



King's Research Portal

DOI:

[10.1111/obes.12311](https://doi.org/10.1111/obes.12311)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Massacci, D. (2019). Unstable Diffusion Indexes: With an Application to Bond Risk Premia. *OXFORD BULLETIN OF ECONOMICS AND STATISTICS*, 81(6), 1376-1400. <https://doi.org/10.1111/obes.12311>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Unstable Diffusion Indexes: With an Application to Bond Risk Premia*

DANIELE MASSACCI

*Bank of England, Threadneedle Street London, EC2R 8AH, UK
(e-mail: daniele.massacci@gmail.com)*

Abstract

This paper studies the empirically relevant problem of estimation and inference in diffusion index forecasting models with structural instability. Factor model and factor augmented regression both experience a structural change with different unknown break dates. In the factor model, we estimate factors and loadings by principal components. We consider least squares estimation of the factor augmented regression and propose a break test. The empirical application uncovers instabilities in the linkages between bond risk premia and macroeconomic factors.

I. Introduction

The diffusion index forecasting model of Stock and Watson (1998, 2002) and Bai and Ng (2006) is a regression model with observable and latent covariates, where the latter are the common factors in the variables of a large-scale data set. The model is usually estimated under the assumption of structural stability in the loadings of the factor model and in the slope coefficients of the factor augmented regression: Connor and Korajczyk (1986, 1988, 1993), Bai and Ng (2002), Stock and Watson (2002) and Bai (2003) deal with linear static factor models; Forni *et al.* (2000, 2004), Forni and Lippi (2001) and Forni *et al.* (2015) study the linear generalized dynamic factor model; Stock and Watson (2002) and Bai and Ng (2006) focus on stable factor augmented regressions. Violation of the stability assumption may affect empirical results. For example, out-of-sample forecasts may not be accurate due

JEL Classification numbers: C12, C13, C38, C52, G12.

*This paper was started when the author was Franco Modigliani Research Fellow in Economics and Finance at the Einaudi Institute for Economics and Finance (EIEF). The views are the author's and do not necessarily reflect those of the Bank of England or its policy committees. The author is indebted to Marco Lippi for introducing him to factor models and for several enlightening conversations. This paper benefits from comments from participants at the International Association for Applied Econometrics (IAAE) 2016 Annual Conference, at the Symposium on Statistical Penalization Methods and Dimension Reduction Methods for Economic and Financial Analysis at the University of York, at the Vienna Workshop on High-Dimensional Time Series in Macroeconomics and Finance 2017, and at the seminar series at The University of Nottingham Granger Centre for Time Series Econometrics; and from conversations with Laura Coroneo, Pasquale Della Corte, Domenico Giannone, Christian Gourieroux, Refet Gürkaynak, George Kapetanios, Andrew Meldrum and Hashem Pesaran. Errors and omissions are the author's responsibility. The financial support from the Associazione Borsisti Marco Fanno and from UniCredit and Universities Foundation is gratefully acknowledged.

to instabilities in the factor augmented regression (Stock and Watson, 2002, 2009) or in the factor model (Giannone, 2007; Banerjee, Marcellino and Masten, 2008). In the latter, structural breaks enlarge the factor space (Breitung and Eickmeier, 2011; Chen, Dolado and Gonzalo, 2014) without conveying more information: including additional factors in the factor augmented regression increases estimation noise and deteriorates forecast performance (Han and Inoue, 2015).

A number of contributions analyses large dimensional factor models with structural instabilities. Bates *et al.* (2013) study the robustness of the principal components estimator as applied to factor models under neglected instability. Breitung and Eickmeier (2011), Chen *et al.* (2014), Han and Inoue (2015), Yamamoto and Tanaka (2015) and Barigozzi and Trapani (2017) develop statistical tools to detect breaks. Chen (2015), Cheng, Liao and Schorfheide (2016), Bai, Han and Shi (2017) and Baltagi, Kao and Wang (2017) focus on estimation under the single break assumption. Baltagi, Kao and Wang (2016), Ma and Su (2016) and Barigozzi, Cho and Fryzlewicz (2018) allow for multiple breaks. Massacci (2017) studies large dimensional threshold factor models with regime shifts in the loadings driven by a covariate. These contributions do not consider the whole diffusion index forecasting model. Corradi and Swanson (2014) propose a test for the joint null hypothesis of stable factor model and factor-augmented regression. Wang, Cui and Li (2015) estimate unstable factor augmented regressions with factors extracted from a linear model. To the very best of our knowledge, estimation and inference in the set up with instabilities in the factor model and in the factor augmented regression has not been studied: we aim at filling this gap.

We start from the single break factor model. Let N and T denote the cross-sectional and time series dimensions respectively. We estimate the model by least squares by minimizing the sum of squared residuals (Baltagi *et al.*, 2017; Massacci, 2017): the resulting principal components estimator for factors and loadings has the same convergence rate $C_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$ as in the linear case (Bai and Ng, 2002). We then turn to the single break factor augmented regression. We estimate the parameters by least squares by replacing the latent factors with their estimates (Bai, 1997; Bai and Ng, 2006): despite the structural instability, the least squares estimator for the slope coefficients is \sqrt{T} consistent and asymptotically normal provided that $\sqrt{T}/N \rightarrow 0$ as $N, T \rightarrow \infty$. We then propose a Lagrange multiplier test for the null hypothesis of stability: we show that the critical values provided in Andrews (1993) remain valid also in the presence of latent factors estimated from an unstable large dimensional model.

Finally, we apply our methodology to bond risk premia (Fama and Bliss, 1987; Cochrane and Piazzesi, 2005; Ludvigson and Ng, 2009). Common factors extracted from a large set of macroeconomic series help predicting bond excess returns (Ludvigson and Ng, 2009); however, the loadings are not stable over time (Breitung and Eickmeier, 2011; Chen *et al.*, 2014; Cheng *et al.*, 2016). Pricing equations for bond risk premia experience structural breaks (Smith and Taylor, 2009; Bikbov and Chernov, 2010). We thus use our model to uncover instabilities in the linkages between bond market risk premia and macroeconomic fundamentals. Depending on the maturity of the bond, we show that: predictive regressions for bond risk premia are stable for most of the Great Moderation (Joslin, Priebsch and Singleton, 2014); a break occurred in the early 1980s (Smith and Taylor, 2009).

The paper is organized as follows. Section II describes the unstable diffusion index forecasting model. Section III gives relevant information on estimation and model selection in large dimensional factor models subject to structural break. Section IV looks at estimation and inference in the unstable factor augmented regression. Section V discusses the Monte Carlo analysis. Section VI provides the application to bond risk premia. Section VII outlines possible extensions and modifications. Section VIII concludes. In the Online Appendix, Section A provides details for the Monte Carlo analysis and Section B collects technical proofs.

Concerning notation, $\mathbb{I}(\cdot)$ denotes the indicator function; given a square matrix \mathbf{A} , $\text{tr}(\mathbf{A})$ denotes the trace of \mathbf{A} ; the norm of a generic matrix \mathbf{A} is $\|\mathbf{A}\| = [\text{tr}(\mathbf{A}'\mathbf{A})]^{1/2}$; for a given scalar A , $|A|$, \mathbf{I}_A and $\mathbf{0}_A$ are the absolute value of A , the $A \times A$ identity matrix and the $A \times A$ zero matrix, respectively; \xrightarrow{p} denotes convergence in probability; \xrightarrow{d} denotes convergence in distribution; \Rightarrow denotes weak convergence. Without round brackets inside, $[\cdot]$ is the integer part of the argument.

II. The unstable diffusion index forecasting model

We consider

$$\mathbf{x}_t = \mathbb{I}(t/T \leq \pi_x)\mathbf{\Lambda}_1\mathbf{f}_t + \mathbb{I}(t/T > \pi_x)\mathbf{\Lambda}_2\mathbf{f}_t + \mathbf{e}_t, \quad t = 1, \dots, T, \tag{1}$$

$$y_{t+h} = \mathbb{I}(t/T \leq \pi_y)(\gamma'_1\mathbf{f}_t + \boldsymbol{\beta}'_1\mathbf{w}_t) + \mathbb{I}(t/T > \pi_y)(\gamma'_2\mathbf{f}_t + \boldsymbol{\beta}'_2\mathbf{w}_t) + \varepsilon_{t+h}, \quad t = 1, \dots, T, \quad h \geq 0, \tag{2}$$

where T is the time series dimension of the sample. Starting from equation (1), $\mathbf{x}_t = (x_{1t}, \dots, x_{Nt})' \in \mathfrak{R}^N$ is the $N \times 1$ vector of observable dependent variables; $\mathbf{f}_t = (f_{1t}, \dots, f_{Rt})' \in \mathfrak{R}^R$ is the $R \times 1$ vector of latent factors; $\mathbf{e}_t = (e_{1t}, \dots, e_{Nt})' \in \mathfrak{R}^N$ is the $N \times 1$ vector of idiosyncratic errors; π_x is the break fraction; $\mathbf{\Lambda}_j = (\lambda_{j1}, \dots, \lambda_{jN})'$ is the $N \times R$ matrix of factor loadings, with i -th row defined as $\lambda'_{ji} = (\lambda_{ji1}, \dots, \lambda_{jiR})$, for $j = 1, 2$ and $i = 1, \dots, N$. Moving to equation (2), $y_{t+h} \in \mathfrak{R}$ is the scalar-dependent variable; \mathbf{f}_t is the same vector of latent factors as in equation (2); $\mathbf{w}_t \in \mathfrak{R}^K$ is a $K \times 1$ vector of observable variables; $\varepsilon_{t+h} \in \mathfrak{R}$ is the error term; π_y is the break fraction; γ_j and $\boldsymbol{\beta}_j$ are $R \times 1$ and $K \times 1$ vectors of slope coefficients, respectively, for $j = 1, 2$.

The model in equation (1) is a large dimensional factor model with a break in the loadings. Given Assumption C3 in section ‘Consistency’, we follow Chamberlain and Rothschild (1983) and allow for some degree of correlation in the idiosyncratic components on each side of the breakpoint: (i) then is an *approximate breakpoint factor model*; it is more general than an exact breakpoint factor model, which would extend the arbitrage pricing theory of Ross (1976) and would not allow for any correlation in the idiosyncratic components on any side of the breakpoint. Equation (2) is a factor augmented regression with a break in the slope coefficients of \mathbf{f}_t or \mathbf{w}_t (or both). Together with equation (1), equation (2) forms an *unstable diffusion index forecasting model*: it extends the linear set up of Stock and Watson (1998, 2002) and Bai and Ng (2006) by introducing structural instability in the factor model and in the factor augmented regression. Notice that the break fractions π_x and π_y in equations (1) and (2), respectively, are not constrained to be equal.

III. The breakpoint factor model

The paper studies diffusion index forecasting models with instability in the factor model *and* in the factor augmented regression: estimation and model selection in equation (1) is functional to section IV. We adapt the methodology of Massacci (2017) for threshold factor models in sections ‘Assumptions’ and ‘Estimation and model selection’.

Assumptions

Let $\mathbb{1}_{1t}(\pi_x) = \mathbb{1}(t/T \leq \pi_x)$ and $\mathbb{1}_{2t}(\pi_x) = \mathbb{1}(t/T > \pi_x)$. Denote $R^0, \Lambda_j^0 = (\lambda_{j1}^0, \dots, \lambda_{jN}^0)'$, π_x^0 and \mathbf{f}_t^0 the true values of R, Λ_j, π_x and \mathbf{f}_t , respectively, for $j = 1, 2$. Define $\mathbf{f}_{jt}^0(\pi_x) = \mathbb{1}_{jt}(\pi_x)\mathbf{f}_t^0$, for $j = 1, 2$ and $t = 1, \dots, T$, and let $\delta_{xi}^0 = \lambda_{2i}^0 - \lambda_{1i}^0$, for $i = 1, \dots, N$. We collect the assumptions into three sets: Section ‘Identification’ states the assumption needed to identify (1) from a linear factor model; Section ‘Consistency’ lists the assumptions that are sufficient to ensure consistency of the estimator for factors, loadings and break fraction; Section ‘Convergence rates’ states the additional assumption required to obtain the convergence rates of the estimators.

Identification

Assumption I – Breakpoint Factor Model. For $0.5 < \alpha^0 \leq 1, \delta_{xi}^0 \neq 0$ for $i = 1, \dots, [N^{\alpha^0}]$, and $\sum_{i=[N^{\alpha^0}]+1}^N \|\delta_{xi}^0\| = O(1)$.

According to Assumption I, *at least* a fraction $O(N^{\alpha^0})$ of the N series experiences a break, for $0.5 < \alpha^0 \leq 1$. Bates *et al.* (2013) show that if at most $O(N^{0.5})$ series undergo a break, the principal components estimator applied to the misspecified linear model achieves the Bai and Ng (2002) convergence rate $C_{NT} = \min \{ \sqrt{N}, \sqrt{T} \}$: Assumption I ensures that enough series experience a break so that equation (1) is identified from a linear factor model when factors and loadings are estimated by principal components. Assumption I relates to Assumption 1 in Chen *et al.* (2014): at least $O(N^{\alpha^0})$ breaks are *big*, as defined in Assumption 1(a); the remaining breaks are *small*, according to Assumption 1(b) with N and T of the same order.

Consistency

Assumption C1 – Factors. $E\|\mathbf{f}_t^0\|^4 < \infty$; for $j = 1, 2, T^{-1} \sum_{t=1}^T \mathbf{f}_{jt}^0(\pi_x)\mathbf{f}_{jt}^0(\pi_x)' \xrightarrow{P} \Sigma_{jf}^0(\pi_x, \pi_x^0)$ as $T \rightarrow \infty$ for all π_x and some positive definite matrix $\Sigma_{jf}^0(\pi_x, \pi_x^0)$.

Assumption C2 – Factor Loadings. For $j = 1, 2$ and $i = 1, \dots, N, \|\lambda_{ji}^0\| \leq \bar{\lambda} < \infty$, and $\|\Lambda_j^{0'} \Lambda_j^0 / N - \mathbf{D}_{\Lambda_j}^0\| \rightarrow 0$ as $N \rightarrow \infty$ for some $R^0 \times R^0$ positive definite matrix $\mathbf{D}_{\Lambda_j}^0$.

Assumption C3 – Time and Cross-Sectional Dependence and Heteroskedasticity. There exists a positive constant $M_x < \infty$ such that for $j = 1, 2$, for all π_x and for all (N, T) ,

(a) $E(e_{it}) = 0$ and $E|e_{it}|^8 \leq M_x$;

(b) $E[\mathbb{1}_{jt}(\pi_x) \mathbb{1}_{jv}(\pi_x) e_{it} e_{iv}] = \tau_{jiv}(\pi_x)$ with $|\tau_{jiv}(\pi_x)| \leq |\tau_{jiv}|$ for some τ_{jiv} and for all i , and $T^{-1} \sum_{t=1}^T \sum_{v=1}^T |\tau_{jiv}| \leq M_x$;

(c) $E \left[T^{-1} \sum_{t=1}^T \mathbb{1}_{jt}(\pi_x) e_{it} e_{it} \right] = \sigma_{jil}(\pi_x), |\sigma_{jil}(\pi_x)| \leq M_x$ for all l , and $N^{-1} \sum_{i=1}^N \sum_{l=1}^N |\sigma_{jil}(\pi_x)| \leq M_x$;

(d) $E \left| T^{-1/2} \sum_{t=1}^T \mathbb{1}_{jt}(\pi_x) e_{it} e_{it} - E[\mathbb{1}_{jt}(\pi_x) e_{it} e_{it}] \right|^4 \leq M_x$ for every (i, l) .

Assumption C4 – Weak Dependence between \mathbf{f}_t^0 and e_{it} . There exists some positive constant $M_x < \infty$ such that for all π_x and for all (N, T) ,

$$E \left\{ N^{-1} \sum_{i=1}^N \left\| T^{-1/2} \left[\sum_{t=1}^T \mathbb{1}_{jt}(\pi_x) \mathbf{f}_t^0 e_{it} \right] \right\|^2 \right\} \leq M_x, j = 1, 2.$$

Assumptions C1–C4 are the natural extensions of Assumptions A–D imposed on linear factor models in Bai and Ng (2002), and accommodate the presence of the breakpoint. Assumption C1 restricts the sequence $\{\mathbf{f}_t^0\}_{t=1}^T$, so that appropriate second moments exist; it also imposes full rank conditions that exclude multicollinearity in the factors. According to Assumption C2, factor loadings are non-stochastic and factors have a non-negligible effect on the variance of \mathbf{x}_t , on each side of the breakpoint. Under Assumption C3, limited degrees of time-series and cross-sectional dependence in the idiosyncratic components as well as heteroskedasticity are allowed: this makes (1) an approximate breakpoint factor model; in particular, Assumption C3(b) is aligned to Assumption C2 in Bai and Ng (2002). Assumption C4 provides an upper bound to the degree of dependence between the factors and the idiosyncratic components: if \mathbf{f}_t^0 and e_{it} are independent, as in Assumption 5(i) in Barigozzi *et al.* (2018), then Assumption C4 is implied by Assumptions C1 and C3(a). Assumptions C1–C4 (including $E|e_{it}|^8 \leq M_x$ in C3(a)) ensure that π_x^0 can be consistently estimated, so that consistency of the estimators for factors and loadings can also be achieved.

Convergence rates

Assumption CR – Mixing Condition and Moment Bounds: For $i = 1, \dots, N$ and $t = 1, \dots, T$,

(a) $\{\mathbf{f}_t^0, \mathbf{e}_t\}_{t=1}^T$ is ρ -mixing, with ρ -mixing coefficients satisfying $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$;

(b) $E(\|\mathbf{f}_t^0 e_{it}\|^4) \leq C_x$ for some $C_x < \infty$.

Assumption CR is analogous to Assumption 1 in Hansen (2000). Assumption CR(a) allows for non-stationarity and suitably restricts the memory of the process $\{\mathbf{f}_t^0, \mathbf{e}_t\}_{t=1}^T$, and thus of $\{\mathbf{x}_t\}_{t=1}^T$, so that Lemma 3.4 in Peligrad (1982) can be used: structural breaks are allowed, whereas unit roots are ruled out. Assumption CR(b) imposes an unconditional moment bound.

Estimation and model selection

We study estimation and model selection in sections ‘Principal components estimation’ and ‘Selecting the Number of Factors’ respectively.

Principal components estimation

We estimate factors and loadings by principal components, and π_x^0 by concentrated least squares: the latter requires minimizing an objective function that depends only on π_x . Define the $N \times 2R$ matrix of loadings $\Lambda^R = (\Lambda_1^R, \Lambda_2^R)$ and the $R \times T$ matrix of factors $\mathbf{F}^R = (\mathbf{f}_1^R, \dots, \mathbf{f}_T^R)$, where the superscript R denotes dependence on the number of factors. Let the $N \times 2R^0$ matrix $\Lambda^0 = (\Lambda_1^0, \Lambda_2^0)$ and the $R^0 \times T$ matrix $\mathbf{F}^0 = (\mathbf{f}_1^0, \dots, \mathbf{f}_T^0)$ be the true value of Λ^R and \mathbf{F}^R respectively. For given R in equation (1), the objective function in terms of Λ^R , \mathbf{F}^R and π_x is

$$S(\Lambda^R, \mathbf{F}^R, \pi_x) = (NT)^{-1} \sum_{t=1}^T \left\{ [\mathbf{x}_t - \mathbb{1}_{1t}(\pi_x) \Lambda_1^R \mathbf{f}_t^R - \mathbb{1}_{2t}(\pi_x) \Lambda_2^R \mathbf{f}_t^R]' \times [\mathbf{x}_t - \mathbb{1}_{1t}(\pi_x) \Lambda_1^R \mathbf{f}_t^R - \mathbb{1}_{2t}(\pi_x) \Lambda_2^R \mathbf{f}_t^R] \right\}. \tag{3}$$

For given π_x , define

$$\hat{\Sigma}_{jx}(\pi_x) = \left[(NT)^{-1} \sum_{t=1}^T \mathbb{1}_{jt}(\pi_x) \mathbf{x}_t \mathbf{x}_t' \right], j = 1, 2. \tag{4}$$

Given the constraints $N^{-1}(\Lambda_1^R \Lambda_1^R) = N^{-1}(\Lambda_2^R \Lambda_2^R) = \mathbf{I}_R$, from equation (3) the estimator $\hat{\pi}_x^R$ for π_x^0 is

$$\hat{\pi}_x^R = \arg \min_{\pi_x} (NT)^{-1} \sum_{t=1}^T \mathbf{x}_t' \left\{ \mathbf{I}_N - N^{-1} \left[\begin{array}{l} \mathbb{1}_{1t}(\pi_x) \hat{\Lambda}_1^R(\pi_x) \hat{\Lambda}_1^R(\pi_x)' \\ + \mathbb{1}_{2t}(\pi_x) \hat{\Lambda}_2^R(\pi_x) \hat{\Lambda}_2^R(\pi_x)' \end{array} \right] \right\} \mathbf{x}_t, \tag{5}$$

where $\hat{\Lambda}_j^R(\pi_x) = [\hat{\lambda}_{j1}^R(\pi_x), \dots, \hat{\lambda}_{jN}^R(\pi_x)]'$ is the estimator for Λ_j^0 given π_x , for $j = 1, 2$; $\hat{\Lambda}_j^R(\pi_x)$ is \sqrt{N} times the $N \times R$ matrix of eigenvectors of $\hat{\Sigma}_{jx}(\pi_x)$ corresponding to its R largest eigenvalues. The estimator $\hat{\mathbf{F}}^R(\pi_x) = [\hat{\mathbf{f}}_1^R(\pi_x), \dots, \hat{\mathbf{f}}_T^R(\pi_x)]$ for \mathbf{F}^0 for given π_x is

$$\hat{\mathbf{f}}_t^R(\pi_x) = N^{-1} \left[\mathbb{1}_{1t}(\pi_x) \hat{\Lambda}_1^R(\pi_x) + \mathbb{1}_{2t}(\pi_x) \hat{\Lambda}_2^R(\pi_x) \right]' \mathbf{x}_t, t = 1, \dots, T.$$

Given $\hat{\pi}_x^R$, the estimators for Λ^0 and \mathbf{F}^0 are $\hat{\Lambda}^R(\hat{\pi}_x^R) = [\hat{\Lambda}_1^R(\hat{\pi}_x^R), \hat{\Lambda}_2^R(\hat{\pi}_x^R)]$ and $\hat{\mathbf{F}}^R(\hat{\pi}_x^R)$ respectively: if R^0 is known, the estimators for Λ^0 , \mathbf{F}^0 and π_x^0 are $\hat{\Lambda} = \hat{\Lambda}^{R^0}(\hat{\pi}_x^{R^0})$, $\hat{\mathbf{F}} = \hat{\mathbf{F}}^{R^0}(\hat{\pi}_x^{R^0})$ and $\hat{\pi}_x = \hat{\pi}_x^{R^0}$, respectively. If R^0 is unknown we proceed as in section ‘Selecting the Number of Factors’ below.

Selecting the number of factors

There exist several procedures to determine the unknown number of factors R^0 . Bai and Ng (2002), Alessi, Barigozzi and Capasso (2010), Kapetanios (2010), Onatski (2010), Ahn and Horenstein (2013) and Caner and Han (2014) consider the static factor model. Amengual and Watson (2007) look at the restricted dynamic case. Hallin and Liška (2007) and Onatski (2009) focus upon the generalized dynamic factor model. Breitung and Eickmeier (2011)

show that neglecting structural breaks in the loadings leads to overestimation of the number of factors. We robustify Bai and Ng (2002) selection criteria to account for the unknown break fraction.

Let \bar{R} be any *a priori* number of factors $R = \bar{R}$ such that $\bar{R} \geq R^0$. From equation (5), the estimator for π_x^0 is $\hat{\pi}_x^{\bar{R}}$. Given equation (3), the breakpoint robust Bai and Ng (2002) information criteria are

$$\begin{aligned}
 IC_{p1}(R, R) &= \ln S \left[\hat{\Lambda}^R \left(\hat{\pi}_x^{\bar{R}} \right), \hat{\mathbf{F}}^R \left(\hat{\pi}_x^{\bar{R}} \right), \hat{\pi}_x^{\bar{R}} \right] + (R + R) \left(\frac{N + T}{NT} \right) \ln \left(\frac{NT}{N + T} \right), \\
 IC_{p2}(R, R) &= \ln S \left[\hat{\Lambda}^R \left(\hat{\pi}_x^{\bar{R}} \right), \hat{\mathbf{F}}^R \left(\hat{\pi}_x^{\bar{R}} \right), \hat{\pi}_x^{\bar{R}} \right] + (R + R) \left(\frac{N + T}{NT} \right) \ln (C_{NT}^2), \\
 IC_{p3}(R, R) &= \ln S \left[\hat{\Lambda}^R \left(\hat{\pi}_x^{\bar{R}} \right), \hat{\mathbf{F}}^R \left(\hat{\pi}_x^{\bar{R}} \right), \hat{\pi}_x^{\bar{R}} \right] + (R + R) \left[\frac{\ln (C_{NT}^2)}{C_{NT}^2} \right]. \tag{6}
 \end{aligned}$$

The following theorem states the validity of the proposed information criteria.

Theorem 1 Under Assumptions I, C1–C4 and CR, the criteria $IC_{p1}(R, R)$, $IC_{p2}(R, R)$ and $IC_{p3}(R, R)$ defined in equation (6) consistently estimate the number of factors R^0 .

The estimator \hat{R} for R^0 is obtained by minimizing the information criteria in equation (6). Theorem 1 implies that \hat{R} involves a two-step estimation strategy: in the first step, for any *a priori* specified number of factors $R = \bar{R}$ such that $\bar{R} \geq R^0$, the estimator $\hat{\pi}_x^{\bar{R}}$ may be obtained from equation (5); in the second step, the criteria in equation (6) may be computed by plugging $\hat{\Lambda}^R \left(\hat{\pi}_x^{\bar{R}} \right)$, $\hat{\mathbf{F}}^R \left(\hat{\pi}_x^{\bar{R}} \right)$ and $\hat{\pi}_x^{\bar{R}}$ in equation (3). In practice, given a bounded integer $R^{\max} \geq R^0$, one may set $\bar{R} = R^{\max}$ in equation (5).

IV. The unstable factor augmented regression

We study estimation and inference in equation (2) under the assumption that the true number of factors R^0 is known: if not, the selection criteria in equation (6) may be used to estimate R^0 . Define $\theta_l = (\gamma_l', \beta_l')'$, and let $\theta_l^0 = (\gamma_l^{0'}, \beta_l^{0'})'$ and π_y^0 be the true values of θ_l and π_y , respectively, for $l = 1, 2$; given $\theta = (\theta_1', \theta_2')'$, let $\theta^0 = (\theta_1^{0'}, \theta_2^{0'})'$ be the true value of θ . In section ‘Testing for structural change’, we consider least squares estimation of the $[(2R^0 + 2K) + 1] \times 1$ vector of coefficients $(\theta^{0'}, \pi_y^0)'$: this goes beyond Wang *et al.* (2015), who estimate unstable factor augmented regressions when factors are extracted from linear models. In section ‘Testing for structural change’, we propose a test for the null hypothesis of structural stability $\mathbb{H}_0: (\theta_1^0 = \theta_2^0)$ in equation (2): this complements Corradi and Swanson (2014) Hausman-type test, which does not identify the source of instability between factor model and factor augmented regression, but it allows for the multiple break scenario under the alternative.

Estimation and inference

Define $\mathbb{1}_{1t}(\pi_y) = \mathbb{1}(t/T \leq \pi_y)$ and $\mathbb{1}_{2t}(\pi_y) = \mathbb{1}(t/T > \pi_y)$; let $\mathcal{W}_{R^0+K}(\cdot)$ be a $(R^0 + K) \times 1$ vector of standard Brownian motions defined on $[0, 1]$.

Assumption FR. Let $\mathbf{z}_t^0 = (\mathbf{f}_t^{0'}, \mathbf{w}_t')'$. Then:

- (a) $T^{-1} \sum_{t=1}^T \mathbb{1}_{l_t}(\pi_y) \mathbf{z}_t^0 \mathbf{z}_t^{0'} \xrightarrow{P} \pi_y \boldsymbol{\Sigma}_z^0$ for all π_y , where $\boldsymbol{\Sigma}_z^0 = E(\mathbf{z}_t^0 \mathbf{z}_t^{0'})$ is a positive definite matrix;
- (b) For $l = 1, 2, T^{-1} \sum_{t=1}^T \mathbb{1}_{l_t}(\pi_y) \mathbf{z}_t^0 \varepsilon_{t+h} \xrightarrow{P} \mathbf{0}$ for all π_y and for any $h \geq 0$;
- (c) $\{\mathbf{z}_t^0, \varepsilon_{t+h}\}_{t=1}^T$ is ρ -mixing, with ρ -mixing coefficients satisfying $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$, for any $h \geq 0$;
- (d) $E(\|\mathbf{z}_t^0 \varepsilon_{t+h}\|^4) \leq C_y$ and $E(\|\mathbf{z}_t^0\|^4) \leq C_y$ for some $C_y < \infty$, for $t = 1, \dots, T$;
- (e) $T^{-1/2} \sum_{t=1}^T \mathbb{1}_{l_t}(\pi_y) \mathbb{1}_{l_t}(\pi_x) \mathbf{z}_t^0 \varepsilon_{t+h} \Rightarrow (\pi_x^0 \boldsymbol{\Sigma}_{z\varepsilon}^0)^{1/2} \mathcal{W}_{R^0+K}(\pi_y)$ for all π_y , and any $h \geq 0$, where $\boldsymbol{\Sigma}_{z\varepsilon}^0$ is a positive definite matrix defined as $\boldsymbol{\Sigma}_{z\varepsilon}^0 = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \sum_{v=1}^T E(\varepsilon_{t+h} \varepsilon_{v+h} \mathbf{z}_t^0 \mathbf{z}_v^{0'})$.

Assumption FR extends Assumption Y1 in Stock and Watson (2002), and Assumption E in Bai and Ng (2006). Assumption FR(a) is a full rank condition. Assumption FR(b) is required to obtain consistent estimators for θ_1^0 and θ_2^0 . Assumptions FR(c) and FR(d) are equivalent to Assumptions CR(a) and CR(b), respectively, as discussed in section ‘Convergence rates’. Assumption FR(e) is needed to derive the asymptotic distribution of the estimators for θ_1^0 and θ_2^0 and of the test statistic for the null hypothesis of stability: it allows for serial correlation and heteroskedasticity in the sequence of error terms $\{\varepsilon_{t+h}\}_{t=1}^T$.

The feasible objective function is the sum of squared residuals (divided by T)

$$L(\hat{\mathbf{F}}, \boldsymbol{\theta}, \pi_y) = T^{-1} \sum_{t=1}^T \left[y_{t+h} - \mathbb{1}_{l_t}(\pi_y) (\gamma_1' \hat{\mathbf{f}}_t + \boldsymbol{\beta}_1' \mathbf{w}_t) - \mathbb{1}_{2t}(\pi_y) (\gamma_2' \hat{\mathbf{f}}_t + \boldsymbol{\beta}_2' \mathbf{w}_t) \right]^2.$$

Let $\hat{\mathbf{z}}_t = (\hat{\mathbf{f}}_t', \mathbf{w}_t')'$. For given π_y , the least squares estimator for θ^0 is $\hat{\boldsymbol{\theta}}(\pi_y) = [\hat{\boldsymbol{\theta}}_1(\pi_y)', \hat{\boldsymbol{\theta}}_2(\pi_y)']'$, where

$$\hat{\boldsymbol{\theta}}_l(\pi_y) = [\hat{\gamma}_l(\pi_y)', \hat{\boldsymbol{\beta}}_l(\pi_y)']' = \left[\sum_{t=1}^T \mathbb{1}_{l_t}(\pi_y) \hat{\mathbf{z}}_t \hat{\mathbf{z}}_t' \right]^{-1} \left[\sum_{t=1}^T \mathbb{1}_{l_t}(\pi_y) \hat{\mathbf{z}}_t y_{t+h} \right], \quad l = 1, 2,$$

is the estimator for θ_l^0 , and $\hat{\gamma}_l(\pi_y)$ and $\hat{\boldsymbol{\beta}}_l(\pi_y)$ are the estimators for γ_l^0 and $\boldsymbol{\beta}_l^0$ respectively: the concentrated loss function is

$$L_\theta(\hat{\mathbf{F}}, \pi_y) = T^{-1} \sum_{t=1}^T \left[y_{t+h} - \mathbb{1}_{l_t}(\pi_y) \hat{\boldsymbol{\theta}}_1(\pi_y)' \hat{\mathbf{z}}_t - \mathbb{1}_{2t}(\pi_y) \hat{\boldsymbol{\theta}}_2(\pi_y)' \hat{\mathbf{z}}_t \right]^2.$$

The estimator $\hat{\pi}_y$ and $\hat{\boldsymbol{\theta}}$ for π_y^0 and $\boldsymbol{\theta}^0$, respectively, then are

$$\hat{\pi}_y = \arg \min_{\pi_y} L_\theta(\hat{\mathbf{F}}, \pi_y), \quad \hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\hat{\pi}_y) = [\hat{\boldsymbol{\theta}}_1(\hat{\pi}_y)', \hat{\boldsymbol{\theta}}_2(\hat{\pi}_y)']'$$

it follows that $\hat{\boldsymbol{\theta}}_l = \hat{\boldsymbol{\theta}}_l(\hat{\pi}_y) = [\hat{\gamma}_l(\hat{\pi}_y)', \hat{\boldsymbol{\beta}}_l(\hat{\pi}_y)']'$ is the estimator for θ_l^0 , and $\hat{\gamma}_l = \hat{\gamma}_l(\hat{\pi}_y)$ and $\hat{\boldsymbol{\beta}}_l = \hat{\boldsymbol{\beta}}_l(\hat{\pi}_y)$ are the estimators for γ_l^0 and $\boldsymbol{\beta}_l^0$, respectively, for $l = 1, 2$.

Consistency

Define the $R^0 \times T$ matrices of state-specific factors $\mathbf{F}_j^0(\pi_x) = [\mathbf{f}_{j1}^0(\pi_x), \dots, \mathbf{f}_{jT}^0(\pi_x)]$, for $j = 1, 2$, such that $\mathbf{F}_1^0(\pi_x) + \mathbf{F}_2^0(\pi_x) = (\mathbf{f}_1^0, \dots, \mathbf{f}_T^0) = \mathbf{F}^0$ and $\mathbf{F}_1^0(\pi_x)\mathbf{F}_2^0(\pi_x)' = \mathbf{0}_{R^0}$. Let $\hat{\mathbf{H}}_{jj}(\pi_x)$ be the rotation matrix

$$\hat{\mathbf{H}}_{jj}(\pi_x) = \frac{\mathbf{F}_j^0(\pi_x^0)\mathbf{F}_j^0(\pi_x)'}{T} \frac{\Lambda_j^{0'} \hat{\Lambda}_j(\pi_x)}{N} \hat{\mathbf{V}}_j(\pi_x)^{-1}, \quad j = 1, 2, \tag{7}$$

where $\hat{\mathbf{V}}_j(\pi_x)$ is the $R^0 \times R^0$ diagonal matrix of the first R^0 largest eigenvalues of $\hat{\Sigma}_{jx}(\pi_x)$ in equation (4) in decreasing order. Let $\hat{\Phi}_{jj}(\pi_x^0) = \text{diag}[\hat{\mathbf{H}}_{jj}(\pi_x^0), \mathbf{I}_K]$ be a $(R^0 + K) \times (R^0 + K)$ block diagonal matrix, for $j = 1, 2$. Define the $(R^0 + K) \times (R^0 + K)$ block diagonal matrices $\hat{\Phi}_{12}(\pi_x^0, \pi_y^0)$ and $\hat{\Phi}_{21}(\pi_x^0, \pi_y^0)$ as

$$\hat{\Phi}_{12}(\pi_x^0, \pi_y^0) = \frac{\min\{\pi_x^0, \pi_y^0\}}{\pi_y^0} \hat{\Phi}_{11}(\pi_x^0) + \mathbb{1}(\pi_x^0 < \pi_y^0) \frac{\pi_y^0 - \pi_x^0}{\pi_y^0} \hat{\Phi}_{22}(\pi_x^0)$$

and

$$\hat{\Phi}_{21}(\pi_x^0, \pi_y^0) = \frac{1 - \max\{\pi_x^0, \pi_y^0\}}{1 - \pi_y^0} \hat{\Phi}_{22}(\pi_x^0) + \mathbb{1}(\pi_x^0 > \pi_y^0) \frac{\pi_x^0 - \pi_y^0}{1 - \pi_y^0} \hat{\Phi}_{11}(\pi_x^0),$$

respectively. The following theorem states the consistency of $\hat{\pi}_y$, $\hat{\theta}_1$ and $\hat{\theta}_2$ as estimators for π_y^0 , θ_1^0 and θ_2^0 respectively.

Theorem 2 Under Assumptions I, C1–C4, FR(a) and FR(b), $\hat{\pi}_y \xrightarrow{P} \pi_y^0$, $\hat{\theta}_1 - \hat{\Phi}_{12}(\pi_x^0, \pi_y^0)' \theta_1^0 \xrightarrow{P} 0$ and $\hat{\theta}_2 - \hat{\Phi}_{21}(\pi_x^0, \pi_y^0)' \theta_2^0 \xrightarrow{P} 0$, as $N, T \rightarrow \infty$.

Theorem 2 extends theorem 2 in Stock and Watson (2002). The factor augmented regression in equation (2) is identified up to a rotation because the latent factors satisfy

$$\begin{aligned} \mathbb{1}_{1t}(\pi_y^0) \gamma_1^{0'} \mathbf{f}_t^0 + \mathbb{1}_{2t}(\pi_y^0) \gamma_2^{0'} \mathbf{f}_t^0 &= \mathbb{1}_{1t}(\pi_x^0) \mathbb{1}_{1t}(\pi_y^0) \gamma_1^{0'} \mathbf{L}_{11} \mathbf{L}_{11}^{-1} \mathbf{f}_t^0 + \mathbb{1}_{2t}(\pi_x^0) \mathbb{1}_{1t}(\pi_y^0) \gamma_1^{0'} \mathbf{L}_{21} \mathbf{L}_{21}^{-1} \mathbf{f}_t^0 \\ &\quad + \mathbb{1}_{1t}(\pi_x^0) \mathbb{1}_{2t}(\pi_y^0) \gamma_2^{0'} \mathbf{L}_{12} \mathbf{L}_{12}^{-1} \mathbf{f}_t^0 + \mathbb{1}_{2t}(\pi_x^0) \mathbb{1}_{2t}(\pi_y^0) \gamma_2^{0'} \mathbf{L}_{22} \mathbf{L}_{22}^{-1} \mathbf{f}_t^0, \end{aligned}$$

for some positive definite matrices \mathbf{L}_{11} , \mathbf{L}_{21} , \mathbf{L}_{12} and \mathbf{L}_{22} . Theorem 2 relates to the difference between $\hat{\theta}_l$ and the space spanned by θ_l^0 , for $l = 1, 2$. The rotation induced around θ_l^0 is a convex linear combination of the rotations induced by $\hat{\Phi}_{11}(\pi_x^0)$ and $\hat{\Phi}_{22}(\pi_x^0)$, which depend on the matrices $\hat{\mathbf{H}}_{11}(\pi_x)$ and $\hat{\mathbf{H}}_{22}(\pi_x)$, respectively, as defined in equation (7); these arise from the rotational indeterminacy stemming from the underlying breakpoint factor model in (1). The rotation around θ_l^0 depends on the relative position of the break fractions π_x^0 and π_y^0 , for $l = 1, 2$. Let us consider $\hat{\theta}_1$ (analogous arguments apply to $\hat{\theta}_2$). If $\pi_x^0 < \pi_y^0$, the break in the factor loadings occurs before the break in the factor augmented regression: the rotation induced around θ_1^0 is affected by the rotations around Λ_1^0 and Λ_2^0 , with weights equal to π_x^0/π_y^0 and $(\pi_y^0 - \pi_x^0)/\pi_y^0$ respectively. If $\pi_x^0 \geq \pi_y^0$, the rotation around θ_1^0 only depends on the rotation around Λ_1^0 .

Rates of convergence

The following theorem states the convergence rate of the concentrated least squares estimator $\hat{\pi}_y$ for the break fraction π_y^0 .

Theorem 3 Under Assumptions I, C1–C4 and FR(a)–FR(d),

$$T (\hat{\pi}_y - \pi_y^0) = O_p(1).$$

The T convergence rate relates to the estimator for the break fraction $\hat{\pi}_y$: Theorem 3 implies that with high probability the difference between the estimator $\hat{T}_y = [T \hat{\pi}_y]$ for the break date and its true value $T_y^0 = [T \pi_y^0]$ is bounded, namely $(\hat{T}_y - T_y^0) = O_p(1)$ (Bai and Perron, 1998). Let $\hat{\mathbf{k}}_{lt}(\pi_y) = \mathbb{1}_l(\pi_y) \hat{\varepsilon}_{t+h}(\pi_y) \hat{\mathbf{z}}_t$, for $l = 1, 2$, where $\hat{\varepsilon}_{t+h}(\pi_y) = y_{t+h} - \mathbb{1}_{1t}(\pi_y) \hat{\boldsymbol{\theta}}_1(\pi_y)' \hat{\mathbf{z}}_t - \mathbb{1}_{2t}(\pi_y) \hat{\boldsymbol{\theta}}_2(\pi_y)' \hat{\mathbf{z}}_t$. From Newey and West (1987), define: $\hat{\mathbf{K}}_{ld}(\pi_y) = T^{-1} \sum_{t=d+1}^T \hat{\mathbf{k}}_{lt}(\pi_y) \hat{\mathbf{k}}_{l,t-d}(\pi_y)'$, for $d = 0, \dots, D_T$, with $D_T = o(T^{1/4})$; $\hat{\boldsymbol{\Omega}}_l(\pi_y) = \hat{\mathbf{K}}_{l0}(\pi_y) + \sum_{d=1}^{D_T} w(d, D_T) [\hat{\mathbf{K}}_{ld}(\pi_y) + \hat{\mathbf{K}}_{ld}(\pi_y)']$, where $w(d, D_T) = [1 - d/(D_T + 1)]$ is the Bartlett kernel. Theorem 3 feeds into the following theorem, which states the asymptotic distribution of the estimators for $\boldsymbol{\theta}_1^0$ and $\boldsymbol{\theta}_2^0$.

Theorem 4 Under Assumptions I, C1–C4, CR and FR, if $\sqrt{T}/N \rightarrow 0$ then

$$\sqrt{T} [\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\Phi}}_{12} (\pi_x^0, \pi_y^0)' \boldsymbol{\theta}_1^0] \xrightarrow{p} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}_1}^0)$$

and

$$\sqrt{T} [\hat{\boldsymbol{\theta}}_2 - \hat{\boldsymbol{\Phi}}_{21} (\pi_x^0, \pi_y^0)' \boldsymbol{\theta}_2^0] \xrightarrow{p} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}_2}^0),$$

with

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}_1}^0 = \frac{1}{\pi_y^0} \left[\begin{array}{c} \frac{\min\{\pi_x^0, \pi_y^0\}}{\pi_y^0} \boldsymbol{\Phi}_{11}^0 (\pi_x^0)' (\boldsymbol{\Sigma}_z^0)^{-1} \boldsymbol{\Sigma}_{z\varepsilon}^0 (\boldsymbol{\Sigma}_z^0)^{-1} \boldsymbol{\Phi}_{11}^0 (\pi_x^0) \\ + \mathbb{1} (\pi_x^0 < \pi_y^0) \frac{\pi_x^0 - \pi_y^0}{\pi_y^0} \boldsymbol{\Phi}_{22}^0 (\pi_x^0)' (\boldsymbol{\Sigma}_z^0)^{-1} \boldsymbol{\Sigma}_{z\varepsilon}^0 (\boldsymbol{\Sigma}_z^0)^{-1} \boldsymbol{\Phi}_{22}^0 (\pi_x^0) \end{array} \right]$$

and

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}_2}^0 = \frac{1}{1 - \pi_y^0} \left[\begin{array}{c} \frac{1 - \max\{\pi_x^0, \pi_y^0\}}{1 - \pi_y^0} \boldsymbol{\Phi}_{22}^0 (\pi_x^0)' (\boldsymbol{\Sigma}_z^0)^{-1} \boldsymbol{\Sigma}_{z\varepsilon}^0 (\boldsymbol{\Sigma}_z^0)^{-1} \boldsymbol{\Phi}_{22}^0 (\pi_x^0) \\ + \mathbb{1} (\pi_x^0 > \pi_y^0) \frac{\pi_x^0 - \pi_y^0}{1 - \pi_y^0} \boldsymbol{\Phi}_{11}^0 (\pi_x^0)' (\boldsymbol{\Sigma}_z^0)^{-1} \boldsymbol{\Sigma}_{z\varepsilon}^0 (\boldsymbol{\Sigma}_z^0)^{-1} \boldsymbol{\Phi}_{11}^0 (\pi_x^0) \end{array} \right],$$

where $\boldsymbol{\Phi}_{jj}^0 (\pi_x^0) = \text{diag} [\boldsymbol{\Sigma}_{j\mathbf{f}}^0 (\pi_x^0, \pi_x^0) \mathbf{Q}_{\lambda_j}^0 (\pi_x^0) \mathbf{V}_j^0 (\pi_x^0)^{-1}, \mathbf{I}_K]$, $\mathbf{Q}_{\lambda_j}^0 (\pi_x^0) = \text{p lim} [\boldsymbol{\Lambda}_j^{0'} \hat{\boldsymbol{\Lambda}}_j (\pi_x^0)]/N$ and $\mathbf{V}_j^0 (\pi_x^0) = \text{p lim} \hat{\mathbf{V}}_j (\pi_x^0)$. Consistent estimators for $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}_1}^0$ and $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}_2}^0$, denoted by $\widehat{\text{Avar}}(\hat{\boldsymbol{\theta}}_1)$ and $\widehat{\text{Avar}}(\hat{\boldsymbol{\theta}}_2)$, respectively, are

$$\widehat{\text{Avar}}(\hat{\boldsymbol{\theta}}_l) = \left[T^{-1} \sum_{t=1}^T \mathbb{1}_l(\hat{\pi}_y) \hat{\mathbf{z}}_t \hat{\mathbf{z}}_t' \right]^{-1} \hat{\boldsymbol{\Omega}}_l(\hat{\pi}_y) \left[T^{-1} \sum_{t=1}^T \mathbb{1}_l(\hat{\pi}_y) \hat{\mathbf{z}}_t \hat{\mathbf{z}}_t' \right]^{-1}, \quad l = 1, 2. \quad (8)$$

Theorem 4 extends theorem 1 in Bai and Ng (2006). It establishes convergence rate and limiting distribution of the least squares estimator $\hat{\theta}_l$ for θ_l^0 , for $l = 1, 2$. The estimated covariance matrix in (8) is robust to heteroskedasticity and autocorrelation. Let $\hat{\sigma}_\varepsilon^2 = T^{-1} \sum_{t=1}^T \hat{\varepsilon}_{t+h}^2$, with $\hat{\varepsilon}_{t+h} = \hat{\varepsilon}_{t+h}(\hat{\pi}_y)$. With homoskedastic and uncorrelated disturbances the estimator in (8) simplifies to

$$\widehat{\text{Avar}}(\hat{\theta}_l) = \hat{\sigma}_\varepsilon^2 \left[T^{-1} \sum_{t=1}^T \mathbb{1}_l(\hat{\pi}_y) \hat{\mathbf{z}}_t \hat{\mathbf{z}}_t' \right]^{-1}, \quad l = 1, 2.$$

Testing for structural change

Testing strategy

We now propose a test for the null hypothesis $\mathbb{H}_0 : (\theta_1^0 = \theta_2^0)$ in equation (2): this extends the literature on testing for a single break to allow for estimated factors (Andrews, 1993; Perron, 2006). We build a Lagrange multiplier statistic. Under the null hypothesis of linearity, the true factor augmented model becomes $y_{t+h} = \theta_1^0 \mathbf{z}_t^0 + \varepsilon_{t+h}$: the least squares estimator for θ_1^0 is $\hat{\theta}_1^{\mathbb{H}_0} = \left(\sum_{t=1}^T \hat{\mathbf{z}}_t \hat{\mathbf{z}}_t' \right)^{-1} \left(\sum_{t=1}^T \hat{\mathbf{z}}_t y_{t+h} \right)$. Let $\hat{\mathbf{k}}_t^{\mathbb{H}_0} = \hat{\varepsilon}_{t+h}^{\mathbb{H}_0} \hat{\mathbf{z}}_t$, where $\hat{\varepsilon}_{t+h}^{\mathbb{H}_0} = y_{t+h} - \hat{\theta}_1^{\mathbb{H}_0} \hat{\mathbf{z}}_t$. From Newey and West (1987), define: $\hat{\mathbf{K}}_d^{\mathbb{H}_0} = T^{-1} \sum_{t=d+1}^T \hat{\mathbf{k}}_t^{\mathbb{H}_0} \hat{\mathbf{k}}_{t-d}^{\mathbb{H}_0}$, for $d = 0, \dots, D_T$, with D_T as in section ‘Rates of convergence’; $\hat{\mathbf{\Omega}}^{\mathbb{H}_0} = \hat{\mathbf{K}}_0^{\mathbb{H}_0} + \sum_{d=1}^{D_T} w(d, D_T) \left(\hat{\mathbf{K}}_d^{\mathbb{H}_0} + \hat{\mathbf{K}}_d^{\mathbb{H}_0'} \right)$, where $w(d, D_T)$ is the Bartlett kernel as in section ‘Rates of convergence’. Following Andrews (1993), the heteroskedasticity and autocorrelation robust Lagrange multiplier test statistic is defined as

$$\mathcal{LM}(\pi_y) = \frac{1}{\pi_y (1 - \pi_y)} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbb{1}_l(\pi_y) \hat{\mathbf{z}}_t \hat{\varepsilon}_{t+h}^{\mathbb{H}_0} \right]' \left(\hat{\mathbf{\Omega}}^{\mathbb{H}_0} \right)^{-1} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbb{1}_l(\pi_y) \hat{\mathbf{z}}_t \hat{\varepsilon}_{t+h}^{\mathbb{H}_0} \right].$$

For known $\pi_y = \pi_y^0$ and under the null hypothesis, $\mathcal{LM}(\pi_y^0)$ has a χ^2 limiting distribution with $(R^0 + K)$ degrees of freedom as $N, T \rightarrow \infty$. However, π_y^0 is generally unknown and not identified under the null hypothesis. As in Andrews (1993), we propose the statistic

$$\sup \mathcal{LM} = \sup_{\pi_y \in [\underline{\pi}_y, 1 - \underline{\pi}_y]} \mathcal{LM}(\pi_y),$$

with $\underline{\pi}_y \in (0.00, 0.50]$: such a definition of π_y ensures the supremum is not evaluated over an empty space. The theorem below states the asymptotic distribution of $\sup \mathcal{LM}$.

Theorem 5 Let Assumptions I, C1–C4, CR, FR(a), FR(b) and FR(e) hold. Further assume that $\sqrt{T}/N \rightarrow 0$ as $N, T \rightarrow \infty$. Then for any $\underline{\pi}_y \in (0.00, 0.50]$, and under the null hypothesis $\mathbb{H}_0 : (\theta_1^0 = \theta_2^0)$,

$$\sup \mathcal{LM} \xrightarrow{P} \sup_{\pi_y \in [\underline{\pi}_y, 1 - \underline{\pi}_y]} \frac{[\mathcal{W}_{R^0+K}(\pi_y) - \pi_y \mathcal{W}_{R^0+K}(1)]' [\mathcal{W}_{R^0+K}(\pi_y) - \pi_y \mathcal{W}_{R^0+K}(1)]}{\pi_y (1 - \pi_y)}$$

as $N, T \rightarrow \infty$, provided that $\hat{\Omega}^{\mathbb{H}_0} \xrightarrow{p} \Omega^{0, \mathbb{H}_0}$ under \mathbb{H}_0 , where

$$\begin{aligned} \Omega^{0, \mathbb{H}_0} &= \left\{ (\pi_x^0)^{1/2} [\Phi_{11}^0(\pi_x^0)]^{-1} + (1 - \pi_x^0)^{1/2} [\Phi_{22}^0(\pi_x^0)]^{-1} \right\} \Sigma_{z\varepsilon}^0 \\ &\quad \times \left\{ (\pi_x^0)^{1/2} [\Phi_{11}^0(\pi_x^0)]^{-1} + (1 - \pi_x^0)^{1/2} [\Phi_{22}^0(\pi_x^0)]^{-1} \right\}' \end{aligned}$$

Theorem 5 extends theorem 3 (b) in Andrews (1993) and implies that the critical values in Andrews (1993) apply to the sup \mathcal{LM} statistic; these values are used also in Breitung and Eickmeier (2011) and Chen *et al.* (2014). The convergence in probability of $\hat{\Omega}^{\mathbb{H}_0}$ to Ω^{0, \mathbb{H}_0} is not stringent: factors are consistently estimated; $\hat{\Omega}^{\mathbb{H}_0}$ is a heteroskedasticity and autocorrelation robust estimator for Ω^{0, \mathbb{H}_0} . With homoskedastic and serially uncorrelated disturbances, $\hat{\Omega}^{\mathbb{H}_0}$ simplifies to $\hat{\Omega}^{\mathbb{H}_0} = (\hat{\sigma}_\varepsilon^{\mathbb{H}_0})^2 [T^{-1} \sum_{t=1}^T \mathbb{1}_{it}(\hat{\pi}_y) \hat{z}_t \hat{z}_t']$, with $(\hat{\sigma}_\varepsilon^{\mathbb{H}_0})^2 = T^{-1} \sum_{t=1}^T (\hat{\varepsilon}_{t+h}^{\mathbb{H}_0})^2$.

By Theorem 2, under $\mathbb{H}_0 : (\theta_1^0 = \theta_2^0)$ it follows that $\hat{\theta}_1 - \hat{\Phi}_{12}(\pi_x^0, \pi_y^0)' \theta_1^0 \xrightarrow{p} 0$ and $\hat{\theta}_2 - \hat{\Phi}_{21}(\pi_x^0, \pi_y^0)' \theta_1^0 \xrightarrow{p} 0$: under the null hypothesis, $\hat{\theta}_1$ and $\hat{\theta}_2$ have different probability limits due to the different rotations around θ_1^0 induced by the structural break in the underlying factor model. However, this feature does not cause any size distortions, as it is fully captured by the asymptotic covariance matrix Ω^{0, \mathbb{H}_0} defined in Theorem 5: the sup \mathcal{LM} statistic thus has the same asymptotic distribution as the one stated in theorem 3(b) in Andrews (1993).

Robustness to factor model representation

The test in section ‘Testing strategy’ requires estimating the factors from the breakpoint model in (1). A factor model with structural instability has an equivalent linear representation (Breitung and Eickmeier, 2011), which one could use to test for stability in the factor augmented regression. Formally, let $\mathbf{f}_t^* = \mathbf{f}_t^*(\pi_x, \mathbf{f}_t) = [\mathbb{1}_{1t}(\pi_x) \mathbf{f}_t', \mathbb{1}_{2t}(\pi_x) \mathbf{f}_t']'$: the equivalent representation for the model in (1) is $\mathbf{x}_t = \Lambda \mathbf{f}_t^* + \mathbf{e}_t$. Let $\mathbf{f}_t^{0*} = \mathbf{f}_t^*(\pi_x^0, \mathbf{f}_t^0) = [\mathbb{1}_{1t}(\pi_x^0) \mathbf{f}_t^{0\prime}, \mathbb{1}_{2t}(\pi_x^0) \mathbf{f}_t^{0\prime}]'$ be the true value of \mathbf{f}_t^* : by Proposition 1 in Chen *et al.* (2014), the principal components estimator $\hat{\mathbf{f}}_t^*$ for the $(R^0 + R^0) \times 1$ vector \mathbf{f}_t^{0*} is consistent up to a $(R^0 + R^0) \times (R^0 + R^0)$ rotation matrix and has convergence rate $C_{NT} = \min \{ \sqrt{N}, \sqrt{T} \}$. A test for structural change in the factor augmented regression based on the equivalent linear factor representation imposes $(R^0 + R^0 + K) \times 1$ restrictions as $N, T \rightarrow \infty$. The test in section ‘Testing strategy’ has $(R^0 + K) \times 1$ restrictions. Estimating factors from the equivalent linear representation requires imposing a higher number of restrictions and results in a test with lower power.

V. Monte Carlo analysis

In section A of the Online Appendix, we conduct a comprehensive Monte Carlo analysis to assess the finite sample properties of our methodology: we consider model selection in the breakpoint factor model; we study estimation, inference and stability testing in the

factor augmented regression. The results show the good finite sample properties of our methodology and its potential usefulness for applied work.

VI. Application to bond risk premia

Section ‘Literature and contribution’ links the paper to the literature. Section ‘Data and implementation’ describes the data. Section ‘Breakpoint factor model: Macro factors’ details the factor model. Section ‘Unstable factor augmented regression: Bond risk premia’ presents the results from the factor augmented regression.

Literature and contribution

We apply our methodology to uncover potential structural instabilities in the empirical linkages between bond market risk premia and macroeconomic fundamentals. Fama and Bliss (1987) show that the one-year excess return on the n -year bond is predicted by the spread between the n -year forward rate and the one-year yield. Campbell and Shiller (1991) find similar results using yield spreads as predictors for yield changes. Fama and Bliss (1987), and Campbell and Shiller (1991), uncover maturity-specific predictors. Cochrane and Piazzesi (2005) show that a single common factor obtained as a linear combination of forward spreads predicts the one-year excess return on the n -year bond: they thus document that a common factor predicts the excess returns of all bonds. Ludvigson and Ng (2009) find that common factors extracted from a large set of macroeconomic series have predictive power for the one-year excess return on the n -year bond beyond the Cochrane and Piazzesi (2005) single factor.

We take Ludvigson and Ng (2009) as a starting point and depart from them in two ways: we extract the latent factors from a large number of macroeconomic variables under the maintained assumption of structural instability in the loadings; we allow for a structural break in the factor augmented pricing equation for bond excess returns. Existing empirical evidence suggests that the degree of cross-sectional dependence among macroeconomic series is not stable over time (Breitung and Eickmeier, 2011; Chen *et al.*, 2014; Cheng *et al.*, 2016): this motivates the inclusion of a structural break in the factor model. Allowing for a potentially unstable factor augmented regression is also desirable: bond risk premia are highly persistent and require long time series of data; the longer the data span, the higher the likelihood of model instability. Bikbov and Chernov (2010) conduct their analysis over 1970–2002: they raise concerns about structural stability due to the 1979–82 monetary policy experiment. In studying 1960–2006, Smith and Taylor (2009) acknowledge the presence of a break in the early 1980s and *a priori* split their sample around that period. Joslin *et al.* (2014) focus on 1985–2007: they estimate a two-state Markov-switching affine term structure model over 1971:11–2007:12 to show that the period 1985–2007 is well approximated by a single regime. We complement Joslin *et al.* (2014) by applying to bond risk premia the unstable diffusion index forecasting model proposed in this paper: Markov-switching models assumes that ‘history repeats’ (Timmermann, 2008); our unstable model does not. Gürkaynak and Wright (2012) discuss the consequences of structural instability on the learning process of investors.

Data and implementation

We study the period 1965:01–2007:12, which extends the samples studied in Fama and Bliss (1987), Cochrane and Piazzesi (2005) and Ludvigson and Ng (2009): the endpoint of the sample coincides with the end of the Great Moderation and it is the same as in Joslin *et al.* (2014). Our chosen period allows us to avoid the zero lower bound regime started in December 2008, during which the Federal Reserve has relied on unconventional monetary policy tools: an assessment of the impact of these tools on the yield curve has been a challenging task (Wu and Xia, 2016 and references therein). Monthly bond return data are from the Fama–Bliss data set available from the Center for Research in Securities Prices (CRSP): the data set has prices for zero-coupon US Treasury bonds for maturities between one to five years; yields, forward rates and returns are constructed as detailed in Cochrane and Piazzesi (2005).

We estimate the factors from the FRED-MD balanced panel of monthly macroeconomic series described in McCracken and Ng (2016).¹ The data belong to eight groups: (i) output and income; (ii) labour market; (iii) consumption and housing; (iv) orders and inventories; (v) money and credit; (vi) interest rate and exchange rates; (vii) prices; (viii) stock market. We use the 2016:06 vintage. We apply to the original data set described in McCracken and Ng (2016) the changes outlined in the document ‘Changes to FRED-MD’.² We make the resulting 123 series stationary by applying the transformations provided in McCracken and Ng (2016).³

We formally implement the factor augmented pricing equation

$$rx_{t+12}^{(n)} = \mathbb{1}(t/T \leq \pi_y^{(n)}) \gamma_1^{(n)'} \mathbf{f}_t + \mathbb{1}(t/T > \pi_y^{(n)}) \gamma_2^{(n)'} \mathbf{f}_t + \beta_1^{(n)} + \beta_2^{(n)} CP_t + \varepsilon_{t+12}^{(n)}, \quad (9)$$

for $t = 1964:01, \dots, 2006:12$ and $n = 2, 3, 4, 5$: the superscript (n) denotes dependence on the maturity of the bond. The variable $rx_{t+12}^{(n)}$ is the excess log return for the n -year discount bond defined as in Cochrane and Piazzesi (2005). The vector \mathbf{f}_t contains R latent factors: based on the Monte Carlo results, we obtain the estimate \hat{R} for R^0 using the criterion $IC_{p_2}(R, R)$ in equation (6). The observable covariate CP_t is the single forward Cochrane and Piazzesi (2005) factor (i.e. a linear combination of forward rates), estimated over the period 1964:01–2006:12. We estimate the true break fraction $\pi_y^{(n),0}$, $R^0 \times 1$ vectors $\gamma_1^{(n),0}$ and $\gamma_2^{(n),0}$, intercept $\beta_1^{(n),0}$ and slope coefficient $\beta_2^{(n),0}$ as detailed in section ‘The unstable factor augmented regression’: confidence intervals and stability testing are performed using the Newey and West (1987) correction with 18 lags to account for the MA(12) structure due to the overlapping data (Cochrane and Piazzesi, 2005; Ludvigson and Ng, 2009).

We assume that $\beta_1^{(n),0}$ and $\beta_2^{(n),0}$ are not subject to structural instability: the information content of forward rates is stable over time (Cochrane and Piazzesi, 2005). The CP_t factor is a linear combination of forward rates, and it is obtained by assuming that returns at all maturities are linear functions of forward rates: this implies that $\beta_2^{(n),0}$ is stable over time. If we were to allow for a break in $\beta_2^{(n),0}$, we would have to extend the procedure to estimate

¹The data set is available at <https://research.stlouisfed.org/econ/mccracken/fred-databases/>.

²The document is available at <https://research.stlouisfed.org/econ/mccracken/fred-databases/>.

³A file containing all macroeconomic series used to estimate the factors is available upon request.

TABLE 1
Empirical results, bond excess returns, summary statistics, 1965:01–2007:12

	$n=2$	$n=3$	$n=4$	$n=5$
Panel A: Levels				
Mean (rx_t^n)	0.004	0.007	0.009	0.009
Std. Dev. (rx_t^n)	0.019	0.034	0.047	0.058
Corr (rx_t^n, rx_{t-1}^n)	0.931	0.933	0.932	0.922
Corr (rx_t^n, rx_{t-12}^n)	0.215	0.147	0.109	0.071
Panel B: Absolute values				
Mean ($ rx_t^n $)	0.015	0.027	0.037	0.045
Std. Dev. ($ rx_t^n $)	0.012	0.022	0.031	0.037
Corr ($ rx_t^n , rx_{t-1}^n $)	0.857	0.862	0.856	0.836
Corr ($ rx_t^n , rx_{t-12}^n $)	0.070	0.129	0.150	0.157

Notes: This table presents summary statistics for yearly bond excess returns expressed in levels (Panel A) and absolute values (Panel B) computed over the maturities $n=2, 3, 4, 5$. The data sample ranges from January 1965 to December 2007, for a sample size of 516 monthly observations. Mean and standard deviation are displayed in decimals per annum (i.e. 0.01 equals 1 percentage point).

the Cochrane and Piazzesi (2005) factor accordingly: this goes beyond the purpose of the paper. The stability of $\beta_2^{(n),0}$ across all maturities is supported by the results in section ‘Unstable factor augmented regression: Bond risk premia’ below. The null hypothesis of stability then is $\mathbb{H}_0 : (\gamma_1^{(n),0} = \gamma_2^{(n),0})$. Under the alternative hypothesis $\mathbb{H}_1 : (\gamma_1^{(n),0} \neq \gamma_2^{(n),0})$, instabilities arise in the linkages between bond risk premia and macroeconomic factors.

Summary statistics for excess log returns are shown in Table 1 for levels and absolute values (Panels A and B, respectively). Both mean and standard deviation of bond excess returns increase with maturity: the former ranges between 0.4% and 0.9%, the latter between 1.9% and 5.8%. Bond risk premia are highly persistent: the first order autocorrelation is above 0.9 over all maturities, which is consistent with the MA (12) structure due to the overlapping data previously discussed; the autocorrelation of order 12 is positive and monotonically declines with maturity, ranging between 0.215 for $n=2$ and 0.071 for $n=5$. Absolute values of bond excess returns proxy bond risk premia volatilities: they are highly persistent, with mean and standard deviation increasing with maturity.

Breakpoint factor model: Macro factors

We set $R^{\max} = 8$; following the Monte Carlo results, we apply the $IC_{p2}(R, R)$ criterion in equation (6) and estimate $\hat{R} = 2$ factors. The estimated break fraction $\hat{\pi}_x = 0.455$ corresponds to the estimated break date $\hat{T}_x = 1983 : 06$: this is just before the beginning of the Great Moderation in early 1985. Our estimated number of factors is in line with what suggested in studies using data sets similar to ours: Stock and Watson (2012) *a priori* select five factors without accounting for structural breaks, which augment the factor space (Breitung and Eickmeier, 2011; Chen *et al.*, 2014); Cheng *et al.* (2016) select one factor during the period ranging from 1985 to 2007, which coincides with our post-break sample.

TABLE 2

Empirical results, estimated factors, average marginal R², 1964:01–2006:12

	Group 1: Output and income	Group 2: Labour market
\hat{f}_{1t}	0.298	0.166
\hat{f}_{2t}	0.107	0.078
	Group 3: Consumption and orders	Group 4: Orders and inventories
\hat{f}_{1t}	0.211	0.100
\hat{f}_{2t}	0.046	0.086
	Group 5: Money and credit	Group 6: Interest rate and exchange rates
\hat{f}_{1t}	0.032	0.084
\hat{f}_{2t}	0.025	0.263
	Group 7: Prices	Group 8: Stock market
\hat{f}_{1t}	0.249	0.002
\hat{f}_{2t}	0.004	0.185

Notes: For each of the eight groups listed in section ‘Data and implementation’, this table displays the average of the R^2 obtained from the breakpoint regression with $\hat{\pi}_x = 0.455$ of each variable within the group on a given factor. The data sample ranges from January 1964 to December 2006, for a sample size of 516 monthly observations.

We interpret the estimated factors by studying their predictive power (Ludvigson and Ng, 2009). For each of the eight groups listed in section ‘Data and implementation’, we set $\hat{\pi}_x = 0.455$ (i.e. the estimate for π_x^0) and compute the average of the R^2 obtained from the breakpoint regression of each variable within the group on a given factor. The results in Table 2 show that the first estimated factor \hat{f}_{1t} captures the real side of the economy and inflation: the average R^2 from regressions of macroeconomic variables in Groups 1, 2, 3 and 7 are equal to 0.298, 0.166, 0.211 and 0.249 respectively. The second estimated factor \hat{f}_{2t} loads on variables in Groups 6 and 8: it describes the financial side of the economy; it is also correlated with the CP_t factor, with correlation coefficient equal to -0.248 . The two factors exhibit time series dependence: the first order autocorrelation for \hat{f}_{1t} and \hat{f}_{2t} are equal to 0.485 and 0.630 respectively.

Unstable factor augmented regression: Bond risk premia

Table 3 displays results from four specifications nested in equation (9): (a) the Cochrane and Piazzesi (2005) single factor model, with $\gamma_1^{(n),0} = \gamma_2^{(n),0} = \mathbf{0}$; (b) a linear model with macroeconomic factors only, with $\gamma_1^{(n),0} = \gamma_2^{(n),0}$ and $\beta_2^{(n),0} = \mathbf{0}$; (c) a factor model under the null hypothesis of linearity $\mathbb{H}_0 : (\gamma_1^{(n),0} = \gamma_2^{(n),0})$; (d) the unstable factor augmented regression in equation (9).

The estimates for $\beta_2^{(n),0}$ from model (a) are 0.455, 0.857, 1.235 and 1.453 for $n = 2, 3, 4, 5$ respectively: these resemble the estimates in Cochrane and Piazzesi (2005) and do not contradict the assumption of stability imposed in equation (9) on $\beta_2^{(n),0}$. Based on the adjusted R^2 , the CP_t factor explains between 34% and 38% of the variation in bond excess returns: these values are in line with those in Cochrane and Piazzesi (2005). Model (b) shows that macroeconomic factors have less explanatory power than the CP_t factor across all maturities. Judging from the values of the adjusted R^2 from the *linear* model (c),

TABLE 3
Empirical results, bond excess returns, least squares estimation, 1965:01–2007:12

	n = 2				n = 3				n = 4				n = 5			
	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
$\pi_y^{(n)0}$	-	-	-	0.510	-	-	-	0.510	-	-	-	0.510	-	-	-	0.510
$\gamma_{11}^{(n)0}$	-	0.008	0.005	0.007	-	0.012	0.006	0.009	-	0.015	0.005	0.009	-	0.017	0.005	0.029***
$\gamma_{12}^{(n)0}$	-	(0.006)	(0.004)	(0.006)	-	(0.011)	(0.008)	(0.010)	-	(0.014)	(0.010)	(0.013)	-	(0.017)	(0.012)	(0.015)
$\gamma_{21}^{(n)0}$	-	-0.009*	-0.001	0.010***	-	-0.021**	-0.006	0.013*	-	-0.028**	-0.007	0.016*	-	-0.032**	-0.007	0.003
$\gamma_{22}^{(n)0}$	-	(0.005)	(0.006)	(0.004)	-	(0.008)	(0.009)	(0.007)	-	(0.011)	(0.011)	(0.009)	-	(0.013)	(0.012)	(0.014)
$\beta_1^{(n)0}$	-	-	-	-0.003	-	-	-	-0.005	-	-	-	-0.006	-	-	-	-0.016
$\beta_2^{(n)0}$	-	-	-	(0.002)	-	-	-	(0.003)	-	-	-	(0.005)	-	-	-	(0.015)
R^2	-	-	-	-0.021***	-	-	-	-0.035***	-	-	-	-0.043***	-	-	-	-0.013
sup \mathcal{LM}	-	-	-	(0.005)	-	-	-	(0.009)	-	-	-	(0.012)	-	-	-	(0.014)
	0.001	0.004*	0.001	0.000	0.001	0.008*	0.001	-0.001	0.000	0.010	0.000	-0.002	-0.002	0.010	-0.001	0.000
	(0.002)	(0.003)	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.007)	(0.006)	(0.005)
	0.455***	-	0.438***	0.467***	0.857***	-	0.823***	0.868***	1.235***	-	1.199***	1.256***	1.453***	-	1.417***	1.464***
	(0.052)	-	(0.058)	(0.058)	(0.103)	-	(0.115)	(0.118)	(0.150)	-	(0.165)	(0.168)	(0.192)	-	(0.208)	(0.202)
	0.344	0.102	0.352	0.415	0.358	0.095	0.362	0.404	0.382	0.084	0.383	0.415	0.347	0.065	0.347	0.370
	-	-	-	10.756*	-	-	-	9.868	-	-	-	10.295*	-	-	-	11.224*

Notes: This table reports results from least squares estimation of the following models for n -year bond excess returns $r_{t,t+12}^{(n)}$, for $t = 1964:01, \dots, 2006:12$:

- (a) $r_{t,t+12}^{(n)} = \beta_1^{(n)} + \beta_2^{(n)} CP_t + \varepsilon_{t,t+12}^{(n)}$;
- (b) $r_{t,t+12}^{(n)} = \gamma_{11}^{(n)} \hat{f}_{1t} + \gamma_{12}^{(n)} \hat{f}_{2t} + \beta_1^{(n)} + \varepsilon_{t,t+12}^{(n)}$;
- (c) $r_{t,t+12}^{(n)} = \gamma_{11}^{(n)} \hat{f}_{1t} + \gamma_{12}^{(n)} \hat{f}_{2t} + \beta_1^{(n)} + \beta_2^{(n)} CP_t + \varepsilon_{t,t+12}^{(n)}$;
- (d) $r_{t,t+12}^{(n)} = \mathbb{1}(tT \leq \pi_y^{(n)}) \left(\gamma_{21}^{(n)} \hat{f}_{1t} + \gamma_{22}^{(n)} \hat{f}_{2t} \right) + \beta_1^{(n)} + \beta_2^{(n)} CP_t + \varepsilon_{t,t+12}^{(n)}$.

The vector of factors $\hat{f}_t = (\hat{f}_{1t}, \hat{f}_{2t})'$ is estimated as detailed in section 'Principal components estimation'; the number of factors $\hat{R} = 2$ (i.e. the dimension of \hat{f}_t) is determined using the $J_{C_{\gamma_2}}(R, R)$ criterion in equation (6). The factors are extracted from the 123 macroeconomic series described in section 'Data and implementation' over the period 1964:01–2006:12. CP_t is the Cochrane and Piazzesi (2005) factor, namely a linear combination of five forward spreads estimated over the period 1964:01–2006:12. Newey and West (1987) standard errors with 18 months lag order are reported in parentheses. \hat{R}^2 , **, * and *** denote significance at 10%, 5% and 1% level respectively. Critical values for the sup \mathcal{LM} statistic with $\bar{x}_y = 15\%$ trimming and $\hat{R} = 2$ restrictions are 10.01, 11.79 and 15.51 for 10%, 5% and 1% level respectively (Andrews, 1993).

macroeconomic factors seem to have limited predictive power in excess of the CP_t factor during the period of interest.

We estimate model (d) over the interval $[\underline{\pi}_y, 1 - \underline{\pi}_y]$, with $\underline{\pi}_y = 0.15$. The adjusted R^2 shows that allowing for structural instability provides a great deal of additional explanatory power with respect to the linear factor augmented regression (c). The sup \mathcal{LM} test rejects the null hypothesis of linearity at 10% level for $n = 2, 4, 5$: this provides statistical evidence in favour of breaks in the factor loadings in the pricing equation for bond excess returns. The estimated break fraction for $n = 2, 4$ is $\hat{\pi}_y^{(n)} = 0.510$, with corresponding estimated break date $\hat{T}_y^{(n)} = \lceil T \hat{\pi}_y^{(n)} \rceil = 1986 : 11$; for $n = 5$, we have $\hat{\pi}_y^{(n)} = 0.355$ and $\hat{T}_y^{(n)} = 1980 : 03$. When structural breaks are allowed for, macroeconomic factors become statistically significant at least on one side of the breakpoint: their predictive ability thus varies over time.

According to our results, the location of the break depends on the maturity of the bond. Our findings relate to Joslin *et al.* (2014), who show that, with the exception of the first year and three isolated months, the Great Moderation period between 1985 and 2007 is well approximated by a single regime; they are also in line with Smith and Taylor (2009), who acknowledge the presence of a break in the early 1980s due to a shift in the monetary policy rule. More generally, we contribute to the literature by shedding further light on the issue of structural instability in bond risk premia dynamics. Our findings run parallel to existing evidence of instabilities in stock returns prediction models (Rapach and Zhou, 2013).

In conclusion, the proposed unstable diffusion index forecasting model has both statistical and economic advantages with respect to existing competing models for bond risk premia: it generates higher adjusted R^2 ; it captures the effects of events such as the Great Moderation and shifts in monetary policy rules on bond excess returns.

Alternative estimation strategy

The vector $(\mathbf{x}'_t, rx_{t+12}^{(n)})'$ admits a common factor representation, for $n = 1, 2, 3, 4$, where \mathbf{x}_t collects the macroeconomic variables. The common components could be estimated by projecting $(\mathbf{x}'_t, rx_{t+12}^{(n)})'$ on the estimated factors: see section 3.3 in Forni *et al.* (2005). Only π_x^0 and not π_y^0 would be identified though, as the contribution of $rx_{t+12}^{(n)}$ to the estimator of the common components would be negligible as $N \rightarrow \infty$. The formulation for the factor augmented regression

$$rx_{t+12}^{(n)} = \mathbb{1}(t/T \leq \pi_x) \gamma_1^{(n)'} \mathbf{f}_t + \mathbb{1}(t/T > \pi_x) \gamma_2^{(n)'} \mathbf{f}_t + \beta_1^{(n)} + \beta_2^{(n)} CP_t + \varepsilon_{t+12}^{(n)}$$

could be implemented, where π_x^0 and \mathbf{f}_t^0 would be estimated from the factor model.

Table 4 collects results from this estimation strategy. For maturities $n = 2, 3, 4$, the results are aligned with the homologous findings in Table 3, whereas for $n = 5$ the differences are more pronounced. Each adjusted R^2 in Table 4 is lower than the homologous value in Table 3. We can conclude that the estimator proposed in this paper is likely to provide more accurate information about the pricing equation for bond risk premia.

TABLE 4
*Alternative estimation strategy, bond excess returns,
 1965:01–2007:12*

	<i>n</i> = 2	<i>n</i> = 3	<i>n</i> = 4	<i>n</i> = 5
$\gamma_{11}^{(n)\hat{D}}$	0.007 (0.006)	0.009 (0.010)	0.017 (0.010)	0.009 (0.015)
$\gamma_{12}^{(n)\hat{D}}$	0.010*** (0.004)	0.013* (0.007)	0.017* (0.010)	0.016 (0.012)
$\gamma_{21}^{(n)\hat{D}}$	-0.002 (0.002)	-0.004 (0.003)	-0.006 (0.004)	-0.007 (0.005)
$\gamma_{22}^{(n)\hat{D}}$	-0.016*** (0.006)	-0.028*** (0.009)	-0.035*** (0.012)	-0.035*** (0.014)
$\beta_1^{(n)\hat{D}}$	0.0001 (0.002)	0.000 (0.004)	-0.001 (0.005)	-0.003 (0.006)
$\beta_2^{(n)\hat{D}}$	0.482*** (0.060)	0.895*** (0.119)	1.290*** (0.168)	1.506*** (0.213)
\bar{R}^2	0.398	0.396	0.409	0.364

Notes: This table reports results from least squares estimation of the following model for *n*-year bond excess returns $rx_{t+12}^{(n)}$, for $t = 1964 : 01, \dots, 2006 : 12$:

$$rx_{t+12}^{(n)} = \mathbb{1}(t/T \leq \hat{\pi}_x) \left(\gamma_{11}^{(n)} \hat{f}_{1t} + \gamma_{12}^{(n)} \hat{f}_{2t} \right) + \mathbb{1}(t/T > \hat{\pi}_x) \left(\gamma_{21}^{(n)} \hat{f}_{1t} + \gamma_{22}^{(n)} \hat{f}_{2t} \right) + \beta_1^{(n)} + \beta_2^{(n)} CP_t + \varepsilon_{t+12}^{(n)}$$

The vector of factors $\hat{\mathbf{f}}_t = (\hat{f}_{1t}, \hat{f}_{2t})'$ and the break fraction $\hat{\pi}_x$ are estimated as detailed in section ‘Principal components estimation’; the number of factors $\hat{R} = 2$ (i.e. the dimension of $\hat{\mathbf{f}}_t$) is determined using the $IC_{p2}(R, R)$ criterion in equation (6). The factors are extracted from the 123 macroeconomic series described in section ‘Data and implementation’ over the period 1964:01–2006:12. CP_t is the Cochrane and Piazzesi (2005) factor, namely a linear combination of five forward spreads estimated over the period 1964:01–2006:12. Newey and West (1987) standard errors with 18 months lag order are reported in parentheses. \bar{R}^2 is the adjusted R^2 . * and *** denote significance at 10% and 1% level respectively.

VII. Extensions and modifications

We now consider possible extensions and modifications to the model in (1) and (2): Section ‘Diffusion index threshold forecasting model’ introduces the diffusion index threshold forecasting model; Section ‘Diffusion index forecasting model with multiple breaks’ discusses the model with multiple breaks; Section ‘Alternative specification of parameter instability’ proposes an alternative strategy to allow for parameter instability in the diffusion index forecasting model.

Diffusion index threshold forecasting model

The specification in (1) and (2) allows for structural instability in the factor model and in the forecasting equation respectively. Most results in this paper hold for the following *diffusion index threshold forecasting model*

$$\begin{aligned} \mathbf{x}_t &= \mathbb{1}(g_t \leq \varphi_x) \Lambda_1 \mathbf{f}_t + \mathbb{1}(g_t > \varphi_x) \Lambda_2 \mathbf{f}_t + \mathbf{e}_t, \quad t = 1, \dots, T, \\ y_{t+h} &= \mathbb{1}(g_t \leq \varphi_y) (\gamma'_1 \mathbf{f}_t + \beta'_1 \mathbf{w}_t) + \mathbb{1}(g_t > \varphi_y) (\gamma'_2 \mathbf{f}_t + \beta'_2 \mathbf{w}_t) + \varepsilon_{t+h}, \quad t = 1, \dots, T, \quad h \geq 0, \end{aligned}$$

where $g_t \in \mathfrak{R}$ is the common observable threshold random variable, and φ_x and φ_y are the unknown threshold values. The model lets y_{t+h} depend upon \mathbf{f}_t , the vector of latent factors driving the comovement in \mathbf{x}_t according to the threshold factor model of Massacci (2017). Let $\hat{\varphi}_x$ and $\hat{\varphi}_y$ be the least squares estimators for φ_x^0 and φ_y^0 , respectively, namely for the true values of φ_x and φ_y respectively. The break fractions π_x^0 and π_y^0 in (1) and (2), respectively, are threshold values as applied to t/T . Under suitable assumptions, it can be shown that the results in Theorems 2–4 hold for the diffusion index threshold forecasting model with respect to $\hat{\varphi}_x$ and $\hat{\varphi}_y$. However, Theorem 5 no longer is valid: g_t generally is a random variable; a test for threshold effect in the forecasting equation can be implemented by suitably generalizing the results in Hansen (1996).

Diffusion index forecasting model with multiple breaks

The theoretical results in this paper accommodate one break in the factor model and one break in the factor augmented regression. A more empirically plausible scenario would be

$$\begin{aligned} \mathbf{x}_t &= \sum_{j_x=1}^{J_x^0+1} \mathbb{1}(\pi_{x,j_x-1} < t/T \leq \pi_{x,j_x}) \Lambda_{j_x} \mathbf{f}_t + \mathbf{e}_t, \quad t = 1, \dots, T, \\ y_{t+h} &= \sum_{j_y=1}^{J_y^0+1} \mathbb{1}(\pi_{y,j_y-1} < t/T \leq \pi_{y,j_y}) (\gamma'_{j_y} \mathbf{f}_t + \beta'_{j_y} \mathbf{w}_t) + \varepsilon_{t+h}, \quad t = 1, \dots, T, \quad h \geq 0, \end{aligned}$$

where $\pi_{x,0} = \pi_{y,0} = 0$ and $\pi_{x,J_x+1} = \pi_{y,J_y+1} = 1$: this allows for $J_x^0 \geq 1$ breaks in the factor model and $J_y^0 \geq 1$ breaks in the factor augmented regression, where for ease of exposition we assume that J_x^0 and J_y^0 are both known. To aid to the understanding of dealing with this more general model, consider first $J_x^0 = 1$ and $J_y^0 = 2$, in which case (2) generalizes to

$$\begin{aligned} y_{t+h} &= \mathbb{1}(t/T \leq \pi_{y,1}) (\gamma'_1 \mathbf{f}_t + \beta'_1 \mathbf{w}_t) \\ &\quad + \mathbb{1}(\pi_{y,1} < t/T \leq \pi_{y,2}) (\gamma'_2 \mathbf{f}_t + \beta'_2 \mathbf{w}_t) \\ &\quad + \mathbb{1}(t/T > \pi_{y,2}) (\gamma'_3 \mathbf{f}_t + \beta'_3 \mathbf{w}_t) + \varepsilon_{t+h}, \quad t = 1, \dots, T, \quad h \geq 0. \end{aligned}$$

Let $\theta_{j_y}^0 = (\gamma_{j_y}^{0r}, \beta_{j_y}^{0r})'$ be the true value of $\theta_{j_y} = (\gamma'_{j_y}, \beta'_{j_y})'$, for $j_y = 1, 2, 3$. Define the estimator $\hat{\theta}_{j_y} = (\hat{\gamma}'_{j_y}, \hat{\beta}'_{j_y})'$ for $\theta_{j_y}^0 = (\gamma_{j_y}^{0r}, \beta_{j_y}^{0r})'$, for $j_y = 1, 2, 3$. Let $\pi_{y,1}^0$, $\pi_{y,2}^0$ and $\pi_{y,3}^0$ be the true values of $\pi_{y,1}$, $\pi_{y,2}$ and $\pi_{y,3}$ respectively. The results in Theorem 2 generalize as

$$\hat{\theta}_1 - \left[\frac{\min \{ \pi_{x,1}^0, \pi_{y,1}^0 \}}{\pi_{y,1}^0} \hat{\Phi}_{11}(\pi_{x,1}^0) + \mathbb{1}(\pi_{x,1}^0 < \pi_{y,1}^0) \frac{\pi_{y,1}^0 - \pi_{x,1}^0}{\pi_{y,1}^0} \hat{\Phi}_{22}(\pi_{x,1}^0) \right]' \theta_1^0 = o_p(1),$$

$$\hat{\theta}_2 - \left[\mathbb{1}(\pi_{y,1}^0 < \pi_{x,1}^0) \frac{\min \{ \pi_{x,1}^0, \pi_{y,2}^0 \} - \pi_{y,1}^0}{\pi_{y,2}^0 - \pi_{y,1}^0} \hat{\Phi}_{11}(\pi_{x,1}^0) + \mathbb{1}(\pi_{x,1}^0 \leq \pi_{y,2}^0) \frac{\pi_{y,2}^0 - \max \{ \pi_{x,1}^0, \pi_{y,1}^0 \}}{\pi_{y,2}^0 - \pi_{y,1}^0} \hat{\Phi}_{22}(\pi_{x,1}^0) \right]' \theta_2^0 = o_p(1),$$

$$\hat{\theta}_3 - \left[\frac{1 - \max \{ \pi_{x,1}^0, \pi_{y,2}^0 \}}{1 - \pi_{y,2}^0} \hat{\Phi}_{22}(\pi_{x,1}^0) + \mathbb{1}(\pi_{x,1}^0 > \pi_{y,2}^0) \frac{\pi_{x,1}^0 - \pi_{y,2}^0}{1 - \pi_{y,2}^0} \hat{\Phi}_{11}(\pi_{x,1}^0) \right]' \theta_3^0 = o_p(1).$$

If we then allowed for $J_x^0 = 2$ breaks, the factor model in (1) would become

$$\mathbf{x}_t = \mathbb{1}(t/T \leq \pi_{x,1}) \Lambda_1 \mathbf{f}_t + \mathbb{1}(\pi_{x,1} < t/T \leq \pi_{x,2}) \Lambda_2 \mathbf{f}_t + \mathbb{1}(t/T > \pi_{x,2}) \Lambda_3 \mathbf{f}_t + \mathbf{e}_t, \quad t = 1, \dots, T :$$

in this case, the results regarding $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\theta}_3$, stated above would have to be extended to account for the additional degree of rotational indeterminacy in the factor model due to the presence of the additional break. This strategy would allow us to generalize Theorems 2 and 4 regarding $\hat{\theta}_{j_y}$, for $j_y = 1, \dots, J_y^0$, for any $J_x^0 \geq 1$ and $J_y^0 \geq 1$. In addition, in the case where the number of breaks J_y^0 no longer is known, Theorem 5 is not valid: the seminal contribution by Bai and Perron (1998) would be a natural starting point to conduct inference on the number of factors in the forecasting equation.

Alternative specification of parameter instability

The model in (1) and (2) allows for different break fractions π_x and π_y in the factor model and in the factor augmented regression, respectively, where π_y is the same for latent factors \mathbf{f}_t and observable covariates \mathbf{w}_t . Alternatively, one could assume a common break fraction π^f for \mathbf{f}_t in the factor model and in the factor augmented regression, and a break fraction π^w for \mathbf{w}_t : formally, the model would be

$$\mathbf{x}_t = \mathbb{1}(t/T \leq \pi^f) \Lambda_1 \mathbf{f}_t + \mathbb{1}(t/T > \pi^f) \Lambda_2 \mathbf{f}_t + \mathbf{e}_t, \quad t = 1, \dots, T,$$

$$y_{t+h} = \mathbb{1}(t/T \leq \pi^f) \gamma'_1 \mathbf{f}_t + \mathbb{1}(t/T > \pi^f) \gamma'_2 \mathbf{f}_t$$

$$+ \mathbb{1}(t/T \leq \pi^w) \beta'_1 \mathbf{w}_t + \mathbb{1}(t/T > \pi^w) \beta'_2 \mathbf{w}_t + \varepsilon_{t+h}, \quad t = 1, \dots, T, \quad h \geq 0.$$

Let $\pi^{f,0}$ and $\pi^{w,0}$ denote the true values of π^f and π^w respectively. Under the maintained assumption of structural instability in the factor model, $\pi^{f,0}$ can be estimated from the factor model only, as the contribution to the underlying loss function coming from the factor augmented regression would be infinitesimal as $N \rightarrow \infty$. This has implications for estimation and inference in the factor augmented regression: only $\pi^{w,0}$ needs to be estimated; the critical values of the test statistic for the null hypothesis $\gamma_1^0 = \gamma_2^0$ are obtained from the $\chi^2(\hat{R})$ distribution; the null hypothesis $\beta_1^0 = \beta_2^0$ still needs to be tested using the procedure detailed in Andrews (1993).

More generally, one could also allow for different break fractions for \mathbf{f}_t in the factor model and in the factor augmented regression: formally, the model would be

$$\begin{aligned} \mathbf{x}_t &= \mathbb{1}(t/T \leq \pi_x^f) \Lambda_1 \mathbf{f}_t + \mathbb{1}(t/T > \pi_x^f) \Lambda_2 \mathbf{f}_t + \mathbf{e}_t, \quad t = 1, \dots, T, \\ y_{t+h} &= \mathbb{1}(t/T \leq \pi_y^f) \gamma_1' \mathbf{f}_t + \mathbb{1}(t/T > \pi_y^f) \gamma_2' \mathbf{f}_t \\ &\quad + \mathbb{1}(t/T \leq \pi_y^w) \beta_1' \mathbf{w}_t + \mathbb{1}(t/T > \pi_y^w) \beta_2' \mathbf{w}_t + \varepsilon_{t+h}, \quad t = 1, \dots, T, \quad h \geq 0. \end{aligned}$$

The null hypotheses $\gamma_1^0 = \gamma_2^0$ and $\beta_1^0 = \beta_2^0$ could be tested separately by using a procedure analogous to that described in section ‘Diffusion index forecasting model with multiple breaks’: the former and the latter would require testing \hat{R} and K restrictions, respectively. A test for the null hypothesis $\theta_1^0 = \theta_2^0$ would involve the two nuisance parameters π_y^f and π_y^w : we leave this interesting problem to future research.

VIII. Conclusion

We study estimation and inference in unstable diffusion index forecasting models: the application uncovers breaks in the linkages between bond risk premia and macroeconomic factors. The novelty of the problem and the related technical difficulties make us analyse the single break case: the multiple break scenario is on our list of priorities (Bai and Perron, 1998).

The paper focuses on in-sample analysis. Diffusion indexes are widely used for out-of-sample forecasting. They may perform poorly in practice due to instabilities in the factor model (Giannone, 2007; Banerjee *et al.*, 2008) or in the factor augmented regression (Stock and Watson, 2002, 2009). Forecasting techniques robust to breaks would be desirable, especially if a break is located close to the end of the sample: one possible solution is to extend the Pesaran and Timmermann (2007) procedure to select the estimation windows to include estimated factors.

Final Manuscript Received: April 2019

References

- Ahn, S. C. and Horenstein, A. R. (2013). ‘Eigenvalue ratio test for the number of factors’, *Econometrica*, Vol. 81, pp. 1203–1227.
- Alessi, L., Barigozzi, M. and Capasso, M. (2010). ‘Improved penalization for determining the number of factors in approximate factor models’, *Statistics and Probability Letters*, Vol. 80, pp. 1806–1813.
- Amengual, D. and Watson, M. W. (2007). ‘Consistent estimation of the number of dynamic factors in a large N and T panel’, *Journal of Business and Economic Statistics*, Vol. 25, pp. 91–96.
- Andrews, D. W. K. (1993). ‘Tests for parameter instability and structural change with unknown change point’, *Econometrica*, Vol. 61, pp. 821–856.
- Bai, J. (1997). ‘Estimation of a change point in multiple regression models’, *Review of Economics and Statistics*, Vol. 79, pp. 551–563.
- Bai, J. (2003). ‘Inferential theory for factor models of large dimensions’, *Econometrica*, Vol. 71, pp. 135–171.
- Bai, J. and Ng, S. (2002). ‘Determining the number of factors in approximate factor models’, *Econometrica*, Vol. 70, pp. 191–221.
- Bai, J. and Ng, S. (2006). ‘Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions’, *Econometrica*, Vol. 74, pp. 1133–1150.

- Bai, J. and Perron, P. (1998). 'Estimation and testing linear models with multiple structural changes', *Econometrica*, Vol. 66, pp. 47–78.
- Bai, J., Han, X. and Shi, Y. (2017). *Estimation and Inference of Change Points in High-Dimensional Factor Models*, Working Paper, Columbia University, City University of Hong Kong and Texas A&M University.
- Baltagi, B. H., Kao, C. and Wang, F. (2016). *Estimating and Testing High Dimensional Factor Models with Multiple Structural Changes*, Working Paper, Syracuse University.
- Baltagi, B. H., Kao, C. and Wang, F. (2017). 'Identification and estimation of a large factor model with structural instability', *Journal of Econometrics*, Vol. 197, pp. 87–100.
- Banerjee, A., Marcellino, M. and Masten, I. (2008). *Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Change*, EUI Working Paper, ECO 2008/17.
- Barigozzi, M. and Trapani, L. (2017). *Sequential Testing for Structural Stability in Approximate Factor Models*, Working Paper, London School of Economics and Cass Business School.
- Barigozzi, M., Cho, H. and Fryzlewicz, P. (2018). 'Simultaneous multiple change-point and factor analysis for high-dimensional time series', *Journal of Econometrics*, Vol. 206, pp. 187–225.
- Bates, B. J., Plagborg-Møller, M., Stock, J. H. and Watson, M. W. (2013). 'Consistent factor estimation in dynamic factor models with structural instability', *Journal of Econometrics*, Vol. 177, pp. 289–304.
- Bikbov, R. and Chernov, M. (2010). 'No-arbitrage macroeconomic determinants of the yield curve', *Journal of Econometrics*, Vol. 159, pp. 166–182.
- Breitung, J. and Eickmeier, S. (2011). 'Testing for structural breaks in dynamic factor models', *Journal of Econometrics*, Vol. 163, pp. 71–84.
- Campbell, J. Y., and Shiller, R. J. (1991). 'Yield spreads and interest rate movements: a bird's eye view', *Review of Economic Studies*, Vol. 58, pp. 495–514.
- Caner, M. and Han, X. (2014). 'Selecting the correct number of factors in approximate factor models: The large panel case with group bridge estimators', *Journal of Business & Economic Statistics*, Vol. 32, pp. 359–374.
- Chamberlain, G. and Rothschild, M. (1983). 'Arbitrage, factor structure, and mean-variance analysis on large asset markets', *Econometrica*, Vol. 51, pp. 1281–1304.
- Chen, L. (2015). 'Estimating the common break date in large factor models', *Economics Letters*, Vol. 131, pp. 70–74.
- Chen, L., Dolado, J. J. and Gonzalo, J. (2014). 'Detecting big structural breaks in large factor models', *Journal of Econometrics*, Vol. 180, pp. 30–48.
- Cheng, X., Liao, Z. and Schorfheide, F. (2016). 'Shrinkage estimation of high-dimensional factor models with structural instabilities', *Review of Economic Studies*, Vol. 83, pp. 1511–1543.
- Cochrane, J. H. and Piazzesi, M. (2005). 'Bond risk premia', *American Economic Review*, Vol. 95, pp. 138–160.
- Connor, G. and Korajczyk, R. A. (1986). 'Performance measurement with the arbitrage pricing theory: A new framework for analysis', *Journal of Financial Economics*, Vol. 15, pp. 373–394.
- Connor, G. and Korajczyk, R. A. (1988). 'Risk and return in an equilibrium APT: Application of a new test methodology', *Journal of Financial Economics*, Vol. 21, pp. 255–289.
- Connor, G. and Korajczyk, R. A. (1993). 'A test for the number of factors in an approximate factor model', *Journal of Finance*, Vol. 48, pp. 1263–1288.
- Corradi, V. and Swanson, N. R. (2014). 'Testing for structural stability of factor augmented forecasting models', *Journal of Econometrics*, Vol. 182, pp. 100–118.
- Fama, E. F. and Bliss, R. R. (1987). 'The information in long-maturity forward rates', *American Economic Review*, Vol. 77, pp. 680–692.
- Forni, M. and Lippi, M. (2001). 'The generalized dynamic factor model: Representation theory', *Econometric Theory*, Vol. 17, pp. 1113–1141.
- Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). 'The generalized dynamic factor model: Identification and estimation', *Review of Economics and Statistics*, Vol. 82, pp. 540–554.
- Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2004). 'The generalized dynamic factor model: Consistency and rates', *Journal of Econometrics*, Vol. 119, pp. 231–255.
- Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005). 'The generalized dynamic factor model: One-sided estimation and forecasting', *Journal of the American Statistical Association*, Vol. 100, pp. 830–840.
- Forni, M., Hallin, M., Lippi, M. and Zaffaroni, P. (2015). 'Dynamic factor models with infinite-dimensional factor spaces: one-sided representations', *Journal of Econometrics*, Vol. 185, pp. 359–371.

- Giannone, D. (2007). 'Discussion on: "Forecasting in dynamic factor models subject to structural instability"', Stock J. H. and Watson M. W. (eds), *New Developments in Dynamic Factor Modelling*, Bank of England, Centre for Central Banking Studies, London.
- Gürkaynak, R. S. and Wright, J. H. (2012). 'Macroeconomics and the term structure', *Journal of Economic Literature*, Vol. 50, pp. 331–367.
- Hallin, M. and Liška, R. (2007). 'Determining the number of factors in the general dynamic factor model', *Journal of the American Statistical Association*, Vol. 102, pp. 603–617.
- Han, X. and Inoue, A. (2015). 'Tests for parameter instability in dynamic factor models', *Econometric Theory*, Vol. 31, pp. 1117–1152.
- Hansen, B. E. (1996). 'Inference when a nuisance parameter is not identified under the null hypothesis', *Econometrica*, Vol. 64, pp. 413–430.
- Hansen, B. E. (2000). 'Sample splitting and threshold estimation', *Econometrica*, Vol. 68, pp. 575–603.
- Joslin, S., Priebsch, M. and Singleton, K. J. (2014). 'Risk premiums in dynamic term structure models with unspanned macro risks', *Journal of Finance*, Vol. 69, pp. 1197–1233.
- Kapetanios, G. (2010). 'A testing procedure for determining the number of factors in approximate factor models with large datasets', *Journal of Business and Economic Statistics*, Vol. 28, pp. 397–409.
- Ludvigson, S. C., and Ng, S. (2009). 'Macro factors in bond risk premia', *Review of Financial Studies*, Vol. 22, pp. 5027–5067.
- Ma, S., and Su, L. (2016). *Estimation of Large Dimensional Factor Models with an Unknown Number of Breaks*, Working Paper, University of California, Riverside and Singapore Management University.
- Massacci, D. (2017). 'Least squares estimation of large dimensional threshold factor models', *Journal of Econometrics*, Vol. 197, pp. 101–129.
- McCracken, M. W., and Ng, S. (2016). 'FRED-MD: A monthly database for macroeconomic research', *Journal of Business and Economics Statistics*, Vol. 34, pp. 574–589.
- Newey, W. K. and West, K. D. (1987). 'A simple, positive semi-definite, heteroskedasticity and autorrelation consistent covariance matrix', *Econometrica*, Vol. 55, pp. 703–708.
- Onatski, A. (2009). 'Testing hypotheses about the number of factors in large factor models', *Econometrica*, Vol. 77, pp. 1447–1479.
- Onatski, A. (2010). 'Determining the number of factors from empirical distribution of eigenvalues', *Review of Economics and Statistics*, Vol. 92, pp. 1004–1016.
- Peligrad, M. (1982). 'Invariance principles for mixing sequences of random variables', *Annals of Probability*, Vol. 10, pp. 968–980.
- Perron, P. (2006). 'Dealing with structural breaks', in Patterson K. and Mills T. C. (eds.), *Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory*, Palgrave Macmillan, pp. 278–352.
- Pesaran, M. H. and Timmermann, A. (2007). 'Selection of estimation window in the presence of breaks', *Journal of Econometrics*, Vol. 137, pp. 134–161.
- Rapach, D. E., and Zhou, G. (2013). 'Forecasting stock returns', in Elliott G. and Timmermann A. (eds.), *Handbook of Economic Forecasting, Volume 2, Part A*, North-Holland, pp. 329–383.
- Ross, S. (1976). 'The arbitrage theory of capital pricing', *Journal of Economic Theory*, Vol. 13, pp. 341–360.
- Smith, J. M., and Taylor, J. B. (2009). 'The term structure of policy rules', *Journal of Monetary Economics*, Vol. 56, pp. 907–917.
- Stock, J. H. and Watson, M. W. (1998). *Diffusion Indexes*, NBER Working Paper No. 6702.
- Stock, J. H., and Watson, M. W. (2002). 'Forecasting using principal components from a large number of predictors', *Journal of the American Statistical Association*, Vol. 97, pp. 1167–1179.
- Stock, J. H. and Watson, M. W. (2009). 'Forecasting in dynamic factor models subject to structural instability', in Castle J. and Shephard N. (eds.), *The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry*, Oxford: Oxford University Press.
- Stock, J. H. and Watson, M. W. (2012). 'Generalized shrinkage methods for forecasting using many predictors', *Journal of Business and Economics Statistics*, Vol. 30, pp. 481–493.
- Timmermann, A. (2008). 'Elusive return predictability', *International Journal of Forecasting*, Vol. 24, pp. 1–18.
- Wang, S., Cui, G. and Li, K. (2015). 'Factor-augmented regression models with structural change', *Economics Letters*, Vol. 130, pp. 124–127.

- Wu, C. J. and Xia, F. D. (2016). 'Measuring the macroeconomic impact of monetary policy at the zero lower bound', *Journal of Money, Credit and Banking*, Vol. 48, pp. 253–291.
- Yamamoto, Y. and Tanaka, S. (2015). 'Testing for factor loading structural change under common breaks', *Journal of Econometrics*, Vol. 189, pp. 187–206.

Supporting Information

Additional supporting information may be found in the online version of this article:

Appendix A. Monte Carlo Analysis.

Appendix B. Proofs of Theorems.