = o(T^{\frac{1}{4}})$, $w(j, m) = 1 - \frac{j}{m+1}$ is the Bartlett kernel, and

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

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

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

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

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

where

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

then we have the following results:

Proposition 3. Assume that Assumptions 1 to 11 hold, under the null $H_0 : k_1 = 0$, we have

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

Proof. Given Assumption 11, it suffices to show that

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

It is easy to see that:

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

For the right-hand side we obtain that:

$$\begin{aligned} & \sup_{0 \leq j \leq m} \|\hat{\Gamma}_j(\hat{F}) - \hat{\Gamma}_j(\mathcal{F})\| \\ & \leq \sup_{0 \leq j \leq m} \frac{1}{T} \sum_{t=j+1}^T \|\hat{F}_{-1t} \hat{F}_{1t} \hat{F}_{1t-j} \hat{F}_{-1t-j}' - \mathcal{F}_{-1t} \mathcal{F}_{1t} \mathcal{F}_{1t-j} \mathcal{F}_{-1t-j}'\| \end{aligned}$$

$$\begin{aligned}
&\leq \sup_{0 \leq j \leq m} \frac{1}{T} \sum_{t=j+1}^T \left\| (\hat{F}_{1t} \hat{F}_{-1t} - \mathcal{F}_{1t} \mathcal{F}_{-1t}) \hat{F}_{1t-j} \hat{F}'_{-1t-j} \right\| + \sup_{0 \leq j \leq m} \frac{1}{T} \sum_{t=j+1}^T \left\| \mathcal{F}_{1t} \mathcal{F}_{-1t} (\hat{F}_{1t-j} \hat{F}'_{-1t-j} - \mathcal{F}_{1t-j} \mathcal{F}'_{-1t-j}) \right\| \\
&\leq \sqrt{\frac{1}{T} \sum_{t=1}^T \left\| \hat{F}_{1t} \hat{F}_{-1t} - \mathcal{F}_{1t} \mathcal{F}_{-1t} \right\|^2} \left(\sqrt{\frac{1}{T} \sum_{t=1}^T \left\| \hat{F}_{1t} \hat{F}_{-1t} \right\|^2} + \sqrt{\frac{1}{T} \sum_{t=1}^T \left\| \mathcal{F}_{1t} \mathcal{F}_{-1t} \right\|^2} \right). \\
&\frac{1}{T} \sum_{t=1}^T \left\| \hat{F}_{1t} \hat{F}_{-1t} \right\|^2 \leq \frac{1}{T} \sum_{t=1}^T \left\| \hat{F}_t \right\|^4 = O_p(1) \text{ by Lemma 5 of HI (2012), and } \frac{1}{T} \sum_{t=1}^T \left\| \mathcal{F}_{1t} \mathcal{F}_{-1t} \right\|^2 \leq \\
&\frac{1}{T} \sum_{t=1}^T \left\| \mathcal{F}_t \right\|^4 = O_p(1) \text{ by Assumption 2. Furthermore,}
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{T} \sum_{t=1}^T \left\| \hat{F}_{1t} \hat{F}_{-1t} - \mathcal{F}_{1t} \mathcal{F}_{-1t} \right\|^2 \\
&\leq \frac{1}{T} \sum_{t=1}^T \left\| (\hat{F}_{1t} - \mathcal{F}_{1t}) \hat{F}_{-1t} \right\|^2 + \frac{1}{T} \sum_{t=1}^T \left\| (\hat{F}_{-1t} - \mathcal{F}_{-1t}) \mathcal{F}_{1t} \right\|^2 \\
&\leq \sqrt{\frac{1}{T} \sum_{t=1}^T \left\| \hat{F}_t - \mathcal{F}_t \right\|^4} \left(\sqrt{\frac{1}{T} \sum_{t=1}^T \left\| \hat{F}_t \right\|^4} + \sqrt{\frac{1}{T} \sum_{t=1}^T \left\| \mathcal{F}_t \right\|^4} \right).
\end{aligned}$$

It can be proved that $\frac{1}{T} \sum_{t=1}^T \left\| \hat{F}_t - \mathcal{F}_t \right\|^4 = O_p(1/T) + O_p(1/N^2)$ (very similar to Theorem 2 of HI, 2012). Then, under the assumption that $\sqrt{T}/N \rightarrow 0$, it follows that

$$\sup_{0 \leq j \leq m} \frac{1}{T} \sum_{t=j+1}^T \left\| \hat{F}_{1t} \hat{F}_{-1t} \hat{F}_{1t-j} \hat{F}'_{-1t-j} - \mathcal{F}_{1t} \mathcal{F}_{-1t} \mathcal{F}_{1t-j} \mathcal{F}'_{-1t-j} \right\| = O_p(T^{-1/4}),$$

which implies $\left\| \hat{S}(\hat{F}) - \hat{S}(\mathcal{F}) \right\| = o_p(1)$ given that $m = o(T^{1/4})$. □

References

- Ahn, S. and A. Horenstein. Eigenvalue ratio test for the number of factors. *Econometrica*. Forthcoming.
- Andrews, D. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* 61(4), 821–856.
- Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* 71(1), 135–171.
- Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica* 77(4), 1229–1279.
- Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* 70(1), 191–221.
- Bai, J. and S. Ng (2006a). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica* 74(4), 1133–1150.
- Bai, J. and S. Ng (2006b). Evaluating latent and observed factors in macroeconomics and finance. *Journal of Econometrics* 131(1-2), 507–537.
- Bai, J. and S. Ng (2011). Principal components estimation and identification of the factors. *Columbia University, Working paper*.
- Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* 66(1), 47–78.
- Bai, J. and P. Perron (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* 18(1), 1–22.
- Banerjee, A., M. Marcellino, and I. Masten (2008). Forecasting macroeconomic variables using diffusion indexes in short samples with structural change. *CEPR. DP. 6706*.
- Bates, B., M. Plagborg-Møller, J. Stock, and M. Watson (2013). Consistent factor estimation in dynamic factor models with structural instability. *Journal of Econometrics*. Forthcoming.
- Bernanke, B., J. Boivin, and P. Elias (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics* 120(1), 387–422.
- Breitung, J. and S. Eickmeier (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics* 163(1), 71–84.
- Brown, R., J. Durbin, and J. Evans (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, 149–192.
- Charnavoki, V. and J. Dolado (2012). The effects of global shocks on small commodity-exporting economies: New evidence from Canada. *CEPR. DP. 8825*.
- Chen, B. and Y. Hong (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica* 80(3), 1157–1183.
- Chen, L. (2012). Identifying observed factors in high dimensional factor models. *Universidad Carlos III de Madrid, Working paper*.
- Chen, N., R. Roll, and S. Ross (1986). Economic forces and the stock market. *Journal of Business*, 383–403.
- Diebold, F. and C. Chen (1996). Testing structural stability with endogenous breakpoint a size comparison of analytic and bootstrap procedures. *Journal of Econometrics* 70(1), 221–241.
- Fernández-Villaverde, J., P. A. Guerrón-Quintana, and J. Rubio-Ramírez (2010). Reading the recent monetary history of the US, 1959-2007. *Federal Reserve Board of St. Louis Review*, 92, 1–28.

- 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 (2012). Tests for parameter instability in dynamic factor models. *North Carolina State University, Working paper*.
- Hansen, B. (2000). Testing for structural change in conditional models. *Journal of Econometrics* 97(1), 93–115.
- Kim, D. and P. Perron (2009). Unit root tests allowing for a break in the trend function at an unknown time under both the null and alternative hypotheses. *Journal of Econometrics* 148(1), 1–13.
- Newey, W. and K. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3), 703–708.
- Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica* 77(5), 1447–1479.
- Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics* 92(4), 1004–1016.
- Perron, P. (2006). Dealing with structural breaks. *Palgrave handbook of econometrics* 1, 278–352.
- Rudebusch, G. and T. Wu (2008). A macro-finance model of the term structure, monetary policy and the economy. *The Economic Journal* 118(530), 906–926.
- Shanken, J. and M. Weinstein (2006). Economic forces and the stock market revisited. *Journal of Empirical Finance* 13(2), 129–144.
- Stock, J. and M. Watson (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97(460), 1167–1179.
- Stock, J. and M. Watson (2009). Forecasting in dynamic factor models subject to structural instability. *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry*, 173–205.

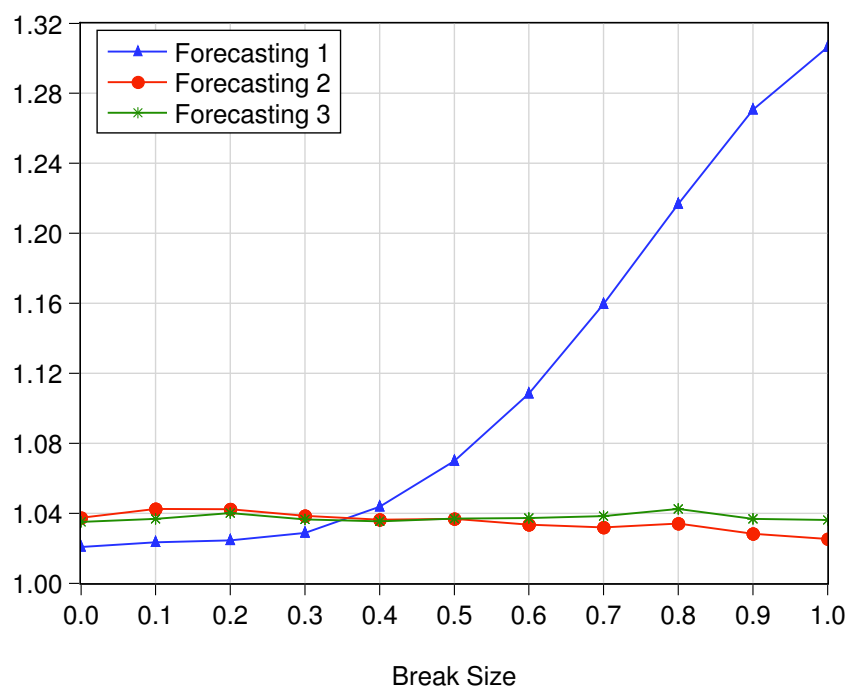


Figure 1: The MSEs of different forecasting methods in the presence of big breaks (see Section 5.1).

Table 1: Empirical Sizes of the Nominal 5% Size Tests for Different Choices of \bar{r} when $r = 3$.

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

Notes: The DGP is $X_{it} = \sum_{k=1}^3 \alpha_{ik} F_{kt} + e_{it}$ where $F_{kt} = \phi_k F_{k,t-1} + v_{kt}$, $\alpha_{ik}, e_{it}, v_{kt} \sim i.i.d N(0, 1)$, and $[\phi_1, \phi_2, \phi_3] = [0.8, 0.5, 0.2]$ (See Section 5.2).

Table 2: Empirical Size of the Nominal 5% Size Tests for Different Choices of \bar{r} when $r = 3$, and when the idiosyncratic errors are cross sectionally and serially correlated.

	$N = 100, T = 50$						$N = 100, T = 100$						$N = 100, T = 200$					
	Wald			Sup Wald			Wald			Sup Wald			Wald			Sup Wald		
	$\bar{r} = 2$	$\bar{r} = 3$	$\bar{r} = 4$	$\bar{r} = 2$	$\bar{r} = 3$	$\bar{r} = 4$	$\bar{r} = 2$	$\bar{r} = 3$	$\bar{r} = 4$	$\bar{r} = 2$	$\bar{r} = 3$	$\bar{r} = 4$	$\bar{r} = 2$	$\bar{r} = 3$	$\bar{r} = 4$	$\bar{r} = 2$	$\bar{r} = 3$	$\bar{r} = 4$
$\rho = 0.2, \beta = J = 0$	5.9	3.8	6.3	11.4	25.0	45.2	6.0	2.9	2.2	4.3	8.2	13.8	5.3	4.0	4.0	3.7	3.0	4.0
$\rho = 0.5, \beta = J = 0$	6.9	3.6	13.2	11.7	24.2	54.1	5.6	3.2	4.0	4.7	8.4	21.6	5.1	4.1	4.4	4.2	3.5	6.2
$\rho = 0.7, \beta = J = 0$	8.6	7.0	30.2	12.7	26.2	63.7	6.0	3.1	11.0	4.7	7.8	37.7	5.1	3.7	7.6	3.7	3.8	14.0
$\rho = 0, \beta = 0.2, J = 5$	5.9	4.6	4.0	12.8	22.9	46.2	5.5	2.8	1.4	4.7	6.9	13.5	4.9	3.6	3.9	3.7	3.4	3.5
$\rho = 0, \beta = 0.2, J = 10$	6.1	3.7	4.3	12.3	26.1	48.3	5.2	2.3	1.4	5.9	8.4	13.8	4.8	3.6	2.9	3.6	3.8	5.0
$\rho = 0, \beta = 0.5, J = 10$	5.2	3.4	4.1	11.4	26.7	52.7	5.7	3.2	2.2	4.9	6.8	14.5	6.2	4.5	4.0	4.6	3.4	4.7
$\rho = 0.2, \beta = 0.2, J = 5$	6.3	4.6	4.7	12.1	23.3	44.2	5.4	2.9	1.4	4.7	6.9	12.9	5.1	3.7	3.6	4.1	3.8	4.5
$\rho = 0.2, \beta = 0.5, J = 10$	6.2	3.3	4.1	11.1	26.7	51.6	6.1	4.2	2.3	4.9	7.8	15.7	5.9	4.8	4.2	3.6	3.5	3.5
$\rho = 0.5, \beta = 0.2, J = 5$	7.6	4.4	7.6	12.7	24.7	46.0	5.7	3.0	2.7	5.3	7.6	15.5	5.2	3.8	4.2	3.5	3.6	5.8
$\rho = 0.5, \beta = 0.5, J = 10$	6.2	4.8	7.9	14.7	29.2	51.7	7.2	4.6	3.9	5.8	9.8	16.5	6.1	5.0	4.1	4.7	4.8	6.4

Notes: The DGP is $X_{it} = \sum_{k=1}^3 \alpha_{ik} F_{kt} + e_{it}$ where $F_{kt} = \phi_k F_{k,t-1} + v_{kt}$, and $[\phi_1, \phi_2, \phi_3] = [0.8, 0.5, 0.2]$. $e_{it} = \rho e_{i,t-1} + u_{it} + \beta \sum_{h=i-J}^{i-1} u_{ht}$, where $\alpha_{ik}, u_{it}, v_{kt} \sim i.i.d. N(0, 1)$ (See Section 5.2).

Table 3: Empirical Power of Nominal 5% Size Tests for Different Choices of \bar{r} when $r = 2$ and $k_1 = 2$.

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

Notes: The DGP is $X_{it} = \sum_{k=1}^2 \alpha_{ik} F_{kt} + e_{it}$ where $F_{kt} = \phi_k F_{k,t-1} + v_{kt}$, $\alpha_{ik}, e_{it}, v_{kt} \sim i.i.d N(0, 1)$, and $[\phi_1, \phi_2] = [0.8, 0.2]$. The shifts in the means of the factor loadings are 0.4 and 0.2 at $\tau = T/2$ (See Section 5.3).

Table 4: Empirical Power of Nominal 5% Size Tests for Different Choices of \bar{r} when $r = 2$, $k_1 = 1$, and \hat{F}_{1t} is used as the regressand.

N	T	$\bar{r} = 2$				$\bar{r} = 3$				$\bar{r} = 4$			
		LM	Sup LM	Wald	Sup Wald	LM	Sup LM	Wald	Sup Wald	LM	Sup LM	Wald	Sup Wald
50	50	9.7	4.5	64.1	66.5	2.0	3.7	99.0	99.9	1.4	3.5	98.7	100
50	100	13.9	4.3	60.5	64.0	8.7	1.4	98.3	99.5	2.0	0.7	98.2	99.7
50	150	15.3	7.8	56.9	62.1	15.2	4.6	98.4	99.7	8.8	1.2	98.4	99.8
50	200	17.6	11.9	56.5	62.4	19.5	9.4	97.2	100	12.3	3.4	97.6	100
100	50	10.6	4.1	69.6	72.3	3.0	4.6	99.5	100	2.3	3.7	99.5	100
100	100	14.2	5.1	67.1	70.3	9.0	2.1	99.8	100	3.5	0.9	99.8	100
100	150	15.4	10.3	63.1	68.4	13.5	4.0	99.4	100	8.0	1.2	99.2	100
100	200	17.2	11.3	67.5	72.1	16.5	7.7	99.7	100	10.5	2.2	99.7	100
150	50	9.7	5.3	73.9	76.8	2.9	3.9	99.9	100	1.2	2.9	100	100
150	100	12.2	5.3	70.5	74.1	7.8	2.1	99.8	100	1.8	0.7	99.8	100
150	150	12.2	7.3	69.6	74.7	9.3	3.0	99.7	100	5.2	0.9	99.9	100
150	200	12.3	10.7	64.9	70.5	11.7	7.6	99.9	100	6.6	2.2	99.9	100
200	50	9.4	5.8	75.3	77.3	2.9	4.3	100	100	1.9	3.5	100	100
200	100	10.8	5.7	71.6	76.4	8.0	2.1	99.8	100	2.6	0.8	99.9	100
200	150	13.1	10.4	70.9	79.0	9.4	4.7	99.8	100	5.6	0.9	99.9	100
200	200	13.8	10.2	73.1	76.8	13.7	7.4	100	100	8.4	2.0	100	100

Notes: The DGP is $X_{it} = \alpha_{i1}F_{1t} + \alpha_{i2}F_{2t} + e_{it}$ for $t = 1, \dots, T/2$, and $X_{it} = \alpha_{i1}F_{1t} + \beta_{i2}F_{2t} + e_{it}$ for $t = T/2 + 1, \dots, T$, where $F_{kt} = \phi_k F_{k,t-1} + v_{kt}$, $\alpha_{i1}, \alpha_{i2}, \beta_{i2}, e_{it}, v_{kt} \sim i.i.d. N(0, 1)$, and $[\phi_1, \phi_2] = [0.8, 0.2]$ (See Section 5.3).

Table 5: Empirical Power of Nominal 5% Size Tests for Different Choices of \bar{r} when $r = 2$, $k_1 = 1$, and \hat{F}_{2t} is used as the regressand.

N	T	$\bar{r} = 2$				$\bar{r} = 3$				$\bar{r} = 4$			
		LM	Sup LM	Wald	Sup Wald	LM	Sup LM	Wald	Sup Wald	LM	Sup LM	Wald	Sup Wald
50	50	9.7	4.5	20.2	32.1	17.8	2.5	94.4	80.6	3.0	2.7	87.1	80.2
50	100	13.9	4.3	16.6.5	23.3	87.8	2.1	98.9	88.9	78.7	0.3	98.6	84.7
50	150	15.3	7.8	18.0	20.9	92.5	78.4	99.7	100	89.9	21.9	99.5	97.0
50	200	17.6	11.9	19.5	22.3	95.1	90.9	99.7	100	93.0	85.9	99.7	100
100	50	10.6	6.1	19.5	33.7	18.8	2.8	96.0	84.8	2.6	3.3	91.6	83.8
100	100	14.2	5.1	19.5	25.1	87.4	1.9	99.6	90.6	79.9	1.4	99.5	87.1
100	150	15.4	10.3	18.3	21.1	93.8	81.6	99.4	100	90.9	25.3	99.2	97.1
100	200	17.2	11.3	19.2	20.6	94.7	90.9	99.6	100	93.3	85.2	99.5	100
150	50	9.7	5.3	17.4	32.2	20.8	2.9	97.1	87.9	3.1	2.8	93.0	86.4
150	100	12.2	5.3	17.1	23.0	88.4	2.1	99.9	93.3	80.3	0.8	99.9	90.6
150	150	12.2	7.3	14.3	23.3	93.3	79.1	99.7	100	91.2	23.0	99.6	98.1
150	200	12.3	10.7	14.7	19.3	94.5	90.7	99.7	100	93.6	85.7	99.7	100
200	50	9.4	5.8	17.7	32.7	20.0	3.9	96.5	87.8	2.4	4.7	92.4	87.0
200	100	10.8	5.7	15.1	24.1	87.4	2.4	100	94.4	78.4	1.1	99.9	90.9
200	150	13.1	10.4	15.3	23.3	93.1	80.7	99.9	100	91.2	26.2	99.8	98.4
200	200	13.8	10.2	15.8	20.3	93.9	89.6	99.7	100	92.3	84.9	99.7	100

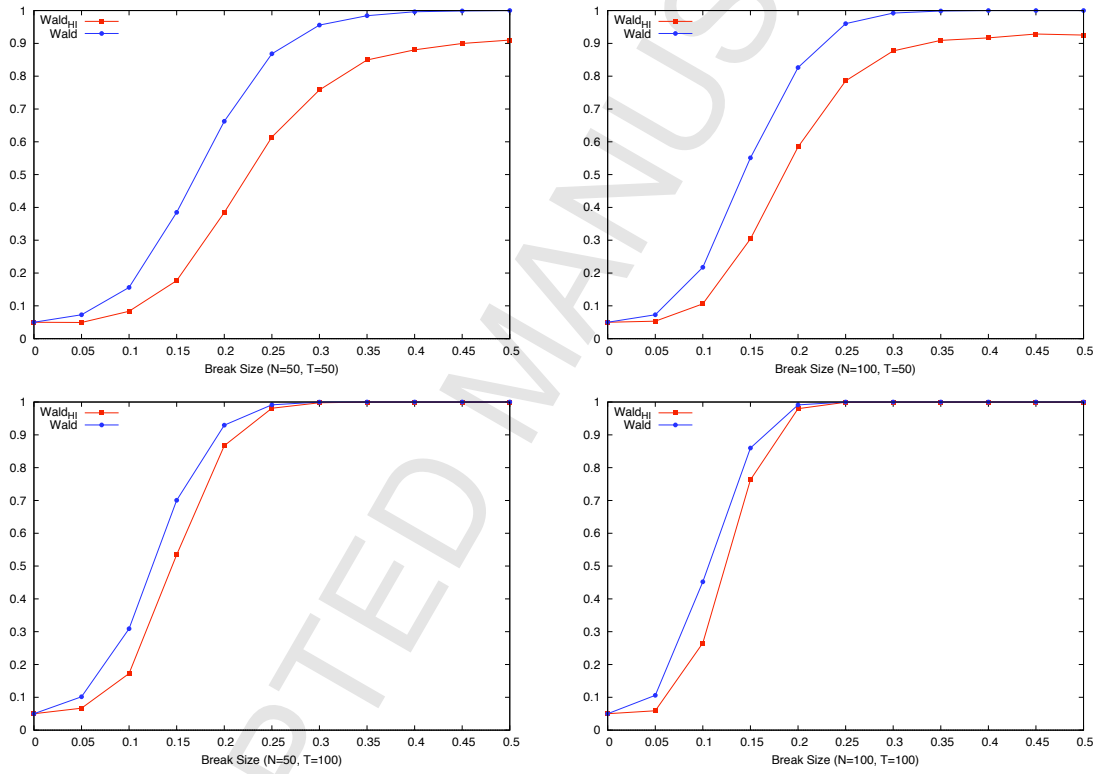
Notes: DGP of Table 4 (See Section 5.3).

Table 6: Size and Power Comparisons of BE (2011) and our Wald Tests at Nominal 5% Size for $r = 2$.

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

Notes: DGP of Table 3. The shift in the mean of the factor loadings is either zero (no break) or 0.1 (break) (See Section 5.4).

Table 7: Power Comparison of HI (2012) and our Wald Tests for $N, T \leq 100$ when $r = 1$.



Notes: The DGP is $X_{it} = A_{1i}F_t + e_{it}$ for $t = 1, \dots, T/2$, and $X_{it} = (A_{1i} + b)F_t + e_{it}$ for $t = T/2 + 1, \dots, T$, where $F_t = \phi F_{t-1} + u_t$, $A_{1i}, u_t, e_{it} \sim i.i.d N(0, 1)$, and $\phi = 0.8$. The reported curves are the size adjusted power curves of our Wald test (blue) and HI's Wald test (red) when the break size b increases from 0 to 0.5 (See Section 5.5).

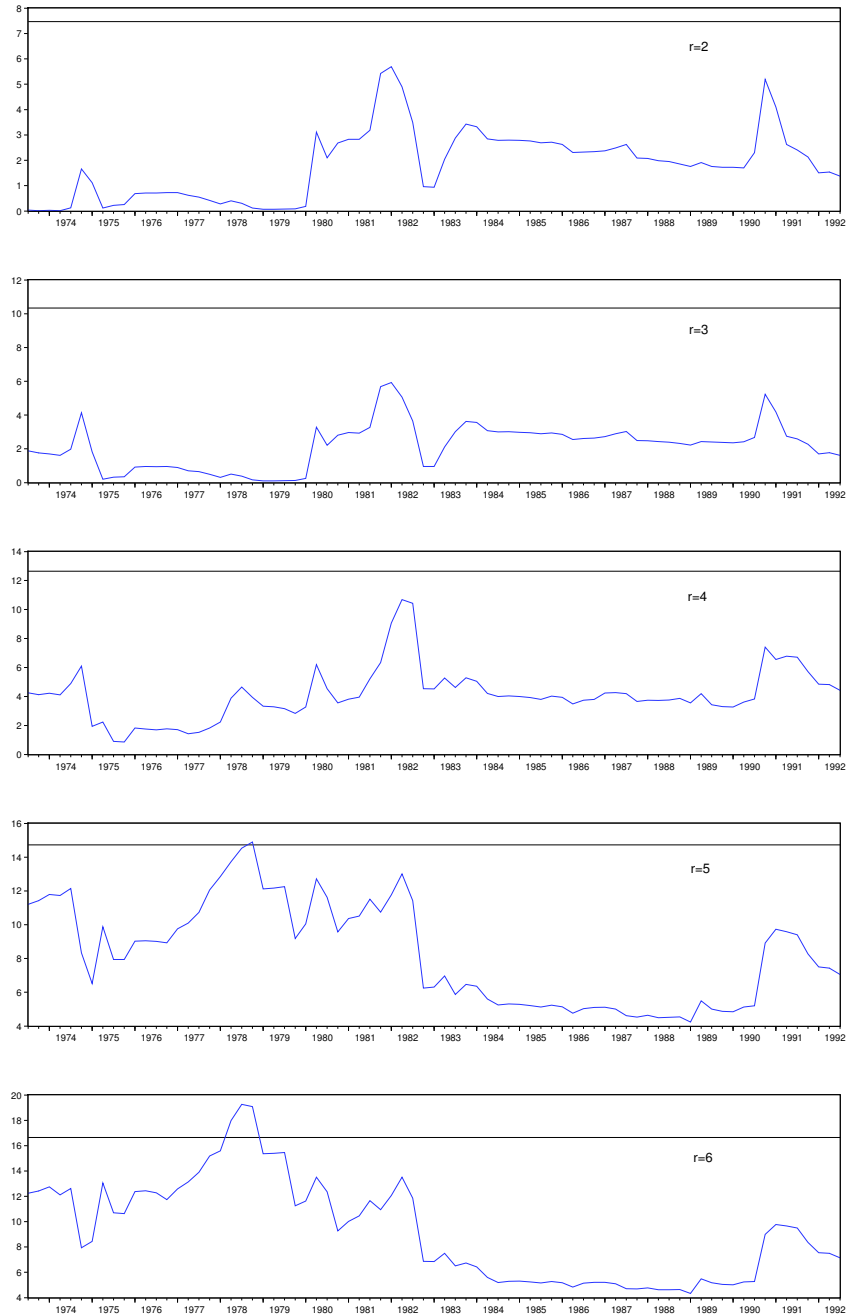


Figure 2: US data set of Stock and Watson (2009), from 1959:Q1 to 2006:Q4. The trimming $\Pi = [0.3, 0.7]$ is used for the Wald tests with $\bar{r} = 2$ to 6 (from top to bottom), and the horizontal lines are the corresponding 5% asymptotic critical values for the Sup-Wald Test.