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, NBER, and Visiting Scholar Federal Reserve Bank of Philadelphia

December 2013

Abstract

In high-dimensional factor models, both the factor loadings and the number of factors may change over time. This paper proposes a shrinkage estimator that detects and disentangles these instabilities. The new method simultaneously and consistently estimates the number of pre- and post-break factors, which liberates researchers from sequential testing and achieves uniform control of the family-wise model selection errors over an increasing number of variables. The shrinkage estimator only requires the calculation of principal components and the solution of a convex optimization problem, which makes its computation efficient and accurate. The finite sample performance of the new method is investigated in Monte Carlo simulations. In an empirical application, we study the change in factor loadings and emergence of new factors during the Great Recession.

JEL Classification: C13, C33, C52

Keywords: Consistent Model Selection, Factor Model, Great Recession, High-dimensional Model, Large Data Sets, LASSO, 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 (Penn) provided excellent research assistance. Many thanks to Ataman Ozylidirim for granting us with access to a selected set of time series published by The Conference Board. We also thank Xu Han and seminar participants at the University of Pennsylvania, Yale University, the 2013 Montreal Time Series Conference, the 2013 Tsinghua Econometrics Conference, the 2013 NSF-NBER Time Series Conference, and the 2013 New York Area Econometrics Colloquium for helpful comments and suggestions. Schorfheide gratefully acknowledges financial support from the National Science Foundation under Grant SES 1061725. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. This paper is available free of charge at www.philadelphiafed.org/research-and-data/publications/working-papers/.

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. Our leading empirical example of a potential structural break is the beginning of 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. Throughout this paper, we distinguish between two types of factor model instabilities: changes in the factor loadings, which alter the response of macroeconomic variables to a fixed number of underlying factors (type-1 instability), and changes in the number of factors (type-2 instability). Since only the product of factors and loadings is identifiable, we use a normalization that attributes changes in this product to changes in loadings.

The main contribution of this paper is in providing a novel econometric procedure that allows researchers to consistently estimate the number of pre- and post-break factors simultaneously and to consistently detect changes in factor loadings in cases in which the number of factors remaines unchanged. Our estimation method relies on a penalized least squares (PLS) criterion function in which adaptive group-LASSO penalties are attached to pre-break factor loadings, to coefficients that parameterize changes in factor loadings after the break, and to coefficients that are associated with new post-break factors. The PLS estimator is a shrinkage estimator because, as compared with the unrestricted least squares estimator, it sets small coefficient estimates equal to zero. The number of factors is determined based on the number of nonzero columns in the loading matrices, and we refer to the determination of the number of factors and the presence of breaks as model selection.

Our procedure substantially differs from existing methods in several dimensions. First, to detect instabilities in factor loadings, our method does not require knowledge of the factors before and/or after the break. Second, our procedure automatically determines whether a structural break belongs to the type-1 or type-2 category. Third, for type-2 instabilities, we decompose the structural change into the contribution of new factors and the contribution of changes of loadings for the old factors.

Fourth, the consistency of our procedure is preserved even if the break date is unknown.

It is known in the literature (e.g., Breitung and Eickmeier (2011)) that the presence of instabilities leads to the overestimation of the number of factors. In turn, a misspecification of a break date leads to the overestimation of pre- or post-break factors. Thus, roughly speaking, we search for break dates that minimize the sum of the number of pre- and post-break factors. While this procedure itself does not deliver a consistent estimate of the break date, it does generate consistent estimates of the number of pre- and post-break factors — even if the "true" break date is unknown to the researcher.

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. Our empirical results differ. Using a similar set of time series, but sampled at a monthly frequency, we find evidence of a type-2 instability at the beginning of the Great Recession (i.e., the emergence of a new factor, which can be interpreted as a financial factor). Our estimation results also 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, the change in loadings in our analysis mirrors the increase in factor volatility in their analysis.

We are building on a large body of literature on the analysis of factor models. In a seminal paper, Bai and Ng (2002) provide information criteria to select the number of factors in stable factor models. We use their assumptions about cross-sectional and temporal dependence and heteroskedasticity in the idiosyncratic errors as a starting point for our theoretical analysis. In subsequent work, several authors (e.g., Onatski (2010), Alessi, Barigozzi, and Capasso (2010), Kapetanios (2010), Caner and Han (2012), Ahn and Horenstein (2013), and Choi (2013)) proposed alternative methods for estimating the number of factors in a stable environment.

Stock and Watson (2002) and, more recently, 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. Our paper focuses on large structural breaks that render the principal component estimator inconsistent. As shown, for instance, in Breitung and Eickmeier (2011), a factor model with big structural breaks can always be written as a stable model with a larger number of pseudo-true factors. These pseudo-true factors are comparable to the factors that are being estimated with a static factor model, if the "true" factors are current and lagged values of a lower-dimensional vector of dynamic factors (e.g., Forni, Hallin, Lippi, and Reichlin (2000), Amengual and Watson (2007), Bai and Ng (2007), Hallin and Lika (2007), Onatski (2009), and Breitung and Pigorsch (2013)). The above-mentioned methods for estimating the number of factors in a stable environment, are only able to determine the number of pseudo-true factors, but not the actual number of pre- and post-break factors. However, in many applications (e.g., forecasting with factor-augmented autoregressive models) the consistent estimation of the post-break factors is crucial.

Several structural break tests for factor models have been developed in the literature. Most papers, including Stock and Watson (2009), Breitung and Eickmeier (2011), Chen, Dolado, and Gonzalo (2011), Han and Inoue (2011), and Corradi and Swanson (2013), consider the null hypothesis of no change in loadings versus an alternative of a single break in the loadings, assuming that the number of factors stays constant throughout the sample. The key challenges are to ensure that the estimation error with the factors is asymptotically negligible and to control the family-wise error as the number of factor loadings, and hence the number of parameters on which the null hypothesis is imposed, tends to infinity in the large sample approximations. Our PLS estimation approach automatically controls the family-wise error over an increasing number of variables. Furthermore, to achieve consistency, we only require that the number of the 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 is negligible.

If the break point is known one, could study the emergence of new post-break factors simply by applying one of the existing methods for determining the number of factors in a stable environment to the pre- and post-break subsamples. Alternatively, Stock and Watson (2012) test for the presence of a factor in the errors associated with the forecasts of the post-break observations based on extensions of the pre-break factors. However, these approaches have not been extended to the unknown-break-date case, and they would be unable to simultaneously detect type-1 instabilities if the data provide no evidence in favor of a change in the number of factors.

The model selection mechanism in this paper employs the adaptive group LASSO estimator (Tibshirani (1994), Zou (2006), and Yuan and Lin (2006)) of a high-dimensional sparse system. This sparse system unifies the factor structures and the specification of structural changes. Specifically, we consider penalized estimation of the factor loadings and changes in factor loadings in an augmented auxiliary model where the factors are replaced by a large number of orthonormal regressors. Consistent model selection is obtained by combining superefficient estimation of the zero components and consistent estimation of the nonzero components in the sparse system. Theoretical results in the paper provide bounds on the penalization tuning parameters for consistent model selection. A practical algorithm is suggested for empirical applications. In recent work, Bai and Liao (2012), Caner and Han (2012), and Lu and Su (2013) provide important results on shrinkage estimation of stable factor models. With observed regressors, Lee, Seo, and Shin (2012) and Qian and Su (2013) propose using shrinkage estimation to detect structural breaks. The main challenge in our paper is that the factor structure is both unobserved and unstable.

The remainder of this paper is organized as follows. Section 2 describes the factor model and the types of instabilities considered in this paper. The proposed shrinkage estimator and model selection procedure are presented in Section 3 under the assumption that the break date is known. Section 4 develops the asymptotic theory for our estimator and establishes the consistency of the model selection procedure. The selection of the tuning parameters as well as the practical implementation of the shrinkage estimation are addressed in Section 5. The extension to the case in which the break date is unknown is presented in Section 6. The Monte Carlo results are reported in Section 7, and Section 8 contains the empirical application. Finally, Section 9 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. Using the break date T_0, we split the full sample into two stable subsamples: The first one contains the first T_0 observations of the N series, and the second one contains the last $T_1 = T - T_0$ observations of the N series. In the remainder of this section, we introduce the (nonidentifiable) data generating process (DGP) and an identifiable version of the DGP.

Moreover, we distinguish between two types of structural changes, the occurrence of which we will consistently estimate.

2.1 The (Nonidentifiable) 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

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

Throughout this paper, we use $C \in R$ to denote a generic positive constant. 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. We assume the factors and their loadings satisfy Assumptions A and B below.

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. $\square$

Assumptions A and B 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. Assumption B(ii) is the same as Assumption G of Bai (2003).[1]

In the remainder of this paper, we assume $r_b \geq r_a$. 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:

$$\begin{aligned} \text{type-1 change} &: \quad r_b = r_a \text{ and } \Gamma_1^0 \neq 0 \\ \text{type-2 change} &: \quad r_b > r_a. \end{aligned} \tag{2.5}$$

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

If the break date T_0 is known, the number 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. Intuitively, the change is identifiable if either the space spanned by the factor loadings or the scaling of the factor loadings changes. Formally, for any square matrix A, we use $\rho_\ell(A)$ to denote its ℓ-th largest eigenvalue. Define a $(r_a+r_b)\times(r_a+r_b)$

[1]Assumption A is sufficient for the identification conditions in Assumption ID below. 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 6.1.

augmented covariance matrix

$$\Sigma_{\Lambda\Psi}^{+} = \begin{bmatrix} \Sigma_{\Lambda} & \Sigma_{\Lambda\Psi} \\ \Sigma_{\Lambda\Psi}' & \Sigma_{\Psi} \end{bmatrix}. \tag{2.6}$$

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. 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.

2.2 An Identifiable Version of the DGP

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. This normalization also helps to further clarify identifiable type-1 structural changes. Because our estimation will be based on principal-components analysis, we normalize the factors to have an identity covariance matrix and the vectors of factor loadings to be orthogonal and sorted according to length.[2]

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

$$\Upsilon_a'(\Sigma_a^{1/2})'\Sigma_F \Sigma_a^{1/2}\Upsilon_a = V_a, \tag{2.7}$$

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 (2.7) holds for large N. Now define the transformation matrix

$$R_a = \Sigma_a^{1/2}\Upsilon_a V_a^{-1/2}. \tag{2.8}$$

[2] 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.

We can rewrite the DGP before T_0 as

$$X_a = F_a R_a R_a^{-1} \Lambda^{0\prime} + e_a = F_a^R \Lambda^{R\prime} + e_a, \tag{2.9}$$

where $F_a^R = F_a R_a$ and $\Lambda^R = \Lambda^0 (R_a^{-1})'$. 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 (2.7) by $\Sigma_{\overline{F}}$, and otherwise replace a subscripts by b subscripts. The second transformation matrix R_b is defined as

$$R_b = \Sigma_b^{1/2} \Upsilon_b V_b^{-1/2} \tag{2.10}$$

such that the second DGP can be rewritten as[3]

$$X_b = F_b R_b R_b^{-1} \Psi^{0\prime} + e_b = F_b^R \Psi^{R\prime} + e_b. \tag{2.11}$$

Assumption ID(i) implies that the column spaces of Λ^R and Ψ^R in (2.9) and (2.11) are different, whereas Assumption ID(ii) translates into $V_a \neq V_b$. 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 $V_b = V_a(1-\rho_b^2)/(1-\rho_a^2)$.

2.3 A Useful Decomposition of Structural Changes

The (nonidentifiable) 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, $F_{b,1}\Gamma_1^{0\prime}$. We can mechanically rewrite the normalized version of the post-break DGP in (2.11) as

$$X_b = F_b^R \Psi^{R\prime} + e_b = F_{b,1}^R (\Lambda^R + \Gamma_1^R)' + F_{b,2}^R \Gamma_2^{R\prime} + e_b, \tag{2.12}$$

[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.

where $F_b^R = (F_{b,1}^R, F_{b,2}^R)$, and we define $\Gamma^R = (\Gamma_1^R, \Gamma_2^R)$. However, the components $F_{b,1}^R \Gamma_1^{R'}$ and $F_{b,2}^R \Gamma_2^{R'}$ are difficult to interpret because after imposing the normalization, there is no sense in which $F_{b,1}$ can be viewed as the post-break extension of F_a.

Nonetheless, 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. To do so, 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 $\Psi^R \Omega_a$:

$$\Omega_a = \text{argmax}_{\tilde{\Omega}_a \in \mathcal{O}} \, tr[\Lambda^{R'} \Psi^R \tilde{\Omega}_a], \tag{2.13}$$

where $\mathcal{O}$ is the class of $r_b \times r_a$ matrices with orthonormal columns. The solution is given by (see Cliff (1966)) $\Omega_a = VU'$, where V is an $r_b \times r_a$ and U an $r_a \times r_a$ orthogonal matrix obtained from the singular value decomposition $\Lambda^{R'} \Psi^R = UDV'$. 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., $F_b^{R\Omega'} F_b^{R\Omega} / T_b = I$. Partitioning $F_b^{R\Omega} = (F_{b,1}^{R\Omega}, F_{b,2}^{R\Omega})$ and $\Psi^{R\Omega} = (\Psi_1^{R\Omega}, \Psi_2^{R\Omega})$, we can decompose X_b as follows:

$$X_b = F_b^R \Omega \Omega' \Psi^{R'} + e_b = \underbrace{F_{b,1}^{R\Omega} \Lambda^{R'}}_{\text{old loadings}} + \underbrace{F_{b,1}^{R\Omega} (\Psi_1^{R\Omega} - \Lambda^R)'}_{\text{change in loadings}} + \underbrace{F_{b,2}^{R\Omega} \Psi_2^{R\Omega'}}_{\text{new factor}} + e_b. \tag{2.14}$$

2.4 Model Classes

The main contribution of this paper is to develop a procedure that consistently detects the occurrence of type-1 and type-2 structural changes. Let $\mathcal{S}_0 \in \{0, 1\}$ be a binary variable such that $\mathcal{S}_0 = 0$ indicates that there is no structural instability (i.e., $\Gamma^{(0)} = 0$ in (2.3) and $r_a = r_b$). If $\mathcal{S}_0 = 1$ and $r_a = r_b$, then the DGP exhibits a type-1 instability. Finally, $\mathcal{S}_0 = 1$ 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

$$\mathcal{M}_0 = (r_a, r_b, \mathcal{S}_0). \tag{2.15}$$

We propose a consistent model selection procedure for $\mathcal{M}_0$ based on the simultaneous estimation of r_a, r_b, and $\mathcal{S}_0$. For the consistent determination of $\mathcal{M}_0$, it suffices to estimate the normalized version of the factor model in (2.9) and (2.11), because $\Gamma^0 = 0$ if and only if $\Gamma^R = 0$, where Γ_1^R and Γ_2^R are (implicitly) defined in (2.12). Moreover, $\Gamma_2^0 = 0$ if and only if $\Gamma_2^R = 0$.

Our procedure differs from the existing methods in two very important dimensions. First, our method not only detects structural instabilities but also automatically determines their type. Second, to detect instabilities 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 Model Selection with Known Break Date

In this section, we assume that the date of the potential structural break, T_0, is known. We divide the full sample into a pre-break and a post-break subsample. Let T_a and T_b denote the number of periods in the two subsamples, respectively. With a known break date, $T_a = T_0$ and $T_b = T_1$. Since we treat the number of factors as unknown, we define $k \geq r_b$ to be the number of potential factors. In order to motivate the criterion function in the shrinkage estimation, we rewrite the normalized DGP in (2.9) and (2.11) as the following augmented system:

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

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

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.

3.1 Estimation Objective Function

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 $(NT_j)^{-1} X_j X_j'$ 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

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

$$P_1(\Lambda) = \alpha_{NT} \sum_{\ell=1}^{k} \omega_\ell^\lambda \left\| \Lambda_\ell \right\| \text{ and } P_2(\Gamma) = \beta_{NT} \sum_{\ell=1}^{k} \omega_\ell^\gamma \left\| \Gamma_\ell \right\|, \tag{3.3}$$

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:

$$\omega_\ell^\lambda = \left(N^{-1} ||\widetilde{\Lambda}_\ell||^2 \mathcal{I}_{\{\widetilde{\Lambda}_\ell \neq 0_{N\times 1}\}} + N^{-1} ||\widetilde{\Lambda}_{\ell,LS}||^2 \mathcal{I}_{\{\widetilde{\Lambda}_\ell = 0_{N\times 1}\}} \right)^{-d},$$

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

Here $\mathcal{I}_{\{x=a\}}$ is the indicator function that is equal to one if $x = a$ and equal to zero otherwise; d is a positive constant; and $\widetilde{\Lambda} \in R^{N \times k}$ 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. The simplest preliminary estimator available is the unrestricted least square estimator (i.e., $\widetilde{\Lambda} = \widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma}_{LS} = \widetilde{\Gamma}_{LS}$). An alternative preliminary estimator can be a shrinkage estimator that is based on a rough choice of tuning parameters. Since such a shrinkage estimator may set some of the columns of $\tilde{\Lambda}$ and $\tilde{\Gamma}$ equal to zero, we use the indicator functions to replace zero columns with the corresponding columns of the unrestricted least square estimators. This model selection procedure is easy to compute because it is a convex optimization problem after the principal components are used for dimension reduction.

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. It is easy to obtain preliminary estimators such that $||\widetilde{\Lambda}_\ell|| \to_p 0$ if and only if the ℓ-th column of Λ^{R+} is zero and $||\widetilde{\Gamma}_\ell|| \to_p 0$ if and only if the ℓ-th column of Γ^{R+} is zero. The unrestricted least square estimator is one simple example. If such requirement is satisfied by the preliminary estimator, we expect $N^{-1}\|\widetilde{\Lambda}_\ell\|$ 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 have zero factor loadings for all series. As such, the group-LASSO estimator automatically controls the group-wise model-selection error, which is challenging if the model-selection is performed sequentially.

3.2 Model Selection

Model selection for $\mathcal{M}_0$ is based on the column norms of $\widehat{\Lambda}$ and $\widehat{\Gamma}$. The estimators of r_a and r_b are

$$\begin{aligned}
\widehat{r}_a &= \min \mathcal{J}_a, \text{ where } \mathcal{J}_a = \left\{ j : ||\widehat{\Lambda}_\ell||^2 = 0 \text{ for all } \ell > j \right\} \\
\widehat{r}_b &= \max\left(\min \mathcal{J}_b, \widehat{r}_a\right), \text{ where } \mathcal{J}_b = \left\{ j : ||\widehat{\Gamma}_\ell||^2 = 0 \text{ for all } \ell > j \right\}. \qquad (3.5)
\end{aligned}$$

Note that $\min \mathcal{J}_a$ is the last nonzero column of $\widehat{\Lambda}$ and $\min \mathcal{J}_b$ is the last nonzero column of $\widehat{\Gamma}$. The estimator of $\mathcal{S}_0$ is

$$\widehat{\mathcal{S}} = \begin{cases} 0 & \text{if } \widehat{\Gamma} = 0, \\ 1 & \text{otherwise.} \end{cases} \qquad (3.6)$$

The procedure selects a model with no structural instability if $\widehat{\mathcal{S}} = 0$. When $\widehat{\mathcal{S}} \neq 0$, the change is type-1 if $\widehat{r}_b = \widehat{r}_a$ and it is type-2 if $\widehat{r}_b > \widehat{r}_a$.[4] In all cases, we not only detect the

[4] If there is a type-2 change, the first $\widehat{r}_a$ columns of $\widehat{\Gamma}$ do not provide any implications on the stability of the "original factor loadings" in the first subsample, because they involve a rotation of all factor loadings, including the new ones.

instabilities and identify their sources but also simultaneously estimate the number of factors before and after the break. In sum, the model selected by the shrinkage estimator is

$$\widehat{\mathcal{M}} = (\widehat{r}_a, \widehat{r}_b, \widehat{\mathcal{S}}). \tag{3.7}$$

In the following sections, we show that

$$\Pr(\widehat{\mathcal{M}} = \mathcal{M}_0) \to 1 \text{ as } N, T \to \infty \tag{3.8}$$

provided that the tuning parameters α_{NT} and β_{NT} are chosen within the bounds specified below.

3.3 Estimation of Factor Loadings

In applications such as forecasting both the dimension of the factor space and the values of the factor loadings are of interest. If $\hat{\mathcal{S}} = 0$ (i.e., there is no evidence of a structural instability), then we recommend to reestimate the factors and their loadings using the full sample to improve efficiency and reduce the bias introduced by penalization. In this case, let $\widetilde{F} \in R^{T \times \widehat{r}_a}$ be the orthonormalized left eigenvectors of $(NT)^{-1}XX'$ associated with its first $\widehat{r}_a$ largest eigenvalues. The factor loading for the full sample is $\widehat{\Psi}_F = T^{-1}X'\widetilde{F}$.

If, on the other hand, there is evidence of a structural break, $\hat{\mathcal{S}} = 1$, then one could either use the pre- and post-break shrinkage estimators $\widehat{\Lambda}$ and $\widehat{\Lambda} + \widehat{\Gamma}$ of the factor loadings or re-estimate the factor model conditional on the selected number of factors $\hat{r}_a$ and $\hat{r}_b$. We call this latter estimator a post-model-selection (PMS) estimator. It is formally defined as

$$\widehat{\Lambda}_{PMS} = \left(\overline{\Lambda},\, 0_\Lambda\right), \; \widehat{\Psi}_{PMS} = \left(\overline{\Psi},\, 0_\Psi\right) \text{ and } \widehat{\Gamma}_{PMS} = \widehat{\Psi}_{PMS} - \widehat{\Lambda}_{PMS}, \tag{3.9}$$

where $\overline{\Lambda}$ denotes the first $\widehat{r}_a$ columns of $\widetilde{\Lambda}_{LS}$, $\overline{\Psi}$ denotes the first $\widehat{r}_b$ columns of $\widetilde{\Psi}_{LS}$, 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. For the first $\widehat{r}_a$ columns, the PMS estimator is identical to the unrestricted least square estimator because the columns of $\widetilde{F}_a$ are orthogonal by construction. The same argument applies to $\widehat{\Psi}_{LS}$. In finite samples, the penalization on the nonzero columns may further reduce the variance of the shrinkage estimator, but at the same time, it introduces extra bias. Whether this feature of the shrinkage estimator is preferable to the PMS estimator depends on the specific bias-variance trade-off, and we provide some simulation evidence in Section 7. For both the shrinkage estimators and the PMS estimators of the factor loadings, their corresponding factor estimators are $\widetilde{F}_a$ for the first subsample and $\widetilde{F}_b$ for the second subsample.

4 Asymptotic Theory

This section establishes the large sample properties of the shrinkage estimators $(\widehat{\Lambda}, \widehat{\Gamma})$ and shows the consistency of the proposed model selection procedure when the break date is known. We begin by stating additional assumptions.

4.1 Additional 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.

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). Assumptions P1 and P2 below are high-level conditions on the stochastic order of the preliminary estimators. They are useful in studying the asymptotic properties of the data-dependent weights ω_ℓ^λ and ω_ℓ^γ defined in (3.4). In practice, we consider the least square estimators $\widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma}_{LS}$, as well as shrinkage estimators as preliminary estimators (see Section 5 for details).

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 $\ell = 1, ..., k$. $\square$

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

Assumption P1 is imposed on any preliminary estimators of Λ^R and Γ^R. If the preliminary estimators are different from $\widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma}_{LS}$, 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 $\widetilde{\Gamma}_\ell$.

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}^{2d}. These large penalties in the second category yield shrinkage estimators that are equal to 0 w.p.a.1.

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) $N^{-1/2}C_{NT}^{-(2d+1)} = o(\alpha_{NT})$ and $N^{-1/2}C_{NT}^{-(2d+1)} = o(\beta_{NT})$. $\square$

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 upper bound is larger than the lower bound provided that d is positive. In the simulation and empirical application, we use $d = 2$. Under the theoretical guidance by Assumption T, we discuss the practical choice of α_{NT} and β_{NT} in Section 5.

4.2 Asymptotic Behavior of the PLS Estimator

We begin by defining 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 (2.9) and (2.11). The transformation matrices R_a and R_b, defined in (2.8) and (2.10), 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. As previously stated, let the subscript ℓ denote the ℓ-th column of a matrix.

Theorem 1 *Suppose Assumptions* A-D*,* 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.

To obtain results in Parts (a) and (c) of Theorem 1, only the upper bound on the convergence rate of the tuning parameters in Assumption T(i) is necessary. The lower bound in Assumption T(ii) is necessary for the superefficiency results stated in Parts (b), (d), and (e) of Theorem 1. Because the unrestricted least square estimators are special cases of the PLS estimators with zero penalties, Parts (a) and (c) of Theorem 1 apply to $\widehat{\Lambda}_\ell = \widetilde{\Lambda}_{\ell,LS}$ and $\widehat{\Gamma}_\ell = \widetilde{\Gamma}_{\ell,LS}$ for $\ell = 1, \ldots, k$, regardless of the specification of the model.

4.3 Consistent Model Selection

In the previous subsection, 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 Λ_ℓ^R and Γ_ℓ^R in Parts (a) and (c) of Theorem 1 are bounded away from zero w.p.a.1.

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

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

(b) *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) *If* $r_b = r_a$ *and* $\mathrm{rank}(\Sigma_{\Lambda\Psi}^+) > r_a$*,* $N^{-1}\Gamma^{R\prime}\Gamma^R \to \Sigma_\Gamma$ *for some* $\Sigma_\Gamma \neq 0$ *as* $N \to \infty$*;*

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

Lemma 1(a) follows from the definition of the transformation matrix R_a in (2.8) and, together with Theorem 1(a) and (c), implies that $||\widehat{\Lambda}_\ell|| \neq 0$ for $\ell \leq r_a$ and $||\widehat{\Lambda}_\ell|| = 0$ for $\ell > r_a$ w.p.a.1. Therefore, a consistent estimator of r_a is the last nonzero column in $\widehat{\Lambda}$. For the type-2 instability with $r_b > r_a$, the same arguments apply by combining Lemma 1(b) with Theorem 1(c) and (e). It follows that a consistent estimator of r_b is the last nonzero column in $\widehat{\Gamma}$.

To consistently detect a type-1 instability (i.e., $r_a = r_b$ and $\Gamma_1^0 \neq 0$), it is sufficient to show that $N^{-1}||\Gamma^R||^2$ is bounded away from 0 asymptotically. Lemma 1(c) focuses on the change of the space spanned by the factor loadings. One sufficient condition for Lemma 1(c) is Assumption ID(i), which holds if and only if Λ^0 and Ψ^0 do not span the same column space asymptotically. Lemma 1(d) suggests that an alternative sufficient condition for identification of the structural instability is Assumption ID(ii), which focuses on the change in the scaling of factor loadings.

Theorem 2 *Suppose Assumptions* A-D*,* ID*,* P1*,* P2*, and* T *hold. Then the model selection is consistent:*

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

Theorem 2 provides asymptotic consistency for 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-D*,* ID*, and* T*.*

If the preliminary estimator is a shrinkage estimator based on a rough choice of the tuning parameters, then Theorems 1 and 2 can be applied repeatedly to a multistep shrinkage estimation procedure, which in each step fine-tunes the penalty terms based on the estimation results of the previous step. Such a multistep estimation procedure is discussed in detail in Section 5.

5 Practical Guidance for Implementation

We provide a practical procedure for choosing the tuning parameters α_{NT} and β_{NT} in Section 5.1. Moreover, in Section 5.2, we propose a two-step shrinkage estimation procedure that fine-tunes the penalties in the second-stage based on the first-stage shrinkage estimation results. This procedure improves the finite-sample performance of the PLS estimator, and we show in Section 5.3 that it also leads to consistent model selection.

5.1 Choosing the Penalty Weights α_{NT} and β_{NT}

According to Theorem 2, consistent model selection requires α_{NT} and β_{NT} to converge to 0 at least as fast as $N^{-1/2}C_{NT}^{-1}$ and slower than $N^{-1/2}C_{NT}^{-(2d+1)}$. In practice, we choose α_{NT} and β_{NT} to balance these two rates. For two scaling constants κ_1 and κ_2, let

$$\alpha_{NT} = \kappa_1 N^{-1/2} C_{NT_a}^{-d-1} \text{ and } \beta_{NT} = \kappa_2 N^{-1/2} C_{NT_b}^{-d-1}, \tag{5.1}$$

where $C_{NT_a} = \min(N^{1/2}, T_a^{1/2})$ and $C_{NT_b} = \min(N^{1/2}, T_b^{1/2})$. Although C_{NT_a}, C_{NT_b}, and C_{NT} are all of the same asymptotic order, we use C_{NT_a} and C_{NT_b} rather than C_{NT} in (5.1) to improve the finite-sample accuracy. 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.

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

$$\left\| e_a(\widehat{\Lambda})\widetilde{F}_{a,\ell} + e_b(\widehat{\Lambda}+\widehat{\Gamma})\widetilde{F}_{b,\ell} \right\| < NT\alpha_{NT}\omega_\ell^\lambda/2, \tag{5.2}$$

where the residual matrices are

$$e_a(\Lambda) = X_a - \widetilde{F}_a\Lambda' \text{ and } e_b(\Lambda+\Gamma) = X_b - \widetilde{F}_b(\Lambda+\Gamma)'. \tag{5.3}$$

The inequality in (5.2) 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 (5.2) holds. Therefore, to standardize the left-hand side of (5.2), a reasonable choice of κ_1 is

$$\kappa_1 = \frac{1}{\zeta}\left\{ (NT_a)^{-1/2} \left\| e_a(\widetilde{\Lambda}) \right\| + (NT_b)^{-1/2} \left\| e_b(\widetilde{\Lambda}+\widetilde{\Gamma}) \right\| \right\}, \tag{5.4}$$

where $\zeta = 1$ by default.

In practice, we also consider $\zeta = 2$, 4, and 6 to check the sensitivity of the estimation results to the penalty function. Although our asymptotic theory does not apply to the weak factors considered in Onatski (2012), the adjustment of ζ corresponds to the strength of the factors in finite samples. Generally speaking, a larger ζ works better on detecting a weak factor and a small break, while a smaller ζ works better for strong factors and large breaks.

If the break occurs at the end of the sample, simulation results suggest that a larger ζ, e.g., $\zeta = 4$, tends to work better than $\zeta = 1$. By similar arguments, the choice of κ_2 is

$$\kappa_2 = \frac{1}{\zeta}(NT_b)^{-1/2}\left\|e_b(\widetilde{\Lambda}+\widetilde{\Gamma})\right\|. \tag{5.5}$$

In sum, the recommended tuning parameters are:

$$\begin{aligned}\alpha_{NT} &= \frac{1}{\zeta}\left((NT_a)^{-1/2}\left\|e_a(\widetilde{\Lambda})\right\| + (NT_b)^{-1/2}\left\|e_b(\widetilde{\Lambda}+\widetilde{\Gamma})\right\|\right)N^{-1/2}C_{NT_a}^{-d-1},\\ \beta_{NT} &= \frac{1}{\zeta}\left((NT_b)^{-1/2}\left\|e_b(\widetilde{\Lambda}+\widetilde{\Gamma})\right\|\right)N^{-1/2}C_{NT_b}^{-d-1}.\end{aligned} \tag{5.6}$$

5.2 Two-Step Estimation Procedure

We recommend a two-step estimation procedure that uses as preliminary estimators in the second step the shrinkage estimators from the first step. 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 (5.6) are more accurate when $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ are based on first-step shrinkage estimators. 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) *Execute the following steps:*

1. *Construct the first-stage shrinkage estimator as follows:*

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

   (b) *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) and (5.6) with* $\widetilde{\Lambda} = \widetilde{\Lambda}^{(1)}$ *and* $\widetilde{\Gamma} = \widetilde{\Gamma}^{(1)}$.

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

   (d) *Estimate* r_a *and* r_b *following (3.5) with* $\widehat{\Lambda} = \widehat{\Lambda}^{(1)}$ *and* $\widehat{\Gamma} = \widehat{\Lambda}^{(1)}$ *and call the estimators* $\widehat{r}_a^{(1)}$ *and* $\widehat{r}_b^{(1)}$. *Let* $\overline{\Lambda}^{(1)}$ *denote the first* $\widehat{r}_a^{(1)}$ *columns of* $\widetilde{\Lambda}_{LS}$ *and* $\overline{\Psi}^{(1)}$

denote the first $\widehat{r}_b^{(1)}$ *columns of* $\widetilde{\Psi}_{LS}$. *Following the definition of the PMS estimator in (3.9),*

$$\widehat{\Lambda}_{PMS}^{(1)} = \left(\overline{\Lambda}^{(1)},\, 0_{\Lambda^{(1)}}\right) \in R^{N\times k} \text{ and } \widehat{\Psi}_{PMS}^{(1)} = \left(\overline{\Psi}^{(1)},\, 0_{\Psi^{(1)}}\right) \in R^{N\times k}, \tag{5.7}$$

where $0_{\Lambda^{(1)}}$ *is a* $N \times (k - \widehat{r}_a^{(1)})$ *zero matrix and* $0_{\Psi^{(1)}}$ *is a* $N \times (k - \widehat{r}_b^{(1)})$ *zero matrix.*

(e) *If* $\widehat{r}_b^{(1)} = \widehat{r}_a^{(1)}$, *transform the columns of* $\overline{\Psi}^{(1)}$ *as follows: Let* $\overline{\Lambda}^{(1)\prime}\overline{\Psi}^{(1)} = UDV'$ *be the singular value decomposition of* $\overline{\Lambda}^{(1)\prime}\overline{\Psi}^{(1)} = UDV'$. *Define the transformed factor loading as*

$$\overline{\Psi}_R^{(1)} = \overline{\Psi}^{(1)} Q, \tag{5.8}$$

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

$$\widehat{\Psi}_{PMS-R}^{(1)} = \left(\overline{\Psi}_R^{(1)},\, 0_{\Psi^{(1)}}\right) \in R^{N\times k}, \tag{5.9}$$

which is a modification of $\widehat{\Psi}_{PMS}^{(1)}$ *with* $\overline{\Psi}^{(1)}$ *replaced by* $\overline{\Psi}_R^{(1)}$.

2. *Construct the second-stage shrinkage estimator as follows:*

   (a) *Let*

$$\widetilde{\Lambda}^{(2)} = \widehat{\Lambda}_{PMS}^{(1)},\ \ \widetilde{\Psi}^{(2)} = \begin{cases} \widehat{\Psi}_{PMS-R}^{(1)} & \text{if } \widehat{r}_b^{(1)} = \widehat{r}_a^{(1)} \\ \widehat{\Psi}_{PMS}^{(1)} & \text{if } \widehat{r}_b^{(1)} > \widehat{r}_a^{(1)} \end{cases},\ \ \widetilde{\Gamma}^{(2)} = \widetilde{\Psi}^{(2)} - \widetilde{\Lambda}^{(2)} \tag{5.10}$$

   and calculate ω_ℓ^λ, ω_ℓ^γ, α_{NT}, *and* β_{NT} *following (3.4) and (5.6) with* $\widetilde{\Lambda} = \widetilde{\Lambda}^{(2)}$ *and* $\widetilde{\Gamma} = \widetilde{\Gamma}^{(2)}$.

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

   (c) *Estimate* r_a, r_b, *and* $\mathcal{S}_0$ *following (3.5)-(3.7) with* $\widehat{\Lambda} = \widehat{\Lambda}^{(2)}$ *and* $\widehat{\Gamma} = \widehat{\Lambda}^{(2)}$.

The preliminary estimators in the second step are based on the PMS estimators of the first step. The rotation in Step 1(e) 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 $\Gamma^0 = 0$. 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 $\overline{\Psi}^{(1)}Q$ (see

Section 2.3). The solution is $Q = VU'$ (see Schönemann (1966)), where U and V are obtained from the singular value decomposition $\overline{\Lambda}^{(1)'}\overline{\Psi}^{(1)} = UDV'$.

In the remainder of this subsection, 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.

5.3 Large Sample Behavior of Two-Step Estimation Procedure

Under Assumption ID, Lemma 1(c) and Lemma 1(d) imply that, in the presence of a type-1 change, there exists a set of columns such that

$$\mathcal{Z} = \{\ell : N^{-1}||\Gamma_\ell^R||^2 = N^{-1}||\Psi_\ell^R - \Lambda_\ell^R||^2 \geq C\}. \tag{5.11}$$

The columns in the set $\mathcal{Z}$ are crucial for the identification of a type-1 instability. We now state 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 model selection consistency established in Theorem 2 can be extended to the two-step estimation procedure, 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-D*,* ID*,* R*, and* T*.*

Recall that Assumptions A-D, ID, and R are primitive conditions on the factor models. The recommended tuning parameters α_{NT} and β_{NT} defined in (5.6) are constructed to ensure that Assumption T is satisfied. As a result, one can continue to use the recommended formulas of α_{NT} and β_{NT} throughout the two-step estimation procedure.

6 Model Selection with Unknown Break Date

We now extend the model selection procedure to study a potential one-time abrupt break at an unknown time T_0, allowing for both type-1 and type-2 instabilities. 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 break date. We assume that the exact value of π_0 is unknown, but it is known that $\pi_0 \in \Pi$, where Π is some closed subset in the interior of $[0,1]$.

For any $\pi \in \Pi$, 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 $T_b = T - T_a$. In Section 6.1, we study the consequences of misspecifying the break date and present an identification condition for the unknown break date. To account for the unknown break date, we slightly modify the penalty terms of the PLS estimator in Section 6.2. Finally, we show in Section 6.3 that the consistency result of Theorem 2 extends to the case with an unknown break date.

6.1 Identification of the Change Point

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 $\pi \neq \pi_0$. In a nutshell, if the break date is misspecified, then one of the two subsamples contains one or more additional factors. Thus, the break date can be identified by minimizing the number of estimated pre- and post-break factors by varying π.

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

$$\begin{aligned} X_a(\pi) &= F_a(\pi)\Lambda^{0\prime} + e_a(\pi), \text{ where} \\ F_a(\pi) &= \left(F_1^0, \ldots, F_{T_a}^0\right)' \in R^{T_a \times r_a}, \\ e_a(\pi) &= (e_1, \ldots, e_{T_a})' \in R^{T_a \times N}. \end{aligned} \tag{6.1}$$

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

$$
\begin{aligned}
X_b(\pi) &= F_a^+(\pi)\Lambda^{0\prime} + F_b(\pi)\Psi^{0\prime} + e_b(\pi), \text{ where} \\
F_a^+(\pi) &= (F^0_{T_a+1}, ..., F^0_{T_0}, 0_{r_a\times(T-T_0)})' \in R^{T_b\times r_a}, \\
F_b(\pi) &= (0_{r_b\times(T_0-T_a)}, \overline{F}^0_{T_0+1}, ..., \overline{F}^0_T)' \in R^{T_b\times r_b}, \\
e_b(\pi) &= (e_{T_a+1}, ..., e_T)' \in R^{T_b\times N}.
\end{aligned} \tag{6.2}
$$

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 (6.1) and (6.2) reduce to (2.2) and (2.3), respectively, if the break date is known.

We now replace Assumptions A and C with Assumptions A* and C* such that the weak dependence and stationarity hold for any subsamples considered.

Assumption A*. $\mathbb{E}\,\|F_t^0\|^4 \le C$, $\mathbb{E}||\overline{F}^0_t||^4 \le C$ 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 $\pi \le \pi_0$ and $T^{-1}\sum_{t=\lfloor T\pi\rfloor+1}^{T} \overline{F}^0_t\overline{F}^{0\prime}_t = (1-\pi)\Sigma_{\overline{F}} + O_p(T^{-1/2})$ for $\pi \ge \pi_0$, where both $O_p(T^{-1/2})$ terms are uniform over $\pi \in \Pi$. □

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 $\pi \in \Pi$. □

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 nonvanishing eigenvalues of $(NT)^{-1}X_a(\pi)'X_a(\pi)$ and $(NT)^{-1}X_b(\pi)'X_b(\pi)$, respectively, as $N, T \to \infty$. Under Assumptions A*, B, C*, and D, the DGP in (6.1) is a factor model with r_a factors (i.e., $r_a(\pi) = r_a$ for $\pi \le \pi_0$). To study the number of factors in (6.2) for the second subsample, note that

$$
\begin{aligned}
&T^{-1}(F_a^+(\pi), F_b(\pi))'(F_a^+(\pi), F_b(\pi)) \to_p \Sigma_F^+(\pi), \\
&N^{-1}\left(\Lambda^0, \Psi^0\right)'\left(\Lambda^0, \Psi^0\right) \to_p \Sigma^+_{\Lambda\Psi}, \text{ where} \\
&\Sigma_F^+(\pi) = \begin{bmatrix} (\pi_0-\pi)\Sigma_F & 0_{r_a\times r_b} \\ 0_{r_b\times r_a} & (1-\pi_0)\Sigma_{\overline{F}} \end{bmatrix}
\end{aligned} \tag{6.3}
$$

and $\Sigma^+_{\Lambda\Psi}$ is defined in (2.6). 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., $r_b(\pi) = rank(\Sigma^+_{\Lambda\Psi})$ for $\pi < \pi_0$). If $\Lambda^0 = \Psi^0$, we know that $rank(\Sigma^+_{\Lambda\Psi}) = r_b$. If, on the other hand, the column spaces generated by Ψ^0 and Λ^0 do not overlap, we have $rank(\Sigma^+_{\Lambda\Psi}) = r_a + r_b$. Typically, we would expect the column spaces generated by Ψ^0 and Λ^0 to be overlapping but non-nested, which means $rank(\Sigma^+_{\Lambda\Psi}) > r_b \geq r_a$ and hence $r_b(\pi) > r_b$ for $\pi < \pi_0$.

Now consider $\pi > \pi_0$. For the first subsample, we have $r_a(\pi) = rank(\Sigma^+_{\Lambda\Psi})$, which implies $r_a(\pi) \geq r_b \geq r_a$, while for the second subsample, we simply get $r_b(\pi) = r_b$. The following lemma provides a summary:

Lemma 2 *Suppose that Assumptions* A*, B, C*, *and* D *hold. Then,*

$$r_a(\pi) = \begin{cases} r_a & \pi \leq \pi_0 \\ \operatorname{rank}(\Sigma^+_{\Lambda\Psi}) & \pi > \pi_0 \end{cases} \quad \textit{and } r_b(\pi) = \begin{cases} \operatorname{rank}(\Sigma^+_{\Lambda\Psi}) & \pi < \pi_0 \\ r_b & \pi \geq \pi_0 \end{cases}, \tag{6.4}$$

where $\operatorname{rank}(\Sigma^+_{\Lambda\Psi}) \geq r_b \geq r_a$.

It follows from Lemma 2 that

$$r_a(\pi) + r_b(\pi) = \begin{cases} r_a + rank(\Sigma^+_{\Lambda\Psi}) & \pi < \pi_0 \\ r_a + r_b & \pi = \pi_0 \\ r_b + rank(\Sigma^+_{\Lambda\Psi}) & \pi > \pi_0 \end{cases}. \tag{6.5}$$

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 $r_a(\pi) + r_b(\pi)$ achieve the smallest value $r_a + r_b$ as

$$\mathcal{D} = \{\pi \in \Pi : r_a(\pi) + r_b(\pi) = r_a + r_b\}. \tag{6.6}$$

By definition, we know that $\pi_0 \in \mathcal{D}$ and hence $\mathcal{D}$ is a well-defined nonempty set. If $rank(\Sigma^+_{\Lambda\Psi}) > r_b$ (which holds if the column space generated by Λ^0 is not contained by that generated by Ψ^0 asymptotically), we can deduce that π_0 is the unique minimizer. The result is summarized in the following corollary:

Corollary 3 *Suppose that Assumptions* A*, B, C*, *and* D *hold.*

(a) *If* $\operatorname{rank}(\Sigma^+_{\Lambda\Psi}) > r_b$, *then* $\pi_0 = \mathcal{D}$;

(b) *If* $\operatorname{rank}(\Sigma^+_{\Lambda\Psi}) = r_b > r_a$, *then* $\pi_0 = \max\{\pi : \pi \in \mathcal{D}\}$.

If the structural break is due to the type-2 instability, Corollary 3 shows that the break date π_0 can always be identified by the maximum value in the set $\mathcal{D}$. π_0 can still be identified under type-1 instability, provided that Assumption ID(i) holds. In that case, we have

$$rank(\Sigma^+_{\Lambda\Psi}) > r_a = r_b, \tag{6.7}$$

which combined with Corollary 3(a) implies that $\{\pi_0\} = \mathcal{D}$. The only case in which $\mathcal{D}$ does not identify π_0 is if $r_a = r_b$ and Λ^0 and Ψ^0 span the same column spaces. However, this type-1 instability itself can still be identified as long as Assumption ID(ii) holds.

6.2 Estimation with Unknown Break Date

For any $\pi \in \Pi$, 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 $\widetilde{F}_b \in R^{T_b \times k}$ be the orthonormalized left eigenvectors of $(NT_b)^{-1}X_b(\pi)X_b(\pi)'$ associated with its first k largest eigenvalues. We assume $k \geq r_a + r_b$, the largest possible number of factors in any subsample. 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 Section 5 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 our empirical application, 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 improve the finite sample performance of the PLS estimation with an unknown break date, we propose an averaging penalty. Based on this averaging penalty, the shrinkage estimator depends on π only through the least square criterion function. Uniform convergence of the least square criterion function over $\pi \in \Pi$ follows from Assumptions A* and C*. In the remainder of this subsection, we describe the construction of the averaging penalty.

The shrinkage estimator with an averaging penalty is obtained by minimizing a PLS criterion function indexed by π:

$$(\widehat{\Lambda}(\pi), \widehat{\Gamma}(\pi)) = \underset{\Lambda \in R^{N\times k}, \Gamma \in R^{N \times k}}{\arg\min} \left[M(\Lambda, \Gamma; \pi) + P_1^*(\Lambda) + P_2^*(\Gamma)\right], \tag{6.8}$$

where

$$M(\Lambda, \Gamma; \pi) = (NT)^{-1}\left[\left\|X_a(\pi) - \widetilde{F}_a(\pi)\Lambda'\right\|^2 + \left\|X_b(\pi) - \widetilde{F}_b(\pi)(\Lambda + \Gamma)'\right\|^2\right]. \tag{6.9}$$

The averaging penalty functions $P_1^*(\Lambda)$ and $P_2^*(\Lambda)$ are

$$P_1^*(\Lambda) = \sum_{\ell=1}^{k} \mathbb{E}_\xi[\alpha_{NT}(\xi)\omega_\ell^{\lambda*}(\xi)]\,\|\Lambda_\ell\|\,,$$
$$P_2^*(\Gamma) = \sum_{\ell=1}^{k} \mathbb{E}_\xi[\beta_{NT}(\xi)\omega_\ell^{\gamma*}(\xi)]\,\|\Gamma_\ell\|\,, \tag{6.10}$$

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 (6.10) is replaced by an average. The tuning parameters $\alpha_{NT}(\pi)$ and $\beta_{NT}(\pi)$ are

$$\alpha_{NT}(\pi) = \kappa_1(\pi)N^{-1/2}C_{NT_a}^{-d-1} \text{ and } \beta_{NT}(\pi) = \kappa_2(\pi)N^{-1/2}C_{NT_b}^{-d-1}, \tag{6.11}$$

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 (5.1). In practice, we can choose $\kappa_1(\pi)$ and $\kappa_2(\pi)$ as in (5.6) but with $\widetilde{\Lambda}$ and $\widetilde{\Gamma}$ replaced by $\widetilde{\Lambda}(\pi)$ and $\widetilde{\Gamma}(\pi)$, respectively.

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

$$\omega_\ell^{\lambda*}(\pi) = \left(N^{-1}||\widetilde{\Lambda}_\ell(\pi)||^2\mathcal{I}_{\{\widetilde{\Lambda}_\ell(\pi)\neq 0_{N\times 1}\}} + N^{-1}||\widetilde{\Lambda}_{\ell,LS}(\pi)||^2\mathcal{I}_{\{\widetilde{\Lambda}_\ell(\pi)=0_{N\times 1}\}}\right)^{-d},$$
$$\omega_\ell^{\gamma*}(\pi) = \left(N^{-1}\min\{||\widetilde{\Gamma}_\ell(\pi)||^2, ||\widetilde{\Psi}_\ell(\pi)||^2\}\mathcal{I}_{\{\widetilde{\Gamma}_\ell(\pi)\neq 0_{N\times 1}\}}\right)^{-d}$$
$$+\left(N^{-1}\min\{||\widetilde{\Gamma}_{\ell,LS}(\pi)||^2, ||\widetilde{\Psi}_{\ell,LS}(\pi)||^2\}\mathcal{I}_{\{\widetilde{\Gamma}_\ell(\pi)=0_{N\times 1}\}}\right)^{-d}. \tag{6.12}$$

Comparing the weights in (6.12) 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 modified weights can also be used if the break date is known, because $N^{-1}||\widetilde{\Gamma}_\ell(\pi_0)||^2$ and $N^{-1}||\widetilde{\Psi}_\ell(\pi_0)||^2$ are either of the same order of magnitude or the former is smaller than the latter.

[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.

6.3 Model Selection with Unknown Break Date

If the break date is unknown, the model selection for $\mathcal{M}_0$ is based on the zero and nonzero columns in $\widehat{\Lambda}(\pi)$ and $\widehat{\Gamma}(\pi)$ for all $\pi \in \Pi$. First, we generalize the estimators $\widehat{r}_a$ and $\widehat{r}_b$ in (3.5) to

$$\begin{aligned} \widehat{r}_a(\pi) &= \min \mathcal{J}_a(\pi), \text{ where } \mathcal{J}_a(\pi) = \left\{ j : ||\widehat{\Lambda}_\ell(\pi)||^2 = 0 \text{ for all } \ell > j \right\}, \\ \widehat{r}_b(\pi) &= \max\left(\min \mathcal{J}_b, \widehat{r}_a(\pi)\right), \text{ where } \mathcal{J}_b(\pi) = \left\{ j : ||\widehat{\Gamma}_\ell(\pi)||^2 = 0 \text{ for all } \ell > j \right\}. \end{aligned} \quad (6.13)$$

With an unknown break date, the estimators of r_a and r_b are

$$\widehat{r}_a^* = \min_{\pi \in \Pi} \widehat{r}_a(\pi) \text{ and } \widehat{r}_b^* = \min_{\pi \in \Pi} \widehat{r}_b(\pi). \quad (6.14)$$

The estimator of $\mathcal{S}_0$ is

$$\widehat{\mathcal{S}}^* = \begin{cases} 0 & \text{if } \sup_{\pi \in \Pi} ||\widehat{\Gamma}(\pi)|| = 0 \\ 1 & \text{otherwise} \end{cases}. \quad (6.15)$$

The selected model is

$$\widehat{\mathcal{M}}^* = (\widehat{r}_a^*, \widehat{r}_b^*, \widehat{\mathcal{S}}^*). \quad (6.16)$$

As in the known-break-date case, we consider a two-step procedure for model selection. Follow the steps in Section 5.2 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 (6.8); and replacing the estimators $\widehat{r}_a$ and $\widehat{r}_b$ in (3.5) with those in (6.14). Note that the first-step estimators $\widehat{r}_a^{(1)}$ and $\widehat{r}_b^{(1)}$ do not vary with π following the definition in (6.14). 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 $\widehat{\Gamma}^{(2)}(\pi)$.

To show the model selection consistency for the two-step PLS estimator described in the previous subsection, 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

$$\begin{aligned} X_a(\pi) &= F_a^R(\pi)\Lambda^R(\pi)' + e_a(\pi), \\ X_b(\pi) &= F_b^R(\pi)\Psi^R(\pi)' + e_b(\pi), \end{aligned} \quad (6.17)$$

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

Assumption R.*. (i) If $r_a = r_b$, then $\inf_{\pi \in \Pi} \inf_{\|w\|=1} N^{-1} \|\Psi^R(\pi) w - \Lambda_\ell^R(\pi)\|^2 \geq C$ 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, these assumptions are not restrictive because $\Lambda_\ell^R(\pi)$ and $\Psi_\ell^R(\pi)$ are some specific matrices transformed from the structural factors. Assumptions R and R* are compatible in applications where the loadings of some structural factors change and some do not.

Theorem 3 *Suppose that Assumptions* A*, B, C*, D, ID, *and* R* *hold. Then the model selection with an unknown break date is consistent:*

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

The identification result in Section 6.1 is used constructively in the proof of Theorem 3. Among all the models indexed by π, only the one with the smallest number of factors is not over-penalized by the averaging penalty. Since the model with the true break date has the smallest number of factors by Lemma 2, the PLS estimator with the averaging penalty can consistently select the true model without knowing the break date.

The proof strategy of Theorem 3 is different from that of Theorem 2 due to the averaging penalty. 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 that $\Pr(\min_{\pi \in \Pi} \widehat{r}_b(\pi) = r_b) \to 1$. 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.

The averaging penalty enables consistent estimation of r_a and r_b but does not yield consistent estimation of $\widehat{r}_a(\pi)$ and $\widehat{r}_b(\pi)$ for $\pi \neq \pi_0$. Therefore, the resulting shrinkage estimation does not simultaneously produce a consistent estimator of the break date. In practice, researchers tend to have a conjecture for the break date, which is denoted by π_c, and used to center the set Π in the previous subsection. The shrinkage estimator can be used to verify whether a conjecture break date π_c is a consistent estimator of π_0. If the data provide evidence against this conjecture, π_c can be revised accordingly. We denote the revised break date by π_{rc}, and it is defined as follows: Let $\widehat{\mathcal{D}}$ denote the set of $\pi \in \Pi$ at which $\widehat{r}_a(\pi) + \widehat{r}_b(\pi)$ achieves its minimum, which typically is an interval in finite samples. Corollary 3 suggests that π_c is a reasonable conjecture only if $\pi_c \in \widehat{\mathcal{D}}$. Hence, if $\pi_c \in \widehat{\mathcal{D}}$, we do not revise the conjectured break date and we define $\pi_{rc} = \pi_c$. On the other hand, if $\pi_c \notin \widehat{\mathcal{D}}$, we choose $\pi_{rc} \in \widehat{\mathcal{D}}$ such that $|\pi_{rc} - \pi_c| = \min_{\pi \in \widehat{\mathcal{D}}} |\pi - \pi_c|$. However, this procedure does not ensure that π_{rc} is a consistent estimator of the break date.

Once the number of factors r_a and r_b have been consistently estimated by the shrinkage estimator, the classical least squares method in Bai (1997) can be applied to obtain a consistent estimator of the break date. While without knowing the numbers of the pre- and post-break factors, one can not consistently estimate the break date using the least squares criterion, our model selection procedure provides a bridge linking the break-date estimation with observed regressors with break-date estimation in latent factor models.

7 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 7.1 describes the DGPs used in the experiment. Section 7.2 discusses various estimators of the empirical factor models. The simulation results are presented in Section 7.3.

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

$$
\begin{aligned}
\text{Pre-break:}\quad & X_{it} = \lambda_i' F_t + e_{it}, \quad F_{t,\ell} = \rho_a F_{t-1,\ell} + \eta_a u_{t,\ell}, \\
& t = 1, \ldots, \lfloor T\pi_0 \rfloor, \quad \ell = 1, \ldots, r_a, \\
\text{Post-break:}\quad & X_{it} = \psi_i' \overline{F}_t + e_{it}, \quad \overline{F}_{t,\ell} = \rho_b \overline{F}_{t-1,\ell} + \eta_b u_{t,\ell}, \\
& t = \lfloor T\pi_0 \rfloor + 1, \ldots, T, \quad \ell = 1, \ldots, r_b,
\end{aligned}
\tag{7.1}
$$

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

$$
e_{it} = \alpha e_{it-1} + v_{it},\ v_t = (v_{1t}, ..., v_{Nt})' \sim N(0, \Omega), \tag{7.2}
$$

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 with the standard normal distribution. 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 $r_b > r_a$, $\overline{F}_{T_0} = (F_{T_0}', F_{T_0}^{*\prime})'$, where each element of $F_{T_0}^*$ is drawn independently from the distribution of $F_{t,\ell}$. The parameters $\{N, T, \pi_0, r_a, r_b, \rho_a, \rho_b, \eta_a, \eta_b, \alpha, \beta\}$ 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.[6]

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.

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

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.

7.2 Estimators

We have described the principal component estimators of the factors in the previous sections. The time series variables are normalized to have zero means and unit variances before estimating these factors.[7] We set the maximum number of factors $k = 8$ and standardize the data before getting the principal component estimators. For the data-dependent weight, we set $d = 2$.

For experiments with known break dates, model selection is based on the two-step PLS estimator following the algorithm described in Section 5. To investigate the model selection accuracy, we report the probability that the "true" model $\mathcal{M}_0 = (r_a, r_b, \mathcal{S}_0)$ is selected, the probabilities of $\widehat{r}_a = r_a$, $r_a - 1$ and $r_a + 1$, and the probabilities of $\widehat{r}_b = r_b$, $r_b - 1$ and $r_b + 1$, respectively. In addition, we report MSEs of four different estimators. The first two are the PMS estimator and the PLS estimator described at the end of Section 3. Both of them switch from a full-sample estimation to a subsample estimation only if a break is detected and the number of factors before and after the break are obtained by the shrinkage estimation. In contrast, the third estimator always uses the full sample, and the fourth estimator always uses the post-break subsample. The last two estimators are standard principal component estimators, where the number of factors is selected by the IC_{p2} information criterion of Bai and Ng (2002). The third estimator coincides with the infeasible benchmark estimator when there is no structural instability, and the fourth estimator coincides with the infeasible benchmark estimator when there is a large change. For the convenience of comparison, the MSE of the first estimator is normalized to be 1.

[7]Without normalization, the idiosyncratic errors of each series have the same variance. When standardizing the variance of all series, those with low regression R^2 receive more weight. Thus, in the present simulation setup, the procedure performs much better without normalization in the heterogeneous R^2 case.

For simulations in which the break date is not assumed to be known, the model selection is based on the procedure described in Section 6.3, where Π is approximated by a discrete set Π_d. To make them similar to the empirical example investigated in Section 8, we conduct simulations with the true break date at $\pi_0 = 0.8$, and the regression R^2 is calibrated. 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 conjectured break date π_c. We consider both the correct specification case in which $\pi_c = \pi_0 = 0.8$ and the misspecification case in which $\pi_c = 0.78$, which is half a year previous to the real break date in the application. We report model selection probabilities and the MSEs of the four estimators discussed above. To define a post-break subsample for the first two estimators, the revised conjectured break date π_{rc} is used because π_0 is unknown. However, to define the post-break subsample for the fourth estimator, we continue to use π_0 because this estimator serves as an infeasible benchmark when there is a large break. For all four estimators, the reported MSEs are based on the subsample from the upper end of Π_d to the end of the full sample. This subsample ensures the availability of the post-break subsample estimator, no matter where the break date is specified.

7.3 Monte Carlo Results

The Monte Carlo results are summarized in four tables. Tables 1 and 2 present results when the break date is known and the regression R^2 is homogenous across series. The break date is in the middle of the sample in Table 1 ($\pi_0 = 0.5$) and at the end of the sample in Table 2 ($\pi_0 = 0.8$). Table 3 also assumes the break date is known, but the regression R^2 is calibrated from the data set in the empirical application and thus is heterogeneous across series. Table 4 allows the break date to be unknown, and the regression R^2 is calibrated. When the regression R^2 is calibrated, we only conduct simulations with the true break date at the end of the sample, because the potential break occurs around $\pi = 0.8$ in the data set of our empirical application. Each table contains three panels, corresponding to no instability, type-1 instability, and type-2 instability, respectively. For a type-1 instability, we consider $\mathbf{w} = 0.2$, 0.5 and 1 in the DGPs for the scenarios of small, medium, and large changes in the factor loadings, respectively. 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. For each DGP,

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

DGP Configuration						$\widehat{r}_a - r_a$			$\widehat{r}_b - r_b$			MSE			
r_a	r_b	$\mathbf{w}$	N	T	$\widehat{\mathcal{M}}$	0	-1	1	0	-1	1	PMS	PLS	Full	Sub
Panel A. No Change															
3	3		100	100	0.77	(0.79	0.21	0.00)	(0.96	0.04	0.00)	1.00	1.07	0.83	1.35
3	3		150	150	0.99	(0.99	0.01	0.00)	(1.00	0.00	0.00)	1.00	1.01	0.99	1.59
3	3		200	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.01	1.00	1.60
Panel B. Type-1 Change															
3	3	0.2	100	100	0.12	(0.88	0.12	0.00)	(0.94	0.06	0.00)	1.00	1.17	0.84	1.34
3	3	0.2	150	150	0.13	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.30	0.95	1.30
3	3	0.2	200	200	0.13	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.42	0.97	1.20
3	3	0.5	100	100	0.90	(0.90	0.10	0.00)	(0.94	0.06	0.00)	1.00	1.11	1.42	1.15
3	3	0.5	150	150	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.19	2.09	1.01
3	3	0.5	200	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.15	2.65	1.00
3	3	1.0	100	100	0.61	(0.61	0.38	0.00)	(0.99	0.01	0.00)	1.00	0.99	1.47	1.00
3	3	1.0	150	150	0.97	(0.97	0.03	0.00)	(1.00	0.00	0.00)	1.00	0.97	1.31	1.00
3	3	1.0	200	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.98	1.30	1.00
Panel C. Type-2 Change															
1	2		100	100	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.98	1.15	1.00
1	2		150	150	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.98	1.13	1.00
1	2		200	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.98	1.13	1.00
3	4		100	100	0.23	(0.61	0.39	0.00)	(0.41	0.58	0.00)	1.00	1.11	1.39	0.81
3	4		150	150	0.90	(0.97	0.03	0.00)	(0.93	0.07	0.00)	1.00	1.22	1.24	0.95
3	4		200	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.13	1.23	1.00

Notes: Parameters $\alpha = \beta = 0.2, \rho_a = \rho_b = 0.5, \eta_a = \eta_b = 1, \zeta = 1$. The first five columns are parameters in the DGPs. The next column is the probability that the true model is selected. The next six columns are the probabilities $\widehat{r}_a = r_a, r_a - 1, r_a + 1$, and $\widehat{r}_b = r_b, r_b - 1, r_b + 1$. The last four columns are MSE for the PMS estimator, the PLS estimator, the full-sample estimator, and the post-break subsample estimator.

we report the model selection results and the MSEs of the four estimators introduced above. All results are based on averages over 5,000 simulation repetitions.

Tables S-1 to S-4 in the Appendix contain supplemental results that serve as benchmarks or for robustness checks. Table S-1 is similar to Table 1, but with i.i.d. idiosyncratic errors. Table S-2 is similar to Table 2, but with a different constant ζ that is associated with larger penalty. Table S-3 is similar to Tables 1 and 2, but the break is in the factor dynamics instead of the loadings, which confirms our early discussion below Assumption ID. Table S-4 is similar to Table 4, but the conjectured break date is equal to the true break date instead

Table 2: Known Break Date, Homogeneous $R^2, \pi_0 = 0.8$

DGP Configuration						$\widehat{r}_a - r_a$			$\widehat{r}_b - r_b$			MSE			
r_a	r_b	$\mathbf{w}$	N	T	$\widehat{\mathcal{M}}$	0	−1	1	0	−1	1	PMS	PLS	Full	Sub
Panel A. No Change															
3	3		100	200	0.99	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.08	0.98	2.75
3	3		150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.02	1.00	2.72
3	3		200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.02	1.00	2.71
Panel B. Type 1 Change															
3	3	0.2	100	200	0.12	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.52	0.88	1.84
3	3	0.2	150	300	0.11	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.75	0.93	1.46
3	3	0.2	200	400	0.13	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	2.04	0.96	1.24
3	3	0.5	100	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.20	1.68	1.18
3	3	0.5	150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.15	2.57	1.04
3	3	0.5	200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.09	3.46	1.00
3	3	1.0	100	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.95	2.83	1.01
3	3	1.0	150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.97	2.03	1.00
3	3	1.0	200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.98	1.39	1.00
Panel C. Type 2 Change															
1	2		100	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.96	1.68	1.00
1	2		150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.97	1.18	1.00
1	2		200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.98	1.14	1.00
3	4		100	200	0.60	(1.00	0.00	0.00)	(0.60	0.40	0.00)	1.00	1.05	2.83	0.97
3	4		150	300	0.96	(1.00	0.00	0.00)	(0.96	0.04	0.00)	1.00	1.08	2.92	0.99
3	4		200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.03	2.24	1.00

Notes: Parameters $\alpha = \beta = 0.2, \rho_a = \rho_b = 0.5, \eta_a = \eta_b = 1, \zeta = 4$.

of being misspecified as in Table 4.

We now discuss the model selection and estimation results. First, Table 1 shows that our procedure is quite accurate in estimating r_a and r_b if the break date is in the middle of the sample, even if the number of periods in each subsample is as small as 75. To detect a type-1 instability, the method works well except when the magnitude of the break is as small as $\mathbf{w} = 0.2$. However, the MSE comparison shows that, with a small change of this magnitude, the full-sample estimator yields the smallest MSE, and the PMS estimator benefits from not detecting the break. In the same spirit, when the sample size is as small as 50 in each subsample, our procedure sometimes favors a more parsimonious model by underestimating

Table 3: Known Break Date, Heterogeneous $R^2, \pi_0 = 0.8$

DGP Configuration						$\widehat{r}_a - r_a$			$\widehat{r}_b - r_b$			MSE			
r_a	r_b	$\mathbf{w}$	N	T	$\widehat{\mathcal{M}}$	0	-1	1	0	-1	1	PMS	PLS	Full	Sub
Panel A. No Change															
3	3		100	200	0.97	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	0.95	3.70
3	3		150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	3.49
3	3		200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	3.08
Panel B. Type-1 Change															
3	3	0.2	100	200	0.11	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.02	0.78	2.66
3	3	0.2	150	300	0.01	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.04	0.98	3.12
3	3	0.2	200	400	0.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.01	1.00	2.55
3	3	0.5	100	200	0.98	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.97	0.70	0.99
3	3	0.5	150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.39	1.13	1.36
3	3	0.5	200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.57	1.50	1.39
3	3	1.0	100	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.85	2.63	1.39
3	3	1.0	150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.91	4.12	1.18
3	3	1.0	200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.93	4.74	1.03
Panel C. Type-2 Change															
1	2		100	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.94	2.80	1.02
1	2		150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.96	1.60	1.00
1	2		200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.98	1.18	1.00
3	4		100	200	0.54	(1.00	0.00	0.00)	(0.54	0.46	0.00)	1.00	0.97	3.13	1.41
3	4		150	300	0.93	(1.00	0.00	0.00)	(0.93	0.07	0.00)	1.00	1.02	3.84	1.15
3	4		200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.99	3.53	1.01

Notes: Parameters $\alpha = \beta = 0.2, \rho_a = \rho_b = 0.5, \eta_a = \eta_b = 1, \zeta = 4$.

of the number of factors. Results in Table 1 are for idiosyncratic errors with both temporal and cross-sectional dependence and the default choice of the penalty with $\zeta = 1$. If i.i.d. idiosyncratic errors are considered in Table S-1, the procedure works quite well even if each subsample only contains 50 periods and the break is small.

Second, if the break date is at the end of the sample, Table 2 shows a pattern similar to that in Table 1. In particular, when the true model is not selected for a small break or a small sample, the misspecified model typically yields a smaller MSE. Comparing Table 2 with Table S-2, it is clear that $\zeta = 4$, which gives a smaller penalty than $\zeta = 1$, is preferred

Table 4: Unknown Break Date, Heterogeneous $R^2, \pi_0 = 0.8$

DGP Configuration						$\widehat{r}_a - r_a$			$\widehat{r}_b - r_b$			MSE			
r_a	r_b	$\mathbf{w}$	N	T	$\widehat{\mathcal{M}}$	0	-1	1	0	-1	1	PMS	PLS	Full	Sub
Panel A. No Change															
3	3		100	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.01	1.00	3.90
3	3		150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	3.49
3	3		200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	3.08
Panel B. Type-1 Change															
3	3	0.2	100	200	0.05	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.06	0.92	3.16
3	3	0.2	150	300	0.01	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.03	0.99	3.16
3	3	0.2	200	400	0.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.01	1.00	2.54
3	3	0.5	100	200	0.78	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.26	0.80	1.14
3	3	0.5	150	300	0.94	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.92	1.19	1.43
3	3	0.5	200	400	0.99	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	2.32	1.57	1.45
3	3	1.0	100	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.13	2.59	1.36
3	3	1.0	150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.12	4.06	1.17
3	3	1.0	200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.07	4.69	1.02
Panel C. Type-2 Change															
1	2		100	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.93	2.78	1.01
1	2		150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.95	1.58	1.00
1	2		200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.97	1.18	1.00
3	4		100	200	0.39	(1.00	0.00	0.00)	(0.54	0.46	0.00)	1.00	0.98	3.00	1.36
3	4		150	300	0.86	(1.00	0.00	0.00)	(0.93	0.07	0.00)	1.00	1.04	3.71	1.11
3	4		200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	3.50	1.00

Notes: Parameters $\alpha = \beta = 0.2, \rho_a = \rho_b = 0.5, \eta_a = \eta_b = 1, \zeta = 4$. The conjecture break date π_c is misspecified, $\pi_c = 0.78$.

if the break date is moved from the middle of the sample to the end of the sample. This is particularly important to detect a type-2 instability, because the new factors are only in the post-break subsample, which has a small sample size. If there is no structural break or if there is a type-1 instability, model selection results are more robust to the choice of ζ.

Third, Table 3 and Table 4 show that heterogeneous regression R^2 and unknown break date make the model selection procedure less accurate than that in Table 2, but it still works quite well in general. In all cases, r_a can be accurately estimated, even if the post-break subsample is as small as 40. In a factor model with type-2 instability, r_b might be underestimated if the post-break subsample is 40, but this underestimation issue is minor if the size of the post-break subsample is larger than 80. As in other cases, type-1 instability can be detected except when $\mathbf{w} = 0.2$. Table 4 and Table S-4 suggest that the results are similar with different specifications of the conjectured break date.

Finally, the PMS estimator (the first column) switches between the full-sample estimator (the third column) and the post-break subsample estimator (the fourth column) and when evaluated by its MSE, the PMS estimator may have better finite sample performances than the other two estimators in some scenarios. For example, in Panel B of Table 1 with $\mathbf{w} = 0.5$, $N = 100$, and $T = 100$, the MSEs of the full-sample estimator and the post-break subsample estimator, respectively, are 42% and 15% larger than that of the PMS estimator. The PLS estimator (second column) may have a smaller MSE than the PMS estimator. This could happen for a type-1 instability with a large $\mathbf{w}$ or a type-2 instability where the factor loadings of the new factors are large. On the other hand, when the shrinkage causes much more bias than desired, the PLS estimator can be worse than the PMS estimator.

8 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 8.1 describes the data set. Estimates of the number of pre-2007 and post-2007 factors are presented in Section 8.2. Finally, we make some identification assumptions and provide an interpretation of the estimated factors in Section 8.3.

8.1 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

Table 5: 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

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 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, which are grouped into the 12 categories listed in Table 5. The sample begins after the Great Moderation and ranges from 1985:M1 to 2013:M1 ($T = 337$).

Table 6: Model Selection, T_c is 2007:12

Penalty		Estimates			
Scaling ζ	Interval	Break T_{rc}	$\hat{r}_a$	$(\hat{r}_b - \hat{r}_a)$	$\lVert\widehat{\Gamma}_l\rVert^2$
Panel A. Known Break Date					
1	N/A	2007:M12	1	1	3.66
4	N/A	2007:M12	4	1	5.02
6	N/A	2007:M12	6	0	5.52
Panel B. Unknown Break Date					
1	$T_c \pm 6$	2007:M12	1	1	12.98
1	$T_c \pm 12$	2007:M12	1	1	15.38
4	$T_c \pm 6$	2007:M12	3	0	19.20
4	$T_c \pm 12$	2007:M3	3	0	24.31
6	$T_c \pm 6$	2007:M6	5	1	19.86
6	$T_c \pm 12$	2007:M5	5	1	25.19

Notes: The PLS estimator uses the averaging penalty functions $P_1^*(\Lambda)$ and $P_2^*(\Lambda)$ defined in (6.10) where the average is taken over the interval specified in the second column of this table.

8.2 The Number of Factors Before and After 2007:M12

The empirical analysis is based on the two-step estimation procedure described in Section 5.2. The starting point is a conjectured break date T_c = 2007:M12, which is the beginning of the Great Recession, according to the business cycle dating of the National Bureau of Economic Research (NBER). We use the extensions described in Section 6 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 we set the constant that controls the rate of decay (as a function of the sample size) of the penalty term to $d = 2$, as we did for the Monte Carlo analysis in Section 7. To document the sensitivity of the empirical results to the magnitude of the penalty terms $P_1(\Lambda)$ and $P_2(\Lambda)$ (or $P_1^*(\Lambda)$ and $P_2^*(\Lambda)$ for the case of unknown break point), we vary the scaling constant ζ that appears in the definition of the penalty weights α_{NT} and β_{NT} in (5.6).

The model selection results are summarized in Table 6 for different choices of the penalty scaling factor ζ. We distinguish the case of treating the break date as known (Panel A) from the case of an unknown break date (Panel B). Following the procedure described in

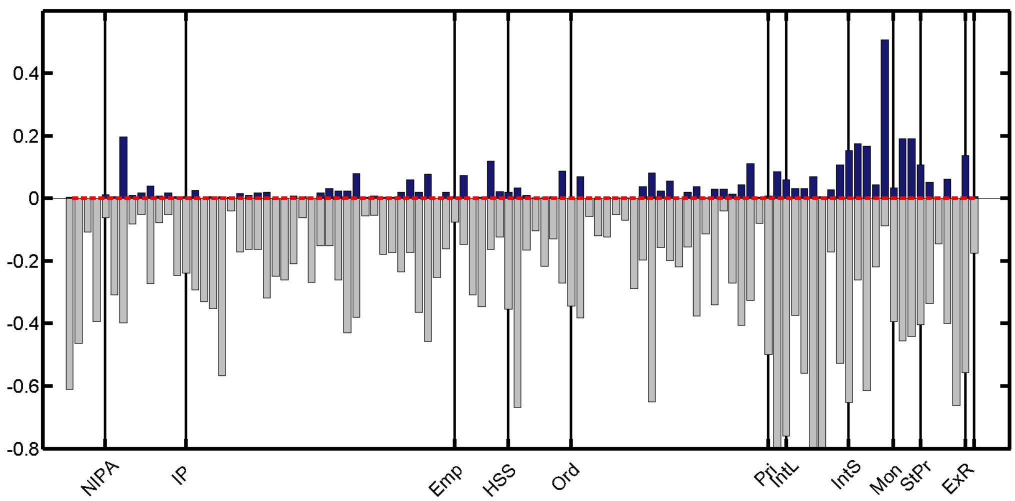
Section 6.3, we consider break dates T_a over the range $T_c \pm \bar{\tau}$, where $\bar{\tau}$ is either six or 12 months, and generate estimates of $\hat{r}_a(T_a/T)$ and $\hat{r}_b(T_b/T)$, keeping the overall sample size $T = T_a + T_b$ fixed. If $\hat{r}_a(T_a/T) + \hat{r}_b(T_b/T)$ achieves its minimum at T_c then we do not revise the conjectured break date. If the minimum is achieved elsewhere in the interval $T_c \pm \bar{\tau}$, we use the date closest to the conjectured break date at which the minimum is achieved. For $\zeta = 1$, the conjectured break date is not revised, whereas for larger values of ζ, in particular $\zeta = 6$, the break date is shifted several months backward in time to the first half of 2007.

We subsequently focus on the case of an unknown break date. The estimates of the number of factors is robust to the choice of $\bar{\tau} \in \{6, 12\}$. The overall number of factors is increasing in the scaling factor ζ, because the larger ζ the smaller the penalty for nonzero coefficients. By choosing different values for ζ, we are essentially setting different thresholds for the increase in goodness-of-fit that an additional factor must generate to justify its inclusion. For the pre-2007 sample, the number of factors ranges from $\hat{r}_a = 1$ for $\zeta = 1$ to $\hat{r}_a = 5$ for $\zeta = 6$. For comparison, we also estimated the number of pre-break factors using the Bai and Ng (2002) criteria: IC_1 and IC_2 deliver the estimate $\tilde{r}_a = 1$, whereas IC_3 generates either $\tilde{r}_a = 6$ (sample ending in 2007:M5 or 2007:M6) or $\tilde{r}_a = 7$ (sample ending in 2007:M3 or 2007:M12).

While the estimation of the overall number of factors is fairly sensitive to the choice of ζ, the estimate of the change in the number of factors, $\hat{r}_b - \hat{r}_a$ is quite stable. For $\zeta = 1$ and $\zeta = 6$, we detect a type-2 instability and find that the number of factors post-2007 has increased by one. For $\zeta = 4$, our procedure detects a type-1 instability, meaning that the loadings change but the estimated number of factors stays constant.

Using the estimates for $\zeta = 6$ and $T_c \pm 12$, we now decompose the effect of the structural change into the effect of the change in loadings on the old factors and the effect of the new factor. The decomposition is based on (2.14). As a baseline, we compute R^2 values for each individual series based on the variation explained by $F_{b,1}^{R\Omega}\Lambda^{R'} + F_{b,1}^{R\Omega}(\Psi_1^{R\Omega} - \Lambda^R)'$ (*new loadings only*). We compare the baseline R^2s to R^2s associated with $F_{b,1}^{R\Omega}\Lambda^{R'}$ (*old loadings*) and R^2s associated with all three terms in (2.14) (i.e., *new loadings and factor*). The results are plotted in Figure 1. 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 the time series categories listed in Table 5.

Figure 1: R^2 Gains from New Loadings and Factor: $(\hat{r}_1, \hat{r}_2) = (5, 6)$

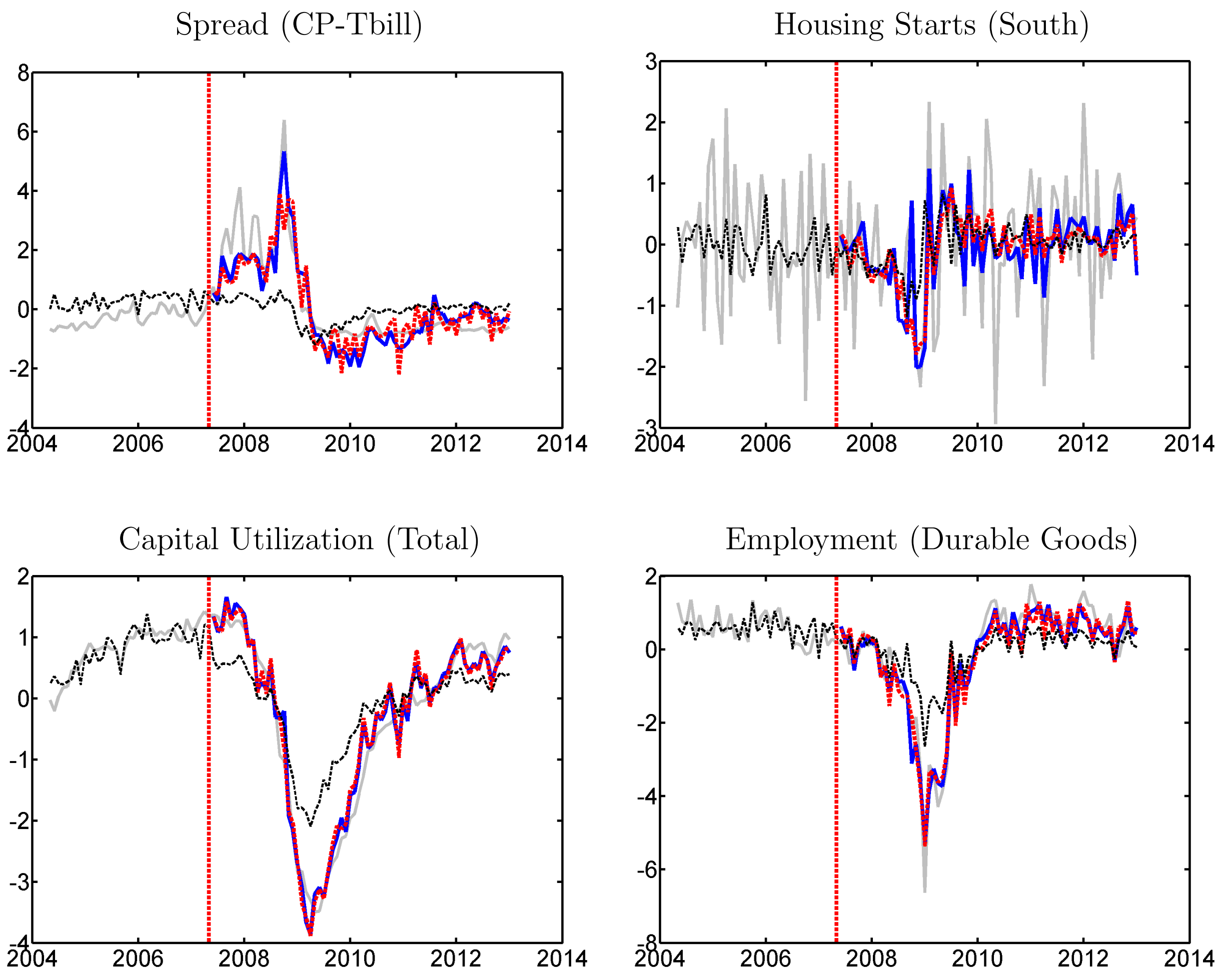


Notes: Base line case *new loadings only*. Dark bars (above zero) indicate R^2 gain due to new factor relative to *new loadings only*. Light grey bars (below zero) indicate R^2 losses from using the old loading matrices. The procedure with $\zeta = 6$ and interval $2007 : M12 \pm 12$ selects the break date 2007:M5.

Two observations stand out. First, our estimates attribute most of the structural change to a change in the loadings of the existing factors, in the sense that the contribution of $F^{R\Omega}_{b,1}(\Psi^{R\Omega}_1 - \Lambda^R)'$ to the overall R^2 exceeds the contribution of $F^{R\Omega}_{b,2}\Psi^{R\Omega\prime}_2$. This is consistent with the fact that for $\zeta = 4$, our estimate of the change in the number of factors is equal to zero. The effect of the loadings change dominates. Second, the new factor mainly affects financial variables, namely those series in the two interest rate categories (IntL, IntS), the money and credit group (Mon), and the stock price and wealth group (StPr). While there are some spillovers to the real side (i.e., industrial production (IP), employment (EMP), and housing starts (HSS) variables), these spillovers are relatively small and affect only a fraction of the series.

Figure 2 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

Figure 2: 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. Break date is 2007:M5 and $(r_1, r_2) = (5, 6)$. Our procedure with $\zeta = 6$ and interval $2007M12 \pm 12$ selects $(r_1, r_2, \text{dates}) = (5, 6, 2007\text{:M5})$.

throughout 2008. As suggested by Figure 1, 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 2.

At first glance, the results in Figure 2 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 1 and 2. If we normalize the length of the loadings before and after the break to one, then the increase in the volatility after 2007 is absorbed in an increase in the factor volatility. The ratio of pre- to post-break factor variance under this alternative normalization ranges from 1.26 to 1.93 for the five factors that were active prior to the break. 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 drastically, generally implying a stronger comovement of the series after 2007. There is also some 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. While the estimate of the total number of factors is sensitive to the scaling of the penalty terms in the objective function of the PLS estimator, the estimate of the number of new factors, which is either one or zero, is much more robust to the tuning of the penalty terms.

8.3 Interpreting the Old and New Factors

Previously, our analysis focused on determining the number of factors pre- and post 2007:M12 and the type of the structural instability. Moreover, for the detected type-2 structural changes we decomposed the overall break effect into the contribution of breaks in the loadings and the emergence of a new factor. Due to the normalization imposed on the DGP, the estimated pre- and post-break factors have identity covariance matrices and no specific economic interpretation. We will now use a nonsingular $r_a \times r_a$ matrix B ($r_b \times r_b$ matrix B) to transform the estimated pre-break factors $\tilde{F}_a$ (post-break factors $\tilde{F}_b$) into factors $\widetilde{F}_a^{\dagger}$ ($\widetilde{F}_b^{\dagger}$) that resemble the first principal components of the variable groups listed in Table 5 and therefore can be interpreted as NIPA, IP, EMP, etc.[8]

To simplify the notation, we drop the "hats" and "tildes" from matrices associated with sample estimates. Without loss of generality, we assume that the columns of X_a are arranged according to the $J = 11$ categories listed in Table 5, such that we can partition $X_a = [X_{a,1}, \ldots, X_{a,J}]$, where $X_{a,j}$ contains the series associated with the j'th category of variables, e.g., $X_{a,1}$ comprises the five NIPA series. For each $X_{a,j}$ we calculate the first principal component, denoted by the $T \times 1$ vector $x_{a,j}$, which can be interpreted as a group-specific factor. We then project each group-specific factor on the space spanned by F_a:

$$\xi_{a,j} = F_a B_j + \text{resid}, \quad j = 1, \ldots, J \tag{8.1}$$

and refer to the $T \times 1$ vector of predicted values $F_a \hat{B}_j$ as the NIPA factor for $j = 1$, the IP factor for $j = 2$, and so forth. For each regression j in (8.1), we compute the associated R^2 and then select the r_a categories that deliver the highest R^2 values. The associated coefficient estimates are $\hat{B}_{(1)}, \ldots, \hat{B}_{(r_a)}$, where each $\hat{B}_{(j)}$ is a $r_a \times 1$ vector. In turn, we define

$$B = (\hat{B}_{(1)}, \ldots, \hat{B}_{(r_a)}), \quad F_a^{\dagger} = F_a B, \quad \text{and} \quad \Lambda^{\dagger\prime} = B^{-1}\Lambda'. \tag{8.2}$$

The transformed factors are labeled according to the associated $\xi_{a,(j)}$ vectors.

For the post-break period, we have a $T \times r_b$ matrix of estimated factors F_b, which we will now transform into a set of economically interpretable factors $F_b^{\dagger} = F_b C$ using a nonsingular $r_b \times r_b$ matrix C. First, we extend the transformed pre-break factors $F_a^{\dagger}$ into the post-break period. Note that $F_a^{\dagger}$ consists of linear combinations of X_a, say, $F_a^{\dagger} = X_a \Upsilon_a^{\dagger}$. Thus, we can define $F_{b|a}^{\dagger} = X_b \Upsilon_a^{\dagger}$. The extended factors $F_{b|a}^{\dagger}$ need not fall into the space spanned by the post-break factors F_b. Thus, we project the extended factors on the space spanned by F_b using the following regression:

$$F_{b|a}^{\dagger} = F_b C_0 + \text{resid}. \tag{8.3}$$

Here the matrix C_0 is of dimension $r_b \times r_a$. If $r_b > r_a$, we proceed by computing post-break group-specific factors $\xi_{b,j}$ based on the first principal components. Focusing on the variable

[8] An alternative method of constructing interpretable factors is developed in Dobrev and Schaumburg (2013).

Table 7: Rotated Factors, Unknown Break Date, T_c is 2007:12

Penalty		Pre-Break Factors					New Post-Break
Scaling ζ	Interval	F_1	F_2	F_3	F_4	F_5	F_1^*
1	$T_c \pm 12$	Emp					IntS
6	$T_c \pm 12$	Emp	IP	IntS	ExR	Pri	Mon

groups j that were not used in the construction of the rotated pre-break factors, we estimate

$$\xi_{b,j} = F_b C_j + \text{resid} \tag{8.4}$$

and compute the R^2 associated with this regression. Let $(r_a+1), \ldots, (r_b)$ denote the indices that generate the highest R^2 values. We define the post-break transformation and the associated transformed factors and loadings as

$$C = (\hat{C}_0, \hat{C}_{(r_a+1)}, \ldots, \hat{C}_{(r_b)}), \qquad F_b^{\dagger} = F_b C, \quad \text{and} \quad \Psi^{\dagger'} = C^{-1}\Psi'. \tag{8.5}$$

Table 7 summarizes the labels for the transformed pre-break factors and the additional post-break factors for the two choices of ζ under which a new factor was detected. Since the results were identical for $T_c \pm 6$ and $T_c \pm 12$, we only report the latter. For $\zeta = 1$, we identify the pre-break factor as an employment factor. By assumption, the employment factor continues to be active from 2008:M1 onward but a second factor, namely an interest rate spread factor, emerges. This is consistent with the narrative of a financial crisis in which drastic increases in spreads coincide with substantial drops in real activity.

For $\zeta = 6$, our identification scheme labels the five pre-break factors as employment, industrial production, interest rates, exchange rates, and prices. We verified that the transformed factors track the group-specific factors well. For the categories employment, industrial production, and money and credit, the correlation with the group-specific factors is above 0.9. The new post-break factor is money and credit, which is also broadly consistent with the narrative of the Great Recession. Overall, our analysis suggests that one can construct a plausible identification scheme for the factors under which the new factor that emerged during the Great Recession can be broadly interpreted as a financial factor. This conclusion is consistent with the findings in Figure 1, in which the financial variables are associated with the largest R^2 gains from the new factor.

9 Conclusion

We develop a shrinkage estimation procedure for high-dimensional factor models that generates consistent estimates of the number of pre- and post-break factors. In situations in which the number of factors is constant throughout the sample, the procedure can consistently detect changes in the matrix of factor loadings. Our model selection procedure remains consistent even if the break date is unknown; however, it does not generate a consistent estimate of the break date itself. Nevertheless, once the number of pre- and post-break factors is known, conventional methods can be used to estimate the break date consistently. Our Monte Carlo analysis shows that the procedure has good finite sample properties. In an application to U.S. data, we show that the procedure detects an increase in the number of factors for a large macroeconomic and financial data set at the onset of the Great Recession. After imposing some identification conditions, we show that the new factor can be interpreted as a financial factor, 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.

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.

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.

——— (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.

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.

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 (2012): "Selecting the Correct Number of Factors in Approximate Factor Models: The Large Panel Case with Group Bridge Estimators," *Manuscript, North Carolina State University.*

Chen, L., J. J. Dolado, and J. Gonzalo (2011): "Detecting Big Structural Breaks in Large Factor Models," *Manuscript, Universidad Carlos III de Madrid.*

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. Swanson (2013): “Testing for Structural Stability of Factor Augmented Forecasting Models,” *Journal of Econometrics*, forthcoming.

Dobrev, D., and E. Schaumburg (2013): “Robust Forecasting by Regularization,” *Manuscript, Board of Governors of the Federal Reserve System and Federal Reserve Bank of New York*.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000): “The Generalized Dynamic Factor Model: Identification and Estimation,” *Review of Economics and Statistics*, 82(4), 540–554.

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 (2011): “Tests for Parameter Instability in Dynamic Factor Models,” *Manuscript, North Carolina State University*.

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 (2012): “The Lasso for High-Dimensional Regression with a Possible Change-Point,” *ArXiv e-prints http://arxiv.org/abs/12094875*.

Lu, X., and L. Su (2013): “Shrinkage Estimation of Dynamic Panel Data Models with Interactive Fixed Effects,” *Manuscript, Hong Kong University of Science & Technology and Singapore Management University*.

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 (2013): "Shrinkage Estimation of Regression Models with Multiple Structural Change," *Manuscript, Shanghai Jiao Tong University and Singapore Management University*.

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.

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 Supplemental Tables

Tables S-1, S-2, S-3, and S-4 provide some additional Monte Carlo results.

Tables S-5 to S-7 provide a list of variables used in the empirical application.

Table S-1: Known Break Date, Homogeneous $R^2, \pi_0 = 0.5$, i.i.d. errors

DGP Configuration						$\widehat{r}_a - r_a$			$\widehat{r}_b - r_b$			MSE			
r_a	r_b	$\mathbf{w}$	N	T	$\widehat{\mathcal{M}}$	0	-1	1	0	-1	1	PMS	PLS	Full	Sub
Panel A. No Change															
3	3		100	100	0.88	(0.88	0.12	0.00)	(0.98	0.02	0.00)	1.00	1.04	0.90	1.42
3	3		150	150	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	1.56
3	3		200	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	1.55
Panel B. Type-1 Change															
3	3	0.2	100	100	0.03	(0.94	0.06	0.00)	(0.97	0.03	0.00)	1.00	1.06	0.94	1.37
3	3	0.2	150	150	0.02	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.06	0.99	1.26
3	3	0.2	200	200	0.02	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.08	1.00	1.14
3	3	0.5	100	100	0.96	(0.96	0.04	0.00)	(0.97	0.03	0.00)	1.00	1.31	1.67	1.17
3	3	0.5	150	150	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.34	2.44	1.01
3	3	0.5	200	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.25	3.14	1.00
3	3	1.0	100	100	0.71	(0.71	0.29	0.00)	(1.00	0.00	0.00)	1.00	0.98	1.45	1.00
3	3	1.0	150	150	0.99	(0.99	0.01	0.00)	(1.00	0.00	0.00)	1.00	0.97	1.34	1.00
3	3	1.0	200	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.98	1.33	1.00
Panel C. Type-2 Change															
1	2		100	100	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.99	1.16	1.00
1	2		150	150	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.98	1.15	1.00
1	2		200	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.98	1.15	1.00
3	4		100	100	0.36	(0.71	0.29	0.00)	(0.53	0.47	0.00)	1.00	1.17	1.43	0.79
3	4		150	150	0.96	(0.99	0.01	0.00)	(0.98	0.03	0.00)	1.00	1.26	1.26	0.98
3	4		200	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.13	1.25	1.00

Notes: Parameters $\alpha = \beta = 0, \rho_a = \rho_b = 0.5, \eta_a = \eta_b = 1, \zeta = 1$.

Table S-2: KNOWN BREAK DATE, HOMOGENEOUS $R^2, \pi_0 = 0.8, \zeta = 1$

DGP Configuration						$\widehat{r}_a - r_a$			$\widehat{r}_b - r_b$			MSE			
r_a	r_b	$\mathbf{w}$	N	T	$\widehat{\mathcal{M}}$	0	-1	1	0	-1	1	PMS	PLS	Full	Sub
Panel A. No Change															
3	3		100	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	2.80
3	3		150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	2.73
3	3		200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	2.72
Panel B. Type-1 Change															
3	3	0.2	100	200	0.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	2.08
3	3	0.2	150	300	0.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	1.57
3	3	0.2	200	400	0.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	1.30
3	3	0.5	100	200	0.50	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.80	1.26	0.90
3	3	0.5	150	300	0.90	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	3.61	2.25	0.91
3	3	0.5	200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	4.49	3.41	0.99
3	3	1.0	100	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.06	2.84	1.01
3	3	1.0	150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	2.03	1.00
3	3	1.0	200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.98	1.38	1.00
Panel C. Type-2 Change															
1	2		100	200	0.97	(1.00	0.00	0.00)	(0.97	0.03	0.00)	1.00	1.27	1.61	0.95
1	2		150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.10	1.19	1.00
1	2		200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.02	1.14	1.00
3	4		100	200	0.00	(1.00	0.00	0.00)	(0.00	1.00	0.00)	1.00	1.05	2.22	0.75
3	4		150	300	0.07	(1.00	0.00	0.00)	(0.07	0.93	0.00)	1.00	1.04	1.63	0.55
3	4		200	400	0.40	(1.00	0.00	0.00)	(0.40	0.60	0.00)	1.00	1.22	1.29	0.57

Notes: Parameters $\alpha = \beta = 0.2, \rho_a = \rho_b = 0.5, \eta_a = \eta_b = 1, \zeta = 1$.

Table S-3: Known Break Date, Homogeneous R^2, Change in Factor Dynamics

DGP Configuration							$\widehat{r}_a - r_a$			$\widehat{r}_b - r_b$			MSE			
ρ_a	ρ_b	η_a	η_b	N	T	$\widehat{\mathcal{M}}$	0	-1	1	0	-1	1	PMS	PLS	Full	Sub
Panel A. $\pi_0 - 0.5$.																
0	0	1	2	100	100	0.95	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.13	0.86	1.01
0	0	1	2	150	150	0.99	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.06	0.85	1.00
0	0	1	2	200	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.02	0.85	1.00
0	0.8	1	1	100	100	0.53	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.18	0.90	1.16
0	0.8	1	1	150	150	0.62	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.22	0.88	1.12
0	0.8	1	1	200	200	0.72	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.23	0.86	1.09
0	0.8	1	0.6	100	100	0.02	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.02	0.99	1.64
0	0.8	1	0.6	150	150	0.01	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.01	1.00	1.63
0	0.8	1	0.6	200	200	0.01	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.01	0.99	1.61
Panel B. $\pi_0 = 0.8$																
0	0	1	2	100	200	0.90	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.54	0.63	1.06
0	0	1	2	150	300	0.98	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.35	0.61	1.01
0	0	1	2	200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.17	0.61	1.00
0	0.8	1	1	100	200	0.38	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.49	0.78	1.53
0	0.8	1	1	150	300	0.49	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.70	0.72	1.39
0	0.8	1	1	200	400	0.58	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.79	0.68	1.30
0	0.8	1	0.6	100	200	0.01	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.04	0.99	2.79
0	0.8	1	0.6	150	300	0.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.02	1.00	2.74
0	0.8	1	0.6	200	400	0.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.01	1.00	2.72

Notes: Parameters: $\psi_i = \lambda_i, r_a = r_b = 1, \alpha = \beta = 0, \zeta = 1$. In the last three rows of each panel, the change from (ρ_a, η_a) to (ρ_b, η_b) does not result in a change in the factor variance, and such a change cannot be identified.

Table S-4: Unknown Break Date, Heterogeneous $R^2, \pi_0 = 0.8$

DGP Configuration						$\widehat{r}_a - r_a$			$\widehat{r}_b - r_b$			MSE			
r_a	r_b	$\mathbf{w}$	N	T	$\widehat{\mathcal{M}}$	0	-1	1	0	-1	1	PMS	PLS	Full	Sub
Panel A. No Change															
3	3		100	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.01	1.01	3.91
3	3		150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	3.49
3	3		200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	3.10
Panel B. Type-1 Change															
3	3	0.2	100	200	0.02	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.03	0.97	3.29
3	3	0.2	150	300	0.01	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.02	0.99	3.15
3	3	0.2	200	400	0.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.01	1.00	2.54
3	3	0.5	100	200	0.70	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.24	0.79	1.11
3	3	0.5	150	300	0.93	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.93	1.13	1.36
3	3	0.5	200	400	0.99	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	2.35	1.50	1.39
3	3	1.0	100	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.19	2.63	1.40
3	3	1.0	150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.14	4.13	1.18
3	3	1.0	200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.09	4.71	1.03
Panel C. Type-2 Change															
1	2		100	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.94	2.78	1.01
1	2		150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.96	1.57	1.00
1	2		200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.97	1.18	1.00
3	4		100	200	0.21	(1.00	0.00	0.00)	(0.21	0.79	0.00)	1.00	0.98	2.88	1.31
3	4		150	300	0.71	(1.00	0.00	0.00)	(0.71	0.29	0.00)	1.00	1.05	3.47	1.04
3	4		200	400	0.97	(1.00	0.00	0.00)	(0.97	0.03	0.00)	1.00	1.03	3.45	1.00

Notes: Parameters $\alpha = \beta = 0.2, \rho_a = \rho_b = 0.5, \eta_a = \eta_b = 1, \zeta = 4$. The conjecture break date π_c is correctly specified.

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).

B Some Auxiliary Results

We first present a lemma on the transformation matrices R_a and R_b defined in (2.8) and (2.10) 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 $(T_0N)^{-1}X_aX_a'$ 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 $(T_1N)^{-1}X_bX_b'$ in a decreasing order. Under Assumptions A-D, Theorem 1 of Bai and Ng (2002) shows that

$$T_0^{-1}||\widetilde{F}_a^r - F_aH_a||^2 = O_p(C_{NT_0}^{-2}) \text{ and } T_1^{-1}||\widetilde{F}_b^r - F_bH_b||^2 = O_p(C_{NT_1}^{-2}), \tag{B.1}$$

where

$$H_a = \Sigma_a \frac{F_a'\widetilde{F}_a^r}{T_0}\widetilde{V}_a^{-1} \text{ and } H_b = \Sigma_b \frac{F_b'\widetilde{F}_b^r}{T_1}\widetilde{V}_b^{-1}. \tag{B.2}$$

Lemma 3 *Suppose that Assumptions* A-D *hold. Then,*

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

Proof of Lemma 3. Note that R_a is invertible w.p.a.1. Hence, we can write

$$F_a\Lambda^{0\prime} = F_aR_aR_a^{-1}\Lambda^{0\prime} = F_a^R\Lambda^{R\prime}, \text{ where } F_a^R = F_aR_a \text{ and } \Lambda^{R\prime} = R_a^{-1}\Lambda^{0\prime}. \tag{B.3}$$

The transformed factors satisfy

$$\begin{aligned}\frac{F_a^{R\prime}F_a^R}{T_0} &= V_a^{-1/2}\Upsilon_a'\Sigma_a^{1/2}\frac{F_a'F_a}{T_0}\Sigma_a^{1/2}\Upsilon_aV_a^{-1/2} \\ &= V_a^{-1/2}\left(\Upsilon_a'\Sigma_a^{1/2}\Sigma_F\Sigma_a^{1/2}\Upsilon_a\right)V_a^{-1/2} + O_p(T_0^{-1/2}) \\ &= V_a^{-1/2}\left(V_a\right)V_a^{-1/2} + O_p(T_0^{-1/2}) = I_{r_a} + O_p(T_0^{-1/2}),\end{aligned} \tag{B.4}$$

where the first equality follows from $F_a^R = F_aR_a$ and $R_a = \Sigma_a^{1/2}\Upsilon_aV_a^{-1/2}$, 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 (2.7). The transformed loadings satisfy

$$\frac{\Lambda^{R\prime}\Lambda^R}{N} = V_a^{1/2}\Upsilon_a^{-1}\Sigma_a^{-1/2}\frac{\Lambda^{0\prime}\Lambda^0}{N}\Sigma_a^{-1/2}\Upsilon_a'^{-1}V_a^{1/2} = V_a^{1/2}\Upsilon_a^{-1}\Upsilon_a'^{-1}V_a^{1/2} = V_a, \tag{B.5}$$

where the first equality follows from $\Lambda^{R\prime} = R_a^{-1}\Lambda^{0\prime}$ and $R_a = \Sigma_a^{1/2}\Upsilon_aV_a^{-1/2}$, 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

$$L_a = \frac{\Lambda^{R\prime}\Lambda^R}{N}\frac{F_a^{R\prime}\widetilde{F}_a^r}{T_0}\widetilde{V}_a^{-1}, \tag{B.6}$$

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

$$\begin{aligned} H_a &= R_a \frac{R_a^{-1}\Lambda^{0\prime}\Lambda^0 R_a^{\prime -1}}{N}\frac{R_a^\prime F_a^\prime \widetilde{F}_a^r}{T_0}\widetilde{V}_a^{-1} \\ &= R_a \frac{\Lambda^{R\prime}\Lambda^R}{N}\frac{F_a^{R\prime}\widetilde{F}_a^r}{T_0}\widetilde{V}_a^{-1} = R_a L_a, \end{aligned} \tag{B.7}$$

where the first equality follows from the definition of H_a in (B.2), the second equality follows from $F_a^R = F_a R_a$ and $\Lambda^{R\prime} = R_a^{-1}\Lambda^{0\prime}$, the third equality follows from the definition of L_a in (B.6).

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 (B.4) and (B.5), 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

$$L_a = I_{r_a} + O_p(C_{NT_0}^{-1}), \tag{B.8}$$

with two modifications to the proof in Bai and Ng (2013): (i) $T_0^{-1}(\widetilde{F}_a^r - F_a^R L_a)^\prime F_a^R = O_p(C_{NT_0}^{-2})$ in Bai and Ng (2013) is changed to $T_0^{-1}(\widetilde{F}_a^r - F_a^R L_a)^\prime F_a^R = O_p(C_{NT_0}^{-1})$, which follows from $F_a^R L_a = F_a H_a$, (B.1), 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 (B.8). 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 (B.8).

Combining the results in (B.7) and (B.8), we obtain $H_a - R_a = O_p(C_{NT}^{-1})$ because $T_0/T \to \pi_0 \in (0,1)$. Similar arguments give $H_b - R_b = O_p(C_{NT}^{-1})$. $\square$

C Proof of Results in Section 4

Recall that we have defined

$$\Lambda^R = \Lambda^0(R_a^{-1})' \in R^{N\times r_a},\ \Psi^R = \Psi^0(R_b^{-1})' \in R^{N\times r_b} \text{ and } \Gamma^R = \left(\Psi_1^R - \Lambda^R,\ \Psi_2^R\right) \tag{C.1}$$

in (2.9), (2.11), and (2.12), respectively. For the ease of notation, we also define

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

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 (5.11) and the definition of Γ^*,

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

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

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

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

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

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.,

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

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

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

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

$$P_1^r(\widehat{\Lambda}) = \alpha_{NT} \sum_{\ell=1}^{r_a} \omega_\ell^\lambda ||\widehat{\Lambda}_\ell|| \leq \alpha_{NT} \sum_{\ell=1}^{k} \omega_\ell^\lambda ||\widehat{\Lambda}_\ell|| = P_1(\widehat{\Lambda}),$$
$$P_2^r(\widehat{\Gamma}) = \beta_{NT} \sum_{\ell \in \mathcal{Z}} \omega_\ell^\gamma ||\widehat{\Gamma}_\ell|| \leq \beta_{NT} \sum_{\ell=1}^{k} \omega_\ell^\gamma ||\widehat{\Gamma}_\ell|| = P_2(\widehat{\Gamma}). \tag{C.7}$$

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 (C.6) satisfy

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

following the inequalities in (C.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

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

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

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

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

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

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

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

Next, we consider the left-hand side of (C.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

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

Some equivalent relationships are useful in the calculation below

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

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

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

where the first equality follows from (C.14), the second equality follows from adding and subtracting $\widetilde{F}_a \Lambda^{*\prime}$, and the third equality follows from (C.14). The difference between the true common component $F_a \Lambda^{0\prime}$ and the estimated common component $\widetilde{F}_a \widehat{\Lambda}^\prime$ are decomposed into two pieces by the calculation in (C.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 (C.12) satisfies

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

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

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

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 (C.14) and the definition of $M_a(\Lambda, F)$ in (C.5). The second term M_2 is

$$\begin{aligned} M_2 &= (NT)^{-1} \left\| \widetilde{F}_a(\widehat{\Lambda} - \Lambda^*)^\prime \right\|^2 \\ &= (NT)^{-1} tr\left((\widehat{\Lambda} - \Lambda^*)\widetilde{F}_a^\prime \widetilde{F}_a(\widehat{\Lambda} - \Lambda^*)^\prime \right) \\ &= \frac{T_0}{T} N^{-1} \left\| \widehat{\Lambda} - \Lambda^* \right\|^2 = \frac{T_0}{T} Z_\lambda^2, \end{aligned} \tag{C.18}$$

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

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

By the Cauchy-Schwarz inequality,

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

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,

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

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

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

The fourth term M_4 is

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

To investigate M_4, we note that

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

by the Cauchy-Schwarz inequality, (B.1), and Lemma 3. Applying the Cauchy-Schwarz inequality, we have

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Assumptions P1 and P2 imply that

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

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

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

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

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

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

Under Assumptions P1 and P2,

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

Define the residual matrices

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

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

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

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 (C.38) is larger than the marginal gain on the left-hand side of (C.38). The inequality in (C.38) can be equivalently written as

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

which holds provided that

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

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

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

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

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

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 (B.1); (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 (C.34). Putting them together with (C.41), we have

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

By the same arguments, we have

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

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

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

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

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

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

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

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 (C.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 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 (C.33) and (C.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 (C.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. $\square$

Proof of Lemma 1. 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

$$\frac{\Lambda^{R\prime}\Lambda^R}{N} = V_a^{1/2}\Upsilon_a'\Sigma_a^{-1/2}\frac{\Lambda^{0\prime}\Lambda^0}{N}\Sigma_a^{-1/2}\Upsilon_a V_a^{1/2} = V_a \text{ and } \frac{\Psi^{R\prime}\Psi^R}{N} = V_b. \tag{C.48}$$

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

$$N^{-1}\left\|\Lambda_\ell^R\right\|^2 = a_\ell'\left(N^{-1}\Lambda^{R\prime}\Lambda^R\right)a_\ell = a_\ell' V_a a_\ell = \rho_\ell(\Sigma_\Lambda\Sigma_F) + o(1). \tag{C.49}$$

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

$$N^{-1}||\Gamma_\ell^R||^2 = a_\ell' \left(N^{-1}\Psi^{R\prime}\Psi^R\right) a_\ell = a_\ell' V_b a_\ell = \rho_\ell(\Sigma_\Psi \Sigma_{\overline{F}}) + o(1). \tag{C.50}$$

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

$$N^{-1}\Gamma^{R\prime}\Gamma^R = N^{-1}(\Psi^R - \Lambda^R)'(\Psi^R - \Lambda^R) = \mathbf{e}'\Sigma_{\Lambda\Psi}^{+}\mathbf{e} + o(1), \tag{C.51}$$

where $\mathbf{e} = \lim_{N\to\infty}\left(R_b^{-1},\ -R_a^{-1}\right)'$ has full rank following Assumptions A and B and $\Sigma_{\Lambda\Psi}^{+}$ is defined in (2.6). 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

$$\begin{aligned} N^{-1}||\Gamma_\ell^R||^2 &= N^{-1}\left\|\Gamma^R a_\ell\right\|^2 = N^{-1}\left\|\Psi^R a_\ell - \Lambda^R a_\ell\right\|^2 \\ &\ge \left(N^{-1/2}\left\|\Psi^R a_\ell\right\| - N^{-1/2}\left\|\Lambda^R a_\ell\right\|\right)^2 \\ &= [(\rho_\ell(\Sigma_\Psi\Sigma_{\overline{F}}))^{1/2} - (\rho_\ell(\Sigma_\Lambda\Sigma_F))^{1/2}]^2 + o(1), \end{aligned} \tag{C.52}$$

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 (C.48). □

Proof of Theorem 2. First, Theorem 1(a) for $\ell = r_a$ and Lemma 1(a) imply that $\Pr(||\widehat{\Lambda}_\ell|| > 0) \to 1$ for $\ell = r_a$ and thus $\Pr(\widehat{r}_a \ge r_a) \to 1$. Theorem 1(b) implies that $\Pr(\widehat{r}_a \le r_a) \to 1$. 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 1(b) imply that $\Pr(||\widehat{\Gamma}_\ell|| > 0) \to 1$ for $\ell = r_b$ and thus $\Pr(\widehat{r}_b \ge r_b) \to 1$. Theorem 1(e) implies that $\Pr(\widehat{r}_b \le r_b) \to 1$. Hence, $\Pr(\widehat{r}_b = r_b) \to 1$ for a type-2 change, which, together with part (a), also implies $\Pr(\widehat{\mathcal{S}} = 1) \to 1$ for a type-2 change because by definition, $\widehat{\mathcal{S}} = 1$ if $\widehat{r}_b > \widehat{r}_a$.

Third, for a type-1 change where $r_b = r_a$ and $\mathcal{S}_0 = 1$, Theorem 1(c), Lemmas 1(c) and 1(d), and Assumption ID imply that $\Pr(||\widehat{\Gamma}_\ell|| > 0) \to 1$ for some $\ell \le r_a$ and thus $\Pr(\widehat{\mathcal{S}} = 1) \to 1$. Note that by definition in (3.5), we have $\widehat{r}_b \ge \widehat{r}_a$. Thus, part (a) and $r_a = r_b$ imply that $\Pr(\widehat{r}_b \ge r_b) \to 1$. On the other hand, Theorem 1(e) implies that $\Pr(\widehat{r}_b \le r_b) \to 1$. 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 $\mathcal{S}_0 = 0$, Theorems 1(d) and 1(e) imply that $\Pr(\widehat{\Gamma} = 0) \to 1$. Thus, $\Pr(\widehat{\mathcal{S}} = 0) \to 1$ by (3.6) and $\Pr(\widehat{r}_b = r_b) \to 1$ by (3.5) and part (a). $\square$

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 (C.32),

$$N^{-1}||\widetilde{\Lambda}_{LS} - \Lambda^*||^2 = O_p(C_{NT}^{-2}) \text{ and } N^{-1}||\widetilde{\Gamma}_{LS} - \Gamma^*||^2 = O_p(C_{NT}^{-2}), \tag{C.53}$$

which combined with the definitions of Λ^* and Γ^* and Lemma 1 imply that

$$\Pr(N^{-1}||\widetilde{\Lambda}_{LS,\ell}||^2 \geq C) \to 1 \text{ for } \ell = 1, \ldots, r_a, \;\; \Pr(N^{-1}||\widetilde{\Gamma}_{LS,\ell}||^2 \geq C) \to 1 \text{ for } \ell \in \mathcal{Z} \tag{C.54}$$

and

$$N^{-1}||\widetilde{\Lambda}_{LS,\ell}||^2 = O_p(C_{NT}^{-2}) \text{ for } \ell > r_a \text{ and } N^{-1}||\widetilde{\Gamma}_{LS,\ell}||^2 = O_p(C_{NT}^{-2}) \text{ for } \ell \in \mathcal{Z}^C. \tag{C.55}$$

Next, we show that (C.33) and (C.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 (C.54) imply that (C.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 (C.55) 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 (C.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. $\square$

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 $\widetilde{\Lambda}^{(2)} = \widehat{\Lambda}_{PMS}^{(1)}$, 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

$$\omega_\ell^\lambda = (N^{-1}||\widetilde{\Lambda}_{LS,\ell}||^2)^{-d} \text{ for } \ell = 1, \ldots, k, \tag{C.56}$$

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

$$\omega_\ell^\gamma = (N^{-1}||\widetilde{\Gamma}_{LS,\ell}||^2)^{-d} \text{ for } \ell = 1, \ldots, k, \tag{C.57}$$

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

$$\widetilde{\Gamma}^{(2)} = \left(\widetilde{\Psi}_{LS}^- Q - \widetilde{\Lambda}_{LS}^-, \, 0_{N\times(k-r_a)}\right), \tag{C.58}$$

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

$$\Gamma^Q = \left(\Psi^R Q - \Lambda^R, \, 0_{N\times(k-r_a)}\right). \tag{C.59}$$

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 (C.58) and (C.59), w.p.a.1,

$$\begin{aligned} N^{-1}||\widetilde{\Gamma}^{(2)} - \Gamma^Q||^2 &= N^{-1}\left\|(\widetilde{\Psi}_{LS}^- - \Psi^R)Q - (\widetilde{\Lambda}_{LS}^- - \Lambda^R)\right\|^2 \\ &= N^{-1}\left\|\left(\widetilde{\Gamma}_{LS}^- - \Gamma^R\right)Q + \left(\widetilde{\Lambda}_{LS}^- - \Lambda^R\right)(Q - I_{r_a})\right\|^2 \\ &= O_p(C_{NT}^{-2}), \end{aligned} \tag{C.60}$$

where the last equality follows from the triangle inequality and (C.53). 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

$$N^{-1}||\Gamma_\ell^Q||^2 \geq C \text{ if } \ell \in \mathcal{Z} \tag{C.61}$$

because $\Gamma_\ell^Q = \Psi^R Q_\ell - \Lambda_\ell^R$ and $||Q_\ell|| = 1$. Therefore, (C.60) and (C.61) imply that

$$\omega_\ell^\gamma = O_p(1) \text{ for } \ell \in \mathcal{Z} \text{ when there is a type-1 change.} \tag{C.62}$$

If there is no structural change, by (C.53), $N^{-1}||\widetilde{\Lambda}_{LS}^- - \Lambda^R||^2 = O_p(C_{NT}^{-2})$ 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

$$N^{-1}\left\|\widetilde{\Psi}_{LS}^- Q - \widetilde{\Lambda}_{LS}^-\right\|^2 \leq N^{-1}\left\|\widetilde{\Psi}_{LS}^- - \widetilde{\Lambda}_{LS}^-\right\|^2 = O_p(C_{NT}^{-2}), \tag{C.63}$$

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 (C.58) and (C.63), we obtain

$$N^{-1}||\widetilde{\Gamma}^{(2)}||^2 = O_p(C_{NT}^{-2}) \text{ when } \Gamma^0 = 0, \tag{C.64}$$

which together with (C.53) and $\Gamma^* = 0$ implies that

$$(\omega_\ell^\gamma)^{-1} = O_p(C_{NT}^{-2d}) \text{ for } \ell = 1, \ldots, k \text{ when there is no structural change.} \tag{C.65}$$

Next, we show that (C.33) and (C.36) hold without imposing Assumptions P1 and P2, so that the proof of Theorem 1 follows without these two assumptions. To show (C.33), note that $\omega_\ell^\lambda = O_p(1)$ for $\ell = 1, \ldots, r_a$ is implied by (C.54) and (C.56), $\omega_\ell^\gamma = O_p(1)$ for $\ell \in \mathcal{Z}$ is implied by (C.54) and (C.57) for a type-2 change, and $\omega_\ell^\gamma = O_p(1)$ for $\ell \in \mathcal{Z}$ is proved in (C.62) for a type-1 change.

To show (C.36), note that: (i) $\{r_a + 1, \ldots, k\} \subseteq \mathcal{L}_a$ holds by (C.55) and (C.56); (ii) $\{r_b + 1, \ldots, k\} \subseteq \mathcal{L}_b$ holds by (C.55) and (C.57); and (iii) if $\Gamma^0 = 0$, $\{1, \ldots, k\} = \mathcal{L}_b$ follows from (C.53) and (C.65).

Because (C.33) and (C.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. □

D Proof of Results in Section 6

Proof of Lemma 2. For $\pi \leq \pi_0$, the result follows from the representation in (6.3) and Assumptions A-D. Analogous arguments yield results for $\pi > \pi_0$. □

Proof of Corollary 3. This corollary is implied by Lemma 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 $\pi \in \Pi$.

Define $r^+ = rank(\Sigma_{\Lambda\Psi}^+)$, $T_a = \lfloor T\pi \rfloor$, and $T_b = T - T_a$. First, consider the second subsample $X_b(\pi)$. When $\pi < \pi_0$, following the model in (6.2), the variance of the factor loadings is

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

By Assumption B and the continuous mapping theorem, we know that Σ_{ab}^{+} has rank r^{+} w.p.a.1, which implies that the rank of $(\Lambda^0,\ \Psi^0)$ is r^{+} w.p.a.1. Thus, there is a $(r_a + r_b) \times (r_a + r_b)$ orthogonal matrix S such that the first r^{+} columns of $(\Lambda^0,\ \Psi^0)\,S$ have full rank and the last $(r_a + r_b - r_{+})$ columns are 0 w.p.a.1. As such, the model in (6.2) can be written as an approximate factor model with r^{+} factors, and the factors and their loadings both have full ranks asymptotically. With a transformation analogous to that in (2.10) to standardize the factors and diagonalize the loadings, the DGP in (6.2) can be written as

$$X_b(\pi) = F_b^R(\pi)\Psi^R(\pi)' + e_b(\pi), \tag{D.2}$$

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

$$\begin{aligned} T_b^{-1} F_b^R(\pi)' F_b^R(\pi) &= I_{r^+} + O_{p\pi}(T^{-1/2}), \\ N^{-1}\Psi^R(\pi)'\Psi^R(\pi) &= \Lambda_b(\pi), \end{aligned} \tag{D.3}$$

where $\Lambda_b(\pi)$ is a $r^{+} \times r^{+}$ diagonal matrix whose diagonal elements are the positive eigenvalues of $\Sigma_F^{+}(\pi)\Sigma_{ab}^{+}$ in a decreasing order. This is analogous to the transformation considered in (B.3)-(B.5) in the proof of Lemma 3 except $\pi < \pi_0$ rather than $\pi = \pi_0$. When $\pi \geq \pi_0$, the DGP in (6.2) can be written as in (D.2) and (D.3) 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 (6.1) can be written as

$$X_a(\pi) = F_a^R(\pi)\Lambda^R(\pi)' + e_a(\pi), \tag{D.4}$$

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

$$\begin{aligned} T_a^{-1} F_a^R(\pi)' F_a^R(\pi) &= I_{r^+} + O_{p\pi}(T^{-1/2}), \\ N^{-1}\Lambda^R(\pi)'\Lambda^R(\pi) &= \Lambda_a(\pi), \end{aligned} \tag{D.5}$$

where $\Lambda_a(\pi)$ is a $r^{+} \times r^{+}$ diagonal matrix with positive eigenvalues. When $\pi \leq \pi_0$, the DGP in (6.1) can be written as that in (D.4) and (D.5) 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,

$$N^{-1}||\Lambda_\ell^R(\pi)||^2 \geq C \text{ for } \ell = 1, \ldots, r_a,\ \ N^{-1}||\Psi_\ell^R(\pi)||^2 \geq C \text{ for } \ell = 1, \ldots, r_b. \tag{D.6}$$

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 (6.12) satisfy

$$\begin{aligned}
\omega_\ell^{\lambda*(i)}(\pi) &= \left(N^{-1}||\widetilde{\Lambda}_{\ell,LS}(\pi)||^2\right)^{-d} \text{ for } i = 1 \text{ and } 2,\\
\omega_\ell^{\gamma*(1)}(\pi) &= \max\left\{\left(N^{-1}||\widetilde{\Gamma}_{\ell,LS}(\pi)||^2\right)^{-d}, \left(N^{-1}||\widetilde{\Psi}_{\ell,LS}(\pi)||^2\right)^{-d}\right\},\\
\omega_\ell^{\gamma*(2)}(\pi) &= \omega_\ell^{\gamma*(1)}(\pi) \text{ if (i) } \widehat{r}_a^{(1)} < \widehat{r}_b^{(1)} \text{ or (ii) } \widehat{r}_a^{(1)} = \widehat{r}_b^{(1)} \text{ and } \ell > \widehat{r}_a^{(1)},\\
\omega_\ell^{\gamma*(2)}(\pi) &= \max\left\{\left(N^{-1}||\widetilde{\Psi}_{\ell,LS}^{-}(\pi)\mathrm{w}(\pi) - \widetilde{\Lambda}_{\ell,LS}(\pi)||^2\right)^{-d}, \left(N^{-1}||\widetilde{\Psi}_{\ell,LS}(\pi)||^2\right)^{-d}\right\} \text{ otherwise,}
\end{aligned} \tag{D.7}$$

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

$$\Pr(\min_{\pi\in\Pi} \widehat{r}_a^{(i)}(\pi) \geq r_a) \to 1 \text{ for } i = 1 \text{ and } 2. \tag{D.8}$$

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

$$\Lambda^\dagger(\pi) = \left(\Lambda_{1:r_a}^R(\pi), \widehat{\Lambda}(\pi)_{r_a+1:k}\right), \Gamma^\dagger(\pi) = \widehat{\Gamma}(\pi), \text{ and } \Psi^\dagger(\pi) = \Lambda^\dagger(\pi) + \Gamma^\dagger(\pi). \tag{D.9}$$

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

$$Z_{\lambda}^2(\pi) = N^{-1} \left\| \widehat{\Lambda}(\pi) - \Lambda^{\dagger}(\pi) \right\|^2, \; Z_{\psi}^2(\pi) = N^{-1} \left\| \widehat{\Psi}(\pi) - \Psi^{\dagger}(\pi) \right\|^2, \; Z_{\gamma}^2(\pi) = N^{-1} \left\| \widehat{\Gamma}(\pi) - \Gamma^{\dagger}(\pi) \right\|^2. \tag{D.10}$$

The criterion function for the shrinkage estimator can be written as

$$Q(\Lambda, \Gamma; \pi) = M_a(\Lambda, \widetilde{F}_a(\pi)) + M_b(\Psi, \widetilde{F}_b(\pi)) + P_1^*(\Lambda) + P_2^*(\Gamma), \tag{D.11}$$

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

$$\begin{aligned} M_a(\Lambda, F_a) &= (NT)^{-1} \left\| X_a(\pi) - F_a \Lambda' \right\|^2, \text{ and} \\ M_b(\Psi, F_b) &= (NT)^{-1} \left\| X_b(\pi) - F_b(\Lambda + \Gamma)' \right\|^2. \end{aligned} \tag{D.12}$$

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

$$\phi_\ell^\lambda = \mathbb{E}_\xi[\alpha_{NT}(\xi)\omega_\ell^{\lambda *}(\xi)] \text{ and } \phi_\ell^\gamma = \mathbb{E}_\xi[\beta_{NT}(\xi)\omega_\ell^{\gamma *}(\xi)], \tag{D.13}$$

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 $Q(\widehat{\Lambda}(\pi), \widehat{\Gamma}(\pi)) \le Q(\Lambda^\dagger(\pi), \Gamma^\dagger(\pi))$, i.e.,

$$\begin{aligned} &\left[M_a(\widehat{\Lambda}(\pi), \widetilde{F}_a(\pi)) - M_a(\Lambda^\dagger(\pi), \widetilde{F}_a(\pi)) \right] + \left[M_b(\widehat{\Psi}(\pi), \widetilde{F}_b(\pi)) - M_b(\Psi^\dagger(\pi), \widetilde{F}_b(\pi)) \right] \\ \le &\left[P_1^*(\Lambda^\dagger(\pi)) - P_1^*(\widehat{\Lambda}(\pi)) \right] + \left[P_2^*(\Gamma^\dagger(\pi)) - P_2^*(\widehat{\Gamma}(\pi)) \right], \end{aligned} \tag{D.14}$$

where $\widehat{\Psi}(\pi) = \widehat{\Lambda}(\pi) + \widehat{\Gamma}(\pi)$. We start with the right-hand side of (D.14). 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

$$P_1^*(\Lambda^\dagger(\pi)) - P_1^*(\widehat{\Lambda}(\pi)) = \sum_{\ell=1}^{r_a} \phi_\ell^\lambda \left(|\Lambda_\ell^\dagger(\pi)| - |\widehat{\Lambda}_\ell(\pi)| \right) \le \bar{b}_\Lambda Z_\lambda(\pi), \text{ where } \bar{b}_\Lambda = N^{1/2} \left(\sum_{\ell=1}^{r_a} (\phi_\ell^\lambda)^2 \right)^{1/2}. \tag{D.15}$$

Because $\Gamma^\dagger(\pi) = \widehat{\Gamma}(\pi)$, the second term on the right-hand side of (D.14) is 0.

Next, we consider the left-hand side of (D.14). 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

$$\begin{aligned}
e_a(\widehat{\Lambda}(\pi),\widetilde{F}_a(\pi)) &= X_a(\pi)-\widetilde{F}_a(\pi)\widehat{\Lambda}(\pi)' \\
&= X_a(\pi)-\widetilde{F}_a^r(\pi)\widehat{\Lambda}^r(\pi)'-\widetilde{F}_a^\perp(\pi)\widehat{\Lambda}^\perp(\pi)' \\
&= \left(X_a(\pi)-\widetilde{F}_a^r(\pi)\Lambda^r(\pi)'-\widetilde{F}_a^\perp(\pi)\widehat{\Lambda}^\perp(\pi)'\right)-\widetilde{F}_a^r(\pi)\left(\widehat{\Lambda}^r(\pi)-\Lambda^r(\pi)\right)' \\
&= e_a(\Lambda^\dagger(\pi),\widetilde{F}_a(\pi))-\widetilde{F}_a^r(\pi)\left(\widehat{\Lambda}^r(\pi)-\Lambda^r(\pi)\right)', \qquad \text{(D.16)}
\end{aligned}$$

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 $r_a(\pi)\geq r_a$, 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 $T_a\times(r_a(\pi)-r_a)$ 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 $N\times(r_a(\pi)-r_a)$ submatrices $\Lambda^r(\pi)$ and $\Lambda^{r+}(\pi)$. Following the partition, we can write

$$X_a(\pi)=e_a(\pi)+F_a^r(\pi)\Lambda^r(\pi)'+F_a^{r+}(\pi)\Lambda^{r+}(\pi)'. \qquad \text{(D.17)}$$

The second expansion is

$$\begin{aligned}
e_a(\Lambda^\dagger(\pi),\widetilde{F}_a(\pi)) &= X_a(\pi)-\widetilde{F}_a^r(\pi)\Lambda^r(\pi)'-\widetilde{F}_a^\perp(\pi)\widehat{\Lambda}^\perp(\pi)' \qquad \text{(D.18)}\\
&= e_a(\pi)+\left(F_a^r(\pi)-\widetilde{F}_a^r(\pi)\right)\Lambda^r(\pi)'+F_a^{r+}(\pi)\Lambda^{r+}(\pi)'-\widetilde{F}_a^\perp(\pi)\widehat{\Lambda}^\perp(\pi)',
\end{aligned}$$

where first equality holds by definition and the second equality follows from (D.17). With the first expansion in (D.16), we have

$$\begin{aligned}
M_a(\widehat{\Lambda}(\pi),\widetilde{F}_a(\pi)) &= (NT)^{-1}\left\|e_a(\widehat{\Lambda}(\pi),\widetilde{F}_a(\pi))\right\|^2 \\
&= (NT)^{-1}\left\|e_a(\Lambda^\dagger(\pi),\widetilde{F}_a(\pi))\right\|^2+(NT)^{-1}\left\|\widetilde{F}_a^r\left(\widehat{\Lambda}^r(\pi)-\Lambda^r(\pi)\right)'\right\|^2 \\
&\quad -2(NT)^{-1}tr\left[e_a(\Lambda^\dagger(\pi),\widetilde{F}_a(\pi))'\widetilde{F}_a^r(\pi)\left(\widehat{\Lambda}^r(\pi)-\Lambda^r(\pi)\right)\right] \\
&= M_a(\Lambda^\dagger(\pi),\widetilde{F}_a(\pi))+K_0+K_1+K_2+K_3+K_4, \qquad \text{(D.19)}
\end{aligned}$$

where

$$\begin{aligned}
K_0 &= (NT)^{-1}\left\|\widetilde{F}_a^r\left(\widehat{\Lambda}^r(\pi)-\Lambda^r(\pi)\right)'\right\|^2 \\
&= \frac{T_a}{T}\frac{1}{N}tr\left[\left(\widehat{\Lambda}^r(\pi)-\Lambda^r(\pi)\right)\frac{\widetilde{F}_a^{r\prime}\widetilde{F}_a^r}{T_a}\left(\widehat{\Lambda}^r(\pi)-\Lambda^r(\pi)\right)'\right] \\
&= \frac{T_a}{T}Z_\lambda^2(\pi) \qquad \text{(D.20)}
\end{aligned}$$

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 (D.18), and they are specified below. The first term is

$$K_1 = -2(NT)^{-1} tr \left[e_a(\pi)' \widetilde{F}_a^r(\pi) \left(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi) \right) \right] = \frac{T_a}{T} O_{p\pi}(C_{NT}^{-1}) Z_\lambda(\pi), \tag{D.21}$$

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

$$\begin{aligned} K_2 &= -2(NT)^{-1} tr \left(\Lambda^r(\pi)(F_a^r(\pi) - \widetilde{F}_a^r(\pi))' \widetilde{F}_a^r(\pi)(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi))' \right) \\ &= \frac{T_a}{T} O_{p\pi}(C_{NT}^{-1}) Z_\lambda(\pi) \end{aligned} \tag{D.22}$$

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

$$\begin{aligned} K_3 &= -2(NT)^{-1} tr \left(\Lambda^{r+}(\pi) F_a^{r+}(\pi)' \widetilde{F}_a^r(\pi)(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi))' \right) \\ &= -2(NT)^{-1} tr \left(\Lambda^{r+}(\pi) \left(F_a^{r+}(\pi) - \widetilde{F}_a^{r+}(\pi) \right)' \widetilde{F}_a^r(\pi)(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi))' \right) \\ &= \frac{T_a}{T} O_{p\pi}(C_{NT}^{-1}) Z_\lambda(\pi), \end{aligned} \tag{D.23}$$

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 (C.25) and (C.26). The forth term is

$$K_4 = 2(NT)^{-1} tr \left[\widehat{\Lambda}^\perp(\pi) \widetilde{F}_a^\perp(\pi)' \widetilde{F}_a^r(\pi) \left(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi) \right) \right] = 0 \tag{D.24}$$

because $\widetilde{F}_a^\perp(\pi)' \widetilde{F}_a^r(\pi) = 0$ by construction. Combining (D.19)-(D.24), we obtain

$$M_a(\widehat{\Lambda}(\pi), \widetilde{F}_a(\pi)) - M_a(\Lambda^\dagger(\pi), \widetilde{F}_a(\pi)) = \frac{T_a}{T} Z_\lambda^2(\pi) + O_{p\pi}(C_{NT}^{-1}) Z_\lambda(\pi). \tag{D.25}$$

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

$$M_b(\widehat{\Psi}(\pi), \widetilde{F}_b(\pi)) - M_b(\Psi^\dagger(\pi), \widetilde{F}_b(\pi)) = \frac{T_b}{T} Z_\psi^2(\pi) + O_{p\pi}(C_{NT}^{-1}) Z_\psi(\pi). \tag{D.26}$$

Plugging (D.25) and (D.26) into the left-hand side of (D.14), we obtain

$$\frac{T_a}{T} Z_\lambda^2(\pi) + O_{p\pi}(C_{NT}^{-1}) Z_\lambda(\pi) + \frac{T_b}{T} Z_\psi^2(\pi) + O_{p\pi}(C_{NT}^{-1}) Z_\psi \leq \bar{b}_\Lambda Z_\lambda(\pi), \tag{D.27}$$

which further implies that

$$Z_\lambda(\pi) = O_{p\pi}(\bar{b}_\Lambda + C_{NT}^{-1}). \tag{D.28}$$

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 (D.28), which together with (D.6) 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 (6.11), $\alpha_{NT}(\pi) = \kappa_1(\pi) N^{-1/2} C_{NT_a}^{-d-1}$, where $\kappa_1(\pi) \leq \overline{\kappa}_1$. Thus, we have

$$N^{1/2}\phi_\ell^\lambda = N^{1/2}\mathbb{E}_\xi[\alpha_{NT}(\xi)\omega_\ell^{\lambda*}(\xi)] = O_p(C_{NT}^{-1}) \tag{D.29}$$

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

$$\overline{b}_\Lambda = O_p(C_{NT}^{-1}) \tag{D.30}$$

for both the first- and second-step PLS estimation. It follows from (D.28) 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,

$$\Pr(\min_{\pi\in\Pi} \widehat{r}_b^{(i)}(\pi) \geq r_b) \to 1 \text{ if } r_b > r_a. \tag{D.31}$$

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 (D.6) for $\ell = r_b$. To show (D.31), 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 (D.9) as

$$\Lambda^\dagger(\pi) = \widehat{\Lambda}(\pi),\ \Gamma^\dagger(\pi) = \left(\widehat{\Gamma}(\pi)_{1:r_b-1},\ \Gamma_{r_b}^R(\pi),\ \widehat{\Gamma}(\pi)_{r_b+1:k}\right) \text{ and } \Psi^\dagger(\pi) = \Lambda^\dagger(\pi) + \Gamma^\dagger(\pi) \tag{D.32}$$

and keep the definitions of $Z_\lambda(\pi)$, $Z_\psi(\pi)$, $Z_\gamma(\pi)$ in (D.10) unchanged. Now consider the inequality in (D.14). Because $\Lambda^\dagger(\pi) = \widehat{\Lambda}(\pi)$, the right-hand side of (D.14) becomes for $\ell = r_b$,

$$P_2^*(\Gamma^\dagger(\pi)) - P_2^*(\widehat{\Gamma}(\pi)) = \phi_\ell^\gamma\left(|\Gamma_\ell^R(\pi)| - |\widehat{\Gamma}_\ell(\pi)|\right) \leq \overline{b}_{\Gamma b} Z_\gamma(\pi), \text{ where } \overline{b}_{\Gamma b} = N^{1/2}\phi_\ell^\gamma. \tag{D.33}$$

By arguments analogous to those used to show (D.25) and (D.26), the left-hand side of (D.14) becomes

$$M_b(\widehat{\Psi}(\pi), \widetilde{F}_b(\pi)) - M_b(\Psi^\dagger(\pi), \widetilde{F}_b(\pi)) = \frac{T_b}{T} Z_\psi^2(\pi) + O_{p\pi}(C_{NT}^{-1}) Z_\psi(\pi). \tag{D.34}$$

Putting (D.33) and (D.34) together with (D.14), we get

$$Z_\psi(\pi) = O_{p\pi}(\overline{b}_{\Gamma b} + C_{NT}^{-1}). \tag{D.35}$$

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 (D.29), we have $\bar{b}_{\Gamma b} = O_p(C_{NT}^{-1})$ for the first-step estimator, which immediately implies that $Z_\psi(\pi) = O_{p\pi}(C_{NT}^{-1})$ and

$$N^{-1}||\widehat{\Gamma}_\ell(\pi) - \Gamma_\ell^R(\pi)||^2 = O_{p\pi}(C_{NT}^{-2}) \text{ for } \ell = r_b. \tag{D.36}$$

This proof (D.31) 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 (D.7) with $r_b > r_a$ w.p.a.1, which is the same as that for the first-step estimator. Hence, $\bar{b}_{\Gamma b} = O_p(C_{NT}^{-1})$ for the second-step estimator and it follows that (D.31) holds for $i = 2$ as well.

Step 3. We prove

$$\Pr(\widehat{r}_a^{(1)} = r_a) \to 1 \tag{D.37}$$

by showing that the inequalities in (D.8) 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

$$\left\| e^a(\widehat{\Lambda}(\pi_0))'\widetilde{F}_{a,\ell}(\pi_0)\right\| + \left\| e^b(\widehat{\Lambda}(\pi_0) + \widehat{\Gamma}(\pi_0))'\widetilde{F}_{b,\ell}(\pi_0)\right\| < \frac{NT}{2}\phi_\ell^\lambda, \tag{D.38}$$

which is similar to (C.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

$$N^{-1/2}||\widehat{\Lambda}(\pi_0) - \Lambda^*|| = O_p(\bar{b}_\Lambda + \bar{b}_\Gamma + C_{NT}^{-1}) \text{ and } N^{-1/2}||\widehat{\Gamma}(\pi_0) - \Gamma^*|| = O_p(\bar{b}_\Lambda + \bar{b}_\Gamma + C_{NT}^{-1}), \tag{D.39}$$

where

$$\bar{b}_\Lambda = N^{1/2}(\sum_{\ell=1}^{r_a}(\phi_\ell^\lambda)^2)^{1/2} \text{ and } \bar{b}_\Gamma = N^{1/2}(\sum_{\ell\in\mathcal{Z}}(\phi_\ell^\gamma)^2)^{1/2}. \tag{D.40}$$

We have shown $\bar{b}_\Lambda = O_p(C_{NT}^{-1})$ in (D.30) for both the first- and second-step estimators. By similar arguments under Assumption R*(i) and (D.6), we also have $\bar{b}_\Gamma = O_p(C_{NT}^{-1})$ for the first step estimator. Because $\bar{b}_\Lambda = O_p(C_{NT}^{-1})$ and $\bar{b}_\Gamma = O_p(C_{NT}^{-1})$,

$$N^{-1/2}||\widehat{\Lambda}^{(1)}(\pi_0) - \Lambda^*|| = O_p(C_{NT}^{-1}) \text{ and } N^{-1/2}||\widehat{\Gamma}^{(1)}(\pi_0) - \Gamma^*|| = O_p(C_{NT}^{-1}). \tag{D.41}$$

Following the arguments used to show (C.43) and (C.44), (D.38) holds provided that

$$N^{-1/2}C_{NT}^{-1} = o_p(\phi_\ell^\lambda), \tag{D.42}$$

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

$$\phi_\ell^\lambda = \mathbb{E}_\xi[\alpha_{NT}(\xi)\omega_\ell^{\lambda*(1)}(\xi)] \geq \underline{\kappa}_1 N^{-1/2}C_{NT}^{-d-1}\mathbb{E}_\xi[\omega_\ell^{\lambda*(1)}(\xi)\mathcal{I}_{\{\xi\leq\pi_0\}}], \tag{D.43}$$

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 (C.55). Therefore,

$$\sup_{\pi\leq\pi_0}\left(\omega_\ell^{\lambda*(1)}(\pi)\right)^{-1} = \sup_{\pi\leq\pi_0}[N^{-1}||\widetilde{\Lambda}_{LS,\ell}(\pi)||^2]^d = O_p(C_{NT}^{-2d}) \text{ for } \ell > r_a. \tag{D.44}$$

Thus, for $\ell > r_a$,

$$\begin{aligned} N^{-1/2}C_{NT}^{-1}(\phi_\ell^\lambda)^{-1} &\leq \underline{\kappa}_1^{-1}C_{NT}^d\left(\mathbb{E}_\xi[\omega_\ell^{\lambda*(1)}(\xi)\mathcal{I}_{\{\xi\leq\pi_0\}}]\right)^{-1} \\ &\leq \underline{\kappa}_1^{-1}C_{NT}^d\left(\inf_{\pi\leq\pi_0}[\omega_\ell^{\lambda*(1)}(\pi)]\mathbb{E}_\xi[\mathcal{I}_{\{\xi\leq\pi_0\}}]\right)^{-1} \\ &= \frac{C_{NT}^d\sup_{\pi\leq\pi_0}\left(\omega_\ell^{\lambda*(1)}(\pi)\right)^{-1}}{\underline{\kappa}_1\mathbb{E}_\xi[\mathcal{I}_{\{\xi\leq\pi_0\}}]} = O_p(C_{NT}^{-d}), \end{aligned} \tag{D.45}$$

where the last equality is by (D.44) 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

$$\Pr(\widehat{r}_a^{(1)}(\pi_0) \leq r_a) \to 1. \tag{D.46}$$

Combining (D.8) 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 (D.37).

Step 4. We prove

$$\Pr(\widehat{r}_b^{(1)} = r_b) \to 1 \tag{D.47}$$

by showing that the inequalities in (D.31) 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

$$\left\|e^b(\widehat{\Lambda}(\pi_0) + \widehat{\Gamma}(\pi_0))'\widetilde{F}_{b,\ell}(\pi_0)\right\| < \frac{NT}{2}\phi_\ell^\gamma. \tag{D.48}$$

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

$$\phi_\ell^\gamma = \mathbb{E}_\xi[\beta_{NT}(\xi)\omega_\ell^{\gamma*(1)}(\xi)] \geq \underline{\kappa}_2 N^{-1/2}C_{NT}^{-d-1}\mathbb{E}_\xi[\omega_\ell^{\gamma*(1)}(\xi)\mathcal{I}_{\{\xi\geq\pi_0\}}], \tag{D.49}$$

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 (C.55). Therefore,

$$\sup_{\pi > \pi_0} \left(\omega_\ell^{\gamma*(1)}(\pi)\right)^{-1} \leq \sup_{\pi > \pi_0} [N^{-1}||\widetilde{\Psi}_{LS,\ell}(\pi)||^2]^d = O_p(C_{NT}^{-2d}) \text{ for } \ell > r_b. \tag{D.50}$$

Thus, for $\ell > r_b$,

$$\begin{aligned} N^{-1/2} C_{NT}^{-1} (\phi_\ell^\gamma)^{-1} &\leq \underline{\kappa}_2^{-1} C_{NT}^d \left(\mathbb{E}_\xi [\omega_\ell^{\gamma*(1)}(\xi) \mathcal{I}_{\{\xi \geq \pi_0\}}]\right)^{-1} \\ &\leq \underline{\kappa}_2^{-1} C_{NT}^d \left(\inf_{\pi > \pi_0} \left(\omega_\ell^{\gamma*(1)}(\pi)\right) \mathbb{E}_\xi [\mathcal{I}_{\{\xi \geq \pi_0\}}]\right)^{-1} \\ &= \frac{C_{NT}^d \sup_{\pi > \pi_0} \left(\omega_\ell^{\gamma*(1)}(\pi)\right)^{-1}}{\underline{\kappa}_2 \mathbb{E}_\xi [\mathcal{I}_{\{\xi \geq \pi_0\}}]} = O_p(C_{NT}^{-d}), \end{aligned} \tag{D.51}$$

following from (D.50) 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

$$\Pr(\widehat{r}_b^{(1)}(\pi_0) \leq r_b) \to 1. \tag{D.52}$$

When $r_b > r_a$, (D.31) and (D.52) imply that

$$\Pr(\widehat{r}_b^{(1)} = \min_{\pi \in \Pi} \widehat{r}_b^{(1)}(\pi) = r_b) \to 1. \tag{D.53}$$

On the other hand, if $r_b = r_a$, we can use (D.52) to deduce that

$$\Pr(\min_{\pi \in \Pi} \widehat{r}_b^{(1)}(\pi) \leq r_a) \to 1, \tag{D.54}$$

which together with the definition of $\widehat{r}_b^{(1)}$ and (D.37) implies that

$$\Pr(\widehat{r}_b^{(1)} = \widehat{r}_a^{(1)} = r_b) \to 1. \tag{D.55}$$

This completes the proof of Step 4.

Step 5. We show

$$\Pr(\widehat{r}_a^{(2)} = r_a) \to 1 \text{ and } \Pr(\widehat{r}_b^{(2)} = r_b) \to 1. \tag{D.56}$$

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 (D.7). 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 $\{\widehat{r}_a^{(1)} = r_a$ 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,

$$\Pr(\widehat{\Gamma}^{(2)}(\pi_0) \neq 0) \to 1. \tag{D.57}$$

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 (D.39) for the second-step estimator, which holds by the same arguments as in Step 3 conditioning on the event $\{\widehat{r}_a^{(1)} = r_a$ and $\widehat{r}_b^{(1)} = r_b\}$. Following Steps 3 and 4, this event occurs w.p.a.1.

The result in (D.57) and Step 5 together imply that $\Pr(\widehat{\mathcal{S}} = 1) \to 1$ if $\mathcal{S}_0 = 1$.

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

$$\Pr(\sup_{\pi\in\Pi} ||\widehat{\Gamma}^{(2)}(\pi)|| = 0) \to 1. \tag{D.58}$$

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

$$N^{-1/2}||\widehat{\Lambda}^{(2)}(\pi) - \Lambda^*|| = O_{p\pi}(\bar{b}_\Lambda + C_{NT}^{-1}) \text{ and } N^{-1/2}||\widehat{\Gamma}^{(2)}(\pi) - \Gamma^*|| = O_{p\pi}(\bar{b}_\Lambda + C_{NT}^{-1}), \tag{D.59}$$

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 (D.30). Revoking the proof of Theorem 1 with π_0 replaced by π, a sufficient condition for (D.58) is

$$N^{-1/2}C_{NT}^{-1} = o_p(\phi_\ell^\gamma) \text{ for } \ell = 1, \ldots, k, \tag{D.60}$$

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 $\{\widehat{r}_a^{(1)} = r_a$ 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

$$\phi_\ell^\gamma = \mathbb{E}_\xi[\beta_{NT}(\xi)\omega_\ell^{\gamma*(2)}(\xi)] \geq \underline{\kappa}_2 N^{-1/2}C_{NT}^{-d-1}\mathbb{E}_\xi[\omega_\ell^{\gamma*(2)}(\xi)]. \tag{D.61}$$

Using the formula of $\omega_\ell^{\gamma*(2)}(\pi)$ in (D.7), for $\ell > r_a$,

$$\left(\omega_\ell^{\gamma*(2)}(\pi)\right)^{-1} = \left(\omega_\ell^{\gamma*(1)}(\pi)\right)^{-1} \leq \left(N^{-1}||\widetilde{\Gamma}_{\ell,LS}(\pi)||^2\right)^d = O_{p\pi}(C_{NT}^{-2d}) \tag{D.62}$$

w.p.a.1, where the last equality holds by arguments analogous to (C.55). On the other hand, for $\ell \leq r_a$,

$$\left(\omega_\ell^{\gamma*(2)}(\pi)\right)^{-1} \leq \left(N^{-1}||\widetilde{\Psi}_{\ell,LS}^{-}(\pi)\mathrm{w}(\pi) - \widetilde{\Lambda}_{\ell,LS}(\pi)||^2\right)^d = O_{p\pi}(C_{NT}^{-2d}) \tag{D.63}$$

w.p.a.1, where the equality follows from arguments analogous to (C.63) under Assumption R*(i). Combining the results in (D.62) and (D.63), we deduce that

$$\sup_{\pi\in\Pi}\left(\omega_\ell^{\gamma*(2)}(\pi)\right)^{-1} = O_p(C_{NT}^{-2d}) \text{ for } \ell = 1,\ldots,k. \tag{D.64}$$

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

$$\begin{aligned} N^{-1/2}C_{NT}^{-1}(\phi_\ell^\gamma)^{-1} &\leq \underline{\kappa}_2^{-1}C_{NT}^d\left(\mathbb{E}_\xi[\omega_\ell^{\gamma*(2)}(\xi)]\right)^{-1} \\ &\leq \underline{\kappa}_2^{-1}C_{NT}^d\left(\inf_{\pi\in\Pi}\left(\omega_\ell^{\gamma*(2)}(\pi)\right)\right)^{-1} \\ &= \underline{\kappa}_2^{-1}C_{NT}^d\sup_{\pi\in\Pi}\left(\omega_\ell^{\gamma*(2)}(\pi)\right)^{-1} = O_p(C_{NT}^{-d}), \end{aligned} \tag{D.65}$$

following from (D.61) and (D.64). The condition in (D.60) follows from (D.65), and it is sufficient for the desired result. Therefore, if $\mathcal{S}_0 = 0$, we have $\Pr(\widehat{\mathcal{S}}_0 = 0) \to 1$. This completes the proof. $\square$