Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities

Xu Cheng* — *University of Pennsylvania*

Zhipeng Liao — *University of California, Los Angeles*

Frank Schorfheide — *University of Pennsylvania and NBER*

This version: November 2015

Abstract

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 paper 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.

JEL Classification: C13, C33, C52

Keywords: Great Recession, High-dimensional Model, Large Data Sets, LASSO, Latent Factor Model, Model Selection, Shrinkage Estimation, Structural Break

* Correspondence: X. Cheng and F. Schorfheide: Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104, USA. Z. Liao: Department of Economics, University of California, Los Angeles, 8379 Bunche Hall, Mail Stop: 147703, Los Angeles, CA 90095. Email: xucheng@sas.upenn.edu (Cheng); zhipeng.liao@econ.ucla.edu (Liao); schorf@ssc.upenn.edu (Schorfheide). 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.

1 Introduction

High-dimensional factor models are widely used to analyze macroeconomic and financial panel data, where a small number of unobserved factors drive the comovement of a large number of time series. This paper focuses on the complications in the estimation of factor models that arise from potential structural breaks such as the 2007-2009 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, Dolado, and Gonzalo (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 paper 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 paper, 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. Second, 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 paper is based on the minimization of a penalized least squares (PLS) criterion function in which adaptive group-LASSO penalties (Tibshirani (1994), Zou (2006), and Yuan and Lin (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. Second, 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 penalized least squares 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. Third, 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 paper 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 paper and left for future research.

The empirical analysis in this paper 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-modeled 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.

[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, Plagborg-Møller, Stock, and Watson (2013) show that in the presence of small structural instabilities of the factor loadings the principal component estimator of the factors remains consistent.

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 dynamic factors includes Amengual and Watson (2007), Bai and Ng (2007), Hallin and Lika (2007), Onatski (2009), Onatski (2010), Alessi, Barigozzi, and Capasso (2010), Kapetanios (2010), Caner and Han (2014), Ahn and Horenstein (2013), Breitung and Pigorsch (2013), and Choi (2013). 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 paper 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, Dolado, and Gonzalo (2014), Han and Inoue (2014), and Corradi and Swanson (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. Second, 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. Third, the procedure controls the model selection error jointly by treating all time series variables as a group. This is particularly desirable for large-scale datasets.

There exist 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, Seo, and Shin (2015) and Qian and Su (2015b,a)). Our paper differs from the above-mentioned work because the factor structure is both unobserved and unstable. Recently Baltagi, Kao, and Wang (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 modeling heterogeneity in panel data with latent group structure. Various classification and shrinkage methods have been proposed, see Bonhomme and Manresa (2015), Lin and Ng (2012), Ando and Bai (2015), Su, Shi, and Phillips (2014). These papers consider a latent structure that is different from the one that is estimated in the current paper. The structure in this paper 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 paper is organized as follows. Section 2 describes the factor model and the types of instabilities considered in this paper. 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 Appendix.

2 A Factor Model with Structural Break

We observe panel data $\{X_{it} \in R : i = 1, \ldots, N, t = 1, \ldots, T\}$. Let $X_t = (X_{1t}, \ldots, X_{Nt})' \in R^{N\times 1}$ denote the observations at time period t. For $t = 1, \ldots, T_0$, 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 DGP before T_0 is

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

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 $X_a = (X_1, \ldots, X_{T_0})' \in R^{T_0\times N}$, $F_a = (F_1^0, \ldots, F_{T_0}^0)' \in R^{T_0\times r_a}$, and $e_a = (e_1, \ldots, e_{T_0})' \in R^{T_0\times N}$. The matrices F_a and Λ^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 $X_b = (X_{T_0+1}, \ldots, X_T)'$, $F_{b,1} = (F_{T_0+1}^0, \ldots, F_T^0)'$, $F_{b,2} = (F_{T_0+1}^*, \ldots, F_T^*)'$, and $e_b = (e_{T_0+1}, \ldots, e_T)'$. 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 Γ_1^0 captures possible changes in the loadings of the pre-break factors F_t^0, whereas the matrix Γ_2^0 contains the loadings for the new factors F_t^*. The changes in the factor loadings are summarized in $\Gamma^0 = (\Gamma_1^0, \Gamma_2^0)$. 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 $F_b = (F_{b,1}, F_{b,2})$ with factor loadings $\Psi^0 = (\Lambda^0 + \Gamma_1^0, \Gamma_2^0)$. 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 behavior 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 nontrivial contributions to the variance of the data.[2] For $t > T_0$, let $\overline{F}_t^0 = (F_t^{0\prime}, F_t^{*\prime})' \in R^{r_b}$ denote the r_b factors after the break. Throughout this paper, we use $C \in \mathbb{R}$ to denote a generic positive constant.

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

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 Σ_Λ, Σ_Ψ 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 Σ_Λ and Σ_Ψ 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 (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 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]

[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.

[3] $T^{-1}F^{R\prime}F^R$ is not defined as an exact identity matrix because the limiting covariance matrices Σ_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

Note that our normalization interprets changes in the law of motion of the factors F_a and F_b as changes in the loadings Λ^R and Ψ^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 ρ_a to ρ_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 Λ^0 and Ψ^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 paper, 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 labeling 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.

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,

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.

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 ℓ-th largest eigenvalue of a square matrix A. The following assumption, stated in terms of the coefficients of the DGP in (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 Λ^0 and Ψ^0 do not span the same column space asymptotically. It implies that the column spaces of Λ^R and Ψ^R in (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 π_0, rather than T_0, the "true" break date and assume that $\pi_0 \in \Pi$, where Π 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 π_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 π. We verify the following relationship between the conjectured break date and the numbers of pre- and post-break factors in the Appendix:

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

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

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

Because $rank(\Sigma^+_{\Lambda\Psi}) \geq r_b \geq r_a$, we see that $r_a(\pi)+r_b(\pi)$ is minimized at π_0, with the minimum value $r_a + r_b$. Define the set of values π 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 nonempty set. In order to ensure that π_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 Λ^0 is asymptotically not contained in the space generated by Ψ^0, which leads to the stronger Assumption ID*.

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

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 (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), Λ^{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 (3.1), because they are equivalent to consistent selection of the zero and and nonzero components in Λ^{R+} and Γ^{R+}. Hence, for consistent model selection, it is key to obtain estimators that can consistently distinguish zeros from nonzeros in Λ^{R+} and Γ^{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 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 square 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 Λ^{R+} and Γ^{R+} by minimizing a PLS criterion function:

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

where

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

Λ_ℓ and Γ_ℓ are the ℓ-th column of Λ and Γ, respectively, α_{NT} and β_{NT} are two sequences of positive constants that depend on N and T, and ω_ℓ^λ and ω_ℓ^γ are data-dependent weights defined as:

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

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 Λ^+ and Γ^+, where the ℓ subscript denotes the ℓ-th column of the matrices.[4]

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

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 nonzeros altogether. This feature is particularly useful for large-scale factor models because the irrelevant factors

[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 Λ or Γ equal to zero, which is why we introduced the indicator function notation.

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 (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 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 (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 paper, 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 (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 (2.5) as

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

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 α_{NT} and β_{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. Second, 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 π_0 replaced by π, 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 π. 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 π only through the least squares criterion function. The averaging penalty functions $P_1^*(\Lambda)$ and $P_2^*(\Lambda)$ are

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

where ξ has a uniform distribution on Π and $\mathbb{E}_\xi[\cdot]$ denotes the expectation with respect to ξ.[5] In practice, Π is approximated by a set of equally spaced grid points Π_d, and the expectation in (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 π and the sequences can vary with π. 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 (3.14) with those in (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.

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

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

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

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

[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 Π. Note that the above two terms depend on N and T.

where $\widehat{r}_a(\pi)$ and $\widehat{r}_b(\pi)$ are defined as in (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 Post Model Selection Estimation

In addition to estimating the model specification our shrinkage estimator also provides an estimate of the loading matrices Λ and Γ. 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 sense), we recommend to re-estimate the loadings using least squares conditional on the selected model specification. We refer to the resulting estimator as post model selection (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

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

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

[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 (π)-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 $\widetilde{F}$.

Building on work by Bai (1997), we will show below that the break date π_0 can be estimated consistently by using a least-squares objective function. Let

$$\widehat{\pi} = \arg\min_{\pi \in \Pi} Q_{NT}(\pi; \widehat{r}_a^*, \widehat{r}_b^*). \tag{3.19}$$

where

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

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 α_{NT} and β_{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 ω_ℓ^λ and ω_ℓ^γ also on the tuning parameters α_{NT} and β_{NT}. Roughly speaking, α_{NT} is the weight attached to the penalty on the coefficients related to X_a, whereas β_{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 κ_1 and κ_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 α_{NT} and β_{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 α_{NT} and β_{NT} to balance these two rates and replace the overall sample size T by the subsample sizes T_a and T_b. We set κ_1 and κ_2 equal to

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

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 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 (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. Second, a better preliminary estimator $\widetilde{\Gamma}$ is obtained through a rotation of the factor loadings Λ^R and Ψ^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 $\widehat{\Lambda}^{(i)}_{PMS}$, $\widehat{\Psi}^{(i)}_{PMS}$ and $\widehat{\Gamma}^{(i)}_{PMS}$ 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* ω_ℓ^λ*,* ω_ℓ^γ*,* α_{NT} *and* β_{NT} *following (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 (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 $\widehat{\Lambda}_{PMS}^{(1)}$ *and* $\widehat{\Psi}_{PMS}^{(1)}$ *using the definition in (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*

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

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

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

2. Second-stage Shrinkage Estimation:

2.1 Let

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

and calculate ω_ℓ^λ*,* ω_ℓ^γ*,* α_{NT}*, and* β_{NT} *following (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 $\widehat{\mathcal{M}}^{(2)} = \left(\widehat{\mathcal{B}}_0^{(2)}, \widehat{r}_a^{(2)}, \widehat{r}_b^{(2)}\right)$ *construct the PMS estimators* $\widehat{\Lambda}_{PMS}^{(2)}$ *and* $\widehat{\Psi}_{PMS}^{(2)}$ *using the definition in (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

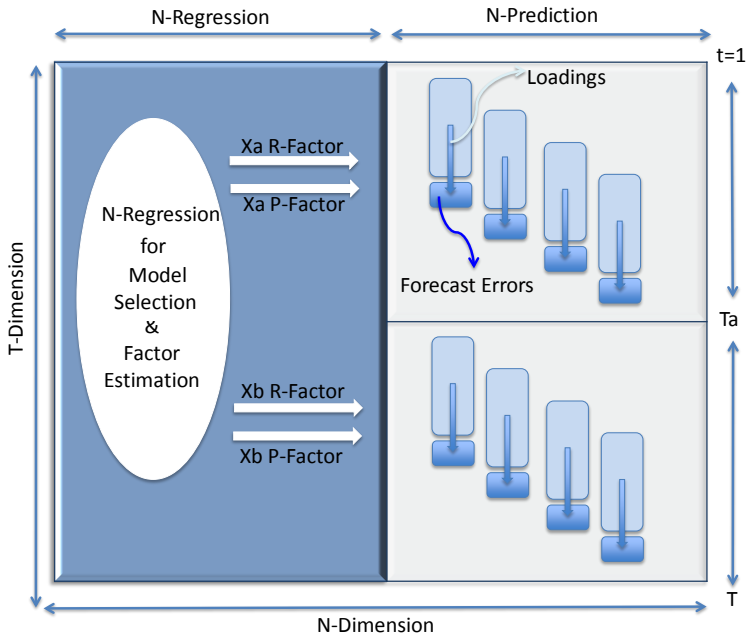
of reporting a "true" break, we found in our simulation experiments that overall it leads to an improved finite-sample performance. The rotation does not affect the asymptotic validity of our procedure. Formally, the problem is to find an orthogonal matrix Q such that $||\overline{\Lambda}^{(1)} - \overline{\Psi}^{(1)}Q||_2$ 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 $\overline{\Lambda}^{(1)\prime}\overline{\Psi}^{(1)} = UDV'$. In Section 5, we show that if there is indeed a type-1 instability, the Q rotation will not eliminate the difference between Λ_ℓ^R and Ψ_ℓ^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 (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 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.

Figure 1: Graphical Representation of Algorithm 2



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.

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 (4.1). In practice, we can choose $\kappa_1(\pi)$ and $\kappa_2(\pi)$ as in (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 ω_ℓ^λ, ω_ℓ^γ, α_{NT}, and β_{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 (3.8) with those in (3.16). Note that the first-step estimators $\widehat{r}_a^{(1)}$ and $\widehat{r}_b^{(1)}$ do not vary with π following the definition in (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 π, the $X_{(-j_N)}$ subsamples are designed in the same way as in the known-break-date case, with π_0 replaced by π. For each c we obtain a selected model, which does not depend on π by definition. From the validation sample $X_{(j_N)}$ we first eliminate the observations that lie outside of the conjectured break interval Π 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 Π, 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 Π 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 Π is closely tied to the application. The interval could be centered around 1984 if the goal is to detect breaks associated with the Great Moderation; or centered 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 behavior 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 Appendix.

Second, 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 α_{NT} and β_{NT}. We begin with the assumptions on the stochastic order of the preliminary estimators, which affect the data-dependent weights ω_ℓ^λ and ω_ℓ^γ defined in (3.4). Define $C_{NT} = \min(T^{1/2}, N^{1/2})$, where C_{NT} is the convergence rate of the unrestricted least square 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, ω_ℓ^λ and ω_ℓ^γ, 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 ω_ℓ^λ and ω_ℓ^γ 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 Λ^R and Γ^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 ω_ℓ^λ and ω_ℓ^γ 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 ω_ℓ^λ and ω_ℓ^γ determine the relative penalties of different columns of factor loadings, the tuning parameters α_{NT} and β_{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 α_{NT} and β_{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 α_{NT} and β_{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 ω_ℓ^λ and ω_ℓ^γ are stochastically bounded, the penalties on the nonzero 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 α_{NT} and β_{NT} converge to zero not too fast. The choice of α_{NT} and β_{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 (2.5). As before, let the subscript ℓ denote the ℓ-th column of a matrix.

Theorem 1 *Suppose Assumptions* A, B, C *(see Appendix),* D *(see 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 zero 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 $\ell = 1, \ldots, r_a$, the PLS estimator $\widehat{\Lambda}_\ell$ converges in probability to the factor loadings Λ_ℓ^R of the transformed DGP. Parts (c) to (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 Appendix),* D *(see Appendix),* ID, P1, P2, *and* T *hold. Then the estimated model is consistent:*

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

Theorem 2 provides model selection consistency for the shrinkage estimation based on any set of preliminary estimators that satisfy Assumption 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 Appendix),* D *(see 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

$$\mathcal{Z} = \{\ell : N^{-1}||\Gamma_\ell^R||^2 = N^{-1}||\Psi_\ell^R - \Lambda_\ell^R||^2 \geq C\}. \tag{5.1}$$

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 Λ_ℓ^R is not in the column space generated by Ψ^R. Assumption R is imposed on the loadings Λ^R of the normalized version of the DGP rather than on the loadings Λ^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 Appendix),* D *(see Appendix),* ID, R, *and* T.

5.2 Unknown Break Date

Next, we show that the shrinkage estimator based on the averaging penalty in (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 behavior of the shrinkage estimator $\widehat{\Lambda}(\pi)$ and $\widehat{\Gamma}(\pi)$ for all π 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 π. 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

$$\begin{aligned} X_a(\pi) &= F_a^R(\pi)\Lambda^R(\pi)' + e_a(\pi), \\ X_b(\pi) &= F_b^R(\pi)\Psi^R(\pi)' + e_b(\pi), \end{aligned} \tag{5.2}$$

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 $\ell = r_b$. $\square$

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 π_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 Appendix),* B, C* *(see Appendix),* D *(see 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 (i) if $\pi = \pi_0$, the shrinkage estimator with the averaging penalty behaves similarly to that in Theorem 1 and (ii) 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 π substantially differs from π_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 (3.19) delivers a consistent estimate of the break date.

Theorem 4 *Suppose that Assumptions* A* *(see Appendix),* B, C* *(see Appendix),* D *(see 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 analyzing 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 mean squared errors (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, Plagborg-Møller, Stock, and Watson (2013), with the additional flexibility to accommodate both type-1 and type-2 instabilities

[8] As is common in the time series literature, we proved a consistency result for π_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, Kao, and Wang (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 paper.

and the shift of focus from small breaks to large breaks. The DGP takes the form

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

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 Ω 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 Σ_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 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 $t \leq T_0$ 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 λ_i.

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

Depending on the type of the instabilities, we consider two different ways of constructing the post-break factor loadings ψ_i. For a type-1 instability, we set $\psi_i = (1 - \mathbf{w})\lambda_i + \mathbf{w}\lambda_i^*$, where λ_i^* and λ_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, ψ_i is drawn independently of everything else with a distribution that is similar to that of λ_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 Π is approximated by a discrete set Π_d. The grid size in Π_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 π_0. To define a post-break subsample for the PMS estimator, the least square estimator of the break date described in Section 3.4 is used because π_0 is unknown.

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

In addition to assessing the probability of selecting the correct model specification we also compute mean-squared errors for out-of-sample forecasts (MSFE) generated by the selected model. The series to be 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 ϵ_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 φ_a^* and φ_a are independent and have the same distribution. For a type-2 change, φ_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. Second, 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 φ_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).

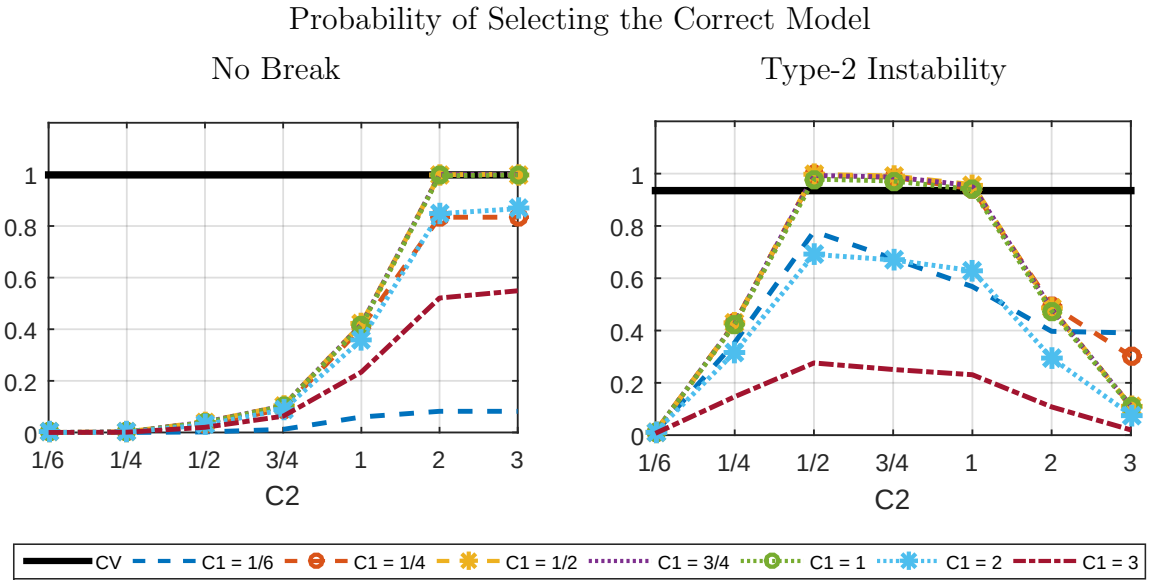
6.2 Results for Shrinkage Estimator

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 toward 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.

Table 1: Monte Carlo Experiments

Exp.	R^2	π_0	α, β	Break Date
1	homogeneous	0.5	0.2	Known
2	heterogeneous	0.8	0.2	Known, Unknown
3	homogeneous	0.5	0.5	Known

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.

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 (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

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 mean-squared-forecast error 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.

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.

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 one favor 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 3 to 4 factors and a sample size of $N = T = 100$ is the selection probability less than one. 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,

[12] As a robustness check, we repeat Experiment 1 using $k = 12$. The results are reported in 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 mean-squared-forecast error 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 $\text{MSE}(\widehat{\pi}) = \mathbb{E}\big[\big(T(\widehat{\pi} - \pi_0)\big)^2\big]$.

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., $\text{MSE}(\widehat{\pi}) = \mathbb{E}\big[\big(T(\widehat{\pi} - \pi_0)\big)^2\big]$. 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 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 one 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 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 paper tackles all unknown aspects of the model specification simultaneously.

We only made comparisons with an alternative procedure in cases where (i) the alternative is implementable and (ii) the alternative specifies the unknown parts of the model correctly. Thus, our experimental designs generally favor 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, Dolado, and Gonzalo (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 2007-2009 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: (i) We extend the series in the SW132 data set to 2012:M12, using May 2013 data vintages. (ii) We replace the quarterly

Table 4: MODEL SELECTION, T_c IS 2007:12

Interval	Factors		Break Dates		Tuning Const		Time (Min)
Size	$\widehat{r}_a$	$(\widehat{r}_b - \widehat{r}_a)$	Least Sq.	Revised	c_1	c_2	
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 center the interval Π at 2007:M12 and use the averaging penalty functions $P_1^*(\Lambda)$ and $P_2^*(\Lambda)$ defined in (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 run on a single Intel Core i7 2.93 Ghz processor.

series in SW132 by their monthly counterparts, if available. This is possible for consumption of nondurables, services, and durables; for nonresidential investment; and for 16 price series. We remove the remaining quarterly series for which no monthly observations are available. (iii) We add two GDP components that are available at monthly frequency: change in private inventory and wage and salary disbursements. (iv) Following Stock and Watson (2012), we remove local means from all series using a biweight kernel with a bandwidth of 100 months. The local means are approximately the same as the ones obtained by a centered moving average of ± 70 months. After making these modifications, our data set consists of $N = 102$ series of monthly macroeconomic 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 6.3. The model selection results are summarized in Table 4. We consider four different sets of potential break dates Π, which are centered 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 Π 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 Π 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 (3.19). We minimize the least squares criterion over the interval Π, characterized in the first column of the table. In 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 Π 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 Π 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 Π is approximately 30 minutes.

7.3 Decomposing the Structural Change

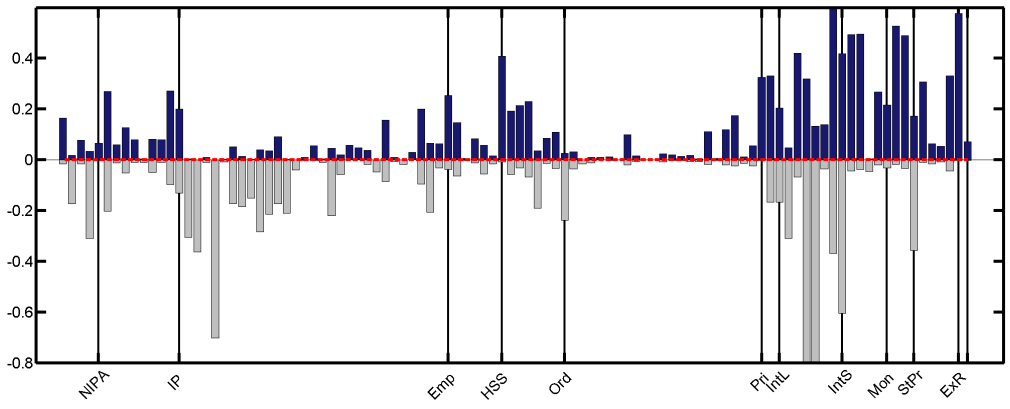
In empirical applications, it is interesting and useful to decompose type-2 changes into the contribution of the new factors and changes in the effects of old factors. While the (non-identifiable) DGP in (2.2) and (2.3) provides a natural decomposition of type-2 structural changes into changes resulting from the new factors, $F_{b,2}\Gamma_2^{0\prime}$, and changes associated with the effect of the extended versions of the old factors, $F_{b,1}\Gamma_1^{0\prime}$, 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 Λ^R and the new loadings $\Psi^R\Omega_a$:

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

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

Figure 3: R^2 Gains from New Loadings and Factor



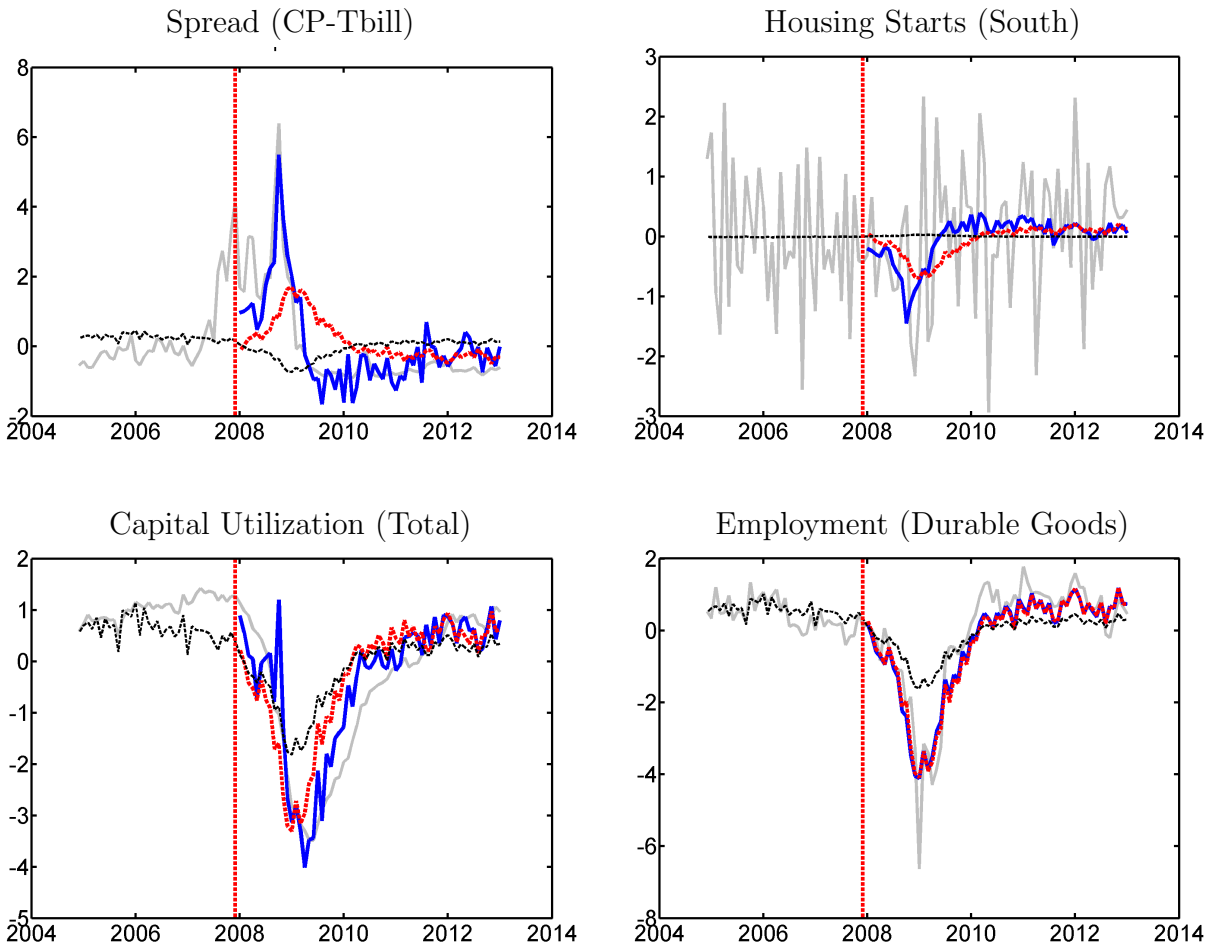
Notes: Base line case *new loadings only*. Dark bars (above zero) indicate R^2 gain due to new factor relative to *new loadings only*. Light grey bars (below zero) indicate R^2 losses from using the old loading matrices. Results are based on $\widehat{r}_a = 1$, $\widehat{r}_b = 2$, and a break date 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 singular value decomposition $\Lambda^{R\prime}\Psi^R = UDV'$ (see Cliff (1966)). Let $\Omega_\perp$ be the null space of Ω_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., $F_b^{R\Omega\prime}F_b^{R\Omega}/T_b = I_{r_b\times r_b}$. 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:

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

As a baseline, we compute R^2 values for each individual series based on the variation explained by $F_{b,1}^{R\Omega}\Lambda^{R\prime} + F_{b,1}^{R\Omega}(\Psi_1^{R\Omega} - \Lambda^R)'$ (*new loadings only*). We compare the baseline R^2s to R^2s associated with $F_{b,1}^{R\Omega}\Lambda^{R\prime}$ (*old loadings*) and R^2s associated with all three terms in (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 corresponds to an individual series, and the vertical lines delimit groups of variables (see Notes for Figure 3).

Figure 4: In-sample Prediction: Individual series



Notes: Gray line: actual; black dashed: old loadings; thick dashed red: new loadings only; thick blue: new loadings and factor. Results are based on $\widehat{r}_a = 1$, $\widehat{r}_b = 2$, and a break date of 2007:M12.

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 $F_{b,1}^{R\Omega}(\Psi_1^{R\Omega} - \Lambda^R)'$ to the overall R^2 is similar to the contribution of $F_{b,2}^{R\Omega}\Psi_2^{R\Omega\prime}$. Second, 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 toward 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 2007-2009 recession.

References

Ahn, S. C., and A. R. Horenstein (2013): "Eigenvalue Ratio Test for the Number of Factors," *Econometrica*, 81(3), 1203–1227.

Alessi, L., M. Barigozzi, and M. Capasso (2010): "Improved Penalization for Determining the Number of Factors in Approximate Factor Models," *Statistics & Probability Letters*, 80(23-24), 1806–1813.

Amengual, D., and M. W. Watson (2007): "Consistent Estimation of the Number of Dynamic Factors in a Large N and T Panel," *Journal of Business & Economic Statistics*, 25(1), 91–96.

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

Andrews, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61(4), 821–56.

Bai, J. (1997): "Estimation of a Change Point in Multiple Regression Models," *Review of Economics and Statistics*, 79(4), 551–563.

——— (2003): "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71(1), 135–171.

——— (2010): "Common breaks in means and variances for panel data," *Journal of Econometrics*, 157(1), 78 – 92, Nonlinear and Nonparametric Methods in Econometrics.

Bai, J., and Y. Liao (2012): "Efficient Estimation of Approximate Factor Models via Regularized Maximum Likelihood," *Manuscript, Columbia University and University of Maryland*.

Bai, J., and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70(1), 191–221.

——— (2006): "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," *Econometrica*, 74(4), 1133–1150.

——— (2007): "Determining the Number of Primitive Shocks in Factor Models," *Journal of Business & Economic Statistics*, 25, 52–60.

——— (2013): "Principal Components Estimation and Identification of Static Factors," *Journal of Econometrics*, 176(1), 18–29.

Baltagi, B. H., C. Kao, and F. Wang (2015): "Change point estimation in large heterogeneous panels," *Manuscript, Syracuse University*.

Bates, B. J., M. Plagborg-Møller, J. H. Stock, and M. W. Watson (2013): "Consistent Factor Estimation in Dynamic Factor Models with Structural Instability," *Journal of Econometrics*, 177(2), 289–304.

Bhatia, R. (1997): *Matrix Analysis*. Springer-Verlag New York.

Bonhomme, S., and E. Manresa (2015): "Grouped Patterns of Heterogeneity in Panel Data," *Econometrica*, 83(3), 1147–1184.

Breitung, J., and S. Eickmeier (2011): "Testing for Structural Breaks in Dynamic Factor Models," *Journal of Econometrics*, 163(1), 71–84.

Breitung, J., and U. Pigorsch (2013): "A Canonical Correlation Approach for Selecting the Number of Dynamic Factors," *Oxford Bulletin of Economics and Statistics*, 75(1), 23–36.

Bühlmann, P., and S. van de Geer (2011): *Statistics for High-Dimensional Data: Methods, Theory and Applications*. New York: Springer.

Caner, M., and X. Han (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(3), 359–374.

Chen, L. (2015): "Estimating the common break date in large factor models," *Economics Letters*, 131(C), 70–74.

Chen, L., J. J. Dolado, and J. Gonzalo (2014): "Detecting big structural breaks in large factor models," *Journal of Econometrics*, 180(1), 30 – 48.

Cheng, X., and P. C. Phillips (2012): "Cointegrating rank selection in models with time-varying variance," *Journal of Econometrics*, 169(2), 155 – 165.

Choi, I. (2013): "Model Selection for Factor Analysis: Some New Criteria and Performance Comparisons," *Working Paper, Sogang University Research Institute for Market Economy*, 1209(1209).

Cliff, N. (1966): "Orthogonal Rotation to Congruence," *Psychometrika*, 31(1), 33–42.

Corradi, V., and N. R. Swanson (2014): "Testing for structural stability of factor augmented forecasting models," *Journal of Econometrics*, 182(1), 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(478), 603–617.

Han, X., and A. Inoue (2014): "Tests for Parameter Instability in Dynamic Factor Models," *Econometric Theory*, FirstView, 1–36.

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(3), 397–409.

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

Lin, C., and S. Ng (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 L. Su (2015): "Shrinkage estimation of dynamic panel data models with interactive fixed effects," *Journal of Econometrics*, pp. –.

Onatski, A. (2009): "Testing Hypotheses About the Number of Factors in Large Factor Models," *Econometrica*, 77(5), 1447–1479.

——— (2010): "Determining the Number of Factors from Empirical Distribution of Eigenvalues," *Review of Economics and Statistics*, 92(4), 1004–1016.

——— (2012): "Asymptotics of the Principal Components Estimator of Large Factor Models with Weakly Influential Factors," *Journal of Econometrics*, 168(2), 244–258.

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

——— (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), 1–10.

Stock, J. H., and W. M. Watson (2002): "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, 97(460), 1167–1179.

——— (2009): "Forecasting in Dynamic Factor Models Subject to Structural Instability," in *The Methodology and Practice of Econometrics: Festschrift in Honor of D.F. Hendry*, ed. by N. Shephard, and J. Castle, pp. 1–57. Oxford University Press.

——— (2012): "Disentangling the Channels of the 2007-09 Recession," *Brookings Papers on Economic Activity*, pp. 81–156.

Su, L., Z. Shi, and P. C. B. Phillips (2014): "Identifying Latent Structures in Panel Data," *Manuscript, Singapore Management University and Yale University*.

Su, L., and X. Wang (2015): "On Time-Varying Factor Models: Estimation and Testing," Discussion paper, Singapore Management University and University of Chinese Academy of Sciences.

Tibshirani, R. (1994): "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 58(1), 267–288.

Yuan, M., and Y. Lin (2006): "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 68(1), 49–67.

Zou, H. (2006): "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101(476), 1418–1429.

Supplemental Appendix: Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities

Xu Cheng, Zhipeng Liao, and Frank Schorfheide

A The Cross-Validation Algorithm

In order to fine-tune the constants $c = (c_1, c_2) \in \mathcal{C}$ that appear in the penalty weights (4.2) we recommend the cross-validation procedure described in Algorithm 2 below. The algorithm consists of an initialization phase in which the time series X_i are ordered and assigned to n_N groups. (Unlike in the main text, we now use the subscript of X to indicate the series i rather than the period t. Thus, $X_i = (X_{i1}, \ldots, X_{iT})'$.) In unreported simulations we found that the initial ordering of the series improves the performance in situations in which the idiosyncratic components of the X_i series have fairly strong cross-sectional correlation. The main loop computes a mean squared forecast error criterion as a function of the tuning constants c_1 and c_2. The goal is to choose these constants to minimize the forecast error criterion.

Algorithm 2 (Cross Validation)

Initialization:

1. *Start with an arbitrary series and denote it by* X_1. *For* $i = 2$ *to* N*: Choose* X_i *among the remaining* $N - i + 1$ *series such that* $|Corr(X_i, X_{i-1})| \geq |Corr(X_j, X_{i-1})|$ *for any* $j > i$.

2. *Partition* X *in the* N *dimension into* n_N *(approximately) equal size subsamples* $X_{(j_N)}$, $j_N = 1, \ldots, n_N$ *such that* $X = (X_{(1)}, \ldots, X_{(n_N)})$.

Main Loop: *For each* $c = (c_1, c_2) \in \mathcal{C}$ *execute the following steps:*

1. *Iterate over N-prediction samples* $X_{(j_N)}$. *For* $j_N = 1, \ldots, n_N$*:*

   1.1 *Define the N-regression sample* $X_{(-j_N)} = (X_{(1)}, \ldots, X_{(j_N-1)}, X_{(j_N+1)}, \ldots, X_{(n_N)})$.

1.2 Model selection based on N-regression sample. Apply Algorithm 1 to $X_{(-j_N)}$. Denote the selected model by
$\widehat{\mathcal{M}}(-j_N, c) = \big(\widehat{\mathcal{B}}(-j_N, c), \widehat{r}_a(-j_N, c), \widehat{r}_b(-j_N, c)\big)$.

1.3 Factor Estimation based on N-regression sample.
If $\widehat{\mathcal{B}}(-j_N, c) = 0$ (no break), let $F(-j_N, c)$ be the first $\widehat{r}_a(-j_N, c)$ principal components of $X_{(-j_N)}$ (full sample). Split $F(-j_N, c)$ into $F_a(-j_N, c)$ and $F_b(-j_N, c)$. If $\widehat{\mathcal{B}}(-j_N, c) = 1$ (break), let $F_a(-j_N, c)$ be the first $\widehat{r}_a(-j_N, c)$ principal components of $X_{a,(-j_N)}$ (pre-break sample) and $F_b(-j_N, c)$ be the first $\widehat{r}_b(-j_N, c)$ principal components of $X_{b,(-j_N)}$ (post-break sample).

1.4 Rolling Forecasts along T dimension for N-prediction sample. For $j_T = 1, \ldots, n_T$:

1.4.1 Partition $X_{(j_N)}$ into $X_{a,(j_N)}$, $X_{b,(j_N)}$. Extract regression (R) and prediction (P) samples along the T dimension from $X_{a,(j_N)}$, $X_{b,(j_N)}$, $F_a(-j_N, c)$, and $F_b(-j_N, c)$ as follows (to simplify the notation we assume that $T_a/(2n_T)$ and $T_b/(2n_T)$ are integers):

$$X_a^R(j_T, j_N, c),\ F_a^R(j_T, -j_N, c) \ : \ t = 1 + (j_T - 1)\frac{T_a}{2n_T}, \ldots, \frac{T_a}{2} + (j_T - 1)\frac{T_a}{2n_T}$$

$$X_a^P(j_T, j_N, c),\ F_a^P(j_T, -j_N, c) \ : \ t = \frac{T_a}{2} + (j_T - 1)\frac{T_a}{2n_T} + 1, \ldots, \frac{T_a}{2} + j_T\frac{T_a}{2n_T}$$

The first regression sample is $t = 1, \ldots, T_a/2$ and the first prediction sample is $t = T_a/2+1, \ldots, T_a/2+T_a/(2n_T)$. The subsequent samples are shifted forward by $T_a/(2n_T)$ observations. The definitions of $X_b^R(j_T, j_N, c)$, $X_b^P(j_T, j_N, c)$, $F_b^R(j_T, j_N, c)$, $F_b^P(j_T, j_N, c)$ are obtained by replacing the a subscripts with b subscripts.

1.4.2 Estimate $\widehat{\Lambda}(j_T, j_N, c)$ and $\widehat{\Psi}(j_T, j_N, c)$ by regressing $X_a^R(j_T, j_N, c)$ on $F_a^R(j_T, -j_N, c)$ and $X_b^R(j_T, j_N, c)$ on $F_b^R(j_T, -j_N, c)$, respectively. If $\widehat{\mathcal{B}}(-j_N, c) = 0$ combine the two samples.

1.4.3 Compute the forecasts
$\widehat{X}_a^P(j_T, j_N, c) = F_a^P(j_T, -j_N, c)\widehat{\Lambda}'(j_T, j_N, c)$ and
$\widehat{X}_b^P(j_T, j_N, c) = F_b^P(j_T, -j_N, c)\widehat{\Psi}'(j_T, j_N, c)$ and
the forecast errors (omitting subscripts) $X^P - \widehat{X}^P$.

2. Compute the mean-squared forecast error

$$MSFE(c) = \sum_{j_N=1}^{n_N}\sum_{j_T=1}^{n_T}\left(\left\|X_a^P(j_T, j_N, c) - \widehat{X}_a^P(j_T, j_N, c)\right\|^2 + \left\|X_b^P(j_T, j_N, c) - \widehat{X}_b^P(j_T, j_N, c)\right\|^2\right).$$

Table S-1: KNOWN BREAK DATE, HOMOGENEOUS $R^2, \pi_0 = 0.5,\ k = 12$

DGP Configuration					Model Selection			Forecast
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	0.88	1.00	0.88	0.94
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 mean-squared-forecast error 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.

B Additional Monte Carlo Results

Table S-1 provides some additional Monte Carlo results. The DGP for Table S-1 is identical to that for Table 2 but the maximum number of factors is changed from $k = 8$ to $k = 12$. It turns out that the results reported in the main text are robust to an increase in k.

Tables S-2 and S-3 provide comparisons with alternative procedures. Three factor estimation procedures are 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 and report the probabilities of $\widehat{r}_a = r_a$ and $\widehat{r}_b = r_b$. For Bai and Ng (2002), we report results from the IC_{p2} criterion,

Table S-2: Comparison with Factor Number Estimation Procedures

DGP Config.				$\Pr(\widehat{r}_a = r_a)$				$\Pr(\widehat{r}_b = r_b)$			
r_a	r_b	N	T	CLS	BN	ON	AH	CLS	BN	ON	AH
Panel A. Heterogeneous R^2, $\pi_0 = 0.8$, $\alpha = \beta = 0.2$											
3	3	100	200	1.00	0.90	1.00	0.98	1.00	0.04	0.58	0.35
3	3	150	300	1.00	1.00	0.99	1.00	1.00	0.45	0.98	0.78
1	2	100	200	1.00	1.00	0.98	1.00	1.00	0.99	0.98	0.98
1	2	150	300	1.00	1.00	0.98	1.00	1.00	1.00	0.99	1.00
3	4	100	200	0.99	0.58	0.99	0.95	0.44	0.18	0.89	0.61
3	4	150	300	1.00	1.00	0.99	1.00	0.98	0.87	1.00	0.96
Panel B. Homogenous R^2, $\pi_0 = 0.5$, $\alpha = \beta = 0.5$											
3	3	100	100	0.94	0.70	0.87	0.78	0.94	0.72	0.88	0.78
3	3	200	200	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00
1	2	100	100	0.99	0.79	0.93	1.00	0.95	0.75	0.96	0.98
1	2	200	200	1.00	1.00	0.97	1.00	1.00	1.00	0.98	1.00
3	4	100	100	0.89	0.70	0.92	0.85	0.68	0.68	0.65	0.62
3	4	200	200	0.99	1.00	0.99	1.00	1.00	1.00	0.99	1.00

Notes: Known break date, temporal correlation $\rho_a = \rho_b = 0.5$. CLS is our proposed shrinkage estimator; BN is Bai and Ng (2002); ON is Onatski (2010); AH is Ahn and Horenstein (2013).

which is widely applied in practice. For Onatski (2010), the result is based on the edge distribution algorithm suggested in the paper. For Ahn and Horenstein (2013), we compute both the eigenvalue ratio and the growth ratio results and report the latter for its better performance in most of the experiments. Because the alternative factor selection procedures are not designed to handle unknown break dates we focus on the known-break-date case.

The results for Experiment 2 (known break date) and Experiment 3 are summarized in Table S-2. We do not tabulate results for Experiment 1, because under the homogeneous R^2 design with $\pi_0 = 0.5$ all procedures performed very well and are able to determine the true number of pre- and post-break factors with probability close to one. Under the heterogeneous R^2 design some important differences emerge. Except for the case of $r_a = 3$, $r_b = 4$, $N = 100$ and $T = 200$, our procedure does very well in determining r_a and r_b and weakly dominates all the alternative procedures.

Most procedures select the correct number of pre-break factors with probability 0.98 or higher. The determination of the post-break factors is more difficult because the subsample is short. While our shrinkage procedure is able to determine the correct number of factors

with probability one, for the case of $r_a = r_b = 3$ and $N = 100$ and $T = 200$ the success rates of the alternative procedures are below 60%. Only if the number of factors changes from 3 to 4 is our procedure dominated when it comes to the estimation of the post-break factors. For the small sample size, ON and AH have success rates of 89% and 61%, respectively, whereas our shrinkage estimator determines the correct number of post-break factors with probability 44%.[13]

The second panel of Table S-2 considers the case of strong cross-sectional correlation with a homogeneous R^2. Compared to the almost perfect selection in Experiment 1, which only had modest cross-sectional correlation, the error rates increase with the degree of cross-sectional correlation. Nonetheless, the shrinkage procedure performs overall quite well. In most of the configurations, our shrinkage estimator determines the correct number of factors with probability of 95% or more. Moreover, it tends to dominate the alternative procedures – sometimes by a wide margin.

We now turn to the comparison with existing break test procedures. For Breitung and Eickmeier (2011), we reports result based on both the series-by-series dynamic LM test and the pooled dynamic LM test.[14] For Chen, Dolado, and Gonzalo (2014) and Han and Inoue (2014), the reported results are for the Wald test.[15] For each test, we report the probability of rejecting the null hypothesis, setting the nominal size of the tests to $\alpha = 0.05$. For the series-by-series test of Breitung and Eickmeier (2011), the rejection probability is averaged across series. The Bai and Ng (2002) IC_{p2} criterion is used to select the number of factors for all of these procedures.[16]

Table S-3 reports rejection probability of the null hypothesis that there is no break in the loadings. For our shrinkage procedure, a rejection corresponds to $\widehat{\mathcal{B}} = 1$. Under

[13]For the comparison between our procedure and the alternatives, the estimation of r_a is more fair than that of r_b. The reason is that the change in loadings is used to detect type-1 instability as well as estimate r_b for a type-2 instability. The goal to detect a type-1 instability reduces its power to estimate r_b correctly for a type-2 instability, see the rotation in Step 1.5 of Algorithm 1.

[14]The lag order is selected by the Bayesian information criterion (BIC) with a maximum of three lags.

[15]In Breitung and Eickmeier (2011), the HAC estimator of the covariance uses the truncation rule suggested in the paper. For Chen, Dolado, and Gonzalo (2014) and Han and Inoue (2014), the Bartlett kernel is used following suggestions in these papers.

[16]For Breitung and Eickmeier (2011) test, Bai and Ng (2002)'s method is first applied to estimate the number of factors in pre- and post-break subsamples. These two numbers are the same by assumption. However, when the two estimators are different, the larger one is used as the number of factors in the subsequent test.

Table S-3: Comparison with Testing Procedures

DGP Configuration					Reject. Prob. of No-Break Hyp.				
r_a	r_b	**w**	N	T	CLS	BE(S)	BE(P)	CDG	HI
Panel A. Known Break Date. Heterogeneous R^2, $\pi = 0.8$, $\alpha = \beta = 0.2$									
3	3	0	100	200	0.00	0.09	0.07	0.07	0.04
3	3	0	150	300	0.00	0.09	0.06	0.07	0.06
3	3	0.5	100	200	0.00	0.30	0.26	0.15	0.25
3	3	0.5	150	300	0.01	0.38	0.33	0.18	0.66
3	3	1	100	200	0.93	0.43	0.38	0.20	0.47
3	3	1	150	300	1.00	0.50	0.45	0.31	0.89
Panel B. Known Break Date. Homogenous R^2, $\pi_0 = 0.5$, $\alpha = \beta = 0.5$									
3	3	0	100	100	0.01	0.23	0.07	0.09	0.04
3	3	0	200	200	0.00	0.20	0.06	0.09	0.06
3	3	0.2	100	100	0.03	0.31	0.13	0.10	0.13
3	3	0.2	200	200	0.00	0.36	0.22	0.09	0.39
3	3	0.5	100	100	0.91	0.60	0.48	0.25	0.60
3	3	0.5	200	200	1.00	0.79	0.76	0.56	0.98
3	3	1	100	100	1.00	0.68	0.60	1.00	0.17
3	3	1	200	200	1.00	0.87	0.85	1.00	0.60
Panel C. Unknown Break Date. Heterogeneous R^2, $\pi = 0.8$, $\alpha = \beta = 0.2$									
3	3	0	100	200	0.00	0.08	0.08	0.06	0.04
3	3	0	150	300	0.00	0.09	0.06	0.07	0.06
3	3	0.5	100	200	0.00	0.24	0.20	0.16	0.36
3	3	0.5	150	300	0.00	0.37	0.32	0.21	0.35
3	3	1	100	200	0.69	0.42	0.37	0.30	0.39
3	3	1	150	300	0.97	0.50	0.44	0.48	0.31

Notes: Temporal correlation $\rho_a = \rho_b = 0.5$. CLS is our proposed shrinkage estimator; BN is Bai and Ng (2002); BE is Breitung and Eickmeier (2011), (S) is series-by-series and (P) is pooled; CDG is Chen, Dolado, and Gonzalo (2014); HI is Han and Inoue (2014). The nominal size for the hypothesis tests is $\alpha = 0.05$.

Experiment 2 (known break date), our procedure is able to detect the absence of a break if $\mathcal{B}_0 = 0$ and the presence of a break if the break in loadings is large ($\mathbf{w} = 1$). Because consistent model selection procedures generally have low power in distinguishing very similar model specifications, the shrinkage estimator is unable to detect small changes in the loadings ($\mathbf{w} = 0.5$). The hypothesis testing procedures, by design, generate a non-zero type-1 error and in return have some local power. In turns out that our shrinkage procedure is more likely to detect large breaks than any of the competing test procedures. We obtain qualitatively

similar results for Experiment 3. The ranking among the competing procedures is generally not uniform. For instance, in Experiment 2 the HI test has a lot of power against large breaks, whereas in Experiment 3 the CDG test is able to beat the other three testing procedures for $\mathbf{w} = 1$. The BE(S) test performs well in terms of power for $\mathbf{w} = 0.2$ and $\mathbf{w} = 0.5$ but also suffers from size distortions under the null hypothesis.

While the results in Panels A and B of Table S-3 pertain to the case in which the break date is known, we also consider the case of unknown break date in Panel C. For the alternative test procedures we follow the authors' recommendations on how to implement the unknown-break-date case. The results in Panel C are very similar to the known-break-date case in Panel A.

C Supplemental Tables for Empirical Analysis

The variables used in the empirical analysis are grouped into 12 distinct categories, which are summarized in Table S-4. Tables S-5 to S-7 provide a complete list of variables.

Table S-4: Categories of Time Series

Symbol	Description	Series
NIPA	National Income and Product Accounts	5
IP	Industrial Production	9
Emp	Employment and Unemployment	30
HSS	Housing Starts	6
Ord	Orders, Inventories, and Sales	7
Pri	Productivity	22
IntL	Interest Rates (Level)	2
IntS	Interest Rates (Spread)	7
Mon	Money and Credit	5
StPr	Stock Prices and Wealth	3
ExR	Exchange Rates	5
Others	Consumer Expectation	1

Table S-5: List of Variables - Part 1

Name	Category	TC	Long Description
Cons: Dur	NIPA	5	Real Personal Consumption Expenditures: Durable Goods
Cons: Svc	NIPA	5	Real Personal Consumption Expenditures: Services
Cons: NonDur	NIPA	5	Real Personal Consumption Expenditures: Nondurable Goods
Real InvtCh	NIPA	1	Component for Change in Private Inventories, deflated by JCXFE
Real WageG	NIPA	5	Component for Government GDP: Wage and Salary Disbursements by Industry, Government, NIPA Tables 2.7A and 2.7B, deflated by JCXFE
IP: DurGds materials	IP	5	Industrial Production: Durable Materials
IP: NondurGds materials	IP	5	Industrial Production: Nondurable Materials
IP: DurConsGoods	IP	5	Industrial Production: Durable Consumer Goods
IP: Auto	IP	5	IP: Automotive products
IP: NonDurConsGoods	IP	5	Industrial Production: Nondurable Consumer Goods
IP: BusEquip	IP	5	Industrial Production: Business Equipment
IP: EnergyProds	IP	5	IP: Consumer Energy Products
CapU Tot	IP	1	Capacity Utilization: Total Industry
CapU Man	IP	1	Capacity Utilization: Manufacturing (FRED past 1972)
Emp: DurGoods	Emp	5	All Employees: Durable Goods Manufacturing
Emp: Const	Emp	5	All Employees: Construction
Emp: Edu&Health	Emp	5	All Employees: Education & Health Services
Emp: Finance	Emp	5	All Employees: Financial Activities
Emp: Infor	Emp	5	All Employees: Information Services
Emp: Bus Serv	Emp	5	All Employees: Professional & Business Services
Emp: Leisure	Emp	5	All Employees: Leisure & Hospitality
Emp: OtherSvcs	Emp	5	All Employees: Other Services
Emp: Mining/NatRes	Emp	5	All Employees: Natural Resources & Mining
Emp: Trade&Trans	Emp	5	All Employees: Trade, Transportation & Utilities
Emp: Retail	Emp	5	All Employees: Retail Trade
Emp: Wholesal	Emp	5	All Employees: Wholesale Trade
Emp: Gov(Fed)	Emp	5	All Employees: Government: Federal
Emp: Gov (State)	Emp	5	All Employees: Government: State Government
Emp: Gov (Local)	Emp	5	All Employees: Government: Local Government
URate: Age16-19	Emp	2	Unemployment Rate - 16-19 yrs
URate: Age > 20 Men	Emp	2	Unemployment Rate - 20 yrs. & over, Men
URate: Age > 20 Women	Emp	2	Unemployment Rate - 20 yrs. & over, Women
U: Dur < 5wks	Emp	5	Number Unemployed for Less than 5 Weeks
U: Dur 5-14wks	Emp	5	Number Unemployed for 5-14 Weeks
U: Dur > 15-26wks	Emp	5	Civilians Unemployed for 15-26 Weeks
U: Dur > 27wks	Emp	5	Number Unemployed for 27 Weeks & over
U: Job Losers	Emp	5	Unemployment Level - Job Losers
U: LF Reentry	Emp	5	Unemployment Level - Reentrants to Labor Force
U: Job Leavers	Emp	5	Unemployment Level - Job Leavers
U: New Entrants	Emp	5	Unemployment Level - New Entrants

Notes: TC is transformation code; see Stock and Watson (2012).

Table S-6: List of Variables - Part 2

Name	Category	TC	Long Description
Emp: SlackWk	Emp	5	Employment Level - Part-Time for Economic Reasons, All Industries
AWH Man	Emp	1	Average Weekly Hours: Manufacturing
AWH Privat	Emp	2	Average Weekly Hours: Total Private Industrie
AWH Overtime	Emp	2	Average Weekly Hours: Overtime: Manufacturing
HPermits	HSS	5	New Private Housing Units Authorized by Building Permit
Hstarts: MW	HSS	5	Housing Starts in Midwest Census Region
Hstarts: NE	HSS	5	Housing Starts in Northeast Census Region
Hstarts: S	HSS	5	Housing Starts in South Census Region
Hstarts: W	HSS	5	Housing Starts in West Census Region
Constr. Contracts	HSS	4	Construction contracts (mil. sq. ft.) (Copyright, McGraw-Hill)
Ret. Sale	Ord	5	Sales of retail stores (mil. Chain 2000 $)
Orders (DurMfg)	Ord	5	Mfrs' new orders durable goods industries (bil. chain 2000 $)
Orders (ConsumerGoods/Mat.)	Ord	5	Mfrs' new orders, consumer goods and materials (mil. 1982 $)
UnfOrders (DurGds)	Ord	5	Mfrs' unfilled orders durable goods indus. (bil. chain 2000 $)
Orders (NonDefCap)	Ord	5	Mfrs' new orders, nondefense capital goods (mil. 1982 $)
VendPerf	Ord	1	Index of supplier deliveries – vendor performance (pct.)
MT Invent	Ord	5	Manufacturing and trade inventories (bil. Chain 2005 $)
PCED-MotorVec	Pri	6	Motor vehicles and parts
PCED-DurHousehold	Pri	6	Furnishings and durable household equipment
PCED-Recreation	Pri	6	Recreational goods and vehicles
PCED-OthDurGds	Pri	6	Other durable goods
PCED-Food-Bev	Pri	6	Food and beverages purchased for off-premises consumption
PCED-Clothing	Pri	6	Clothing and footwear
PCED-Gas-Enrgy	Pri	6	Gasoline and other energy goods
PCED-OthNDurGds	Pri	6	Other nondurable goods
PCED-Housing-Utilities	Pri	6	Housing and utilities
PCED-HealthCare	Pri	6	Health care
PCED-TransSvg	Pri	6	Transportation services
PCED-RecServices	Pri	6	Recreation services
PCED-FoodServ-Acc.	Pri	6	Food services and accommodations
PCED-FIRE	Pri	6	Financial services and insurance
PCED-OtherServices	Pri	6	Other services
PPI: FinConsGds	Pri	6	Producer Price Index: Finished Consumer Goods
PPI: FinConsGds(Food)	Pri	6	Producer Price Index: Finished Consumer Foods
PPI: IndCom	Pri	6	Producer Price Index: Industrial Commodities
PPI: IntMat	Pri	6	Producer Price Index: Intermediate Materials: Supplies & Components
NAPM ComPrice	Pri	1	NAPM COMMODITY PRICES INDEX (PERCENT)
Real Price: NatGas	Pri	5	PPI: Natural Gas, deflated by PCEPILFE
Real Price: Oil	Pri	5	PPI: Crude Petroleum, deflated by PCEPILFE

Notes: TC is transformation code; see Stock and Watson (2012).

Table S-7: List of Variables - Part 3

Name	Category	TC	Long Description
FedFunds	IntL	2	Effective Federal Funds Rate
TB-3Mth	IntL	2	3-Month Treasury Bill: Secondary Market Rate
BAA-GS10	IntS	1	BAA-GS10 Spread
MRTG-GS10	IntS	1	Mortg-GS10 Spread
TB6m-TB3m	IntS	1	tb6m-tb3m
GS1-TB3m	IntS	1	GS1-Tb3m
GS10-TB3m	IntS	1	GS10-Tb3m
CP-TB Spread	IntS	1	CP-Tbill Spread: CP3FM-TB3MS
Ted-Spread	IntS	1	MED3-TB3MS (Version of TED Spread)
Real C&I Loan	Mon	5	Commercial and Industrial Loans at All Commercial BanksDefl by PCEPILFE
Real ConsLoans	Mon	5	Consumer (Individual) Loans at All Commercial Banks
			Outlier Code because of change in data in April 2010 see FRB H8 ReleasDefl by PCEPILFE
Real NonRevCredit	Mon	5	Total Nonrevolving Credit Owned and Securitized, OutstandingDefl by PCEPILFE
Real LoansRealEst	Mon	5	Real Estate Loans at All Commercial BanksDefl by PCEPILFE
Real RevolvCredit	Mon	5	Total Revolving Credit OutstandingDefl by PCEPILFE
S&P500	StPr	5	S&P'S COMMON STOCK PRICE INDEX: COMPOSITE (1941-43=10)
DJIA	StPr	5	COMMON STOCK PRICES: DOW JONES INDUSTRIAL AVERAGE
VXO	StPr	1	VXO (Linked by N. Bloom) .. Average daily VIX from 2009
Ex rate: Major	ExR	5	FRB Nominal Major Currencies Dollar Index (Linked to EXRUS in 1973:1)
Ex rate: Switz	ExR	5	FOREIGN EXCHANGE RATE: SWITZERLAND (SWISS FRANC PER USD)
Ex rate: Japan	ExR	5	FOREIGN EXCHANGE RATE: JAPAN (YEN PER USD)
Ex rate: UK	ExR	5	FOREIGN EXCHANGE RATE: UNITED KINGDOM (CENTS PER POUND)
EX rate: Canada	ExR	5	FOREIGN EXCHANGE RATE: CANADA (CAD PER USD)
Cons. Expectations	Others	1	Consumer expectations NSA (Copyright, University of Michigan)

Notes: TC is transformation code; see Stock and Watson (2012).

D Auxiliary Results

D.1 Normalization of the DGP

The matrices R_a and R_b that normalize the DGP in (2.5) are constructed as follows. Let $\Sigma_a = \Lambda^{0\prime}\Lambda^0/N \in R^{r_a \times r_a}$, let $\Sigma_a^{1/2}$ be the Cholesky factor of Σ_a, and let Υ_a be a matrix of orthonormal eigenvectors such that

$$\Upsilon_a'(\Sigma_a^{1/2})'\Sigma_F\Sigma_a^{1/2}\Upsilon_a = V_a, \tag{D.1}$$

where V_a is a diagonal matrix of eigenvalues, ordered from largest to smallest. Note that by Assumptions A and B, the matrix $(\Sigma_a^{1/2})'\Sigma_F\Sigma_a^{1/2}$ has positive and distinct eigenvalues with large N, which means that (D.1) holds for large N. Now define the transformation matrix

$$R_a = \Sigma_a^{1/2}\Upsilon_a V_a^{-1/2}. \tag{D.2}$$

For the post-break DGP, we let $\Sigma_b = \Psi^{0\prime}\Psi^0/N \in R^{r_b \times r_b}$, substitute Σ_F in (D.1) by $\Sigma_{\overline{F}}$, and otherwise replace a subscripts by b subscripts. The second transformation matrix R_b is defined as

$$R_b = \Sigma_b^{1/2}\Upsilon_b V_b^{-1/2}. \tag{D.3}$$

It can be verified that the transformation induces the desired normalization. For the pre-break period, using Assumption A and the fact that Υ_a is a finite matrix, we have

$$T_a^{-1}F_a^{R\prime}F_a^R = V_a^{-1/2}\Upsilon_a'\Sigma_a^{1/2}\Sigma_F\Sigma_a^{1/2}\Upsilon_a V_a^{-1/2} + O_p(T_a^{-1/2}) = I_{r_a \times r_a} + O_p(T_a^{-1/2}).$$

Moreover, by definition of Σ_a, $N^{-1}\Lambda^{R\prime}\Lambda^R = V_a^{1/2}\Upsilon_a'\Sigma_a^{-1/2}\Sigma_a(\Sigma_a^{-1/2})'\Upsilon_a V_a^{1/2} = V_a$, which is a diagonal matrix, as desired.

The transformation matrices R_a and R_b, defined in (D.2) and (D.3), that were used to normalize the DGP are related to, but essentially different from, their counterparts considered in the literature, such as those in Bai and Ng (2002) and Bai (2003). In the definitions of R_a and R_b, one subtle point is that Σ_a and Σ_b are averages that depend on N, whereas Σ_F and $\Sigma_{\overline{F}}$ are asymptotic limits as $T \to \infty$. This subtle difference is crucial for deriving asymptotic limits of the PLS estimators if potential structural change is considered. In the absence of structural instabilities, $R_a = R_b$ by construction. This immediately implies that $\Gamma^R = 0$ for any N, instead of $\Gamma^R \to 0$, as both N and T go to infinity.

We first present a lemma on the transformation matrices R_a and R_b defined in (D.2) and (D.3) of the main text. This lemma is used in the proof of Theorem 1. Let $\widetilde{F}_a^r \in R^{T_0 \times r_a}$ and $\widetilde{F}_b^r \in R^{(T-T_0)\times r_b}$ denote the first r_a and r_b columns of $\widetilde{F}_a$ and $\widetilde{F}_b$, respectively. The $r_a \times r_a$ diagonal matrix $\widetilde{V}_a$ consists of the first r_a largest eigenvalues of $(T_0 N)^{-1} X_a X_a'$ in a decreasing order, and the $r_b \times r_b$ diagonal matrix $\widetilde{V}_b$ consists of the first r_b largest eigenvalues of $(T_1 N)^{-1} X_b X_b'$ in a decreasing order. Under Assumptions A-D, Theorem 1 of Bai and Ng (2002) shows that

$$T_0^{-1}||\widetilde{F}_a^r - F_a H_a||^2 = O_p(C_{NT_0}^{-2}) \text{ and } T_1^{-1}||\widetilde{F}_b^r - F_b H_b||^2 = O_p(C_{NT_1}^{-2}), \tag{D.4}$$

where

$$H_a = \Sigma_a \frac{F_a' \widetilde{F}_a^r}{T_0} \widetilde{V}_a^{-1} \text{ and } H_b = \Sigma_b \frac{F_b' \widetilde{F}_b^r}{T_1} \widetilde{V}_b^{-1}. \tag{D.5}$$

Lemma 1 *Suppose that Assumptions* A-D *hold. Then,*

$$H_a - R_a = O_p(C_{NT}^{-1}) \text{ and } H_b - R_b = O_p(C_{NT}^{-1}).$$

Proof of Lemma 1. Note that R_a is invertible w.p.a.1. Hence, we can write

$$F_a \Lambda^{0\prime} = F_a R_a R_a^{-1} \Lambda^{0\prime} = F_a^R \Lambda^{R\prime}, \text{ where } F_a^R = F_a R_a \text{ and } \Lambda^{R\prime} = R_a^{-1} \Lambda^{0\prime}. \tag{D.6}$$

The transformed factors satisfy

$$\begin{aligned} \frac{F_a^{R\prime} F_a^R}{T_0} &= V_a^{-1/2} \Upsilon_a' \Sigma_a^{1/2} \frac{F_a' F_a}{T_0} \Sigma_a^{1/2} \Upsilon_a V_a^{-1/2} \\ &= V_a^{-1/2} \left(\Upsilon_a' \Sigma_a^{1/2} \Sigma_F \Sigma_a^{1/2} \Upsilon_a \right) V_a^{-1/2} + O_p(T_0^{-1/2}) \\ &= V_a^{-1/2} \left(V_a \right) V_a^{-1/2} + O_p(T_0^{-1/2}) = I_{r_a} + O_p(T_0^{-1/2}), \end{aligned} \tag{D.7}$$

where the first equality follows from $F_a^R = F_a R_a$ and $R_a = \Sigma_a^{1/2} \Upsilon_a V_a^{-1/2}$, the second equality follows from $F_a' F_a / T_0 - \Sigma_F = O_p(T_0^{-1/2})$ in Assumption A, and the third equality follows from (D.1). The transformed loadings satisfy

$$\frac{\Lambda^{R\prime} \Lambda^R}{N} = V_a^{1/2} \Upsilon_a^{-1} \Sigma_a^{-1/2} \frac{\Lambda^{0\prime} \Lambda^0}{N} \Sigma_a^{-1/2} \Upsilon_a'^{-1} V_a^{1/2} = V_a^{1/2} \Upsilon_a^{-1} \Upsilon_a'^{-1} V_a^{1/2} = V_a, \tag{D.8}$$

where the first equality follows from $\Lambda^{R\prime} = R_a^{-1} \Lambda^{0\prime}$ and $R_a = \Sigma_a^{1/2} \Upsilon_a V_a^{-1/2}$, the second equality follows from $\Sigma_a = \Lambda^{0\prime} \Lambda^0 / N$ by definition, the third equality holds because $\Upsilon_a' \Upsilon_a = I_{r_a}$.

Let L_a be a $r_a \times r_a$ matrix defined as

$$L_a = \frac{\Lambda^{R\prime}\Lambda^R}{N}\frac{F_a^{R\prime}\widetilde{F}_a^r}{T_0}\widetilde{V}_a^{-1}, \tag{D.9}$$

which is a transformation matrix analogous to H_a but with F_a and Λ^0 replaced by F_a^R and Λ^R, respectively. Stock and Watson (2002) and Bai and Ng (2002) show that L_a is invertible w.p.a.1 and $\widetilde{F}_a^r$ is a consistent estimator of $F_a^R L_a$. The transformation matrix H_a and the new transformation matrix L_a satisfy

$$\begin{aligned} H_a &= R_a\frac{R_a^{-1}\Lambda^{0\prime}\Lambda^0 R_a^{\prime-1}}{N}\frac{R_a^{\prime}F_a^{\prime}\widetilde{F}_a^r}{T_0}\widetilde{V}_a^{-1} \\ &= R_a\frac{\Lambda^{R\prime}\Lambda^R}{N}\frac{F_a^{R\prime}\widetilde{F}_a^r}{T_0}\widetilde{V}_a^{-1} = R_aL_a, \end{aligned} \tag{D.10}$$

where the first equality follows from the definition of H_a in (D.5), the second equality follows from $F_a^R = F_aR_a$ and $\Lambda^{R\prime} = R_a^{-1}\Lambda^{0\prime}$, the third equality follows from the definition of L_a in (D.9).

Equation (2) of Bai and Ng (2013) shows that $L_a = I_{r_a}$ if the underlying factor matrix F_a^R satisfies $F_a^{R\prime}F_a^R/T_0 = I_r$, and the underlying loading matrix Λ^R satisfies that $\Lambda^{R\prime}\Lambda^R$ is a diagonal matrix with distinct elements. By (D.7) and (D.8), we know that these conditions are satisfied asymptotically by the transformation above. Following the arguments for equation (2) of Bai and Ng (2013), we obtain

$$L_a = I_{r_a} + O_p(C_{NT_0}^{-1}), \tag{D.11}$$

with two modifications to the proof in Bai and Ng (2013): (i) $T_0^{-1}(\widetilde{F}_a^r - F_a^RL_a)'F_a^R = O_p(C_{NT_0}^{-2})$ in Bai and Ng (2013) is changed to $T_0^{-1}(\widetilde{F}_a^r - F_a^RL_a)'F_a^R = O_p(C_{NT_0}^{-1})$, which follows from $F_a^RL_a = F_aH_a$, (D.4), and the Cauchy-Schwarz inequality, and (ii) $F_a^{R\prime}F_a^R/T_0 = I_{r_a}$ is changed to $F_a^{R\prime}F_a^R/T_0 = I_{r_a} + O_p(T_0^{-1/2})$ and the $O_p(T_0^{-1/2})$ term is absorbed in $O_p(C_{NT_0}^{-1})$ in (D.11). The reason for the first change is that Assumptions A-D in this paper are comparable to Assumptions A – D of Bai and Ng (2002), which are weaker than similar assumptions in Bai and Ng (2013). The Assumptions in Bai and Ng (2013) are needed to obtain asymptotic distributions of the estimated factors and loadings, which is not the purpose here. After making these two modifications above, the rest of the arguments for equation (2) of Bai and Ng (2013) follow directly to yield the result in (D.11).

Combining the results in (D.10) and (D.11), we obtain $H_a - R_a = O_p(C_{NT}^{-1})$ because $T_0/T \to \pi_0 \in (0,1)$. Similar arguments give $H_b - R_b = O_p(C_{NT}^{-1})$. □

D.2 Identification of the Break Date

We will discuss the case of $\pi < \pi_0$ in detail. For the first subsample $X_a(\pi)$, the DGP is the same as that in (2.2), which can be written as

$$\begin{aligned} X_a(\pi) &= F_a(\pi)\Lambda^{0\prime} + e_a(\pi), \text{ where} \\ F_a(\pi) &= \left(F_1^0, \ldots, F_{T_a}^0\right)' \in R^{T_a \times r_a}, \\ e_a(\pi) &= (e_1, \ldots, e_{T_a})' \in R^{T_a \times N}. \end{aligned} \tag{D.12}$$

For the second subsample $X_b(\pi)$, which includes observations for $t = T_a + 1, \ldots, T_0, \ldots, T$, the DGP corresponds to (2.2) for $t \leq T_0$ and to (2.3) for $t > T_0$. Thus, the DGP for $X_b(\pi)$ can be written as

$$\begin{aligned} X_b(\pi) &= F_a^+(\pi)\Lambda^{0\prime} + F_b(\pi)\Psi^{0\prime} + e_b(\pi), \text{ where} \\ F_a^+(\pi) &= (F_{T_a+1}^0, ..., F_{T_0}^0, 0_{r_a \times (T-T_0)})' \in R^{T_b \times r_a}, \\ F_b(\pi) &= (0_{r_b \times (T_0 - T_a)}, \overline{F}_{T_0+1}^0, ..., \overline{F}_T^0)' \in R^{T_b \times r_b}, \\ e_b(\pi) &= (e_{T_a+1}, ..., e_T)' \in R^{T_b \times N}. \end{aligned} \tag{D.13}$$

Here the r_a factors in $F_a^+(\pi)$ with loadings Λ^0 are only for observations before the true break date, and the r_b factors in $F_b(\pi)$ with loadings Ψ^0 are only for observations after the true break. By construction, $F_a^+(\pi)$ and $F_b(\pi)$ are orthogonal to each other. By definition, $F_a(\pi_0) = F_a$, $F_a^+(\pi_0) = 0$, and $F_b(\pi_0) = F_b$. The DGPs in (D.12) and (D.13) reduce to (2.2) and (2.3), respectively, if $\pi = \pi_0$ and $T_a = T_0$.

Let $r_a(\pi)$ and $r_b(\pi)$ denote the number of factors in $X_a(\pi)$ and $X_b(\pi)$, respectively. By definition, they are 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 $N, T \to \infty$. We extend Assumption A to Assumption A* for the uniform convergence of the factor covariance matrices over time.

Assumption A*. $\mathbb{E}\left\|F_t^0\right\|^4 \leq C$, $\mathbb{E}||\overline{F}_t^0||^4 \leq C$ and there exist some positive definite matrices Σ_F and $\Sigma_{\overline{F}}$ such that $T^{-1}\sum_{t=1}^{\lfloor T\pi \rfloor} F_t^0 F_t^{0\prime} = \pi\Sigma_F + O_p(T^{-1/2})$ for $\pi \leq \pi_0$ and $T^{-1}\sum_{t=\lfloor T\pi \rfloor+1}^{T} \overline{F}_t^0 \overline{F}_t^{0\prime} = (1-\pi)\Sigma_{\overline{F}} + O_p(T^{-1/2})$ for $\pi \geq \pi_0$, where both $O_p(T^{-1/2})$ terms are uniform over $\pi \in \Pi$. □

Assumption A* assumes covariance stationarity, as in the instability test by Andrews (1993). It can be generalized to time-varying variance as in Cheng and Phillips (2012) under some additional conditions. Section 5 provides regularity conditions (Assumptions C* and

D) on the idiosyncratic errors similar to those in Bai and Ng (2002). We analyze $r_a(\pi)$ and $r_b(\pi)$ under these assumptions.

For the first subsample, the DGP in (D.12) is a factor model with r_a factors (i.e., $r_a(\pi) = r_a$ for $\pi \leq \pi_0$) because there is no break in this subsample. To study the number of factors in (D.13) for the second subsample, note that

$$\begin{aligned} &T^{-1}(F_a^+(\pi),F_b(\pi))'(F_a^+(\pi),F_b(\pi)) \to_p \Sigma_F^+(\pi), \\ &N^{-1}\left(\Lambda^0,\Psi^0\right)'\left(\Lambda^0,\Psi^0\right) \to_p \Sigma_{\Lambda\Psi}^+, \text{ where} \\ &\Sigma_F^+(\pi) = \begin{bmatrix} (\pi_0-\pi)\Sigma_F & 0_{r_a\times r_b} \\ 0_{r_b\times r_a} & (1-\pi_0)\Sigma_{\overline{F}} \end{bmatrix} \end{aligned} \tag{D.14}$$

and $\Sigma_{\Lambda\Psi}^+$ is defined in (2.7). Because $\Sigma_F^+(\pi)$ has full rank by construction, the number of factors in $X_b(\pi)$ depends on the rank of $\Sigma_{\Lambda\Psi}^+$ (i.e., $r_b(\pi) = \mathit{rank}(\Sigma_{\Lambda\Psi}^+)$ for $\pi < \pi_0$). If $\Lambda^0 = \Psi^0$, we know that $\mathit{rank}(\Sigma_{\Lambda\Psi}^+) = r_b$. If, on the other hand, the column spaces generated by Ψ^0 and Λ^0 do not overlap, we have $\mathit{rank}(\Sigma_{\Lambda\Psi}^+) = r_a + r_b$. Typically, we would expect the column spaces generated by Ψ^0 and Λ^0 to be overlapping but non-nested, which means $\mathit{rank}(\Sigma_{\Lambda\Psi}^+) > r_b \geq r_a$ and hence $r_b(\pi) > r_b$ for $\pi < \pi_0$.

Note that $r_a(\pi)$ and $r_b(\pi)$ can be derived for the case $\pi > \pi_0$ using the same steps as above. For the first subsample, we have $r_a(\pi) = \mathit{rank}(\Sigma_{\Lambda\Psi}^+)$, which implies $r_a(\pi) \geq r_b \geq r_a$, while for the second subsample, we simply get $r_b(\pi) = r_b$.

D.3 Choosing κ_1 and κ_2

In choosing the scaling constants κ_1 and κ_2, we consider the optimality conditions that lead the PLS estimators to have zero solutions for some columns in Λ and/or Γ. Intuitively, the criterion function in (3.2) is minimized at 0 if the marginal penalty for deviating from 0 is larger than the marginal gain on the least square criterion function. Translated into our notation, $||\widehat{\Lambda}_\ell|| = 0$ for $\ell > r_a$ if[17]

$$\left\| e_a'(\widehat{\Lambda})\widetilde{F}_{a,\ell} + e_b'(\widehat{\Lambda}+\widehat{\Gamma})\widetilde{F}_{b,\ell} \right\| < NT\alpha_{NT}\omega_\ell^\lambda/2, \tag{D.15}$$

where the residual matrices were defined as

$$e_a(\Lambda) = X_a - \widetilde{F}_a\Lambda' \text{ and } e_b(\Lambda+\Gamma) = X_b - \widetilde{F}_b(\Lambda+\Gamma)'. \tag{D.16}$$

[17]See Lemma 4.2 of Bühlmann and van de Geer (2011).

The inequality in (D.15) suggests that doubling every element in the residual matrices $e_a(\Lambda)$ and $e_b(\Lambda+\Gamma)$ has to be compensated for by doubling κ_1 to ensure that the inequality in (D.15) holds. Therefore, to standardize the left-hand side of (D.15), a reasonable choice of κ_1 is

$$\kappa_1 = \left\{ (NT_a)^{-1/2} \left\| e_a(\widetilde{\Lambda}) \right\| + (NT_b)^{-1/2} \left\| e_b(\widetilde{\Lambda}+\widetilde{\Gamma}) \right\| \right\}. \tag{D.17}$$

A similar argument can be made for κ_2. We introduce constants c_1 and c_2 that can potentially differ from 1 to fine-tune the penalty in small samples.

D.4 Additional Technical Assumptions

Suppose $T_0/T \to \tau_0$ for some constant $\tau_0 \in (0,1)$ as $T \to \infty$. We assume the following assumptions in addition to Assumptions A and B. Let $e = [e_1, \ldots, e_T] \in R^{N\times T}$ be the matrix of idiosyncratic errors and e_{it} denote the (i,t) element of e that is associated with series i in period t.

Assumption C. (i). $\mathbb{E}[e_{it}] = 0$, $\mathbb{E}[|e_{it}|^8] \leq C$;

(ii). $\mathbb{E}[N^{-1}\sum_{i=1}^N e_{is}e_{it}] = \sigma_N(s,t)$, $|\sigma_N(s,s)| \leq C$ for all s, $T^{-1}\sum_{s=1}^T \sum_{t=1}^T |\sigma_N(s,t)| \leq C$;

(iii). $\mathbb{E}[e_{it}e_{jt}] = \tau_{ij,t}$ with $|\tau_{ij,t}| \leq |\tau_{ij}|$ for some τ_{ij} and for all t, and $N^{-1}\sum_{i=1}^N \sum_{j=1}^N |\tau_{ij}| \leq C$;

(iv). $\mathbb{E}[e_{it}e_{js}] = \tau_{ij,ts}$ and $(NT)^{-1}\sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T |\tau_{ij,ts}| \leq C$;

(v). For every (t,s), $\mathbb{E}[|N^{-1/2}\sum_{i=1}^N [e_{is}e_{it} - \mathbb{E}[e_{is}e_{it}]]|^4] \leq C$;

(vi). $\rho_1((NT)^{-1}e_a e_a') = O_p(\max[N^{-1}, T^{-1}])$ and $\rho_1((NT)^{-1}e_b e_b') = O_p(\max[N^{-1}, T^{-1}])$. □

Assumption D. $\mathbb{E}[N^{-1}\sum_{i=1}^N ||T^{-1/2}(\sum_{t=1}^{T_0} F_t^0 e_{it} + \sum_{t=T_0+1}^{T} \overline{F}_t^0 e_{it})||^2] \leq C$. □

Assumptions C and D are analogous to Assumptions C and D of Bai and Ng (2002). Assumption C allows for time-series and cross-sectional weak dependence in the idiosyncratic errors. Assumption C(vi) or a similar condition is needed for the consistent selection of the number of factors (see Amengual and Watson (2007)). Assumption D allows for weak dependence between the factors and the idiosyncratic errors.

To handle the unknown break date case, we strengthen Assumption C on the idiosyncratic errors to Assumption C* such that the weak dependence and stationarity hold for any subsamples considered.

Assumption C*. Assumption C holds with e_a and e_b replaced by $e_a(\pi)$ and $e_b(\pi)$ and Assumption C(vi) holds uniformly over $\pi \in \Pi$. □

D.5 Technical Results Used To Prove Main Theorems

The following Lemma provides a link between the identification assumptions, in particular Assumption ID, and the consistent model determination.

Lemma 2 *Suppose Assumptions* A-D *hold. Then,*

(a) *Pre-break factors:* $N^{-1}\left\|\Lambda_\ell^R\right\|^2 = \rho_\ell(\Sigma_\Lambda\Sigma_F) + o(1)$ *for* $\ell = 1, \ldots, r_a$*;*

(b) *New factors: If* $r_b > r_a$*,* $N^{-1}||\Gamma_\ell^R||^2 = \rho_\ell(\Sigma_\Psi\Sigma_{\overline{F}}) + o(1)$ *for* $\ell = r_a + 1, \ldots, r_b$*;*

(c) *Change to column space of loadings: If* $r_b = r_a$ *and* $\text{rank}(\Sigma^+_{\Lambda\Psi}) > r_a$*,* $N^{-1}\Gamma^{R\prime}\Gamma^R \to \Sigma_\Gamma$ *for some* $\Sigma_\Gamma \neq 0$ *as* $N \to \infty$*;*

(d) *Change to scaling of loadings:* $N^{-1}||\Gamma_\ell^R||^2 \geq [\sqrt{\rho_\ell(\Sigma_\Psi\Sigma_{\overline{F}})} - \sqrt{\rho_\ell(\Sigma_\Lambda\Sigma_F)}]^2 + o(1)$ *for* $\ell = 1, \ldots, r_a$*.*

Proof of Lemma 2. Because $\Lambda^R = \Lambda^0 R_a^{-1\prime}$ and $\Psi^R = \Psi^0 R_b^{-1\prime}$ with $R_a = \Sigma_a^{1/2}\Upsilon_a V_a^{-1/2}$ and $R_b = \Sigma_b^{1/2}\Upsilon_b V_b^{-1/2}$, we have

$$\frac{\Lambda^{R\prime}\Lambda^R}{N} = V_a^{1/2}\Upsilon_a'\Sigma_a^{-1/2}\frac{\Lambda^{0\prime}\Lambda^0}{N}\Sigma_a^{-1/2}\Upsilon_a V_a^{1/2} = V_a \text{ and } \frac{\Psi^{R\prime}\Psi^R}{N} = V_b. \tag{D.18}$$

By definition, V_a is a diagonal matrix and its ℓ-th diagonal element is the ℓ-th largest eigenvalue of $\Sigma_a^{1/2}\Sigma_F\Sigma_a^{1/2}$, which is the same as the ℓ-th largest eigenvalue of $\Sigma_a\Sigma_F$. Following Assumption B and the continuity of the eigenvalue (with respect to the matrix), it converges to the ℓ-th largest eigenvalue of $\Sigma_\Lambda\Sigma_F$, denoted by $\rho_\ell(\Sigma_\Lambda\Sigma_F)$. Similarly, the ℓ-th diagonal element of V_b converges to the ℓ-th largest eigenvalue of $\Sigma_\Psi\Sigma_{\overline{F}}$, denoted by $\rho_\ell(\Sigma_\Psi\Sigma_{\overline{F}})$.

Let a_ℓ be a selection vector that selects the ℓ-th column of a matrix. Part (a) holds because

$$N^{-1}\left\|\Lambda_\ell^R\right\|^2 = a_\ell'\left(N^{-1}\Lambda^{R\prime}\Lambda^R\right)a_\ell = a_\ell' V_a a_\ell = \rho_\ell(\Sigma_\Lambda\Sigma_F) + o(1). \tag{D.19}$$

To prove part (b), note that for $r_a < \ell \leq r_b$, the ℓ-th column of Γ^R is equivalent to the ℓ-th column of Ψ^R. Hence,

$$N^{-1}||\Gamma_\ell^R||^2 = a_\ell'\left(N^{-1}\Psi^{R\prime}\Psi^R\right)a_\ell = a_\ell' V_b a_\ell = \rho_\ell(\Sigma_\Psi\Sigma_{\overline{F}}) + o(1). \tag{D.20}$$

To show part (c), first note that if $r_a = r_b$, we have

$$N^{-1}\Gamma^{R\prime}\Gamma^R = N^{-1}(\Psi^R - \Lambda^R)'(\Psi^R - \Lambda^R) = \mathbf{e}'\Sigma^+_{\Lambda\Psi}\mathbf{e} + o(1), \tag{D.21}$$

where $\mathbf{e} = \lim_{N\to\infty}\left(R_a^{-1},\ -R_b^{-1}\right)'$ has full rank following Assumptions A and B and $\Sigma^+_{\Lambda\Psi}$ is defined in (2.7). By a Cholesky decomposition, write $\Sigma^+_{\Lambda\Psi} = (\Sigma^+_{\Lambda\Psi})^{1/2}(\Sigma^+_{\Lambda\Psi})^{1/2}$ with $rank((\Sigma^+_{\Lambda\Psi})^{1/2}) = rank(\Sigma^+_{\Lambda\Psi}) > r_a$. For a $2r_a \times 2r_a$ matrix $(\Sigma^+_{\Lambda\Psi})^{1/2}$, the rank of the null space of $(\Sigma^+_{\Lambda\Psi})^{1/2}$ is smaller than r_a. It follows that $(\Sigma^+_{\Lambda\Psi})^{1/2}\mathbf{e} \neq 0$ because $rank(\mathbf{e}) = r_a$, and this immediately implies that part (c) holds with $\Sigma_\Gamma = \mathbf{e}'\Sigma^+_{\Lambda\Psi}\mathbf{e} \neq 0$.

To prove part (d), write

$$\begin{aligned} N^{-1}||\Gamma^R_\ell||^2 &= N^{-1}\left\|\Gamma^R a_\ell\right\|^2 = N^{-1}\left\|\Psi^R a_\ell - \Lambda^R a_\ell\right\|^2 \\ &\geq \left(N^{-1/2}\left\|\Psi^R a_\ell\right\| - N^{-1/2}\left\|\Lambda^R a_\ell\right\|\right)^2 \\ &= [(\rho_\ell(\Sigma_\Psi\Sigma_{\overline{F}}))^{1/2} - (\rho_\ell(\Sigma_\Lambda\Sigma_F))^{1/2}]^2 + o(1), \end{aligned} \tag{D.22}$$

where the first two equalities follow from the definition of a_ℓ and Γ^R, the inequality follows from the triangle inequality, and the last equality holds by (D.18). □

The following theorem is used in the proof of Theorem 4. Let $\|\cdot\|_{op}$ denote the operator norm:

$$\|A\|_{op} = \sup_{\gamma'\gamma\leq 1}\|A\gamma\| \tag{D.23}$$

for any real matrix A.

Theorem 5 (Weyl's Eigenvalue Perturbation Theorem) *Let A and B are $k\times k$ symmetric real matrices. Then*

$$\max_{1\leq\ell\leq k}|\rho_\ell(A) - \rho_\ell(B)| \leq \|A - B\|_{op}.$$

Theorem 5 is a simplified version of Corollary III.2.6 in Bhatia (1997).

E Proof of Results in Section 5

Recall that we have defined

$$\Lambda^R = \Lambda^0(R_a^{-1})' \in R^{N\times r_a},\ \Psi^R = \Psi^0(R_b^{-1})' \in R^{N\times r_b} \text{ and } \Gamma^R = \left(\Psi^R_1 - \Lambda^R,\ \Psi^R_2\right) \tag{E.1}$$

in (2.5) and (3.6), respectively. For the ease of notation, we also define

$$\Lambda^* = \left(\Lambda^R,\ \mathbf{0}_{N\times(k-r_a)}\right),\ \Psi^* = \left(\Psi^R,\ \mathbf{0}_{N\times(k-r_b)}\right) \text{ and } \Gamma^* = \Psi^* - \Lambda^*. \tag{E.2}$$

If $N^{-1}||\Psi_\ell^R - \Lambda_\ell^R||^2 \to 0$ as $N \to \infty$ for some ℓ, we replace the definition of Γ_ℓ^R and Γ_ℓ^* above with 0. The augmented matrices Λ^* and Ψ^* are transformed from Λ^+ and $\Psi^+ = \Lambda^+ + \Gamma^+$ defined in (3.1). Generally speaking, for the rest of the proof, the superscript 0 represents the true factor loadings, the superscript R represents transformed factor loadings, and the superscript asterisk represents augmented transformed factor loadings.

Following the definition of $\mathcal{Z}$ in (5.1) and the definition of Γ^*,

$$\mathcal{Z} = \{\ell \in \{1, ..., k\} : \Gamma_\ell^* \neq 0\} \text{ and } \mathcal{Z}^C = \{\ell \in \{1, ..., k\} : \Gamma_\ell^* = 0\}. \tag{E.3}$$

By the definition of Γ^*, $\{r_b + 1, ..., k\} \subseteq \mathcal{Z}^C$ and $\mathcal{Z} \subseteq \{1, ..., r_b\}$. We allow $\ell \in \mathcal{Z}^C$ for some $\ell \leq r_b$ in the proofs below.

Recall $\widehat{\Lambda}$ and $\widehat{\Gamma}$ are the PLS estimators. Write $\widehat{\Psi} = \widehat{\Lambda} + \widehat{\Gamma}$. Define

$$Z_\lambda^2 = N^{-1} \left\| \widehat{\Lambda} - \Lambda^* \right\|^2, \; Z_\psi^2 = N^{-1} \left\| \widehat{\Psi} - \Psi^* \right\|^2, \; Z_\gamma^2 = N^{-1} \left\| \widehat{\Gamma} - \Gamma^* \right\|^2. \tag{E.4}$$

Proof of Theorem 1. The criterion function for the shrinkage estimator can be written as

$$\begin{aligned} Q(\Lambda, \Gamma) &= M_a(\Lambda, \widetilde{F}_a) + M_b(\Psi, \widetilde{F}_b) + P_1(\Lambda) + P_2(\Gamma), \text{ where} \\ M_a(\Lambda, F_a) &= (NT)^{-1} \left\| X_a - F_a \Lambda' \right\|^2, \\ M_b(\Psi, F_b) &= (NT)^{-1} \left\| X_b - F_b (\Lambda + \Gamma)' \right\|^2, \\ P_1(\Lambda) &= \alpha_{NT} \sum_{\ell=1}^k \omega_\ell^\lambda ||\Lambda_\ell|| \text{ and } P_2(\Gamma) = \beta_{NT} \sum_{\ell=1}^k \omega_\ell^\gamma ||\Gamma_\ell||, \end{aligned} \tag{E.5}$$

with $\Psi = \Lambda + \Gamma$. For notational simplicity, the dependence on N and T is suppressed. Because the shrinkage estimators $\widehat{\Lambda}$ and $\widehat{\Gamma}$ minimize the criterion function $Q(\Lambda, \Gamma)$, we have $Q(\widehat{\Lambda}, \widehat{\Gamma}) \leq Q(\Lambda^*, \Gamma^*)$, i.e.,

$$\begin{aligned} &\left[M_a(\widehat{\Lambda}, \widetilde{F}_a) - M_a(\Lambda^*, \widetilde{F}_a) \right] + \left[M_b(\widehat{\Psi}, \widetilde{F}_b) - M_b(\Psi^*, \widetilde{F}_b) \right] \\ \leq &\left[P_1(\Lambda^*) - P_1(\widehat{\Lambda}) \right] + \left[P_2(\Gamma^*) - P_2(\widehat{\Gamma}) \right], \end{aligned} \tag{E.6}$$

where $\widehat{\Psi} = \widehat{\Lambda} + \widehat{\Gamma}$.

We start with the right-hand side of (E.6). Define

$$p_1 = P_1(\Lambda^*) - P_1^r(\widehat{\Lambda}) \text{ and } p_2 = \begin{cases} P_2(\Gamma^*) - P_2^r(\widehat{\Gamma}) & \text{if } \Gamma^0 \neq 0 \\ 0 & \text{if } \Gamma^0 = 0 \end{cases}, \text{ where}$$

$$P_1^r(\widehat{\Lambda}) = \alpha_{NT} \sum_{\ell=1}^{r_a} \omega_\ell^\lambda ||\widehat{\Lambda}_\ell|| \leq \alpha_{NT} \sum_{\ell=1}^{k} \omega_\ell^\lambda ||\widehat{\Lambda}_\ell|| = P_1(\widehat{\Lambda}),$$

$$P_2^r(\widehat{\Gamma}) = \beta_{NT} \sum_{\ell \in \mathcal{Z}} \omega_\ell^\gamma ||\widehat{\Gamma}_\ell|| \leq \beta_{NT} \sum_{\ell=1}^{k} \omega_\ell^\gamma ||\widehat{\Gamma}_\ell|| = P_2(\widehat{\Gamma}). \tag{E.7}$$

If $\Gamma^0 = 0$, we have $\Gamma^* = 0$ and $P_2(\Gamma^*) - P_2^r(\widehat{\Gamma}) \leq 0$ because $P_2(\Gamma^*) = 0$ and $P_2^r(\Gamma) \geq 0$. The penalty terms on the right-hand side of (E.6) satisfy

$$P_1(\Lambda^*) - P_1(\widehat{\Lambda}) \leq p_1 \text{ and } P_2(\Gamma^*) - P_2(\widehat{\Gamma}) \leq p_2 \tag{E.8}$$

following the inequalities in (E.7).

We have $\Lambda_\ell^* = 0$ for $\ell = r_a + 1, \ldots, k$ and $\Gamma_\ell^* = 0$ for $\ell \in \mathcal{Z}^C$, which implies that

$$P_1(\Lambda^*) = \alpha_{NT} \sum_{\ell=1}^{r_a} \omega_\ell^\lambda ||\Lambda_\ell^*|| \text{ and } P_2(\Gamma^*) = \beta_{NT} \sum_{\ell \in \mathcal{Z}} \omega_\ell^\gamma ||\Gamma_\ell^*||. \tag{E.9}$$

Following (E.7), (E.9), the triangle inequality, and the Cauchy-Schwarz inequality, we have

$$p_1 \leq \alpha_{NT} \sum_{\ell=1}^{r_a} \omega_\ell^\lambda \left\| \widehat{\Lambda}_\ell - \Lambda_\ell^* \right\| \leq b_\Lambda Z_\lambda, \text{ where } b_\Lambda = N^{1/2} \alpha_{NT} \left[\sum_{\ell=1}^{r_a} (\omega_\ell^\lambda)^2 \right]^{1/2} \tag{E.10}$$

and Z_λ is defined in (E.4). By the same arguments,

$$p_2 \leq b_\Gamma Z_\gamma, \text{ where } b_\Gamma = \begin{cases} N^{1/2} \beta_{NT} \left[\sum_{\ell \in \mathcal{Z}} (\omega_\ell^\gamma)^2 \right]^{1/2} & \text{if } \Gamma^0 \neq 0 \\ 0 & \text{if } \Gamma^0 = 0 \end{cases} \tag{E.11}$$

and Z_γ is in (E.4). Combining (E.6) and (E.8)-(E.11), we obtain

$$\left[M_a(\widehat{\Lambda}, \widetilde{F}_a) - M_a(\Lambda^*, \widetilde{F}_a) \right] + \left[M_b(\widehat{\Psi}, \widetilde{F}_b) - M_b(\Psi^*, \widetilde{F}_b) \right] \leq b_\Lambda Z_\lambda + b_\Gamma Z_\gamma. \tag{E.12}$$

Next, we consider the left-hand side of (E.12). To this end, we first show some useful equalities. Write $\widetilde{F}_a = (\widetilde{F}_a^r, \widetilde{F}_a^\perp) \in R^{T_0 \times k}$, where $\widetilde{F}_a$ is partitioned into a $T_0 \times r_a$ submatrix $\widetilde{F}_a^r$ and a $T_0 \times (k - r_a)$ submatrix $\widetilde{F}_a^\perp$. Replacing $\widetilde{F}_a^r$ with $F_a^R = F_a R_a$, we define

$$F_a^* = (F_a^R, \widetilde{F}_a^\perp) = (F_a R_a, \widetilde{F}_a^\perp) \in R^{T_0 \times k}. \tag{E.13}$$

Some equivalent relationships are useful in the calculation below

$$F_a^*\Lambda^{*\prime} = F_a^R\Lambda^{R\prime} = F_a\Lambda^{0\prime} \text{ and } \widetilde{F}_a\Lambda^{*\prime} = \widetilde{F}_a^r\Lambda^{R\prime}, \tag{E.14}$$

because $\Lambda^* = (\Lambda^R, \mathbf{0}_{N\times(k-r_a)})$. It follows that

$$\begin{aligned} F_a\Lambda^{0\prime} - \widetilde{F}_a\widehat{\Lambda}' &= F_a^*\Lambda^{*\prime} - \widetilde{F}_a\widehat{\Lambda}' \\ &= (F_a^* - \widetilde{F}_a)\Lambda^{*\prime} - \widetilde{F}_a(\widehat{\Lambda} - \Lambda^*)' \\ &= (F_aR_a - \widetilde{F}_a^r)\Lambda^{R\prime} - \widetilde{F}_a(\widehat{\Lambda} - \Lambda^*)', \end{aligned} \tag{E.15}$$

where the first equality follows from (E.14), the second equality follows from adding and subtracting $\widetilde{F}_a\Lambda^{*\prime}$, and the third equality follows from (E.14). The difference between the true common component $F_a\Lambda^{0\prime}$ and the estimated common component $\widetilde{F}_a\widehat{\Lambda}'$ are decomposed into two pieces by the calculation in (E.15), where the first piece focuses on the factor estimation error and the second piece focuses on the factor loading estimation error.

The first term on the left-hand side of (E.12) satisfies

$$\begin{aligned} M_a(\widehat{\Lambda}, \widetilde{F}_a) &= (NT)^{-1}\left\|X_a - \widetilde{F}_a\widehat{\Lambda}'\right\|^2 \\ &= (NT)^{-1}\left\|e_a + (F_a\Lambda^{0\prime} - \widetilde{F}_a\widehat{\Lambda}')\right\|^2 \\ &= (NT)^{-1}\left\|\left(e_a + (F_aR_a - \widetilde{F}_a^r)\Lambda^{R\prime}\right) - \widetilde{F}_a(\widehat{\Lambda} - \Lambda^*)'\right\|^2 \\ &= M_1 + M_2 + M_3 + M_4, \end{aligned} \tag{E.16}$$

where the first equality follows from the definition of $M_a(\Lambda, F_a)$ in (E.5), the second equality follows from $X_a = e_a + F_a\Lambda^{0\prime}$, the third equality holds by the decomposition in (E.15), and M_1, M_2, M_3 and M_4 are defined as follows. The first term M_1 is

$$\begin{aligned} M_1 &= (NT)^{-1}\left\|e_a + (F_aR_a - \widetilde{F}_a^r)\Lambda^{R\prime}\right\|^2 \\ &= (NT)^{-1}\left\|X_a - \widetilde{F}_a\Lambda^{*\prime}\right\|^2 = M_a(\Lambda^*, \widetilde{F}_a), \end{aligned} \tag{E.17}$$

following $X_a = e_a + F_a^{R\prime}\Lambda^{R\prime}$, $\widetilde{F}_a^r\Lambda^{R\prime} = \widetilde{F}_a\Lambda^{*\prime}$ in (E.14) and the definition of $M_a(\Lambda, F)$ in (E.5). The second term M_2 is

$$\begin{aligned} M_2 &= (NT)^{-1}\left\|\widetilde{F}_a(\widehat{\Lambda} - \Lambda^*)'\right\|^2 \\ &= (NT)^{-1}tr\left((\widehat{\Lambda} - \Lambda^*)\widetilde{F}_a'\widetilde{F}_a(\widehat{\Lambda} - \Lambda^*)'\right) \\ &= \frac{T_0}{T}N^{-1}\left\|\widehat{\Lambda} - \Lambda^*\right\|^2 = \frac{T_0}{T}Z_\lambda^2, \end{aligned} \tag{E.18}$$

following $\widetilde{F}_a'\widetilde{F}_a/T_0 = I_{r_a}$ and the definition of Z_λ. The third term M_3 is

$$M_3 = -2(NT)^{-1} tr\left(e_a'\widetilde{F}_a(\widehat{\Lambda} - \Lambda^*)'\right). \tag{E.19}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}(NT)^{-1}\left|tr\left(e_a'\widetilde{F}_a(\widehat{\Lambda} - \Lambda^*)'\right)\right| &\leq (NT)^{-1}\left|tr\left(e_a'\widetilde{F}_a\widetilde{F}_a'e_a\right)\right|^{1/2}\left\|\widehat{\Lambda} - \Lambda^*\right\| \\ &= N^{-1/2}T^{-1}\left|T_0 tr\left(P_{\widetilde{F}_a}e_a e_a'\right)\right|^{1/2} Z_\lambda \\ &\leq N^{-1/2}T^{-1}\left|NT_0^2 k\rho_1((NT_0)^{-1}e_a e_a')\right|^{1/2} Z_\lambda \\ &= \frac{C_{3,n}Z_\lambda}{2}. \end{aligned} \tag{E.20}$$

The first equality holds because $P_{\widetilde{F}_a} = {T_0}^{-1}\widetilde{F}_a\widetilde{F}_a'$, $tr(AB) = tr(BA)$ for two matrices, and because of the definition of Z_λ. The second inequality follows from von Neumann's trace inequality and the fact that the eigenvalues of $P_{\widetilde{F}_a}$ consist of k ones and $T-k$ zeros. By Assumption C(vi) and simple calculations,

$$\begin{aligned} C_{3,n} &= 2N^{-1/2}T^{-1}\left|NT_0^2 k\rho_1((NT_0)^{-1}e_a e_a')\right|^{1/2} \\ &= 2N^{-1/2}T^{-1}\left|NT_0^2 O_p(C_{NT}^{-2})\right|^{1/2} \\ &= \frac{T_0}{T}O_p(C_{NT}^{-1}) = O_p(C_{NT}^{-1}), \end{aligned} \tag{E.21}$$

which together with (E.19) and (E.20) implies

$$|M_3| \leq C_{3,n}Z_\lambda, \text{ where } C_{3,n} = O_p(C_{NT}^{-1}). \tag{E.22}$$

The fourth term M_4 is

$$M_4 = -2(NT)^{-1} tr\left(\Lambda^R(F_aR_a - \widetilde{F}_a^r)'\widetilde{F}_a(\widehat{\Lambda} - \Lambda^*)'\right). \tag{E.23}$$

To investigate M_4, we note that

$$\frac{(F_aR_a - \widetilde{F}_a^r)'\widetilde{F}_a}{T_0} = \frac{(F_aH_a - \widetilde{F}_a^r)'\widetilde{F}_a}{T_0} + \frac{(F_a(R_a - H_a))'\widetilde{F}_a}{T_0} = O_p(C_{NT}^{-1}) \tag{E.24}$$

by the Cauchy-Schwarz inequality, (D.4), and Lemma 1. Applying the Cauchy-Schwarz inequality, we have

$$\begin{aligned} &(NT)^{-1}\left|tr\left(\Lambda^R(F_aR_a - \widetilde{F}_a^r)'\widetilde{F}_a(\widehat{\Lambda} - \Lambda^*)'\right)\right| \\ &\leq (NT)^{-1}\left\|\Lambda^R\right\|\left\|(F_aR_a - \widetilde{F}_a^r)'\widetilde{F}_a\right\|\left\|\widehat{\Lambda} - \Lambda^*\right\| \\ &= \frac{T_0}{T}\left(N^{-1}\left\|\Lambda^R\right\|^2\right)^{1/2}\left\|\frac{(F_aR_a - \widetilde{F}_a^r)'\widetilde{F}_a}{T_0}\right\|\left(N^{-1}\left\|\widehat{\Lambda} - \Lambda^*\right\|^2\right)^{1/2} = \frac{C_{4,n}Z_\lambda}{2}. \end{aligned} \tag{E.25}$$

Using $\Lambda^{R\prime} = R_a^{-1}\Lambda'$, $R_a^{-1} = O_p(1)$, $||N^{-1}\Lambda'\Lambda - \Sigma_\Lambda|| \to 0$ and (E.24), we deduce that

$$C_{4,n} = \frac{2T_0}{T}\left(N^{-1}\left\|\Lambda^R\right\|^2\right)^{1/2}\left\|\frac{(F_aR_a - \widetilde{F}_a^r)'\widetilde{F}_a}{T_0}\right\| = \frac{T_0}{T}O_p(C_{NT}^{-1}) = O_p(C_{NT}^{-1}), \tag{E.26}$$

which together with (E.23) and (E.25) yields

$$|M_4| \leq C_{4,n}Z_\lambda, \text{ where } C_{4,n} = O_p(C_{NT}^{-1}). \tag{E.27}$$

Putting the four terms in (E.17), (E.18), (E.22), and (E.27) into (E.16), we obtain

$$M_a(\widehat{\Lambda}, \widetilde{F}_a) - M_a(\Lambda^*, \widetilde{F}_a) \geq \frac{T_0}{T}Z_\lambda^2 - C_{a,n}Z_\lambda, \text{ where } C_{a,n} = C_{3,n} + C_{4,n} = O_p(C_{NT}^{-1}). \tag{E.28}$$

Replacing the first subsample with the second subsample and the factor loadings Λ with Ψ, we also have

$$M_b(\widehat{\Psi}, \widetilde{F}_b) - M_b(\Psi^*, \widetilde{F}_b) \geq \frac{T_1}{T}Z_\psi^2 - C_{b,n}Z_\psi, \text{ where } C_{b,n} = O_p(C_{NT}^{-1}). \tag{E.29}$$

Plugging (E.28) and (E.29) into the left-hand side of (E.12), we obtain

$$\frac{T_0}{T}Z_\lambda^2 - C_{a,n}Z_\lambda + \frac{T_1}{T}Z_\psi^2 - C_{b,n}Z_\psi \leq b_\Lambda Z_\lambda + b_\Gamma Z_\gamma \leq (b_\Lambda + b_\Gamma)Z_\lambda + b_\Gamma Z_\psi, \tag{E.30}$$

following the triangle inequality. Rearranging (E.30) gives

$$\begin{aligned}&\pi_0\left(Z_\lambda - \frac{C_{a,n} + b_\Lambda + b_\Gamma}{2\pi_0}\right)^2 + \pi_1\left(Z_\psi - \frac{C_{b,n} + b_\Gamma}{2\pi_1}\right)^2 \\ \leq\ &\pi_0\left(\frac{C_{a,n} + b_\Lambda + b_\Gamma}{2\pi_0}\right)^2 + \pi_1\left(\frac{C_{b,n} + b_\Gamma}{2\pi_1}\right)^2,\end{aligned} \tag{E.31}$$

where $\pi_0 = T_0/T \in (0,1)$ and $\pi_1 = 1 - \pi_0$. It follows from (E.31), $C_{a,n} = O_p(C_{NT}^{-1})$, $C_{b,n} = O_p(C_{NT}^{-1})$, and the triangle inequality that

$$\begin{aligned}Z_\lambda &= O_p(b_\Lambda + b_\Gamma + C_{NT}^{-1}), \\ Z_\psi &= O_p(b_\Lambda + b_\Gamma + C_{NT}^{-1}), \\ Z_\gamma &= O_p(b_\Lambda + b_\Gamma + C_{NT}^{-1}).\end{aligned} \tag{E.32}$$

Assumptions P1 and P2 imply that

$$\omega_\ell^\lambda = O_p(1) \text{ for } \ell = 1, \ldots, r_a,\ \omega_\ell^\gamma = O_p(1) \text{ for } \ell \in \mathcal{Z}. \tag{E.33}$$

Assumption T(i) implies that $b_\Lambda = O_p(C_{NT}^{-1})$ and $b_\Gamma = O_p(C_{NT}^{-1})$, following (E.33). It follows from (E.32) that

$$Z_\lambda = O_p(C_{NT}^{-1}) \text{ and } Z_\gamma = O_p(C_{NT}^{-1}). \tag{E.34}$$

Theorems 1(a) and 1(c) follow from the definitions of Z_λ and Z_γ in (E.4) and the results in (E.34).

Next, we show the superefficiency results in Theorems 1(b), 1(d), and 1(e). To this end, first define

$$\mathcal{L}_a = \{\ell : (\omega_\ell^\lambda)^{-1} = O_p(C_{NT}^{-2d})\} \text{ and } \mathcal{L}_b = \{\ell : (\omega_\ell^\gamma)^{-1} = O_p(C_{NT}^{-2d})\}. \tag{E.35}$$

Under Assumptions P1 and P2,

$$\{r_a + 1, \ldots, k\} \subseteq \mathcal{L}_a,\ \{r_b + 1, \ldots, k\} \subseteq \mathcal{L}_b, \text{ and if } \Gamma^0 = 0,\ \{1, \ldots, k\} = \mathcal{L}_b. \tag{E.36}$$

Define the residual matrices

$$e_a(\widehat{\Lambda}) = X_a - \widetilde{F}_a\widehat{\Lambda}' \in R^{T_0 \times N} \text{ and } e_b(\widehat{\Lambda} + \widehat{\Gamma}) = X_b - \widetilde{F}_b(\widehat{\Lambda} + \widehat{\Gamma})' \in R^{T_1 \times N}. \tag{E.37}$$

Let $e_t^a(\widehat{\Lambda})$ for $t = 1, \ldots, T_0$ be the rows of $e_a(\widehat{\Lambda})$ and $e_t^b(\widehat{\Lambda} + \widehat{\Gamma})$ for $t = T_0 + 1, \ldots, T$ be the rows of $e_b(\widehat{\Lambda} + \widehat{\Gamma})$. Let $\widetilde{F}_\ell = (\widetilde{F}_{a,\ell}', \widetilde{F}_{b,\ell}')' \in R^{T \times 1}$, where $\widetilde{F}_{a,\ell}$ and $\widetilde{F}_{b,\ell}$ are the ℓ-th columns of $\widetilde{F}_a$ and $\widetilde{F}_b$, respectively, and let $\widetilde{F}_{t,\ell}$ denote the t-th row of $\widetilde{F}_\ell$. By Lemma 4.2 of Bühlmann and van de Geer (2011), a sufficient condition for $\widehat{\Lambda}_\ell = 0$ is

$$2\,(NT)^{-1} \left\| \sum_{t=1}^{T_0} e_t^a(\widehat{\Lambda})\widetilde{F}_{t,\ell} + \sum_{t=T_0+1}^{T} e_t^b(\widehat{\Lambda} + \widehat{\Gamma})\widetilde{F}_{t,\ell} \right\| < \alpha_{NT}\omega_\ell^\lambda, \tag{E.38}$$

where the left-hand side is associated with the partial derivative of $M_a(\Lambda, \widetilde{F}_a) + M_b(\Psi, \widetilde{F}_b)$, with respect to Λ_ℓ evaluated at the PLS estimators, and the right-hand side is the marginal penalty once $\widehat{\Lambda}_\ell$ deviates from 0. Intuitively, the optimal solution is $\widehat{\Lambda}_\ell = 0$ when the marginal penalty on the right-hand side of (E.38) is larger than the marginal gain on the left-hand side of (E.38). The inequality in (E.38) can be equivalently written as

$$\left\| e^a(\widehat{\Lambda})'\widetilde{F}_{a,\ell} + e^b(\widehat{\Lambda} + \widehat{\Gamma})'\widetilde{F}_{b,\ell} \right\| < \frac{NT}{2}\alpha_{NT}\omega_\ell^\lambda, \tag{E.39}$$

which holds provided that

$$\left\| e^a(\widehat{\Lambda})'\widetilde{F}_{a,\ell} \right\| + \left\| e^b(\widehat{\Lambda} + \widehat{\Gamma})'\widetilde{F}_{b,\ell} \right\| < \frac{NT}{2}\alpha_{NT}\omega_\ell^\lambda. \tag{E.40}$$

Next, we study the two terms on the left-hand side of (E.40). The first term satisfies

$$\begin{aligned}\left\|e^a(\widehat{\Lambda})'\widetilde{F}_{a,\ell}\right\| &= \left\|(e_a + F_a\Lambda^{0\prime} - \widetilde{F}_a\widehat{\Lambda}')'\widetilde{F}_{a,\ell}\right\| \\ &= \left\|e_a'\widetilde{F}_{a,\ell} + (F_aR_a - \widetilde{F}_a^r)\Lambda^{R\prime}\widetilde{F}_{a,\ell} - \widetilde{F}_a(\widehat{\Lambda} - \Lambda^*)'\widetilde{F}_{a,\ell}\right\| \\ &\leq \left\|e_a'\widetilde{F}_{a,\ell}\right\| + \left\|F_aR_a - \widetilde{F}_a^r\right\| \left\|\Lambda^R\right\| \left\|\widetilde{F}_{a,\ell}\right\| + \left\|\widetilde{F}_a\right\| \left\|\widehat{\Lambda} - \Lambda^*\right\| \left\|\widetilde{F}_{a,\ell}\right\| \quad \text{(E.41)}\end{aligned}$$

where the second equality follows from (E.15) and the inequality follows from the Cauchy-Schwarz inequality and the triangle inequality. The terms in the last line of (E.41) are: (i)

$$\begin{aligned}||e_a'\widetilde{F}_{a,\ell}|| &= (NT)^{1/2}\sqrt{\widetilde{F}_{a,\ell}'\frac{e_ae_a'}{NT}\widetilde{F}_{a,\ell}} \\ &\leq (NT)^{1/2}T_0^{1/2}\sqrt{\rho_1\left((NT)^{-1}e_ae_a'\right)}\sqrt{\frac{\widetilde{F}_{a,\ell}'\widetilde{F}_{a,\ell}}{T_0}} \\ &= (NT)^{1/2}T_0^{1/2}O_p(C_{NT}^{-1}) = O_p(N^{1/2}TC_{NT}^{-1}), \quad \text{(E.42)}\end{aligned}$$

where the second equality is by $T_0^{-1}\widetilde{F}_{a,\ell}'\widetilde{F}_{a,\ell} = 1$ and Assumption C(vi); (ii) $||F_aR_a - \widetilde{F}_a^r|| = O_p(T^{1/2}C_{NT}^{-1})$ by (D.4); (iii) $\left\|\Lambda^R\right\| = O_p(N^{1/2})$ because $R_a = O_p(1)$ and $||\Lambda'\Lambda/N - \Sigma_\Lambda|| \to 0$; (iv) $||\widetilde{F}_{a,\ell}|| = O(T^{1/2})$ and $||\widetilde{F}_a|| = O(T^{1/2})$ because $T_0^{-1}\widetilde{F}_a'\widetilde{F}_a = I_{r_a}$; (v) $||\widehat{\Lambda} - \Lambda^*|| = O_p(N^{1/2}C_{NT}^{-1})$ by the definition of Z_λ and (E.34). Putting them together with (E.41), we have

$$\left\|e^a(\widehat{\Lambda})'\widetilde{F}_{a,\ell}\right\| = O_p(N^{1/2}TC_{NT}^{-1}). \quad \text{(E.43)}$$

By the same arguments, we have

$$\left\|e^b(\widehat{\Lambda} + \widehat{\Gamma})'\widetilde{F}_{b,\ell}\right\| = O_p(N^{1/2}TC_{NT}^{-1}). \quad \text{(E.44)}$$

Equations (E.43) and (E.44) imply that for the inequality in (E.40) to hold, it suffices to have

$$N^{-1/2}C_{NT}^{-1} = o_p(\alpha_{NT}\omega_\ell^\lambda), \quad \text{(E.45)}$$

which is satisfied for all $\ell \in \mathcal{L}_a$ under Assumption T(ii).

To prove Theorems 1(d) and 1(e), note that a sufficient condition for $\widehat{\Gamma}_\ell = 0$ is

$$2\left(NT\right)^{-1}\left\|\sum_{t=T_0+1}^{T} e_t^b(\widehat{\Lambda} + \widehat{\Gamma})\widetilde{F}_{t,\ell}\right\| < \beta_{NT}\omega_\ell^\gamma. \quad \text{(E.46)}$$

Following (E.44), the inequality in (E.46) holds provided that

$$N^{-1/2}C_{NT}^{-1} = o_p(\beta_{NT}\omega_\ell^\gamma), \tag{E.47}$$

which is satisfied for all $\ell \in \mathcal{L}_b$ under Assumption T(ii). Therefore, Theorems 1(b), 1(d), and 1(e) follow from (E.36).

Some remarks on the proof of Theorem 1 and its relationship to the proofs of Corollaries 1 and 2 below are in order. First, in the proof of Theorem 1, we give general definition of $\mathcal{Z}$, $\mathcal{L}_a$ and $\mathcal{L}_b$ without imposing Assumptions P1 and P2 so that the proof can be recycled when these assumptions are relaxed. Specifically, Theorem 1 can be proved as above without Assumptions P1 and P2 as long as (E.33) and (E.36) can be verified for a given preliminary estimator, as we shall do in the proofs below. Second, Assumptions P1 and P2 are slightly stronger than needed to prove Theorem 1, however, we present them as is for the simplicity of the presentation to convey the idea. These assumptions can be relaxed as follows: Assumption P1(ii) assumes that $\Pr(N^{-1}||\widetilde{\Gamma}_\ell||^2 \geq C) \to 1$ for $\ell = 1, \ldots, r_b$, while we only need this to hold for $\ell \in \mathcal{Z}$ rather than for all $\ell = 1, \ldots, r_b$ in order to verify (E.33). The set $\mathcal{Z}$, associated with the nonzero columns of Γ^R, could be a subset of $\{1, \ldots, r_b\}$ to identify a type-1 or type-2 change. For this reason, the proofs of Corollaries 1 and 2 do not verify Assumptions P1 and P2 but rather show Theorem 1 directly. □

Proof of Theorem 2. First, Theorem 1(a) for $\ell = r_a$ and Lemma 2(a) imply that $\Pr(||\widehat{\Lambda}_\ell|| > 0) \to 1$ for $\ell = r_a$ and thus $\Pr(\widehat{r}_a \geq r_a) \to 1$. Theorem 1(b) implies that $\Pr(\widehat{r}_a \leq r_a) \to 1$. Thus, $\Pr(\widehat{r}_a = r_a) \to 1$.

Second, for a type-2 change where $r_b > r_a$, Theorem 1(c) for $\ell = r_b$ and Lemma 2(b) imply that $\Pr(||\widehat{\Gamma}_\ell|| > 0) \to 1$ for $\ell = r_b$ and thus $\Pr(\widehat{r}_b \geq r_b) \to 1$. Theorem 1(e) implies that $\Pr(\widehat{r}_b \leq r_b) \to 1$. Hence, $\Pr(\widehat{r}_b = r_b) \to 1$ for a type-2 change, which, together with part (a), also implies $\Pr(\widehat{\mathcal{B}} = 1) \to 1$ for a type-2 change because by definition, $\widehat{\mathcal{B}} = 1$ if $\widehat{r}_b > \widehat{r}_a$.

Third, for a type-1 change where $r_b = r_a$ and $\mathcal{B}_0 = 1$, Theorem 1(c), Lemmas 2(c) and 2(d), and Assumption ID imply that $\Pr(||\widehat{\Gamma}_\ell|| > 0) \to 1$ for some $\ell \leq r_a$ and thus $\Pr(\widehat{\mathcal{B}} = 1) \to 1$. Note that by definition in (3.8), we have $\widehat{r}_b \geq \widehat{r}_a$. Thus, part (a) and $r_a = r_b$ imply that $\Pr(\widehat{r}_b \geq r_b) \to 1$. On the other hand, Theorem 1(e) implies that $\Pr(\widehat{r}_b \leq r_b) \to 1$. Hence, $\Pr(\widehat{r}_b = r_b) \to 1$ for a type-1 change.

Finally, for the case where there is no change, i.e., $r_a = r_b$ and $\mathcal{B}_0 = 0$, Theorems 1(d) and 1(e) imply that $\Pr(\widehat{\Gamma} = 0) \to 1$. Thus, $\Pr(\widehat{\mathcal{B}} = 0) \to 1$ by (3.7) and $\Pr(\widehat{r}_b = r_b) \to 1$ by (3.8) and part (a). □

Proof of Corollary 1. We first study the properties of the unrestricted least square estimator $\widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma}_{LS}$. Note that the unrestricted least squares estimator is a special case of the PLS estimator when $\alpha_{NT} = \beta_{NT} = 0$. Therefore, following (E.32),

$$N^{-1}||\widetilde{\Lambda}_{LS} - \Lambda^*||^2 = O_p(C_{NT}^{-2}) \text{ and } N^{-1}||\widetilde{\Gamma}_{LS} - \Gamma^*||^2 = O_p(C_{NT}^{-2}), \tag{E.48}$$

which combined with the definitions of Λ^* and Γ^* and Lemma 2 imply that

$$\Pr(N^{-1}||\widetilde{\Lambda}_{LS,\ell}||^2 \geq C) \to 1 \text{ for } \ell = 1, \ldots, r_a, \ \Pr(N^{-1}||\widetilde{\Gamma}_{LS,\ell}||^2 \geq C) \to 1 \text{ for } \ell \in \mathcal{Z} \tag{E.49}$$

and

$$N^{-1}||\widetilde{\Lambda}_{LS,\ell}||^2 = O_p(C_{NT}^{-2}) \text{ for } \ell > r_a \text{ and } N^{-1}||\widetilde{\Gamma}_{LS,\ell}||^2 = O_p(C_{NT}^{-2}) \text{ for } \ell \in \mathcal{Z}^C. \tag{E.50}$$

Next, we show that (E.33) and (E.36) hold without imposing Assumptions P1 and P2, so that the proof of Theorem 1 follows without these two assumptions. The definition of weights in (3.4) and (E.49) imply that (E.33) holds for the case $\widetilde{\Lambda} = \widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma} = \widetilde{\Gamma}_{LS}$. The definition of $\mathcal{L}_a$ and $\mathcal{L}_b$ together with (E.50) imply that $\mathcal{L}_a = \{r_a + 1, \ldots, k\}$ and $\mathcal{L}_b = \mathcal{Z}^C$. By definition, $\{r_b + 1, \ldots, k\} \subseteq \mathcal{Z}^C$ and, if $\Gamma^0 = 0$, then $\{1, \ldots, k\} = \mathcal{Z}^C$, which implies that (E.36) holds for the case $\widetilde{\Lambda} = \widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma} = \widetilde{\Gamma}_{LS}$. Therefore, Theorem 1 holds without imposing Assumptions P1 and P2 for the one-step estimator $\widetilde{\Lambda} = \widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma} = \widetilde{\Gamma}_{LS}$. Applying Theorem 1, model selection consistency follows from the proof for Theorem 2. □

Proof of Corollary 2. We first study the preliminary estimators $\widetilde{\Lambda}^{(2)}$, $\widetilde{\Psi}^{(2)}$, and $\widetilde{\Gamma}^{(2)}$, and the weights ω_ℓ^λ and ω_ℓ^γ in the second step. Because $\widetilde{\Lambda}^{(2)} = \widehat{\Lambda}^{(1)}_{PMS}$, whose first $\widehat{r}_a^{(1)}$ columns are the same as those of $\widetilde{\Lambda}_{LS}$ and whose last $k - \widehat{r}_a^{(1)}$ columns are zeros, it follows from (3.4) that

$$\omega_\ell^\lambda = (N^{-1}||\widetilde{\Lambda}_{LS,\ell}||^2)^{-d} \text{ for } \ell = 1, \ldots, k, \tag{E.51}$$

which is the same for the first- and second-step estimators. If there is a type-2 change, $\widehat{r}_b^{(1)} > \widehat{r}_a^{(1)}$ w.p.a.1 by Corollary 1, and

$$\omega_\ell^\gamma = (N^{-1}||\widetilde{\Gamma}_{LS,\ell}||^2)^{-d} \text{ for } \ell = 1, \ldots, k, \tag{E.52}$$

which is the same for the first and second step estimations.

If there are no structural instabilities or there is a type-1 change, $\widehat{r}_b^{(1)} = \widehat{r}_a^{(1)} = r_b = r_a$ w.p.a.1 by Corollary 1. Let $\widetilde{\Psi}^-_{LS}$ and $\widetilde{\Lambda}^-_{LS}$ denote the first r_a columns of $\widetilde{\Psi}_{LS}$ and $\widetilde{\Lambda}_{LS}$,

respectively. Given $\widehat{r}_b^{(1)} = \widehat{r}_a^{(1)} = r_a = r_b$, we have $\overline{\Psi}^{(1)} = \widetilde{\Psi}_{LS}^-$, $\overline{\Lambda}^{(1)} = \widetilde{\Lambda}_{LS}^-$, and the second-step preliminary estimator $\widetilde{\Gamma}^{(2)}$ can be written as

$$\widetilde{\Gamma}^{(2)} = \left(\widetilde{\Psi}_{LS}^- Q - \widetilde{\Lambda}_{LS}^-,\, 0_{N\times(k-r_a)}\right), \tag{E.53}$$

following from $\widetilde{\Gamma}^{(2)} = \widetilde{\Psi}^{(2)} - \widetilde{\Lambda}^{(2)}$ and steps 1d, 1e, and 2a in the algorithm to construct the two-step estimator.

Define

$$\Gamma^Q = \left(\Psi^R Q - \Lambda^R,\, 0_{N\times(k-r_a)}\right). \tag{E.54}$$

Recall that Ψ^R and Λ^R are the transformed factor loadings. In addition, Γ^R and Λ^R are the first r_a columns of Γ^* and Λ^*, respectively, given $r_a = r_b$. By (E.53) and (E.54), w.p.a.1,

$$\begin{aligned} N^{-1}||\widetilde{\Gamma}^{(2)} - \Gamma^Q||^2 &= N^{-1}\left\|(\widetilde{\Psi}_{LS}^- - \Psi^R)Q - (\widetilde{\Lambda}_{LS}^- - \Lambda^R)\right\|^2 \\ &= N^{-1}\left\|\left(\widetilde{\Gamma}_{LS}^- - \Gamma^R\right)Q + \left(\widetilde{\Lambda}_{LS}^- - \Lambda^R\right)(Q - I_{r_a})\right\|^2 \\ &= O_p(C_{NT}^{-2}), \end{aligned} \tag{E.55}$$

where the last equality follows from the triangle inequality and (E.48). To analyze $\widetilde{\Gamma}^{(2)}$ for the second-step estimation, we first discuss the centering term Γ^Q when there is a type-1 change. Assumption R implies that

$$N^{-1}||\Gamma_\ell^Q||^2 \geq C \text{ if } \ell \in \mathcal{Z} \tag{E.56}$$

because $\Gamma_\ell^Q = \Psi^R Q_\ell - \Lambda_\ell^R$ and $||Q_\ell|| = 1$. Therefore, (E.55) and (E.56) imply that

$$\omega_\ell^\gamma = O_p(1) \text{ for } \ell \in \mathcal{Z} \text{ when there is a type-1 change.} \tag{E.57}$$

If there is no structural change, by (E.48), $N^{-1}||\widetilde{\Lambda}_{LS}^- - \Lambda^R||^2 = O_p(C_{NT}^{-2})$ and $N^{-1}||\widetilde{\Psi}_{LS}^- - \Psi^R||^2 = O_p(C_{NT}^{-2})$. Because $\Lambda^R = \Psi^R$ in this case, we have $N^{-1}||\widetilde{\Lambda}_{LS}^- - \widetilde{\Psi}_{LS}^-||^2 = O_p(C_{NT}^{-2})$, which further implies that

$$N^{-1}\left\|\widetilde{\Psi}_{LS}^- Q - \widetilde{\Lambda}_{LS}^-\right\|^2 \leq N^{-1}\left\|\widetilde{\Psi}_{LS}^- - \widetilde{\Lambda}_{LS}^-\right\|^2 = O_p(C_{NT}^{-2}), \tag{E.58}$$

where the inequality holds because the choice of Q solves the orthogonal procrustes problem by minimizing $||\widetilde{\Psi}_{LS}^- Q - \widetilde{\Lambda}_{LS}^-||^2$ among all orthogonal matrices (Schönemann (1966)). Combining (E.53) and (E.58), we obtain

$$N^{-1}||\widetilde{\Gamma}^{(2)}||^2 = O_p(C_{NT}^{-2}) \text{ when } \Gamma^0 = 0, \tag{E.59}$$

which together with (E.48) and $\Gamma^* = 0$ implies that

$$(\omega_\ell^\gamma)^{-1} = O_p(C_{NT}^{-2d}) \text{ for } \ell = 1, \ldots, k \text{ when there is no structural change.} \tag{E.60}$$

Next, we show that (E.33) and (E.36) hold without imposing Assumptions P1 and P2, so that the proof of Theorem 1 follows without these two assumptions. To show (E.33), note that $\omega_\ell^\lambda = O_p(1)$ for $\ell = 1, \ldots, r_a$ is implied by (E.49) and (E.51), $\omega_\ell^\gamma = O_p(1)$ for $\ell \in \mathcal{Z}$ is implied by (E.49) and (E.52) for a type-2 change, and $\omega_\ell^\gamma = O_p(1)$ for $\ell \in \mathcal{Z}$ is proved in (E.57) for a type-1 change.

To show (E.36), note that: (i) $\{r_a + 1, \ldots, k\} \subseteq \mathcal{L}_a$ holds by (E.50) and (E.51); (ii) $\{r_b + 1, \ldots, k\} \subseteq \mathcal{L}_b$ holds by (E.50) and (E.52); and (iii) if $\Gamma^0 = 0$, $\{1, \ldots, k\} = \mathcal{L}_b$ follows from (E.48) and (E.60).

Because (E.33) and (E.36) hold without imposing Assumptions P1 and P2, Theorem 1 holds without imposing Assumptions P1 and P2 for the two-step estimator. Applying Theorem 1, model selection consistency follows from the proof for Theorem 2. □

Proof of Theorem 3. In the proof below, we use $o_{p\pi}(\cdot)$ and $O_{p\pi}(\cdot)$ to represent $o_p(\cdot)$ and $O_p(\cdot)$ that hold uniformly over $\pi \in \Pi$.

Formally, Theorem 3 is proved by first showing the convergence of $\widehat{\Lambda}_{r_a}(\pi)$ and $\widehat{\Gamma}_{r_a}(\pi)$ uniformly over $\pi \in \Pi$. Provided that $\widehat{\Lambda}_{r_a}(\pi)$ uniformly converges to a nonzero limit for all $\pi \in \Pi$, it follows that $\Pr(\min_{\pi\in\Pi} \widehat{r}_a(\pi) \geq r_a) \to 1$. Because $\pi_0 \in \Pi$ and $r_a(\pi_0) = r_a$ by definition, one can show that $\Pr(\widehat{r}_a(\pi_0) = r_a) \to 1$ as long as results like those in Theorem 1 hold for $\widehat{\Lambda}(\pi_0)$. Combining the two results above, we immediately get $\Pr(\min_{\pi\in\Pi} \widehat{r}_a(\pi) = r_a) \to 1$. Similar arguments can be applied to $\widehat{\Gamma}_{r_a}(\pi)$ to show $\Pr(\min_{\pi\in\Pi} \widehat{r}_b(\pi) = r_b) \to 1$. After showing consistency of the estimators of the number of factors, we analyze $\widehat{\Gamma}(\pi_0)$ for consistent detection of type-1 instability, and show that $\Pr(\widehat{\Gamma}(\pi) = 0) \to 1$ uniformly over $\pi \in \Pi$ when there are no structural instabilities.

Define $r^+ = r_a + r_b$, $T_a = \lfloor T\pi \rfloor$, and $T_b = T - T_a$. First, consider the second subsample $X_b(\pi)$. When $\pi < \pi_0$, following the model in (D.13), the variance of the factor loadings is

$$\Sigma_{ab}^+ = N^{-1} \left(\Lambda^0,\, \Psi^0\right)' \left(\Lambda^0,\, \Psi^0\right). \tag{E.61}$$

With a transformation analogous to that in (D.3) to standardize the factors and diagonalize the loadings, the DGP in (D.13) can be written as

$$X_b(\pi) = F_b^R(\pi)\Psi^R(\pi)' + e_b(\pi), \tag{E.62}$$

where $F_b^R(\pi)$ is $T_b \times r^+$, $\Psi^R(\pi)$ is $N \times r^+$, and

$$\begin{aligned} T_b^{-1} F_b^R(\pi)' F_b^R(\pi) &= I_{r^+} + O_{p\pi}(T^{-1/2}), \\ N^{-1} \Psi^R(\pi)' \Psi^R(\pi) &= \Lambda_b(\pi), \end{aligned} \tag{E.63}$$

where $\Lambda_b(\pi)$ is a $r^+ \times r^+$ diagonal matrix whose diagonal elements are the eigenvalues of $\Sigma_F^+(\pi)\Sigma_{ab}^+$ in a decreasing order. Here we allow some eigenvalues to be zero. This is analogous to the transformation considered in (D.6)-(D.8) in the proof of Lemma 1 except $\pi < \pi_0$ rather than $\pi = \pi_0$. When $\pi \geq \pi_0$, the DGP in (D.13) can be written as in (E.62) and (E.63) but with $r^+ = r_b$ and $\Psi^R(\pi) = \Psi^R$, where $\Psi^R = \Psi^0(R_b^{-1})'$.

Next, we consider the first subsample $X_a(\pi)$. Following the transformation discussed above, when $\pi > \pi_0$, the DGP in (D.12) can be written as

$$X_a(\pi) = F_a^R(\pi)\Lambda^R(\pi)' + e_a(\pi), \tag{E.64}$$

where $F_a^R(\pi)$ is $T_a \times r^+$, $\Lambda^R(\pi)$ is $N \times r^+$, and

$$\begin{aligned} T_a^{-1} F_a^R(\pi)' F_a^R(\pi) &= I_{r^+} + O_{p\pi}(T^{-1/2}), \\ N^{-1} \Lambda^R(\pi)' \Lambda^R(\pi) &= \Lambda_a(\pi), \end{aligned} \tag{E.65}$$

where $\Lambda_a(\pi)$ is a $r^+ \times r^+$ diagonal matrix. When $\pi \leq \pi_0$, the DGP in (D.12) can be written as that in (E.64) and (E.65) but with $r^+ = r_a$ and $\Lambda^R(\pi) = \Lambda^R = \Lambda^0(R_a^{-1})'$.

For any $\pi \in \Pi$, $X_a(\pi)$ contains at least the r_a factors in $X_a(\pi_0)$ and $X_b(\pi)$ contains at least the r_b factors in $X_b(\pi_0)$. Therefore,

$$N^{-1}||\Lambda_\ell^R(\pi)||^2 \geq C \text{ for } \ell = 1, \ldots, r_a, \; N^{-1}||\Psi_\ell^R(\pi)||^2 \geq C \text{ for } \ell = 1, \ldots, r_b. \tag{E.66}$$

Note that in the proof of Theorem 1 above, the magnitudes of the approximation errors are developed under Assumptions A-D. After Assumptions A and C are replaced by Assumptions A* and C*, Assumptions A*, B, C*, and D are all uniform over $\pi \in \Pi$. As a result, replacing π_0 with π, asymptotic results as those in Theorem 1 hold uniformly over $\pi \in \Pi$. We use such uniform convergence in the analysis below.

Below we analyze model selection based on the two-step procedure. Recall that $\widehat{r}_a^{(i)}(\pi)$ for $i = 1$ and 2 denotes the estimator of $r_a(\pi)$ by the first- and second-step PLS estimator.

Let $\omega_\ell^{\lambda*(i)}(\pi)$ and $\omega_\ell^{\gamma*(i)}(\pi)$ denote the weights in step i. Let $\widetilde{\Psi}_{LS}^{-}(\pi)$ denote the first $\widehat{r}_a^{(1)}$ columns of $\widetilde{\Psi}_{LS}(\pi)$. By construction, the adaptive weights in (3.14) satisfy

$$\begin{aligned}
\omega_\ell^{\lambda*(i)}(\pi) &= \left(N^{-1}||\widetilde{\Lambda}_{\ell,LS}(\pi)||^2\right)^{-d} \text{ for } i = 1 \text{ and } 2, \\
\omega_\ell^{\gamma*(1)}(\pi) &= \max\left\{\left(N^{-1}||\widetilde{\Gamma}_{\ell,LS}(\pi)||^2\right)^{-d}, \left(N^{-1}||\widetilde{\Psi}_{\ell,LS}(\pi)||^2\right)^{-d}\right\}, \\
\omega_\ell^{\gamma*(2)}(\pi) &= \omega_\ell^{\gamma*(1)}(\pi) \text{ if (i) } \widehat{r}_a^{(1)} < \widehat{r}_b^{(1)} \text{ or (ii) } \widehat{r}_a^{(1)} = \widehat{r}_b^{(1)} \text{ and } \ell > \widehat{r}_a^{(1)}, \\
\omega_\ell^{\gamma*(2)}(\pi) &= \max\left\{\left(N^{-1}||\widetilde{\Psi}_{\ell,LS}^{-}(\pi)\mathrm{w}(\pi) - \widetilde{\Lambda}_{\ell,LS}(\pi)||^2\right)^{-d}, \left(N^{-1}||\widetilde{\Psi}_{\ell,LS}(\pi)||^2\right)^{-d}\right\} \text{ otherwise,}
\end{aligned} \tag{E.67}$$

where the vector $\mathrm{w}(\pi)$ satisfies $\|\mathrm{w}(\pi)\| = 1$ and is obtained by the orthogonal transformation to minimize the difference between the first $\widehat{r}_a^{(1)}$ columns of $\widetilde{\Lambda}_{LS}(\pi)$ and $\widetilde{\Psi}_{LS}(\pi)$.

In the proof below, if notations and results are not specified to be the first step or the second step, they apply to both. We typically do not distinguish between them until discussing the penalties.

Step 1. We show

$$\Pr(\min_{\pi\in\Pi} \widehat{r}_a^{(i)}(\pi) \geq r_a) \to 1 \text{ for } i = 1 \text{ and } 2. \tag{E.68}$$

To this end, it is sufficient to show $N^{-1}||\widehat{\Lambda}_\ell(\pi) - \Lambda_\ell^R(\pi)||^2 = o_{p\pi}(1)$ for $\ell = r_a$ in both steps. The proof strategy is different from that in Theorem 1 because here we do not require the convergence of $\widehat{\Lambda}_\ell(\pi)$ to $\Lambda_\ell^R(\pi)$ for $\ell > r_a$. Let $X_{a:b}$ denote a submatrix of X that contains the columns from a to b. For any $\pi \in \Pi$, define

$$\Lambda^\dagger(\pi) = \left(\Lambda_{1:r_a}^R(\pi), \widehat{\Lambda}(\pi)_{r_a+1:k}\right), \Gamma^\dagger(\pi) = \widehat{\Gamma}(\pi), \text{ and } \Psi^\dagger(\pi) = \Lambda^\dagger(\pi) + \Gamma^\dagger(\pi). \tag{E.69}$$

For notational simplicity, define $\Lambda^r(\pi) = \Lambda_{1:r_a}^R(\pi)$. Note that the definition of $\Lambda^\dagger(\pi)$ is different from that of Λ^* used in the proof of Theorem 1 even when $\pi = \pi_0$, because the former involves the PLS estimator but the latter does not. Define

$$Z_\lambda^2(\pi) = N^{-1}\left\|\widehat{\Lambda}(\pi) - \Lambda^\dagger(\pi)\right\|^2, Z_\psi^2(\pi) = N^{-1}\left\|\widehat{\Psi}(\pi) - \Psi^\dagger(\pi)\right\|^2, Z_\gamma^2(\pi) = N^{-1}\left\|\widehat{\Gamma}(\pi) - \Gamma^\dagger(\pi)\right\|^2. \tag{E.70}$$

The criterion function for the shrinkage estimator can be written as

$$Q(\Lambda, \Gamma; \pi) = M_a(\Lambda, \widetilde{F}_a(\pi)) + M_b(\Psi, \widetilde{F}_b(\pi)) + P_1^*(\Lambda) + P_2^*(\Gamma), \tag{E.71}$$

where $\Psi = \Lambda + \Gamma$,

$$\begin{aligned} M_a(\Lambda, F_a) &= (NT)^{-1} \left\| X_a(\pi) - F_a \Lambda' \right\|^2, \text{ and} \\ M_b(\Psi, F_b) &= (NT)^{-1} \left\| X_b(\pi) - F_b (\Lambda + \Gamma)' \right\|^2. \end{aligned} \tag{E.72}$$

For notational simplicity, we do not write $M_a(\Lambda, F_a)$ and $M_b(\Psi, F_b)$ indexed by π, although they are by definition. Define

$$\phi_\ell^\lambda = \mathbb{E}_\xi[\alpha_{NT}(\xi)\omega_\ell^{\lambda *}(\xi)] \text{ and } \phi_\ell^\gamma = \mathbb{E}_\xi[\beta_{NT}(\xi)\omega_\ell^{\gamma *}(\xi)], \tag{E.73}$$

where ξ has a uniform distribution on Π and $\mathbb{E}_\xi[\cdot]$ is taken w.r.t. ξ. As such, $P_1^*(\Lambda) = \sum_{\ell=1}^k \phi_\ell^\lambda \|\Lambda_\ell\|$ and $P_2^*(\Gamma) = \sum_{\ell=1}^k \phi_\ell^\gamma \|\Gamma_\ell\|$.

Because the shrinkage estimators $\widehat{\Lambda}(\pi)$ and $\widehat{\Gamma}(\pi)$ minimize the criterion function $Q(\Lambda, \Gamma; \pi)$, we have $Q(\widehat{\Lambda}(\pi), \widehat{\Gamma}(\pi)) \leq Q(\Lambda^\dagger(\pi), \Gamma^\dagger(\pi))$, i.e.,

$$\begin{aligned} &\left[M_a(\widehat{\Lambda}(\pi), \widetilde{F}_a(\pi)) - M_a(\Lambda^\dagger(\pi), \widetilde{F}_a(\pi)) \right] + \left[M_b(\widehat{\Psi}(\pi), \widetilde{F}_b(\pi)) - M_b(\Psi^\dagger(\pi), \widetilde{F}_b(\pi)) \right] \\ \leq &\left[P_1^*(\Lambda^\dagger(\pi)) - P_1^*(\widehat{\Lambda}(\pi)) \right] + \left[P_2^*(\Gamma^\dagger(\pi)) - P_2^*(\widehat{\Gamma}(\pi)) \right], \end{aligned} \tag{E.74}$$

where $\widehat{\Psi}(\pi) = \widehat{\Lambda}(\pi) + \widehat{\Gamma}(\pi)$. We start with the right-hand side of (E.74). Because the last $(k - r_a)$ columns of $\Lambda^\dagger(\pi)$ and $\widehat{\Lambda}(\pi)$ are the same, by the triangle inequality and the Cauchy-Schwarz inequality, we have

$$P_1^*(\Lambda^\dagger(\pi)) - P_1^*(\widehat{\Lambda}(\pi)) = \sum_{\ell=1}^{r_a} \phi_\ell^\lambda \left(|\Lambda_\ell^\dagger(\pi)| - |\widehat{\Lambda}_\ell(\pi)| \right) \leq \overline{b}_\Lambda Z_\lambda(\pi), \text{ where } \overline{b}_\Lambda = N^{1/2} \left(\sum_{\ell=1}^{r_a} (\phi_\ell^\lambda)^2 \right)^{1/2}. \tag{E.75}$$

Because $\Gamma^\dagger(\pi) = \widehat{\Gamma}(\pi)$, the second term on the right-hand side of (E.74) is 0.

Next, we consider the left-hand side of (E.74). Write $\widetilde{F}_a(\pi) = (\widetilde{F}_a^r(\pi), \widetilde{F}_a^\perp(\pi)) \in R^{T_a \times k}$, where $\widetilde{F}_a(\pi)$ is partitioned into the $T_a \times r_a$ and $T_a \times (k - r_a)$ submatrices $\widetilde{F}_a^r(\pi)$ and $\widetilde{F}_a^\perp(\pi)$. Similarly, write $\widehat{\Lambda}(\pi) = (\widehat{\Lambda}^r(\pi), \widehat{\Lambda}^\perp(\pi))$, where $\widehat{\Lambda}(\pi)$ is partitioned into the $N \times r_a$ and $N \times (k - r_a)$ submatrices $\widehat{\Lambda}^r(\pi)$ and $\widehat{\Lambda}^\perp(\pi)$. With this partition, we can write $\Lambda^\dagger(\pi) = (\Lambda^r(\pi), \widehat{\Lambda}^\perp(\pi))$. Define $e_a(\Lambda(\pi), F(\pi)) = X_a(\pi) - F(\pi)\Lambda(\pi)'$. For the calculation below, we first show two expansions. The first is

$$\begin{aligned} e_a(\widehat{\Lambda}(\pi), \widetilde{F}_a(\pi)) &= X_a(\pi) - \widetilde{F}_a(\pi)\widehat{\Lambda}(\pi)' \\ &= X_a(\pi) - \widetilde{F}_a^r(\pi)\widehat{\Lambda}^r(\pi)' - \widetilde{F}_a^\perp(\pi)\widehat{\Lambda}^\perp(\pi)' \\ &= \left(X_a(\pi) - \widetilde{F}_a^r(\pi)\Lambda^r(\pi)' - \widetilde{F}_a^\perp(\pi)\widehat{\Lambda}^\perp(\pi)' \right) - \widetilde{F}_a^r(\pi) \left(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi) \right)' \\ &= e_a(\Lambda^\dagger(\pi), \widetilde{F}_a(\pi)) - \widetilde{F}_a^r(\pi) \left(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi) \right)', \end{aligned} \tag{E.76}$$

where the first and last equalities hold by definition, the second equality follows from the partition of $\widetilde{F}_a(\pi)$ and $\widehat{\Lambda}(\pi)$, and the third equality follows from subtracting and adding $\widetilde{F}_a^r(\pi)\Lambda^r(\pi)'$. Because $r^+ \geq r_a$, we write $F_a^R(\pi) = (F_a^r(\pi),\, F_a^{r+}(\pi))$, where $F_a^R(\pi)$ is partitioned into the $T_a \times r_a$ and $T_a \times (r^+ - r_a)$ submatrices $F_a^r(\pi)$ and $F_a^{r+}(\pi)$. Similarly, write $\Lambda^R(\pi) = (\Lambda^r(\pi),\, \Lambda^{r+}(\pi))$, where $\Lambda^R(\pi)$ is partitioned into the $N \times r_a$ and $N \times (r^+ - r_a)$ submatrices $\Lambda^r(\pi)$ and $\Lambda^{r+}(\pi)$. Following the partition, we can write

$$X_a(\pi) = e_a(\pi) + F_a^r(\pi)\Lambda^r(\pi)' + F_a^{r+}(\pi)\Lambda^{r+}(\pi)'. \tag{E.77}$$

The second expansion is

$$\begin{aligned} e_a(\Lambda^\dagger(\pi), \widetilde{F}_a(\pi)) &= X_a(\pi) - \widetilde{F}_a^r(\pi)\Lambda^r(\pi)' - \widetilde{F}_a^\perp(\pi)\widehat{\Lambda}^\perp(\pi)' \\ &= e_a(\pi) + \left(F_a^r(\pi) - \widetilde{F}_a^r(\pi)\right)\Lambda^r(\pi)' + F_a^{r+}(\pi)\Lambda^{r+}(\pi)' - \widetilde{F}_a^\perp(\pi)\widehat{\Lambda}^\perp(\pi)', \end{aligned} \tag{E.78}$$

where first equality holds by definition and the second equality follows from (E.77). With the first expansion in (E.76), we have

$$\begin{aligned} M_a(\widehat{\Lambda}(\pi), \widetilde{F}_a(\pi)) &= (NT)^{-1}\left\| e_a(\widehat{\Lambda}(\pi), \widetilde{F}_a(\pi))\right\|^2 \\ &= (NT)^{-1}\left\| e_a(\Lambda^\dagger(\pi), \widetilde{F}_a(\pi))\right\|^2 + (NT)^{-1}\left\| \widetilde{F}_a^r\left(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi)\right)'\right\|^2 \\ &\quad -2(NT)^{-1} tr\left[e_a(\Lambda^\dagger(\pi), \widetilde{F}_a(\pi))'\widetilde{F}_a^r(\pi)\left(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi)\right)\right] \\ &= M_a(\Lambda^\dagger(\pi), \widetilde{F}_a(\pi)) + K_0 + K_1 + K_2 + K_3 + K_4, \end{aligned} \tag{E.79}$$

where

$$\begin{aligned} K_0 &= (NT)^{-1}\left\| \widetilde{F}_a^r\left(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi)\right)'\right\|^2 \\ &= \frac{T_a}{T}\frac{1}{N} tr\left[\left(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi)\right)\frac{\widetilde{F}_a^{r\prime}\widetilde{F}_a^r}{T_a}\left(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi)\right)'\right] \\ &= \frac{T_a}{T} Z_\lambda^2(\pi) \end{aligned} \tag{E.80}$$

by definition and the fact that $T_a^{-1}(\widetilde{F}_a^{r\prime}\widetilde{F}_a^r) = I_{r_a \times r_a}$. The terms K_1 to K_4 follow from the second expansion in (E.78), and they are specified below. The first term is

$$K_1 = -2(NT)^{-1} tr\left[e_a(\pi)'\widetilde{F}_a^r(\pi)\left(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi)\right)\right] = \frac{T_a}{T} O_{p\pi}(C_{NT}^{-1}) Z_\lambda(\pi), \tag{E.81}$$

following calculations analogous to those in (E.20) and (E.21). The second term is

$$\begin{aligned} K_2 &= -2(NT)^{-1} tr\left(\Lambda^r(\pi)(F_a^r(\pi) - \widetilde{F}_a^r(\pi))'\widetilde{F}_a^r(\pi)(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi))'\right) \\ &= \frac{T_a}{T} O_{p\pi}(C_{NT}^{-1}) Z_\lambda(\pi) \end{aligned} \tag{E.82}$$

following calculations analogous to those in (E.25) and (E.26). The third term is

$$\begin{aligned} K_3 &= -2(NT)^{-1} tr\left(\Lambda^{r+}(\pi) F_a^{r+}(\pi)' \widetilde{F}_a^r(\pi)(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi))'\right) \\ &= -2(NT)^{-1} tr\left(\Lambda^{r+}(\pi)\left(F_a^{r+}(\pi) - \widetilde{F}_a^{r+}(\pi)\right)' \widetilde{F}_a^r(\pi)(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi))'\right) \\ &= \frac{T_a}{T} O_{p\pi}(C_{NT}^{-1}) Z_\lambda(\pi), \end{aligned} \tag{E.83}$$

where $\widetilde{F}_a^{r+}(\pi)$ is a submatrix of $\widetilde{F}_a(\pi)$ with columns associated with those in $F_a^{r+}(\pi)$, the second equality holds because $\widetilde{F}_a^{r+}(\pi)$ and $\widetilde{F}_a^r(\pi)$ are orthogonal by construction, and the third equality holds by arguments analogous to those in (E.25) and (E.26). The fourth term is

$$K_4 = 2(NT)^{-1} tr\left[\widehat{\Lambda}^\perp(\pi)\widetilde{F}_a^\perp(\pi)' \widetilde{F}_a^r(\pi)\left(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi)\right)\right] = 0 \tag{E.84}$$

because $\widetilde{F}_a^\perp(\pi)' \widetilde{F}_a^r(\pi) = 0$ by construction. Combining (E.79)-(E.84), we obtain

$$M_a(\widehat{\Lambda}(\pi), \widetilde{F}_a(\pi)) - M_a(\Lambda^\dagger(\pi), \widetilde{F}_a(\pi)) = \frac{T_a}{T} Z_\lambda^2(\pi) + O_{p\pi}(C_{NT}^{-1}) Z_\lambda(\pi). \tag{E.85}$$

Replacing the first subsample with the second subsample and applying similar arguments, we also have

$$M_b(\widehat{\Psi}(\pi), \widetilde{F}_b(\pi)) - M_b(\Psi^\dagger(\pi), \widetilde{F}_b(\pi)) = \frac{T_b}{T} Z_\psi^2(\pi) + O_{p\pi}(C_{NT}^{-1}) Z_\psi(\pi). \tag{E.86}$$

Plugging (E.85) and (E.86) into the left-hand side of (E.74), we obtain

$$\frac{T_a}{T} Z_\lambda^2(\pi) + O_{p\pi}(C_{NT}^{-1}) Z_\lambda(\pi) + \frac{T_b}{T} Z_\psi^2(\pi) + O_{p\pi}(C_{NT}^{-1}) Z_\psi \leq \overline{b}_\Lambda Z_\lambda(\pi), \tag{E.87}$$

which further implies that

$$Z_\lambda(\pi) = O_{p\pi}(\overline{b}_\Lambda + C_{NT}^{-1}). \tag{E.88}$$

The unrestricted least square estimator for any $\pi \in \Pi$ can be viewed as a PLS estimator with 0 penalty. Therefore, $N^{-1}||\widetilde{\Lambda}_{LS,\ell}(\pi) - \Lambda_\ell^R(\pi)||^2 = O_{p\pi}(C_{NT}^{-2})$ for $\ell = 1, \ldots, r_a$ by (E.88), which together with (E.66) implies that $N^{-1}||\widetilde{\Lambda}_{LS,\ell}(\pi)||^2 \geq C^{-1}$ w.p.a.1. for $\ell = 1, \ldots, r_a$. For $i = 1$ and 2, we have $\omega_\ell^{\lambda*(i)}(\pi) = (N^{-1}||\widetilde{\Lambda}_{LS,\ell}(\pi)||^2)^{-d} \leq C^d$ w.p.a.1 for $\ell = 1, \ldots, r_a$. Following the specification in (4.7), $\alpha_{NT}(\pi) = \kappa_1(\pi) N^{-1/2} C_{NT_a}^{-d-1}$, where $\kappa_1(\pi) \leq \overline{\kappa}_1$. Thus, we have

$$N^{1/2}\phi_\ell^\lambda = N^{1/2}\mathbb{E}_\xi[\alpha_{NT}(\xi)\omega_\ell^{\lambda*}(\xi)] = O_p(C_{NT}^{-1}) \tag{E.89}$$

for $\ell = 1, \ldots, r_a$, which implies

$$\overline{b}_\Lambda = O_p(C_{NT}^{-1}) \tag{E.90}$$

for both the first- and second-step PLS estimation. It follows from (E.88) that $Z_\lambda(\pi) = O_{p\pi}(C_{NT}^{-1})$. This completes the proof of $\Pr(\min_{\pi\in\Pi} \widehat{r}_a^{(i)}(\pi) \geq r_a) \to 1$ for $i = 1, 2$.

Step 2. We show for $i = 1$ and 2,

$$\Pr(\min_{\pi\in\Pi} \widehat{r}_b^{(i)}(\pi) \geq r_b) \to 1 \text{ if } r_b > r_a. \tag{E.91}$$

In this case, $N^{-1}||\Gamma_\ell^R(\pi)||^2 \geq C$ by Assumption R*(ii) and $N^{-1}||\Psi_\ell^R(\pi)||^2 \geq C$ by (E.66) for $\ell = r_b$. To show (E.91), it is sufficient to prove $N^{-1}||\widehat{\Gamma}_\ell(\pi) - \Gamma_\ell^R(\pi)||^2 = o_{p\pi}(1)$ for $\ell = r_b$ for both the first and second step estimators. To this end, we redefine $\Lambda^\dagger(\pi)$ and $\Gamma^\dagger(\pi)$ in (E.69) as

$$\Lambda^\dagger(\pi) = \widehat{\Lambda}(\pi),\ \Gamma^\dagger(\pi) = \left(\widehat{\Gamma}(\pi)_{1:r_b-1},\ \Gamma_{r_b}^R(\pi),\ \widehat{\Gamma}(\pi)_{r_b+1:k}\right) \text{ and } \Psi^\dagger(\pi) = \Lambda^\dagger(\pi) + \Gamma^\dagger(\pi) \tag{E.92}$$

and keep the definitions of $Z_\lambda(\pi)$, $Z_\psi(\pi)$, $Z_\gamma(\pi)$ in (E.70) unchanged. Now consider the inequality in (E.74). Because $\Lambda^\dagger(\pi) = \widehat{\Lambda}(\pi)$, the right-hand side of (E.74) becomes for $\ell = r_b$,

$$P_2^*(\Gamma^\dagger(\pi)) - P_2^*(\widehat{\Gamma}(\pi)) = \phi_\ell^\gamma \left(|\Gamma_\ell^R(\pi)| - |\widehat{\Gamma}_\ell(\pi)|\right) \leq \overline{b}_{\Gamma b} Z_\gamma(\pi), \text{ where } \overline{b}_{\Gamma b} = N^{1/2}\phi_\ell^\gamma. \tag{E.93}$$

By arguments analogous to those used to show (E.85) and (E.86), the left-hand side of (E.74) becomes

$$M_b(\widehat{\Psi}(\pi), \widetilde{F}_b(\pi)) - M_b(\Psi^\dagger(\pi), \widetilde{F}_b(\pi)) = \frac{T_b}{T} Z_\psi^2(\pi) + O_{p\pi}(C_{NT}^{-1}) Z_\psi(\pi). \tag{E.94}$$

Putting (E.93) and (E.94) together with (E.74), we get

$$Z_\psi(\pi) = O_{p\pi}(\overline{b}_{\Gamma b} + C_{NT}^{-1}). \tag{E.95}$$

Note that we can show the consistency of $\widehat{\Lambda}(\pi)$ and $\widehat{\Psi}(\pi)$ column by column because $\widetilde{F}_a(\pi)$ and $\widetilde{F}_b(\pi)$ both have orthogonal regressors by construction. Now following the arguments used to show (E.89), we have $\overline{b}_{\Gamma b} = O_p(C_{NT}^{-1})$ for the first-step estimator, which immediately implies that $Z_\psi(\pi) = O_{p\pi}(C_{NT}^{-1})$ and

$$N^{-1}||\widehat{\Gamma}_\ell(\pi) - \Gamma_\ell^R(\pi)||^2 = O_{p\pi}(C_{NT}^{-2}) \text{ for } \ell = r_b. \tag{E.96}$$

This proof (E.91) holds for $i = 1$ and also implies that $\widehat{r}_b^{(1)} = \min_{\pi\in\Pi} \widehat{r}_b^{(1)}(\pi) \geq r_b > r_a$ w.p.a.1. Thus, for the second-step estimator, $\omega_\ell^{\gamma*(2)}(\pi)$ takes the form in (E.67) with $r_b > r_a$ w.p.a.1, which is the same as that for the first-step estimator. Hence, $\overline{b}_{\Gamma b} = O_p(C_{NT}^{-1})$ for the second-step estimator and it follows that (E.91) holds for $i = 2$ as well.

Step 3. We prove

$$\Pr(\widehat{r}_a^{(1)} = r_a) \to 1 \tag{E.97}$$

by showing that the inequalities in (E.68) become equalities when $\pi = \pi_0$. To this end, it is sufficient to show $\Pr(\widehat{\Lambda}_\ell(\pi_0) = 0) \to 1$ for $\ell > r_a$ in the first-step estimation. (We use generic notation below without superscript (1) for notational simplicity.) By the proof of Theorem 1, to obtain $\Pr(\widehat{\Lambda}_\ell(\pi_0) = 0) \to 1$, it is sufficient to show

$$\left\| e^a(\widehat{\Lambda}(\pi_0))' \widetilde{F}_{a,\ell}(\pi_0) \right\| + \left\| e^b(\widehat{\Lambda}(\pi_0) + \widehat{\Gamma}(\pi_0))' \widetilde{F}_{b,\ell}(\pi_0) \right\| < \frac{NT}{2} \phi_\ell^\lambda, \tag{E.98}$$

which is similar to (E.40). Replacing $\widehat{\Lambda}$ and $\widehat{\Gamma}$ in the proof of Theorem 1 with $\widehat{\Lambda}(\pi_0)$ and $\widehat{\Gamma}(\pi_0)$, respectively, we have

$$N^{-1/2}||\widehat{\Lambda}(\pi_0) - \Lambda^*|| = O_p(\overline{b}_\Lambda + \overline{b}_\Gamma + C_{NT}^{-1}) \text{ and } N^{-1/2}||\widehat{\Gamma}(\pi_0) - \Gamma^*|| = O_p(\overline{b}_\Lambda + \overline{b}_\Gamma + C_{NT}^{-1}), \tag{E.99}$$

where

$$\overline{b}_\Lambda = N^{1/2}(\sum_{\ell=1}^{r_a} (\phi_\ell^\lambda)^2)^{1/2} \text{ and } \overline{b}_\Gamma = N^{1/2}(\sum_{\ell \in \mathcal{Z}} (\phi_\ell^\gamma)^2)^{1/2}. \tag{E.100}$$

We have shown $\overline{b}_\Lambda = O_p(C_{NT}^{-1})$ in (E.90) for both the first- and second-step estimators. By similar arguments under Assumption R*(i) and (E.66), we also have $\overline{b}_\Gamma = O_p(C_{NT}^{-1})$ for the first step estimator. Because $\overline{b}_\Lambda = O_p(C_{NT}^{-1})$ and $\overline{b}_\Gamma = O_p(C_{NT}^{-1})$,

$$N^{-1/2}||\widehat{\Lambda}^{(1)}(\pi_0) - \Lambda^*|| = O_p(C_{NT}^{-1}) \text{ and } N^{-1/2}||\widehat{\Gamma}^{(1)}(\pi_0) - \Gamma^*|| = O_p(C_{NT}^{-1}). \tag{E.101}$$

Following the arguments used to show (E.43) and (E.44), (E.98) holds provided that

$$N^{-1/2} C_{NT}^{-1} = o_p(\phi_\ell^\lambda), \tag{E.102}$$

where $\phi_\ell^\lambda = \mathbb{E}_\xi[\alpha_{NT}(\xi)\omega_\ell^{\lambda*(1)}(\xi)]$. Using $\alpha_{NT}(\pi) = \kappa_1(\pi) N^{-1/2} C_{NT_a}^{-d-1}$, we have

$$\phi_\ell^\lambda = \mathbb{E}_\xi[\alpha_{NT}(\xi)\omega_\ell^{\lambda*(1)}(\xi)] \geq \underline{\kappa}_1 N^{-1/2} C_{NT}^{-d-1} \mathbb{E}_\xi[\omega_\ell^{\lambda*(1)}(\xi)\mathcal{I}_{\{\xi \leq \pi_0\}}], \tag{E.103}$$

where $\underline{\kappa}_1$ is the lower bound of $\kappa_1(\pi)$. For $\pi \leq \pi_0$, $X_a(\pi)$ has r_a factors. Thus, the unrestricted least square estimator $N^{-1}||\widetilde{\Lambda}_{LS,\ell}(\pi)||^2 = O_{p\pi}(C_{NT}^{-2})$ for $\ell > r_a$, by arguments analogous to (E.50). Therefore,

$$\sup_{\pi \leq \pi_0} \left(\omega_\ell^{\lambda*(1)}(\pi) \right)^{-1} = \sup_{\pi \leq \pi_0} [N^{-1}||\widetilde{\Lambda}_{LS,\ell}(\pi)||^2]^d = O_p(C_{NT}^{-2d}) \text{ for } \ell > r_a. \tag{E.104}$$

Thus, for $\ell > r_a$,

$$\begin{aligned} N^{-1/2}C_{NT}^{-1}(\phi_\ell^\lambda)^{-1} &\leq \underline{\kappa}_1^{-1}C_{NT}^d\left(\mathbb{E}_\xi[\omega_\ell^{\lambda*(1)}(\xi)\mathcal{I}_{\{\xi\leq\pi_0\}}]\right)^{-1} \\ &\leq \underline{\kappa}_1^{-1}C_{NT}^d\left(\inf_{\pi\leq\pi_0}[\omega_\ell^{\lambda*(1)}(\pi)]\mathbb{E}_\xi[\mathcal{I}_{\{\xi\leq\pi_0\}}]\right)^{-1} \\ &= \frac{C_{NT}^d\sup_{\pi\leq\pi_0}\left(\omega_\ell^{\lambda*(1)}(\pi)\right)^{-1}}{\underline{\kappa}_1\mathbb{E}_\xi[\mathcal{I}_{\{\xi\leq\pi_0\}}]} = O_p(C_{NT}^{-d}), \end{aligned} \tag{E.105}$$

where the last equality is by (E.104) and $\underline{\kappa}_1\mathbb{E}_\xi[\mathcal{I}_{\{\xi\leq\pi_0\}}] > C > 0$ for some fixed constant C. It follows that $\Pr(\widehat{\Lambda}_\ell^{(1)}(\pi_0) = 0) \to 1$ for $\ell > r_a$, which implies that

$$\Pr(\widehat{r}_a^{(1)}(\pi_0) \leq r_a) \to 1. \tag{E.106}$$

Combining (E.68) with the result above, we obtain $\Pr(\min_{\pi\in\Pi}\widehat{r}_a^{(1)}(\pi) = \widehat{r}_a^{(1)}(\pi_0) = r_a) \to 1$. This proves (E.97).

Step 4. We prove

$$\Pr(\widehat{r}_b^{(1)} = r_b) \to 1 \tag{E.107}$$

by showing that the inequalities in (E.91) become equalities when $\pi = \pi_0$. To this end, it is sufficient to show $\Pr(\widehat{\Gamma}_\ell^{(1)}(\pi_0) = 0) \to 1$ for $\ell > r_b$. (We use generic notation below without superscript (1) for notational simplicity.) By the proof of Theorem 1, to obtain $\Pr(\widehat{\Gamma}_\ell(\pi_0) = 0) \to 1$, it is sufficient to show

$$\left\|e^b(\widehat{\Lambda}(\pi_0) + \widehat{\Gamma}(\pi_0))'\widetilde{F}_{b,\ell}(\pi_0)\right\| < \frac{NT}{2}\phi_\ell^\gamma. \tag{E.108}$$

To this end, it is sufficient to show $N^{-1/2}C_{NT}^{-1} = o_p(\phi_\ell^\gamma)$. Using $\beta_{NT}(\pi) = \kappa_2(\pi)N^{-1/2}C_{NT_b}^{-d-1}$, we have

$$\phi_\ell^\gamma = \mathbb{E}_\xi[\beta_{NT}(\xi)\omega_\ell^{\gamma*(1)}(\xi)] \geq \underline{\kappa}_2 N^{-1/2}C_{NT}^{-d-1}\mathbb{E}_\xi[\omega_\ell^{\gamma*(1)}(\xi)\mathcal{I}_{\{\xi\geq\pi_0\}}], \tag{E.109}$$

where $\underline{\kappa}_2$ is the lower bound of $\kappa_2(\pi)$. For $\pi \geq \pi_0$, $X_b(\pi)$ has r_b factors, thus $N^{-1}||\widetilde{\Psi}_{LS,\ell}(\pi)||^2 = O_{p\pi}(C_{NT}^{-2})$ for $\ell > r_b$ by arguments analogous to (E.50). Therefore,

$$\sup_{\pi>\pi_0}\left(\omega_\ell^{\gamma*(1)}(\pi)\right)^{-1} \leq \sup_{\pi>\pi_0}[N^{-1}||\widetilde{\Psi}_{LS,\ell}(\pi)||^2]^d = O_p(C_{NT}^{-2d}) \text{ for } \ell > r_b. \tag{E.110}$$

Thus, for $\ell > r_b$,

$$\begin{aligned} N^{-1/2}C_{NT}^{-1}(\phi_\ell^\gamma)^{-1} &\leq \underline{\kappa}_2^{-1}C_{NT}^d\left(\mathbb{E}_\xi[\omega_\ell^{\gamma*(1)}(\xi)\mathcal{I}_{\{\xi\geq\pi_0\}}]\right)^{-1} \\ &\leq \underline{\kappa}_2^{-1}C_{NT}^d\left(\inf_{\pi>\pi_0}\left(\omega_\ell^{\gamma*(1)}(\pi)\right)\mathbb{E}_\xi[\mathcal{I}_{\{\xi\geq\pi_0\}}]\right)^{-1} \\ &= \frac{C_{NT}^d\sup_{\pi>\pi_0}\left(\omega_\ell^{\gamma*(1)}(\pi)\right)^{-1}}{\underline{\kappa}_2\mathbb{E}_\xi[\mathcal{I}_{\{\xi\geq\pi_0\}}]} = O_p(C_{NT}^{-d}), \end{aligned} \tag{E.111}$$

following from (E.110) and $\underline{\kappa}_2\mathbb{E}_\xi[\mathcal{I}_{\{\xi\geq\pi_0\}}] > C > 0$ for some fixed constant C. It follows that $\Pr(\widehat{\Gamma}_\ell^{(1)}(\pi_0) = 0) \to 1$ for $\ell > r_b$, which implies

$$\Pr(\widehat{r}_b^{(1)}(\pi_0) \leq r_b) \to 1. \tag{E.112}$$

When $r_b > r_a$, (E.91) and (E.112) imply that

$$\Pr(\widehat{r}_b^{(1)} = \min_{\pi\in\Pi} \widehat{r}_b^{(1)}(\pi) = r_b) \to 1. \tag{E.113}$$

On the other hand, if $r_b = r_a$, we can use (E.112) to deduce that

$$\Pr(\min_{\pi\in\Pi} \widehat{r}_b^{(1)}(\pi) \leq r_a) \to 1, \tag{E.114}$$

which together with the definition of $\widehat{r}_b^{(1)}$ and (E.97) implies that

$$\Pr(\widehat{r}_b^{(1)} = \widehat{r}_a^{(1)} = r_b) \to 1. \tag{E.115}$$

This completes the proof of Step 4.

Step 5. We show

$$\Pr(\widehat{r}_a^{(2)} = r_a) \to 1 \text{ and } \Pr(\widehat{r}_b^{(2)} = r_b) \to 1. \tag{E.116}$$

Following Steps 3 and 4, we know that the event $\{\widehat{r}_a^{(1)} = r_a \text{ and } \widehat{r}_b^{(1)} = r_b\}$ has probability approaching 1. If $r_b > r_a$, $\omega_\ell^{\lambda*(i)}$ and $\omega_\ell^{\gamma*(i)}$ are the same for $i = 1, 2$ following (E.67). Hence, all arguments in Steps 3 and 4 apply to the second-step estimator, which completes the proof immediately.

Next, we consider $r_a = r_b$. Conditioning on the event $\{\widehat{r}_a^{(1)} = r_a \text{ and } \widehat{r}_b^{(1)} = r_b\}$, the proofs in Step 3 and Step 4 apply to the second-step estimator as well, and this gives the desired results.

Step 6. We show that when there is a type-1 change,

$$\Pr(\widehat{\Gamma}^{(2)}(\pi_0) \neq 0) \to 1. \tag{E.117}$$

To this end, it is sufficient to show $N^{-1}||\widehat{\Gamma}_\ell^{(2)}(\pi_0) - \Gamma_\ell^R(\pi_0)||^2 \to_p 0$ for some $\ell \in \mathcal{Z}$. This follows from (E.99) for the second-step estimator, which holds by the same arguments as in Step 3 conditioning on the event $\{\widehat{r}_a^{(1)} = r_a \text{ and } \widehat{r}_b^{(1)} = r_b\}$. Following Steps 3 and 4, this event occurs w.p.a.1.

The result in (E.117) and Step 5 together imply that $\Pr(\widehat{\mathcal{B}} = 1) \to 1$ if $\mathcal{B}_0 = 1$.

Step 7. When there is no structural instability, i.e., $\Gamma^0 = 0$, we show

$$\Pr(\sup_{\pi \in \Pi} ||\widehat{\Gamma}^{(2)}(\pi)|| = 0) \to 1. \tag{E.118}$$

Replacing $\widehat{\Lambda}$ and $\widehat{\Gamma}$ in the proof of Theorem 1 with $\widehat{\Lambda}^{(2)}(\pi)$ and $\widehat{\Gamma}^{(2)}(\pi)$, we have uniform consistency

$$N^{-1/2}||\widehat{\Lambda}^{(2)}(\pi) - \Lambda^*|| = O_{p\pi}(\bar{b}_\Lambda + C_{NT}^{-1}) \text{ and } N^{-1/2}||\widehat{\Gamma}^{(2)}(\pi) - \Gamma^*|| = O_{p\pi}(\bar{b}_\Lambda + C_{NT}^{-1}), \tag{E.119}$$

where $\bar{b}_\Lambda = N^{1/2}(\sum_{\ell=1}^{r_a}(\phi_\ell^\lambda)^2)^{1/2}$. We have shown $\bar{b}_\Lambda = O_p(C_{NT}^{-1})$ in (E.90). Revoking the proof of Theorem 1 with π_0 replaced by π, a sufficient condition for (E.118) is

$$N^{-1/2}C_{NT}^{-1} = o_p(\phi_\ell^\gamma) \text{ for } \ell = 1, \ldots, k, \tag{E.120}$$

where the left-hand side follows from uniform convergence rate of the criterion function and the right-hand side is based on the averaging penalty. Following Steps 3 and 4, we know that the event $\{\widehat{r}_a^{(1)} = r_a \text{ and } \widehat{r}_b^{(2)} = r_b\}$ has probability approaching 1. Using $\beta_{NT}(\pi) = \kappa_2(\pi)N^{-1/2}C_{NT_b}^{-d-1}$, we have

$$\phi_\ell^\gamma = \mathbb{E}_\xi[\beta_{NT}(\xi)\omega_\ell^{\gamma*(2)}(\xi)] \geq \underline{\kappa}_2 N^{-1/2}C_{NT}^{-d-1}\mathbb{E}_\xi[\omega_\ell^{\gamma*(2)}(\xi)]. \tag{E.121}$$

Using the formula of $\omega_\ell^{\gamma*(2)}(\pi)$ in (E.67), for $\ell > r_a$,

$$\left(\omega_\ell^{\gamma*(2)}(\pi)\right)^{-1} = \left(\omega_\ell^{\gamma*(1)}(\pi)\right)^{-1} \leq \left(N^{-1}||\widetilde{\Gamma}_{\ell,LS}(\pi)||^2\right)^d = O_{p\pi}(C_{NT}^{-2d}) \tag{E.122}$$

w.p.a.1, where the last equality holds by arguments analogous to (E.50). On the other hand, for $\ell \leq r_a$,

$$\left(\omega_\ell^{\gamma*(2)}(\pi)\right)^{-1} \leq \left(N^{-1}||\widetilde{\Psi}_{\ell,LS}^-(\pi)\mathrm{w}(\pi) - \widetilde{\Lambda}_{\ell,LS}(\pi)||^2\right)^d = O_{p\pi}(C_{NT}^{-2d}) \tag{E.123}$$

w.p.a.1, where the equality follows from arguments analogous to (E.58) under Assumption R*(i). Combining the results in (E.122) and (E.123), we deduce that

$$\sup_{\pi \in \Pi}\left(\omega_\ell^{\gamma*(2)}(\pi)\right)^{-1} = O_p(C_{NT}^{-2d}) \text{ for } \ell = 1, \ldots, k. \tag{E.124}$$

Thus, for $\ell = 1, \ldots, k$,

$$\begin{aligned} N^{-1/2}C_{NT}^{-1}(\phi_\ell^\gamma)^{-1} &\leq \underline{\kappa}_2^{-1}C_{NT}^d\left(\mathbb{E}_\xi[\omega_\ell^{\gamma*(2)}(\xi)]\right)^{-1} \\ &\leq \underline{\kappa}_2^{-1}C_{NT}^d\left(\inf_{\pi \in \Pi}\left(\omega_\ell^{\gamma*(2)}(\pi)\right)\right)^{-1} \\ &= \underline{\kappa}_2^{-1}C_{NT}^d \sup_{\pi \in \Pi}\left(\omega_\ell^{\gamma*(2)}(\pi)\right)^{-1} = O_p(C_{NT}^{-d}), \end{aligned} \tag{E.125}$$

following from (E.121) and (E.124). The condition in (E.120) follows from (E.125), and it is sufficient for the desired result. Therefore, if $\mathcal{B}_0 = 0$, we have $\Pr(\widehat{\mathcal{B}}_0 = 0) \to 1$. This completes the proof. □

Proof of Theorem 4. For any small number $\varepsilon > 0$, we define

$$\begin{aligned} \Pi_{0,\varepsilon}^{-} &= \{\pi \in \Pi_0 : \pi \leq \pi_0 - \varepsilon\}, \\ \Pi_{0,\varepsilon}^{+} &= \{\pi \in \Pi_0 : \pi \geq \pi_0 + \varepsilon\}. \end{aligned} \tag{E.126}$$

Let $\Pi_{0,\varepsilon} = \Pi_{0,\varepsilon}^{-} \cup \Pi_{0,\varepsilon}^{+} = \{\pi \in \Pi_0 : |\pi - \pi_0| \geq \varepsilon\}$. To show $\widehat{\pi}$ is a consistent estimator of π_0, it is sufficient to show that for any $\varepsilon > 0$, $\Pr(\widehat{\pi} \in \Pi_{0,\varepsilon}) \to 0$ as $N, T \to \infty$. The proof is divided into 5 steps.

Step 1. We show that

$$\sup_{\pi \in \Pi_0} |Q_{NT}(\pi; \widehat{r}_a, \widehat{r}_b) - Q_{NT}(\pi; r_a, r_b)| = o_p(1). \tag{E.127}$$

By definition, we can write

$$\begin{aligned} Q_{NT}(\pi; \widehat{r}_a, \widehat{r}_b) &= \sum_{i=1}^{k} \sum_{j=1}^{k} Q_{NT}(\pi; i, j) I\{(\widehat{r}_a, \widehat{r}_b) = (i, j)\} \\ &= Q_{NT}(\pi; r_a, r_b) + Q_{NT}(\pi; r_a, r_b) \left[I\{(\widehat{r}_a, \widehat{r}_b) = (r_a, r_b)\} - 1\right] \\ &\quad + \sum_{i,j=1,(i,j)\neq(r_a,r_b)}^{k} Q_{NT}(\pi; i, j) I\{(\widehat{r}_a, \widehat{r}_b) = (i, j)\}, \end{aligned} \tag{E.128}$$

where the second first equality holds because $\widehat{r}_a$ and $\widehat{r}_b$ are random variables that take integer values between 1 and k and the second equality separates the event $(\widehat{r}_a, \widehat{r}_b) = (r_a, r_b)$ from the events $(\widehat{r}_a, \widehat{r}_b) \neq (r_a, r_b)$.

In view of (E.128), to show (E.127), we see that it is sufficient to show

$$I\{(\widehat{r}_a, \widehat{r}_b) = (r_a, r_b)\} = 1 + o_p(1) \text{ and} \tag{E.129}$$

$$I\{(\widehat{r}_a, \widehat{r}_b) = (i, j)\} = o_p(1) \tag{E.130}$$

for any $1 \leq i, j \leq k$ and $(i, j) \neq (r_a, r_b)$, and

$$\sup_{\pi \in \Pi_0} Q_{NT}(\pi; i, j) = O_p(1) \tag{E.131}$$

for $1 \leq i, j \leq k$. Because $\widehat{r}_a$ and $\widehat{r}_b$ are consistent estimators of r_a and r_b, respectively, (E.129) and (E.130) follow from the Markov's inequality and

$$E\left[|I\{(\widehat{r}_a, \widehat{r}_b) = (r_a, r_b)\} - 1|\right] = 1 - \Pr((\widehat{r}_a, \widehat{r}_b) = (r_a, r_b)) = o(1). \tag{E.132}$$

To show the equality in (E.131), we note that

$$\begin{aligned}
& Q_{NT}(\pi; i, j) && \text{(E.133)}\\
= \;& (NT)^{-1}\left(||X_a(\pi)||^2 + ||X_b(\pi)||^2 - \sum_{\ell=1}^{i} \rho_\ell(X_a(\pi)X_a(\pi)') - \sum_{\ell=1}^{j} \rho_\ell(X_b(\pi)X_b(\pi)') \right)
\end{aligned}$$

because $\widetilde{F}_a(\pi)$, $\widetilde{F}_b(\pi)$, $\widetilde{\Lambda}_{LS}(\pi)$ and $\widetilde{\Psi}_{LS}(\pi)$ are the principal component estimators of factors and loadings. Thus, it suffices to show $(NT)^{-1}||X||^2 = (NT)^{-1}(||X_a||^2 + ||X_b||^2) = O_p(1)$.

Because $X_a = F_a\Lambda^{0\prime} + e_a$, we have

$$\begin{aligned}
(NT)^{-1}||X_a||^2 &= (NT)^{-1} tr(\Lambda_0'\Lambda_0 F_a' F_a) + 2(NT)^{-1} tr(F_a\Lambda^{0\prime} e_a') + (NT)^{-1} tr(e_a e_a') \\
&\le 2(NT)^{-1/2}\frac{\|\Lambda^0\|}{N^{1/2}}\left| tr(\frac{F_a F_a'}{T} e_a e_a')\right|^{1/2} + O_p(1) \\
&\le 2\frac{\|\Lambda^0\|}{N^{1/2}}\sqrt{(NT)^{-1}\rho_1(e_a e_a')}\sqrt{tr(T^{-1}F_a F_a')} + O_p(1) = O_p(1), \qquad \text{(E.134)}
\end{aligned}$$

where the first equality is a simple expansion, the first inequality holds by the Cauchy-Schwarz inequality and Assumptions A, B, C, the second inequality follows from the Von Neumann's trace inequality, and the last equality holds by Assumptions A, B, C. Similarly, we can show $(NT)^{-1}||X_b||^2 = O_p(1)$. This completes the proof of step 1.

Step 2. We show that

$$\max_{\ell=1,\ldots,r_a}\sup_{\pi\in\Pi_{0,\varepsilon}^-}\left|\rho_\ell\left(\frac{X_a(\pi)X_a(\pi)'}{NT}\right) - \pi\rho_\ell(\Sigma_\Lambda\Sigma_F)\right| = o_p(1), \qquad \text{(E.135)}$$

$$\max_{\ell=1,\ldots,(r_a+r_b)}\sup_{\pi\in\Pi_{0,\varepsilon}^-}\left|\rho_\ell\left(\frac{X_b(\pi)X_b(\pi)'}{NT}\right) - \rho_\ell\left(\Sigma_F^+\Sigma_{\Lambda\Psi}^+\right)\right| = o_p(1). \qquad \text{(E.136)}$$

Note that for $\pi \in \Pi_{0,\varepsilon}^-$, we have $\pi \le \pi_0 - \varepsilon$ for $\varepsilon > 0$.

Let γ be any $T_a \times 1$ real vector with $\gamma'\gamma \le 1$. Because $X_a(\pi) = F_a(\pi)\Lambda^{0\prime} + e_a(\pi)$ for $\pi < \pi_0$, we have

$$\begin{aligned}
& \left\|\left(\frac{X_a(\pi)X_a(\pi)'}{NT} - \frac{F_a(\pi)\Lambda^{0\prime}\Lambda^0 F_a'(\pi)}{NT}\right)\gamma\right\| \\
= \;& \left\|\frac{F_a(\pi)\Lambda^{0\prime}e_a'(\pi)\gamma}{NT} + \frac{e_a(\pi)\Lambda^0 F_a'(\pi)\gamma}{NT} + \frac{e_a(\pi)e_a'(\pi)\gamma}{NT}\right\| \\
\le \;& \left\|\frac{F_a(\pi)\Lambda^{0\prime}e_a'(\pi)\gamma}{NT}\right\| + \left\|\frac{e_a(\pi)\Lambda^0 F_a'(\pi)\gamma}{NT}\right\| + \left\|\frac{e_a(\pi)e_a'(\pi)\gamma}{NT}\right\| \qquad \text{(E.137)}
\end{aligned}$$

by the triangle inequality. Using Assumptions A*, B and C*(vi), we obtain

$$\begin{aligned} & \sup_{\gamma'\gamma\leq 1}\left\|\frac{F_a(\pi)\Lambda^{0\prime}e_a'(\pi)\gamma}{NT}\right\|^2 \\ = & \sup_{\gamma'\gamma\leq 1}\frac{\gamma' e_a(\pi)\Lambda^0 F_a'(\pi)F_a(\pi)\Lambda^{0\prime}e_a'(\pi)\gamma}{N^2T^2} \\ \leq & \rho_1\left(\frac{F_a'(\pi)F_a(\pi)}{T}\right)\rho_1\left(\frac{\Lambda^0\Lambda^{0\prime}}{N}\right)\rho_1\left(\frac{e_a(\pi)e_a'(\pi)}{NT}\right) \\ = & o_p(1) \end{aligned} \tag{E.138}$$

uniformly over $\pi\in\Pi_{0,\varepsilon}^-$. Similarly we can show that

$$\sup_{\gamma'\gamma\leq 1}\left\|\frac{e_a(\pi)\Lambda^0 F_a'(\pi)\gamma}{NT}\right\|^2 = o_p(1) \tag{E.139}$$

uniformly over $\pi\in\Pi_{0,\varepsilon}^-$. By Assumption C*(vi),

$$\begin{aligned} \sup_{\gamma'\gamma\leq 1}\left\|\frac{e_a(\pi)e_a'(\pi)\gamma}{NT}\right\|^2 &= \sup_{\gamma'\gamma\leq 1}\frac{\gamma' e_a(\pi)e_a'(\pi)e_a(\pi)e_a'(\pi)\gamma}{N^2T^2} \\ &= \left(\rho_1\left(\frac{e_a(\pi)e_a'(\pi)}{NT}\right)\right)^2 = o_p(1) \end{aligned} \tag{E.140}$$

uniformly over $\pi\in\Pi_{0,\varepsilon}^-$. Combining the results in (E.137), (E.138), (E.139) and (E.140), and applying Theorem 5, we get

$$\max_{\ell=1,\ldots,r_a}\sup_{\pi\in\Pi_{0,\varepsilon}^-}\left|\rho_\ell\left(\frac{X_a(\pi)X_a(\pi)'}{NT}\right)-\rho_\ell\left(\frac{F_a(\pi)\Lambda^{0\prime}\Lambda^0F_a'(\pi)}{NT}\right)\right| = o_p(1). \tag{E.141}$$

By Assumptions A* and B, we have

$$\sup_{\pi\in\Pi_{0,\varepsilon}^-}\left\|\frac{\Lambda^{0\prime}\Lambda^0F_a'(\pi)F_a(\pi)}{NT}-\pi\Sigma_\Lambda\Sigma_F\right\| = o_p(1) \tag{E.142}$$

which together with $\|\cdot\|_{op}\leq\|\cdot\|$ and Theorem 5 implies that

$$\max_{\ell=1,\ldots,r_a}\sup_{\pi\in\Pi_{0,\varepsilon}^-}\left|\rho_\ell\left(\frac{\Lambda^{0\prime}\Lambda^0F_a'(\pi)F_a(\pi)}{NT}\right)-\pi\rho_\ell\left(\Sigma_\Lambda\Sigma_F\right)\right| = o_p(1). \tag{E.143}$$

Combining (E.141), (E.143), and

$$\rho_\ell\left(\frac{F_a(\pi)\Lambda^{0\prime}\Lambda^0F_a'(\pi)}{NT}\right) = \rho_\ell\left(\frac{\Lambda^{0\prime}\Lambda^0F_a'(\pi)F_a(\pi)}{NT}\right) \tag{E.144}$$

for $\ell=1,\ldots,r_a$, and applying the triangle inequality, we obtain (E.135).

For any $\ell = 1, \ldots, r_a + r_b$, we have

$$\rho_\ell \left(\frac{X_b(\pi) X_b(\pi)'}{NT} \right) = \rho_\ell \left(\frac{X_b(\pi)' X_b(\pi)}{NT} \right). \tag{E.145}$$

By $X_b(\pi) = F_a^+(\pi)\Lambda^{0\prime} + F_b(\pi)\Psi^{0\prime} + e_b(\pi)$, we have

$$\begin{aligned} \frac{X_b(\pi)' X_b(\pi)}{NT} &= \frac{\Lambda^0 F_a^{+\prime}(\pi) F_a^+(\pi) \Lambda^{0\prime}}{NT} + \frac{\Psi^0 F_b'(\pi) F_b(\pi) \Psi^{0\prime}}{NT} \\ &\quad + \frac{e_b'(\pi) F_a^+(\pi) \Lambda^{0\prime}}{NT} + \frac{e_b'(\pi) F_b(\pi) \Psi^{0\prime}}{NT} + \frac{e_b'(\pi) e_b(\pi)}{NT} \end{aligned} \tag{E.146}$$

for any $\pi < \pi_0$. Using arguments similar to those that derive (E.138), (E.139) and (E.140), we can show that

$$\sup_{\pi \in \Pi_{0,\varepsilon}^-} \left[\left\| \frac{e_b'(\pi) F_a^+(\pi) \Lambda^{0\prime}}{NT} \right\|_{op} + \left\| \frac{e_b'(\pi) F_b(\pi) \Psi^{0\prime}}{NT} \right\|_{op} + \left\| \frac{e_b'(\pi) e_b(\pi)}{NT} \right\|_{op} \right] = o_p(1) \tag{E.147}$$

which together with (E.145), the triangle inequality and Theorem 5 implies that

$$\max_{\ell=1,\ldots,r_a+r_b} \sup_{\pi \in \Pi_{0,\varepsilon}^-} \left| \rho_\ell \left(\frac{X_b(\pi) X_b(\pi)'}{NT} \right) - \rho_\ell \left(\frac{\Lambda^0 F_a^{+\prime}(\pi) F_a^+(\pi) \Lambda^{0\prime}}{NT} + \frac{\Psi^0 F_b'(\pi) F_b(\pi) \Psi^{0\prime}}{NT} \right) \right| = o_p(1). \tag{E.148}$$

Note that we can write

$$\begin{aligned} &\frac{\Lambda^0 F_a^{+\prime}(\pi) F_a^+(\pi) \Lambda^{0\prime}}{NT} + \frac{\Psi^0 F_b'(\pi) F_b(\pi) \Psi^{0\prime}}{NT} \\ &= \frac{(\Lambda^0, \Psi^0)}{N} \begin{pmatrix} \frac{F_a^{+\prime}(\pi) F_a^+(\pi)}{T} & 0 \\ 0 & \frac{F_b'(\pi) F_b(\pi)}{T} \end{pmatrix} \frac{(\Lambda^0, \Psi^0)'}{N}. \end{aligned} \tag{E.149}$$

Under Assumptions A* and B, we have

$$\sup_{\pi \in \Pi_{0,\varepsilon}^-} \left\| \begin{pmatrix} \frac{F_a^{+\prime}(\pi) F_a^+(\pi)}{T} & 0 \\ 0 & \frac{F_b'(\pi) F_b(\pi)}{T} \end{pmatrix} \begin{pmatrix} \frac{\Lambda^{0\prime}\Lambda^0}{N} & \frac{\Lambda^{0\prime}\Psi^0}{N} \\ \frac{\Psi^{0\prime}\Lambda^0}{N} & \frac{\Psi^{0\prime}\Psi^0}{N} \end{pmatrix} - \Sigma_F^+ \Sigma_{\Lambda\Psi}^+ \right\| = o_p(1) \tag{E.150}$$

which combined with $\|\cdot\|_{op} \leq \|\cdot\|$ and Theorem 5 implies that

$$\max_{\ell=1,\ldots,r_a+r_b} \sup_{\pi \in \Pi_{0,\varepsilon}^-} \left| \rho_\ell \left(\frac{\Lambda^0 F_a^{+\prime}(\pi) F_a^+(\pi) \Lambda^{0\prime}}{NT} + \frac{\Psi^0 F_b'(\pi) F_b(\pi) \Psi^{0\prime}}{NT} \right) - \rho_\ell \left(\Sigma_F^+ \Sigma_{\Lambda\Psi}^+ \right) \right| = o_p(1). \tag{E.151}$$

Collecting the results in (E.148) and (E.151), and applying the triangle inequality, we immediately get (E.136).

Step 3. Using the arguments similar to those in step 2, we also have

$$\max_{\ell=1,...,r_a} \left| \rho_\ell \left(\frac{X_a(\pi_0) X_a(\pi_0)'}{NT} \right) - \pi_0 \rho_\ell(\Sigma_\Lambda \Sigma_F) \right| = o_p(1), \tag{E.152}$$

$$\max_{\ell=1,...,r_b} \left| \rho_\ell \left(\frac{X_b(\pi_0) X_b(\pi_0)'}{NT} \right) - (1-\pi_0) \rho_\ell \left(\Sigma_\Psi \Sigma_{\overline{F}} \right) \right| = o_p(1). \tag{E.153}$$

Step 4. We show that

$$\Pr(\widehat{\pi} \in \Pi^-_{0,\varepsilon}) \to 0 \text{ as } N, T \to \infty. \tag{E.154}$$

By the definition of $\widehat{\pi}$,

$$\begin{aligned}
\Pr(\widehat{\pi} &\in \Pi^-_{0,\varepsilon}) \\
&\leq \Pr\left(\min_{\pi \in \Pi^-_{0,\varepsilon}} Q_{NT}(\pi; \widehat{r}_a, \widehat{r}_b) \leq Q_{NT}(\pi_0; \widehat{r}_a, \widehat{r}_b) \right) \\
&= \Pr\left(\min_{\pi \in \Pi^-_{0,\varepsilon}} Q_{NT}(\pi; r_a, r_b) \leq Q_{NT}(\pi_0; r_a, r_b) + o_p(1) \right) \\
&= \Pr\left(\min_{\pi \in \Pi^-_{0,\varepsilon}} \left(\begin{array}{c} \sum_{\ell=1}^{r_a} \left[\rho_\ell \left(\frac{X_a(\pi_0) X_a(\pi_0)'}{NT} \right) - \rho_\ell \left(\frac{X_a(\pi) X_a(\pi)'}{NT} \right) \right] \\ + \sum_{\ell=1}^{r_b} \left[\rho_\ell \left(\frac{X_b(\pi_0) X_b'(\pi_0)}{NT} \right) - \rho_\ell \left(\frac{X_b(\pi) X_b(\pi)'}{NT} \right) \right] \end{array} \right) \leq o_p(1) \right) \\
&= \Pr\left(\min_{\pi \in \Pi^-_{0,\varepsilon}} \left(\begin{array}{c} \sum_{\ell=1}^{r_a} \left[\pi_0 \rho_\ell(\Sigma_\Lambda \Sigma_F) - \pi \rho_\ell(\Sigma_\Lambda \Sigma_F) \right] \\ + \sum_{\ell=1}^{r_b} \left[(1-\pi_0) \rho_\ell \left(\Sigma_\Psi \Sigma_{\overline{F}} \right) - \rho_\ell \left(\Sigma_F^+ \Sigma_{\Lambda\Psi}^+ \right) \right] \end{array} \right) \leq o_p(1) \right)
\end{aligned} \tag{E.155}$$

where the first equality is by (E.127), the second equality follows from (E.133), and the last equality is by (E.135), (E.136), (E.152) and (E.153). It is clear that

$$\begin{aligned}
\sum_{\ell=1}^{r_b} \rho_\ell \left(\Sigma_F^+ \Sigma_{\Lambda\Psi}^+ \right) &= tr(\Sigma_F^+ \Sigma_{\Lambda\Psi}^+) - \sum_{\ell=r_b+1}^{r_a+r_b} \rho_\ell \left(\Sigma_F^+ \Sigma_{\Lambda\Psi}^+ \right) \\
&= (\pi_0 - \pi) tr(\Sigma_\Lambda \Sigma_F) + (1-\pi_0) tr(\Sigma_\Psi \Sigma_{\overline{F}}) - \sum_{\ell=r_b+1}^{r_a+r_b} \rho_\ell \left(\Sigma_F^+ \Sigma_{\Lambda\Psi}^+ \right) \\
&= (\pi_0 - \pi) \sum_{\ell=1}^{r_a} \rho_\ell(\Sigma_\Lambda \Sigma_F) + (1-\pi_0) \sum_{\ell=1}^{r_b} \rho_\ell \left(\Sigma_\Psi \Sigma_{\overline{F}} \right) - \sum_{\ell=r_b+1}^{r_a+r_b} \rho_\ell \left(\Sigma_F^+ \Sigma_{\Lambda\Psi}^+ \right)
\end{aligned} \tag{E.156}$$

which together with (E.155) implies that

$$\Pr(\widehat{\pi} \in \Pi^-_{0,\varepsilon}) \leq \Pr\left(\min_{\pi \in \Pi^-_{0,\varepsilon}} \sum_{\ell=r_b+1}^{r_a+r_b} \rho_\ell \left(\Sigma_F^+ \Sigma_{\Lambda\Psi}^+ \right) \leq o_p(1) \right). \tag{E.157}$$

By Assumption ID*, $\rho_{r_b+1}\left(\Sigma_F^+\Sigma_{\Lambda\Psi}^+\right) > 0$ implies

$$\Pr\left(\min_{\pi\in\Pi_{0,\varepsilon}^-}\sum_{\ell=r_b+1}^{r_a+r_b}\rho_\ell\left(\Sigma_F^+\Sigma_{\Lambda\Psi}^+\right) \leq o_p(1)\right) \to 0 \text{ as } N,T\to\infty. \tag{E.158}$$

From (E.157) and (E.158), we immediately get (E.154).

Step 5. We show that

$$\Pr(\widehat{\pi}\in\Pi_{0,\varepsilon}) \to 0 \text{ as } N,T\to\infty. \tag{E.159}$$

Note that we can use the arguments similar to those in Step 2 to show that

$$\begin{aligned}\max_{\ell=1,\ldots,r_a+r_b}\sup_{\pi\in\Pi_{0,\varepsilon}^+}\left|\rho_\ell\left(\frac{X_a(\pi)X_a(\pi)'}{NT}\right)-\rho_\ell\left(\Sigma_F^+\Sigma_{\Lambda\Psi}^+\right)\right| &= o_p(1),\\ \max_{\ell=1,\ldots,r_b}\sup_{\pi\in\Pi_{0,\varepsilon}^+}\left|\rho_\ell\left(\frac{X_b(\pi)X_b(\pi)'}{NT}\right)-(1-\pi)\rho_\ell\left(\Sigma_{\overline{F}}\Sigma_\Psi\right)\right| &= o_p(1),\end{aligned} \tag{E.160}$$

which combined with the arguments similar to those in Step 4 can be used to show that

$$\Pr(\widehat{\pi}\in\Pi_{0,\varepsilon}^+) \to 0 \text{ as } N,T\to\infty. \tag{E.161}$$

Collecting the results in (E.154) and (E.161), we deduce that

$$\Pr(\widehat{\pi}\in\Pi_{0,\varepsilon}) \leq \Pr(\widehat{\pi}\in\Pi_{0,\varepsilon}^-)+\Pr(\widehat{\pi}\in\Pi_{0,\varepsilon}^+) \to 0 \text{ as } N,T\to\infty. \quad \square \tag{E.162}$$