Review of Economic Studies (2016) **83**, 1511–1543 doi:10.1093/restud/rdw005

Advance access publication 21 January 2016

# Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities

XU CHENG

*University of Pennsylvania*

ZHIPENG LIAO

*University of California, Los Angeles*

and

FRANK SCHORFHEIDE

*University of Pennsylvania and NBER*

*First version received September* 2014*; final version accepted November* 2015 (*Eds.*)

In large-scale panel data models with latent factors the number of factors and their loadings may change over time. Treating the break date as unknown, this article proposes an adaptive group-LASSO estimator that consistently determines the numbers of pre- and post-break factors and the stability of factor loadings if the number of factors is constant. We develop a cross-validation procedure to fine-tune the data-dependent LASSO penalties and show that after the number of factors has been determined, a conventional least-squares approach can be used to estimate the break date consistently. The method performs well in Monte Carlo simulations. In an empirical application, we study the change in factor loadings and the emergence of new factors in a panel of U.S. macroeconomic and financial time series during the Great Recession.

*Key words*: Great Recession, High-dimensional model, Large data sets, LASSO, Latent factor model, Model selection, Shrinkage estimation, Structural break

*JEL Codes*: C13, C33, C52

## 1. INTRODUCTION

High-dimensional factor models are widely used to analyse macroeconomic and financial panel data, where a small number of unobserved factors drive the comovement of a large number of time series. This article focuses on the complications in the estimation of factor models that arise from potential structural breaks such as the 2007–9 Great Recession, which, unlike other post-war U.S. recessions, was characterized by a severe disruption of financial markets, a slow recovery, and a lasting episode of zero nominal interest rates and unconventional monetary policies. The empirical application raises a number of interesting and important questions: did the Great Recession trigger a long-lasting change in business cycle dynamics? In the context of a factor model representation for macroeconomic and financial indicators, was the Great Recession associated with the emergence of new factors, *e.g.* a financial or a credit factors? Did the loading on existing factors change? When exactly did this change occur: during the subprime mortgage crisis in mid-2007, during the Bear Stearns rescue in March 2008, or during the Lehman Brothers collapse in September 2008?

None of the existing econometric techniques for factor models can answer all of these questions simultaneously. Existing methods to determine the number of factors, *e.g.* Bai and Ng (2002), Onatski (2010), or Ahn and Horenstein (2013), would require the knowledge of the break date and are unable to detect changes in loadings only. Structural break tests for factor loadings, *e.g.* Breitung and Eickmeier (2011), Chen *et al*. (2014), or Han and Inoue (2014) do not provide estimates of the number of factors and are not designed to detect a change in the number of factors. Conventional residual-based procedures to determine the break date, *e.g.* Bai (1997), require the number of factors to be known. Stock and Watson (2012) assess the evidence for a break in the number of factors during the Great Recession by testing for the presence of a factor structure in the errors associated with the forecasts of the post-break observations based on extensions of the pre-break factors. However, their approach requires knowledge of the break date and is not designed to distinguish a change in the number of factors from a change in factor loadings.

The main contribution of this article is to develop an econometric procedure that consistently detects changes in factor loadings, determines the numbers of pre- and post-break factors, and estimates the break date if it is unknown. Formally, we consider two types of factor model instabilities: large changes in the factor loadings when the number of factors is constant (type-1 instability), and changes in the number of strong factors (type-2 instability). Beyond the particular application studied in this article, a consistent estimator for the unobserved factor structure in large-scale panel data models is essential in many other empirical contexts. In general, ignoring a break point leads to an overestimation of the numbers of pre- and post-break factors and distorts any subsequent econometric analysis that conditions on the number of factors. In forecasting applications, using unnecessary predictors leads to imprecise forecasts. In a structural dynamic factor analysis that uses factor models to trace out the effect of structural innovations such as monetary and technology shocks on a large set of macroeconomic and financial indicators an incorrect estimate of the number of factors and their loadings makes it infeasible to recover the "true" impulse responses.

Our estimator utilizes two novel identification results for factor models. Because only the product of factors and loadings is identifiable, we use a normalization that attributes changes in this product to changes in the loadings. First, we show that a structural change is identifiable if either the space spanned by the factor loadings or the scaling of the factor loadings changes. Secondly, we show that the unknown break point is determined by the dimensionality of the factor model. It has been previously shown in the literature (*e.g.* Breitung and Eickmeier, 2011) that the presence of large breaks leads to the overestimation of the number of factors. As a consequence, the sum of the numbers of pre- and post-break factors is minimized if the break date is correctly specified. We exploit this insight to robustify the inference about the number of pre- and post-break factors against lack of knowledge of the exact break date. Moreover, we show that once the number of factors has been determined, one can estimate the location of the break date using a traditional sum-of-squared residuals criterion (*e.g.* Bai, 1997 in a model with observed regressors).

The estimator developed in this article is based on the minimization of a penalized least-squares (PLS) criterion function in which adaptive group-LASSO penalties (Tibshirani, 1994; Yuan and Lin, 2006 and Zou, 2006) are attached to pre-break factor loadings and to changes in factor loadings. The PLS estimator is a shrinkage estimator because, compared to the unrestricted least-squares estimator, it sets small coefficient estimates equal to zero. The numbers of pre- and post-break factors are determined based on the number of non-zero columns in the matrices of factor loadings and the change of factor loadings. A new factor appears if a column of zero loadings becomes non-zero after the break.

Although the idea of data-dependent penalty for LASSO originates from Zou (2006), our specific penalty is novel in three dimensions. First, to improve the finite-sample performance, the adaptive LASSO penalty is determined in a two-step procedure, in which the second-stage penalty is computed based on a first-stage shrinkage estimator. Importantly, in the second step, an orthogonal procrustes problem is solved to match the columns space of the pre- and post-break loadings obtained in the first-stage estimation. Secondly, to handle the unknown break date case, the penalty is constructed as an average over the penalties computed conditional on each potential break date. Unlike break-date-specific penalties, this averaging penalty ensures uniform convergence of the PLS criterion over all break dates, similar to the uniform results in Andrews (1993), and robustifies the estimated number of factors against small perturbations of the break date in finite samples. Thirdly, we develop a cross-validation procedure that lets the user fine-tune the LASSO penalties to improve the finite-sample performance of the shrinkage estimator.

Our theoretical results establish consistency of the estimation of the numbers of pre- and post-break factors, the detection of changes in loadings in case of type-1 instabilities, and the estimation of the break date. The results are obtained under large $N$ and $T$ asymptotics. The inference problem is high dimensional because the number of elements in each column of the loadings vector goes to infinity asymptotically and this rate can be faster than the rate at which the number of time periods diverges. Throughout this article, we assume that the number of factors is fixed as the sample size increases, that the factors are strong, and that the breaks in the loadings large in the sense that they do not shrink with the sample size.[1] Extensions to small breaks, weak factors, and numbers of factors that increase with sample size are beyond the scope of this article and left for future research.

The empirical analysis in this article revisits a recent study by Stock and Watson (2012), who investigated whether new factors appeared at the onset of the Great Recession, considering a large data set of macroeconomic and financial times series. In a nutshell, Stock and Watson (2012) extended the pre-break factor to the post-break period and examined whether there was evidence of an un-modelled factor in the residuals of the post-break sample. They found no such evidence. Using a similar set of time series, but sampled at a monthly frequency, and being agnostic about the specific break date, we find evidence of a type-2 instability at the beginning of the Great Recession, *i.e.* the emergence of a new factor which mostly affects financial variables, but also has spill-over effects on real activity variables. Conditional on the normalization of the pre- and post-break factors our estimation results indicate that the factor loadings changed drastically during this episode. Because Stock and Watson (2012) normalized the size of the loadings rather than the variance of the factors in their analysis, some but not all of the change in loadings in our analysis mirrors the increase in factor volatility in their analysis.

Our work is related to, but in several important dimensions distinctly different from, the existing literatures on factor model estimation, structural break testing, and LASSO estimation. If the break date were known, one could study the emergence of new post-break factors by applying one of the existing methods for determining the number of factors in a stable environment to the pre- and post-break subsamples. In a seminal paper, Bai and Ng (2002) provide information criteria to consistently determine the number of factors in time-invariant static factor models. Subsequent work, often distinguishing between the number of static and

1. Onatski (2012) analyses a model in which some of the factors only have a weak influence on the observables. Stock and Watson (2002) and Bates *et al*. (2013) show that in the presence of small structural instabilities of the factor loadings, the principal component estimator of the factors remains consistent.

dynamic factors includes Amengual and Watson (2007), Bai and Ng (2007), Hallin and Lika (2007), Onatski (2009, 2010), Alessi *et al*. (2010), Kapetanios (2010), Ahn and Horenstein (2013), Breitung and Pigorsch (2013), Choi (2013), and Caner and Han (2014). However, as discussed above, if the break date is unknown the direct application of these econometric procedures will overestimate the number of factors as soon as the break date is misspecified and pick up "pseudo-true" factors.

The procedure developed in this article not only allows researchers to consistently estimate the numbers of pre- and post-break factors but also consistently detect changes in factor loading if the number of factors stays constant. Several structural break tests for factor loadings have been developed, including Stock and Watson (2009), Breitung and Eickmeier (2011), Chen *et al*. (2014), Corradi and Swanson (2014), and Han and Inoue (2014). Our procedure differs in several dimensions. First, it detects the instabilities without requiring any knowledge of the numbers of pre- and post-break factors. This is important because a consistent estimator of these factors is not available in the literature when the break point is unknown. Secondly, to achieve consistency, we only require that the number of time-series variables and the number of time periods are both large without any restriction on their relative rates, whereas structural break tests in the literature typically restrict their relative rates to ensure that the generated-regressor effect from the estimation of unobserved factors is negligible. Thirdly, the procedure controls the model selection error jointly by treating all time-series variables as a group. This is particularly desirable for large-scale data sets.

There exists some recent work that utilizes shrinkage methods to estimate stable factor models (*e.g*. Bai and Liao, 2012; Caner and Han, 2014; and Lu and Su, 2015) and to detect structural breaks in models with observed regressors (*e.g*. Lee *et al*., 2015 and Qian and Su, 2015a,b). Our article differs from the above-mentioned work because the factor structure is both unobserved and unstable. Recently Baltagi *et al*. (2015) and Chen (2015) consider break-point estimation in large-scale factor models and Su and Wang (2015) develop an information criterion to estimate the number of factors in a factor model with slowly time-varying loadings. There is also a growing literature on modelling heterogeneity in panel data with latent group structure. Various classification and shrinkage methods have been proposed, see Lin and Ng (2012), Su *et al*. (2014), Ando and Bai (2015), Bonhomme and Manresa (2015). These papers consider a latent structure that is different from the one that is estimated in the current article. The structure in this article is comparable to time-varying interactive fixed effects. There are no additional exogenous regressors in our model because the factor structure is the parameter of interest instead of the nuisance parameter. A general form of time-varying group heterogeneity is considered by Bonhomme and Manresa (2015).

The remainder of this article is organized as follows. Section 2 describes the factor model and the types of instabilities considered in this article. It also provides identification conditions for changes in the loadings and identification results for the break point if it is unknown. Section 3 presents the shrinkage estimator and the model selection method. The selection of the tuning parameters as well as the practical implementation of the shrinkage estimation are addressed in Section 4. Section 5 develops the asymptotic theory for our estimator and establishes the consistency of the estimator of the number of the pre- and post-break factors, the stability of the loadings, and the break date. Monte Carlo results on the finite-sample performance of the proposed shrinkage estimator are reported in Section 6. These results include comparisons with existing procedures that are designed to determine the number of factors in a stable environment and procedures that test for the stability of loadings coefficients if the number of factors is known and stable. Section 7 contains the empirical application. Finally, Section 8 concludes. All proofs as well as additional simulation and empirical results are relegated to the Online Appendix.

## 2. A FACTOR MODEL WITH STRUCTURAL BREAK

We observe panel data $\{X_{it} \in R: i=1,\dots,N, t=1,\dots,T\}$. Let $X_t=(X_{1t},\dots,X_{Nt})' \in R^{N\times 1}$ denote the observations at time period $t$. For $t=1,\dots,T_0$, the observed $N$ series are driven by $r_a$ unobserved common factors. At time period $T_0$, the number of factors and/or the magnitude of the factor loadings may change. We assume that there are no further breaks after $T_0$. In general, the break point $T_0$ is unknown. In Section 2.1 we introduce the data generating process, and in Section 2.2 we discuss the identification of a structural change and its date.

### 2.1. *The data generating process*

The data generating process (DGP) before $T_0$ is

$$X_t = \Lambda^0 F_t^0 + e_t, \text{ for } t=1,\dots,T_0, \tag{2.1}$$

where $\Lambda^0 \in R^{N\times r_a}$ denotes the factor loadings and $e_t \in R^N$ denotes the idiosyncratic errors. Using matrix notation, we write

$$X_a = F_a \Lambda^{0\prime} + e_a, \tag{2.2}$$

where $X_a=(X_1,\dots,X_{T_0})' \in R^{T_0\times N}$, $F_a=(F_1^0,\dots,F_{T_0}^0)' \in R^{T_0\times r_a}$, and $e_a=(e_1,\dots,e_{T_0})' \in R^{T_0\times N}$. The matrices $F_a$ and $\Lambda^0$ are both unknown and they are not separately identified.

To take into account the potential structural break in period $T_0$, we write the post-break DGP in matrix form as

$$X_b = F_{b,1}(\Lambda^0+\Gamma_1^0)' + F_{b,2}\Gamma_2^{0\prime} + e_b, \tag{2.3}$$

where $X_b=(X_{T_0+1},\dots,X_T)'$, $F_{b,1}=(F_{T_0+1}^0,\dots,F_T^0)'$, $F_{b,2}=(F_{T_0+1}^*,\dots,F_T^*)'$, and $e_b=(e_{T_0+1},\dots,e_T)'$. Here the $T_1\times r_a$ matrix $F_{b,1}$ extends the pre-break factors to the post-break period, whereas the $T_1\times(r_b-r_a)$ matrix $F_{b,2}$ collects the new factors that may emerge after the break. The matrix $\Gamma_1^0$ captures possible changes in the loadings of the pre-break factors $F_t^0$, whereas the matrix $\Gamma_2^0$ contains the loadings for the new factors $F_t^*$. The changes in the factor loadings are summarized in $\Gamma^0=\left(\Gamma_1^0, \Gamma_2^0\right)$. If the loadings of the old factors stay constant, then $\Gamma_1^0=0$. Likewise, in the absence of new factors $\Gamma_2^0=0$. After $T_0$, there are $r_b$ factors $F_b=\left(F_{b,1}, F_{b,2}\right)$ with factor loadings $\Psi^0=\left(\Lambda^0+\Gamma_1^0, \Gamma_2^0\right)$. Thus, the model in (2.3) can be equivalently written as

$$X_b = F_b \Psi^{0\prime} + e_b. \tag{2.4}$$

We now state some technical assumptions on the large sample behaviour of the factors and the loadings. These assumptions are analogous to Assumptions A and B of Bai and Ng (2002) with the modification to accommodate additional factors and changes of factor loadings at $T_0$. They ensure that all $r_a$ factors before the break and $r_b$ factors after the break make non-trivial contributions to the variance of the data.[2] For $t>T_0$, let $\overline{F}_t^0=\left(F_t^{0\prime}, F_t^{*\prime}\right)' \in R^{r_b}$ denote the $r_b$ factors after the break. Throughout this article, we use $C \in \mathbb{R}$ to denote a generic positive constant.

**Assumption A.** $\mathbb{E}[\,\|F_t^0\|^4] \le C$, $\mathbb{E}[||\overline{F}_t^0||^4] \le C$ and there exist positive definite matrices $\Sigma_F$ and $\Sigma_{\overline{F}}$ such that $T_0^{-1}\sum_{t=1}^{T_0} F_t^0 F_t^{0\prime} = \Sigma_F + O_p(T_0^{-1/2})$ and $T_1^{-1}\sum_{t=T_0+1}^{T} \overline{F}_t^0 \overline{F}_t^{0\prime} = \Sigma_{\overline{F}} + O_p(T_1^{-1/2})$. □

2. Assumption A is sufficient for the identification conditions in Assumption ID. It is also one of the sufficient conditions for consistent model selection with a known break date. For consistent model selection with an unknown break date, Assumption A is strengthened to Assumption A* in Section 2.2.

Write $\Lambda^0 = (\lambda_1^0, \ldots, \lambda_N^0)'$, where $\lambda_i^0 \in R^{r_a \times 1}$ is the factor loading for series $i$ before the break. Similarly, write $\Psi^0 = (\psi_1^0, \ldots, \psi_N^0)'$, where $\psi_i^0 \in R^{r_b \times 1}$ is the factor loading for series $i$ after the break.

**Assumption B.** (i) $\|\lambda_i^0\| \leq C$, $||\psi_i^0|| \leq C$ and there exist matrices $\Sigma_\Lambda$, $\Sigma_\Psi$ and $\Sigma_{\Lambda\Psi}$ such that $||\Lambda^{0\prime}\Lambda^0/N - \Sigma_\Lambda|| \to 0$, $||\Psi^{0\prime}\Psi^0/N - \Sigma_\Psi|| \to 0$, and $||\Lambda^{0\prime}\Psi^0/N - \Sigma_{\Lambda\Psi}|| \to 0$ as $N \to \infty$, where $\Sigma_\Lambda$ and $\Sigma_\Psi$ are positive definite. (ii) The matrices $\Sigma_\Lambda \Sigma_F$ and $\Sigma_\Psi \Sigma_{\overline{F}}$ both have distinct eigenvalues. □

The factors and their loadings in equations (2.2) and (2.4) are not separately identified. In order to develop an estimation theory for the factor model, we have to impose normalization restrictions. We rewrite the DGP as

$$X_a = F_a R_a R_a^{-1} \Lambda^{0\prime} + e_a = F_a^R \Lambda^{R\prime} + e_a, \quad X_b = F_b R_b R_b^{-1} \Psi^{0\prime} + e_b = F_b^R \Psi^{R\prime} + e_b. \tag{2.5}$$

The transformation matrices $R_a$ and $R_b$ are formally defined in the Online Appendix such that the factors have an identity covariance matrix in the sense (omitting $a$ and $b$ subscripts) that $T^{-1} F^{R\prime} F^R = I_{r \times r} + O_p(T^{-1/2})$ and the vectors of factor loadings are orthogonal and sorted according to length in the sense that $N^{-1} \Lambda^{R\prime} \Lambda^R = V_a$ and $N^{-1} \Psi^{R\prime} \Psi^R = V_b$, where $V_a$ and $V_b$ are diagonal matrices.[3]

Note that our normalization interprets changes in the law of motion of the factors $F_a$ and $F_b$ as changes in the loadings $\Lambda^R$ and $\Psi^R$. For example, consider a DGP with $r_a = r_b = 1$, constant factor loadings $\Lambda = \Psi$, and a break in the persistence of the factor, which follows an AR(1) process $F_t = \rho_a F_{t-1} + \varepsilon_t$ for $t \leq T_0$ and $F_t = \rho_b F_{t-1} + \varepsilon_t$ for $T > T_0$, where $\varepsilon_t \sim i.i.d. N(0,1)$ for all $t$. The change of the autocorrelation of $F_t$ from $\rho_a$ to $\rho_b$ in our setting translates into a change of the transformed factor loadings from $\Lambda^R = \Lambda / \sqrt{1-\rho_a^2}$ to $\Psi^R = \Lambda / \sqrt{1-\rho_b^2}$. This leads to $V_b = V_a(1-\rho_b^2)/(1-\rho_a^2)$. If $r_a = r_b > 1$, a model with changes in factor loadings can only be reparameterized to attribute all the changes to the factors if $\Lambda^0$ and $\Psi^0$ span the same column space, that is, there exists a full rank matrix $P$ such that $\Psi^0 = \Lambda^0 P$ and, therefore, $F_b \Psi^{0\prime} = (F_b P')\Lambda^{0\prime} = F_b^* \Lambda^{0\prime}$. We will come back to this issue in the empirical application in Section 7.

## 2.2. *Identification of a structural change and its date*

In the remainder of this article, we assume that the number of post-break factors is not smaller than the number of pre-break factors: $r_b \geq r_a$. If the application suggests that $r_b \leq r_a$, then labelling the subsample before $T_0$ as $X_b$ and the subsample after $T_0$ as $X_a$ maintains the validity of the proposed method. We distinguish between two types of instabilities:

$$\begin{aligned} &\text{type-1 instability}: r_b = r_a \text{ and } \Gamma_1^0 \neq 0 \\ &\text{type-2 instability}: r_b > r_a. \end{aligned} \tag{2.6}$$

Under a type-1 instability, the number of factors is constant, but there is a change in the factor loadings. For a type-2 instability, new factors appear in the model after $T_0$, while some of the loadings of the old factors also may change.

3. $T^{-1} F^{R\prime} F^R$ is not defined as an exact identity matrix because the limiting covariance matrices $\Sigma_F$ and $\Sigma_{\overline{F}}$, rather than the sample covariance matrices pre and post-break, are used to define the rotation matrices pre- and post-break, respectively. This definition ensures that the rotation matrices pre- and post-break are identical as long as $\Sigma_F = \Sigma_{\overline{F}}$ and the factor loadings are constant. In addition, the signs of the factors and loadings need to be normalized. However, because this sign normalizaton is immaterial for our analysis, we do not provide further details.

**Known break data $T_0$:** The numbers of pre- and post-break factors $r_a$ and $r_b$ are identified and can be consistently estimated using existing methods, *e.g.* the model selection criteria proposed by Bai and Ng (2002). The strict inequality $r_b > r_a$ identifies type-2 instabilities without further assumptions on the DGP. To identify type-1 instabilities, further restrictions are necessary. Given our normalization of the factor covariance matrix, a type-1 change is, intuitively, identifiable if either the space spanned by the factor loadings or the scaling of the factor loadings changes. Define a $(r_a+r_b)\times(r_a+r_b)$ augmented covariance matrix

$$\Sigma^{+}_{\Lambda\Psi} = \begin{bmatrix} \Sigma_\Lambda & \Sigma_{\Lambda\Psi} \\ \Sigma'_{\Lambda\Psi} & \Sigma_\Psi \end{bmatrix}. \tag{2.7}$$

Let $\rho_\ell(A)$ be the $\ell$-th largest eigenvalue of a square matrix $A$. The following assumption, stated in terms of the coefficients of the DGP in equations (2.2) and (2.3), is sufficient for identifying type-1 structural instabilities.

**Assumption ID.** One of the following two conditions holds:
(i) $rank(\Sigma^{+}_{\Lambda\Psi}) > r_a$;
(ii) $\rho_\ell(\Sigma_F \Sigma_\Lambda) \neq \rho_\ell(\Sigma_{\overline{F}} \Sigma_\Psi)$ for some $\ell \leq r_a$. □

Assumption ID(i) holds if and only if $\Lambda^0$ and $\Psi^0$ do not span the same column space asymptotically. It implies that the column spaces of $\Lambda^R$ and $\Psi^R$ in equation (2.5) are different. Assumption ID(ii) focuses on the scaling of the loadings and provides an alternative identification condition through the eigenvalues of $\Sigma_\Lambda \Sigma_F$ and $\Sigma_\Psi \Sigma_{\overline{F}}$. This condition does not put restrictions on the asymptotic column spaces generated by the factor loadings. It translates into $V_a \neq V_b$, where the $V$'s are the diagonal covariance matrices of the rotated pre- and post-break loadings.

**Unknown break date $T_0$:** Let $\pi_0 = T_0/T$, where $T$ is the number of periods in the sample. For simplicity, we call $\pi_0$, rather than $T_0$, the "true" break date and assume that $\pi_0 \in \Pi$, where $\Pi$ is some closed subset in the interior of $[0,1]$. For any potential break date $\pi \in \Pi$, we split the full sample into two subsamples $X_a(\pi) = (X_1,...,X_{T_a})' \in R^{T_a \times N}$ and $X_b(\pi) = (X_{T_a+1},...,X_T)' \in R^{T_b \times N}$, where $T_a = \lfloor T\pi \rfloor$ is the integer part of $T\pi$ and $T_b = T - T_a$. To obtain an identification condition for the unknown break date $\pi_0$, we now study the number of factors in $X_a(\pi)$ and $X_b(\pi)$ when $\pi \neq \pi_0$. We denote the number of factors by $r_a(\pi)$ and $r_b(\pi)$. They are defined as the number of non-vanishing eigenvalues of $(NT)^{-1}X_a(\pi)'X_a(\pi)$ and $(NT)^{-1}X_b(\pi)'X_b(\pi)$, respectively, as $N,T \to \infty$.

Building on previous results in the literature, *e.g.* Breitung and Eickmeier (2011), if the break date is misspecified, then the subsample that consists of pre- and post-break observations contains one or more additional factors. Thus, the break date can be identified by minimizing the sum of the numbers of pre- and post-break factors by varying the potential break date $\pi$. We verify the following relationship between the conjectured break date and the numbers of pre- and post-break factors in the Online Appendix:

$$r_a(\pi) = \begin{cases} r_a & \pi \leq \pi_0 \\ \text{rank}(\Sigma^{+}_{\Lambda\Psi}) & \pi > \pi_0 \end{cases} \text{ and } r_b(\pi) = \begin{cases} \text{rank}(\Sigma^{+}_{\Lambda\Psi}) & \pi < \pi_0 \\ r_b & \pi \geq \pi_0 \end{cases}, \tag{2.8}$$

where $\text{rank}(\Sigma^{+}_{\Lambda\Psi}) \geq r_b \geq r_a$ and the matrix $\Sigma^{+}_{\Lambda\Psi}$ was defined in equation (2.7). It follows from equation (2.8) that

$$r_a(\pi) + r_b(\pi) = \begin{cases} r_a + \text{rank}(\Sigma^{+}_{\Lambda\Psi}) & \pi < \pi_0 \\ r_a + r_b & \pi = \pi_0 \\ r_b + \text{rank}(\Sigma^{+}_{\Lambda\Psi}) & \pi > \pi_0 \end{cases}. \tag{2.9}$$

Because $\text{rank}(\Sigma_{\Lambda\Psi}^{+}) \geq r_b \geq r_a$, we see that $r_a(\pi)+r_b(\pi)$ is minimized at $\pi_0$, with the minimum value $r_a+r_b$. Define the set of values $\pi$ such that $r_a(\pi)+r_b(\pi)$ achieves the smallest value $r_a+r_b$ as

$$\mathcal{D}=\{\pi \in \Pi : r_a(\pi)+r_b(\pi)=r_a+r_b\}. \tag{2.10}$$

By definition, we know that $\pi_0 \in \mathcal{D}$ and hence $\mathcal{D}$ is a well-defined non-empty set. In order to ensure that $\pi_0$ is the unique minimizer of $r_a(\pi)+r_b(\pi)$, *i.e.* $\pi_0=\mathcal{D}$, we need to assume that the column space generated by $\Lambda^0$ is asymptotically not contained in the space generated by $\Psi^0$, which leads to the stronger Assumption ID*.

**Assumption ID***. $\text{rank}(\Sigma_{\Lambda\Psi}^{+}) > r_b$.

## 3. SHRINKAGE ESTIMATION

Starting point of the proposed estimation procedure is a conjectured break date $T_a$. If the break date is correctly specified then $T_a=T_0$. We define $T_b=T-T_a$. Because we treat the number of factors as unknown, we introduce a user-selected upper bound $k$ on the sum of pre- and post-break factors: $r_a+r_b \leq k$. In order to motivate the criterion function in the shrinkage estimation, we rewrite the normalized DGP in equation (2.5) as the following augmented system:

$$X_a = \begin{bmatrix} F_a^R \; F_{a,1}^{R\perp} \; F_{a,2}^{R\perp} \end{bmatrix} \begin{bmatrix} \Lambda^{R\prime} \\ 0_{(r_b-r_a)\times N} \\ 0_{(k-r_b)\times N} \end{bmatrix} + e_a = F_a^{R+}(\Lambda^{R+})' + e_a,$$

$$X_b = \begin{bmatrix} F_{b,1}^R \; F_{b,2}^R \; F_b^{R\perp} \end{bmatrix} \begin{bmatrix} \Lambda^{R\prime}+\Gamma_1^{R\prime} \\ \Gamma_2^{R\prime} \\ 0_{(k-r_b)\times N} \end{bmatrix} + e_b = F_b^{R+}(\Lambda^{R+}+\Gamma^{R+})' + e_b. \tag{3.1}$$

Here, $F_a^{R\perp}$ denotes a $T\times(k-r_a)$ orthogonal complement of $F_a^R$. We partition $F_a^{R\perp}$ into $T\times(r_b-r_a)$ and $T\times(k-r_b)$ submatrices $F_{a,1}^{R\perp}$ and $F_{a,2}^{R\perp}$. Likewise, $F_b^{R\perp}$ is an orthogonal complement of $F_b^R$. Below, we call $F_a^R$ and $F_b^R$ the "true" and $F_a^{R\perp}$ and $F_b^{R\perp}$ the irrelevant factors. In the augmented model (3.1), $\Lambda^{R+}$ and $(\Lambda^{R+}+\Gamma^{R+})$ are the factor loadings before and after the break, respectively. Estimating the number of factors and detecting instability in factor loadings can be executed simultaneously in equation (3.1), because they are equivalent to consistent selection of the zero and and non-zero components in $\Lambda^{R+}$ and $\Gamma^{R+}$. Hence, for consistent model selection, it is key to obtain estimators that can consistently distinguish zeros from non-zeros in $\Lambda^{R+}$ and $\Gamma^{R+}$. The shrinkage estimator proposed below is designed to achieve such consistency.

Although the main theoretical innovations lie in the analysis of the unknown break date case, we first present the main idea of the estimation method under the assumption that the break date is known and $T_a=T_0$. The estimation objective function and the shrinkage estimator are introduced in Section 3.1. The consistent estimation of the numbers of pre- and post-break factors and the occurrence of a break in the loadings are discussed in Section 3.2. Section 3.3 provides the extension to the unknown break date. Finally, Section 3.4 discusses the post model-selection (PMS) estimation of the factor loadings and the break date.

### 3.1. *Estimation objective function (known break date)*

The $k$ potential factors are estimated by the principal component estimator in each subsample. Specifically, for subsample $j \in \{a,b\}$, let $\widetilde{F}_j \in R^{T_j\times k}$ be the orthonormalized eigenvectors of $(NT_j)^{-1}X_jX_j'$ associated with its first $k$ largest eigenvalues. For both subsamples, estimating an overfitted model with $k$ factors gives the unrestricted least-squares estimators of the factor loading matrices $\widetilde{\Lambda}_{LS}=T_a^{-1}X_a'\widetilde{F}_a$, $\widetilde{\Psi}_{LS}=T_b^{-1}X_b'\widetilde{F}_b$ and $\widetilde{\Gamma}_{LS}=\widetilde{\Psi}_{LS}-\widetilde{\Lambda}_{LS}$. Given $\widetilde{F}_a$ and $\widetilde{F}_b$, we propose shrinkage estimators of $\Lambda^{R+}$ and $\Gamma^{R+}$ by minimizing a PLS criterion function:

$$(\widehat{\Lambda},\widehat{\Gamma})=\underset{\Lambda\in R^{N\times k},\Gamma\in R^{N\times k}}{\arg\min}\ [M(\Lambda,\Gamma)+P_1(\Lambda)+P_2(\Gamma)], \tag{3.2}$$

where

$$M(\Lambda,\Gamma) = (NT)^{-1}\Big[\big\|X_a-\widetilde{F}_a\Lambda'\big\|^2+\big\|X_b-\widetilde{F}_b(\Lambda+\Gamma)'\big\|^2\Big],$$
$$P_1(\Lambda) = \alpha_{NT}\sum_{\ell=1}^{k}\omega_\ell^{\lambda}\,\|\Lambda_\ell\| \text{ and } P_2(\Gamma)=\beta_{NT}\sum_{\ell=1}^{k}\omega_\ell^{\gamma}\,\|\Gamma_\ell\|, \tag{3.3}$$

$\Lambda_\ell$ and $\Gamma_\ell$ are the $\ell$-th column of $\Lambda$ and $\Gamma$, respectively, $\alpha_{NT}$ and $\beta_{NT}$ are two sequences of positive constants that depend on $N$ and $T$, and $\omega_\ell^{\lambda}$ and $\omega_\ell^{\gamma}$ are data-dependent weights defined as:

$$\omega_\ell^{\lambda} = \Big(N^{-1}||\widetilde{\Lambda}_\ell||^2\mathcal{I}_{\{\widetilde{\Lambda}_\ell\neq 0_{N\times 1}\}}+N^{-1}||\widetilde{\Lambda}_{\ell,LS}||^2\mathcal{I}_{\{\widetilde{\Lambda}_\ell=0_{N\times 1}\}}\Big)^{-2},$$
$$\omega_\ell^{\gamma} = \Big(N^{-1}||\widetilde{\Gamma}_\ell||^2\mathcal{I}_{\{\widetilde{\Gamma}_\ell\neq 0_{N\times 1}\}}+N^{-1}||\widetilde{\Gamma}_{\ell,LS}||^2\mathcal{I}_{\{\widetilde{\Gamma}_\ell=0_{N\times 1}\}}\Big)^{-2}. \tag{3.4}$$

Here, $\mathcal{I}_{\{x=a\}}$ is the indicator function that is equal to one if $x=a$ and equal to zero otherwise. $\widetilde{\Lambda}\in R^{N\times k}$ and $\widetilde{\Gamma}\in R^{N\times k}$ are some preliminary estimators of $\Lambda^{+}$ and $\Gamma^{+}$, where the $\ell$ subscript denotes the $\ell$-th column of the matrices.[4]

In this adaptive estimation, the data-dependent weights $\omega_\ell^{\lambda}$ and $\omega_\ell^{\gamma}$ are designed to differentiate the zero columns of $\Lambda^{R+}$ and $\Gamma^{R+}$ from the non-zero columns. Assuming that the preliminary estimators have the property that $N^{-1}||\widetilde{\Lambda}_\ell||^2\to_p 0$ if and only if the $\ell$-th column of $\Lambda^{R+}$ is zero and $N^{-1}||\widetilde{\Gamma}_\ell||^2\to_p 0$ if and only if the $\ell$-th column of $\Gamma^{R+}$ is zero, we expect $N^{-1}\|\widetilde{\Lambda}_\ell\|^2$ to converge to a positive constant for $\ell\le r_a$ and to converge to zero for $\ell>r_a$. In the latter case, $\omega_\ell^{\lambda}$ diverges to infinity, which delivers strong penalization in the shrinkage estimation (3.2) to the estimators of the zero columns in $\Lambda^0$. The weights, $\omega_\ell^{\gamma}$, have similar effects on the estimation of $\Gamma^{+}$.

The penalty functions $P_1(\Lambda)$ and $P_2(\Gamma)$, defined in terms of the column norms $\|\Lambda_\ell\|$ and $\|\Gamma_\ell\|$, are group-LASSO penalties (cf., Yuan and Lin, 2006). A group-LASSO estimator either sets all the elements in a group equal to zero or estimates them as non-zeros altogether. This feature is particularly useful for large-scale factor models because the irrelevant factors have zero factor loadings for all series. As such, the group-LASSO estimator automatically controls the group-wise model-selection error over all series, which is challenging if the model selection is performed series by series. The solution to the minimization problem in equation (3.2) can be computed efficiently, because it is a convex optimization problem after the first $k$ principal components of the data set have been calculated.

### 3.2. *Consistent model selection (known break date)*

The shrinkage estimator defined above is used to determine the numbers of pre- and post-break factors and to detect the occurrence of type-1 and type-2 structural changes. Let $\mathcal{B}_0\in\{0,1\}$ be a

4. The simplest preliminary estimators are the unrestricted least-squares estimators $\widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma}_{LS}$. Other preliminary estimators may set columns of $\Lambda$ or $\Gamma$ equal to zero, which is why we introduced the indicator function notation.

binary variable such that $\mathcal{B}_0=0$ indicates that there is no structural break (*i.e.* $\Gamma^{(0)}=(\Gamma_1^0,\Gamma_2^0)=0$ in equation (2.3)). If $\mathcal{B}_0=1$ and $r_a=r_b$, then the DGP exhibits a type-1 instability. $\mathcal{B}_0=1$ and $r_a<r_b$ corresponds to a type-2 instability. For the remainder of this article, we refer to a model as a collection of DGPs that are associated with the triplet

$$\mathcal{M}_0=(\mathcal{B}_0,r_a,r_b). \tag{3.5}$$

We propose consistent estimation of $\mathcal{M}_0$ based on the simultaneous estimation of $\mathcal{B}_0$, $r_a$, and $r_b$. For the consistent determination of $\mathcal{B}_0$, it suffices to estimate the normalized version of the factor model in equation (2.5), because $\Gamma^0=0$ if and only if $\Gamma^R=0$, where $\Gamma^R=(\Gamma_1^R,\Gamma_2^R)$ are defined by rewriting the normalized version of the post-break DGP in equation (2.5) as

$$X_b=F_b^R\Psi^{R\prime}+e_b=F_{b,1}^R(\Lambda^R+\Gamma_1^R)'+F_{b,2}^R\Gamma_2^{R'}+e_b. \tag{3.6}$$

Estimation of $\mathcal{M}_0$ is based on the column norms of $\widehat{\Lambda}$ and $\widehat{\Gamma}$. The estimator of the break indicator $\mathcal{B}_0$ is given by

$$\widehat{\mathcal{B}}=\mathcal{I}_{\{\|\widehat{\Gamma}\|>0\}}. \tag{3.7}$$

The estimators of $r_a$ and $r_b$ are obtained by finding the last non-zero columns of $\widehat{\Lambda}$ and $\widehat{\Gamma}$:

$$\begin{aligned}\widehat{r}_a &= \min\left\{j:||\widehat{\Lambda}_\ell||^2=0 \text{ for all } \ell>j\right\}\\ \widehat{r}_b &= \max\left(\hat{r}_a,\min\left\{j:||\widehat{\Gamma}_\ell||^2=0 \text{ for all } \ell>j\right\}\right).\end{aligned} \tag{3.8}$$

The model selected by the shrinkage estimator is

$$\widehat{\mathcal{M}}=(\widehat{\mathcal{B}},\widehat{r}_a,\widehat{r}_b). \tag{3.9}$$

In Section 5, we formally show that

$$\Pr(\widehat{\mathcal{M}}=\mathcal{M}_0)\to 1 \text{ as } N,T\to\infty \tag{3.10}$$

provided that the tuning parameters $\alpha_{NT}$ and $\beta_{NT}$ are chosen within the bounds specified below. Even for a known break date, our procedure differs from the existing methods in some very important dimensions. First, our method not only detects a structural break but also automatically determines its type. Secondly, to detect a break in factor loadings, our method does not require knowledge of the number of factors before and/or after the break. Instead, it determines the pre- and post-break factors structures simultaneously.

### 3.3. *Estimation and model selection with unknown break date*

If the break date is unknown, the factor model has to be estimated for a range of hypothetical break dates $\pi\in\Pi=[\underline{\pi},\bar{\pi}]$. Formally, we assume that $\underline{\pi}>0$ and $\bar{\pi}<1$. However, in practice, the break dates cannot be too close to the boundaries of zero and one, because it is difficult to estimate the factor model on samples with a very small time dimension. Following the literature on the estimation of models with unknown break dates, we recommend to set $\underline{\pi}\geq 0.15$ and $\bar{\pi}\leq 0.85$.

Let $\widetilde{F}_a(\pi)\in R^{T_a\times k}$ be the orthonormalized eigenvectors of $(NT_a)^{-1}X_a(\pi)X_a(\pi)'$ associated with its first $k$ largest eigenvalues. Similarly, let $\widetilde{F}_b(\pi)\in R^{T_b\times k}$ be the orthonormalized left eigenvectors of $(NT_b)^{-1}X_b(\pi)X_b(\pi)'$ associated with its first $k$ largest eigenvalues.

The unrestricted estimators of the factor loadings are $\widetilde{\Lambda}_{LS}(\pi) = T_a^{-1} X_a(\pi)' \widetilde{F}_a(\pi)$, $\widetilde{\Psi}_{LS}(\pi) = T_b^{-1} X_b(\pi)' \widetilde{F}_b(\pi)$, and $\widetilde{\Gamma}_{LS}(\pi) = \widetilde{\Psi}_{LS}(\pi) - \widetilde{\Lambda}_{LS}(\pi)$.

By applying the procedure in Sections 3.1 and 3.2 with $\pi_0$ replaced by $\pi$, we obtain a shrinkage estimator indexed by $\pi \in \Pi$, which yields consistent estimators of $r_a(\pi)$ and $r_b(\pi)$ for any $\pi \in \Pi$. In preliminary work, we found that this simple procedure is undesirable because the estimators of $r_a(\pi)$ and $r_b(\pi)$ are highly sensitive to $\pi$. To stabilize the estimator in finite samples, we propose the following shrinkage estimator with averaging penalty:

$$(\widehat{\Lambda}(\pi), \widehat{\Gamma}(\pi)) = \underset{\Lambda \in R^{N\times k}, \Gamma \in R^{N \times k}}{\arg\min} \left[ M(\Lambda, \Gamma; \pi) + P_1^*(\Lambda) + P_2^*(\Gamma) \right], \tag{3.11}$$

where

$$M(\Lambda, \Gamma; \pi) = (NT)^{-1} \left[ \left\| X_a(\pi) - \widetilde{F}_a(\pi) \Lambda' \right\|^2 + \left\| X_b(\pi) - \widetilde{F}_b(\pi)(\Lambda + \Gamma)' \right\|^2 \right]. \tag{3.12}$$

This estimator depends on $\pi$ only through the least-squares criterion function. The averaging penalty functions $P_1^*(\Lambda)$ and $P_2^*(\Lambda)$ are

$$P_1^*(\Lambda) = \sum_{\ell=1}^{k} \mathbb{E}_\xi [\alpha_{NT}(\xi) \omega_\ell^{\lambda *}(\xi)] \| \Lambda_\ell \|, \quad P_2^*(\Gamma) = \sum_{\ell=1}^{k} \mathbb{E}_\xi [\beta_{NT}(\xi) \omega_\ell^{\gamma *}(\xi)] \| \Gamma_\ell \|, \tag{3.13}$$

where $\xi$ has a uniform distribution on $\Pi$ and $\mathbb{E}_\xi[\cdot]$ denotes the expectation with respect to $\xi$.[5] In practice, $\Pi$ is approximated by a set of equally spaced grid points $\Pi_d$, and the expectation in equation (3.13) is replaced by an average.

The tuning parameters $\alpha_{NT}(\pi)$ and $\beta_{NT}(\pi)$ are two sequences of constants that depend on $N$ and $T$ for each $\pi$ and the sequences can vary with $\pi$. The specific choices are provided in Section 4.4. For each $\pi \in \Pi$, let $\widetilde{\Lambda}(\pi)$, $\widetilde{\Psi}(\pi)$, and $\widetilde{\Gamma}(\pi)$ be some preliminary estimators. We define adaptive weights $\omega_\ell^{\lambda *}(\pi)$ and $\omega_\ell^{\gamma *}(\pi)$ as

$$\begin{aligned}
\omega_\ell^{\lambda *}(\pi) &= \left( N^{-1} ||\widetilde{\Lambda}_\ell(\pi)||^2 \mathcal{I}_{\{\widetilde{\Lambda}_\ell(\pi) \neq 0_{N\times 1}\}} + N^{-1} ||\widetilde{\Lambda}_{\ell,LS}(\pi)||^2 \mathcal{I}_{\{\widetilde{\Lambda}_\ell(\pi) = 0_{N\times 1}\}} \right)^{-2}, \\
\omega_\ell^{\gamma *}(\pi) &= \left( N^{-1} \min \left\{ ||\widetilde{\Gamma}_\ell(\pi)||^2, ||\widetilde{\Psi}_\ell(\pi)||^2 \right\} \mathcal{I}_{\{\widetilde{\Gamma}_\ell(\pi) \neq 0_{N\times 1}\}} \right)^{-2} \\
&\quad + \left( N^{-1} \min \left\{ ||\widetilde{\Gamma}_{\ell,LS}(\pi)||^2, ||\widetilde{\Psi}_{\ell,LS}(\pi)||^2 \right\} \mathcal{I}_{\{\widetilde{\Gamma}_\ell(\pi) = 0_{N\times 1}\}} \right)^{-2}.
\end{aligned} \tag{3.14}$$

Comparing the weights in equation (3.14) with those in equation (3.4), we see that $\omega_\ell^{\lambda *}(\pi_0) = \omega_\ell^{\lambda}$ but $\omega_\ell^{\gamma *}(\pi_0) \neq \omega_\ell^{\gamma}$. If the break date is unknown, it is crucial to use $\omega_\ell^{\gamma *}(\pi)$ for consistent estimation of $r_b$ because, for $\pi > \pi_0$ and $\ell > r_b$, $N^{-1}||\widetilde{\Psi}_{\ell,LS}(\pi)||^2$ converges (in probability) to 0, but $N^{-1}||\widetilde{\Gamma}_{\ell,LS}(\pi)||^2$ may not converge (in probability) to 0. Thus, the modified adaptive weights can deliver larger penalties, when needed.

5. By definition,

$$\mathbb{E}_\xi[\alpha_{NT}(\xi)\omega_\ell^{\lambda}(\xi)] = \int_{\underline{\pi}}^{\overline{\pi}} \alpha_{NT}(\xi)\omega_\ell^{\lambda}(\xi) \frac{1}{\overline{\pi} - \underline{\pi}} d\xi \text{ and } \mathbb{E}_\xi[\beta_{NT}(\xi)\omega_\ell^{\gamma}(\xi)] = \int_{\underline{\pi}}^{\overline{\pi}} \beta_{NT}(\xi)\omega_\ell^{\gamma}(\xi) \frac{1}{\overline{\pi} - \underline{\pi}} d\xi,$$

where $\underline{\pi}$ and $\overline{\pi}$ are the lower and upper bounds of $\Pi$. Note that the above two terms depend on $N$ and $T$.

The model specification estimator $\widehat{\mathcal{M}}^*=(\widehat{\mathcal{B}}^*,\widehat{r}_a^*,\widehat{r}_b^*)$ can be obtained as follows. First, let

$$\widehat{\mathcal{B}}^*=\mathcal{I}_{\{\sup_{\pi\in\Pi}||\widehat{\Gamma}(\pi)||>0\}} \tag{3.15}$$

Secondly, the number of pre- and post-break factors can be estimated according to

$$\widehat{r}_a^*=\min_{\pi\in\Pi}\widehat{r}_a(\pi) \quad \text{and} \quad \widehat{r}_b^*=\min_{\pi\in\Pi}\widehat{r}_b(\pi), \tag{3.16}$$

where $\widehat{r}_a(\pi)$ and $\widehat{r}_b(\pi)$ are defined as in equation (3.8), replacing $\widehat{\Lambda}$ and $\widehat{\Gamma}$ by $\widehat{\Lambda}(\pi)$ and $\widehat{\Gamma}(\pi)$, respectively. In Section 5, we show that

$$\Pr(\widehat{\mathcal{M}}^*=\mathcal{M}_0)\to 1 \text{ as } N,T\to\infty \tag{3.17}$$

for the suggested choice of the tuning parameters. To the best of our knowledge this is the first estimator of the "true" number of factors that is robust to both type-1 and type-2 instabilities at an unknown date. In addition, it detects instabilities in a large number of time series ($N\to\infty$) as a group.

### 3.4. *PMS estimation*

In addition to estimating the model specification our shrinkage estimator also provides an estimate of the loading matrices $\Lambda$ and $\Gamma$. However, because the penalty terms of the estimator are not optimized to estimate the non-zero coefficients of the loading matrices efficiently (*e.g.* in a mean-squared error (MSE) sense), we recommend to re-estimate the loadings using least-squares conditional on the selected model specification. We refer to the resulting estimator as PMS estimator.

If $\widehat{\mathcal{B}}^*=0$ (no break) then the factor model should be re-estimated on the full sample.[6] In this case, let $\widetilde{F}\in R^{T\times k}$ be the (orthonormalized) first $k$ principal components constructed from the full sample. Let $\overline{\Lambda}$ denote the first $\hat{r}_a^*$ columns of the full-sample least-squares estimator $\widetilde{\Lambda}_{LS}=T^{-1}X'\widetilde{F}$ and set $\overline{\Psi}=\overline{\Lambda}$.[7] Alternatively, if $\widehat{\mathcal{B}}^*=1$, then the factors and the loadings should be re-estimated for the two subsamples separately. Let $\widetilde{F}_a(\pi)$ and $\widetilde{F}_b(\pi)$ denote the factor estimates for the two subsamples. Moreover, let $\overline{\Lambda}(\pi)$ be the first $\widehat{r}_a^*$ columns of the least-squares estimator $\widetilde{\Lambda}_{LS}(\pi)=T^{-1}X_a'(\pi)\widetilde{F}_a(\pi)$ and $\overline{\Psi}(\pi)$ be the first $\widehat{r}_b^*$ columns of $\widetilde{\Psi}_{LS}(\pi)=T^{-1}X_b'(\pi)\widetilde{F}_b(\pi)$. The PMS estimator can be defined as follows

$$\widehat{\Lambda}_{\text{PMS}}(\pi)=\left(\overline{\Lambda}(\pi),0_\Lambda\right) \quad \text{and} \quad \widehat{\Psi}_{\text{PMS}}(\pi)=\left(\overline{\Psi}(\pi),0_\Psi\right), \tag{3.18}$$

where $0_\Lambda$ is a $N\times(k-\widehat{r}_a^*)$ zero matrix, and $0_\Psi$ is a $N\times(k-\widehat{r}_b^*)$ zero matrix.

Building on work by Bai (1997), we will show below that the break date $\pi_0$ can be estimated consistently by using a least-squares objective function. Let

$$\widehat{\pi}=\operatorname*{argmin}_{\pi\in\Pi} Q_{NT}(\pi;\widehat{r}_a^*,\widehat{r}_b^*), \tag{3.19}$$

6. We adopt the notation for the unknown break date case. For the known break date case, one can simply drop the *-superscripts and the $(\pi)$-arguments.

7. Because the columns of $\widetilde{F}$ are orthogonal by construction, $\overline{\Lambda}$ is identical of the OLS estimator obtained by regressing $X$ on the first $\widehat{r}_a^*$ columns of $\widetilde{F}$.

where

$$Q_{NT}(\pi;\widehat{r}_a^*,\widehat{r}_b^*) \tag{3.20}$$
$$= (NT)^{-1}\left[\left\|X_a(\pi)-\widetilde{F}_a(\pi)\widehat{\Lambda}'_{\text{PMS}}(\pi)\right\|^2+\left\|X_b(\pi)-\widetilde{F}_b(\pi)\widehat{\Psi}'_{\text{PMS}}(\pi)\right\|^2\right].$$

In practice, this estimator should only be computed if the shrinkage estimator detects a break, *i.e.* $\widehat{\mathcal{B}}=1$.

## 4. PRACTICAL GUIDANCE FOR IMPLEMENTATION

We first introduce the estimation algorithm with a known break date and extend it to the unknown break date subsequently. Section 4.1 provides a practical procedure for choosing the tuning parameters $\alpha_{NT}$ and $\beta_{NT}$. Section 4.2 describes a two-step shrinkage estimation procedure proposed in which the second-stage tuning improves the finite-sample performance of the estimation procedure while maintaining its asymptotic validity. A cross-validation procedure to fine-tune the penalty weights is presented in Section 4.3. Finally, Section 4.4 discusses the choice of penalty weights and the two-step estimation algorithm if the break date is unknown.

### 4.1. *Choosing the penalty weights (known break date)*

The penalty functions $P_1(\Lambda)$ and $P_2(\Gamma)$ depend in addition to $\omega_\ell^\lambda$ and $\omega_\ell^\gamma$ also on the tuning parameters $\alpha_{NT}$ and $\beta_{NT}$. Roughly speaking, $\alpha_{NT}$ is the weight attached to the penalty on the coefficients related to $X_a$, whereas $\beta_{NT}$ is the penalty weight on the coefficients of $X_b$. We suggest choosing these factors as

$$\alpha_{NT}=\kappa_1 N^{-1/2}C_{NT_a}^{-3} \text{ and } \beta_{NT}=\kappa_2 N^{-1/2}C_{NT_b}^{-3}, \tag{4.1}$$

where $\kappa_1$ and $\kappa_2$ are two constants, $C_{NT_a}=\min(N^{1/2},T_a^{1/2})$, and $C_{NT_b}=\min(N^{1/2},T_b^{1/2})$. These rates are justified by the asymptotic results in Section 5. Specifically, Theorem 2 in Section 5 states that consistent estimation of the model requires $\alpha_{NT}$ and $\beta_{NT}$ to converge to 0 at least as fast as $N^{-1/2}C_{NT}^{-1}$ and slower than $N^{-1/2}C_{NT}^{-5}$. In practice, we choose $\alpha_{NT}$ and $\beta_{NT}$ to balance these two rates and replace the overall sample size $T$ by the subsample sizes $T_a$ and $T_b$. We set $\kappa_1$ and $\kappa_2$ equal to

$$\kappa_1 = c_1\left\{(NT_a)^{-1/2}\left\|e_a(\widetilde{\Lambda})\right\|+(NT_b)^{-1/2}\left\|e_b(\widetilde{\Lambda}+\widetilde{\Gamma})\right\|\right\} \tag{4.2}$$
$$\kappa_2 = c_2(NT_b)^{-1/2}\left\|e_b(\widetilde{\Lambda}+\widetilde{\Gamma})\right\|,$$

where $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ are preliminary estimators and the residual matrices $e_a(\Lambda)$ and $e_b(\Lambda+\Gamma)$ are defined as

$$e_a(\Lambda)=X_a-\widetilde{F}_a\Lambda' \text{ and } e_b(\Lambda+\Gamma)=X_b-\widetilde{F}_b(\Lambda+\Gamma)'. \tag{4.3}$$

A justification for this choice is provided in the Online Appendix. Our default choice for the constants $c_1$ and $c_2$ is $c_1=c_2=1$, but we develop a cross-validation procedure to fine-tune these constants over a fixed interval in finite samples.

### 4.2. *Two-step estimation procedure (known break date)*

We recommend a two-step estimation procedure. The preliminary estimator obtained in the first step is used to fine-tune the penalty terms of the second-step shrinkage estimator. The two-step procedure improves the finite sample performance through two channels. First, the tuning parameters are better calibrated in the second step because the residual matrices in equation (4.2) are more accurate when $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ are based on a first-step model selection rather than the estimation of an unrestricted model with $k$ factors. Secondly, a better preliminary estimator $\widetilde{\Gamma}$ is obtained through a rotation of the factor loadings $\Lambda^R$ and $\Psi^R$. For $i=1$ and 2, let $\widetilde{\Lambda}^{(i)}$, $\widetilde{\Psi}^{(i)}$, and $\widetilde{\Gamma}^{(i)}$ denote the preliminary estimators; $\widehat{\Lambda}^{(i)}$, $\widehat{\Psi}^{(i)}$ and $\widehat{\Gamma}^{(i)}$ denote the PLS estimators; and $\widehat{\Lambda}^{(i)}_{\text{PMS}}$, $\widehat{\Psi}^{(i)}_{\text{PMS}}$ and $\widehat{\Gamma}^{(i)}_{\text{PMS}}$ denote the PMS estimators in Step $i$. The two-step estimation can be implemented with the following algorithm:

**Algorithm 1 Two-step estimation procedure**.

*(1) First-stage shrinkage estimation:*

*(1.1) Compute the unrestricted least-squares estimators $\widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma}_{LS}$.*

*(1.2) Let $\widetilde{\Lambda}^{(1)}=\widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma}^{(1)}=\widetilde{\Gamma}_{LS}$. Calculate $\omega^{\lambda}_{\ell}$, $\omega^{\gamma}_{\ell}$, $\alpha_{NT}$ and $\beta_{NT}$ following equations (3.4), (4.1), and (4.2) with $\widetilde{\Lambda}=\widetilde{\Lambda}^{(1)}$ and $\widetilde{\Gamma}=\widetilde{\Gamma}^{(1)}$.*

*(1.3) Compute the shrinkage estimator $\widehat{\Lambda}^{(1)}$ and $\widehat{\Gamma}^{(1)}$ by minimizing the criterion function (3.2).*

*(1.4) Estimate $r_a$ and $r_b$ based on equation (3.8) with $\widehat{\Lambda}=\widehat{\Lambda}^{(1)}$ and $\widehat{\Gamma}=\widehat{\Gamma}^{(1)}$. Call the estimators $\widehat{r}_a^{(1)}$ and $\widehat{r}_b^{(1)}$.*

*(1.5) Construct subsample PMS estimators $\widehat{\Lambda}^{(1)}_{PMS}$ and $\widehat{\Psi}^{(1)}_{PMS}$ using the definition in equation (3.18). If $\widehat{r}_b^{(1)}=\widehat{r}_a^{(1)}$, transform the columns of $\overline{\Psi}^{(1)}$ as follows: let $\overline{\Lambda}^{(1)'}\overline{\Psi}^{(1)}=UDV'$ be the singular value decomposition of $\overline{\Lambda}^{(1)'}\overline{\Psi}^{(1)}$. Define the transformed factor loading as*

$$\overline{\Psi}^{(1)}_R=\overline{\Psi}^{(1)}Q, \tag{4.4}$$

*where $Q=VU'$. Define the modified PMS estimator of $\Psi$ as*

$$\widehat{\Psi}^{(1)}_{PMS-R}=\left(\overline{\Psi}^{(1)}_R,\, 0_{\Psi^{(1)}}\right)\in R^{N\times k}. \tag{4.5}$$

*(2) Second-stage shrinkage estimation:*

*(2.1) Let*

$$\widetilde{\Lambda}^{(2)}=\widehat{\Lambda}^{(1)}_{\text{PMS}},\ \widetilde{\Psi}^{(2)}=\begin{cases}\widehat{\Psi}^{(1)}_{\text{PMS}-R} & \text{if } \widehat{r}_b^{(1)}=\widehat{r}_a^{(1)}\\ \widehat{\Psi}^{(1)}_{\text{PMS}} & \text{if } \widehat{r}_b^{(1)}>\widehat{r}_a^{(1)}\end{cases},\ \widetilde{\Gamma}^{(2)}=\widetilde{\Psi}^{(2)}-\widetilde{\Lambda}^{(2)} \tag{4.6}$$

*and calculate $\omega^{\lambda}_{\ell}$, $\omega^{\gamma}_{\ell}$, $\alpha_{NT}$, and $\beta_{NT}$ following equations (3.4), (4.1), and (4.2) with $\widetilde{\Lambda}=\widetilde{\Lambda}^{(2)}$ and $\widetilde{\Gamma}=\widetilde{\Gamma}^{(2)}$.*

*(2.2) Compute the shrinkage estimators $\widehat{\Lambda}^{(2)}$ and $\widehat{\Gamma}^{(2)}$ by minimizing the criterion function (3.2).*

*(2.3) Compute* $\widehat{\mathcal{B}}_0^{(2)}$, $\widehat{r}_a^{(2)}$, *and* $\widehat{r}_b^{(2)}$ *based on (3.8)–(3.9) with* $\widehat{\Lambda}=\widehat{\Lambda}^{(2)}$ *and* $\widehat{\Gamma}=\widehat{\Gamma}^{(2)}$.

*(2.4) Conditional on the selected model* $\widehat{\mathcal{M}}^{(2)}=\big(\widehat{\mathcal{B}}_0^{(2)},\widehat{r}_a^{(2)},\widehat{r}_b^{(2)}\big)$ *construct the PMS estimators* $\widehat{\Lambda}_{\mathrm{PMS}}^{(2)}$ *and* $\widehat{\Psi}_{\mathrm{PMS}}^{(2)}$ *using the definition in equation (3.18).*

The preliminary estimators in the second step are based on the PMS estimators of the first step. The rotation in Step 1.5 minimizes the risk of falsely reporting a structural break when there is no instability in the data. It is designed to match the column spaces of $\overline{\Lambda}^{(1)}$ and $\overline{\Psi}^{(1)}$. This leads to a smaller $\widetilde{\Gamma}$ if $\Gamma^0=0$. While this rotation may also reduce the probability of reporting a "true" break, we found in our simulation experiments that overall it leads to an improved finite-sample performance. The rotation does not affect the asymptotic validity of our procedure. Formally, the problem is to find an orthogonal matrix $Q$ such that $||\overline{\Lambda}^{(1)}-\overline{\Psi}^{(1)}Q||_2$ is minimized. This is an orthogonal procrustes problem. It is equivalent to maximizing the correlation between the columns of $\overline{\Lambda}^{(1)}$ and $\overline{\Psi}^{(1)}Q$. The solution is $Q=VU'$ (see Schönemann, 1966), where $U$ and $V$ are obtained from the singular value decomposition $\overline{\Lambda}^{(1)'}\overline{\Psi}^{(1)}=UDV'$. In Section 5, we show that if there is indeed a type-1 instability, the $Q$ rotation will not eliminate the difference between $\Lambda_\ell^R$ and $\Psi_\ell^R$. Moreover, we show that the asymptotic theory we established in the previous section applies to the two-step shrinkage estimator.

### 4.3. *Cross validation*

We recommend to fine-tune the constants $c=(c_1,c_2)\in\mathcal{C}$ that appear in the penalty weights in equation (4.2) using a cross-validation procedure. Because we are operating in an environment in which the regressors, *i.e.* the factors, are unobserved, we need to partition the sample in both the cross-sectional as well as the time-series dimension. A graphical illustration of the algorithm is provided in Figure 1 and a formal description of the of the algorithm can be found in the Online Appendix.

We first partition the data matrix in the cross-sectional dimension, creating disjoint subsamples $X_{(-j_N)}$ ($N$-regression) and $X_{(j_N)}$ ($N$-prediction). Given a particular $c$ we apply the model selection procedure to $X_{(-j_N)}$, which yields an estimated model specification $\widehat{\mathcal{M}}(-j_N,c)$, and estimate the unobserved factors. We then partition the hold-out sample $X_{(j_N)}$ along the $T$ dimension into regression and prediction samples. If $\widehat{\mathcal{M}}(-j_N,c)$ corresponds to a specification with structural break, then the regression and prediction samples are constructed separately for the pre- and post-break periods. Using the factor estimates from the $X_{(-j_N)}$ sample as "observed" regressors, we estimate the factor loadings based on the regression sample using OLS. Given the estimates of the factors and the loadings we can then generate pseudo-out-of-sample forecasts (and forecast errors) for the prediction sample. Throughout, we condition on the previously selected model specification $\widehat{\mathcal{M}}(-j_N,c)$. We use separate rolling pseudo-out-of-sample forecasting schemes for the pre- and post-break samples.

Our cross-validation criterion is based on mean-squared forecast errors (MSFE). The tuning constants are chosen to minimize the MSFE with respect to $c$. It is important that this minimization is carried out over a bounded set $\mathcal{C}$. According to the theoretical results presented in Section 5, the model selection procedure is consistent for each fixed $c\in\mathcal{C}$ as $(N,T)\to\infty$. Our cross-validation procedure helps researchers to fine-tune the constant to achieve good finite-sample performance—as we will demonstrate in the Monte Carlo experiments presented in Section 6.

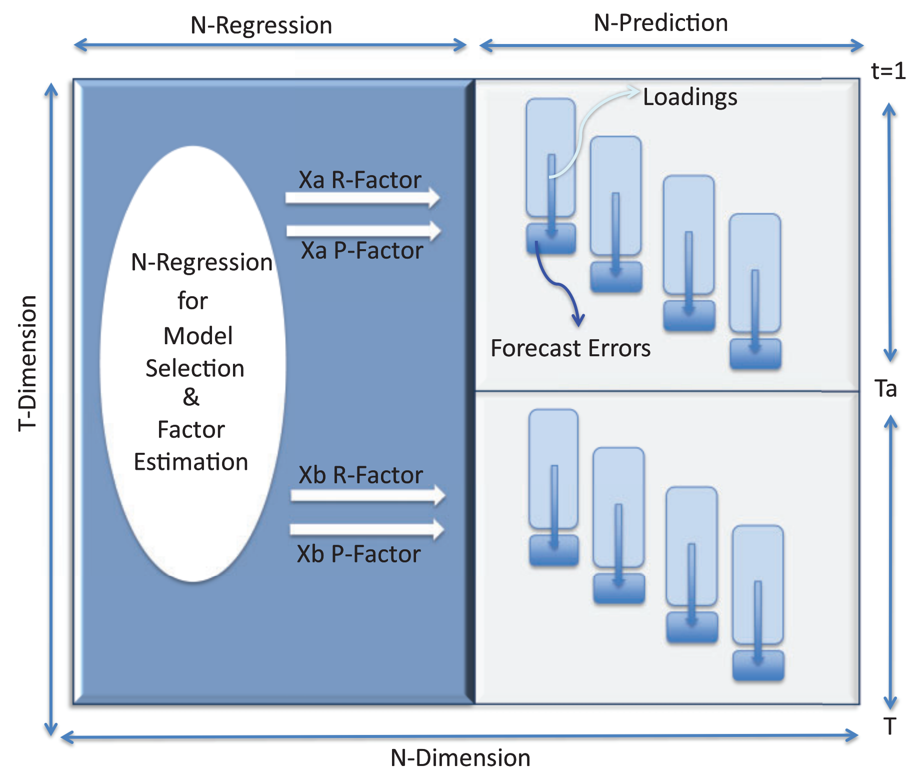


FIGURE 1

Graphical representation of Algorithm 2.

### 4.4. *Implementation for unknown break date*

Next, we extend the estimation algorithm to the case with an unknown break date. The recommended tuning parameters are

$$\alpha_{NT}(\pi)=\kappa_1(\pi)N^{-1/2}C_{NT_a}^{-3} \text{ and } \beta_{NT}(\pi)=\kappa_2(\pi)N^{-1/2}C_{NT_b}^{-3}, \tag{4.7}$$

where $\kappa_1(\pi)\in[\underline{\kappa}_1,\overline{\kappa}_1]$ and $\kappa_2(\pi)\in[\underline{\kappa}_2,\overline{\kappa}_2]$ for some $\underline{\kappa}_1,\underline{\kappa}_2>0$ and $\overline{\kappa}_1,\overline{\kappa}_2<\infty$. They are analogous to those in equation (4.1). In practice, we can choose $\kappa_1(\pi)$ and $\kappa_2(\pi)$ as in equation (4.2) but with $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ replaced by $\widetilde{\Lambda}(\pi)$ and $\widetilde{\Gamma}(\pi)$, respectively.

As in the known break date case, we consider a two-step procedure to estimate the true model. Follow the steps in Section 4.2 by setting $\pi_0=\pi$, $\widetilde{\Lambda}^{(1)}(\pi)=\widetilde{\Lambda}_{LS}(\pi)$, $\widetilde{\Psi}^{(1)}(\pi)=\widetilde{\Psi}_{LS}(\pi)$, and $\widetilde{\Gamma}^{(1)}(\pi)=\widetilde{\Gamma}_{LS}(\pi)$; replacing $\omega_\ell^\lambda$, $\omega_\ell^\gamma$, $\alpha_{NT}$, and $\beta_{NT}$ with $\omega_\ell^{\lambda*}(\pi)$, $\omega_\ell^{\gamma*}(\pi)$, $\alpha_{NT}(\pi)$, and $\beta_{NT}(\pi)$, respectively; replacing the PLS criterion (3.2) with (3.11); and replacing the estimators $\widehat{r}_a$ and $\widehat{r}_b$ in equation (3.8) with those in equation (3.16). Note that the first-step estimators $\widehat{r}_a^{(1)}$ and $\widehat{r}_b^{(1)}$ do not vary with $\pi$ following the definition in equation (3.16). Therefore, one should first obtain the first-step estimator $\widehat{\Lambda}^{(1)}(\pi)$ and $\widehat{\Gamma}^{(1)}(\pi)$ for each $\pi\in\Pi$ and get $\widehat{r}_a^{(1)}$ and $\widehat{r}_b^{(1)}$, and then obtain the second-step estimator $\widehat{\Lambda}^{(2)}(\pi)$ and $\widehat{\Gamma}^{(2)}(\pi)$ for each $\pi\in\Pi$. The selected model $\widehat{\mathcal{M}}^*$ is based on the two-step PLS estimator $\widehat{\Lambda}^{(2)}(\pi)$ and $\widehat{\Gamma}^{(2)}(\pi)$ following the specifications in Section 3.3.

The cross-validation procedure also has a straightforward extension to the unknown break date case. We choose a common $c$ for all potential break dates. For each $\pi$, the $X_{(-j_N)}$ subsamples are designed in the same way as in the known break date case, with $\pi_0$ replaced by $\pi$. For each $c$ we obtain a selected model, which does not depend on $\pi$ by definition. From the validation sample $X_{(j_N)}$ we first eliminate the observations that lie outside of the conjectured break interval $\Pi$ and then proceed with Step 1.4 of Algorithm 2.

In order to use the proposed shrinkage estimator, the user has to make four choices: the maximum number of potential factors $k$, the break date interval $\Pi$, the domain $\mathcal{C}$ for the tuning constants, and the number of sample partitions $n_N$ and $n_T$. In practice, the choice of $k$ and $\Pi$ is likely to be based on some preliminary examination of the data. For instance, many researchers have estimated the number of factors in variants of the Stock and Watson (2012) data set, which can help choosing $k$. Asset pricing theory often gives some indication of how many factors to expect in panels of financial data. Choosing an unreasonably high value of $k$ delivers a large number of potential regressors and may lead to a deterioration of the performance of the shrinkage estimator. If $\widehat{r}_b = k$ that might indicate that $k$ was chosen too small. The choice of $\Pi$ is closely tied to the application. The interval could be centred around 1984 if the goal is to detect breaks associated with the Great Moderation; or centred around 2007 if the Great Recession is the topic of interest. Choosing an overly large interval is likely to lead to a deterioration of the performance of our estimator. Finally, we provide a particular choice for $\mathcal{C}$, $n_N$, and $n_T$ that performs well in our Monte Carlo study under a variety of data generating processes.

## 5. ASYMPTOTIC THEORY

We now provide a formal asymptotic theory for the proposed shrinkage estimator. We first consider the case of known break date in Section 5.1 and then generalize the results to allow for an unknown break date in Section 5.2.

### 5.1. *Known break date*

To derive the asymptotic behaviour of the PLS estimator and establish that the proposed model selection procedure is consistent some additional assumptions are necessary. First, we need to control the degree of time-series and cross-sectional dependence in the idiosyncratic errors as well as the degree of dependence between the factors and the idiosyncratic errors. Here, we follow the literature and make assumptions that are analogous to Assumptions C and D of Bai and Ng (2002). These assumptions are formally stated in the Online Appendix.

Secondly, we will make high-level assumptions on the large sample properties of the preliminary estimators $\tilde{\Lambda}$ and $\tilde{\Gamma}$ and on the convergence rates of the sequences $\alpha_{NT}$ and $\beta_{NT}$. We begin with the assumptions on the stochastic order of the preliminary estimators, which affect the data-dependent weights $\omega_\ell^\lambda$ and $\omega_\ell^\gamma$ defined in equation (3.4). Define $C_{NT} = \min(T^{1/2}, N^{1/2})$, where $C_{NT}$ is the convergence rate of the unrestricted least-squares estimator in Bai and Ng (2002).

**Assumption P1.** The preliminary estimators $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ satisfy

(i) $\Pr(N^{-1}||\widetilde{\Lambda}_\ell||^2 \geq C) \to 1$ for $\ell = 1, \ldots, r_a$, $N^{-1}||\widetilde{\Lambda}_\ell||^2 = O_p(C_{NT}^{-2})$ for $\ell = r_a+1, \ldots, k$;

(ii) If $\Gamma^0 \neq 0$, $\Pr(N^{-1}||\widetilde{\Gamma}_\ell||^2 \geq C) \to 1$ for $\ell = 1, ..., r_b$, $N^{-1}||\widetilde{\Gamma}_\ell||^2 = O_p(C_{NT}^{-2})$ for $\ell = r_b + 1, \ldots, k$;

(iii) If $\Gamma^0 = 0$, $N^{-1}||\widetilde{\Gamma}_\ell||^2 = O_p(C_{NT}^{-2})$ for $\ell = 1, ..., k$. □

**Assumption P2.** Assumption P1 holds with $\widetilde{\Lambda} = \widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma} = \widetilde{\Gamma}_{LS}$. □

Under the conditions in Assumption P1, the columns of the preliminary estimators are divided into two categories. For the first category, $\Pr(N^{-1}||\widetilde{\Lambda}_\ell||^2 \geq C) \to 1$ and $\Pr(N^{-1}||\widetilde{\Gamma}_\ell||^2 \geq C) \to 1$ such that the data-dependent weights, $\omega_\ell^\lambda$ and $\omega_\ell^\gamma$, are stochastically bounded. For the second category, $N^{-1}||\widetilde{\Lambda}_\ell||^2 = O_p(C_{NT}^{-2})$ and $N^{-1}||\widetilde{\Gamma}_\ell||^2 = O_p(C_{NT}^{-2})$, which implies that $\omega_\ell^\lambda$ and $\omega_\ell^\gamma$ diverge in probability faster than $C_{NT}^4$. These large penalties in the second category yield shrinkage estimators that are equal to 0 w.p.a.1. Assumption P1 is imposed on any preliminary estimators of $\Lambda^R$ and $\Gamma^R$. In the second step of the two-step estimator (Algorithm 1), the preliminary estimators are different from $\widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma}_{LS}$. However, Assumption P2 is still necessary because $\omega_\ell^\lambda$ and $\omega_\ell^\gamma$ depend on $\widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma}_{LS}$ whenever $\widetilde{\Lambda}$ or $\widetilde{\Gamma}$ has zero columns. Note that $\widetilde{\Lambda}_\ell = 0$ is a special case of $N^{-1}||\widetilde{\Lambda}_\ell||^2 = O_p(C_{NT}^{-2})$ in Assumption P1, and the same argument applies to $\widetilde{\Gamma}_\ell$.

While the data-dependent weights $\omega_\ell^\lambda$ and $\omega_\ell^\gamma$ determine the relative penalties of different columns of factor loadings, the tuning parameters $\alpha_{NT}$ and $\beta_{NT}$ determine the overall penalization. We make the following assumptions about the rates at which the tuning parameters vanish asymptotically.

**Assumption T.** The tuning parameters $\alpha_{NT}$ and $\beta_{NT}$ satisfy

(i) $\alpha_{NT} = O(N^{-1/2}C_{NT}^{-1})$ and $\beta_{NT} = O(N^{-1/2}C_{NT}^{-1})$;

(ii) $N^{-1/2}C_{NT}^{-5} = o(\alpha_{NT})$ and $N^{-1/2}C_{NT}^{-5} = o(\beta_{NT})$. □

Assumption T imposes bounds on the tuning parameters $\alpha_{NT}$ and $\beta_{NT}$. These bounds control the magnitudes of penalization on all columns and are designed for consistent model selection. The upper bound in Assumption T(i) ensures that if the data-dependent weights $\omega_\ell^\lambda$ and $\omega_\ell^\gamma$ are stochastically bounded, the penalties on the non-zero columns are small such that the shrinkage bias is negligible asymptotically. On the other hand, we aim to shrink the estimators of zero columns to zero. For this purpose, the lower bound in Assumption T(ii) requires that the tuning parameters $\alpha_{NT}$ and $\beta_{NT}$ converge to zero not too fast. The choice of $\alpha_{NT}$ and $\beta_{NT}$ made in Section 4.1 satisfies Assumption T.

We are now in a position to state the asymptotic limits of the PLS estimators $\widehat{\Lambda}$ and $\widehat{\Gamma}$. The estimators converge to the coefficients of the normalized version of the DGP in equation (2.5). As before, let the subscript $\ell$ denote the $\ell$-th column of a matrix.

**Theorem 1.** *Suppose Assumptions* A, B, C *(see Online Appendix),* D *(see Online Appendix),* P1–P2*, and* T *hold. Then,*

*(a) pre-break loadings of relevant factors:* $N^{-1}||\widehat{\Lambda}_\ell - \Lambda_\ell^R||^2 = O_p(C_{NT}^{-2})$ *for* $\ell = 1,\ldots,r_a$*;*

*(b) pre-break loadings of irrelevant factors:* $\Pr(||\widehat{\Lambda}_\ell||^2 = 0$ *for* $\ell = r_a+1,\ldots,k) \to 1$*;*

*(c) post-break changes in loadings of relevant factors: if* $\Gamma^0 \neq 0$, $N^{-1}||\widehat{\Gamma}_\ell - \Gamma_\ell^R||^2 = O_p(C_{NT}^{-2})$ *for* $\ell = 1,\ldots,r_b$*;*

*(d) no-break: If* $\Gamma^0 = 0$, $\Pr(||\widehat{\Gamma}_\ell||^2 = 0$ *for* $1,\ldots,r_b) \to 1$*;*

*(e) post-break changes in loadings of irrelevant factors:* $\Pr(||\widehat{\Gamma}_\ell||^2 = 0$ *for* $\ell = r_b+1,\ldots,k) \to 1$.

Parts (a) and (b) of Theorem 1 characterize the limits of the PLS estimators of the pre-break factor loadings. Due to the penalization, the factor loadings of the irrelevant factors are estimated as exactly 0 w.p.a.1. This superefficiency result cannot be achieved by the unrestricted least square estimators. In contrast, for the true factors, the penalization does not affect the consistency and the convergence rate of their estimators. For $\ell = 1,\dots,r_a$, the PLS estimator $\widehat{\Lambda}_\ell$ converges in probability to the factor loadings $\Lambda_\ell^R$ of the transformed DGP. Parts (c)–(e) of Theorem 1 characterize asymptotic properties of the PLS estimators of the changes in the factor loadings, which is essential to detecting structural instabilities. In the absence of structural instabilities, the PLS estimators of the changes are equal to 0 w.p.a.1. In the presence of a structural instability, the superefficiency in part (e) of Theorem 1 only applies to the redundant factors, which pins down the number of factors after the break.

Thus far, we showed that the factor loadings of the irrelevant factors are estimated as zeros w.p.a.1. We also showed that the changes in the loadings of the relevant factors are estimated as zero w.p.a.1, if their loadings are not subjected to any instability. Hence, to establish the model selection consistency for the PLS estimation, it is sufficient to show that the asymptotic limits $N^{-1}\|\Lambda_\ell^R\|^2$ and $N^{-1}\|\Gamma_\ell^R\|^2$ in parts (a) and (c) of Theorem 1 are bounded away from zero, which requires the identification Assumption ID. The consistency result is stated in the following theorem.

**Theorem 2.** *Suppose Assumptions* A, B, C *(see Online Appendix),* D *(see Online Appendix),* ID, P1, P2*, and* T *hold. Then the estimated model is consistent:*

$$\Pr(\widehat{\mathcal{M}} = \mathcal{M}_0) \to 1 \textit{ as } N,T \to \infty.$$

Theorem 2 provides model selection consistency for the shrinkage estimation based on any set of preliminary estimators that satisfy Assumptions P1 and P2. If the unrestricted least-squares estimators are used as preliminary estimators, our model selection procedure is consistent under a set of primitive conditions that do not involve Assumptions P1 and P2.

**Corollary 1.** *If* $(\widetilde{\Lambda},\widetilde{\Gamma}) = (\widetilde{\Lambda}_{LS},\widetilde{\Gamma}_{LS})$*, then Theorem 2 holds under Assumptions* A, B, C *(see Online Appendix),* D *(see Online Appendix),* ID*, and* T.

We now extend the consistency result in Theorem 2 to the two-step estimator described in Algorithm 1. It can be shown that under our assumptions, in the presence of a type-1 change, there exists a set of columns such that

$$\mathcal{Z} = \{\ell : N^{-1}||\Gamma_\ell^R||^2 = N^{-1}||\Psi_\ell^R - \Lambda_\ell^R||^2 \geq C\}. \tag{5.8}$$

The columns in the set $\mathcal{Z}$ are crucial for the identification of a type-1 instability. Due to the orthogonal rotation in Step 1.5 of Algorithm 1, we need the following additional assumption:

**Assumption R**. If $r_a = r_b$, then $\inf_{\|\mathrm{w}\|=1} N^{-1}||\Psi^R \mathrm{w} - \Lambda_\ell^R||^2 \geq C$ for $\ell \in \mathcal{Z}$. □

Assumption R is not restrictive. It holds whenever $\Lambda_\ell^R$ is not in the column space generated by $\Psi^R$. Assumption R is imposed on the loadings $\Lambda^R$ of the normalized version of the DGP rather than on the loadings $\Lambda^0$ of the DGP itself. Assumption R allows the loadings of some of the "structural" factors in the unnormalized DGP to remain constant while the loadings of other "structural" factors change. In the absence of structural instabilities, $\mathcal{Z}$ is empty and Assumption R is not necessary. Using Assumption R, the general consistency result established in Theorem 2 can be extended to the two-step estimation procedure described in Section 4.2, as summarized in the following corollary:

**Corollary 2.** *If* $\widetilde{\Lambda} = \widetilde{\Lambda}^{(2)}$ *and* $\widetilde{\Gamma} = \widetilde{\Gamma}^{(2)}$*, then Theorem 2 holds under Assumptions* A, B, C *(see Online Appendix),* D *(see Online Appendix),* ID, R*, and* T.

### 5.2. *Unknown break date*

Next, we show that the shrinkage estimator based on the averaging penalty in equation (3.13) yields consistent estimation of the model. The tuning parameters and the two-step estimation algorithm follow the specifications in Section 4.4. For the case with an unknown break date, we establish the consistency of $\widehat{\mathcal{M}}^*$ directly rather than by first establishing the asymptotic behaviour of the shrinkage estimator $\widehat{\Lambda}(\pi)$ and $\widehat{\Gamma}(\pi)$ for all $\pi$ as in Theorem 1. The main reason is that the shrinkage estimator with the averaging penalty does not yield consistent estimation of $r_a(\pi)$ and $r_b(\pi)$ for all $\pi$. The averaging penalty tends to over-penalize for $\pi \neq \pi_0$. However, we can still obtain consistent estimation of $r_a$ and $r_b$ because $r_a \leq r_a(\pi)$ and $r_b \leq r_b(\pi)$.

To show that the two-step PLS estimator described in Section 4.4 yields consistent estimation of the model, we strengthen Assumption R to take into account the unknown break date and the averaging penalty. For any $\pi \in \Pi$, we can write the normalized system as

$$\begin{aligned} X_a(\pi) &= F_a^R(\pi)\Lambda^R(\pi)' + e_a(\pi), \\ X_b(\pi) &= F_b^R(\pi)\Psi^R(\pi)' + e_b(\pi), \end{aligned} \tag{5.9}$$

where $F_a^R(\pi)$ and $\Lambda^R(\pi)$ are $T_a \times (r_a + r_b)$ and $N \times (r_a + r_b)$ matrices, respectively, and $F_b^R(\pi)$ and $\Psi^R(\pi)$ are $T_b \times (r_a + r_b)$ and $N \times (r_a + r_b)$ matrices, respectively.

**Assumption R*.** (i) If $r_a = r_b$, then $\inf_{\pi \in \Pi} \inf_{\|\mathrm{w}\|=1} N^{-1} || \Psi^R(\pi)\mathrm{w} - \Lambda_\ell^R(\pi) ||^2 \geq C$ for $\ell \in \mathcal{Z}$; (ii) if $r_b > r_a$, then $\inf_{\pi > \pi_0} N^{-1} || \Psi_\ell^R(\pi) - \Lambda_\ell^R(\pi) ||^2 \geq C$ for $\ell = r_b$. □

Assumption R*(i) generalizes Assumption R from $\pi = \pi_0$ to any $\pi \in \Pi$. Assumption R*(ii) is not necessary if the break date $\pi_0$ is known because $\Lambda_\ell^R(\pi_0) = 0$ for $\ell = r_b > r_a$. Similar to Assumption R, we do not view the modified Assumption R* as restrictive because in many applications the matrices $\Lambda_\ell^R(\pi)$ and $\Psi_\ell^R(\pi)$ are transformations of loading matrices for factors with a structural interpretation. Assumptions R and R* are compatible in applications where the loadings of some "structural" factors change and the loadings of the other "structural" factors do not. The following theorem states that even with an unknown break date we can still estimate the occurrence of a break and the numbers of pre- and post-break factors consistently.

**Theorem 3.** *Suppose that Assumptions* A* *(see Online Appendix),* B*,* C* *(see Online Appendix),* D *(see Online Appendix),* ID*, and* R* *hold. Then the model selected by the two-step estimator in Algorithm 1 is consistent:*

$$\Pr(\widehat{\mathcal{M}}^* = \mathcal{M}_0) \to 1 \text{ as } N, T \to \infty.$$

The proof strategy of Theorem 3 is different from that of Theorem 2 due to the averaging penalty. To establish consistency of $\widehat{\mathcal{M}}^*$, it is sufficient to show that (1) if $\pi = \pi_0$, the shrinkage estimator with the averaging penalty behaves similarly to that in Theorem 1 and (2) if $\pi \neq \pi_0$, the estimated model is not smaller than the true model. In this regard, the identification result in Section 2.2 is used constructively. Whenever $\pi$ substantially differs from $\pi_0$, the averaging penalty tends to over-penalize those loadings that would be set to zero for $\pi = \pi_0$. This means that there is a tendency to underestimate either $r_a(\pi)$ or $r_b(\pi)$ if the conjectured break point is incorrect. At $\pi = \pi_0$, the averaging penalty is smaller than the pointwise penalty, but still sufficiently large to ensure consistency.

Using the estimates $\widehat{r}_a^*$ and $\widehat{r}_b^*$, the least-squares objective function in equation (3.19) delivers a consistent estimate of the break date.

**Theorem 4.** *Suppose that Assumptions* A* *(see Online Appendix),* B, C* *(see Online Appendix),* D *(see Online Appendix),* ID*, *and* R* *hold. Then,* $\widehat{\pi} \to_p \pi_0$ *as* $N,T \to \infty$.

The consistency of $\widehat{\pi}$ shown by Theorem 4 complements the consistency of $\widehat{\mathcal{M}}^* = (\widehat{\mathcal{B}}^*, \widehat{r_a}^*, \widehat{r_b}^*)$. Deriving the asymptotic distribution of $\widehat{\pi}$ requires analysing the generated regressors issue due to latent factors. Some results along this line are developed in Bai (2003) and Bai and Ng (2006). Incorporating these results in break-point estimation is left for future research.[8]

## 6. MONTE CARLO SIMULATIONS

In this section, we conduct Monte Carlo simulations to illustrate the accuracy of the proposed model selection procedure, and the MSEs of the shrinkage estimators and the PMS estimators in finite samples. Section 6.1 describes the DGPs and the estimators used in the experiments. The simulation results are presented in Section 6.2 and we report some comparisons to existing factor selection and structural break test procedures in Section 6.3.

### 6.1. *Design*

The design of the DGPs roughly follows that in Bates *et al*. (2013), with the additional flexibility to accommodate both type-1 and type-2 instabilities and the shift of focus from small breaks to large breaks. The DGP takes the form

$$\begin{aligned} \text{pre-break: } & X_{it} = \lambda_i' F_t + e_{it}, \quad F_{t,\ell} = \rho_a F_{t-1,\ell} + u_{t,\ell}, \\ & t = 1,\ldots,\lfloor T\pi_0 \rfloor, \quad \ell = 1,\ldots,r_a, \\ \text{post-break: } & X_{it} = \psi_i' \overline{F}_t + e_{it}, \quad \overline{F}_{t,\ell} = \rho_b \overline{F}_{t-1,\ell} + u_{t,\ell}, \\ & t = \lfloor T\pi_0 \rfloor + 1,\ldots,T, \quad \ell = 1,\ldots,r_b, \end{aligned} \tag{6.1}$$

where $i = 1,\ldots,N$, $F_t = (F_{t,1},\ldots,F_{t,r_a})'$, $\overline{F}_t = (\overline{F}_{t,1},\ldots,\overline{F}_{t,r_b})'$, and $u_{t,\ell} \sim N(0,1)$. To model the temporal and cross-sectional dependence of the idiosyncratic errors, we consider

$$e_{it} = \alpha e_{it-1} + v_{it}, \; v_t = (v_{1t},...,v_{Nt})' \sim N(0,\Omega), \tag{6.2}$$

where the $(i,j)$-th element of $\Omega$ is $\beta^{|i-j|}$. The processes $\{u_{t,\ell} : \ell = 1,...,r_b\}$ and $\{v_{it}\}$ are mutually independent and are i.i.d. across $t$. All innovations are normally distributed. The initial values $F_0$ and $e_0 = (e_{10},...,e_{N0})'$ are drawn from their stationary distribution. When $r_b = r_a$, $\overline{F}_{T_0} = F_{T_0}$. When $r_b > r_a$, $\overline{F}_{T_0} = (F_{T_0}', F_{T_0}^{*'})'$, where each element of $F_{T_0}^*$ is drawn independently from the distribution of $F_{t,\ell}$. The parameters $\{N,T,\pi_0,r_a,r_b,\rho_a,\rho_b,\alpha,\beta\}$ are specified below.

The pre-break factor loadings $\{\lambda_i : i = 1,\ldots,N\}$ are independent across $i$ and independent of the factors and the idiosyncratic errors. Let $\lambda_i \sim N(0,\Sigma_i)$, where $\Sigma_i$ is a diagonal matrix with diagonal elements $\sigma_i^2(1),\ldots,\sigma_i^2(r_a)$. These diagonal elements are distinct to ensure that Assumption ID holds, and their sum controls the population regression $R^2$ of $X_{it}$ on the factors. To this end, we

8. As is common in the time-series literature, we proved a consistency result for $\pi_0$ instead of $T_0 = \pi_0 T$. In the context of a panel data model with observed regressors, Bai (2010) provides a consistency result with respect to $T_0$. Recently, Baltagi *et al*. (2015), in the context of a factor model, showed that $\hat{T}_0 - T_0 = O_p(1)$. While it might be possible to strengthen our result to a statement about $\hat{T}_0$, it is beyond the scope of this article.

set $\sigma_i^2(\ell)=0.9^{(\ell-1)}\sigma_i^2(1)$ and $\sum_{\ell=1}^{r_a}\sigma_i^2(\ell)=\sigma^*(R_i^2)$, where the scalar $\sigma^*(R_i^2)$ is chosen such that $\mathbb{E}[(\lambda_i' F_t)^2]/\mathbb{E}[X_{it}^2]=R_i^2$ for $t\le T_0$ and $R_i^2$ is the pre-specified regression $R^2$ of the $i$-th series.[9]

We consider two different ways of choosing $R_i^2$ for $i=1,\ldots,N$. One is the homogeneous case of $R_i^2=0.5$, which is considered in Bai and Ng (2002) to assess their information criteria and the benchmark DGP in our simulations. Another is the heterogeneous case in which $R_i^2$ is calibrated to match the distribution of $R^2$ values in the data sets used in the empirical applications. Taking the data set before December 2007, which is the conjectured break date of the recent recession, we regress each time-series variable on the principal component estimators of five factors and obtain the empirical distribution of the regression $R^2$. We then draw $R_i^2$ for $i=1,\ldots,N$ independently from this empirical distribution and use the realized $R_i^2$ to construct the pre-break factor loadings $\lambda_i$.

Depending on the type of the instabilities, we consider two different ways of constructing the post-break factor loadings $\psi_i$. For a type-1 instability, we set $\psi_i=(1-\mathbf{w})\lambda_i+\mathbf{w}\lambda_i^*$, where $\lambda_i^*$ and $\lambda_i$ are independent and have the same distribution. We vary the scalar $\mathbf{w}$ to control the size of the instability, with $\mathbf{w}=0$ corresponding to the special case of no break in the factor loadings. For a type-2 instability, $\psi_i$ is drawn independently of everything else with a distribution that is similar to that of $\lambda_i$, except that $r_a$ is changed to $r_b$, $\mathbb{E}[(\psi_i'\overline{F}_t)^2]/\mathbb{E}[X_{it}^2]=R_i^2$ for $t>T_0$, and the post-December 2007 subsample is used to calibrate $R_i^2$ in the heterogeneous $R^2$ case.

We normalize the simulated time series to have zero means and unit variance before using principal components analysis to extract a maximum of $k=8$ potential factors from either the subsamples or the full sample.[10] For experiments with known break dates, the model selection is based on the two-step PLS estimator described in Algorithm 1. The cross-validation Algorithm 2 is used to choose the tuning constants $c_1$ and $c_2$ from the set

$$\mathcal{C}=\left\{\frac{1}{6},\frac{1}{4},\frac{1}{2},\frac{3}{4},1,2,3\right\}\otimes\left\{\frac{1}{6},\frac{1}{4},\frac{1}{2},\frac{3}{4},1,2,3\right\}, \tag{6.3}$$

where $\otimes$ here denote the Cartesian product. We set $n_N=5$ and we choose $n_T=10$.[11] In general, the cross-sectional division is computationally more costly because the model selection procedure has to be applied to each cross-sectional regression sample. Conditional on the selected model, the time-series rolling window forecast is fast because it only requires least-squares regressions with orthogonalized regressors.

For simulations in which the break date is not assumed to be known, the model selection and estimation are based on modified versions of Algorithms 1 and 2, described in Section 4.4, where $\Pi$ is approximated by a discrete set $\Pi_d$. The grid size in $\Pi_d$ is $\tau=0.01$, a shift by a quarter for a monthly data set of 300 periods, like the data set in our empirical application. We consider $\Pi_d=\{\pi_c-4\tau,\pi_c-3\tau,\ldots,\pi_c,\ldots,\pi_c+3\tau,\pi_c+4\tau\}$, which spans a two-year interval and is symmetric around the true break date $\pi_0$. To define a post-break subsample for the PMS estimator, the least-squares estimator of the break date described in Section 3.4 is used because $\pi_0$ is unknown.

In addition to assessing the probability of selecting the correct model specification, we also compute MSFE for out-of-sample forecasts generated by the selected model. The series to be

9. The choice is $\sigma^*(R_i^2)=\frac{1-\rho_a^2}{(1-\alpha^2)}\frac{R_i^2}{1-R_i^2}$.

10. As a robustness check we also report some results for $k=12$ in the Online Appendix.

11. Due to computational constraints, we did not conduct an extensive exploration for the optimal choice of $n_N$ and $n_T$. In the context of structural break tests for factor augmented forecasting models, Corradi and Swanson (2014) considered different sample splitting schemes in the time-series dimension.

TABLE 1
*Monte Carlo experiments*

| Exp. | $R^2$ | $\pi_0$ | $\alpha, \beta$ | Break date |
|---|---|---|---|---|
| 1 | Homogeneous | 0.5 | 0.2 | Known |
| 2 | Heterogeneous | 0.8 | 0.2 | Known, unknown |
| 3 | Homogeneous | 0.5 | 0.5 | Known |

forecast follows the law of motion

$$\begin{array}{ll} \text{pre-break:} & y_{t+1} = \varphi_a' F_t + \epsilon_{t+1}, \quad t = 1, \ldots, T_a \\ \text{post-break:} & y_{t+1} = \varphi_b' \bar{F}_t + \epsilon_{t+1}, \quad t = T_a + 1, \ldots, T_a + T_b. \end{array} \tag{6.4}$$

The $\epsilon_t$'s are i.i.d. $N(0,1)$ distributed and independent of the processes $\{u_{t,\ell}\}$ and $v_{it}$. The loading vector is generated from the distribution $\varphi_a \sim N(0, I_{r_a})$. In a stable model, $\varphi_b = \varphi_a$. For a type-1 change, $\varphi_b = (1-\mathbf{w})\varphi_a + \mathbf{w}\varphi_a^*$, where $\varphi_a^*$ and $\varphi_a$ are independent and have the same distribution. For a type-2 change, $\varphi_b$ is drawn independently according to $\varphi_b \sim N(0, I_{r_b})$. The out-of-sample forecasts are generated as follows. We first determine the selected model and the factors based on the $X$ sample. Secondly, under the no-break scenario we estimate $\varphi_b = \varphi_a$ based on the full sample $t = 1, \ldots, T_a + T_b - 1$ and evaluate the MSFE associated with the prediction of $y_{T_a+T_b+1}$. Under break scenario we estimate $\varphi_b$ based on the subsample $t = T_a + 1, \ldots, T_a + T_b - 1$ and evaluate the MSFE associated with the prediction of $y_{T_a+T_b+1}$.

We compare the MSFE of the predictor based on the PMS estimator to the MSFE of a predictor that is based on full-sample estimation. The full-sample estimator is defined as the first $r$ columns of the full-sample least-squares estimator $\widetilde{\Lambda}_{LS} = T^{-1} X' \widetilde{F}$, where $r = r_a$ if $\mathcal{B}_0 = 0$ (no break) and $r = r_a + r_b$ if $\mathcal{B}_0 \neq 0$ (break).

## 6.2. *Results for shrinkage estimator*

In the remainder of Section 6, we present results from three types of Monte Carlo experiments, which are summarized in Table 1. In the first experiment, the regression $R^2$ is homogeneous across all series, the break date is located in the middle of the sample ($\pi_0 = 0.5$) and the cross-sectional correlation is modest. The second experiment is considerably more challenging as it is designed to mimic the problem: the regression $R^2$ is heterogeneous across the series and the break takes place towards the end of the sample ($\pi_0 = 0.8$). The third experiment is similar to the first, but the cross-sectional correlation among the series is stronger. We conduct Experiments 1 and 3 under the assumption that the break date is known and consider the known and unknown break date case for Experiment 2. In all three experiments, we set the temporal correlation to $\rho_a = \rho_b = 0.5$. All results reported below are based on averages over 1,000 Monte Carlo.

We begin with an illustration of the cross-validation procedure for Experiment 2 (known break date) in Figure 2. The figure depicts the probability of selecting the correct model specification in a setting in which no break occurs and in a setting with type-2 structural change. The solid horizontal line in each panel indicates the success rate of cross-validation Algorithm 2, whereas the other lines correspond to different combinations of $c_1$ and $c_2$ taken from the set $\mathcal{C}$ defined in equation (6.3). The figure shows that some choices of the tuning constants, *e.g.* $c_1 = 1/6$ or $c_1 = 3$, can lead to poor performance of the shrinkage-based model selection procedure. Other choices, *e.g.* $c_1 = 1/2$ or $c_1 = 1$, combined with $c_2 \geq 2$ for the no change case or $1/2 \leq c_2 \leq 1$ for the type-2 change case, lead to perfect model selection. Our cross-validation algorithm is able to select $c_1$ and $c_2$ values that lead to the selection of the correct specification with probability close to one in this experiment.

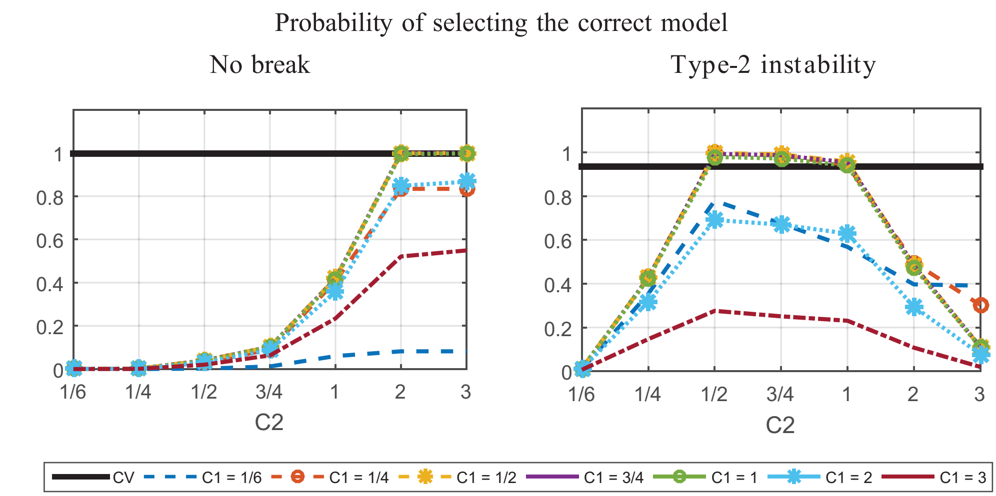


FIGURE 2

Choice of tuning constants and cross validation.

*Notes:* Known break date; heterogeneous $R^2$, $\pi_0 = 0.8$; cross-sectional correlation $\alpha = \beta = 0.2$; temporal correlation $\rho_a = \rho_b = 0.5$; sample size $N = 150$, $T = 150$; true number of factors $r_a = 3$, and $r_b = 3$ (no break) or $r_b = 4$ (type-2 instability). The solid horizontal line in each panel indicates the success rate of cross-validation Algorithm 2.

TABLE 2

*Known break date, homogeneous $R^2$, $\pi_0 = 0.5$*

| DGP configuration | | | | | Model selection | | | Relative |
|---|---|---|---|---|---|---|---|---|
| $r_a$ | $r_b$ | **w** | $N$ | $T$ | $\Pr(\widehat{\mathcal{M}} = \mathcal{M})$ | $\Pr(\widehat{r}_a = r_a)$ | $\Pr(\widehat{r}_b = r_b)$ | MSFE |
| Panel A: No break | | | | | | | | |
| 3 | 3 | 0 | 100 | 100 | 1.00 | 1.00 | 1.00 | 1.00 |
| 3 | 3 | 0 | 150 | 150 | 1.00 | 1.00 | 1.00 | 1.00 |
| 3 | 3 | 0 | 200 | 200 | 1.00 | 1.00 | 1.00 | 1.00 |
| Panel B: Type-1 instability | | | | | | | | |
| 3 | 3 | 0.2 | 100 | 100 | 0.00 | 1.00 | 1.00 | 0.99 |
| 3 | 3 | 0.2 | 150 | 150 | 0.00 | 1.00 | 1.00 | 0.98 |
| 3 | 3 | 0.2 | 200 | 200 | 0.00 | 1.00 | 1.00 | 1.00 |
| 3 | 3 | 0.5 | 100 | 100 | 0.79 | 0.99 | 0.99 | 0.94 |
| 3 | 3 | 0.5 | 150 | 150 | 1.00 | 1.00 | 1.00 | 0.94 |
| 3 | 3 | 0.5 | 200 | 200 | 1.00 | 1.00 | 1.00 | 0.94 |
| Panel C: Type-2 instability | | | | | | | | |
| 1 | 2 | 0 | 100 | 100 | 1.00 | 1.00 | 1.00 | 0.97 |
| 1 | 2 | 0 | 150 | 150 | 1.00 | 1.00 | 1.00 | 0.97 |
| 1 | 2 | 0 | 200 | 200 | 1.00 | 1.00 | 1.00 | 0.99 |
| 3 | 4 | 0 | 100 | 100 | 0.89 | 1.00 | 0.90 | 0.92 |
| 3 | 4 | 0 | 150 | 150 | 1.00 | 1.00 | 1.00 | 0.93 |
| 3 | 4 | 0 | 200 | 200 | 1.00 | 1.00 | 1.00 | 0.94 |

*Notes:* Cross-sectional correlation $\alpha = \beta = 0.2$; temporal correlation $\rho_a = \rho_b = 0.5$. The last column contains the MSFE relative to a forecast based on full-sample estimation of the factors, where the number of factors is set to $r_a$ for Panel (A). and to $r_a + r_b$ for Panels (B) and (C). A number less than 1 means that the proposed PMS predictor is more accurate.

Detailed Monte Carlo results for Experiment 1 are presented in Table 2. The table contains three panels, corresponding to no break, type-1 instability, and type-2 instability, respectively. For a type-1 instability, we consider $\mathbf{w}=0.2$ and 0.5. For a type-2 instability, we consider the changes of the number of factors from 1 to 2 and 3 to 4. Various values of $N$ and $T$ are considered. We report the probability of correctly determining $\mathcal{M}$, $r_a$, and $r_b$ as well as the MSFE of the predictor based on the PMS estimator relative to the predictor based on the full-sample least-squares estimator. Values less than 1 favour our proposed PMS predictor.

Table 2 shows that our procedure is overall very accurate in estimating the model specification $\mathcal{M}_0$ if the break date is known and located in the middle of the sample. The only notable inaccuracy arises under a small type-1 change with $\mathbf{w}=0.2$. Because our selection procedure is designed to be consistent for the no-break case, it has low power against small changes in the loadings. Here the break in the factor loadings is so small, that it remains undetected. Once the break in the factor loadings is increased to $\mathbf{w}=0.5$ our procedure detects the structural change with probability 0.79 for the sample size $N=T=100$ and with probability one for the two larger sample sizes. Our procedure has generally no problem detecting the type-2 changes. Only for a change from three to four factors and a sample size of $N=T=100$ is the selection probability less than 1. Here we sometimes underestimate $r_b$.[12]

The last column of Table 2 shows the relative MSFEs. None of the relative MSFEs is greater than one, indicating that our PMS predictor weakly dominates the full-sample predictor. For the no-break design, our procedure selects the correct model specification with probability one, which means that the PMS predictor is identical to the full-sample predictor. If there is a small type-I instability, the shrinkage estimator is unable to detect the break and the PMS predictor corresponds to a full-sample predictor with $\widehat{r}_a=\widehat{r}_b=3$ factors. Our benchmark full-sample predictor, on the other hand, is based on $r_a+r_b=6$ factors. Due to the larger number of estimated parameters this predictor is slightly less accurate than the PMS predictor. If the type-I instability is large or the instability is caused by a change in the number of factors, the PMS predictor attains a substantially lower MFSE than the full-sample predictor.

Table 3 is based on Experiment 2 (unknown break date) and shows that a heterogeneous regression $R^2$ and an unknown break date make the model selection procedure less accurate. Under the no-change scenario our procedure correctly determines the model specification for all three sample sizes. Under the type-1 change we now need a larger break in the loadings ($\mathbf{w}=1$ instead of $\mathbf{w}=0.5$) for the break to be detectable and a relatively large sample size for detection probabilities above 0.9. While our procedure has no problems detecting a type-2 change from one to two factors, it has some difficulties correctly determining the number of post-break factors if the number of factors changes from three to four. However, once we increase the sample size to $N=200$ and $T=400$, the probability of selecting the correct model becomes close to one.

If the break date is unknown the ranking of our PMS predictor and the full-sample predictor is ambiguous. Under the no-break scenario the shrinkage procedure correctly determines the absence of a break and the PMS predictor is equivalent to the full-sample predictor. Under a small type-I instability the full-sample predictor leads to a slightly lower MFSE, whereas our shrinkage procedure dominates for larger sample sizes and more easily detectable breaks.

In the last column of Table 3 we are reporting the MSE for the break date estimator, which is measured in terms of number of time periods, *i.e.* $\text{MSE}(\widehat{\pi})=\mathbb{E}\left[\left(T(\widehat{\pi}-\pi_0)\right)^2\right]$. Suppose that the break date estimator is approximately unbiased and normally distributed. Then under this MSE definition a value of 1 would imply that with 95% probability the break date estimate lies

12. As a robustness check, we repeat Experiment 1 using $k=12$. The results are reported in Supplementary Table S-1 and virtually identical to those obtained for $k=8$.

TABLE 3
*Unknown break date, heterogeneous $R^2$, $\pi_0 = 0.8$*

| DGP configuration | | | | | Model selection | | | Relative | $\widehat{\pi}$ |
|---|---|---|---|---|---|---|---|---|---|
| $r_a$ | $r_b$ | $\mathbf{w}$ | $N$ | $T$ | $\Pr(\widehat{\mathcal{M}} = \mathcal{M})$ | $\Pr(\widehat{r}_a = r_a)$ | $\Pr(\widehat{r}_b = r_b)$ | MSFE | MSE |
| Panel A: No break | | | | | | | | | |
| 3 | 3 | 0 | 100 | 200 | 1.00 | 1.00 | 1.00 | 1.00 | N/A |
| 3 | 3 | 0 | 100 | 300 | 1.00 | 1.00 | 1.00 | 1.00 | N/A |
| 3 | 3 | 0 | 150 | 300 | 1.00 | 1.00 | 1.00 | 1.00 | N/A |
| Panel B: Type-1 instability | | | | | | | | | |
| 3 | 3 | 0.5 | 100 | 200 | 0.00 | 0.98 | 0.98 | 1.05 | 3.66 |
| 3 | 3 | 0.5 | 100 | 300 | 0.00 | 1.00 | 1.00 | 1.04 | 3.69 |
| 3 | 3 | 0.5 | 150 | 300 | 0.00 | 1.00 | 1.00 | 1.07 | 0.66 |
| 3 | 3 | 1 | 100 | 200 | 0.57 | 0.88 | 0.91 | 1.09 | 0.17 |
| 3 | 3 | 1 | 100 | 300 | 0.92 | 0.98 | 0.98 | 0.65 | 0.13 |
| 3 | 3 | 1 | 150 | 300 | 0.97 | 0.99 | 0.99 | 0.79 | 0.05 |
| Panel C: Type-2 instability | | | | | | | | | |
| 1 | 2 | 0 | 100 | 200 | 0.90 | 0.99 | 0.91 | 1.00 | 0.94 |
| 1 | 2 | 0 | 100 | 300 | 1.00 | 1.00 | 1.00 | 0.93 | 1.12 |
| 1 | 2 | 0 | 150 | 300 | 1.00 | 1.00 | 1.00 | 0.95 | 0.34 |
| 3 | 4 | 0 | 100 | 200 | 0.10 | 0.93 | 0.10 | 1.08 | 0.14 |
| 3 | 4 | 0 | 100 | 300 | 0.46 | 1.00 | 0.46 | 0.96 | 0.11 |
| 3 | 4 | 0 | 200 | 400 | 0.97 | 1.00 | 0.97 | 0.95 | 0.03 |

*Notes:* Cross-sectional correlation $\alpha = \beta = 0.2$; temporal correlation $\rho_a = \rho_b = 0.5$. The second-to-last column contains the MSFE relative to a forecast based on full-sample estimation of the factors, where the number of factors is set to $r_a$ for Panel (A) and to $r_a + r_b$ for Panels (B) and (C). A number less than one means that the proposed PMS predictor is more accurate. In the last column, we report $\mathrm{MSE}(\widehat{\pi}) = \mathbb{E}\big[\big(T(\widehat{\pi} - \pi_0)\big)^2\big]$.

in the interval $T_0 \pm 2$. We only report MSE for the simulation in which there is an instability (Panels (B) and (C)). In the simulations, the break date estimator is applied regardless of whether our shrinkage estimator detects a break or not. Most of the MSEs in Table 3 are less than 1 and they become smaller as the sample size increases. Overall, the break date estimates are very precise. The only exception is the case of a small type-1 change—but here we fail to correctly determine the presence of a break in the first place.

### 6.3. *Comparison with alternative procedures*

We briefly summarize a comparison of our model selection procedure with two groups of alternative procedures. Detailed numerical results are available in the Online Appendix. The first group includes procedures that estimate the number of factors in a stable model. If the break date is known, these procedures can be applied separately to the two subsamples $X_a$ and $X_b$ to estimate $r_a$ and $r_b$, respectively. These procedures do not consider the possibility of a type-1 instability. The second group includes testing procedures for the null hypothesis of the absence of a break versus the alternative hypothesis of a type-1 instability and do not produce an estimate of the number of factors. Procedures belonging to the second group also do not allow for a type-2 instability. All of the alternative procedures estimate or test some aspect of the model specification assuming that the rest of the specification is known. In contrast, our article tackles all unknown aspects of the model specification simultaneously.

We only made comparisons with an alternative procedure in cases where (1) the alternative is implementable and (2) the alternative specifies the unknown parts of the model correctly. Thus, our experimental designs generally favour the alternative procedures. Three factor estimation procedures were considered for comparison: Bai and Ng (2002; henceforth BN), Onatski (2010; henceforth ON), and Ahn and Horenstein (2013; henceforth AH). We apply each of these procedures to the pre- and post-break subsamples assuming that the break date is known. Using as an evaluation criterion the probability that $r_a$ and $r_b$, respectively, are correctly determined, we find that our shrinkage procedure tends to dominate (in some cases weakly, in other cases strongly) the alternative procedures.

We also compared our shrinkage procedure with three break tests: Breitung and Eickmeier (2011), Chen *et al*. (2014), and Han and Inoue (2014). For each test, we computed the probability of rejecting the null hypothesis that there is no break. In cases in which the null hypothesis is correct, the rejection probability of the break tests is approximately equal to the nominal size and our shrinkage procedure detects the absence of a break with probability close to one. For simulation designs in which there is a break, the ranking of the procedures generally depends on the magnitude of the break. Because consistent model selection procedures generally have low power in distinguishing very similar model specifications, our shrinkage estimator is unable to detect small changes in the loadings. The hypothesis testing procedures, by design, generate a non-zero type-1 error and in return have some local power. Our shrinkage procedure is more likely to detect large breaks than any of the competing test procedures.

## 7. STRUCTURAL CHANGES DURING THE GREAT RECESSION

Unlike in other post-war recessions, the disruption of borrowing and lending played an important role in the 2007–9 recession. Narratives emphasize a collapse of the U.S. housing market; massive devaluations of mortgage-backed securities that spilled over to other asset markets and ultimately led to a large-scale disruption of financial intermediation; a drop in real activity caused by the crisis in the financial sector; and an extended period of zero nominal interest rates in combination with unconventional monetary policy interventions. We use the shrinkage methods developed in the preceding sections to investigate the stability of factor loadings and the emergence of new factors. Section 7.1 describes the data set and the empirical findings are presented in Sections 7.2 and 7.3.

### 7.1. *Data set*

The data set used for the empirical analysis is based on Stock and Watson (2012), who compiled a set of 200 macroeconomic and financial indicators. These 200 series contain both high-level aggregates and disaggregated components. To avoid double counting, Stock and Watson retained 132 of the 200 series, and we refer to the resulting data set as SW132. Using SW132 as starting point, our data set is constructed as follows: (1) We extend the series in the SW132 data set to 2012:M12, using May 2013 data vintages. (2) We replace the quarterly series in SW132 by their monthly counterparts, if available. This is possible for consumption of non-durables, services, and durables; for non-residential investment; and for sixteen price series. We remove the remaining quarterly series for which no monthly observations are available. (3) We add two GDP components that are available at monthly frequency: change in private inventory and wage and salary disbursements. (4) Following Stock and Watson (2012), we remove local means from all series using a bi-weight kernel with a bandwidth of 100 months. The local means are approximately the same as the ones obtained by a centred moving average of $\pm 70$ months. After making these modifications, our data set consists of $N = 102$ series of monthly macroeconomic

TABLE 4
*Model selection, $T_c$ is 2007:12*

| Interval size | Factors $\widehat{r}_a$ | Factors $(\widehat{r}_b-\widehat{r}_a)$ | Break dates Least sq. | Break dates Revised | Tuning const $c_1$ | Tuning const $c_2$ | Time (min) |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 1 | 2007:M12 | 2007:M12 | 1/2 | 1 | 2 |
| 3 | 1 | 1 | 2007:M9 | 2007:M12 | 1/2 | 1 | 12 |
| 6 | 1 | 1 | 2007:M6 | 2007:M12 | 1/2 | 2 | 21 |
| 9 | 1 | 2 | 2007:M3 | 2007:M12 | 1/2 | 2 | 30 |

*Notes:* We centre the interval $\Pi$ at 2007:M12 and use the averaging penalty functions $P_1^*(\Lambda)$ and $P_2^*(\Lambda)$ defined in equation (3.13) where the average is taken over the interval 2007:M12 $\pm$ size. The run times (in minutes) are based on MATLAB code run on a single Intel Core i7 2.93 Ghz processor.

and financial indicators. The sample begins after the Great Moderation and ranges from 1985:M1 to 2013:M1 ($T=337$).

### 7.2. *The number of factors before and after 2007:M12*

The empirical analysis is based on the two-step estimation procedure described in Section 4.2. We use Algorithms 1 and 2 with the adjustments described in Section 4.4 to account for the fact that the "true" break date is unknown. Throughout the empirical analysis, we fix the number of potential factors to $k=8$ and use the cross-valudation procedure with $n_N=5$ and $n_T=10$ to choose the penalty tuning constants among the set $\mathcal{C}$ defined in equation (6.3). The model selection results are summarized in Table 4. We consider four different sets of potential break dates $\Pi$, which are centred around the conjectured break date $T_c$ = 2007:M12. $T_c$ is the beginning of the Great Recession, according to the business cycle dating of the National Bureau of Economic Research (NBER). For Size$=$0 the set $\Pi$ corresponds to a single month, 2007:M12, meaning that we are treating the break date essentially as known. For Size$=$9, the set of potential break dates ranges from 2007:M3 to 2008:M9.

For each choice of $\Pi$ we obtain a single pre-break factor ($\widehat{r}_a=1$) and two or three post-break factors. Thus, our procedure finds clear evidence of a structural change in the number of factors. In column 4 of Table 4 we report the least-squares estimate of the break date defined in equation (3.19). We minimize the least-squares criterion over the interval $\Pi$, characterized in the first column of the table. It turns out that the minimum is always attained at the boundary, which may be an indication that Assumption ID* may not hold. The analysis in Section 2.2 implies that at the "true" break date the sum of pre- and post-break factors is minimized. Thus, for each break date in a given $\Pi$ we compute $\widehat{r}_a+\widehat{r}_b$ and check whether the minimum over this interval is attained at the conjectured break date $T_c$ = 2007:M12. If it is, we set the revised break date equal to the conjectured break date. If the minimum is not attained at $T_c$, then we define the revised break date as the date closest to the conjectured break date at which the minimum is achieved. For all choices of $\Pi$ there is no evidence in the data that leads us to revise the conjectured break date. The run time on a single core processor for the largest interval $\Pi$ is approximately 30 minutes.

### 7.3. *Decomposing the structural change*

In empirical applications, it is interesting and useful to decompose type-2 changes into the contribution of the new factors and changes in the effects of old factors. While the (nonidentifiable) DGP in equations (2.2) and (2.3) provides a natural decomposition of type-2 structural changes into changes resulting from the new factors, $F_{b,2}\Gamma_2^{0'}$, and changes associated with the effect of

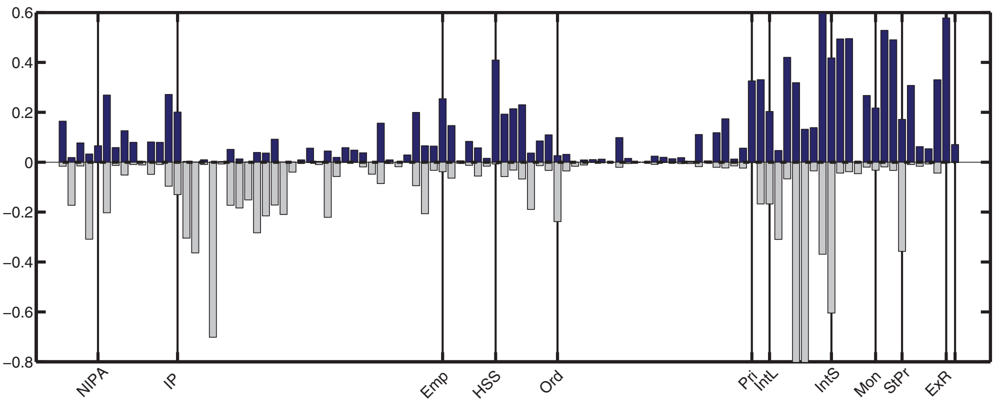


FIGURE 3

$R^2$ Gains from new loadings and factor.

*Notes:* Base line case *new loadings only*. Dark bars (above zero) indicate $R^2$ gain due to new factor relative to *new loadings only*. Light grey bars (below zero) indicate $R^2$ losses from using the old loading matrices. Results are based on $\widehat{r}_a=1$, $\widehat{r}_b=2$, and a break date of 2007:M12. The series are ordered according to the following categories: National Income and Product Accounts (NIPA), Industrial Production (IP), Employment and Unemployment (Emp), Housing Starts (HSS), Orders, Inventories, and Sales (Ord), Productivity (Pri), Interest Rate Levels (IntL), Interest Rate Spreads (IntS), Money and Credit (Mon), Stock Prices and Wealth (StPr), Exchange Rates (ExR), and Others.

the extended versions of the old factors, $F_{b,1}\Gamma_1^{0'}$, after normalizing the pre- and post-break factors to have unit variance there is no sense in which the first $\widehat{r}_a$ post-break factors can be viewed as extensions of the pre-break factors.

To obtain a meaningful decomposition, we proceed as follows. We construct an $r_b \times r_a$ matrix with orthogonal columns by maximizing the correlation between the old normalized loadings $\Lambda^R$ and the new loadings $\Psi^R\Omega_a$:

$$\Omega_a = \text{argmax}_{\tilde{\Omega}_a \in \mathcal{O}}\, tr[\Lambda^{R'}\Psi^R\tilde{\Omega}_a], \tag{7.1}$$

where $\mathcal{O}$ is the class of $r_b \times r_a$ matrices with orthonormal columns. The solution is given by $\Omega_a = VU'$, where $V$ is an $r_b \times r_a$ and $U$ an $r_a \times r_a$ orthogonal matrix obtained from the singular value decomposition $\Lambda^{R'}\Psi^R = UDV'$ (see Cliff, 1966). Let $\Omega_\perp$ be the null space of $\Omega_a'$ and define $\Omega = (\Omega_a, \Omega_\perp)$. Moreover, define the rotated loadings and factors $F_b^{R\Omega} = F^R\Omega$ and $\Psi^{R\Omega} = \Psi^R\Omega$. This rotation preserves the normalization of the factors, *i.e.* $F_b^{R\Omega'}F_b^{R\Omega}/T_b = I_{r_b \times r_b}$. Partitioning $F_b^{R\Omega} = (F_{b,1}^{R\Omega}, F_{b,2}^{R\Omega})$ and $\Psi^{R\Omega} = (\Psi_1^{R\Omega}, \Psi_2^{R\Omega})$, we can decompose $X_b$ as follows:

$$X_b = F_b^R\Omega\Omega'\Psi^{R'} + e_b = \underbrace{F_{b,1}^{R\Omega}\Lambda^{R'}}_{\text{old loadings}} + \underbrace{F_{b,1}^{R\Omega}(\Psi_1^{R\Omega} - \Lambda^R)'}_{\text{change in loadings}} + \underbrace{F_{b,2}^{R\Omega}\Psi_2^{R\Omega'}}_{\text{new factor}} + e_b. \tag{7.2}$$

As a baseline, we compute $R^2$ values for each individual series based on the variation explained by $F_{b,1}^{R\Omega}\Lambda^{R'} + F_{b,1}^{R\Omega}(\Psi_1^{R\Omega} - \Lambda^R)'$ (*new loadings only*). We compare the baseline $R^2$s to $R^2$s associated with $F_{b,1}^{R\Omega}\Lambda^{R'}$ (*old loadings*) and $R^2$s associated with all three terms in equation (7.2) (*i.e. new loadings and factor*). The results are plotted in Figure 3. Bars below the zero baseline indicate the $R^2$ loss due to ignoring the change in loadings. Bars above the zero line indicate the $R^2$ gain from also accounting for the effect of the new factor. Each set of bars

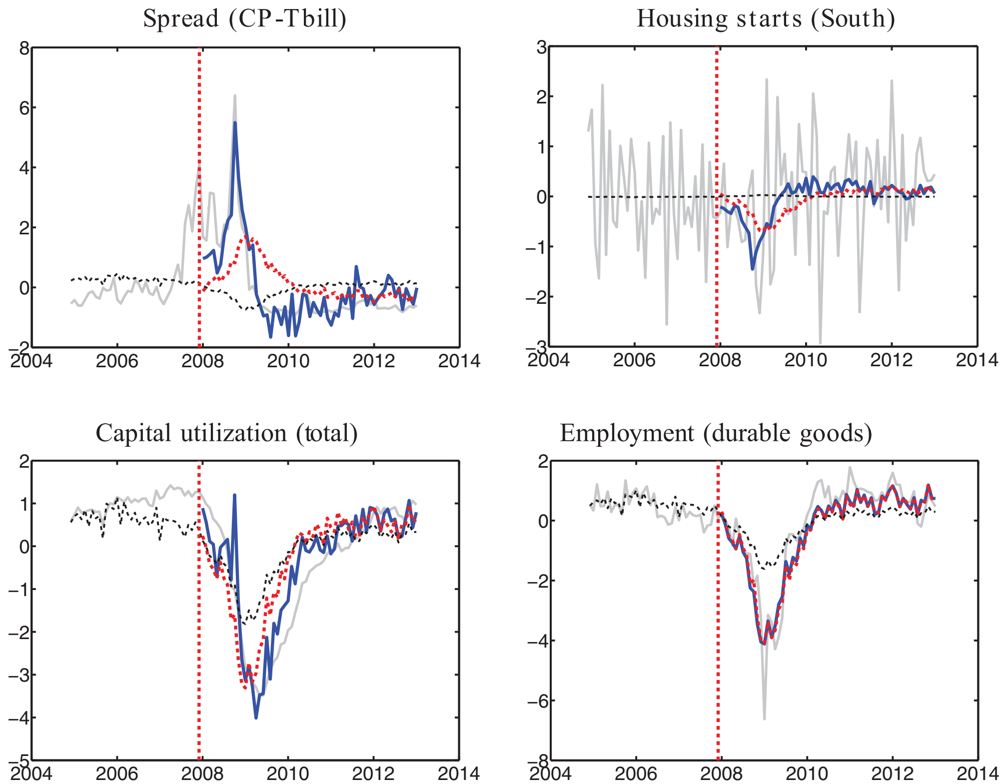


FIGURE 4

In-sample prediction: individual series.

*Notes:* Grey solid line: actual; dark dashed line: old loadings; grey (red) dashed line: new loadings only; dark (blue) solid line: new loadings and factor (a colored graph is available in the web version of this article). Results are based on $\widehat{r}_a = 1$, $\widehat{r}_b = 2$, and a break date of 2007:M12.

corresponds to an individual series, and the vertical lines delimit groups of variables (see Notes for Figure 3).

Two observations stand out. First, capturing the change in the loadings and the change in the number of factors is approximately equally important, in the sense that the contribution of $F^{R\Omega}_{b,1}(\Psi^{R\Omega}_1 - \Lambda^R)'$ to the overall $R^2$ is similar to the contribution of $F^{R\Omega}_{b,2}\Psi^{R\Omega\prime}_2$. Secondly, the new factor predominantly affects financial variables, namely those series in the two interest rate categories (IntL, IntS), the money and credit group (Mon), the stock price and wealth group (StPr), and the exchange rates group (ExR). While there are some spillovers to the real side (*i.e.* industrial production (IP) and orders, inventories, and sales (ORD), the $R^2$ differential is generally lower than for the financial variables.

Figure 4 depicts the fitted time path of four series: the spread between commercial paper and Treasury bills, housing starts in the southern Census district, capital utilization, and employment in durable goods manufacturing. We overlay the actual sample paths with three (in-sample) predicted paths, which, as before, we refer to as *old loadings*, *new loadings only*, *new loadings and factor*. The spread starts to rise towards the end of 2007. This rise is not captured by the path predicted under the pre-break loadings, which stays fairly constant throughout 2008. As suggested by Figure 3, the discrepancy between the *old loadings* and the *new loadings only* paths is substantial during the Great Recession period. Once the loadings are allowed to change, the predicted spread rises drastically throughout 2008, and even more so once the new factor is accounted for. Capital utilization and employment drop drastically in the second half of 2008 and only start recovering in 2010. The *old loadings* path is unable to capture the large drop in real activity. With the *new loadings only,* on the other hand, the model is able to track both capital utilization and employment quite well during and after the recession, and the additional factor has not altered the predicted paths of these series. However, there are real series that show a noticeable effect of the new factor, one of them being the housing starts series in the top-right panel of Figure 4.

At first glance, the results in Figure 4 look very different from those presented in Figure 2 of Stock and Watson (2012). Part of the discrepancy is due to the different normalization schemes. We are normalizing the variance of the factors to one, whereas Stock and Watson (2012) normalized the length of the loading vectors to one (*i.e.* $\Lambda'\Lambda/N = I_r$). To be able to explain the macroeconomic dynamics during and after the Great Recession with factors that have unit variance, a big change in the loadings is required. This is evident from our Figures 3 and 4. If we normalize the length of the loadings vector before and after the break to one, then the increase in the volatility after 2007 is absorbed by an increase in the factor volatility. The ratio of pre- to post-break factor standard deviation under this alternative normalization is approximately 2. Stock and Watson (2012) interpret this phenomenon as an unchanged response to “old” factors combined with large innovations to the “old” factors in the post-2007 sample. In the absence of the emergence of a new factor, we would interpret this phenomenon as a type-1 instability of the factor model.

To summarize, our model selection procedure provides strong evidence that the loadings in the normalized factor model changed, generally implying a stronger comovement of the series after 2007. There is also evidence of the emergence of a new factor, which to a large extent seems to capture important co-movements among financial series but also spills over into the real activity variables.

## 8. CONCLUSION

We develop a shrinkage estimation procedure for high-dimensional factor models that generates consistent estimates of the numbers of pre- and post-break factors. The estimator is appealing because it is robust to instabilities at an unknown break date. Moreover, in situations in which the number of factors is constant throughout the sample, the procedure can consistently detect changes in the factor loadings. We show that once the numbers of pre- and post-break factors have been consistently estimated, one can use a conventional least-squares approach to determine the break date consistently. Our Monte Carlo analysis demonstrates that the shrinkage procedure has good finite sample properties and either is competitive with or strictly dominates existing procedures that are either designed to determine the number of factors in a no-break environment or designed to test for a break in the loadings if the number of factors is known. In an application to U.S. data, our procedure detects an increase in the number of factors for a large macroeconomic and financial data set at the onset of the Great Recession and a substantial change in the factor loadings. The new factor mainly affects financial variables but also generates spillovers to the real economy, which is consistent with the narratives of the 2007–9 recession.

*Acknowledgments*. Minchul Shin and Irina Pimenova (Penn) provided excellent research assistance. Many thanks to Ataman Ozylidirim for granting us access to a selected set of time series published by The Conference Board and to Xu Han and Atsushi Inoue for sharing the code that implements their break test. We thank Stephane Bonhomme (co-editor), four anonymous referees, and participants at various seminars and conferences for helpful comments and suggestions. Schorfheide gratefully acknowledges financial support from the National Science Foundation under Grant SES 1061725.

**Supplementary Data**

Supplementary Data are available at *Review of Economic Studies* online.

## REFERENCES


AHN, S. C. and HORENSTEIN, A. R. (2013), "Eigenvalue Ratio Test for the Number of Factors", *Econometrica*, **81**, 1203–1227.

ALESSI, L., BARIGOZZI, M. and CAPASSO, M. (2010), "Improved Penalization for Determining the Number of Factors in Approximate Factor Models", *Statistics & Probability Letters*, **80**, 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 & Economic Statistics*, **25**, 91–96.

ANDO, T. and BAI, J. (2015), "Panel Data Models with Grouped Factor Structure Under Unknown Group Membership", *Journal of Applied Econometrics*, forthcoming.

ANDREWS, D. W. K. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point", *Econometrica*, **61**, 821–56.

BAI, J. (1997), "Estimation of a Change Point in Multiple Regression Models", *Review of Economics and Statistics*, **79**, 551–563.

BAI, J. (2003), "Inferential Theory for Factor Models of Large Dimensions", *Econometrica*, **71**, 135–171.

BAI, J. (2010), "Common Breaks in Means and Variances for Panel Data", *Journal of Econometrics*, **157**, 78–92,

BAI, J. and LIAO, Y. (2012), "Efficient Estimation of Approximate Factor Models via Regularized Maximum Likelihood" (Manuscript, Columbia University and University of Maryland).

BAI, J. and NG, S. (2002), "Determining the Number of Factors in Approximate Factor Models", *Econometrica*, **70**, 191–221.

BAI, J. and NG, S. (2006), "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions", *Econometrica*, **74**, 1133–1150.

BAI, J. and NG, S. (2007), "Determining the Number of Primitive Shocks in Factor Models", *Journal of Business & Economic Statistics*, **25**, 52–60.

BALTAGI, B. H., KAO, C. and WANG, F. (2015), "Change Point Estimation in Large Heterogeneous Panels" (Manuscript, Syracuse University).

BATES, B. J., PLAGBORG-MØLLER, M., STOCK, J. H., *et al*. (2013), "Consistent Factor Estimation in Dynamic Factor Models with Structural Instability", *Journal of Econometrics*, **177**, 289–304.

BONHOMME, S. and MANRESA, E. (2015), "Grouped Patterns of Heterogeneity in Panel Data", *Econometrica*, **83**, 1147–1184.

BREITUNG, J. and EICKMEIER, S. (2011), "Testing for Structural Breaks in Dynamic Factor Models", *Journal of Econometrics*, **163**, 71–84.

BREITUNG, J. and PIGORSCH, U. (2013), "A Canonical Correlation Approach for Selecting the Number of Dynamic Factors", *Oxford Bulletin of Economics and Statistics*, **75**, 23–36.

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*, **32**, 359–374.

CHEN, L. (2015), "Estimating the Common Break Date in Large Factor Models", *Economics Letters*, **131**, 70–74.

CHEN, L., DOLADO, J. J. and GONZALO, J. (2014), "Detecting Big Structural Breaks in Large Factor Models", *Journal of Econometrics*, **180**, 30–48.

CHOI, I. (2013), "Model Selection for Factor Analysis: Some New Criteria and Performance Comparisons" (Working Paper No. 1209, Sogang University Research Institute for Market Economy), **1209**.

CLIFF, N. (1966), "Orthogonal Rotation to Congruence", *Psychometrika*, **31**, 33–42.

CORRADI, V. and SWANSON, N. R. (2014), "Testing for Structural Stability of Factor Augmented Forecasting Models", *Journal of Econometrics*, **182**, 100–118.

HALLIN, M., AND R. LIKA (2007), "Determining the Number of Factors in the General Dynamic Factor Model", *Journal of the American Statistical Association*, **102**, 603–617.

HAN, X. and INOUE, A. (2014), "Tests for Parameter Instability in Dynamic Factor Models", *Econometric Theory*, **31**, 1117–1152.

KAPETANIOS, G. (2010), "A Testing Procedure for Determining the Number of Factors in Approximate Factor Models With Large Datasets", *Journal of Business & Economic Statistics*, **28**, 397–409.

LEE, S., SEO, M. H. and SHIN, Y. (2015), "The Lasso for High-Dimensional Regression with a Possible Change-Point", *Journal of the Royal Statistical Society: Series B*, forthcoming.

LIN, C. and NG, S. (2012), "Estimation of Panel Data Models with Parameter Het- erogeneity When Group Membership is Unknown", *Journal of Econometric Methods*, **1**, 42–55.

LU, X. and SU, L. (2015), "Shrinkage Estimation of Dynamic Panel Data Models with Interactive Fixed Effects", *Journal of Econometrics*, **190**, 148–175.

ONATSKI, A. (2009), "Testing Hypotheses About the Number of Factors in Large Factor Models", *Econometrica*, **77**, 1447–1479.

ONATSKI, A. (2010), "Determining the Number of Factors from Empirical Distribution of Eigenvalues", *Review of Economics and Statistics*, **92**, 1004–1016.

ONATSKI, A. (2012), "Asymptotics of the Principal Components Estimator of Large Factor Models with Weakly Influential Factors", *Journal of Econometrics*, **168**, 244–258.

QIAN, J. and SU, L. (2015a), "Shrinkage Estimation of Common Breaks in Panel Data Models via Adaptive Group Fused Lasso", *Journal of Econometrics*, Forthcoming.

QIAN, J. and SU, L. (2015b), "Shrinkage Estimation of Regression Models with Multiple Structural Changes", *Econometric Theory*, Forthcoming.

SCHÖNEMANN, P. (1966), "A Generalized Solution of the Orthogonal Procrustes Problem", *Psychometrika*, **31**, 1–10.

STOCK, J. H. and WATSON, W. M. (2002), "Forecasting Using Principal Components From a Large Number of Predictors", *Journal of the American Statistical Association*, **97**, 1167–1179.

STOCK, J. H. and WATSON, W. M. (2009), "Forecasting in Dynamic Factor Models Subject to Structural Instability", in Shephard, N. and Castle, J. (eds), *The Methodology and Practice of Econometrics: Festschrift in Honor of D.F. Hendry*. (Oxford University Press) 1–57.

STOCK, J. H. and WATSON, W. M. (2012), "Disentangling the Channels of the 2007–09 Recession", *Brookings Papers on Economic Activity*, Spring, 81–156.

SU, L., SHI, Z. and PHILLIPS, P. C. B. (2014), "Identifying Latent Structures in Panel Data", (Manuscript, Singapore Management University and Yale University).

SU, L. and WANG, X. (2015), "On Time-Varying Factor Models: Estimation and Testing", (Manuscript, Singapore Management University).

TIBSHIRANI, R. (1994), "Regression Shrinkage and Selection Via the Lasso", *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, **58**, 267–288.

YUAN, M. and LIN, Y. (2006), "Model Selection and Estimation in Regression with Grouped Variables", *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, **68**, 49–67.

ZOU, H. (2006), "The Adaptive Lasso and Its Oracle Properties", *Journal of the American Statistical Association*, **101**, 1418–1429.