Testing for common structures in high-dimensional factor models [*,†,‡]

Marie-Christine Düker
FAU Erlangen-Nürnberg

Vladas Pipiras
University of North Carolina

March 28, 2024

Abstract

This work proposes a novel procedure to test for common structures across two high-dimensional factor models. The introduced test allows to uncover whether two factor models are driven by the same loading matrix up to some linear transformation. The test can be used to discover inter individual relationships between two data sets. In addition, it can be applied to test for structural changes over time in the loading matrix of an individual factor model. The test aims to reduce the set of possible alternatives in a classical change-point setting. The theoretical results establish the asymptotic behavior of the introduced test statistic. The theory is supported by a simulation study showing promising results in empirical test size and power. A data application investigates changes in the loadings when modeling the celebrated US macroeconomic data set of Stock and Watson.

1 Introduction

High-dimensional factor models are used to describe data which are driven by a relatively small number of latent time series. It is assumed that the temporal and cross-sectional dependence in the data can be attributed to these so-called common factors. The co-movements of the observed component time series are assumed to be driven by these factors and their dynamic structures. This kind of behavior is reasonable to assume in many applications (Sargent and Sims, 1977) and provides an effective tool for dimension reduction in high-dimensional time series. Originally introduced in finance (Markowitz, 1952; Cox and Ross, 1976; Ross, 1977; Feng, Giglio, and Xiu, 2020), factor models have found their way into a wide range of disciplines including economics (Stock and Watson, 2002, 2009; Han, 2018), psychometrics (Timmerman and Kiers, 2003; Song and Zhang, 2014) and genomics (Carvalho, Chang, Lucas, Nevins, Wang, and West, 2008).

This work focuses on discovering common structures in two high-dimensional factor models from their respective time series data. In many of the disciplines mentioned above, one is not only interested in studying intra individual differences for one factor model but also inter individual similarities across factor models. Examples include resemblances between macroeconomic indices of different countries or behavioral/neurological measurements for different individuals. While it is unlikely for two high-dimensional series to share the exact same loading matrix, one can ask whether they share a common structure in the form of linear combinations of loading matrices. We

[*]AMS subject classification. Primary: 62H25, 62M10. Secondary: 62F03.

[†]Keywords: High-dimensional time series, factor models, multi-task learning, change-point analysis.

[‡]The authors would like to thank the participants of the NBER-NSF Time Series Conference 2023 in Montreal for their helpful comments and suggestions. Marie Düker's reseach was supported by NSF grant DMS-1934985, and Vladas Pipiras acknowledges NSF grants DMS-2113662 and DMS-2134107.

propose a novel procedure allowing one to test whether the loadings in one series can be expressed as a linear combination of the other.

The questions of interest in high-dimensional factor model analysis are manyfold and have led to several lines of research in the statistics/econometrics literature. In particular, the estimation of the loading matrix and the latent factors is well understood and their estimators' theoretical properties have been studied in several works; see Bai (2003), Bai and Ng (2008), Lam, Yao, and Bathia (2011), Doz, Giannone, and Reichlin (2011, 2012). Other questions of interest include estimating the number of factors (Bai and Ng, 2002; Onatski, 2009; Lam and Yao, 2012; Li, Li, and Shi, 2017), high-dimensional covariance estimation and sparsity (Fan, Fan, and Lv, 2008; Fan, Masini, and Medeiros, 2023), using a factor model approach to describe multivariate count data (Wedel, Böckenholt, and Kamakura, 2003; Jung, Liesenfeld, and Richard, 2011; Wang and Wang, 2018) and extensions to networks (Bräuning and Koopman, 2016, 2020).

The questions considered in this work are closely related to change-point analysis in high-dimensional factor models. Recent literature has proposed ways to test for changes in the loadings over time. Changes in the loading matrix of a factor model are observationally equivalent to changes in the second-order structure of their factors. This way, a high-dimensional change-point problem can be reduced to a low-dimensional problem. Han and Inoue (2015, 2014), Bai, Duan, and Han (2022) proposed respectively a Wald type test and a likelihood ratio test. Similar approaches have been pursued in Chen, Dolado, and Gonzalo (2014), Ma and Su (2018), Duan, Bai, and Han (2023). More recently, Barigozzi, Cho, and Fryzlewicz (2018) studied the problem of multiple change-points and proposed a wavelet based approach. Baltagi, Kao, and Wang (2021) generalized the method in Han and Inoue (2015) to test for multiple breaks in the loading matrix. (We use the terms "change" and "break" throughout the paper interchangeably.) Su and Wang (2017) established a method to estimate the latent factors and time-varying factor loadings simultaneously.

Much less attention has been given to multi-subject analysis. In many applications, the objective is to estimate several unknowns corresponding to related models; this problem is often referred to as multi-task learning in the machine learning literature. When multivariate repeated measurements are collected from multiple subjects, there is substantial interest in learning the same model parameters across all subjects. For factor models, this has lead to different methods of estimating loading matrices across subjects; see Timmerman and Kiers (2003), De Roover, Ceulemans, Timmerman, Vansteelandt, Stouten, and Onghena (2012), Song and Zhang (2014). Multi-subject learning is slightly better understood for multiple regression problems (Jalali, Ravikumar, and Sanghavi, 2013; Ollier and Viallon, 2014; Gross and Tibshirani, 2016) and multiple vector autoregression models (Fisher, Kim, Fredrickson, and Pipiras, 2022), where one assumes the presence of common effects in the regression vectors across individuals.

Even though there is substantial interest in multi-subject models in the applied sciences, the statistical literature on multi-subject factor models is scarce. We put forward the framework of discovering inter connections for two-subject factor models. A modification of our test can also address situations when the two factor models are driven by different numbers of factors. Our method can also be applied in change-point settings, assuming that structural changes occur over time. The existing literature tests no change in the loadings against the alternative hypothesis that a non-negligible portion of the cross sections have a break in their loadings. Our test allows to discover what kind of change occurs and effectively reduces the number of alternatives.

The rest of this paper is organized as follows. Section 2 formally introduces the hypothesis testing problem for two-subject factor models and for changes in the loading matrix over time. Section 3 motivates and introduces the test statistic including its theoretical properties. Section 4 studies the behavior under the alternative. We also present some variations of our approach and a discussion in Section 5. Section 6 provides simulation results giving some insights into the numerical performance of our test. An application to real data can be found in Section 7 and we conclude with Section 8. Appendices A–D contain the proofs and are complemented by a supplementary document Düker and Pipiras (2023).

Notation: For the reader's convenience, notation used throughout the paper is collected here. Let $\lambda_1(A) \geq \ldots \geq \lambda_k(A)$ denote the eigenvalues of a symmetric matrix A. The maximum and minimum eigenvalues of A are then denoted by $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$, respectively. A range of different norms are used, including the maximum, the spectral and the Frobenius norm, defined respectively as $\|A\|_{\max} = \max_{1\leq i,j\leq d} |A_{ij}|$, $\|A\| = \sqrt{\lambda_{\max}(A'A)}$ and $\|A\|_F^2 = \sum_{i,j=1}^d |A_{ij}|^2$ for a matrix $A = (A_{ij})_{i,j=1,\ldots,d} \in \mathbb{R}^{d\times d}$. For the vectorized version of a matrix A, we write $\mathrm{vec}(A)$. The vec operator transforms a matrix into a vector by stacking its columns one underneath the other. Similarly, $\mathrm{vech}(A)$ denotes the vector that is obtained from $\mathrm{vec}(A)$ by eliminating all supradiagonal elements of A.

2 Hypothesis testing problem

The hypothesis testing problem studied in this paper is motivated by two different subject areas. On the one hand, we study multi-task learning which is concerned with discovering structures across multiple models; see Section 2.1. On the other hand, we are interested in structural changes in the loadings of factor models over time; see Section 2.2. Finally, we formulate the hypothesis testing problem in Section 2.3.

2.1 Two-subject factor model

Multi-subject learning aims to study inter individual differences and similarities in intra individual variability. The concept is formalized here for two subjects.

Let $X_t^1 = (X_{1,t}^1, \ldots, X_{d,t}^1)'$, $t \in \mathbb{Z}$, and $X_t^2 = (X_{1,t}^2, \ldots, X_{d,t}^2)'$, $t \in \mathbb{Z}$, denote two d-dimensional time series. Here, the prime denotes transpose. Both series are assumed to follow a factor model representation such that

$$\begin{aligned} X_t^1 &= \Lambda_1 F_t^1 + \varepsilon_t^1, \quad t = 1, \ldots, T_1, \\ X_t^2 &= \Lambda_2 F_t^2 + \varepsilon_t^2, \quad t = 1, \ldots, T_2, \end{aligned} \tag{2.1}$$

where Λ_1, Λ_2 are $d \times r$ loading matrices with $r < d$, $F_t^k = (F_{1,t}^k, \ldots, F_{r,t}^k)'$, $t \in \mathbb{Z}$, $k = 1, 2$, are r-vector time series of latent factors and $\varepsilon_t^k = (\varepsilon_{1,t}^k, \ldots, \varepsilon_{d,t}^k)'$, $t \in \mathbb{Z}$, $k = 1, 2$, are idiosyncratic errors which are assumed to be independent of the factors F_t^k. We will refer to (2.1) as a two-subject factor model. A typical assumption on the factors is to suppose that they follow a VAR(p) model

$$F_t = \Psi_1 F_{t-1} + \ldots + \Psi_p F_{t-p} + e_t, \quad t \in \mathbb{Z},$$

which is assumed to be stable. The series X_t^1 and X_t^2 are assumed to be stationary.

There has been a lot of research on how to estimate the loading matrices Λ_1 and Λ_2 jointly by imposing some common structure. A typical objective function is

$$\sum_{k=1}^{2} \sum_{t=1}^{T} \|X_t^k - \Lambda F_t^k\|_F^2 \quad \text{subject to} \quad \frac{1}{T} \sum_{t=1}^{T} F_t^k F_t^{k'} = D_k \Sigma D_k \quad \text{for} \quad k = 1, 2, \tag{2.2}$$

and diagonal matrices D_k, $k = 1, 2$, which may be different for each subject and some nonsingular matrix Σ which is assumed to be shared across subjects; see Timmerman and Kiers (2003), De Roover et al. (2012). Note that the constraint in (2.2) is equivalent to having

$$\Lambda_k = \Lambda D_k \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T} F_t^k F_t^{k\prime} = \Sigma \quad \text{for} \quad k = 1, 2. \tag{2.3}$$

While the problem (2.2) has drawn some interest in terms of optimization, there is no literature verifying that the constraint of having some common structure is a reasonable assumption. Our goal here is to develop a test for whether the data admit this kind of structure in the first place. The hypothesis we are interested in is whether the two loading matrices Λ_1, Λ_2 in (2.1) share common structure similar to (2.3). More specifically, can one of the matrices be represented as a linear combination of the other?

2.2 Structural changes over time

The second motivation for our testing problem comes from the literature on testing for structural changes in loading matrices of a single factor model over time. Since Han and Inoue (2015) proposed a classical change-point test to test the null hypothesis of no change against the alternative of structural changes in the factor loadings by applying a CUSUM type statistic, the related literature has grown significantly.

Consider the following factor model that allows for a structural break in the factor loadings:

$$X_t = \begin{cases} \Lambda_1 F_t + \varepsilon_t & 1 \le t \le \lfloor \pi T \rfloor, \\ \Lambda_2 F_t + \varepsilon_t & \lfloor \pi T \rfloor + 1 \le t \le T, \end{cases} \tag{2.4}$$

where, for $\pi \in (0,1)$, $\lfloor \pi T \rfloor$ is a possibly unknown break date. The change-point literature typically distinguishes between three alternatives, expressed in terms of column spaces. Recall that the column space of a matrix $A \in \mathbb{R}^{d\times r}$ is the set of all linear combinations of the columns in A. More formally, for $A = (a_1, \dots, a_r)$, $\mathrm{col}(A) = \{\sum_{i=1}^r \mu_i a_i \mid \mu_i \in \mathbb{R}\}$. The three alternatives are as follows:

Type 1: The columns of Λ_1 and Λ_2 are linearly independent, that is, $\mathrm{col}(\Lambda_1) \cap \mathrm{col}(\Lambda_2) = \{0\}$.

Type 2: Loadings undergo a rotational change, i.e. the columns of Λ_1 and Λ_2 are linearly dependent or, equivalently, $\mathrm{col}(\Lambda_1) = \mathrm{col}(\Lambda_2)$.

Type 3: Some (but not all) columns of Λ_1 and Λ_2 are linearly independent, that is, $\mathrm{col}(\Lambda_1) \cap \mathrm{col}(\Lambda_2) \neq \{0\}$ and $\mathrm{col}(\Lambda_1) \neq \mathrm{col}(\Lambda_2)$.

A rotational change as described by Type 2 can be written as $\Lambda_2 = \Lambda_1 \Phi$ for some nonsingular matrix Φ (this assumes that Λ_1 and Λ_2 have full column rank). Suppose further that $\mathrm{E}\, F_t F_t' = \Sigma_F$ for all $t = 1, \dots, T$. Then, a rotational change by the nonsingular matrix Φ is observationally equivalent to no change in the loadings and a change in the second order structure of the factors:

$$\mathrm{E}\, F_t F_t' = \begin{cases} \Sigma_F & 1 \le t \le \lfloor \pi T \rfloor, \\ \Phi \Sigma_F \Phi' & \lfloor \pi T \rfloor + 1 \le t \le T. \end{cases} \tag{2.5}$$

Note that the different types of breaks are not fully characterized by Types 1–3. For instance, given $\Sigma_F = I_r$ and Φ being an orthogonal matrix, we have $\Sigma_F = \Phi \Sigma_F \Phi'$. Therefore, this is not considered a change. Whenever a Type 2 change is meant, we need to assume $\Sigma_F \neq \Phi \Sigma_F \Phi'$. This is formalized in Assumption 10 in Han and Inoue (2015).

While the change-point literature tends to formalize those three alternatives, there is currently no way of detecting them separately. We propose a test for rotational changes in the loadings. To be more precise, suppose a regular change-point test as that of Han and Inoue (2015) rejects the hypothesis of no change and one estimates a break point π^*. Then, the occurring change can follow either of the three types above. We aim to test for the hypothesis of a change of Type 2 against the alternative of any other kinds of changes. In other words, can one test whether the change in the loadings follows a Type 2 break?

2.3 Hypothesis testing problem

The formulation of a Type 2 break in (2.5) resembles the constraint of interest in the multi-task learning literature as formalized in (2.2). While multi-task learning aims to find structural similarities, the change-point problem seeks for structural differences. Both problems reduce to the same hypothesis.

For the remainder of the paper we focus on the two-subject model (2.1) but have the change-point problem (2.4) in mind as another application. We consider

$$\begin{aligned} X_t^1 &= \Lambda_1 F_t^1 + \varepsilon_t^1, \;\; t = 1, \dots, T_1, \\ X_t^2 &= \Lambda_2 F_t^2 + \varepsilon_t^2, \;\; t = 1, \dots, T_2. \end{aligned} \tag{2.6}$$

Then, the null hypothesis can be formalized as

$$H_0 : \Lambda_k = \Lambda \Phi_k \text{ for } k = 1, 2. \tag{2.7}$$

Here, the matrix $\Lambda \in \mathbb{R}^{d \times r}$ is a matrix of full column rank and the matrices $\Phi_k \in \mathbb{R}^{r \times r}$, $k = 1, 2$, are nonsingular. Set $\Lambda_k' = (\lambda_{k,1}, \dots, \lambda_{k,d})$ with $\lambda_{k,i} \in \mathbb{R}^r$ for $k = 1, 2$. We want to test against the alternative hypothesis that a non-negligible portion of the cross sections have linearly independent loadings, i.e.,

$$H_1 : \lambda_{1,i}' \neq \lambda_{2,i}' \Phi \quad \text{for some } i = 1, \dots, d,$$

and all nonsingular matrices $\Phi \in \mathbb{R}^{r \times r}$.

Before moving on to proposing a test for the stated hypothesis problem, we pause here to get some intuition for what (2.7) means for data which might follow this model. Consider the following simple example.

Example 2.1. Let $\Phi_1 = I_r$ and $\Phi_2 = \operatorname{diag}(\phi_1, \dots, \phi_r)$ such that

$$\Lambda_1 = \Lambda \quad \text{and} \quad \Lambda_2 = \Lambda \operatorname{diag}(\phi_1, \dots, \phi_r). \tag{2.8}$$

In addition, we assume that $\operatorname{E} F_t^k F_t^{k'} = I_r$, $k = 1, 2$. Recall that (2.8) is observationally equivalent to differences in the second-order structure of the factors, i.e.

$$\operatorname{E} F_t^1 F_t^{1'} = I_r \quad \text{and} \quad \operatorname{E} F_t^2 F_t^{2'} = \operatorname{diag}(\phi_1^2, \dots, \phi_r^2).$$

This setting means that only the volatility of the individual factors changes over time. We suspect that this type of change explains why our test does not reject in our data application; see Section 7 for more details.

3 Testing procedure

In this section, we detail our method to test the null hypothesis (2.7). In particular, we give a motivation for our test statistic before formulating it.

We first recall estimation of loadings and factors in individual factor models based on principal component analysis (PCA). In order to estimate the loadings and factors in the two-subject factor model (2.6), we concatenate the series across time and apply PCA following Bai (2003). For $k = 1, 2$, set

$$\begin{pmatrix} X_1^{k'} \\ \vdots \\ X_{T_k}^{k'} \end{pmatrix} = \begin{pmatrix} F_1^{k'} \\ \vdots \\ F_{T_k}^{k'} \end{pmatrix} \Lambda_k' + \begin{pmatrix} \varepsilon_1^{k'} \\ \vdots \\ \varepsilon_{T_k}^{k'} \end{pmatrix} \tag{3.1}$$

and for short $X^k = F^k\Lambda_k' + \varepsilon^k$. Write $\widehat{\Sigma}_{T_k}^k = X^k X^{k'}$ and let $\widehat{Q}_r^k$ be the r eigenvectors corresponding to the r largest eigenvalues of $\widehat{\Sigma}_{T_k}^k$. Then, the PCA estimators are

$$\widehat{\Lambda}_k' = \frac{1}{T_k}\widehat{F}^{k'} X^k, \quad \widehat{F}^k = \sqrt{T_k}\widehat{Q}_r^k. \tag{3.2}$$

These estimators are consistent under appropriate assumptions outlined in Section 3.3 below; see Bai (2003).

3.1 Motivation

The null (2.7) states that the columns of Λ_k lie in the subspace generated by the columns of Λ. Therefore, the null hypothesis can be reformulated as

$$H_0 : P_0\Lambda_k = \Lambda_k, \quad k = 1, 2,$$

where P_0 is the projection matrix onto the subspace generated by the columns of Λ. Since Λ has linearly independent columns, the projection matrix is given by $P_0 = \Lambda(\Lambda'\Lambda)^{-1}\Lambda'$.

This observation inspires the following test. Define the projection matrices of the individual loadings as

$$P_{0,k} = \Lambda_k(\Lambda_k'\Lambda_k)^{-1}\Lambda_k', \quad k = 1, 2. \tag{3.3}$$

The projection matrices $P_{0,k}$ in (3.3) can be estimated through the PCA estimator (3.2) of the loading matrices as

$$\widehat{P}_k = \widehat{\Lambda}_k(\widehat{\Lambda}_k'\widehat{\Lambda}_k)^{-1}\widehat{\Lambda}_k', \quad k = 1, 2. \tag{3.4}$$

Note that projection matrices are invariant under any transformation with a nonsingular matrix $\Phi \in \mathbb{R}^{r\times r}$, i.e.

$$\Lambda_k\Phi(\Phi'\Lambda_k'\Lambda_k\Phi)^{-1}\Phi'\Lambda_k' = \Lambda_k(\Lambda_k'\Lambda_k)^{-1}\Lambda_k', \quad k = 1, 2. \tag{3.5}$$

In particular, under the hypothesis (2.7),

$$P_{0,k} = \Lambda_k(\Lambda_k'\Lambda_k)^{-1}\Lambda_k' = \Lambda\Phi_k(\Phi_k'\Lambda'\Lambda\Phi_k)^{-1}\Phi_k'\Lambda' = \Lambda(\Lambda'\Lambda)^{-1}\Lambda', \quad k = 1, 2. \tag{3.6}$$

Introduce the notation

$$X_{1:T_k}^k = (X_1^k, \ldots, X_{T_k}^k) \in \mathbb{R}^{d\times T_k}, \quad k = 1, 2.$$

The general idea is to split the series $X_{1:T_k}^k$ into two halves and to apply an appropriate transformation such that there is no change in the loadings under the null, but there is a change under the alternative. We propose the following transformation

$$Y_{1:T_2} = \left(\widehat{P}_I^1 X_{1:T_2/2}^2, \widehat{P}_I^2 X_{(T_2/2+1):T_2}^2\right) \quad \text{with } \widehat{P}_I^k = \frac{1}{2}(\widehat{P}_k + I_d), \ k = 1, 2, \tag{3.7}$$

where $\widehat{P}_1, \widehat{P}_2$ are as in (3.4). Note that the roles of the first and the second series are interchangeable. To motivate our choice of transforming the series $X^2_{1:T_2}$ by $\frac{1}{2}(\widehat{P}_1 + I_d)$ and $\frac{1}{2}(\widehat{P}_2 + I_d)$, we consider the transformation at the population level. Then, by (3.6),

$$P_{0,1}X_t^2 \approx P_{0,1}\Lambda_2 F_t^2 = P_{0,1}\Lambda\Phi_2 F_t^2 = \Lambda(\Lambda'\Lambda)^{-1}\Lambda'\Lambda\Phi_2 F_t^2 = \Lambda\Phi_2 F_t^2 = \Lambda_2 F_t^2, \quad \text{for } t = 1, \ldots, T_2/2,$$
$$P_{0,2}X_t^2 \approx P_{0,2}\Lambda_2 F_t^2 = P_{0,2}\Lambda\Phi_2 F_t^2 = \Lambda(\Lambda'\Lambda)^{-1}\Lambda'\Lambda\Phi_2 F_t^2 = \Lambda_2 F_t^2, \quad \text{for } t = T_2/2+1, \ldots, T_2$$

and therefore

$$\begin{aligned} &\frac{1}{2}(P_{0,1} + I_d)X_t^2 \approx \Lambda_2 F_t^2, \quad \text{for } t = 1, \ldots, T_2/2, \\ &\frac{1}{2}(P_{0,2} + I_d)X_t^2 \approx \Lambda_2 F_t^2, \quad \text{for } t = T_2/2+1, \ldots, T_2. \end{aligned} \tag{3.8}$$

Since $P_{0,1} = P_{0,2}$ under the null hypothesis, the idiosyncratic errors of the transformed series $Y_{1:T_2}$ are still stationary. In addition, the errors have positive definite covariances whenever the original series has positive definite covariance since

$$\begin{aligned} &(P_{0,1} + I_d)\,\mathrm{E}(\varepsilon_t^2\varepsilon_t^{2'})(P_{0,1} + I_d)' = (P_0 + I_d)\Sigma_\varepsilon(P_0 + I_d)', \\ &(P_{0,2} + I_d)\,\mathrm{E}(\varepsilon_t^2\varepsilon_t^{2'})(P_{0,2} + I_d)' = (P_0 + I_d)\Sigma_\varepsilon(P_0 + I_d)'; \end{aligned} \tag{3.9}$$

see also Remark 3.2 below.

Due to (3.8), the series $Y_{1:T_2}$ in (3.7) is expected to have the same loadings and factors under the null as $X^2_{1:T_2}$. However, given (3.9), the errors have a different covariance matrix than the original series.

In contrast, when the null hypothesis is not satisfied, we get

$$\begin{aligned} &P_{0,1}X_t^2 \approx P_{0,1}\Lambda_2 F_t^2 = \Lambda_1(\Lambda_1'\Lambda_1)^{-1}\Lambda_1'\Lambda_2 F_t^2 \neq \Lambda_2 F_t^2, \quad \text{for } t = 1, \ldots, T_2/2, \\ &P_{0,2}X_t^2 \approx P_{0,2}\Lambda_2 F_t^2 = \Lambda_2(\Lambda_2'\Lambda_2)^{-1}\Lambda_2'\Lambda_2 F_t^2 = \Lambda_2 F_t^2, \quad \text{for } t = T_2/2+1, \ldots, T_2. \end{aligned} \tag{3.10}$$

The illustrated behavior of $Y_{1:T_2}$ at the population level in (3.8) and (3.10) suggests that for the transformed series $Y_{1:T_2}$ our hypothesis testing problem is equivalent to testing for a change at time $T_2/2$ in the loadings. The roles of $X^1_{1:T_1}$ and $X^2_{1:T_2}$ are interchangeable. Depending on which series we base our test on, we can consider either

$$Y_t = \begin{cases} \frac{1}{2}(\widehat{P}_1 + I_d)X_t^2, & \text{for } t = 1, \ldots, T_2/2, \\ \frac{1}{2}(\widehat{P}_2 + I_d)X_t^2, & \text{for } t = T_2/2+1, \ldots, T_2 \end{cases} \tag{3.11}$$

or

$$Y_t = \begin{cases} \frac{1}{2}(\widehat{P}_1 + I_d)X_t^1, & \text{for } t = 1, \ldots, T_1/2, \\ \frac{1}{2}(\widehat{P}_2 + I_d)X_t^1, & \text{for } t = T_1/2+1, \ldots, T_1. \end{cases} \tag{3.12}$$

Under the hypothesis, i.e., when there is no change in the loadings, the true loadings and factors of the transformed series (3.11) and (3.12) should respectively coincide with the true loadings and factors of the series X_t^2 and X_t^1.

To avoid complicated notation, we base our test on the second series X_t^2 and write from now on

$$F_t \ \text{ instead of } \ F_t^2. \tag{3.13}$$

We keep Λ_2 to denote the corresponding loading matrix and to avoid any confusion with Λ as used to characterize the null hypothesis in (2.7). We also assume from now on that

$$T = T_1 = T_2.$$

Several quantities are defined above before and after the "change-point" $T/2$. We shall sometimes distinguish between them below by using superscript b for "before" and a for "after". For example, in view of (3.13), we shall write

$$F^b = (F_1, \dots, F_{T/2})', \;\; F^a = (F_{T/2+1}, \dots, F_T)'. \tag{3.14}$$

In order to test for changes in the loadings of the transformed series $\{Y_t\}_{t=1,\dots,T}$ we can now employ a change-point test as introduced in Han and Inoue (2015). This is outlined in more detail in the next section. See also Baek, Gates, Leinwand, and Pipiras (2021) for an application of the test in Han and Inoue (2015) to testing for changes in the autocovariances of the latent factors.

3.2 The test statistic

Recall from (3.7) the transformed series $Y_{1:T_2} = \left(\widehat{P}_I^1 X_{1:T/2}^2, \widehat{P}_I^2 X_{(T/2+1):T}^2\right)$. To employ the change-point test of Han and Inoue (2015), we first concatenate (3.11) across time to get for short

$$Y := Y_{1:T}' = \begin{pmatrix} F_{1:T/2}'\Lambda_2'\widehat{P}_I^1 \\ F_{(T/2+1):T}'\Lambda_2'\widehat{P}_I^2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{1:T/2}'\widehat{P}_I^1 \\ \varepsilon_{(T/2+1):T}'\widehat{P}_I^2 \end{pmatrix} =: \begin{pmatrix} F^b\Lambda_2'\widehat{P}_I^1 \\ F^a\Lambda_2'\widehat{P}_I^2 \end{pmatrix} + \begin{pmatrix} \varepsilon^b\widehat{P}_I^1 \\ \varepsilon^a\widehat{P}_I^2 \end{pmatrix} \tag{3.15}$$

with $F' := F_{1:T} = (F_1, \dots, F_T)$, $F_t \in \mathbb{R}^r$. The use of superscripts a and b was discussed before in (3.14). We follow the same procedure as in (3.1) but now for the transformed series (3.15). The PCA estimators are then based on $\widehat{\Sigma}_T = YY'$ such that

$$\widehat{F} = \sqrt{T}\widehat{Q}_r, \;\; \widehat{\Lambda}' = \frac{1}{T}\widehat{F}'Y, \tag{3.16}$$

where $\widehat{Q}_r$ are the r eigenvectors corresponding to the r largest eigenvalues of $\widehat{\Sigma}_T$.

We base the proposed test statistic on the pre- and post-sample means of $\widehat{F}_t\widehat{F}_t'$, where $\widehat{F}_t$ is the PCA estimator of the factors of our transformed series as stated in (3.16). Define

$$V(\widehat{F}) = \operatorname{vech}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T/2}\widehat{F}_t\widehat{F}_t' - \frac{1}{\sqrt{T}}\sum_{t=T/2+1}^{T}\widehat{F}_t\widehat{F}_t'\right). \tag{3.17}$$

To normalize appropriately, we use the following long-run variance estimator as suggested in Section 2.4 of Han and Inoue (2015),

$$\Omega(\widehat{F}) = \widehat{\Gamma}_0(\widehat{F}) + \sum_{j=1}^{T-1}\kappa\left(\frac{j}{b_T}\right)\left(\widehat{\Gamma}_j(\widehat{F}) + \widehat{\Gamma}_j'(\widehat{F})\right), \tag{3.18}$$

where κ is a kernel function, b_T the corresponding bandwidth and

$$\widehat{\Gamma}_j(\widehat{F}) = \frac{1}{T}\sum_{t=1+j}^{T}\operatorname{vech}(\widehat{F}_t\widehat{F}_t' - I_r)\operatorname{vech}(\widehat{F}_{t-j}\widehat{F}_{t-j}' - I_r)'. \tag{3.19}$$

Then, the Wald type test statistic can be defined as

$$W(\widehat{F}) = V(\widehat{F})'\Omega^{-1}(\widehat{F})V(\widehat{F}). \tag{3.20}$$

Our main theoretical contribution is to prove that the PCA estimators of the transformed series Y are still consistent estimators for the loadings and factors.

We show in Appendix B that $\widehat{F}$ behaves as FH or FH_0. Here,

$$H := (\Lambda_2'\Lambda_2/d)(F'\widehat{F}/T)\widehat{V}_r^{-1} \xrightarrow{\text{p}} H_0, \text{ as } d, T \to \infty, \tag{3.21}$$

where $\widehat{V}_r$ is an $r \times r$ diagonal matrix with the r largest eigenvalues of $\frac{1}{Td}\widehat{\Sigma}_T$ on the diagonal. The convergence (3.21) is shown in Appendix B. We also use $\Omega(FH_0)$ which is $\Omega(\cdot)$ as in (3.18) but evaluated at FH_0. Both quantities are shown to be consistent estimators of the true long-run variance

$$\Omega = \lim_{T\to\infty} \operatorname{Var}\left(\operatorname{vech}\left(\frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T} H_0' F_t F_t' H_0 - I_r\right)\right)\right). \tag{3.22}$$

Remark 3.1. We base our test on a CUSUM like statistic as considered in Han and Inoue (2015). However, it is certainly conceivable to use our transformation in combination with other change-point tests. For example, Bai et al. (2022) considered a likelihood ratio based test.

3.3 Assumptions

To study the limiting distribution of our Wald type test statistic, we state some assumptions. Those assumptions are standard in the factor model literature and ensure that the PCA estimators used in our test statistic are consistent.

In the factor model literature, there are different approaches of proving consistency of the PCA estimators. Those different approaches involve different assumptions. We follow here the approach in Doz et al. (2011). The assumptions therein are slightly different than those introduced by Stock and Watson (2002), Bai and Ng (2002) and Bai (2003) but have a similar role.

Note that all assumptions need to be satisfied for both series in (2.6). To avoid unnecessarily complicated notation we state the assumptions for a universal factor model

$$X_t = \Upsilon F_t + \varepsilon_t.$$

Assumption A.1. Suppose $\{X_t\}_{t=1,\dots,T}$ is stationary with $\operatorname{E} X_t = 0$ and $\operatorname{Var}(X_{j,t}) \le M$ for all $j = 1, \dots, d$ and $t = 1, \dots, T$, where the constant M does not depend on d.

Assumption A.2. $\{F_t\}$ and $\{\varepsilon_t\}$ are independent and admit Wold representations:

(i) $F_t = \sum_{k=0}^{\infty} C_k a_{t-k}$, with $\sum_{k=0}^{\infty} \|C_k\| < \infty$ and a white noise series $\{a_t\}$ with finite fourth moments, and $\{F_t\}$ has positive definite covariance matrix $\operatorname{E} F_t F_t' = \Sigma_F$.

(ii) $\varepsilon_t = \sum_{k=0}^{\infty} D_k b_{t-k}$, $\sum_{k=0}^{\infty} \|D_k\| < \infty$, and $\{b_t = (b_{1,t}, \dots, d_{d,t})'\}$ is a white noise series such that $\operatorname{E} b_{j,t}^4 \le M$ for all $j = 1, \dots, d$ and $t = 1, \dots, T$.

Assumption A.2(i) implies that $\frac{1}{T}\sum_{t=1}^{T} F_t F_t' \xrightarrow{\text{p}} \Sigma_F$ as $T \to \infty$; see Theorem 2 in Hannan (1970), p. 203. The convergence aligns with Assumption A in Bai (2003). A convenient way to parameterize the dynamics would be to further assume that the common factors follow a vector autoregression model.

The following Assumption A.3 imposes more restrictions on the correlation structure of idiosyncratic errors of the factor models. In particular, it collects the assumptions which go beyond the assumptions necessary to prove consistent estimation of the factors. Instead, they are needed to infer convergence of the long-run variance matrix in Proposition 3.3. Theses assumptions could also be inferred from more restrictive assumptions on the matrices D_k in Assumption A.2.

Assumption A.3. There exists a positive constant M such that for all d, T,

(i) $\frac{1}{d}\sum_{i,j=1}^{d} |\operatorname{E}(\varepsilon_{i,t_1}\varepsilon_{j,t_2})| \leq M$ for all $t_1, t_2 = 1, \dots, T$;

(ii) $\frac{1}{d^2}\operatorname{E}\|e_t'\varepsilon\Upsilon\|^4 \leq M$ for each $t = 1, \dots, T$ with e_t denoting the tth unit vector in $\mathbb{R}^T$ and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_T)'$;

(iii) $\operatorname{E}(\varepsilon_{i,t}^4) \leq M$ for all $i = 1, \dots, d$ and $t = 1, \dots, T$.

For the following assumptions, let Π_Υ denote the diagonal matrix whose diagonal entries are the eigenvalues of $\Upsilon'\Upsilon$ in decreasing order.

Assumption CR.1. There is a positive definite diagonal matrix $\widetilde{\Pi}$ such that $\|\Pi_\Upsilon/d - \widetilde{\Pi}\| \to 0$ with $\|\Pi_\Upsilon/d - \widetilde{\Pi}\| = \mathcal{O}\left(\frac{1}{\sqrt{d}}\right)$. In particular, one can infer that

(i) $\liminf_{d\to\infty} \frac{\lambda_{\min}(\Upsilon'\Upsilon)}{d} > 0$;

(ii) $\limsup_{d\to\infty} \frac{\lambda_{\max}(\Upsilon'\Upsilon)}{d} = \limsup_{d\to\infty} \frac{\|\Upsilon\|^2}{d} < \infty$.

Assumption CR.2. The eigenvalues of Π_Υ are distinct.

Assumption CR.3. Suppose $\limsup_{d\to\infty} \sum_{h\in\mathbb{Z}} \|\Gamma_\varepsilon(h)\| < \infty$ (with $\Gamma_Z(h) = \operatorname{E} Z_{t+h}Z_t'$ denoting the autocovariance function of a stationary mean-zero series $\{Z_t\}$).

Assumption CR.4. Suppose there is a $\bar{\lambda}$ such that $\|\Upsilon\|_{\max} \leq \bar{\lambda} < \infty$.

Remark 3.2. We conclude this section with a remark on our transformation (3.7) and the assumptions made in Doz et al. (2011) for consistent model estimation. As pointed out in (3.9), our transformation ensures that the error terms have a positive definite covariance matrix with high probability as long as the original representation has errors with positive definite covariance matrix. Doz et al. (2011) formalizes this requirement of positive definite covariances in the errors in their Assumption (CR4). However, this assumption does not seem to be necessary for consistent estimation of loading and factor matrices. The assumption seems to be only needed to prove consistency of the Kalman smoothing which is one of the subjects considered in Doz et al. (2011). We omit the assumption here but emphasize that our transformation can still accommodate further requirements on the errors.

We also note that our transformation expands the treatment under the alternative hypothesis. As outlined in Section 4, using a transformation including the identity matrix requires to distinguish among a larger set of different alternatives.

3.4 Limiting distribution under the null

In this section, we present our theoretical results, in particular, the asymptotic behavior of our test statistic (3.20) under the null hypothesis. For that we need a few more assumptions.

Assumption W.1. (Asymptotics)

(i) Suppose Ω in (3.22) is positive definite and

$$\|\Omega(FH_0) - \Omega\| = o_{\mathrm{p}}(1).$$

(ii) For $W(\cdot)$ in (3.20), suppose

$$W(FH_0) \xrightarrow{\mathrm{d}} \chi^2(r(r+1)/2),$$

where $\chi^2(r(r+1)/2)$ denotes the chi-squared distribution with $r(r+1)/2$ degrees of freedom.

Theorem 3 in Andrews (1993) provides conditions under which Assumption W.1(ii) is satisfied.

Assumption W.2. Suppose κ in (3.18) is the Bartlett kernel ($\kappa(x) = (1 - |x|)\mathbb{1}_{\{|x|\leq 1\}}$) and the bandwidth b_T satisfies $b_T = O\big(T^{1/3}\big)$ and $\frac{T^{\frac{2}{3}}}{d} \to 0$, as $d, T \to \infty$.

The following theorem states the asymptotic behavior of our test statistic.

Proposition 3.1. *Recall the definition of $W(\cdot)$ in (3.20) and suppose Assumptions A.1–A.3, CR.1–CR.4, W.1, W.2, and $\frac{\sqrt{T}}{d} \to 0$, as $d, T \to \infty$. Then, under the null hypothesis (2.7),*

$$W(\widehat{F}) \xrightarrow{\text{d}} \chi^2(r(r+1)/2),$$

where $\chi^2(r(r+1)/2)$ denotes the chi-squared distribution with $r(r+1)/2$ degrees of freedom.

Note that Proposition 3.1 requires $\frac{\sqrt{T}}{d} \to 0$, as $d, T \to \infty$. This assumption on the relationship between dimension and sample size is typical in the setting of testing for structural changes in high-dimensional factor models. In particular, it is imposed in Han and Inoue (2015). The assumption is also covered by the assumptions studied in random matrix theory. Random matrix theory typically distinguishes between $\frac{T}{d} \to c > 0$ and $\frac{T}{d} \to 0$ (also known as the ultra-high dimensional regime), both imply our assumption.

The proof of Proposition 3.1 requires convergence of $V(\cdot)$ in (3.17) and of the long-run variance $\widehat{\Omega}(\cdot)$ in (3.18). The two propositions below formalize that.

Proposition 3.2. *Recall the definition of $V(\cdot)$ in (3.17) and suppose Assumptions A.1–A.3(i), CR.1–CR.4 and $\frac{\sqrt{T}}{d} \to 0$, as $d, T \to \infty$. Then, under the null hypothesis (2.7),*

$$\|V(\widehat{F}) - V(FH_0)\| \xrightarrow{\text{p}} 0.$$

Proposition 3.3. *Recall the definition of $\Omega(\cdot)$ in (3.18) and suppose Assumptions A.1–A.3, CR.1–CR.4, W.2 and $\frac{\sqrt{T}}{d} \to 0$, as $d, T \to \infty$. Then, under the null hypothesis (2.7),*

$$\|\Omega(\widehat{F}) - \Omega(FH_0)\| \xrightarrow{\text{p}} 0.$$

The proofs of the results in this section can be found in the appendix.

4 Power against the alternative hypothesis

We consider the alternative that a portion of the cross sections have linearly independent loadings. In order to study the behavior of our test statistic under the alternative, we follow Han and Inoue (2015) and Duan et al. (2023) and rewrite the model in terms of so-called pseudo-factors.

The possible alternatives are of Types 1 and 3 as described in Section 2.2. As in Section 3.1, we base our analysis on $\{X_t^2\}$, using the transformed series $Y_{1:T_2}$ as introduced in (3.7). For simplicity, we assume that $T = T_1 = T_2$. Then,

$$Y_t = \begin{cases} \widehat{P}_I^1 \Lambda_2 F_t + \widehat{P}_I^1 \varepsilon_t =: \widetilde{\Theta}_1 F_t + \widetilde{\varepsilon}_t, & \text{for } t = 1, \dots, T/2, \\ \widehat{P}_I^2 \Lambda_2 F_t + \widehat{P}_I^2 \varepsilon_t =: \widetilde{\Theta}_2 F_t + \widetilde{\varepsilon}_t, & \text{for } t = T/2+1, \dots, T. \end{cases} \tag{4.1}$$

Recall also the behavior of $Y_{1:T}$ under the alternative at the population level from (3.10). Whenever we use the transformed series (4.1) for estimation, we keep in mind a population model of the form

$$Z_t = \begin{cases} P_I^1 \Lambda_2 F_t + \eta_t & =: \Theta_1 F_t + \eta_t, \quad \text{for } t = 1, \dots, T/2, \\ \Lambda_2 F_t + \eta_t & =: \Theta_2 F_t + \eta_t, \quad \text{for } t = T/2+1, \dots, T, \end{cases} \tag{4.2}$$

where P_I^1, P_I^2 are defined as the population counterparts of (3.7) such that the loadings behave according to (3.10) and with some error series η_t. The pre- and post-break covariances of η_t may be different. We use (4.2) to illustrate the behavior of the estimated loadings under the alternative when (4.1) is used for estimation. As we discuss next, pre- and post-break loadings in (4.2) admit different types of breaks depending on the behavior of Λ_1 and Λ_2.

Recall that $\Lambda_1 = \Lambda_2\Phi$ with nonsingular Φ under the null hypothesis or, equivalently, $\mathrm{col}(\Lambda_1) = \mathrm{col}(\Lambda_2)$. In the subsequent analysis, we aim to shed light on how the relationship of Λ_1 and Λ_2 under the alternative gets translated to $P_I^1\Lambda_2$ and Λ_2 in (4.2).

First, note that our null hypothesis of having a nonsingular matrix Φ such that $\Lambda_1 = \Lambda_2\Phi$, is equivalent to $P_I^1\Lambda_2 = \Lambda_2$ and $\mathrm{rk}(\Lambda_1'\Lambda_2) = r$. The null hypothesis implies the first relation since $P_{0,1} = P_{0,2}$, and the second relation since $\mathrm{rk}(\Lambda_1'\Lambda_2) = \mathrm{rk}(\Phi'\Lambda_2'\Lambda_2)$ and both Φ and $\Lambda_2'\Lambda_2$ have rank r. For the converse, we have

$$P_I^1\Lambda_2 = \frac{1}{2}\Lambda_1\left((\Lambda_1'\Lambda_1)^{-1}\Lambda_1'\Lambda_2\right) + \frac{1}{2}\Lambda_2 = \Lambda_2$$

and therefore $\Lambda_1\Psi = \Lambda_2$ with nonsingular $\Psi = (\Lambda_1'\Lambda_1)^{-1}\Lambda_1'\Lambda_2$.

Then, the alternative hypothesis, i.e., $\Lambda_1 \neq \Lambda_2\Phi$ for all nonsingular Φ, is equivalent to the condition that $P_I^1\Lambda_2 \neq \Lambda_2$ or $\mathrm{rk}(\Lambda_1'\Lambda_2) < r$. In fact, we will argue that *(i)*

$$P_I^1\Lambda_2 = \frac{1}{2}\Lambda_1\left((\Lambda_1'\Lambda_1)^{-1}\Lambda_1'\Lambda_2\right) + \frac{1}{2}\Lambda_2 \neq \Lambda_2 \tag{4.3}$$

and discuss whether *(ii)* $\mathrm{col}(P_I^1\Lambda_2) \cap \mathrm{col}(\Lambda_2) = \{0\}$ and *(iii)* $\mathrm{col}(P_I^1\Lambda_2) = \mathrm{col}(\Lambda_2)$. The different behaviors of the respective column spaces of $P_I^1\Lambda_2$ and Λ_2 are then associated with different types of changes in the model (4.2). That is, if *(i)* holds, then the change in (4.2) is of Type 1, if *(ii)* holds, then the change is of Type 3, and if *(i)* and *(ii)* do not hold, it is of Type 2. To investigate *(i)*–*(iii)*, we distinguish between whether $\mathrm{col}(\Lambda_1) \cap \mathrm{col}(\Lambda_2) = \{0\}$ (Type 1) or $\mathrm{col}(\Lambda_1) \cap \mathrm{col}(\Lambda_2) \neq \{0\}$ and $\mathrm{col}(\Lambda_1) \neq \mathrm{col}(\Lambda_2)$ (Type 3) in the original model.

Case 1: Suppose $\mathrm{col}(\Lambda_1) \cap \mathrm{col}(\Lambda_2) = \{0\}$. We consider the following subcases: $\mathrm{rk}(\Lambda_1'\Lambda_2) = r$, $= 0$ and $= s(\neq 0, r)$.

1a. $\mathrm{rk}(\Lambda_1'\Lambda_2) = r$: Then,

- *(i)* $P_I^1\Lambda_2 \neq \Lambda_2$: Arguing by contradiction, assume equality in (4.3) so that $P_I^1\Lambda_2 = \Lambda_2$. This yields $\Lambda_1 = \Lambda_2(\Lambda_1'\Lambda_2)^{-1}\Lambda_1'\Lambda_1$ which contradicts $\mathrm{col}(\Lambda_1) \cap \mathrm{col}(\Lambda_2) = \{0\}$.
- *(ii)* $\mathrm{col}(P_I^1\Lambda_2) \cap \mathrm{col}(\Lambda_2) = \{0\}$: By contradiction, suppose there are nonzero vectors $v_1, v_2 \in \mathbb{R}^r$ such that $P_I^1\Lambda_2 v_1 = \Lambda_2 v_2$. Then, $\frac{1}{2}\Lambda_1\widetilde{v}_1 = \Lambda_2(v_2 - \frac{1}{2}v_1)$ with $\widetilde{v}_1 = \left((\Lambda_1'\Lambda_1)^{-1}\Lambda_1'\Lambda_2\right)v_1$ which, since $\Lambda_1\widetilde{v}_1 \neq 0$ because Λ_1 has full column rank, contradicts $\mathrm{col}(\Lambda_1) \cap \mathrm{col}(\Lambda_2) = \{0\}$.

This corresponds to a change of Type 1 for the model (4.2).

1b. $\mathrm{rk}(\Lambda_1'\Lambda_2) = 0$: Then $\Lambda_1'\Lambda_2 = 0$ and (4.3) becomes *(i)* $P_I^1\Lambda_2 = \frac{1}{2}\Lambda_2 \neq \Lambda_2$. We can then infer that *(iii)* $\mathrm{col}(P_I^1\Lambda_2) = \mathrm{col}(\Lambda_2)$.

This corresponds to a change of Type 2 for the model (4.2).

1c. $\mathrm{rk}(\Lambda_1'\Lambda_2) = s \neq 0, r$: Then $\Lambda_1'\Lambda_2 = \alpha\beta'$ with $\alpha, \beta \in \mathbb{R}^{r\times s}$ and $\mathrm{rk}(\alpha) = \mathrm{rk}(\beta) = \mathrm{rk}(\Lambda_1'\Lambda_2)$.

- *(i)* $P_I^1\Lambda_2 \neq \Lambda_2$: Arguing by contradiction, equality in (4.3) gives $\Lambda_1\widetilde{\alpha} = \Lambda_2\beta(\beta'\beta)^{-1}$ with $\widetilde{\alpha} = (\Lambda_1'\Lambda_1)^{-1}\alpha$, contradicting $\mathrm{col}(\Lambda_1) \cap \mathrm{col}(\Lambda_2) = \{0\}$.

(ii) $\mathrm{col}(P_I^1\Lambda_2)\cap\mathrm{col}(\Lambda_2)\neq\{0\}$: We show that there are nonzero vectors $v_1,v_2\in\mathbb{R}^r$ such that $P_I^1\Lambda_2 v_1=\Lambda_2 v_2$. Choose $v_1=2v_2$. Then, $P_I^1\Lambda_2 v_1=\Lambda_1\widetilde{\alpha}\beta' v_2+\Lambda_2 v_2=\Lambda_2 v_2$ whenever $\beta' v_2=0$. We can find a nonzero vector v_2 such that $\beta' v_2=0$ since β is $r\times s$, β' is $s\times r$ and $s<r$.

(iii) $\mathrm{col}(P_I^1\Lambda_2)\neq\mathrm{col}(\Lambda_2)$: By contradiction, suppose there is a nonsingular matrix Φ such that $P_I^1\Lambda_2=\Lambda_2\Phi$. Then, $\frac{1}{2}\Lambda_1\widetilde{\alpha}\beta'=\Lambda_2(\Phi-\frac{1}{2}I_r)$ and, choosing a vector $v\in\mathbb{R}^r$ with $\widetilde{v}:=\widetilde{\alpha}\beta' v\neq 0$, we get $\frac{1}{2}\Lambda_1\widetilde{v}=\Lambda_2(\Phi-\frac{1}{2}I_r)v$, which contradicts $\mathrm{col}(\Lambda_1)\cap\mathrm{col}(\Lambda_2)=\{0\}$ or Λ_1 having full column rank.

This corresponds to the change of Type 3 for the model (4.2).

Case 2: Suppose $\mathrm{col}(\Lambda_1)\cap\mathrm{col}(\Lambda_2)\neq\{0\}$ and $\mathrm{col}(\Lambda_1)\neq\mathrm{col}(\Lambda_2)$. We consider the following subcases:

2a. $\mathrm{rk}(\Lambda_1'\Lambda_2)=r$: Then,

(i) $P_I^1\Lambda_2\neq\Lambda_2$: Arguing by contradiction, assume equality in (4.3) such that $P_I^1\Lambda_2=\Lambda_2$. This yields $\Lambda_1=\Lambda_2(\Lambda_1'\Lambda_2)^{-1}\Lambda_1'\Lambda_1$ which contradicts $\mathrm{col}(\Lambda_1)\neq\mathrm{col}(\Lambda_2)$.

(ii) $\mathrm{col}(P_I^1\Lambda_2)\cap\mathrm{col}(\Lambda_2)\neq\{0\}$: Since $\mathrm{col}(\Lambda_1)\cap\mathrm{col}(\Lambda_2)\neq\{0\}$, there are nonzero vectors $v_1,v_2\in\mathbb{R}^r$ such that $\Lambda_1 v_1=\Lambda_2 v_2$. Choose $w_1=2\Psi^{-1}v_1$ and $w_2=v_2+\Psi^{-1}v_1$ with $\Psi=(\Lambda_1'\Lambda_1)^{-1}\Lambda_1'\Lambda_2$. Then, $P_I^1\Lambda_2 w_1=\frac{1}{2}\Lambda_1\Psi w_1+\frac{1}{2}\Lambda_2 w_1=\Lambda_1 v_1+\Lambda_2\Psi^{-1}v_1=\Lambda_2(v_2+\Psi^{-1}v_1)=\Lambda_2 w_2$. Therefore, we found vectors $w_1,w_2\in\mathbb{R}^r$ such that $P_I^1\Lambda_2 w_1=\Lambda_2 w_2$.

(iii) $\mathrm{col}(P_I^1\Lambda_2)\neq\mathrm{col}(\Lambda_2)$: By contradiction, suppose there is a nonsingular matrix Φ such that $P_I^1\Lambda_2=\Lambda_2\Phi$. Then, $\frac{1}{2}\Lambda_1\Psi=\Lambda_2(\Phi-\frac{1}{2}I_r)$ with $\Psi=(\Lambda_1'\Lambda_1)^{-1}\Lambda_1'\Lambda_2$, which contradicts $\mathrm{col}(\Lambda_1)\neq\mathrm{col}(\Lambda_2)$.

This corresponds to a change of Type 3 for the model (4.2).

2b. $\mathrm{rk}(\Lambda_1'\Lambda_2)=0$: Then, $\Lambda_1'\Lambda_2=0$. This implies that all column vectors of Λ_1 and Λ_2 are orthogonal which contradicts $\mathrm{col}(\Lambda_1)\cap\mathrm{col}(\Lambda_2)\neq\{0\}$. That is, this subcase cannot occur.

2c. $\mathrm{rk}(\Lambda_1'\Lambda_2)=s\neq 0,r$: Then,

(i) $P_I^1\Lambda_2\neq\Lambda_2$: We follow the arguments in *1c(i)*. Arguing by contradiction, equality in (4.3) gives $\Lambda_1\widetilde{\alpha}=\Lambda_2\beta(\beta'\beta)^{-1}$ with $\widetilde{\alpha}=(\Lambda_1'\Lambda_1)^{-1}\alpha$, contradicting $\mathrm{col}(\Lambda_1)\neq\mathrm{col}(\Lambda_2)$.

(ii) $\mathrm{col}(P_I^1\Lambda_2)\cap\mathrm{col}(\Lambda_2)\neq\{0\}$: This follows exactly as in *1c(ii)*.

(iii) $\mathrm{col}(P_I^1\Lambda_2)\neq\mathrm{col}(\Lambda_2)$: Arguing as in *1c(iii)*, we get $\frac{1}{2}\Lambda_1\widetilde{v}=\Lambda_2(\Phi-\frac{1}{2}I_r)v$. This now contradicts $\mathrm{col}(\Lambda_1)\neq\mathrm{col}(\Lambda_2)$.

This corresponds to a change of Type 3 for the model (4.2).

Note that Case *1b* results in a Type 2 break, that is, the transformed series admits a rotational change. We point out that this case is due to our transformation with P_I^1 rather than using just the projection matrix $P_{0,1}$. A transformation with $P_{0,1}$ would result in a series without any latent factors and one would still detect a break due to a non existing loading matrix. It might seem counterintuitive to have a change of Type 2 since this is our null hypothesis. However, as long as the Type 2 change occurs under the alternative model (4.2), the application of a change-point test to the transformed series correctly rejects the null hypothesis since the Type 2 change in the alternative model stems from Case *1b* (a change of Type 1) in the original model.

In summary, depending on the type of change and the relationship between Λ_1 and Λ_2 in the original model, we can get any of the three different types of breaks described in Section 2.2 for the model (4.2).

For the remainder of the paper and for the shortness sake, we focus on two subcases above: Case *1a*, which results in a Type 1 break and Case *2a* assuming that $\Lambda_2 = (\Lambda_1\Phi, \Pi_2)$ with $\Phi \in \mathbb{R}^{r\times r_1}$, and $\Pi_2 \in \mathbb{R}^{d\times r_2}$ with $r_2 = r - r_1$ being linearly independent but not orthogonal to Λ_1 which results in a Type 3 break.

For the proofs of our theoretical results under the alternative, one needs to distinguish each subcase, and we will use Cases *1a* and *2a* to illustrate the main ideas. The remaining cases can be handled similarly. From here on, we will refer to those cases as *TA 1* and *TA 2* but keep in mind how those types result from the original model according to Cases *1a* or *2a*.

We aim to rewrite (4.2) in terms of so-called pseudo-factors. Consider the following matrix representations:

$$\begin{aligned} Z^b &= (Z_1, \dots, Z_{T/2})', \;\; Z^a = (Z_{T/2+1}, \dots, Z_T)'; \\ F^b &= (F_1, \dots, F_{T/2})', \;\; F^a = (F_{T/2+1}, \dots, F_T)'; \\ G^b &= (G_1, \dots, G_{T/2})', \;\; G^a = (G_{T/2+1}, \dots, G_T)'; \\ \eta^b &= (\eta_1, \dots, \eta_{T/2})', \;\; \eta^a = (\eta_{T/2+1}, \dots, \eta_T)', \end{aligned} \tag{4.4}$$

where G^b and G^a will refer to pseudo-factors and both have dimensions $T/2\times \mathrm{r}_p$, and r_p is the number of pseudo-factors. The pseudo-factors are chosen such that the pre- and post-break loadings can be modeled as $\Theta_1 = \Theta B$ and $\Theta_2 = \Theta C$ with *square* matrices $B, C \in \mathbb{R}^{\mathrm{r}_p\times \mathrm{r}_p}$ and some $\Theta \in \mathbb{R}^{d\times \mathrm{r}_p}$ with full column rank. Then, (4.2) can be written as

$$\begin{pmatrix} Z^b \\ Z^a \end{pmatrix} = \begin{pmatrix} F^b\Theta_1' \\ F^a\Theta_2' \end{pmatrix} + \begin{pmatrix} \eta^b \\ \eta^a \end{pmatrix} = \begin{pmatrix} G^bB' \\ G^aC' \end{pmatrix}\Theta' + \begin{pmatrix} \eta^b \\ \eta^a \end{pmatrix} =: G\Theta' + \eta. \tag{4.5}$$

In this model, $r = \mathrm{rk}(B) \le \mathrm{r}_p$ and $r = \mathrm{rk}(C) \le \mathrm{r}_p$ denote the numbers of original factors before and after the break, respectively. Note that (4.5) provides an observationally equivalent representation without a change in the loading matrix $\Theta \in \mathbb{R}^{d\times \mathrm{r}_p}$.

We consider now the alternatives of Types 1 and 3 resulting respectively from *TA 1* (Case *1a*) and *TA 2* (Case *2a*) separately to write them in terms of pseudo-factors as in (4.5). For that, we need to choose G and Θ in (4.5) appropriately. Instead of expressing Types 1 and 3 through linear independence of the loadings as described in Section 2.2, we can express the different types in terms of the column rank of the matrix (B, C).

TA 1: We suppose $\mathrm{col}(\Lambda_1)\cap \mathrm{col}(\Lambda_2) = \{0\}$ and $\mathrm{rk}(\Lambda_1'\Lambda_2) = r$. Then, according to Case *1a*, the alternative model admits a change of Type 1 since $\mathrm{col}(P_I^1\Lambda_2)\cap \mathrm{col}(\Lambda_2) = \{0\}$. Let $F^b \in \mathbb{R}^{T/2\times r}$ and $F^a \in \mathbb{R}^{T/2\times r}$ denote the original factors as in (4.4). Then, the model (4.2) can be written as

$$\begin{aligned} \begin{pmatrix} Z^b \\ Z^a \end{pmatrix} &= \begin{pmatrix} F^b\Theta_1' \\ F^a\Theta_2' \end{pmatrix} + \eta = \begin{pmatrix} F^b & 0_{T/2\times r} \\ 0_{T/2\times r} & F^a \end{pmatrix}\begin{pmatrix} (P_I^1\Lambda_2)' \\ \Lambda_2' \end{pmatrix} + \eta \\ &= \begin{pmatrix} \left[F^b : \star\right]B' \\ \left[\star : F^a\right]C' \end{pmatrix}\Theta' + \eta = G\Theta' + \eta \end{aligned} \tag{4.6}$$

with $\Theta = (\Theta_1, \Theta_2) = \frac{1}{\sqrt{2}}(P_I^1\Lambda_2, \Lambda_2)$, $B = \mathrm{diag}(\sqrt{2}I_r, 0_{r\times r})$, $C = \mathrm{diag}(0_{r\times r}, \sqrt{2}I_r)$ and $G^b = \left[F^b : \star\right]$, $G^a = \left[\star : F^a\right]$ with $\star$ denoting some unidentified numbers. The matrix (B, C) has column rank $2r$. The number of original factors is strictly less than that of the pseudo-factors.

TA 2: We suppose $\mathrm{col}(\Lambda_1)\cap \mathrm{col}(\Lambda_2) \ne \{0\}$, $\mathrm{col}(\Lambda_1) \ne \mathrm{col}(\Lambda_2)$ and $\mathrm{rk}(\Lambda_1'\Lambda_2) = r$. Then, according to Case *2a*, the alternative model admits a change of Type 3 since $\mathrm{col}(P_I^1\Lambda_2)\cap \mathrm{col}(\Lambda_2) \ne \{0\}$ and $\mathrm{col}(P_I^1\Lambda_2) \ne \mathrm{col}(\Lambda_2)$. In order to write the model in terms of pseudo-factors, we need to impose some more structural assumptions on the loadings. Suppose $\Lambda_2 = (\Lambda_1\Phi, \Pi_2)$ with $\Phi \in \mathbb{R}^{r\times r_1}$, and $\Pi_2 \in \mathbb{R}^{d\times r_2}$ with $r_2 = r - r_1$ being linearly independent of Λ_1. Then, $P_I^1\Lambda_2 = (\Lambda_1\Phi, P_I^1\Pi_2)$, $P_I^2\Lambda_2 = \Lambda_2$, and some (but not all) columns of $\Theta_1 = P_I^1\Lambda_2$ and $\Theta_2 = \Lambda_2$ are linearly independent. In particular, we have that $\operatorname{col}(P_I^1\Pi_2)\cap\operatorname{col}(\Pi_2) = \{0\}$. By contradiction, suppose that there are nonzero vectors $v_1, v_2 \in \mathbb{R}^s$ such that $P_I^1\Pi_2 v_1 = \Pi_2 v_2$. Then, $\frac{1}{2}\Lambda_1(\Lambda_1'\Lambda_1)^{-1}\Lambda_1'\Pi_2 v_1 = \Pi_2(v_2 - \frac{1}{2}v_1)$ which contradicts $\operatorname{col}(\Lambda_1)\cap\operatorname{col}(\Pi_2) = \{0\}$ since $\operatorname{rk}(\Lambda_1'\Pi_2) = \operatorname{rk}(\Pi_2) = s$.

Then,

$$\begin{aligned}\begin{pmatrix} Z^b \\ Z^a \end{pmatrix} &= \begin{pmatrix} F^b\Theta_1' \\ F^a\Theta_2' \end{pmatrix} + \eta = \begin{pmatrix} F^b(\Lambda_1\Phi, P_I^1\Pi_2)' \\ F^a\Lambda_2' \end{pmatrix} + \eta \\ &= \begin{pmatrix} \big[F^b : \star\big]B'(\Lambda_1, (1/\sqrt{2})\Pi_2, (1/\sqrt{2})P_I^1\Pi_2)' \\ \big[F^a : \star\big]C'(\Lambda_1, (1/\sqrt{2})\Pi_2, (1/\sqrt{2})P_I^1\Pi_2)' \end{pmatrix} + \eta \\ &= \begin{pmatrix} \big[F^b : \star\big]B' \\ \big[F^a : \star\big]C' \end{pmatrix}\begin{pmatrix} \Lambda_1' \\ (1/\sqrt{2})\Pi_2' \\ (1/\sqrt{2})\Pi_2' P_I^{1\prime} \end{pmatrix} + \eta = G\Theta' + \eta\end{aligned} \tag{4.7}$$

with $\Theta = (\Lambda_1, (1/\sqrt{2})\Pi_2, (1/\sqrt{2})P_I^1\Pi_2)$,

$$B' = \begin{pmatrix} \Phi' & 0_{r_1\times r_2} & 0_{r_1\times r_2} \\ 0_{r_2\times r} & 0_{r_2\times r_2} & \sqrt{2}I_{r_2\times r_2} \\ 0_{r_2\times r} & 0_{r_2\times r_2} & 0_{r_2\times r_2} \end{pmatrix}_{(r+r_2)\times(r+r_2)}, \quad C' = \begin{pmatrix} \Phi' & 0_{r_1\times r_2} & 0_{r_1\times r_2} \\ 0_{r_2\times r} & \sqrt{2}I_{r_2\times r_2} & 0_{r_2\times r_2} \\ 0_{r_2\times r} & 0_{r_2\times r_2} & 0_{r_2\times r_2} \end{pmatrix}_{(r+r_2)\times(r+r_2)}$$

and $G^b = \big[F^b : \star\big]$, $G^a = [F^a : \star]$. The matrix (B, C) has column rank between r and $2r$, namely, $r + r_2$.

For both *TA 1* and *TA 2*, the normalization with $\sqrt{2}$ is out of convenience because it allows us to express the covariance matrix of G in terms of Σ_F (the covariance matrix of F) without a multiplicative constant; see for example (4.9) below.

Following Section 3.2, we define the PCA estimator $\widehat{G}$ as in (3.16) using the transformed series $Y := Y_{1:T}'$ in (4.1). Under the alternative hypothesis, however, $\widehat{G}$ is an estimator of factors in the equivalent model (4.5). In particular, we work with r_p factors. In practice, the number of pseudo-factors is estimated, for example, through the information criterion introduced in Bai and Ng (2002). Their information criterion estimates the number of factors consistently under the alternative model; see Proposition 1 in Han and Inoue (2015) and the preceding paragraph for a discussion.

Note that in contrast to Han and Inoue (2015), we choose the notation $\widehat{G}$ for the estimators of the pseudo-factors to emphasize that they have a different dimension than the estimators under the null hypothesis.

We also define an analogous quantity to H in (3.21) as the matrix

$$J = (\Theta'\Theta/d)(G'\widehat{G}/T)\widehat{V}_{\mathrm{r}_p}^{-1} \tag{4.8}$$

with $\widehat{V}_{\mathrm{r}_p}$ being the $\mathrm{r}_p \times \mathrm{r}_p$ diagonal matrix of the r_p largest eigenvalues of $\frac{1}{Td}\widehat{\Sigma}_T$. Then, as discussed in Appendix G.2 of the supplementary document Düker and Pipiras (2023), one can show that there is a matrix J_0 such that $J \xrightarrow{\mathrm{P}} J_0$.

Next we calculate for both *TA 1* and *TA 2* the pre- and post-break covariances.

TA 1: Based on the representation (4.6), we get

$$\begin{aligned} D_1 &:= \frac{1}{T}\operatorname{E}(BG^{b\prime}G^bB') = \frac{1}{T}B\operatorname{E}\Big(\Big[F^b : \star\Big]'\Big[F^b : \star\Big]\Big)B' = \begin{pmatrix} \Sigma_F & 0_{r\times r} \\ 0_{r\times r} & 0_{r\times r} \end{pmatrix}, \\ D_2 &:= \frac{1}{T}\operatorname{E}(CG^{a\prime}G^aC') = \frac{1}{T}C\operatorname{E}\Big(\Big[\star : F^a\Big]'\Big[\star : F^a\Big]\Big)C' = \begin{pmatrix} 0_{r\times r} & 0_{r\times r} \\ 0_{r\times r} & \Sigma_F \end{pmatrix}. \end{aligned} \tag{4.9}$$

TA 2: Based on the representation (4.7), we get

$$
\begin{aligned}
D_1 &:= \frac{1}{T}\,\mathrm{E}(BG^{b'}G^{b}B') = \frac{1}{T}B\,\mathrm{E}\Big(\big[F^b:\star\big]'\big[F^b:\star\big]\Big)B' = B\begin{pmatrix}\frac{1}{2}\Sigma_F & \star_{r\times r_2}\\ \star_{r_2\times r} & \star_{r_2\times r_2}\end{pmatrix}B',\\
D_2 &:= \frac{1}{T}\,\mathrm{E}(CG^{a'}G^{a}C') = \frac{1}{T}C\,\mathrm{E}\Big(\big[F^a:\star\big]'\big[F^a:\star\big]\Big)C' = C\begin{pmatrix}\frac{1}{2}\Sigma_F & \star_{r\times r_2}\\ \star_{r_2\times r} & \star_{r_2\times r_2}\end{pmatrix}C'.
\end{aligned}
\tag{4.10}
$$

For both types, we define the matrix

$$
\mathcal{C} := J_0'(D_1 - D_2)J_0. \tag{4.11}
$$

Here, J_0 and D_1, D_2 can be different depending on the respective type of change as given in (4.9) and (4.10).

Let Π_Θ be the diagonal matrix whose diagonal entries are the eigenvalues of $\Theta'\Theta$ in decreasing order. We state the following assumptions.

Assumption P.1. (Conditions on the break.)

(i) There is a positive definite diagonal matrix $\widetilde{\Pi}$ such that $\|\Pi_\Theta/d - \widetilde{\Pi}\| = \mathcal{O}\Big(\frac{1}{\sqrt{d}}\Big)$.

(ii) The eigenvalues of Π_Θ are distinct.

Assumption P.2. (Conditions on long-run variance estimator.) Recall the matrix $\mathcal{C}$ in (4.11).

(i) Suppose

$$
\lim_{T\to\infty} \inf\{\mathrm{vech}(\mathcal{C})'\big(b_{T/2}\Omega(GJ_0)^{-1}\big)\,\mathrm{vech}(\mathcal{C})\} \xrightarrow{\mathrm{p}} c > 0,
$$

where b_T is the bandwidth parameter.

(ii) Assumption W.1(i) holds for $\Omega(GJ_0)$.

Assumption P.2 regulates the asymptotic behavior of the variance matrices in our statistic.

Proposition 4.1. *Under Assumptions A.1–A.3, CR.1–CR.4 and P.1–P.2:*

(i) For a nonrandom matrix $\mathcal{C} \neq 0$ *in* (4.11),

$$
\frac{1}{T}\sum_{t=1}^{T/2}\widehat{G}_t\widehat{G}_t' - \frac{1}{T}\sum_{t=T/2+1}^{T}\widehat{G}_t\widehat{G}_t' \xrightarrow{\mathrm{p}} \mathcal{C}.
$$

(ii) The test statistic $W(\widehat{G})$ *defined as in* (3.20) *is consistent under the alternative that a portion of cross-sections are linearly independent; i.e.* $W(\widehat{G}) \xrightarrow{\mathrm{p}} \infty$.

We refer to Appendix G in the supplementary material for the proof of Proposition 4.1.

5 Extensions and other observations

In Section 5.1, we discuss how our approach can be generalized to a setting where the two factor models are driven by different numbers of factors. In Section 5.2, we present an extension of our approach using a different type of change-point test and we end with a discussion of our test in Section 5.3.

5.1 Different numbers of factors

In this section, we allow the number of factors to be different for the two series under consideration. Suppose $\Lambda_1 \in \mathbb{R}^{d\times r_1}$ and $\Lambda_2 \in \mathbb{R}^{d\times r_2}$ with $r_2 \leq r_1$ and

$$\Lambda_2 = \Lambda_1 \Phi_2, \tag{5.1}$$

where $\Phi_2 \in \mathbb{R}^{r_1\times r_2}$ is a matrix of full column rank.

Note that our procedure for $r_1 = r_2$ relies heavily on the invariance under invertible transformation of the projection matrices; see (3.6). In particular, it ensures that the error terms of the new series (3.7) are stationary over time; see (3.9). This property is not satisfied when $r_2 \neq r_1$, in particular, even under the null hypothesis,

$$P_{0,1} = \Lambda_1(\Lambda_1'\Lambda_1)^{-1}\Lambda_1' \neq \Lambda_1\Phi_2(\Phi_2'\Lambda_1'\Lambda_1\Phi_2)^{-1}\Phi_2'\Lambda_1' = P_{0,2}.$$

Alternatively, we can take advantage of the relationship

$$\begin{aligned} P_{0,2}P_{0,1} &= \Lambda_2(\Lambda_2'\Lambda_2)^{-1}\Lambda_2'\Lambda_1(\Lambda_1'\Lambda_1)^{-1}\Lambda_1' \\ &= \Lambda_2(\Lambda_2'\Lambda_2)^{-1}\Phi_2'\Lambda_1'\Lambda_1(\Lambda_1'\Lambda_1)^{-1}\Lambda_1' = \Lambda_2(\Lambda_2'\Lambda_2)^{-1}\Lambda_2' = P_{0,2}. \end{aligned} \tag{5.2}$$

Then, we can follow the same approach as in (3.7) and transform the series as

$$Y_{1:T_2} = \left(\frac{1}{2}(\widehat{P}_2\widehat{P}_1 + I_d)X^2_{1:T_2/2}, \frac{1}{2}(\widehat{P}_2 + I_d)X^2_{(T_2/2+1):T_2}\right), \tag{5.3}$$

where $\widehat{P}_1, \widehat{P}_2$ are as in (3.4) such that at the population level

$$\begin{aligned} P_{0,2}P_{0,1}X_t^2 &\approx P_{0,2}P_{0,1}\Lambda_2 F_t^2 = \Lambda_2 F_t^2 \quad \text{for } t = 1, \dots, T_2/2, \\ P_{0,2}X_t^2 &\approx P_{0,2}\Lambda_2 F_t^2 = \Lambda_2 F_t^2 \quad \text{for } t = T_2/2+1, \dots, T_2. \end{aligned} \tag{5.4}$$

Note that the roles of the first and the second series are interchangeable.

Since $P_{0,2}P_{0,1} = P_{0,2}$ under the null hypothesis, the idiosyncratic errors of the transformed series $Y_{1:T_2}$ are still stationary. In addition, the errors have positive definite covariances since

$$\begin{aligned} (P_{0,2}P_{0,1} + I_d)\,\mathrm{E}(\varepsilon_t^2\varepsilon_t^{2'})(P_{0,2}P_{0,1} + I_d)' &= (P_{0,2} + I_d)\Sigma_\varepsilon(P_{0,2} + I_d)', \\ (P_{0,2} + I_d)\,\mathrm{E}(\varepsilon_t^2\varepsilon_t^{2'})(P_{0,2} + I_d)' &= (P_{0,2} + I_d)\Sigma_\varepsilon(P_{0,2} + I_d)'. \end{aligned} \tag{5.5}$$

Due to (5.2), the series $Y_{1:T_2}$ in (5.3) is expected to have the same loadings and factors under the null (5.1) as $X^2_{1:T_2}$. However, given (5.5), the errors have a different covariance matrix than the original series.

Under the alternative, we get

$$\begin{aligned} P_{0,2}P_{0,1}X_t^2 &\approx P_{0,2}P_{0,1}\Lambda_2 F_t^2 \neq \Lambda_2 F_t^2 \quad \text{for } t = 1, \dots, T_2/2, \\ P_{0,2}X_t^2 &\approx P_{0,2}\Lambda_2 F_t^2 = \Lambda_2 F_t^2 \quad \text{for } t = T_2/2+1, \dots, T_2. \end{aligned} \tag{5.6}$$

The illustrated behavior of $Y_{1:T_2}$ at the population level in (5.4) and (5.6) suggests that for the transformed series $Y_{1:T_2}$ our hypothesis testing problem is equivalent to testing for a change at time $T_2/2$ in the loadings. In other words, we can proceed as in Section 3.2 and apply an available change-point test as the one of Han and Inoue (2015).

5.2 Supremum approach

Our approach suggests to split the data into two halves and applies different transformations to each of them. This way, we artificially create a change in the loadings under the alternative, while there is no change under the null hypothesis. In particular, this means that in the case of a change, we know its location. We therefore apply a test comparing the pre- and post-sample means of the covariances of the factors using the exact change-point.

However, a potential alternative approach is to consider the supremum over a certain window around the artificially created change-point location $T/2$. The hope is to improve our test's power while not loosing too much in terms of its size. However, a trade-off between power and size is expected, at least to some degree. See also the corresponding simulation study in Section 6.4 below.

For the supremum approach, we introduce

$$V(\pi, \widehat{F}) = \operatorname{vech}\left(\sqrt{T}\left(\frac{1}{\lfloor \pi T \rfloor}\sum_{t=1}^{\lfloor \pi T \rfloor} \widehat{F}_t \widehat{F}_t' - \frac{1}{T - \lfloor \pi T \rfloor}\sum_{t=\lfloor \pi T \rfloor + 1}^{T} \widehat{F}_t \widehat{F}_t'\right)\right).$$

To construct a Wald type test statistic, we also need the sample long-run variance of $V(FH_0)$. We use the following estimator for the long-run variance (3.22),

$$\Omega(\pi, FH_0) = \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)\Omega(FH_0)$$

with $\Omega(FH_0)$ as in (3.18). Then, the Wald type test statistic across a certain interval can be defined as

$$\sup_{\pi \in [\pi_0, 1-\pi_0]} W(\pi, \widehat{F}) \quad \text{with} \quad W(\pi, \widehat{F}) = V(\pi, \widehat{F})' \Omega^{-1}(\pi, \widehat{F}) V(\pi, \widehat{F}) \tag{5.7}$$

for $\pi_0 \in (0, \frac{1}{2}]$. The test statistic is expected to satisfy the following convergence under the null hypothesis,

$$\sup_{\pi \in [\pi_0, 1-\pi_0]} W(\pi, \widehat{F}) \xrightarrow{\mathrm{d}} \sup_{\pi \in [\pi_0, 1-\pi_0]} Q_p(\pi) \tag{5.8}$$

with $Q_p(\pi) = (B_p(\pi) - \pi B_p(1))'(B_p(\pi) - \pi B_p(1))/(\pi(1-\pi))$ and B_p is a p-vector of independent Brownian motions on $[0, 1]$ restricted to $[\pi_0, 1-\pi_0] \in (0, 1)$, $p = r(r+1)/2$. The respective critical values can be found in Table 1 in Andrews (1993).

5.3 Discussion on alternative approaches

Recall the three different types of breaks in Section 2.2. In order to characterize the behavior of our test statistic under the alternatives, we wrote Types 1 and 2 in terms of so-called pseudo-factors in Section 4. Similarly, one can rewrite what happens under Type 2 (our null hypothesis). Recall from Section 2.1 that Type 2 describes how the loadings undergo a rotational change.

Type 2: Suppose $\Theta_2 = \Theta_1 C$ with C being nonsingular. Then the model can be put into the form (4.5) with the matrix (B, C) in (4.5) having column rank r:

$$\begin{aligned}
\begin{pmatrix} Z^b \\ Z^a \end{pmatrix} &= \begin{pmatrix} F^b \Theta_1' \\ F^a \Theta_2' \end{pmatrix} + \eta = \begin{pmatrix} F^b \Theta_1' \\ F^a C' \Theta_1' \end{pmatrix} + \eta \\
&= \begin{pmatrix} F^b \\ F^a C' \end{pmatrix} \Theta_1' + \eta = G\Theta' + \eta
\end{aligned}$$

with $\Theta = \Theta_1$, $B = I_r$, and $G^b = F^b$, $G^a = F^a$. Given that the different types of breaks are characterized through the column rank of the matrix (B, C), a more natural way of distinguishing seems to be a rank test on the covariances of the latent factors. This would also allow us to test whether only a partial set of vectors is shared across individuals. There are several works discussing issues of testing for the rank of a matrix; e.g., Donald, Fortuna, and Pipiras (2010). A reliable estimator for the rank, however, would still only answer the question of how many vectors are shared across individuals but not which ones.

In particular, a rank-type test approach can be applied directly to the data instead of using the transformed series (3.7). The idea is to concatenate the two series $\{X_t^1\}$ and $\{X_t^2\}$ over time and to do PCA estimation based on $\mathrm{r}_p = 2r$ pseudo-factors with r being the original number of factors. Based on the PCA estimator, one can then estimate the rank of the covariances of the latent factors. The rank takes values from r to $2r$. The rank equal to r corresponds to a rotational change in the loadings as under our null hypothesis. With increasing rank the number of column vectors that are not shared across loadings increases, where a rank of $2r$ corresponds to no column vectors being shared. More formally, let $\Lambda_k = (\Lambda\Phi_k, \Pi_k)$, $k = 1, 2$ with $\Lambda \in \mathbb{R}^{d\times r}$, $\Phi_k \in \mathbb{R}^{r\times r_k}$, $\Pi_k \in \mathbb{R}^{d\times(r-r_k)}$ and $F_{1:T}^k = (F_1^{k'}, \ldots, F_T^{k'})'$. Then,

$$
\begin{aligned}
Y &= \begin{pmatrix} F_{1:T}^1 & 0 \\ 0 & F_{1:T}^2 \end{pmatrix} \begin{pmatrix} \Lambda_1' \\ \Lambda_2' \end{pmatrix} + \begin{pmatrix} \varepsilon^1 \\ \varepsilon^2 \end{pmatrix} = \begin{pmatrix} F_{1:T}^1 & 0 \\ 0 & F_{1:T}^2 \end{pmatrix} \begin{pmatrix} \Phi_1'\Lambda' \\ \Pi_1' \\ \Phi_2'\Lambda' \\ \Pi_2' \end{pmatrix} + \begin{pmatrix} \varepsilon^1 \\ \varepsilon^2 \end{pmatrix} \\
&= \begin{pmatrix} F_{1:s_1,1}^{1'}\Phi_1' & F_{s_1+1:r_1,1}^{1'} & 0 \\ \vdots & \vdots & \\ F_{1:s_1,T}^{1'}\Phi_1' & F_{s_1+1:r_1,T}^{1'} & 0 \\ F_{1:s_2,1}^{2'}\Phi_2' & 0 & F_{s_2+1:r_2,1}^{2'} \\ \vdots & \vdots & \\ F_{1:s_2,T}^{2'}\Phi_2' & 0 & F_{s_2+1:r_2,T}^{2'} \end{pmatrix} \begin{pmatrix} \Lambda' \\ \Pi_1' \\ \Pi_2' \end{pmatrix} + \begin{pmatrix} \varepsilon^1 \\ \varepsilon^2 \end{pmatrix} =: \mathcal{F}\Lambda' + \widetilde{\varepsilon},
\end{aligned}
\tag{5.9}
$$

where we used $F_{a:b,t}^{k'} = (F_{a,t}^k, \ldots, F_{b,t}^k)$ in (5.9). Then, the matrix $\mathrm{E}(\mathcal{F}'\mathcal{F})$ has rank $r + r_1 + r_2$.

Another natural approach is to check directly whether the projection matrices associated with Λ_1 and Λ_2 are equal. We refer to a recent work by Liao and Todorov (2023) who studied a similar problem for continuous factor models and pursued this approach. They suggested a test statistic based on the Frobenius norm of the projection discrepancy between the pre- and post-break loadings.

The work by Silin and Fan (2020) is similar in nature and also studies a certain distance between two projection matrices calculated based on the eigenvectors of two covariance matrices. The suggested test is based on finite sample properties, introducing a novel resampling method as an alternative to using bootstrap. The authors do not suggest any asymptotic limit of the proposed test statistic.

In comparison to the approaches of Liao and Todorov (2023) and Silin and Fan (2020), our test is low dimensional in nature. While Liao and Todorov (2023) and Silin and Fan (2020) base their analyses on the $d \times d$-dimensional projection matrices, we manage to reduce the problem to testing for changes in the low-dimensional factors. In particular, this allows us to derive the asymptotic behavior of our test statistic.

6 Simulation study

We examine our testing procedure in a simulation study considering the two-subject setting (Section 6.1), the change-point analysis (Section 6.2) and the modification for factors with different dimensions (Section 6.3). Section 6.4 presents some simulation results on the supremum approach discussed in Section 5.2. We conclude this section with a discussion on the simulation design (Section 6.5). All simulations are conducted at a 5% significance level and the data were standardized before applying our test.

6.1 Two subjects

We generate data for two subjects with r factors. Recall the notation

$$\begin{aligned} X_t^1 &= \Lambda_1 F_t^1 + \varepsilon_t^1, \quad t = 1, \dots, T, \\ X_t^2 &= \Lambda_2 F_t^2 + \varepsilon_t^2, \quad t = 1, \dots, T, \end{aligned} \tag{6.1}$$

where $\Lambda_k = (\lambda_{k,ij})_{i=1,\dots,d;j=1,\dots,r}$, $k = 1, 2$. Under the null hypothesis, there are nonsingular matrices $\Phi_k \in \mathbb{R}^{r\times r}$ such that $\Lambda_k = \Lambda\Phi_k$, $k = 1, 2$, and some common loading matrix $\Lambda = (\lambda_{ij})_{i=1,\dots,d;j=1,\dots,r}$ with full column rank. We consider the following data generating models for the null hypothesis (NDGP):

NDGP1: Independent factors and errors ($r = 3$): Choose $\lambda_{ij} \sim \mathcal{N}(0, 1)$. Set $\Phi_1 = I_r$ and Φ_2 is randomly generated with eigenvalues in the interval $(0.75, 1.25)$. The factors are $F_{i,t}^k \sim \mathcal{N}(0, 1)$. For the errors, suppose $\varepsilon_{i,t}^k \sim \mathcal{N}(0, 1)$, $k = 1, 2$.

NDGP2: Temporal correlation in the errors ($r = 3$): Choose $\lambda_{ij} \sim \mathcal{N}(0, 1)$. The matrices Φ_k, $k = 1, 2$ are randomly generated with eigenvalues in the interval $(0.75, 1.25)$. The factors are $F_{i,t}^k \sim \mathcal{N}(0, 1)$. For the errors, suppose $\varepsilon_{i,t}^k = \sigma_i v_{i,t}$, $\sigma_i \sim \mathcal{U}(0.5, 1.5)$, $v_{i,t} = 0.5v_{i,t-1} + e_{i,t}$, $e_{i,t} \sim \mathcal{N}(0, 1)$.

NDGP3: Temporal correlation in the factors ($r = 3$): Choose $\lambda_{ij} \sim \mathcal{N}(0, 1)$. The matrices Φ_k, $k = 1, 2$ are randomly generated with eigenvalues in the interval $(0.75, 1.25)$. The factors are generated as $F_{i,t}^k = 0.5F_{i,t-1}^k + e_{i,t}$, $e_{i,t} \sim \mathcal{N}(0, 1 - 0.7^2)$. For the errors, suppose $\varepsilon_{i,t}^k \sim \mathcal{N}(0, 1)$, $k = 1, 2$.

NDGP4: Different DGPs for each subject ($r = 3$): Choose $\lambda_{i,j} \sim \mathcal{N}(0, 1)$. The matrices Φ_k, $k = 1, 2$, are randomly generated with eigenvalues in the interval $(0.75, 1.25)$. The factors are $F_{i,t}^k \sim \mathcal{N}(0, 1)$, $k = 1, 2$. For the errors, suppose $\varepsilon_{i,t}^1 \sim \mathcal{N}(0, 1)$ and $\varepsilon_{i,t}^2 = \sigma_i v_{i,t}$, $\sigma_i \sim \mathcal{U}(0.5, 1.5)$, $v_{i,t} = 0.5v_{i,t-1} + e_{i,t}$, $e_{i,t} \sim \mathcal{N}(0, 1)$.

The corresponding empirical sizes over 1000 replications can be found in Table 1. NDGP1 does not allow for any serial correlation of factors or idiosyncratic errors, while NDGP2–NDGP4 allow for cross-sectional heteroskedasticity. NDGP3 assumes that the factors follow and AR(1) model. All simulations show reasonable results for the test sizes. The nominal level of five percent is only exceeded in NDGP3.

The following list provides the data generating models under the alternative (ADGP):

ADGP1: Additive change across all loading vectors ($r = 3$): Choose $\lambda_{ij} \sim \mathcal{N}(b/2, 1)$ and $\Lambda_1 = \Lambda$, $\Lambda_2 = \Lambda - b$ The factors are $F_{i,t}^k \sim \mathcal{N}(0, 1)$. For the errors, suppose $\varepsilon_{i,t}^k \sim \mathcal{N}(0, 1 + b^2/4)$, $k = 1, 2$.

(d,T)	NDGP1	NDGP2	NDGP3	NDGP4
(200,200)	0.033	0.0285	0.1065	0.027
(500,500)	0.0395	0.0395	0.0945	0.044

Table 1: Empirical test size for the setting with two individuals sharing the same loading matrix up to a rotation with a nonsingular matrix.

ADGP2: Additive change across a partial set of loading vectors ($r = 3$): Choose $\lambda_{ij} \sim \mathcal{N}(b/2, 1)$ and $\Lambda_1 = \Lambda$, $\lambda_{2,ij} = \lambda_{1,ij} - b$ for $i = 1, \dots, a \cdot d$ and $j = 1, \dots, r$ and $\lambda_{2,ij} = \lambda_{1,ij}$ for $i = a \cdot d + 1, \dots, d$ and $j = 1, \dots, r$. The factors are $F_{i,t}^k \sim \mathcal{N}(0, 1)$. For the errors, suppose $\varepsilon_{i,t}^k \sim \mathcal{N}(0, 1 + b^2/4)$, $k = 1, 2$.

ADGP3: Partial change in the loadings plus rotational change ($r = 4$): Set $\Lambda_1 = \Lambda$ and $\Lambda_2 = (\Pi_2 \ \Lambda\Phi)$ with $\lambda_{ij} \sim \mathcal{N}(0, 1)$. Let $\Pi_2 = (\pi_{2,ij})_{i=1,\dots,d; j=1,\dots,(r-c)}$, $\pi_{2,ij} \sim \mathcal{N}(0, 2)$ and $\Phi \in \mathbb{R}^{r \times c}$ randomly generated. The factors are $F_{i,t}^k \sim \mathcal{N}(0, 1)$. For the errors, suppose $\varepsilon_{i,t}^k \sim \mathcal{N}(0, 1)$, $k = 1, 2$.

ADGP4: Loading matrices drawn from different distributions ($r = 4$): Same as ADGP3 but with $\lambda_{2,ij} \sim \mathcal{C}auchy(0, 1)$.

(d,T)	b	ADGP1	a	ADGP2	c	ADGP3	c	ADGP4
(200,200)	1/3	0.089	0.2	0.1035	1	1	1	1
(500,500)	1/3	0.2345	0.2	0.2865	1	1	1	1
(200,200)	2/3	0.813	0.4	0.4275	2	0.941	2	1
(500,500)	2/3	1	0.4	0.927	2	0.9345	2	1
(200,200)	1	0.994	0.6	0.8325	3	0.71	3	0.9795
(500,500)	1	1	0.6	0.9995	3	0.53	3	1
(200,200)	2	0.8015	0.8	0.973				
(500,500)	2	0.999	0.8	1				

Table 2: Empirical test power for the setting with two individuals under different alternatives.

The corresponding empirical powers can be found in Table 2. ADGP1 focusses on an additive change in the loadings, for varying magnitudes $b \in \{1/3, 2/3, 1, 3\}$. The model is very similar to DGP A1 in Han and Inoue (2015) for changes in the loadings over time. Under ADGP1, our test does not have good power when the magnitude of the break is small. With larger magnitude of the break, the test becomes more powerful, though the power is not monotonically increasing in b. This observation aligns with observations made for DGP A1 in Han and Inoue (2015).

Similarly, ADGP2 aligns with DGP A2 in Han and Inoue (2015) and considers additive differences in a fraction of the loadings. The proportion of loadings that are different increases with $a \in \{0.2, 0.4, 0.6, 0.8\}$. Our test struggles to achieve good power for small fraction $a = 0.2$. However, it increases with sample size and a.

In ADGP3, we consider partial rotational changes in the loadings concatenated with linearly independent column vectors. The proportion of the linearly independent vectors decreases with $c \in \{1, 2, 3\}$.

The last model, ADGP4, is the simplest to detect since both loading matrices are drawn from different distributions. Our test does not have any issues to detect those differences.

6.2 Change-point

We generate data with a change at $\lfloor \pi T \rfloor$ and r_1 pre- and r_2 post-break factors:

$$X_t = \begin{cases} \Lambda_1 F_t + \varepsilon_t, & t = 1, \ldots, \lfloor \pi T \rfloor, \\ \Lambda_2 F_t + \varepsilon_t, & t = \lfloor \pi T \rfloor + 1, \ldots, T. \end{cases} \tag{6.2}$$

We consider the following settings for the factors and loadings Λ_1, Λ_2 under the null hypothesis with $r = r_1 = r_2 = 3$:

NDGPcp1: Independent factors and errors: Choose $\lambda_{ij} \sim \mathcal{N}(0,1)$. Set $\Phi_1 = I_r$ and Φ_2 is randomly generated with eigenvalues in the interval $(0.75, 1.25)$. The factors are $F_{i,t} \sim \mathcal{N}(0,1)$. For the error, suppose $\varepsilon_{i,t} \sim \mathcal{N}(0,1)$.

NDGPcp2: Temporal correlation in the errors: Choose $\lambda_{ij} \sim \mathcal{N}(0,1)$. The matrices Φ_k, $k = 1,2$, are randomly generated with eigenvalues in the interval $(0.75, 1.25)$. The factors are $F_{i,t} \sim \mathcal{N}(0,1)$. For the errors, suppose $\varepsilon_{i,t} = \sigma_i v_{i,t}$, $\sigma_i \sim \mathcal{U}(0.5, 1.5)$, $v_{i,t} = 0.5 v_{i,t-1} + e_{i,t}$, $e_{i,t} \sim \mathcal{N}(0,1)$.

Under the alternative with $r = r_1 = r_2 = 3$:

ADGPcp1: Additive change across all loading vectors: Choose $\lambda_{ij} \sim \mathcal{N}(b/2, 1)$ and $\Lambda_1 = \Lambda$, $\Lambda_2 = \Lambda - b$ The factors are $F_{i,t} \sim \mathcal{N}(0,1)$. For the errors, suppose $\varepsilon_{i,t} \sim \mathcal{N}(0, 1 + b^2/4)$, with $b = 1$

ADGPcp2: Additive change across a partial set of loading vectors: Choose $\lambda_{ij} \sim \mathcal{N}(b/2, 1)$ and $\Lambda_1 = \Lambda$, $\lambda_{2,ij} = \lambda_{1,ij} - b$ for $i = 1, \ldots, a \cdot d$ and $j = 1, \ldots, r$ and $\lambda_{2,ij} = \lambda_{1,ij}$ for $i = a \cdot d + 1, \ldots, d$ and $j = 1, \ldots, r$. The factors are $F_{i,t} \sim \mathcal{N}(0,1)$. For the errors, suppose $\varepsilon_{i,t} \sim \mathcal{N}(0, 1 + b^2/4)$, with $b = 1$ and $a = 0.4$.

ADGPcp3: Partial change in the loadings plus rotational change: We suppose here that $r = 4$ Set $\Lambda_1 = \Lambda$ and $\Lambda_2 = (\Pi_2 \; \Lambda\Phi)$ with $\lambda_{ij} \sim \mathcal{N}(0,1)$. Let $\Pi_2 = (\pi_{2,ij})_{i=1,\ldots,d; j=1,\ldots,(r-c)}$, $\pi_{ij} \sim \mathcal{C}auchy(0,1)$ and $\Phi \in \mathbb{R}^{r \times c}$ randomly generated. The factors are $F_{i,t} \sim \mathcal{N}(0,1)$. For the errors, suppose $\varepsilon_{i,t} \sim \mathcal{N}(0,1)$.

Table 3 shows the rejection rates for the different DGPs under the null hypothesis. As observed in a comprehensive comparison of different change-point tests in Su and Wang (2017), the Wald type test of Han and Inoue (2015) tends to under-reject the null hypothesis. This can also be observed in the simulation results for our test.

Note that the effective sample size in the change-point setting is smaller than in the two-subject model setting since the pre- and post-break samples are of length $1 - \lfloor T\pi \rfloor$ and $\lfloor T\pi \rfloor$. The change-point test is then applied to a transformed series of length $1 - \lfloor T\pi \rfloor$. In Table 3, one can see that the test size increases with the sample size and with changes closer to the boundary. A smaller π has the effect that we have more data available to create our artificial change-point by transforming the longer sequence of data $(1 - \lfloor T\pi \rfloor)$. The estimation of the projection matrices happens based on $\lfloor T\pi \rfloor$th data points and does not seem to be impacted by fewer data points.

For comparison, we applied the change-point test of Han and Inoue (2015) to NDGPcp1 with $\pi = 0.5$. The resulting values for $d, T = 200$ and $d, T = 500$ are respectively 0.8225 and 0.9995. This confirms that while our test does not reject due to a rotational change in the loadings, a regular test for structural changes in the loadings does.

In Table 4, one can find the rejections rates under the alternative. For ADGPcp2 the test suffers under the relatively small sample size $T = 200$. However, we achieve good power for the other two DGPs.

(d,T)	π	NDGPcp1	NDGPcp2
(200,200)	0.5	0.0145	0.013
(500,500)	0.5	0.033	0.028
(200,200)	0.4	0.018	0.0195
(500,500)	0.4	0.0275	0.028
(200,200)	0.3	0.0225	0.018
(500,500)	0.3	0.0355	0.033
(200,200)	0.2	0.0225	0.0265
(500,500)	0.2	0.04	0.0345

Table 3: Empirical test size for different change-point locations and different DGPs under the hypothesis.

(d,T)	π	ADGPcp1	ADGPcp2	ADGPcp3
(200,200)	0.5	0.562	0.098	0.8005
(500,500)	0.5	1	0.4885	1
(200,200)	0.4	0.7445	0.1475	0.8935
(500,500)	0.4	1	0.612	1
(200,200)	0.3	0.869	0.191	0.945
(500,500)	0.3	1	0.7255	1
(200,200)	0.2	0.924	0.274	1
(500,500)	0.2	1	0.805	1

Table 4: Empirical test power for different change-point locations.

6.3 Different numbers of factors

We study here a setting allowing for different numbers of factors using the test in Section 5.1. This includes both the two-subjects test with each being modeled by a different number of factors and having a different number of factors pre- and post-break in the change-point setting. We assume for both two subject and change-point setting that $\Lambda \in \mathbb{R}^{d\times r_1}$ with $\lambda_{ij} \sim \mathcal{N}(0,1)$, $\Phi_1 \in \mathbb{R}^{r_1\times r_1}$, $\Phi_2 \in \mathbb{R}^{r_1\times r_2}$ and $r_1 = 4$, $r_2 = 3$.

NDGPdnf1: Two subjects, (6.1): Set $\Phi_1 = I_r$ and Φ_2 is randomly generated with singular values in the interval $(0.75, 1.25)$. The factors are $F^k_{i,t} \sim \mathcal{N}(0,1)$. For the errors, suppose $\varepsilon^k_{i,t} \sim \mathcal{N}(0,1)$, $k = 1,2$.

NDGPdnf2: Two subjects, (6.1): Set $\Phi_1 = I_r$ and Φ_2 is randomly generated with singular values in the interval $(0.75, 1.25)$. The factors for the first subject are generated as $F^1_{i,t} = 0.5F^1_{i,t-1} + e_{i,t}$, $e_{i,t} \sim \mathcal{N}(0, 1-0.7^2)$. The factors for the second subject are $F^2_{i,t} \sim \mathcal{N}(0,1)$. For the errors, suppose $\varepsilon^k_{i,t} \sim \mathcal{N}(0,1)$, $k = 1,2$.

NDGPdnf3: Change-point, (6.2): Φ_1, Φ_2 are randomly generated with singular values in the interval $(0.75, 1.25)$. The factors are generated as $F_{i,t} = 0.5F_{i,t-1} + e_{i,t}$, $e_{i,t} \sim \mathcal{N}(0, 1-0.7^2)$. For the error, suppose $\varepsilon_{i,t} \sim \mathcal{N}(0,1)$.

NDGPdnf4: Change-point, (6.2): Φ_1, Φ_2 are randomly generated with singular values in the interval $(0.75, 1.25)$. The factors are generated as $F_{i,t} \sim \mathcal{N}(0,1)$. For the error, suppose $\varepsilon_{i,t} \sim \mathcal{N}(0,1)$.

	Two-subjects		Change-point	
(d,T)	NDGPdnf1	NDGPdnf2	NDGPdnf3	NDGPdnf4
(200,200)	0.045	0.0470	0.0215	0.0945
(500,500)	0.043	0.0425	0.0370	0.0980

Table 5: Empirical test sizes for two individuals and change-point setting assuming different numbers of factors in pre- and post-break samples.

The corresponding empirical sizes can be found in Table 5. The test performs generally well. In particular in the setting of considering two subjects, the empirical size gets close to the nominal 5% significance level. On the other hand, in the change-point setting, we see slightly under- and oversized results for NDGPdnf3 and NDGPdnf4, respectively. As mentioned in Section 6.2, this could be due to the fact that the effective sample size in the change-point setting is smaller than in the two-subject model setting. We omit the power analysis for different numbers of factors.

6.4 Supremum approach

In Section 5.2, we proposed a supremum approach, by applying a change-point test not only at the artificially created change-point $T/2$ but partitioning the data at locations over a certain interval around $T/2$ and taking the supremum over the resulting test statistics. The expected behavior of the test statistic under the null hypothesis is stated in (5.8).

The interval for the supremum is determined by $\pi_0 = 0.45$ in (5.7). In Table 6, we consider some of the DGPs from Section 6.1 which suffered from low power. As one can see, the power of ADGP3 improved quite a bit. There is no improvement for ADGP2 with 0.2 percent of the loading matrix being different from the pre-break loadings.

(d,T)	a	ADGP2	c	ADGP3
(200,200)	0.2	0.0700	1	1
(500,500)	0.2	0.2515	1	1
(200,200)	0.6	0.7525	3	0.998
(500,500)	0.6	1	3	1

Table 6: Empirical test power using the test statistic (5.7) with $\pi_0 = 0.45$.

In addition, we applied the supremum approach to some of the DGPs from Section 6.1 under the null hypothesis. One can observe the expected trade-off of loosing size while improving power. The size increased for all of the considered NDGPs and exceeds the nominal level of 5%.

(d,T)	NDGP1	NDGP2
(200,200)	0.0950	0.098
(500,500)	0.1155	0.110

Table 7: Empirical test size using the test statistic (5.7) with $\pi_0 = 0.45$.

6.5 Discussion on simulation design

The theoretical results on consistent estimation of the loadings rely (among others) on Assumption CR.1. The assumptions in there ensure that the rescaled singular values of the loading matrices are nonzero and finite, i.e., $\liminf_{d\to\infty} \frac{\lambda_{\min}(\Lambda'\Lambda)}{d} > 0$ and $\limsup_{d\to\infty} \frac{\lambda_{\max}(\Lambda'\Lambda)}{d} < \infty$.

Our simulation study includes the assumptions that the matrices Φ_k, $k = 1, 2$, take only eigenvalues in a certain interval. This interval can certainly be bigger than the one we considered. However, the smallest eigenvalue $\lambda_{\min}(\Phi\Lambda'\Lambda\Phi)$ can not get too close to zero. This should not be surprising since we would move closer to the regime of having a different number of factors. On the population level, a different number of factors results in different loading matrices since their dimensions mismatch. Then, our test statistic is expected to reject the null hypothesis of having the same (or the same up to transformation with a nonsingular matrix) loading matrices across time or individuals.

To illustrate these points, we consider some extreme cases with particularly small and large eigenvalues of Φ_k. We focus here on the two subject case.

DGP: $\Lambda_1 = \Lambda$ and $\Lambda_2 = \Lambda\Phi$ with $\lambda_{ij} \sim \mathcal{N}(0, 1)$. Φ is randomly generated but with fixed smallest and largest eigenvalues $\lambda_{\min}(\Phi) = e$ and $\lambda_{\max}(\Phi) = f$.

In theory, the chosen DGP falls under the null hypothesis, however, one can see from Table 8 that we get a 100% rejection rate.

(d, T)	e	f	$\lambda_{\min}(\Phi\Lambda'\Lambda\Phi)$	$\lambda_{\max}(\Phi\Lambda'\Lambda\Phi)$	DGP
(200,200)	0.01	1	0.140	14.117	1
(500,500)	0.01	1	0.223	22.337	1
(200,200)	0.5	1	7.01	11.15	0.033
(500,500)	0.5	1	14.18	22.40	0.043
(200,200)	0.5	100	7.03	1411.81	1
(500,500)	0.5	100	11.15	2234.62	1

Table 8: Rejection rate under different assumptions on the eigenvalues of the rotation matrix.

7 Data application

To illustrate our procedure, we study the US macroeconomic data set of Stock and Watson (2009). The original data set contains 108 monthly and 79 quarterly time series of US nominal and real variables, including prices, interest rates, money and credit aggregates, stock prices, exchange rates. We use here preprocessed data following the suggestions in Table A.1 in Stock and Watson (2009). The US quarterly data are taken from the DRI/McGraw-Hill Basic Economics database of 1999.[1] Following Stock and Watson (2009), we then removed some high level aggregates related by identities to the lower level sub-aggregates and ended up with $d = 109$ time series, spanning from 1959:Q3–2006:Q4 and having a length of $T = 190$.

This data set has been studied in many methodological papers concerning factor models. We refer to Section 6 in Baltagi et al. (2021) for a detailed analysis of the data set, including the estimation of multiple change-points in the loading matrices and the number of factors in each sub-interval. Baltagi et al. (2021) detected two change-points 1979:Q1 and 1983:Q4. In particular, Baltagi et al. (2021) hypothesized that the first break could be due to the impact of the Iranian revolution on the oil price and US inflation. The second break could be due to the great moderation, which is also considered in Chen et al. (2014) and Ma and Su (2018). The corresponding estimated

[1] https://dataverse.unc.edu/dataset.xhtml?persistentId=hdl:1902.29/D-17267

number of factors in the three regimes are $r_1 = 3, r_2 = 3$ and $r_3 = 4$, respectively. Our test allows one now to study what type of breaks occur.

First, we apply our test modified for the analysis of change-points to the first and second regime. Our test rejects the null hypothesis at a 5% significance level. In other words, we reject the hypothesis that the loading matrices in the first and second regimes can be expressed as linear combinations of each other. This gives evidence that a non-negligible portion of the cross sections have linearly independent loadings.

Secondly, we apply our test for different numbers of factors as presented in Section 5.1 and adapted to the change-point setting to the second and third regimes. In this case, our test does not reject the null hypothesis at a 5% significance level. That is, there is enough evidence to conclude that there is a change in the loading matrices, but not that those loadings are linearly independent. As mentioned above, the second break has been considered in many works, including Chen et al. (2014), Ma and Su (2018), Baltagi et al. (2021). It has been hypothesized to be due to the great moderation. The great moderation refers to a period of decreased macroeconomic volatility experienced in the United States starting in the 1980s. For example, during this period, the standard deviation of quarterly real gross domestic product declined by half and the standard deviation of inflation declined by two-thirds, according to figures reported by former U.S. Federal Reserve Chair Ben Bernanke.[2] Our test might indicate that the changes in volatility happened in a somewhat more "structured" way alluding to similarities between the pre- and post-break volatilities of the factors up to transformation with an invertible matrix. We also refer to Example 2.1 for one possibility of what a more "structured" change in volatility can look like.

8 Conclusions

This work advances the analysis of factor models in several directions. On the one hand, the proposed hypothesis test is motivated by learning relationships across multiple subjects, also known as multi-subject learning. In particular, it allows discovering whether structural differences between two models are due to a linear transformation. On the other hand, our test pushes forward the framework of testing for structural changes over time in high-dimensional factor models. Our test allows to further characterize what type of change occurs.

The questions of interest required the development of a new hypothesis test and its theoretical investigation under the null hypothesis and the alternative. Our simulation study indicates a good numerical performance of our test. In addition, we present interesting new findings on the celebrated US macroeconomic data set of Stock and Watson.

Despite our contributions, there is still a lot of room for future work on partial testing, in particular, on finding an efficient way of testing whether a partial set of vectors in the loadings is shared across individuals or between pre- and post-break point. Another direction is the investigation of other change-point tests. Since our procedure does not rely on a specific change-point test, one could use our proposed data transformation to apply other such tests. Furthermore, it would be of interest to test whether the rotational transformation matrix is of a certain, for example diagonal, form.

[2]The Federal Reserve Board. "Remarks by Governor Ben S. Bernanke. https://www.federalreserve.gov/boarddocs/speeches/2004/20040220/".

A Proofs of results in Section 3.4 and statements of auxiliary results

We state here the general idea of the proof. Recall from Section 3.1 that our testing procedure transforms the data such that under the null hypothesis, there is no change in the loadings, while under the alternative, there is a change in the loadings. The transformation was motivated by the goal of applying an available change-point test as the one introduced in Han and Inoue (2015) to the transformed data. This is also reflected in the proofs which are adapted from those in Han and Inoue (2015). It turns out to be sufficient to prove that the PCA estimators of the transformed series are consistent and subsequently apply the results in Han and Inoue (2015). Set $\delta_{dT} = \min\{\sqrt{d}, \sqrt{T}\}$ and recall the notation $F^b, F^a, \widehat{F}, H$ and $F' = (F^{b'}, F^{a'})$ from Section 3.2. We further introduce $\widetilde{F}\widetilde{H} = (H_1' F^{b'}, H_2' F^{a'})'$ for matrices H_1, H_2 specified in (B.8) below. To help the reader follow the arguments, we display these and other quantities indicating their dimensions:

$$\underbrace{F'}_{r\times T} = (\underbrace{F^{b'}}_{r\times T/2}, \underbrace{F^{a'}}_{r\times T/2}),\ \underbrace{\widehat{F}'}_{r\times T},\ \underbrace{\widetilde{F}\widetilde{H}}_{T\times r} = (\underbrace{H_1' F^{b'}}_{r\times T/2}, \underbrace{H_2' F^{a'}}_{r\times T/2})',\ \underbrace{\widetilde{F}}_{T\times 2r}\underbrace{\widetilde{H}}_{2r\times r},\ \underbrace{H}_{r\times r}.$$

Then, in order to prove Propositions 3.1–3.3, it suffices to establish the following to be stated more formally below:

- Proposition B.1 below states that
$$\frac{1}{T}\|\widehat{F} - \widetilde{F}\widetilde{H}\|_F^2 = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^2}\right), \text{ as } d, T \to \infty, \tag{A.1}$$
which is analogous to Lemma 1 in Han and Inoue (2015).

- Proposition B.2 below states that
$$\frac{1}{T}\|(\widehat{F} - \widetilde{F}\widetilde{H})' F\|_F = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^2}\right), \text{ as } d, T \to \infty, \tag{A.2}$$
which is analogous to Lemma 2 in Han and Inoue (2015).

- Lemma B.1 below states
$$H_1 - H_0 = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}}\right) \text{ and } H_2 - H_0 = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}}\right), \text{ as } d, T \to \infty, \tag{A.3}$$
which is analogous to Lemma 6 in Han and Inoue (2015).

Note that H_1, H_2 have the same limit as H which was introduced in (3.21). In particular, that means asymptotically $\widetilde{F}\widetilde{H} \approx FH \approx FH_0$ with high probability.

The asymptotics (A.1) and (A.2) are the key results to proving the convergence of our test statistic under the null hypothesis. The bottom line is that our main contribution is to show that (A.1) and (A.2) are satisfied for the PCA estimator based on the transformed series Y.

Inferring Propositions 3.1–3.3 from Propositions B.1 and B.2 follows arguments in Han and Inoue (2015). Therefore, we omitted those in the article. For completeness, we provide a supplementary document Düker and Pipiras (2023) with the detailed proofs. Recall also, that our estimation procedure and our proofs for consistent estimation of the underlying factor model follow the representation in Doz et al. (2011). In contrast, Han and Inoue (2015) follow Bai (2003). While both lead to the same results, they differ in their assumptions. The supplementary document Düker and Pipiras (2023) also has the purpose of clarifying that the arguments in Han and Inoue (2015) go through by using the assumptions in Doz et al. (2011).

Proof of Proposition 3.1: By Assumption W.1(ii), it suffices to prove that

$$|W(\widehat{F}) - W(FH_0)| = o_{\mathrm{p}}(1).$$

The result then follows as in the proof of Theorem 3(i) in Han and Inoue (2015) by using Propositions 3.2 and 3.3. (Proposition 3.2 is analogous to Theorem 1 and Proposition 3.3 to Theorem 2 in Han and Inoue (2015).) □

Proof of Proposition 3.2: The proof follows the same arguments as the proof of Theorem 1 in Han and Inoue (2015). It requires our results (A.1) (Proposition B.1), (A.2) (Proposition B.2) and Lemma B.1. □

Proof of Proposition 3.3: The proof follows the same arguments as the proof of Theorem 2 in Han and Inoue (2015). □

The proofs use existing results for the estimation of the model parameters of the individual series as given in Doz et al. (2011). For the reader's convenience, we collect the results of Doz et al. (2011) as needed here.

Set $\widehat{\Sigma}_k = \frac{1}{T}\sum_{t=1}^{T} X_t^k X_t^{k'}$, $k = 1, 2$, and write $\bar{Q}_r^k$ for the eigenvectors corresponding to the r largest eigenvalues of $\widehat{\Sigma}_k$. We further consider the eigendecomposition

$$\Lambda_k' \Lambda_k = R^k \Pi_k R^{k'},$$

where the diagonal matrix Π_k consists of the eigenvalues in decreasing order and the orthogonal matrix R^k consists of the corresponding eigenvectors. Choose the $d \times r$ matrices

$$Q^k = \Lambda_k R^k \Pi_k^{-\frac{1}{2}} \tag{A.4}$$

such that $\Lambda_k \Lambda_k' = Q^k \Pi_k Q^{k'}$ and $Q^{k'} Q^k = I_r$, $k = 1, 2$. Then, under Assumptions A.1–A.2, CR.1–CR.3, we get

DGR.1. By Lemma 2(i) in Doz et al. (2011), $\frac{1}{d}\|\widehat{\Sigma}_k - \Lambda_k \Lambda_k'\| = \mathcal{O}\big(\frac{1}{d}\big) + \mathcal{O}_{\mathrm{p}}\Big(\frac{1}{\sqrt{T}}\Big)$ for $k = 1, 2$.

DGR.2. By Lemma 4(i) in Doz et al. (2011), $\bar{Q}_r^{k'} Q^k - I_r = \mathcal{O}_{\mathrm{p}}\big(\frac{1}{d}\big) + \mathcal{O}_{\mathrm{p}}\Big(\frac{1}{\sqrt{T}}\Big)$ for $k = 1, 2$.

DGR.3. By Lemma 4(ii) in Doz et al. (2011), $\|\bar{Q}_r^k - Q^k\|^2 = \mathcal{O}_{\mathrm{p}}\big(\frac{1}{d}\big) + \mathcal{O}_{\mathrm{p}}\Big(\frac{1}{\sqrt{T}}\Big)$ for $k = 1, 2$.

Note that $\bar{Q}_r^k, Q^k$ and Π^k are denoted by $\widehat{P}, P_0$ and D_0, respectively, in Doz et al. (2011). Note also that DGR.1 implies the analogous results for $\frac{1}{T/2}\sum_{t=1}^{T/2} X_t^k X_t^{k'}$ and $\frac{1}{T/2}\sum_{t=T/2+1}^{T} X_t^k X_t^{k'}$, $k = 1, 2$. We use the notation $\widehat{\Sigma}_2$ for either of those quantities.

Our analysis could as well be based on results similar to DGR.1–DGR.2 in Bai and Ng (2002) and Bai (2003). We refer to Doz et al. (2011) since their statements are expressed more conveniently for our purposes.

B Proofs of auxiliary results under null hypothesis

We first recall some relationships between different PCA estimators used in the factor model literature. Those relationships are collected in a more comprehensive way in Chapter 3 in Bai and Ng (2008). Recall from (3.16) the PCA estimators for the factors and loadings

$$\widehat{F} = \sqrt{T}\widehat{Q}_r, \quad \widehat{\Lambda}' = \frac{1}{T}\widehat{F}'Y, \tag{B.1}$$

where $\widehat{Q}_r$ are the eigenvectors corresponding to the r largest eigenvalues of the $T \times T$ matrix YY'. On the other hand, we can also define

$$\bar{F} = \frac{1}{d}Y\bar{\Lambda}, \quad \bar{\Lambda} = \sqrt{d}\bar{Q}_r \tag{B.2}$$

with $\bar{Q}_r$ being the r eigenvectors corresponding to the r largest eigenvalues of the $d \times d$ matrix $Y'Y$. Note that $\bar{Q}_r$ and $\bar{Q}_r^k$ (from Appendix A) are defined in the same way but for Y and X^k, respectively. We further introduce the matrix $\widehat{V}_r$, which is a diagonal matrix consisting of the r largest eigenvalues of $\frac{1}{dT}YY'$, in particular, $\lambda_i\big(\frac{1}{dT}YY'\big) = \lambda_i\big(\frac{1}{dT}Y'Y\big)$ for all i such that λ_i is nonzero.

By equation (3.2) in Bai and Ng (2008), the two ways of writing the PCA estimators relate to each other as

$$\widehat{F}' = \widehat{V}_r^{-\frac{1}{2}}\bar{F}', \quad \bar{\Lambda} = \widehat{\Lambda}\widehat{V}_r^{-\frac{1}{2}}. \tag{B.3}$$

Then,

$$\widehat{\Lambda} = \bar{\Lambda}\widehat{V}_r^{\frac{1}{2}} = \sqrt{d}\bar{Q}_r\widehat{V}_r^{\frac{1}{2}} = \bar{Q}_r\widehat{\Pi}_r^{\frac{1}{2}}, \tag{B.4}$$

where $\widehat{\Pi}_r$ is the diagonal matrix with the r largest eigenvalues of the $d \times d$ matrix $\frac{1}{T}Y'Y$. The same relation also hold for $k = 1, 2$ when using data X^k rather than Y. This leads to an important observation regarding our estimators (3.4) for the projection matrices:

$$\widehat{P}_k = \widehat{\Lambda}_k(\widehat{\Lambda}_k'\widehat{\Lambda}_k)^{-1}\widehat{\Lambda}_k' = \bar{Q}_r^k\bar{Q}_r^{k'}, \tag{B.5}$$

which follows since projection matrices are invariant under transformation with nonsingular matrices (see (3.5)) and $\bar{Q}_r^{k'}\bar{Q}_r^k = I_r$ due to orthonormality of the eigenvectors.

Before moving on to our auxiliary results, note from (3.7) and $Y = Y_{1:T}'$ in (3.15) that

$$YY' = \begin{pmatrix} X_{1:T/2}^{2'}\widehat{P}_I^{1'}\widehat{P}_I^1X_{1:T/2}^2 & X_{1:T/2}^{2'}\widehat{P}_I^{1'}\widehat{P}_I^2X_{(T/2+1):T}^2 \\ X_{(T/2+1):T}^{2'}\widehat{P}_I^{2'}\widehat{P}_I^1X_{1:T/2}^2 & X_{(T/2+1):T}^{2'}\widehat{P}_I^{2'}\widehat{P}_I^2X_{(T/2+1):T}^2 \end{pmatrix}. \tag{B.6}$$

For instance, by (3.15), the upper right block is

$$\begin{aligned} X_{1:T/2}^{2'}\widehat{P}_I^{1'}\widehat{P}_I^2X_{(T/2+1):T}^2 &= F^b\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^2\Lambda_2F^{a'} + F^b\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^2\varepsilon^{a'} \\ &\quad + \varepsilon^b\widehat{P}_I^{1'}\widehat{P}_I^2\Lambda_2F^{a'} + \varepsilon^b\widehat{P}_I^{1'}\widehat{P}_I^2\varepsilon^{a'}. \end{aligned} \tag{B.7}$$

Similarly to F^b, F^a, set $\widehat{F}^b = \widehat{F}_{1:T/2}'$ and $\widehat{F}^b = \widehat{F}_{(T/2+1):T}'$. We introduce the matrices

$$\begin{aligned} H_1 &:= \frac{1}{dT}(\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^1\Lambda_2F^{b'}\widehat{F}^b + \Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^2\Lambda_2F^{a'}\widehat{F}^a)\widehat{V}_r^{-1}, \\ H_2 &:= \frac{1}{dT}(\Lambda_2'\widehat{P}_I^{2'}\widehat{P}_I^2\Lambda_2F^{b'}\widehat{F}^b + \Lambda_2'\widehat{P}_I^{2'}\widehat{P}_I^1\Lambda_2F^{a'}\widehat{F}^a)\widehat{V}_r^{-1} \end{aligned} \tag{B.8}$$

such that

$$
\begin{aligned}
\widetilde{F}\widetilde{H} &:= \begin{pmatrix} F^b & 0 \\ 0 & F^a \end{pmatrix} \begin{pmatrix} H_1 \\ H_2 \end{pmatrix} = \begin{pmatrix} F^b H_1 \\ F^a H_2 \end{pmatrix} \\
&= \frac{1}{dT}\widetilde{F} \begin{pmatrix} \Lambda_2' \widehat{P}_I^{1\prime} \widehat{P}_I^1 \Lambda_2 & \Lambda_2' \widehat{P}_I^{1\prime} \widehat{P}_I^2 \Lambda_2 \\ \Lambda_2' \widehat{P}_I^{2\prime} \widehat{P}_I^1 \Lambda_2 & \Lambda_2' \widehat{P}_I^{2\prime} \widehat{P}_I^2 \Lambda_2 \end{pmatrix} \widetilde{F}' \widehat{F} \widehat{V}_r^{-1} \\
&=: \frac{1}{dT} \widetilde{F} \Lambda_P \widetilde{F}' \widehat{F} \widehat{V}_r^{-1}. \qquad \text{(B.9)}
\end{aligned}
$$

We now formalize and prove the key results (A.1)–(A.3). Those results are crucial to prove Propositions 3.1–3.3 as stated in Appendix A. Recall the notation $\delta_{dT} = \min\{\sqrt{d}, \sqrt{T}\}$.

Proposition B.1. *Under Assumptions A.1–A.2, CR.1–CR.4,*

$$
\frac{1}{T}\|\widehat{F} - \widetilde{F}\widetilde{H}\|_F^2 = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^2}\right), \text{ as } d, T \to \infty.
$$

Proof: Recall the transformed series $Y_{1:T} = \left(\widehat{P}_I^1 X_{1:T/2}^2, \widehat{P}_I^2 X_{(T/2+1):T}^2\right)$ from (3.7) as well as $Y' := Y_{1:T}$ from (3.15). Due to (B.1), we get $\frac{1}{dT} YY'\widehat{F} = \widehat{F}\widehat{V}_r$ such that $\widehat{F} = \frac{1}{dT} YY'\widehat{F}\widehat{V}_r^{-1}$. Then, given (B.6) and (B.9),

$$
\begin{aligned}
\widehat{F} - \widetilde{F}\widetilde{H} &= \frac{1}{dT} YY'\widehat{F}\widehat{V}_r^{-1} - \frac{1}{dT}\widetilde{F}\Lambda_P \widetilde{F}'\widehat{F}\widehat{V}_r^{-1} \\
&= \frac{1}{dT}(YY' - \widetilde{F}\Lambda_P\widetilde{F}')(\widehat{F} - F)\widehat{V}_r^{-1} + \frac{1}{dT}(YY' - \widetilde{F}\Lambda_P\widetilde{F}')F\widehat{V}_r^{-1}. \qquad \text{(B.10)}
\end{aligned}
$$

Note further from (B.6), (B.7) and (B.8) that

$$
YY' - \widetilde{F}\Lambda_P\widetilde{F}' = \begin{pmatrix} A & C \\ C' & B \end{pmatrix} \qquad \text{(B.11)}
$$

with

$$
\begin{aligned}
A &= F^b \Lambda_2' \widehat{P}_I^{1\prime} \widehat{P}_I^1 \varepsilon^{b\prime} + \varepsilon^b \widehat{P}_I^{1\prime} \widehat{P}_I^1 \Lambda_2 F^{b\prime} + \varepsilon^b \widehat{P}_I^{1\prime} \widehat{P}_I^1 \varepsilon^{b\prime}, \\
B &= F^a \Lambda_2' \widehat{P}_I^{2\prime} \widehat{P}_I^2 \varepsilon^{a\prime} + \varepsilon^a \widehat{P}_I^{2\prime} \widehat{P}_I^2 \Lambda_2 F^{a\prime} + \varepsilon^a \widehat{P}_I^{2\prime} \widehat{P}_I^2 \varepsilon^{a\prime}, \\
C &= F^b \Lambda_2' \widehat{P}_I^{1\prime} \widehat{P}_I^2 \varepsilon^{a\prime} + \varepsilon^b \widehat{P}_I^{1\prime} \widehat{P}_I^2 \Lambda_2 F^{a\prime} + \varepsilon^b \widehat{P}_I^{1\prime} \widehat{P}_I^2 \varepsilon^{a\prime}.
\end{aligned}
$$

After applying the Frobenius norm and triangular inequality, we consider the two summands in (B.10) separately. For the first summand in (B.10), we get

$$
\begin{aligned}
&\frac{1}{T}\|\frac{1}{dT}(YY' - \widetilde{F}\Lambda_P\widetilde{F}')(\widehat{F} - F)\widehat{V}_r^{-1}\|_F^2 \\
&\leq \|\frac{1}{dT}(YY' - \widetilde{F}\Lambda_P\widetilde{F}')\|^2 \|\widehat{V}_r^{-1}\|^2 \frac{1}{T}\|\widehat{F} - F\|_F^2 \qquad \text{(B.12)} \\
&\leq \frac{1}{(dT)^2}(\|A\|^2 + \|B\|^2 + 2\|C\|^2)\|\widehat{V}_r^{-1}\|^2 \frac{1}{T}\|\widehat{F} - F\|_F^2 \qquad \text{(B.13)} \\
&= \frac{1}{(dT)^2}(\|A\|^2 + \|B\|^2 + 2\|C\|^2)\|\widehat{V}_r^{-1}\|^2 \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^2}\right) \qquad \text{(B.14)} \\
&= \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^2}\right), \qquad \text{(B.15)}
\end{aligned}
$$

where (B.12) follows by M.1 in Appendix D below. The relation (B.13) is due to the block representation (B.11) and M.2. Then, (B.14) follows by Proposition C.1 below. For the remaining quantities in (B.14), we focus on $\|A\|^2$ since the matrices B, C can be treated similarly. We have

$$\begin{aligned}\frac{1}{(dT)^2}\|A\|^2\|\widehat{V}_r^{-1}\|^2 &= \frac{1}{(dT)^2}\|F^b\Lambda_2'\widehat{P}_I^{1\prime}\widehat{P}_I^1\varepsilon^{b\prime} + \varepsilon^b\widehat{P}_I^{1\prime}\widehat{P}_I^1\Lambda_2F^{b\prime} + \varepsilon^b\widehat{P}_I^{1\prime}\widehat{P}_I^1\varepsilon^{b\prime}\|^2\|\widehat{V}_r^{-1}\|^2 \\ &\leq 3\|\widehat{P}_I^1\|^2\|\widehat{P}_I^1\|^2\left(\frac{1}{T}\|F\|^2\frac{1}{d}\|\Lambda_2\|^2\frac{1}{dT}\|\varepsilon\|^2 + \frac{1}{(dT)^2}\|\varepsilon\|^4\right)\|\widehat{V}_r^{-1}\|^2 \qquad \text{(B.16)} \\ &= \mathcal{O}_{\mathrm{p}}(1), \qquad \text{(B.17)}\end{aligned}$$

where (B.16) is due to submultiplicativity of the spectral norm, and the fact that for instance, $\|F^b\|^2 = \lambda_{\max}(F^{b\prime}F^b) \leq \lambda_{\max}(F^{b\prime}F^b + F^{a\prime}F^a) = \|F\|^2$ by M.3. For (B.17), observe that

$$\begin{aligned}\|P_{0,k}\| &= \sqrt{\lambda_{\max}(P_{0,k}'P_{0,k})} = \sqrt{\lambda_{\max}(P_{0,k})} = \sqrt{\lambda_{\max}(\Lambda_k(\Lambda_k'\Lambda_k)^{-1}\Lambda_k')} \\ &= \sqrt{\lambda_{\max}(\Lambda_k'\Lambda_k(\Lambda_k'\Lambda_k)^{-1})} = \sqrt{\lambda_{\max}(I_r)} = 1, \quad k = 1, 2,\end{aligned} \qquad \text{(B.18)}$$

and similarly for their estimated counterparts $\|\widehat{P}_k\| = 1$, $k = 1, 2$, so that $\|\widehat{P}_I^k\| = \frac{1}{2}\|I_d + \widehat{P}_k\| \leq 1$. For the remaining terms in (B.16), we get:

S.1. $\frac{1}{T}\|F\|^2 = \mathcal{O}_{\mathrm{p}}(1)$ since $\frac{1}{T}\sum_{t=1}^T F_tF_t' \xrightarrow{\mathrm{p}} \Sigma_F$ by the discussion following Assumption A.2.

S.2. $\|\varepsilon\varepsilon'\| \leq T\|\frac{1}{T}\sum_{t=1}^T \varepsilon_t\varepsilon_t' - \Sigma_\varepsilon\| + T\|\Sigma_\varepsilon\| = \mathcal{O}_{\mathrm{p}}\left(d\sqrt{T}\right) + \mathcal{O}(T)$ which follows by Assumption A.2 (see p. 199 in Doz et al. (2011)) and by Assumption CR.3, respectively.

S.3. $\|\widehat{V}_r^{-1}\| = \|d\widehat{\Pi}_r^{-1}\| \leq d\|\widehat{\Pi}_r^{-1} - \Pi_2^{-1}\| + \|(\Pi_2/d)^{-1}\| = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right) + \mathcal{O}(1) = \mathcal{O}_{\mathrm{p}}(1)$ by Corollary C.1*(ii)*.

For the second summand in (B.10), we can infer, with further explanations given below,

$$\begin{aligned}&\frac{1}{T}\|\frac{1}{dT}(YY' - \widetilde{F}\Lambda_P\widetilde{F}')F\widehat{V}_r^{-1}\|_F^2 \\ &\leq \frac{1}{T}\|\frac{1}{dT}(YY' - \widetilde{F}\Lambda_P\widetilde{F}')F\|_F^2\|\widehat{V}_r^{-1}\|^2 \qquad \text{(B.19)} \\ &= \frac{1}{T}\frac{1}{(dT)^2}\left\|\begin{pmatrix} A & C \\ C' & B\end{pmatrix}F\right\|_F^2\|\widehat{V}_r^{-1}\|^2 \qquad \text{(B.20)} \\ &= \frac{1}{T}\frac{1}{(dT)^2}\left\|\begin{pmatrix} AF^b + CF^a \\ C'F^b + BF^a\end{pmatrix}\right\|_F^2\|\widehat{V}_r^{-1}\|^2 \\ &\leq \frac{1}{T}\frac{1}{(dT)^2}2(\|AF^b\|_F^2 + \|CF^a\|_F^2 + \|C'F^b\|_F^2 + \|BF^a\|_F^2)\|\widehat{V}_r^{-1}\|^2 \qquad \text{(B.21)} \\ &= \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^2}\right), \qquad \text{(B.22)}\end{aligned}$$

where (B.19) follows again by M.1, (B.20) uses the notation (B.11), and (B.21) is due to properties of the Frobenius norm. The term $\|\widehat{V}_r^{-1}\|^2$ in (B.21) can be bounded by S.3 above. For the other terms in (B.21), we focus on the first summand

$$AF^b = F^b\Lambda_2'\widehat{P}_I^{1\prime}\widehat{P}_I^1\varepsilon^{b\prime}F^b + \varepsilon^b\widehat{P}_I^{1\prime}\widehat{P}_I^1\Lambda_2F^{b\prime}F^b + \varepsilon^b\widehat{P}_I^{1\prime}\widehat{P}_I^1\varepsilon^{b\prime}F^b. \qquad \text{(B.23)}$$

For the first summand in (B.23), we get

$$\begin{aligned}\frac{1}{d^2T^3}\|F^b\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^1\varepsilon^{b'}F^b\|_F^2 &\leq \frac{1}{d}\frac{1}{T}\|F\|^2\|\widehat{P}_I^1\|^2\|\widehat{P}_I^1\|^2\frac{1}{d}\|\Lambda_2'\|^2\|\frac{1}{T}\varepsilon^{b'}F^b\|_F^2\\ &= \frac{1}{d}\mathcal{O}_{\mathrm{p}}(1)\mathcal{O}(1)\mathcal{O}_{\mathrm{p}}\left(\frac{d}{T}\right) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{T}\right) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^2}\right),\end{aligned}$$

where the asymptotics of the terms is gotten as in (B.17), except for $\varepsilon^{b'}F^b$. The asymptotics of $(\varepsilon^{b'}F^b)/T$ follow by:

S.4. $\frac{1}{T}\|\varepsilon'F\|_F^2 = \mathcal{O}_{\mathrm{p}}(dT)$ due to the same calculations as done on p. 199 in Doz et al. (2011) as part of their proof of Lemma 2(i) under Assumptions A.2 and CR.3.

For the second summand in (B.23),

$$\frac{1}{d^2T^3}\|\varepsilon^b\widehat{P}_I^{1'}\widehat{P}_I^1\Lambda_2F^{b'}F^b\|_F^2 \leq \frac{1}{d^2T}\|\varepsilon^b\widehat{P}_I^{1'}\widehat{P}_I^1\Lambda_2\|^2\left(\frac{1}{T}\|F\|_F^2\right)^2 = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^2}\right)$$

by S.1, and since

$$\begin{aligned}\frac{1}{d^2T}\|\varepsilon^b\widehat{P}_I^{1'}\widehat{P}_I^1\Lambda_2\|^2 &\leq \frac{2}{d^2T}\|\varepsilon^b\Lambda_2\|^2 + \frac{2}{d^2T}\|\varepsilon^b(\widehat{P}_I^{1'}\widehat{P}_I^1\Lambda_2 - \Lambda_2)\|^2\\ &= \frac{1}{d^2T}\mathcal{O}_{\mathrm{p}}(dT) + \frac{1}{d^2T}\Big(\mathcal{O}_{\mathrm{p}}\Big(d\sqrt{T}\Big) + \mathcal{O}(T)\Big)d\left(\mathcal{O}_{\mathrm{p}}\left(\frac{1}{d^2}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{T}\right)\right) &\text{(B.24)}\\ &= \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^2}\right), &\text{(B.25)}\end{aligned}$$

where (B.24) is obtained from the following observations. The contribution of $\|\varepsilon^b\|^2 \leq \|\varepsilon\|^2 = \|\varepsilon\varepsilon'\|$ by M.3 is bounded by S.2. The term $\|\widehat{P}_I^{1'}\widehat{P}_I^1\Lambda_2 - \Lambda_2\|^2$ can be dealt with as (B.47), with the difference of extra Λ_2' and $\sqrt{d}$. Furthermore, with $\sigma_{\varepsilon,ij}^2$ denoting the ijth element of Σ_ε,

$$\begin{aligned}\mathrm{E}\,\|\varepsilon^b\Lambda_2\|_F^2 &= \sum_{s=1}^{r}\sum_{t=1}^{T/2}\sum_{i,j=1}^{d}\lambda_{2,is}\lambda_{2,js}\,\mathrm{E}(\varepsilon_{t,i}\varepsilon_{t,j})\\ &\leq \bar{\lambda}^2 r\sum_{t=1}^{T}\sum_{i,j=1}^{d}\sigma_{\varepsilon,ij}^2 &\text{(B.26)}\\ &= T\bar{\lambda}^2 r\|\Sigma_\varepsilon\|_F^2 \leq dT\bar{\lambda}^2 r\|\Sigma_\varepsilon\|^2, &\text{(B.27)}\end{aligned}$$

where (B.26) is due to Assumption CR.4. We further need Assumptions A.2 and CR.3 for the last bound to be $\mathcal{O}(1/d)$. The calculations in (B.26) and (B.27) are essentially the same as those done for the proof of Lemma 1(ii) in Bai and Ng (2002). For the last summand in (B.23),

$$\begin{aligned}\frac{1}{d^2T^3}\|\varepsilon^b\widehat{P}_I^{1'}\widehat{P}_I^1\varepsilon^{b'}F^b\|_F^2 &\leq \frac{1}{d^2T}\|\varepsilon\|^2\|\widehat{P}_I^1\|^2\|\widehat{P}_I^1\|^2\|\frac{1}{T}\varepsilon^{b'}F^b\|_F^2\\ &= \frac{1}{d^2T}\Big(\mathcal{O}_{\mathrm{p}}\Big(\sqrt{T}d\Big) + \mathcal{O}(T)\Big)\mathcal{O}_{\mathrm{p}}\left(\frac{d}{T}\right) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^2}\right)\end{aligned}$$

by S.2 and again by the calculations as on p. 199 in Doz et al. (2011) under Assumptions A.2 and CR.3, and $\|\widehat{P}_I^k\| = \frac{1}{2}\|I_d + \widehat{P}_k\| \leq 1$.

The desired result follows from (B.10), (B.15) and (B.22). □

Proposition B.2. *Under Assumptions A.1–A.3, CR.1–CR.4,*

$$\frac{1}{T^2}\|(\widehat{F}-\widetilde{F}\widetilde{H})'F\|_F^2 = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^4}\right), \text{ as } d,T\to\infty.$$

Proof: As in (B.10), we get

$$\begin{aligned} F'(\widehat{F}-\widetilde{F}\widetilde{H}) &= \frac{1}{dT}F'(YY'\widehat{F}\widehat{V}_r^{-1}-\widetilde{F}\Lambda_P\widetilde{F}'\widehat{F}\widehat{V}_r^{-1}) \\ &= \frac{1}{dT}F'(YY'-\widetilde{F}\Lambda_P\widetilde{F}')(\widehat{F}-F)\widehat{V}_r^{-1}+\frac{1}{dT}F'(YY'-\widetilde{F}\Lambda_P\widetilde{F}')F\widehat{V}_r^{-1}. \end{aligned} \tag{B.28}$$

After applying the Frobenius norm and triangle inequality, we consider the two summands in (B.28) separately. For the first summand in (B.28), using the representation (B.11), we proceed similarly to (B.12)–(B.15) but employ different norms on respective quantities. Note that

$$\begin{aligned} &\frac{1}{T^2}\|\frac{1}{dT}F'(YY'-\widetilde{F}\Lambda_P\widetilde{F}')(\widehat{F}-F)\widehat{V}_r^{-1}\|_F^2 \\ &\le \frac{1}{T}\frac{1}{(dT)^2}\left\|F'\begin{pmatrix} A & C \\ C' & B\end{pmatrix}\right\|^2\frac{1}{T}\|\widehat{F}-F\|_F^2\|\widehat{V}_r^{-1}\|^2 && \text{(B.29)} \\ &= \frac{1}{T}\frac{1}{(dT)^2}\left\|\left(F^{b'}A+F^{a'}C',F^{b'}C+F^{a'}B\right)\right\|^2\frac{1}{T}\|\widehat{F}-F\|_F^2\|\widehat{V}_r^{-1}\|^2 \\ &\le \frac{1}{T}\frac{1}{(dT)^2}\left(\|F^{b'}A+F^{a'}C'\|^2+\|F^{b'}C+F^{a'}B\|^2\right)\frac{1}{T}\|\widehat{F}-F\|_F^2\|\widehat{V}_r^{-1}\|^2 && \text{(B.30)} \\ &\le \frac{1}{T}\frac{1}{(dT)^2}2\left(\|F^{b'}A\|^2+\|F^{a'}C'\|^2+\|F^{b'}C\|^2+\|F^{a'}B\|^2\right)\frac{1}{T}\|\widehat{F}-F\|_F^2\|\widehat{V}_r^{-1}\|^2 && \text{(B.31)} \\ &= \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^4}\right), && \text{(B.32)} \end{aligned}$$

where (B.29) follows by M.1 and the inequality (B.31) is due to M.4. The quantities in (B.31) can then be treated separately. We get (B.32) by using S.3 in the proof of Proposition B.1, Proposition C.1 and suitably bounding the four summands in (B.31) as follows. For example, we give the detailed arguments for the second summand in (B.31); the remaining ones can be treated similarly. We have

$$F^{a'}C = F^{a'}F^b\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^2\varepsilon^{b'}+F^{a'}\varepsilon^b\widehat{P}_I^{1'}\widehat{P}_I^2\Lambda_2F^{a'}+F^{a'}\varepsilon^b\widehat{P}_I^{1'}\widehat{P}_I^2\varepsilon^{a'}. \tag{B.33}$$

After applying the spectral norm and triangle inequality, we consider the summands in (B.33) separately. For the first one,

$$\frac{1}{T}\frac{1}{(dT)^2}\|F^{a'}F^b\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^2\varepsilon^{a'}\|^2 \le \frac{1}{T}\|F^{a'}\|^2\frac{1}{T}\|F^b\|^2\frac{1}{d^2T}\|\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^2\varepsilon^{a'}\|^2 = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^2}\right)=\mathcal{O}_{\mathrm{p}}(1),$$

by S.1 and arguing as for (B.25). For the second summand in (B.33),

$$\frac{1}{T}\frac{1}{(dT)^2}\|F^{a'}\varepsilon^b\widehat{P}_I^{1'}\widehat{P}_I^2\Lambda_2F^{a'}\|^2 \le \frac{1}{T^2}\|F^{a'}\|^4\frac{1}{d^2T}\|\varepsilon^b\widehat{P}_I^{1'}\widehat{P}_I^2\Lambda_2\|^2 = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^2}\right)=\mathcal{O}_{\mathrm{p}}(1),$$

by S.1 and (B.25). For the third summand in (B.33),

$$\frac{1}{T}\frac{1}{(dT)^2}\|F^{a'}\varepsilon^b\widehat{P}_I^{1'}\widehat{P}_I^2\varepsilon^{a'}\|^2 \le \frac{1}{T}\|F^{a'}\|^2\frac{1}{dT}\|\varepsilon^b\|^2\|\widehat{P}_I^{1'}\widehat{P}_I^2\|^2\frac{1}{dT}\|\varepsilon^a\|^2 = \mathcal{O}_{\mathrm{p}}(1)$$

by S.1 and S.2 and since $\|\widehat{P}_I^k\| = \frac{1}{2}\|I_d + \widehat{P}_k\| \le 1$.

For the second summand in (B.28), we get as in (B.19)–(B.22), with further explanations given below,

$$\begin{aligned}
&\frac{1}{T^2}\frac{1}{(dT)^2}\|F'(YY' - \widetilde{F}\Lambda_P\widetilde{F}')F\widehat{V}_r^{-1}\|_F^2 \\
&\le \frac{2}{T^2}\frac{1}{(dT)^2}\Big(\|F^{b'}AF^b\|_F^2 + \|F^{a'}C'F^b\|_F^2 + \|F^{b'}CF^a\|_F^2 + \|F^{a'}BF^a\|_F^2\Big)\|\widehat{V}_r^{-1}\|^2 && \text{(B.34)}\\
&= \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^4}\right). && \text{(B.35)}
\end{aligned}$$

For (B.35), we focus on $F^{b'}AF^b$, that is,

$$F^{b'}AF^b = F^{b'}F^b\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^1\varepsilon^{b'}F^b + F^{b'}\varepsilon^b\widehat{P}_I^{1'}\widehat{P}_I^1\Lambda_2F^{b'}F^b + F^{b'}\varepsilon^b\widehat{P}_I^{1'}\widehat{P}_I^1\varepsilon^{b'}F^b. \qquad \text{(B.36)}$$

Before we consider the individual summands in (B.36) to prove (B.35) for $\|F^{b'}AF^b\|_F^2$, note that

$$\begin{aligned}
\frac{1}{d^2T^2}\|\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^1\varepsilon^{b'}F^b\|_F^2 &\le \frac{2}{d^2T^2}\|\Lambda_2'\varepsilon^{b'}F^b\|_F^2 + \frac{2}{d^2T^2}\|(\Lambda_2' - \Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^1)\varepsilon^{b'}F^b\|_F^2 \\
&\le \frac{2}{d^2T^2}\|\Lambda_2'\varepsilon^{b'}F^b\|_F^2 + \frac{2}{d^2T}\|\Lambda_2' - \Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^1\|^2\frac{1}{T}\|\varepsilon^{b'}F^b\|_F^2 && \text{(B.37)}\\
&= \frac{1}{d^2T^2}\mathcal{O}_{\mathrm{p}}(dT) + \frac{1}{d^2T}d\left(\mathcal{O}_{\mathrm{p}}\left(\frac{1}{d^2}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{T}\right)\right)\mathcal{O}_{\mathrm{p}}(dT) && \text{(B.38)}\\
&= \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^4}\right), && \text{(B.39)}
\end{aligned}$$

where (B.37) is due to M.1 and (B.38) can be argued similarly to (B.24). Indeed, the asymptotics of $\|\varepsilon^{b'}F^b\|_F^2$ follow from S.4 and $\Lambda_2' - \Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^1$ can be dealt with as (B.47). For $\Lambda_2'\varepsilon^{b'}F^b = \left(\sum_{i=1}^d\sum_{t=1}^{T/2}\lambda_{2,is_1}\varepsilon_{t,i}F_{t,s_2}\right)_{s_1,s_2=1,\dots,r}$,

$$\begin{aligned}
\mathrm{E}\,\|\Lambda_2'\varepsilon^{b'}F^b\|_F^2 &= \sum_{s_1,s_2=1}^{r}\sum_{t_1,t_2=1}^{T/2}\sum_{i,j=1}^{d}\lambda_{2,is_1}\lambda_{2,js_1}\,\mathrm{E}(F_{t_1,s_1}F_{t_2,s_2}\varepsilon_{t_1,i}\varepsilon_{t_2,j}) \\
&= \sum_{s_1,s_2=1}^{r}\sum_{t_1,t_2=1}^{T/2}\sum_{i,j=1}^{d}\lambda_{2,is_1}\lambda_{2,js_1}\Gamma_{F,s_1s_2}(t_1-t_2)\,\mathrm{E}(\varepsilon_{t_1,i}\varepsilon_{t_2,j}) && \text{(B.40)}\\
&\le \bar{\lambda}^2\sum_{s_1,s_2=1}^{r}T\sum_{k\in\mathbb{Z}}|\Gamma_{F,s_1s_2}(k)|\max_{t_1,t_2=1,\dots,T}\sum_{i,j=1}^{d}|\,\mathrm{E}(\varepsilon_{t_1,i}\varepsilon_{t_2,j})| && \text{(B.41)}\\
&\le \bar{\lambda}^2r^2T\sum_{k\in\mathbb{Z}}\|\Gamma_F(k)\|dM = \mathcal{O}(dT), && \text{(B.42)}
\end{aligned}$$

where (B.40) is due to the independence of factors and errors by Assumption A.2, and (B.41) is due to Assumption CR.4. Finally, (B.42) is due to Assumption A.3(i). Note that (B.42) is what is stated as Assumption 6(b) in Han and Inoue (2015). In Doz et al. (2011) as well as here, this assumption is covered by the additional structural assumptions on the factors in Assumption A.2 which imply the summability of their autocovariances.

For the first summand in (B.36), we get

$$\frac{1}{d^2T^4}\|F^{b'}F^b\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^1\varepsilon^{b'}F^b\|_F^2 \le \frac{1}{T^2}\|F\|^4\frac{1}{d^2T^2}\|\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^1\varepsilon^{b'}F^b\|_F^2$$

$$= \mathcal{O}_{\mathrm{p}}(1)\mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^4}\right) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^4}\right),$$

by (B.39) and S.1 above. For the second summand in (B.36),

$$\begin{aligned}
&\frac{1}{d^2T^4}\|F^{b'}\varepsilon^b\widehat{P}_I^{1'}\widehat{P}_I^1\Lambda_2F^{b'}F^b\|_F^2\\
&\leq \frac{1}{d^2T^2}\|F^{b'}\varepsilon^b\widehat{P}_I^{1'}\widehat{P}_I^1\Lambda_2\|_F^2\left(\frac{1}{T}\|F\|^2\right)^2 = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^4}\right),
\end{aligned}$$

again by (B.39) and S.1. For the last summand in (B.36),

$$\frac{1}{d^2T^4}\|F^{b'}\varepsilon^b\widehat{P}_I^{1'}\widehat{P}_I^1\varepsilon^{b'}F^b\|_F^2 \leq \frac{1}{d^2}\|\frac{1}{T}\varepsilon^{b'}F^b\|_F^4 = \frac{1}{d^2}\mathcal{O}_{\mathrm{p}}\left(\frac{d^2}{T^2}\right) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{T^2}\right) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}^4}\right)$$

by S.4 which yields (B.35).

The desired result follows from (B.28), (B.32) and (B.35). □

Lemma B.1. *Recall H_1, H_2 from (B.8). Under Assumptions A.1–A.2, CR.1–CR.3,*

$$H_1 - H_0 = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}}\right) \text{ and } H_2 - H_0 = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}}\right), \text{ as } d, T \to \infty,$$

where

$$H_0 = Q^2\widetilde{\Pi}_2Q^{2'}\Sigma_F\widetilde{\Pi}_2^{-1}. \tag{B.43}$$

Proof: We only show the result for H_1; the second result follows from analogous considerations. Recall from (3.21) that $H = (\Lambda_2'\Lambda_2/d)(F'\widehat{F}/T)\widehat{V}_r^{-1}$ and consider

$$\|H_1 - H_0\| \leq \|H_1 - H\| + \|H - H_0\|.$$

Given Lemma B.2 below, it is enough to show the asymptotics of $\|H_1 - H\|$.

With further explanations given below, we get

$$\begin{aligned}
&\|H_1 - H\|\\
&= \frac{1}{dT}\left\|\left(\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^1\Lambda_2F^{b'}\widehat{F}^b + \Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^2\Lambda_2F^{a'}\widehat{F}^a\right)\widehat{V}_r^{-1} - \Lambda_2'\Lambda_2F'\widehat{F}\widehat{V}_r^{-1}\right\|\\
&\leq \frac{1}{dT}\left\|\left(\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^1\Lambda_2F^{b'}, \Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^2\Lambda_2F^{a'}\right) - \Lambda_2'\Lambda_2F'\right\|\|\widehat{F}\|\|\widehat{V}_r^{-1}\|\\
&= \frac{1}{dT}\left\|\left(\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^1\Lambda_2F^{b'}, \Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^2\Lambda_2F^{a'}\right) - (\Lambda_2'\Lambda_2F^{b'}, \Lambda_2'\Lambda_2F^{a'})\right\|\|\widehat{F}\|\|\widehat{V}_r^{-1}\|\\
&\leq \frac{1}{d}\frac{1}{\sqrt{T}}\left(\left\|\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^1\Lambda_2F^{b'} - \Lambda_2'\Lambda_2F^{b'}\right\| + \left\|\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^2\Lambda_2F^{a'} - \Lambda_2'\Lambda_2F^{a'}\right\|\right)\frac{1}{\sqrt{T}}\|\widehat{F}\|\|\widehat{V}_r^{-1}\| &\text{(B.44)}\\
&= \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}}\right). &\text{(B.45)}
\end{aligned}$$

Here, (B.44) is due to M.4. We focus on the second summand in (B.44) to derive (B.45). Then,

$$\begin{aligned}
\frac{1}{d}\frac{1}{\sqrt{T}}\left\|\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^2\Lambda_2F^{a'} - \Lambda_2'\Lambda_2F^{a'}\right\| &\leq \frac{1}{d}\left\|\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^2\Lambda_2 - \Lambda_2'\Lambda_2\right\|\frac{1}{\sqrt{T}}\|F^{a'}\|\\
&= \left(\mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right)\right). &\text{(B.46)}
\end{aligned}$$

The asymptotic of $\frac{1}{\sqrt{T}}\|F^{a'}\|$ is due to S.1 in the proof of Proposition B.1. The behavior of the estimated projection matrices for (B.46) follows from

$$\begin{aligned}
&\|\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^2\Lambda_2 - \Lambda_2'\Lambda_2\| \\
&= \|\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^2\Lambda_2 - \Lambda_2'\widehat{P}_I^{1'}\Lambda_2 + \Lambda_2'\widehat{P}_I^{1'}\Lambda_2 - \Lambda_2'\Lambda_2\| \\
&\leq \|\Lambda_2'\widehat{P}_I^{1'}\widehat{P}_I^2\Lambda_2 - \Lambda_2'\widehat{P}_I^{1'}\Lambda_2\| + \|\Lambda_2'\widehat{P}_I^{1'}\Lambda_2 - \Lambda_2'\Lambda_2\| \\
&\leq \|\Lambda_2'\widehat{P}_I^{1'}\|\|\widehat{P}_I^2\Lambda_2 - \Lambda_2\| + \|\Lambda_2\|\|\widehat{P}_I^{1'}\Lambda_2 - \Lambda_2\| \\
&= d\left(\mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right)\right),
\end{aligned} \tag{B.47}$$

as in (C.13) and since $\|\widehat{P}_I^k\| = \frac{1}{2}\|I_d + \widehat{P}_k\| \leq 1$. □

The following lemma was used in the preceding proof and is similar in nature but concerns the matrix H in (3.21).

Lemma B.2. *Under Assumptions A.1–A.2, CR.1–CR.3,*

$$H - H_0 = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}}\right) \quad \text{and} \quad \|H_0\| = \mathcal{O}(1), \text{ as } d, T \to \infty \tag{B.48}$$

with H_0 as in (B.43).

Proof: Recall from (3.21) that $H = (\Lambda_2'\Lambda_2/d)(F'\widehat{F}/T)\widehat{V}_r^{-1}$ and also that there is a matrix $\widetilde{\Pi}_2$ with $\|\Pi_2/d - \widetilde{\Pi}_2\| = \mathcal{O}(1/\sqrt{d})$ by Assumption CR.1. Then, with further explanations given below,

$$\begin{aligned}
&\|H - H_0\| \\
&= \left\|\left(\frac{\Lambda_2'\Lambda_2}{d}\right)\left(\frac{F'\widehat{F}}{T}\right)\widehat{V}_r^{-1} - Q^2\widetilde{\Pi}_2Q^{2'}\Sigma_F\widetilde{\Pi}_2^{-1}\right\| \\
&= \left\|Q^2\left(\frac{\Pi_2}{d}\right)Q^{2'}\left(\frac{F'\widehat{F}}{T}\right)d\widehat{\Pi}_r^{-1} - Q^2\widetilde{\Pi}_2Q^{2'}\Sigma_F\widetilde{\Pi}_2^{-1}\right\| \qquad \text{(B.49)} \\
&\leq \left\|Q^2\left(\frac{\Pi_2}{d} - \widetilde{\Pi}_2\right)Q^{2'}\left(\frac{F'\widehat{F}}{T}\right)d\widehat{\Pi}_r^{-1}\right\| + \left\|Q^2\widetilde{\Pi}_2Q^{2'}\left(\frac{F'\widehat{F}}{T}\right)d\widehat{\Pi}_r^{-1} - Q^2\widetilde{\Pi}_2Q^{2'}\Sigma_F\widetilde{\Pi}_2^{-1}\right\| \\
&\leq \left\|Q^2\left(\frac{\Pi_2}{d} - \widetilde{\Pi}_2\right)Q^{2'}\left(\frac{F'\widehat{F}}{T}\right)d\widehat{\Pi}_r^{-1}\right\| + \left\|Q^2\widetilde{\Pi}_2Q^{2'}\frac{1}{\sqrt{T}}F'\frac{1}{\sqrt{T}}(\widehat{F} - F)d\widehat{\Pi}_r^{-1}\right\| \\
&\qquad + \left\|Q^2\widetilde{\Pi}_2Q^{2'}\left(\frac{F'F}{T} - \Sigma_F\right)d\widehat{\Pi}_r^{-1}\right\| + \left\|Q\widetilde{\Pi}_2Q'\Sigma_F(d\widehat{\Pi}_r^{-1} - \widetilde{\Pi}_2^{-1})\right\| \\
&\leq \|Q\|^2\left\|\frac{\Pi}{d} - \widetilde{\Pi}_2\right\|\frac{1}{\sqrt{T}}\|F\|\frac{1}{\sqrt{T}}\|\widehat{F}\|\|d\widehat{\Pi}_r^{-1}\| + \|Q\|^2\|\widetilde{\Pi}_2\|\frac{1}{\sqrt{T}}\|F\|\frac{1}{\sqrt{T}}\|\widehat{F} - F\|\|d\widehat{\Pi}_r^{-1}\| \\
&\qquad + \|Q\|^2\|\widetilde{\Pi}_2\|\left\|\frac{F'F}{T} - \Sigma_F\right\|\|d\widehat{\Pi}_r^{-1}\| + \|Q\|^2\|\widetilde{\Pi}_2\|\|\Sigma_F\|\|d\widehat{\Pi}_r^{-1} - \widetilde{\Pi}_2^{-1}\| \qquad \text{(B.50)} \\
&= \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\delta_{dT}}\right).
\end{aligned}$$

For (B.49) recall that $\Lambda_2'\Lambda_2 = Q^2\Pi_2Q^{2'}$ in (B.43) and $\widehat{V}_r^{-1} = d\widehat{\Pi}_r^{-1}$. The asymptotics of the quantities in (B.50) are the following:

(i) $\left\|\frac{\Pi_2}{d} - \widetilde{\Pi}_2\right\| = \mathcal{O}\Big(\frac{1}{\sqrt{d}}\Big)$ by Assumption CR.1.

(ii) $\frac{1}{T}\|F\|^2 = \mathcal{O}_{\mathrm{p}}(1)$ by S.1 in the proof of Proposition B.1 and $\frac{1}{\sqrt{T}}\|\widehat{F} - F\| = \mathcal{O}_{\mathrm{p}}\Big(\frac{1}{\delta_{dT}}\Big)$ by Proposition C.1.

(iii) $\left\|\frac{F'F}{T} - \Sigma_F\right\| = \mathcal{O}_{\mathrm{p}}\Big(\frac{1}{\sqrt{T}}\Big)$; see p. 199 in Doz et al. (2011).

(iv) $\Big(\frac{1}{d}\widehat{\Pi}_r\Big)^{-1} = \mathcal{O}_{\mathrm{p}}(1)$ by Corollary C.1*(ii)*.

Finally, we have

$$\|H_0\| = \|Q^2\widetilde{\Pi}_2 Q^{2'}\Sigma_F\widetilde{\Pi}_2^{-1}\| = \mathcal{O}(1).$$

□

C Results for consistency of PCA estimators under null hypothesis

We present some results and their proofs to establish consistent estimation of PCA estimators under the null hypothesis. Those results were used in Appendix B.

Lemma C.1. *Set $\widehat{\Sigma}_Y = \frac{1}{T}Y'Y$. Then, under Assumptions A.1–A.2, CR.1–CR.3,*

$$\frac{1}{d}\|\widehat{\Sigma}_Y - \Lambda_2\Lambda_2'\| = \mathcal{O}_{\mathrm{p}}\Big(\frac{1}{d}\Big) + \mathcal{O}_{\mathrm{p}}\Big(\frac{1}{\sqrt{T}}\Big).$$

Proof: We follow the proof of Lemma 2 in Doz et al. (2011). Consider

$$\frac{1}{d}\|\widehat{\Sigma}_Y - \Lambda_2\Lambda_2'\| \leq \frac{1}{d}\|\widehat{\Sigma}_Y - \Sigma\| + \frac{1}{d}\|\Sigma - \Lambda_2\Lambda_2'\| \tag{C.1}$$

with

$$\begin{aligned}\Sigma &= \Lambda_2\Lambda_2' + P_{0,2}\Sigma_\varepsilon P_{0,2}' \\ &= P_{0,2}\Lambda_2\Lambda_2' P_{0,2}' + P_{0,2}\Sigma_\varepsilon P_{0,2}' \\ &= \frac{1}{2}\Big(P_{0,1}\Lambda_2\Lambda_2' P_{0,1}' + P_{0,1}\Sigma_\varepsilon P_{0,1}'\Big) + \frac{1}{2}\Big(P_{0,2}\Lambda_2\Lambda_2' P_{0,2}' + P_{0,2}\Sigma_\varepsilon P_{0,2}'\Big),\end{aligned} \tag{C.2}$$

where $P_{0,1} = P_{0,2}$ under the null hypothesis. We consider the two summands in (C.1) separately. For the second summand, we get

$$\frac{1}{d}\|\Sigma - \Lambda_2\Lambda_2'\| = \frac{1}{d}\|P_{0,2}\Sigma_\varepsilon P_{0,2}'\| \leq \frac{1}{d}\|P_{0,2}\|^2\|\Sigma_\varepsilon\| = \frac{1}{d}\|\Sigma_\varepsilon\| = O\Big(\frac{1}{d}\Big)$$

due to Assumption CR.3 and our observation (B.18). For the first summand in (C.1), we separate the series according to the transformation (3.7) as

$$\|\widehat{\Sigma}_Y - \Sigma\| = \left\|\frac{1}{2}\widehat{P}_I^1\frac{1}{T/2}\sum_{t=1}^{T/2}X_t^2X_t^{2'}\widehat{P}_I^{1'} + \frac{1}{2}\widehat{P}_I^2\frac{1}{T/2}\sum_{t=T/2+1}^{T}X_t^2X_t^{2'}\widehat{P}_I^{2'} - \Sigma\right\|$$

$$\leq \frac{1}{2}\left\|\widehat{P}_I^1 \frac{1}{T/2}\sum_{t=1}^{T/2} X_t^2 X_t^{2'} \widehat{P}_I^{1'} - \left(P_{0,1}\Lambda_2\Lambda_2' P_{0,1}' + P_{0,1}\Sigma_\varepsilon P_{0,1}'\right)\right\|$$

$$+ \frac{1}{2}\left\|\widehat{P}_I^2 \frac{1}{T/2}\sum_{t=T/2+1}^{T} X_t^2 X_t^{2'} \widehat{P}_I^{2'} - \left(P_{0,2}\Lambda_2\Lambda_2' P_{0,2}' + P_{0,2}\Sigma_\varepsilon P_{0,2}'\right)\right\|, \tag{C.3}$$

where we used (C.2). We consider the two summands in (C.3) separately. For the first summand in (C.3), with explanations given below,

$$\left\|\widehat{P}_I^1 \frac{1}{T/2}\sum_{t=1}^{T/2} X_t^2 X_t^{2'} \widehat{P}_I^{1'} - \left(P_{0,1}\Lambda_2\Lambda_2' P_{0,1}' + P_{0,1}\Sigma_\varepsilon P_{0,1}'\right)\right\|$$

$$\leq \left\|\widehat{P}_I^1\left(\widehat{\Sigma}_2 - \Lambda_2\Lambda_2'\right)\widehat{P}_I^{1'}\right\| + \left\|\widehat{P}_I^1\Lambda_2\Lambda_2'\widehat{P}_I^{1'} - \left(P_{0,1}\Lambda_2\Lambda_2' P_{0,1}' + P_{0,1}\Sigma_\varepsilon P_{0,1}'\right)\right\|$$

$$\leq \left\|\widehat{P}_I^1\left(\widehat{\Sigma}_2 - \Lambda_2\Lambda_2'\right)\widehat{P}_I^{1'}\right\| + \left\|\widehat{P}_I^1\Lambda_2\Lambda_2'\widehat{P}_I^{1'} - P_{0,1}\Lambda_2\Lambda_2' P_{0,1}'\right\| + \|P_{0,1}\Sigma_\varepsilon P_{0,1}'\|$$

$$\leq \|\widehat{P}_I^1\|^2\|\widehat{\Sigma}_2 - \Lambda_2\Lambda_2'\| + \left\|\widehat{P}_I^1\Lambda_2\Lambda_2'\widehat{P}_I^{1'} - P_{0,1}\Lambda_2\Lambda_2' P_{0,1}'\right\| + \|P_{0,1}\|^2\|\Sigma_\varepsilon\| \tag{C.4}$$

$$\leq \|\widehat{\Sigma}_2 - \Lambda_2\Lambda_2'\| + \left\|\widehat{P}_I^1\Lambda_2\Lambda_2'\widehat{P}_I^{1'} - P_{0,1}\Lambda_2\Lambda_2' P_{0,1}'\right\| + \|\Sigma_\varepsilon\| \tag{C.5}$$

$$= d\mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + d\mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right) + \mathcal{O}(1), \tag{C.6}$$

where (C.4) uses the submultiplicativity of the spectral norm, (C.5) is due to (B.18) and similarly for their estimated counterparts $\|\widehat{P}_k\| = 1$, $k = 1, 2$, such that $\|\widehat{P}_I^k\| = \frac{1}{2}\|I_d + \widehat{P}_k\| \leq 1$. Finally, (C.6) is due to DGR.1 in Appendix A, Lemma C.2 below and $\|\Sigma_\varepsilon\| = \mathcal{O}(1)$ by Assumption CR.3. The second summand in (C.3) can be handled by analogous considerations. □

The next result was used in the preceding proof.

Lemma C.2. *Recall $\widehat{P}_I^1 = \frac{1}{2}(\widehat{P}_1 + I_d)$ from (3.7) and the definition of $\widehat{P}_1$ in (3.4). Then, under Assumptions A.1–A.2, CR.1–CR.3,*

$$\frac{1}{d}\left\|\widehat{P}_I^1\Lambda_2\Lambda_2'\widehat{P}_I^{1'} - P_{0,1}\Lambda_2\Lambda_2' P_{0,1}'\right\| = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right).$$

Proof: Since $\widehat{P}_I^1 = \frac{1}{2}(\widehat{P}_1 + I_d)$ and under the null hypothesis, $P_{0,1} = P_{0,2}$ and hence $P_{0,1}\Lambda_2 = P_{0,2}\Lambda_2 = \Lambda_2$, we can infer

$$\left\|\widehat{P}_I^1\Lambda_2\Lambda_2'\widehat{P}_I^{1'} - P_{0,1}\Lambda_2\Lambda_2' P_{0,1}'\right\|$$

$$\leq \frac{1}{4}\left(\left\|\widehat{P}_1\Lambda_2\Lambda_2'\widehat{P}_1' - P_{0,1}\Lambda_2\Lambda_2' P_{0,1}'\right\| + \left\|\widehat{P}_1\Lambda_2\Lambda_2' - P_{0,1}\Lambda_2\Lambda_2'\right\| + \left\|\Lambda_2\Lambda_2'\widehat{P}_1' - \Lambda_2\Lambda_2' P_{0,1}'\right\|\right). \tag{C.7}$$

We consider the three terms in (C.7) separately. Recall from (B.5) that $\widehat{P}_k = \bar{Q}_r^k\bar{Q}_r^{k'}$, where $\bar{Q}_r^k$ are the eigenvectors corresponding to the first r largest eigenvalues of the matrix $X^{k'}X^k$. Recall from (A.4) that $Q^k = \Lambda_k R^k \Pi_k^{-1/2}$, $k = 1, 2$. Then, with explanations given below,

$$\left\|\widehat{P}_1\Lambda_2\Lambda_2'\widehat{P}_1' - P_{0,1}\Lambda_2\Lambda_2' P_{0,1}'\right\|$$

$$= \left\|\bar{Q}_r^1\bar{Q}_r^{1'}\Lambda_2\Lambda_2'\bar{Q}_r^1\bar{Q}_r^{1'} - \Lambda_2\Lambda_2'\right\|$$

$$
\begin{aligned}
&\leq \left\|\bar{Q}_r^1\bar{Q}_r^{1'}\Lambda_2\Lambda_2'\bar{Q}_r^1\bar{Q}_r^{1'} - \bar{Q}_r^1 Q^{1'}\Lambda_2\Lambda_2' Q^1\bar{Q}_r^{1'}\right\| + \left\|\bar{Q}_r^1 Q^{1'}\Lambda_2\Lambda_2' Q^1\bar{Q}_r^{1'} - \Lambda_2\Lambda_2'\right\| \\
&\leq \|\bar{Q}_r^1\|^2 \left\|\bar{Q}_r^{1'}\Lambda_2\Lambda_2'\bar{Q}_r^1 - Q^{1'}\Lambda_2\Lambda_2' Q^1\right\| + \left\|\bar{Q}_r^1 Q_r^{1'}\Lambda_2\Lambda_2' Q^1\bar{Q}_r^{1'} - \Lambda_2\Lambda_2'\right\| \\
&\leq \|(\bar{Q}_r^1 - Q^1)'\Lambda_2\Lambda_2'(\bar{Q}_r^1 - Q^1)\| + \|(\bar{Q}_r^1 - Q^1)'\Lambda_2\Lambda_2' Q^1\| \\
&\qquad + \|Q^{1'}\Lambda_2\Lambda_2'(\bar{Q}_r^1 - Q^1)\| + \|(\bar{Q}_r^1 - Q^1)Q^{1'}\Lambda_2\Lambda_2' Q^1(\bar{Q}_r^1 - Q^1)'\| \\
&\qquad\qquad + \|(\bar{Q}_r^1 - Q^1)Q_r^{1'}\Lambda_2\Lambda_2' Q^1 Q^{1'}\| + \|Q^1 Q^{1'}\Lambda_2\Lambda_2' Q_1(\bar{Q}^1 - Q^1)'\| \qquad (C.8) \\
&\leq \|\bar{Q}_r^1 - Q^1\|^2\|\Lambda_2\Lambda_2'\| + 2\|(\bar{Q}_r^1 - Q^1)'\Lambda_2\Lambda_2' Q^1\| \\
&\qquad + \|\bar{Q}_r^1 - Q^1\|^2\|Q^1\|^2\|\Lambda_2\Lambda_2'\| + 2\|(\bar{Q}_r^1 - Q^1)Q^{1'}\Lambda_2\Lambda_2' Q^1 Q^{1'}\| \qquad (C.9) \\
&= d\left(\mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right)\right).
\end{aligned}
$$

Note that (C.8) follows since $P_{0,1} = Q^1 Q^{1'}$ and therefore $Q^1 Q^{1'}\Lambda_2\Lambda_2' Q^1 Q^{1'} = \Lambda_2\Lambda_2'$. The asymptotics of the first and third summand in (C.9) follow by Assumption CR.1 and DGR.3. For the remaining summands in (C.9), let $Q_\perp^1$ be a $d \times (d-r)$ matrix whose columns consist of vectors forming an orthonormal basis of the orthogonal space of Q^1 such that $I_d = Q^1 Q^{1'} + Q_\perp^1 Q_\perp^{1'}$. Then, for the second summand in (C.9),

$$
\begin{aligned}
\|(\bar{Q}_r^1 - Q^1)'\Lambda_2\Lambda_2' Q^1\| &\leq \|(\bar{Q}_r^1 - Q^1)'\Lambda_2\Lambda_2'\| \qquad (C.10) \\
&= \|(\bar{Q}_r^1 - Q^1)'(Q^1 Q^{1'} + Q_\perp^1 Q_\perp^{1'})\Lambda_2\Lambda_2'\| \\
&= \|\bar{Q}_r^{1'} Q^1 Q^{1'}\Lambda_2\Lambda_2' - Q^{1'}\Lambda_2\Lambda_2' - \bar{Q}_r^{1'} Q_\perp^1 Q_\perp^{1'}\Lambda_2\Lambda_2'\| \\
&= \|(\bar{Q}_r^{1'} Q^1 - I_r)Q^{1'}\Lambda_2\Lambda_2' - \bar{Q}_r^{1'} Q_\perp^1 Q_\perp^{1'}\Lambda_2\Lambda_2'\| \\
&\leq \|\bar{Q}_r^{1'} Q^1 - I_r\|\|Q^{1'}\Lambda_2\Lambda_2'\| + \|\bar{Q}_r^{1'} Q_\perp^1\|\|Q_\perp^{1'}\Lambda_2\Lambda_2'\| \\
&\leq \|\bar{Q}_r^{1'} Q^1 - I_r\|\|\Lambda_2\Lambda_2'\| + \|\bar{Q}_r^{1'} Q_\perp^1\|\|\Lambda_2\Lambda_2'\| \\
&= d\left(\mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right)\right), \qquad (C.11)
\end{aligned}
$$

where (C.11) is due to Assumption CR.1, DGR.2 and Corollary C.3 below. The last summand in (C.9) can be treated similarly.

For the second summand in (C.7), we get

$$
\begin{aligned}
\left\|\widehat{P}_1\Lambda_2\Lambda_2' - \Lambda_2\Lambda_2'\right\| &= \|\bar{Q}_r^1\bar{Q}_r^{1'}\Lambda_2\Lambda_2' - \Lambda_2\Lambda_2'\| \\
&\leq \|\bar{Q}_r^1(\bar{Q}_r^1 - Q^1)'\Lambda_2\Lambda_2'\| + \|(\bar{Q}_r^1 - Q^1)Q^{1'}\Lambda_2\Lambda_2'\| \\
&\leq \|(\bar{Q}_r^1 - Q^1)'\Lambda_2\Lambda_2'\| + \|(\bar{Q}_r^1 - Q^1)Q^{1'}\Lambda_2\Lambda_2'\| \qquad (C.12) \\
&= d\left(\mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right)\right), \qquad (C.13)
\end{aligned}
$$

where the asymptotic behavior of the first summand in (C.12) follows by that of (C.10) above and the second summand in (C.12) can be treated similarly. The third summand in (C.7) is, in fact, the same as the second since $\|A\| = \|A'\|$. □

The next result is a consequence of Lemma C.1. The proof arguments follow those for Lemma 2 in Doz et al. (2011).

Corollary C.1. *Set $\widehat{\Pi}_r = d\widehat{V}_r$ as in (B.4) and recall that $\widehat{V}_r$ is a diagonal matrix consisting of the r largest eigenvalues of $\frac{1}{dT}YY'$. Recall also the diagonal Π_2 from Appendix A. Under Assumptions A.1–A.2, CR.1–CR.4,* *(i)* $\frac{1}{d}\|\widehat{\Pi}_r - \Pi_2\| = \mathcal{O}_\mathrm{p}\big(\frac{1}{d}\big) + \mathcal{O}_\mathrm{p}\Big(\frac{1}{\sqrt{T}}\Big)$,

(ii) $d\|\widehat{\Pi}_r^{-1} - \Pi_2^{-1}\| = \mathcal{O}_\mathrm{p}\big(\frac{1}{d}\big) + \mathcal{O}_\mathrm{p}\Big(\frac{1}{\sqrt{T}}\Big)$,

(iii) $\Pi_2\widehat{\Pi}_r^{-1} - I_r = \mathcal{O}_\mathrm{p}\big(\frac{1}{d}\big) + \mathcal{O}_\mathrm{p}\Big(\frac{1}{\sqrt{T}}\Big)$.

Proof: With $\widehat{\Sigma}_Y = \frac{1}{T}Y'Y$, we have

$\Pi_2 = \mathrm{diag}(\lambda_1(\Lambda_2'\Lambda_2), \dots, \lambda_r(\Lambda_2'\Lambda_2)) = \mathrm{diag}(\lambda_1(\Lambda_2\Lambda_2'), \dots, \lambda_r(\Lambda_2\Lambda_2'))$, $\widehat{\Pi}_r = \mathrm{diag}(\lambda_1(\widehat{\Sigma}_Y), \dots, \lambda_r(\widehat{\Sigma}_Y))$.

(i) By Weyl's Theorem (Theorem 4.3.1 in Horn and Johnson (2012)) and Lemma C.1, for any $j = 1, \dots, r$,

$$|\lambda_j(\widehat{\Sigma}_Y) - \lambda_j(\Lambda_2\Lambda_2')| \le \|\widehat{\Sigma}_Y - \Lambda_2\Lambda_2'\| = d\left(\mathcal{O}\left(\frac{1}{d}\right) + \mathcal{O}_\mathrm{p}\left(\frac{1}{\sqrt{T}}\right)\right),$$

which gives the desired result, since $\|A\| = \max_{j=1,\dots,d}|\lambda_j(A)|$ for diagonal matrix A.

(ii) Write

$$d(\widehat{\Pi}_r^{-1} - \Pi_2^{-1}) = \left(\frac{1}{d}\widehat{\Pi}_r\right)^{-1} \cdot \frac{1}{d}(\Pi_2 - \widehat{\Pi}_r) \cdot \left(\frac{1}{d}\Pi_2\right)^{-1},$$

so that

$$d\|\widehat{\Pi}_r^{-1} - \Pi_2^{-1}\| \le \left\|\left(\frac{1}{d}\widehat{\Pi}_r\right)^{-1}\right\|\left\|\left(\frac{1}{d}\Pi_2\right)^{-1}\right\|\frac{1}{d}\|\Pi_2 - \widehat{\Pi}_r\|.$$

The desired result then follows by (i) and since Assumption CR.1 implies

$$\left(\frac{1}{d}\Pi_2\right)^{-1} = \mathcal{O}(1) \quad \text{and hence} \quad \left(\frac{1}{d}\widehat{\Pi}_r\right)^{-1} = \mathcal{O}_\mathrm{p}(1).$$

(iii) Write

$$\Pi_2\widehat{\Pi}_r^{-1} = \Pi_2(\widehat{\Pi}_r^{-1} - \Pi_2^{-1} + \Pi_2^{-1}) = I_r + \frac{\Pi_2}{d}d\left(\widehat{\Pi}_r^{-1} - \Pi_2^{-1}\right),$$

so that

$$\Pi_2\widehat{\Pi}_r^{-1} - I_r = \frac{\Pi_2}{d}d\left(\widehat{\Pi}_r^{-1} - \Pi_2^{-1}\right)$$

and use *(ii)*. □

The next result and its corollaries are used in this appendix. See also Lemmas 3 and 4 in Doz et al. (2011), and their proofs.

Lemma C.3. *Let $\widehat{A} = (\widehat{a}_{ij})_{i,j=1,\dots,r} = \bar{Q}_r'Q$ with $Q := Q^2$ defined in Appendix A and $\bar{Q}_r$ defined in (B.2). Under Assumptions A.1–A.2, CR.1–CR.3, $\widehat{a}_{ij} = \mathcal{O}_\mathrm{p}\big(\frac{1}{d}\big) + \mathcal{O}_\mathrm{p}\Big(\frac{1}{\sqrt{T}}\Big)$, $i \neq j$, $\widehat{a}_{ii}^2 = 1 + \mathcal{O}_\mathrm{p}\big(\frac{1}{d}\big) + \mathcal{O}_\mathrm{p}\Big(\frac{1}{\sqrt{T}}\Big)$, $i = 1, \dots, r$.*

Proof: For the off diagonal terms $\widehat{a}_{ij}$, write

$$\begin{aligned}
\widehat{A} = \bar{Q}_r' Q &= \widehat{\Pi}_r^{-1}\bar{Q}_r'\widehat{\Sigma}_Y Q \\
&= \widehat{\Pi}_r^{-1}\bar{Q}_r'(\widehat{\Sigma}_Y - \Lambda_2\Lambda_2')Q + \widehat{\Pi}_r^{-1}\bar{Q}_r'\Lambda_2\Lambda_2' Q \\
&= \left(\frac{\widehat{\Pi}_r}{d}\right)^{-1}\bar{Q}_r'\left(\frac{\widehat{\Sigma}_Y - \Lambda_2\Lambda_2'}{d}\right)Q + \left(\frac{\widehat{\Pi}_r}{d}\right)^{-1}\bar{Q}_r' Q\left(\frac{\Pi_2}{d}\right) \\
&= \mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right) + \left(\frac{\Pi_2}{d}\right)^{-1}\widehat{A}\left(\frac{\Pi_2}{d}\right),
\end{aligned}$$

where we used Lemma C.1 and Corollary C.1. This means that $\widehat{a}_{ij} = \frac{\pi_{jj}}{\pi_{ii}}\widehat{a}_{ij} + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right)$ with $\frac{\pi_{jj}}{\pi_{ii}} \neq 1$ by Assumption CR.2 which yields the desired result, when $i \neq j$.

For the diagonal terms $\widehat{a}_{ii}$, write

$$\begin{aligned}
\frac{\Pi_2}{d} &= \frac{\widehat{\Pi}_r}{d} + \mathcal{O}\left(\frac{1}{d}\right) + \mathcal{O}\left(\frac{1}{\sqrt{T}}\right) \\
&= \bar{Q}_r'\frac{\widehat{\Sigma}_Y}{d}\bar{Q}_r + \mathcal{O}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right) \\
&= \bar{Q}_r'\frac{\Lambda_2\Lambda_2'}{d}\bar{Q}_r + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right) \\
&= \bar{Q}_r' Q\frac{\Pi_2}{d}Q'\bar{Q}_r' + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right) \\
&= \widehat{A}\frac{\Pi_2}{d}\widehat{A}' + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}$$

This means that, for $i = 1, \dots, r$, $\frac{\pi_{ii}}{d} = \sum_{k=1}^{r}\frac{\pi_{kk}}{d}\widehat{a}_{ik}^2 + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right)$ which yields the desired result by using the result for the off-diagonal elements. □

Corollary C.2. *Under Assumptions A.1–A.2, CR.1–CR.3, one can take* $\bar{Q}_r$, $Q^2 =: Q$ *such that*

$$\bar{Q}_r' Q = I_r + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right), \quad \|\bar{Q}_r - Q\|^2 = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right).$$

Proof: We consider the two asymptotics separately. For the first result, since the columns of Q are the orthonormal eigenvectors of $\Lambda_2\Lambda_2'$ associated with distinct eigenvalues by assumption, these columns are defined uniquely up to a sign change. Then one can choose them in such a way that the diagonal terms of $\widehat{A} = \bar{Q}_r' Q$ are positive. The desired result follows from Lemma C.3.

For the second result, note that for any $x \in \mathbb{R}^r$ with $\|x\| = 1$,

$$x'(\bar{Q}_r - Q)'(\bar{Q}_r - Q)x = x'(2I_r - \bar{Q}_r' Q - Q'\bar{Q}_r)x = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right),$$

since $\bar{Q}_r'\bar{Q}_r = Q'Q = I_r$. □

Corollary C.3. *Under Assumptions A.1–A.2, CR.1–CR.3,*

$$\bar{Q}_r' Q_\perp = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right).$$

Proof: Note that $\bar{Q}_r = \widehat{\Sigma}_Y \bar{Q}_r \widehat{\Pi}_r^{-1}$. Hence,

$$\begin{aligned}\bar{Q}_r' Q_\perp &= \widehat{\Pi}_r^{-1} \bar{Q}_r' \widehat{\Sigma}_Y Q_\perp \\ &= \left(\frac{\widehat{\Pi}_r}{d}\right)^{-1} \bar{Q}_r' \frac{1}{d}(\widehat{\Sigma}_Y - \Lambda_2\Lambda_2')Q_\perp + \widehat{\Pi}_r^{-1}\bar{Q}_r'\Lambda_2\Lambda_2' Q_\perp \\ &= \mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right)\end{aligned}$$

due to Lemma C.1, Corollary C.1*(ii)* and since $\Lambda_2' Q_\perp = 0$ because of $Q_\perp = Q_\perp^2 = (\Lambda_2 R^2 \Pi_2^{-\frac{1}{2}})_\perp$. □

The next result is a consequence of the preceding results and was used in the proofs of Propositions B.1 and B.2.

Proposition C.1. *Under Assumptions A.1–A.2, CR.1–CR.3,*

$$\frac{1}{T}\|\widehat{F} - F\|_F^2 = \frac{1}{T}\sum_{t=1}^{T}\|\widehat{F}_t - F_t\|_F^2 = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{T}\right), \text{ as } d, T \to \infty.$$

Proof: Using the representations (B.1), (B.2), their relationship (B.3), and the definition (3.7),

$$\widehat{F}_t = \widehat{V}_r^{-\frac{1}{2}}\bar{F}_t = \widehat{V}_r^{-\frac{1}{2}}\frac{1}{d}\bar{\Lambda}' Y_t = \widehat{V}_r^{-\frac{1}{2}}\frac{1}{\sqrt{d}}\bar{Q}_r' Y_t = \begin{cases}\widehat{\Pi}_r^{-1/2}\bar{Q}_r'\widehat{P}_I^1 X_t^2, & \text{for } t = 1,\dots,T/2,\\ \widehat{\Pi}_r^{-1/2}\bar{Q}_r'\widehat{P}_I^2 X_t^2, & \text{for } t = T/2+1,\dots,T.\end{cases}$$

We consider only the case $t = 1,\dots,T/2$. Then,

$$\begin{aligned}\widehat{F}_t - F_t &= \widehat{\Pi}_r^{-1/2}\bar{Q}_r'\widehat{P}_I^1 X_t^2 - F_t \\ &= \widehat{\Pi}_r^{-1/2}\bar{Q}_r'\widehat{P}_I^1(\Lambda_2 F_t + \varepsilon_t) - F_t \\ &= (\widehat{\Pi}_r^{-1/2}\bar{Q}_r'\widehat{P}_I^1\Lambda_2 - I_r)F_t + \widehat{\Pi}_r^{-1/2}\bar{Q}_r'\widehat{P}_I^1\varepsilon_t \\ &= \left(\widehat{\Pi}_r^{-1/2}\bar{Q}_r'\frac{1}{2}(\widehat{P}_1\Lambda_2 - \Lambda_2) + \widehat{\Pi}_r^{-1/2}\bar{Q}_r'\Lambda_2 - I_r\right)F_t + \widehat{\Pi}_r^{-1/2}\bar{Q}_r'\widehat{P}_I^1\varepsilon_t \\ &= \widehat{\Pi}_r^{-1/2}\bar{Q}_r'\frac{1}{2}(\widehat{P}_1\Lambda_2 - \Lambda_2)F_t + (\widehat{\Pi}_r^{-1/2}\bar{Q}_r'\Lambda_2 - I_r)F_t + \widehat{\Pi}_r^{-1/2}\bar{Q}_r'\widehat{P}_I^1\varepsilon_t \\ &= \widehat{\Pi}_r^{-1/2}\bar{Q}_r'\frac{1}{2}(\widehat{P}_1\Lambda_2 - \Lambda_2)F_t + \widehat{\Pi}_r^{-1/2}(\bar{Q}_r' Q - \widehat{\Pi}_r^{1/2}\Pi_2^{-1/2})\Pi_2^{1/2}F_t + \widehat{\Pi}_r^{-1/2}\bar{Q}_r'\widehat{P}_I^1\varepsilon_t. \qquad \text{(C.14)}\end{aligned}$$

We consider the three summands in (C.14) separately. For the first one, note that

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T/2}\|\widehat{\Pi}_r^{-1/2}\bar{Q}_r'\frac{1}{2}(\widehat{P}_1\Lambda_2 - \Lambda_2)F_t\|_F^2 &\leq \|\widehat{\Pi}_r^{-1/2}\bar{Q}_r'\frac{1}{2}(\widehat{P}_1\Lambda_2 - \Lambda_2)\|^2\frac{1}{T}\|F\|_F^2 \\ &\leq d\|\widehat{\Pi}_r^{-1/2}\|^2\|\bar{Q}_r'\|^2\frac{1}{d}\|\frac{1}{2}(\widehat{P}_1\Lambda_2 - \Lambda_2)\|^2\frac{1}{T}\|F\|_F^2 \\ &= \mathcal{O}_{\mathrm{p}}\left(\frac{1}{d^2}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{T}\right),\end{aligned}$$

since $\frac{1}{T}\|F\|_F^2 \leq \frac{r}{T}\|F\|^2 = \mathcal{O}_{\mathrm{p}}(1)$ by S.1 in the proof of Proposition B.1 and $\widehat{\Pi}_r^{-1/2} = \frac{1}{\sqrt{d}}\left(\frac{1}{d}\widehat{\Pi}_r\right)^{-1/2} = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{d}}\right)$ by Corollary C.1. Furthermore, since $P_{0,1} = Q^1 Q^{1'}$ and therefore $Q^1 Q^{1'}\Lambda_2 = \Lambda_2$,

$$\|\widehat{P}_1\Lambda_2 - \Lambda_2\| = \|\bar{Q}_r^1\bar{Q}_r^{1'}\Lambda_2 - \Lambda_2\|$$

$$
\begin{aligned}
&\leq \|\bar{Q}_r^1(\bar{Q}_r^1 - Q^1)'\Lambda_2\| + \|(\bar{Q}_r^1 - Q^1)Q^{1'}\Lambda_2\| \\
&= \sqrt{d}\Big(\mathcal{O}_{\mathrm{p}}\Big(\frac{1}{d}\Big) + \mathcal{O}_{\mathrm{p}}\Big(\frac{1}{\sqrt{T}}\Big)\Big),
\end{aligned} \tag{C.15}
$$

where (C.15) follows by the same arguments as (C.13).

For the second and third summands in (C.14), we do the same calculations as in the proof of Proposition 2 in Doz et al. (2011). That is, for the second summand,

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T/2}\|\widehat{\Pi}_r^{-1/2}(\bar{Q}_r'Q - \widehat{\Pi}_r^{1/2}\Pi_2^{-1/2})\Pi_2^{1/2}F_t\|_F^2 &\leq \|\widehat{\Pi}_r^{-1/2}(\bar{Q}_r'Q - \widehat{\Pi}_r^{1/2}\Pi_2^{-1/2})\Pi_2^{1/2}\|^2\frac{1}{T}\|F\|_F^2 \\
&= \mathcal{O}_{\mathrm{p}}\Big(\frac{1}{d^2}\Big) + \mathcal{O}_{\mathrm{p}}\Big(\frac{1}{T}\Big)
\end{aligned}
$$

by Corollaries C.1, C.2 and $\frac{1}{T}\|F\|_F^2 \leq \frac{r}{T}\|F\|^2 = \mathcal{O}_{\mathrm{p}}(1)$ by S.1 in the proof of Proposition B.1.

For the third summand, let $Q_\perp$ be a $d\times(d-r)$ matrix whose columns consist of vectors forming an orthonormal basis of the orthogonal space of Q. Then $I_d = QQ' + Q_\perp Q_\perp'$ and we can write

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T/2}\|\widehat{\Pi}_r^{-1/2}\bar{Q}_r'\widehat{P}_I^1\varepsilon_t\|_F^2 &= \frac{2}{T}\sum_{t=1}^{T/2}\|\widehat{\Pi}_r^{-1/2}\bar{Q}_r'QQ'\widehat{P}_I^1\varepsilon_t\|_F^2 + \frac{2}{T}\sum_{t=1}^{T/2}\|\widehat{\Pi}_r^{-1/2}\bar{Q}_r'Q_\perp Q_\perp'\widehat{P}_I^1\varepsilon_t\|_F^2 && \text{(C.16)} \\
&= \mathcal{O}_{\mathrm{p}}\Big(\frac{1}{d^2}\Big) + \mathcal{O}_{\mathrm{p}}\Big(\frac{1}{T}\Big). && \text{(C.17)}
\end{aligned}
$$

To show (C.17), note for the first summand in (C.16),

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T/2}\|Q'\widehat{P}_I^1\varepsilon_t\|_F^2 &\leq \frac{2}{T}\sum_{t=1}^{T/2}\|Q'(\widehat{P}_I^1 - P_I^1)\varepsilon_t\|_F^2 + \frac{2}{T}\sum_{t=1}^{T/2}\|Q'P_I^1\varepsilon_t\|_F^2 \\
&= \frac{1}{2T}\sum_{t=1}^{T/2}\|Q'(\bar{Q}^1\bar{Q}^{1'} - Q^1Q^{1'})\varepsilon_t\|_F^2 + \frac{2}{T}\sum_{t=1}^{T/2}\|Q'P_I^1\varepsilon_t\|_F^2 && \text{(C.18)} \\
&\leq \|\bar{Q}^1\bar{Q}^{1'} - Q^1Q^{1'}\|^2\frac{1}{T}\sum_{t=1}^{T/2}\|\varepsilon_t\|_F^2 + \frac{2}{T}\sum_{t=1}^{T/2}\|Q'P_I^1\varepsilon_t\|_F^2 \\
&\leq \|\bar{Q}^1\bar{Q}^{1'} - Q^1Q^{1'}\|^2\mathcal{O}_{\mathrm{p}}(d) + \mathcal{O}_{\mathrm{p}}(1) && \text{(C.19)} \\
&\leq \Big(\|(\bar{Q}^1 - Q^{1'})\bar{Q}^{1'}\|^2 + \|Q^{1'}(\bar{Q}^{1'} - Q^{1'})\|^2\Big)\mathcal{O}_{\mathrm{p}}(d) + \mathcal{O}_{\mathrm{p}}(1) \\
&= \Big(\mathcal{O}_{\mathrm{p}}\Big(\frac{1}{d}\Big) + \mathcal{O}_{\mathrm{p}}\Big(\frac{1}{\sqrt{T}}\Big)\Big)\mathcal{O}_{\mathrm{p}}(d) + \mathcal{O}_{\mathrm{p}}(1) = \mathcal{O}_{\mathrm{p}}(1), && \text{(C.20)}
\end{aligned}
$$

where (C.18) uses the definitions of $\widehat{P}_I^1$ and P_I^1, (C.19) is explained in more detail below and (C.20) follows by DGR.3. For (C.19), we get

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T/2}\mathrm{E}\,\|Q'P_I^1\varepsilon_t\|_F^2 &= \frac{1}{T}\sum_{t=1}^{T/2}\mathrm{E}\,\varepsilon_t'P_I^{1'}QQ'P_I^1\varepsilon_t = \frac{1}{T}\sum_{t=1}^{T/2}\mathrm{E}\,\mathrm{tr}(Q'P_I^1\varepsilon_t\varepsilon_t'P_I^{1'}Q) \\
&= \frac{1}{T}\sum_{t=1}^{T/2}\mathrm{tr}(Q'P_I^1\Sigma_\varepsilon P_I^{1'}Q)
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{1}{T}\sum_{t=1}^{T/2} r\|Q' P_I^1 \Sigma_\varepsilon P_I^{1'} Q\| \leq \frac{r}{2}\|P_I^1 \Sigma_\varepsilon P_I^{1'}\| \\
&\leq \frac{r}{2}\|P_I^1\|^2\|\Sigma_\varepsilon\| \leq \frac{r}{2}\|\Sigma_\varepsilon\| = \mathcal{O}(1), \qquad \text{(C.21)}
\end{aligned}
$$

due to (B.18). The asymptotics in (C.21) follow by Assumption CR.3. Then, it suffices to show that

$$
\|\widehat{\Pi}_r^{-1/2}\bar{Q}_r' Q\|^2 = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{d^2}\right),
$$

which holds due to Corollary C.2 and $\widehat{\Pi}_r^{-1/2} = \frac{1}{\sqrt{d}}\left(\frac{1}{d}\widehat{\Pi}_r\right)^{-1/2} = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{d}}\right)$ by Corollary C.1. A similar argument applies to show that $\frac{1}{T}\sum_{t=1}^{T/2}\|\varepsilon_t\|_F^2 = \mathcal{O}_{\mathrm{p}}(d)$ as also stated in (C.19).

For the second summand in (C.16),

$$
\frac{1}{T}\sum_{t=1}^{T/2}\|Q_\perp' \widehat{P}_I^1 \varepsilon_t\|_F^2 \leq \frac{1}{T}\sum_{t=1}^{T/2}\|Q_\perp'(\widehat{P}_I^1 - P_I^1)\varepsilon_t\|_F^2 + \frac{1}{T}\sum_{t=1}^{T/2}\|Q_\perp' P_I^1 \varepsilon_t\|_F^2 = \mathcal{O}(d),
$$

where the first summand can be treated similarly to (C.20) and for the second summand, we get

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T/2}\mathrm{E}\,\|Q_\perp' P_I^1 \varepsilon_t\|_F^2 &= \frac{1}{T}\sum_{t=1}^{T/2}\mathrm{E}\,\varepsilon_t' P_I^{1'} Q_\perp Q_\perp' P_I^1 \varepsilon_t = \frac{1}{T}\sum_{t=1}^{T/2}\mathrm{E}\,\mathrm{tr}(Q_\perp' P_I^1 \varepsilon_t \varepsilon_t' P_I^{1'} Q_\perp) \\
&= \frac{1}{T}\sum_{t=1}^{T/2}\mathrm{tr}(Q_\perp' P_I^1 \Sigma_\varepsilon P_I^{1'} Q_\perp) \leq \frac{d-r}{2}\|P_I^1 \Sigma_\varepsilon P_I^{1'}\| \\
&\leq \frac{d-r}{2}\|\Sigma_\varepsilon\| = \mathcal{O}(d)
\end{aligned}
$$

following the arguments in (C.21) and by Assumption CR.3. Then, it suffices to show that

$$
\|\widehat{\Pi}_r^{-1/2}\bar{Q}_r' Q_\perp\|^2 = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{d^3}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{dT}\right),
$$

which holds since $\bar{Q}_r' Q_\perp = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{d}\right) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{T}}\right)$ by Corollary C.3 and $\frac{1}{\sqrt{d}}\left(\frac{1}{d}\widehat{\Pi}_r\right)^{-1/2} = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{d}}\right)$ by Corollary C.1. □

D Matrix norm inequalities

We collect here some results on matrix norms that are used throughout the paper.

M.1. $\|AB\|_F \leq \|A\|\|B\|_F$ for matrices A, B; see Theorem 1 in Fang, Loparo, and Feng (1994).

M.2. For a general block matrix (not necessarily positive semidefinite), we know that

$$
\begin{aligned}
\left\|\begin{pmatrix} A & C \\ C' & B \end{pmatrix}\right\|^2 &= \lambda_{\max}\begin{pmatrix} AA' + CC' & AC + CB' \\ C'A' + BC' & BB' + CC' \end{pmatrix} \\
&= \left\|\begin{pmatrix} AA' + CC' & AC + CB' \\ C'A' + BC' & BB' + CC' \end{pmatrix}\right\|
\end{aligned}
$$

$$\leq \|AA' + CC'\| + \|BB' + CC'\| \tag{D.1}$$

$$\leq \|A\|^2 + \|B\|^2 + 2\|C\|^2, \tag{D.2}$$

where (D.1) is due to the unitary invariance of the operator norm and Lemma 1.1. in Bourin, Lee, and Lin (2012). For (D.2), see Weyl's Theorem (Theorem 4.3.1 in Horn and Johnson (2012)).

M.3. Let A, B be Hermitian and suppose that B is positive semidefinite. Then, $\lambda_{\max}(A) \leq \lambda_{\max}(A + B)$ by Corollary 4.3.12 in Horn and Johnson (2012).

M.4. For matrices A, B,

$$\left\| \begin{pmatrix} A \\ B \end{pmatrix} \right\|^2 = \lambda_{\max} \left(A'A + B'B \right) = \|A'A + B'B\| \leq \|A\|^2 + \|B\|^2$$

again by Weyl's Theorem (Theorem 4.3.1 in Horn and Johnson (2012)).

Supplement

We omitted some of the proofs which follow very similar arguments as the analogous results in Han and Inoue (2015). For completeness we put the details regarding those proofs into a supplementary document Düker and Pipiras (2023). This document contains the proofs of Propositions 3.1–3.3 and Proposition 4.1. We also point out that Han and Inoue (2015) omit proofs regarding the alternative hypothesis. We provide some details of the proof of our Proposition 4.1 in the supplementary material.

References

Andrews, D. Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–56, 1993.

Baek, C., Gates, K., Leinwand, B., and Pipiras, V. Two sample tests for high-dimensional autocovariances. *Computational Statistics & Data Analysis*, 153:107067, 2021.

Bai, J. Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171, 2003.

Bai, J. and Ng, S. Determining the number of factors in approximate factor models. *Econometrica*, 70(1): 191–221, 2002.

Bai, J. and Ng, S. *Large Dimensional Factor Analysis*. Now Publishers Inc, 2008.

Bai, J., Duan, J., and Han, X. Likelihood ratio test for structural changes in factor models. *Preprint, Available at arXiv:2206.08052*, 2022.

Baltagi, B., Kao, C., and Wang, F. Estimating and testing high dimensional factor models with multiple structural changes. *Journal of Econometrics*, 220(2):349–365, 2021.

Barigozzi, M., Cho, H., and Fryzlewicz, P. Simultaneous multiple change-point and factor analysis for high-dimensional time series. *Journal of Econometrics*, 206(1):187–225, 2018.

Bourin, J.-C., Lee, E.-Y., and Lin, M. On a decomposition lemma for positive semi-definite block-matrices. *Linear Algebra and its Applications*, 437(7):1906–1912, 2012.

Bräuning, F. and Koopman, S. The dynamic factor network model with an application to global credit-risk. *Preprint, Available at EconPapers*, 2016.

Bräuning, F. and Koopman, S. The dynamic factor network model with an application to international trade. *Journal of Econometrics*, 216(2):494–515, 2020.

Carvalho, C., Chang, J., Lucas, J., Nevins, J., Wang, Q., and West, M. High-dimensional sparse factor modeling: applications in gene expression genomics. *Journal of the American Statistical Association*, 103 (484):1438–1456, 2008.

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

Cox, J. and Ross, S. A survey of some new results in financial option pricing theory. *The Journal of Finance*, 31(2):383–402, 1976.

De Roover, K., Ceulemans, E., Timmerman, M., Vansteelandt, K., Stouten, J., and Onghena, P. Clusterwise simultaneous component analysis for analyzing structural differences in multivariate multiblock data. *Psychological Methods*, 17(1):100, 2012.

Donald, S., Fortuna, N., and Pipiras, V. On rank estimation in semidefinite matrices. *Preprint, Available at EconPapers*, 2010.

Doz, C., Giannone, D., and Reichlin, L. A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, 164(1):188–205, 2011.

Doz, C., Giannone, D., and Reichlin, L. A quasi–maximum likelihood approach for large, approximate dynamic factor models. *Review of Economics and Statistics*, 94(4):1014–1024, 2012.

Duan, J., Bai, J., and Han, X. Quasi-maximum likelihood estimation of break point in high-dimensional factor models. *Journal of Econometrics*, 233(1):209–236, 2023.

Düker, M.-C. and Pipiras, V. Supplement to "Testing for common structures in high-dimensional factor models". *Preprint*, 2024.

Fan, J., Fan, Y., and Lv, J. High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147(1):186–197, 2008.

Fan, J., Masini, R., and Medeiros, M. Bridging factor and sparse models. *The Annals of Statistics*, 51(4): 1692–1717, 2023.

Fang, Y., Loparo, K., and Feng, X. Inequalities for the trace of matrix product. *IEEE Transactions on Automatic Control*, 39(12):2489–2490, 1994.

Feng, G., Giglio, S., and Xiu, D. Taming the factor zoo: A test of new factors. *The Journal of Finance*, 75 (3):1327–1370, 2020.

Fisher, Z., Kim, Y., Fredrickson, B., and Pipiras, V. Penalized estimation and forecasting of multiple subject intensive longitudinal data. *Psychometrika*, 87(2):1–29, 2022.

Gross, S. and Tibshirani, R. Data shared lasso: A novel tool to discover uplift. *Computational Statistics & Data Analysis*, 101:226–235, 2016.

Han, X. and Inoue, A. Tests for parameter instability in dynamic factor models. *Working paper, Available at SSRN 2314338*, 2014.

Han, X. and Inoue, A. Tests for parameter instability in dynamic factor models. *Econometric Theory*, 31: 1117–1152, 2015.

Han, X. Estimation and inference of dynamic structural factor models with over-identifying restrictions. *Journal of Econometrics*, 202(2):125–147, 2018.

Hannan, E. *Multiple Time Series.* Wiley Series in Probability and Statistics. Wiley, New York, 1970.

Horn, R. and Johnson, C. *Matrix Analysis.* Cambridge University Press, 2012.

Jalali, A., Ravikumar, P., and Sanghavi, S. A dirty model for multiple sparse regression. *IEEE Transactions on Information Theory*, 59(12):7947–7968, 2013.

Jung, R. C., Liesenfeld, R., and Richard, J.-F. Dynamic factor models for multivariate count data: An application to stock-market trading activity. *Journal of Business & Economic Statistics*, 29(1):73–85, 2011.

Lam, C., Yao, Q., and Bathia, N. Estimation of latent factors for high-dimensional time series. *Biometrika*, 98(4):901–918, 2011.

Lam, C. and Yao, Q. Factor modeling for high-dimensional time series: inference for the number of factors. *Annals of Statistics*, 40(2):694–726, 2012.

Li, H., Li, Q., and Shi, Y. Determining the number of factors when the number of factors can increase with sample size. *Journal of Econometrics*, 197(1):76–86, 2017.

Liao, Y. and Todorov, V. Changes in the span of systematic risk exposures. *Preprint, Available at SSRN 4354011*, 2023.

Ma, S. and Su, L. Estimation of large dimensional factor models with an unknown number of breaks. *Journal of Econometrics*, 207(1):1–29, 2018.

Markowitz, H. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

Ollier, E. and Viallon, V. Joint estimation of K related regression models with simple L_1–norm penalties. *Preprint, Available at arXiv:1411.1594*, 2014.

Onatski, A. Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77(5): 1447–1479, 2009.

Ross, S. The capital asset pricing model (CAPM), short-sale restrictions and related issues. *The Journal of Finance*, 32(1):177–183, 1977.

Sargent, T. and Sims, C. A. Business cycle modeling without pretending to have too much a priori economic theory. *New Methods in Business Cycle Research*, 1:145–168, 1977.

Silin, I. and Fan, J. Hypothesis testing for eigenspaces of covariance matrix. *Preprint, Available at arXiv:2002.09810*, 2020.

Song, H. and Zhang, Z. Analyzing multiple multivariate time series data using multilevel dynamic factor models. *Multivariate Behavioral Research*, 49(1):67–77, 2014.

Stock, J. and Watson, M. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179, 2002.

Stock, J. and Watson, M. Forecasting in dynamic factor models subject to structural instability. *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry*, 173:205, 2009.

Su, L. and Wang, X. On time-varying factor models: Estimation and testing. *Journal of Econometrics*, 198 (1):84–101, 2017.

Timmerman, M. E. and Kiers, H. A. Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences. *Psychometrika*, 68(1):105–121, 2003.

Wang, F. and Wang, H. Modelling non-stationary multivariate time series of counts via common factors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(4):769–791, 2018.

Wedel, M., Böckenholt, U., and Kamakura, W. Factor models for multivariate count data. *Journal of Multivariate Analysis*, 87(2):356–369, 2003.

Marie-Christine Düker
Dept. of Statistics and Data Science
FAU Erlangen-Nürnberg
Cauerstrasse 11
91058 Erlangen, Germany
marie.dueker@fau.de

Vladas Pipiras
Dept. of Statistics and Operations Research
UNC Chapel Hill
CB#3260, Hanes Hall
Chapel Hill, NC 27599, USA
pipiras@email.unc.edu