Estimating Change-Point Latent Factor Models for High-Dimensional Time Series*

Xialu Liu

San Diego State University

Ting Zhang

Boston University

June 21, 2021

Abstract

We consider estimating a factor model for high-dimensional time series that contains structural breaks in the factor loading space at unknown time points. We first study the case when there is one change point in factor loadings, and propose a consistent estimator for the structural break location, whose convergence rate is shown to depend on an interplay between the dimension of the observed time series and the strength of the underlying factor structure. Our results reveal that the asymptotic behavior of the proposed estimator can be asymmetric in the sense that a larger estimation error can occur toward the regime with weaker factor strength. Based on the proposed estimator for the structural break location, we also consider the problem of estimating the factor loading spaces before and after the structural break. We show that the proposed estimators for change-point location and loading spaces are consistent when the numbers of factors are correctly estimated or overestimated. The algorithm for multiple change-point detection is also developed in the paper. Compared with existing results on change-point factor analyses of high-dimensional time series, a distinguished feature of the current paper is that the noise process is not necessarily assumed to be idiosyncratic and as a result we allow the noise process with potentially strong cross-sectional dependence.

*Xialu Liu is Associate Professor, Management Information Systems Department, San Diego State University, San Diego, CA 92182. Email: xialu.liu@sdsu.edu. Ting Zhang is Assistant Professor, Department of Mathematics and Statistics, Boston University, Boston, MA 02215. E-mail: tingz@bu.edu. Xialu Liu is the corresponding author.

Another advantage for the proposed method is that it is specifically designed for the changes in the factor loading space and the stationarity assumption is not imposed on either the factor or noise process, while most existing methods for change-point detection of high-dimensional time series with/without a factor structure require the data to be stationary or 'close' to a stationary process between two change points, which is rather restrictive. Numerical experiments including a Monte Carlo simulation and a real data application are presented to illustrate the proposed estimators perform well.

KEYWORDS: Change point estimation; high-dimensional time series; large latent factor model; non-stationary process; strong cross-sectional dependence.

1 Introduction

High-dimensional time series has been emerging as a common and important data type in applications from a number of disciplines, including climate science, economics, finance, medical science, and telecommunication engineering among others. Although numerous statistical methods and their associated theory have been developed for the modeling and inference of time series data, existing results mostly focused on the univariate or finite-dimensional multivariate case. The problem of extending existing results developed under low-dimensional settings to handle high-dimensional time series, however, is typically nontrivial and requires significant innovations. For example, when the dimension is larger than the length of the observed time series, the commonly used autoregressive moving-average (ARMA) model in its conventional form may face a serious identification problem as commented by Lam et al. (2011). To handle the phenomenon of high dimensionality, one typically resorts to certain sparsity-type conditions for the purpose of dimension reduction. For example, when considering vector autoregressive (VAR) models in the high-dimensional setting, one typically needs to assume that the coefficient matrices are sparse in a suitable sense in order to obtain their meaningful estimators; see for example Basu and Michailidis (2015), Davis et al. (2016) and references therein for research results in this direction.

Unlike the aforementioned sparse VAR approach that aims at extending existing parametric time series models to their sparse high-dimensional counterparts, a popular approach in the literature for modeling high-dimensional time series is through the use of a factor model; see for example Chamberlain and Rothschild (1983), Stock and Watson (1998), Bai and Ng (2002), Bai (2003) and Forni et al. (2004) among others. The approximate factor model is one of the most widely used models discussed by Bai and Ng (2002) and Bai (2003), and it assumes that most of the variation in high-dimensional time series data can be explained by a few of factors. Serial dependence is allowed to exist in both the factor and noises but the cross-sectional dependence in the noise has to satisfy the condition that $\sum_{i=1}^{p}\sum_{j=1}^{p}|\sigma_{t,ij}| \leqslant Cp$ for any $t = 1, \ldots, n$, where $\sigma_{t,ij}$ is the (i, j)-th entry in the covariance matrix of the noise process at time t, C is a positive constant, p is the dimension of time series, and n is the time length. The common component in such factor models is asymptotically identifiable when the number of time series goes to infinity. On the other hand, Lam et al. (2011) proposed an alternative way to define the factor model for time series data. In their model, the common factors are now viewed as the force that drives all the dynamics and is used to explain the serial dependence in the data. The noise process in this setting can exhibit a strong notion of cross-sectional dependence with $|\sigma_{t,ij}| \leqslant C$ for $i, j = 1, \ldots, p$ and $t = 1, \ldots, n$, and the common component in the resulting factor model becomes identifiable no matter whether the number of time series grows to infinity with the time length. Therefore, both classes of factor models have been proven to be useful in different scenarios.

Change-point detection in the approximate factor model has been well investigated; see for example Breitung and Eickmeier (2011), Chen et al. (2014), Han and Inoue (2015), Barigozzi et al. (2018), Ma and Su (2018) and references therein. However, the model proposed by Lam et al. (2011) in the change-point setting has not been much explored. Related works in this direction include Liu and Chen (2016), which modeled change points as regime shifts between different states of a hidden Markov chain, and Liu and Chen (2020), which discussed a threshold variable approach to modeling the change-point mechanism. The major goal of the current paper is to consider estimating the recently proposed factor model of Lam et al. (2011) in the change-point setting, while not imposing additional structural assumption on the underlying change-point mechanism.

Change-point detection for high-dimensional time series become popular recently. Cho and Fryzlewicz (2012) used the nonparametric locally stationary wavelet model to estimate the number and locations of change points. Xie et al. (2013) described a new approach and introduced the multi-scale model to detect breaks for data with missing values. Cho and Fryzlewicz (2015) proposed the sparsified binary segmentation algorithm to segment the second-order structure of a time series. Cho (2016) used the double CUSUM statistic combined with the binary segmentation algorithm to examine the breakpoints. The existing methods aiming at identifying abrupt changes for high-dimensional time series with/without a factor structure, require the data to be stationary or 'close' to a stationary process within regimes, which is rather restrictive. The algorithm we proposed in this paper focuses exclusively on the changes of the factor loading space and can be applied to the case that the factor and noise processes are non-stationary.

In Section 2, we consider the factor model of Lam et al. (2011) with a single change point, and propose a projection-based change-point estimator whose convergence rate shown in Section 3 depends on an interplay between the dimension of the observed time series and the strength of the underlying factor structure. Furthermore, our results reveal that its asymptotic behavior can be asymmetric in the sense that a larger estimation error can occur toward the regime with weaker factor strength. Based on the proposed estimator for the structural break location, we also consider the problem of estimating the factor loading spaces before and after the structural break. We show that the proposed estimators for change-point location and loading spaces are still consistent when the numbers of factors are correctly estimated or overestimated. Section 4 describes the algorithm to identify and locate multiple change points when the number of change points is unknown. Compared with existing results on change-point detection of high-dimensional time series, one advantage of the current paper is that the stationarity assumption for the factor or noise processes is not necessary and as a result our method performs well when the observed data are non-stationary within regimes. It can be seen from our simulation results in Section 5 that existing results on multiple change-point detection developed for high-dimensional time series without a factor structure may struggle in detecting and locating the change points when the observed process is non-stationary within regimes, while the proposed algorithm works reasonably well. The performance for the proposed methods are further illustrated in Section 6 using real data, and Section 7 concludes the paper. Technical proofs are deferred to the Appendix.

2 Estimating Change-Point Factor Model with a Single Change Point

In this section we first introduce the factor model with a single change point, and then develop the projection-based change point estimator in Section 2.2. The estimation for the numbers of factors is discussed in Section 2.3.

2.1 Change-point factor model for high-dimensional time series with a single change point

Suppose we observe a p-dimensional time series $\mathbf{y}_t$, $t = 1, \ldots, n$, according to the factor model of Lam et al. (2011), then

$$\mathbf{y}_t = \mathbf{A}\mathbf{x}_t + \boldsymbol{\varepsilon}_t, \tag{1}$$

where $\{\mathbf{x}_t\}$ is a latent factor process whose dimension k_0 is typically much smaller than p, $\mathbf{A} \in \mathbb{R}^{p \times k_0}$ is the associated loading matrix, and $\{\boldsymbol{\varepsilon}_t\}$ denotes a noise process. Note that both $\mathbf{A}$ and $\mathbf{x}_t$ are unobserved and can be replaced by $\mathbf{AU}$ and $\mathbf{U}^{-1}\mathbf{x}_t$ for any invertible matrix $\mathbf{U} \in \mathbb{R}^{k_0 \times k_0}$ in the model. Although the loading matrix $\mathbf{A}$ is not identifiable, the space spanned by the columns of $\mathbf{A}$, called loading space and denoted by $\mathcal{M}(\mathbf{A})$, is uniquely defined. Thus, estimation of the loading space instead of the loading matrix is one of the primary goals for factor models. Since its first appearance in the influential work of Lam et al. (2011), the latent factor model (1) has been widely used in the literature for dimension reduction of high-dimensional time series; see for example Lam and Yao (2012), Chang et al. (2015), Liu and Chen (2016) and references therein.

In model (1), the factor loading structure is assumed to remain the same over the whole sampling period, which is very restrictive when the datasets span a long time period and may make forecasting and inference misleading and unreliable (Su and Wang, 2017). For this, we consider the change-point factor model

$$\mathbf{y}_t = \begin{cases} \mathbf{A}_1\mathbf{x}_{t,1} + \boldsymbol{\varepsilon}_t, & \text{if } t \leqslant r_0; \\ \mathbf{A}_2\mathbf{x}_{t,2} + \boldsymbol{\varepsilon}_t, & \text{if } t > r_0, \end{cases} \tag{2}$$

where $\mathbf{x}_{t,i} \in \mathbb{R}^{k_i}$, $i = 1, 2$, represents the underlying latent factor before and after the change point whose location is denoted by r_0, $\mathbf{A}_1$ and $\mathbf{A}_2$ are the associated loading matrices with $\mathcal{M}(\mathbf{A}_1) \neq \mathcal{M}(\mathbf{A}_2)$, and $(\boldsymbol{\varepsilon}_t)$ is an independent process whose covariance matrix is allowed to be time-varying. Recently, there have been efforts in studying the change-point factor model by incorporating certain beliefs or structural assumptions on the change-point mechanism into the analysis. For example, Liu and Chen (2016) modeled the change-point mechanism by a finite-state hidden Markov chain, in which case a change point occurs when there is a regime switching in the hidden state variable. On the other hand, Liu and Chen (2020) considered using a threshold variable to model the change point, where the threshold variable is assumed to be α-mixing and observable up to a small number of unknown parameters. Instead of introducing a Markov chain process or an additional threshold variable, we shall in the current paper focus on the change-point factor model (2) which uses time to naturally divide the observed process into segments before and after the change point. We shall in the following introduce a projection-based change point estimator and study its asymptotic properties.

We shall now introduce some notations. For a matrix $\mathbf{H}$, we use $\|\mathbf{H}\|_F$ and $\|\mathbf{H}\|_2$ to denote its Frobenius and L-2 norms respectively. Let $\sigma_i(\mathbf{H})$ be the i-th largest singular value of $\mathbf{H}$, and $\|\mathbf{H}\|_{\min}$ be the square root of minimum nonzero eigenvalue of $\mathbf{H}'\mathbf{H}$. In addition, we use $\mathrm{tr}(\mathbf{H})$ to denote its trace if $\mathbf{H}$ is a square matrix. Also, we write $a \asymp b$ if $a = O(b)$ and $b = O(a)$, and we use $\lfloor x \rfloor$ and $\lceil x \rceil$ to denote the largest previous and smallest following integers of x.

2.2 A projection-based change point estimator when there exists a single change point

For the change-point factor model (2), we consider the situation where one only observes $(\mathbf{y}_t)$ but not $(\mathbf{x}_{t,i})$ nor $(\boldsymbol{\varepsilon}_t)$. In this case, the loading matrices themselves are not directly identifiable as one can replace $(\mathbf{A}_i, \mathbf{x}_{t,i})$ in (2) by $(\mathbf{A}_i\mathbf{U}_i, \mathbf{U}_i^{-1}\mathbf{x}_t)$ for any invertible matrix $\mathbf{U}_i \in \mathbb{R}^{k_i \times k_i}$ for $i = 1, 2$. However, the linear spaces spanned by columns of the loading matrices, denoted by $\mathcal{M}(\mathbf{A}_i) = \mathcal{M}(\mathbf{A}_i\mathbf{U}_i)$, will not be affected by such a transformation and are indeed uniquely identifiable for $i = 1, 2$. Therefore, compared with the conventional change-point setting where the object of interest is typically a finite-dimensional vector, the current setting can be more challenging as we have to deal with linear spaces spanned by columns of high-dimensional matrices. We in the following propose a projection-based estimator for the change-point location, which is shown to consistently identify the time point at which the underlying loading space undergoes a structural break.

Given $\mathbf{y}_1, \ldots, \mathbf{y}_n$, for any $\gamma \in (0, 1)$ we can split the data into $\mathbf{y}_1, \ldots, \mathbf{y}_{\lfloor \gamma n \rfloor}$ and $\mathbf{y}_{\lfloor \gamma n \rfloor + 1}, \ldots, \mathbf{y}_n$. Let $I_{t,1}(\gamma)$ and $I_{t,2}(\gamma)$ be the associated indicator functions where $I_{t,1}(\gamma) = 1$ if $1 \leqslant t \leqslant \lfloor \gamma n \rfloor$ and $I_{t,2}(\gamma) = 1$ if $\lfloor \gamma n \rfloor < t \leqslant n$. For $i = 1, 2$, we consider the generalized second cross moment matrices

$$\boldsymbol{\Sigma}_{y,i}(h, \gamma) = \frac{1}{n} \sum_{t=1}^{n} \mathrm{E}\left[\mathbf{y}_t \mathbf{y}_{t+h}' I_{t,i}(\gamma) I_{t+h,i}(\gamma)\right],$$

which can be estimated by its sample version

$$\widehat{\boldsymbol{\Sigma}}_{y,i}(h, \gamma) = \frac{1}{n} \sum_{t=1}^{n} \mathbf{y}_t \mathbf{y}_{t+h}' I_{t,i}(\gamma) I_{t+h,i}(\gamma).$$

By borrowing information from different lags, we consider

$$\widehat{\mathbf{M}}_i(\gamma) = \sum_{h=1}^{h_0} \widehat{\boldsymbol{\Sigma}}_{y,i}(h, \gamma) \widehat{\boldsymbol{\Sigma}}_{y,i}(h, \gamma)',$$

which serves as an estimator for

$$\mathbf{M}_i(\gamma) = \sum_{h=1}^{h_0} \boldsymbol{\Sigma}_{y,i}(h,\gamma)\boldsymbol{\Sigma}_{y,i}(h,\gamma)', \tag{3}$$

where h_0 is a pre-specified positive integer. If $\gamma = \gamma_0 = r_0/n$ correctly specifies the change-point location, then one can show that

$$\mathbf{M}_i(\gamma) = \sum_{h=1}^{h_0} \mathbf{A}_i \left[\boldsymbol{\Sigma}_{x,i}(h,\gamma)\mathbf{A}_i'\mathbf{A}_i\boldsymbol{\Sigma}_{x,i}(h,\gamma)'\right]\mathbf{A}_i', \tag{4}$$

where

$$\boldsymbol{\Sigma}_{x,i}(h,\gamma) = \frac{1}{n}\sum_{t=1}^{n} \mathrm{E}\left[\mathbf{x}_{t,i}\mathbf{x}_{t+h,i}' I_{t,i}(\gamma) I_{t+h,i}(\gamma)\right] \tag{5}$$

is the generalized second cross moment matrix for the hidden factor process. In this case, $\mathbf{M}_i(\gamma)$ is a symmetric non-negative definite matrix sandwiched by $\mathbf{A}_i$ and $\mathbf{A}_i'$, and thus its eigenspace associated with nonzero eigenvalues coincides with the loading space $\mathcal{M}(\mathbf{A}_i)$ for $i = 1,2$ if $\boldsymbol{\Sigma}_{x,i}(h,\gamma)$ is full rank for some $h \in [1, h_0]$. This motivates us to consider estimating the change-point location by exploiting the orthogonality between eigenspaces associated with zero and nonzero eigenvalues. To illustrate the idea, we first consider the simple scenario where the number of factors k_i is known; see Section 2.3 for the case when it is unknown. For $i = 1, 2$, let $\mathbf{q}_{i,k}(\gamma)$ be the unit eigenvector of $\mathbf{M}_i(\gamma)$ associated with its k-th largest eigenvalue for $k = 1, \ldots, k_i$, and $\mathbf{q}_{i,k_i+1}, \mathbf{q}_{i,k_i+2}, \ldots, \mathbf{q}_{i,p}$ be the unit eigenvectors of $\mathbf{M}_i(\gamma)$ corresponding to zero eigenvalues, with $\mathbf{1}'\mathbf{q}_{i,j}(\gamma) > 0$ for $j = 1, \ldots, p$. Define

$$\mathbf{Q}_i(\gamma) = (\mathbf{q}_{i,1}(\gamma), \ldots, \mathbf{q}_{i,k_i}(\gamma)), \quad \mathbf{B}_i(\gamma) = (\mathbf{q}_{i,k_i+1}(\gamma), \ldots, \mathbf{q}_{i,p}(\gamma)), \tag{6}$$

which form orthogonal matrices representing eigenvectors for nonzero and zero eigenvalues respectively. For notational ease, we use $\mathbf{Q}_i$, $\mathbf{B}_i$, and $\mathbf{M}_i$ to denote $\mathbf{Q}_i(\gamma_0)$, $\mathbf{B}_i(\gamma_0)$, and $\mathbf{M}_i(\gamma_0)$. To estimate the change-point location, we introduce the projection criterion

$$G(\gamma) = \sum_{i=1}^{2} g_i(\gamma), \quad g_i(\gamma) = \left\|\mathbf{B}_i'\,\mathbf{M}_i(\gamma)\,\mathbf{B}_i\right\|_2. \tag{7}$$

Note that although $\mathbf{B}_i$ is not uniquely defined and subject to any orthogonal transformation, $g_i(\gamma)$ is invariant under such transformations. If we project the cross moment matrices $\{\boldsymbol{\Sigma}_{y,i}(h,\gamma),\ h = 1, \ldots, h_0\}$ onto $\mathcal{M}(\mathbf{B}_i)$, the linear space spanned by columns of $\mathbf{B}_i$, then by (3) we can see that $G(\gamma)$ measures the squared norm of such projections. If $\gamma = \gamma_0$ is correctly specified, then $\mathbf{M}_i(\gamma_0) = \mathbf{M}_i$ for $i = 1, 2$, and by (4) we have

$$G(\gamma_0) = \sum_{i=1}^{2} \left\|\mathbf{B}_i'\,\mathbf{M}_i(\gamma_0)\,\mathbf{B}_i\right\|_2 = \sum_{i=1}^{2} \left\| \sum_{h=1}^{h_0} \left\{\mathbf{B}_i'\,\mathbf{A}_i\left[\boldsymbol{\Sigma}_{x,i}(h,\gamma_0)\mathbf{A}_i'\mathbf{A}_i\boldsymbol{\Sigma}_{x,i}(h,\gamma_0)\right]\mathbf{A}_i'\,\mathbf{B}_i\right\}\right\|_2 = 0.$$

On the other hand, if $\gamma \neq \gamma_0$, then the data are not correctly separated and one of the subsets contains data from different factor loading structures. In fact, we can show that if the amount of misspecification $|\gamma - \gamma_0|$ exceeds a certain rate, then the norm of the aforementioned projection will be strictly positive, namely $G(\gamma) > 0$; see Lemma 6 in the Appendix.

This motivates us to consider estimating the change-point location by minimizing an empirical version of the projection criterion $G(\gamma)$. To be more specific, let $\widehat{\lambda}_{i,1} \geqslant \widehat{\lambda}_{i,2} \geqslant \ldots \geqslant \widehat{\lambda}_{i,p}$ be the p eigenvalues of $\widehat{\mathbf{M}}_i(\gamma)$, and $\widehat{\mathbf{q}}_{i,1}(\gamma), \widehat{\mathbf{q}}_{i,2}(\gamma), \ldots, \widehat{\mathbf{q}}_{i,p}(\gamma)$ be the set of corresponding orthonormal eigenvectors with $\mathbf{1}'\widehat{\mathbf{q}}_{i,j}(\gamma) > 0$, then empirical versions of quantities in (6) are given by

$$\widehat{\mathbf{Q}}_i(\gamma) = (\widehat{\mathbf{q}}_{i,1}(\gamma), \ldots, \widehat{\mathbf{q}}_{i,k_i}(\gamma)), \quad \widehat{\mathbf{B}}_i(\gamma) = (\widehat{\mathbf{q}}_{i,k_i+1}(\gamma), \ldots, \widehat{\mathbf{q}}_{i,p}(\gamma)). \tag{8}$$

For statistical analyses in the change-point setting, it is typically assumed that the change point does not occur in the boundary area, namely there exists $0 < \eta_1 < \eta_2 < 1$ such that $\gamma_0 \in (\eta_1, \eta_2)$. A popular choice for (η_1, η_2) is in the form of $(\varepsilon, 1-\varepsilon)$ for some small ε such as 0.1; see for example the discussions in Zhou and Shao (2013) and Zhang and Lavitas (2018). With data in $[0, \eta_1]$ and $[\eta_2, 1]$, we can obtain consistent estimators for $\mathbf{B}_1(\eta_1)$ and $\mathbf{B}_2(\eta_2)$ (Lam et al., 2011), and then estimate $G(\gamma)$ by

$$\widehat{G}(\gamma) = \sum_{i=1}^{2} \left\| \widehat{\mathbf{B}}_i(\eta_i)' \, \widehat{\mathbf{M}}_i(\gamma) \, \widehat{\mathbf{B}}_i(\eta_i) \right\|_2, \tag{9}$$

and we propose to estimate the change-point location by

$$\widehat{\gamma} = \operatorname*{argmin}_{\gamma \in \{0, \frac{1}{n}, \ldots, 1\} \cap (\eta_1, \eta_2)} \widehat{G}(\gamma). \tag{10}$$

It can be seen from our theoretical results in Section 3 that the proposed estimator $\widehat{\gamma}$ consistently estimates the change-point location, and its convergence rate depends on an interplay between the dimension of the observed time series and the strength of the underlying factor structure. In addition, it may have an asymmetric asymptotic behavior depending on the factor strength in the regimes before and after the change point. Given the change point estimator $\widehat{\gamma}$, one can estimate the loading spaces before and after by $\mathcal{M}[\widehat{\mathbf{Q}}_1(\widehat{\gamma})]$ and $\mathcal{M}[\widehat{\mathbf{Q}}_2(\widehat{\gamma})]$ respectively, whose consistency and convergence rates are also studied in Section 3.

2.3 Estimation when the numbers of factors are unknown

The number of factors is often unknown in factor analysis. Many approaches to identifying the number of factors have been developed in the literature. The factor model is characterized by the presence of a large eigengap between eigenvalues of the covariance matrix (Barigozzi and Cho, 2020). Based on this observation, the scree test introduced by Cattell (1966) uses an eye-ball rule to select the number of factors. Parallel analysis (Horn, 1965; Buja and Eyuboglu, 1998; Dobriban, 2020) and deterministic parallel analysis (Dobriban and Owen, 2019; Dobriban, 2020) are also effective methods to estimate the number of factors designed for data with no serial dependence. Factor analysis for time series data, which involves dependence between observations and brings an extra layer of difficulty, was studied by Forni et al. (2000), Bai and Ng (2002), Onatski. (2010), and Ahn and Horenstein (2013). Bai and Ng (2002) constructed various criterion functions based on the covariance matrix of the observed process, and Ahn and Horenstein (2013) utilized the eigenvalues of the covariance matrix to determine the number of factors. The aforementioned methods cannot handle the situation when there exists strong cross-sectional dependence in the noise process. To solve the problem, Lam and Yao (2012) proposed a ratio-estimator based on the eigenvalues of the covariance matrices at nonzero lags.

Following the approach used in Lam and Yao (2012), we can estimate k_1 and k_2 through the eigenvalue ratios in the current change-point setting as well. To be more specific, for $i = 1, 2$, let $\widehat{\lambda}_{i,k}(\eta_i)$ be the k-th largest eigenvalue of $\widehat{\mathbf{M}}_i(\eta_i)$, then k_i can be estimated by

$$\widehat{k}_i = \underset{1 \leqslant k \leqslant R}{\operatorname{argmin}} \frac{\widehat{\lambda}_{i,k+1}(\eta_i)}{\widehat{\lambda}_{i,k}(\eta_i)}. \tag{11}$$

The search cannot be extended to p because the minimum eigenvalue of $\mathbf{M}_i(\eta_i)$ goes to 0. We follow Lam and Yao (2012) and use $R = p/2$ when $n \geqslant p$; when $n < p$, we let $R = n/2$. We can then plug (11) into the estimation procedure described in Section 2.2 to handle the situation when k_1 and k_2 are unknown. In particular, the projection-based criterion function (9) in this case becomes

$$\widehat{G}(\gamma, \widehat{k}_1, \widehat{k}_2) = \sum_{i=1}^{2} \left\| \widehat{\mathbf{B}}_{i,\widehat{k}_i}(\eta_i)' \, \widehat{\mathbf{M}}_i(r) \, \widehat{\mathbf{B}}_{i,\widehat{k}_i}(\eta_i) \right\|_2,$$

where $\widehat{\mathbf{B}}_{i,\widehat{k}_i}(\eta_i) = (\widehat{\mathbf{q}}_{i,\widehat{k}_i+1}(\eta_i), \ldots, \widehat{\mathbf{q}}_{i,p}(\eta_i))$, and we estimate the change-point location by

$$\widetilde{\gamma} = \arg \min_{\gamma \in \{0, \frac{1}{n}, \ldots, 1\} \bigcap (\eta_1, \eta_2)} \widehat{G}(\gamma, \widehat{k}_1, \widehat{k}_2).$$

Similar to the discussion in Section 2.2, the loading spaces in this case can be estimated by $\mathcal{M}(\widetilde{\mathbf{Q}}_i(\widetilde{\gamma}, \widehat{k}_i))$ where

$$\widetilde{\mathbf{Q}}_i(\widetilde{\gamma}, \widehat{k}_i) = (\widehat{\mathbf{q}}_{i,1}(\widetilde{\gamma}), \ldots, \widehat{\mathbf{q}}_{i,\widehat{k}_i}(\widetilde{\gamma})), \text{ for } i = 1, 2.$$

The consistency results and explicit estimation bounds for the change point $\widetilde{\gamma}$ and loading spaces $\mathcal{M}(\widetilde{\mathbf{Q}}_i(\widetilde{\gamma}, \widehat{k}_i))$ are provided in Theorem 3 and Corollary 1 respectively.

3 Theoretical Properties

We shall here study the asymptotic properties of the estimators proposed in Section 2 for change-point factor modeling of high-dimensional time series. For this, we need to introduce the notion of factor strength, which plays an important role in understanding the theory of factor modeling and has been commonly used in the literature; see for example Bai and Ng (2002), Bai (2003), Doz et al. (2011), Lam et al. (2011), Lam and Yao (2012), Chang et al. (2015), and Liu and Chen (2020) among others. In particular, it assumes that the loading matrix $\mathbf{A}_i$ satisfies

$$\|\mathbf{A}_i\|_2^2 \asymp \|\mathbf{A}_i\|_{\min}^2 \asymp p^{1-\delta_i}$$

for some $0 \leqslant \delta_i \leqslant 1$, and the factor strength is said to be strong if $\delta_i = 0$ and weak if $\delta_i \in (0,1]$. The factor strength measures the relative growth rate of the amount of information carried by the observed process $\mathbf{y}_t$ about the factor process $\mathbf{x}_t$ as the dimension p increases, with respect to the growth rate of the amount of noise process.

When presenting the asymptotic properties of $\mathcal{M}[\widehat{\mathbf{Q}}_i(\widehat{\gamma})]$, we also need to introduce a measure that quantifies the distance between two linear spaces which can then be used to assess the statistical performance of the proposed estimators for loading spaces. In particular, let $\mathbf{S}_1$ and $\mathbf{S}_2$ be full rank matrices in $\mathbb{R}^{p\times q_1}$ and $\mathbb{R}^{p\times q_2}$ respectively with $\max(q_1,q_2) \leqslant p$. Denote $\mathbf{O}_i$ the matrix whose columns form an orthonormal basis of $\mathcal{M}(\mathbf{S}_i)$ for $i = 1,2$, then the distance between column spaces of $\mathbf{S}_1$ and $\mathbf{S}_2$ can be measured by

$$\mathcal{D}\{\mathcal{M}(\mathbf{S}_1),\, \mathcal{M}(\mathbf{S}_2)\} = \left\{1 - \frac{\operatorname{tr}(\mathbf{O}_1\mathbf{O}_1'\mathbf{O}_2\mathbf{O}_2')}{\min(q_1,q_2)}\right\}^{1/2}. \tag{12}$$

The distance measure (12) was first introduced in Liu and Chen (2020), and is a quantity between 0 and 1. In particular, it equals to 0 if $\mathcal{M}(\mathbf{S}_1) \subset \mathcal{M}(\mathbf{S}_2)$ or $\mathcal{M}(\mathbf{S}_2) \subset \mathcal{M}(\mathbf{S}_1)$, and equals to 1 if $\mathcal{M}(\mathbf{S}_1)$ and $\mathcal{M}(\mathbf{S}_2)$ are orthogonal. For the special case when $q_1 = q_2 = q$, the two spaces $\mathbf{S}_1$ and $\mathbf{S}_2$ have the same dimension, and the distance measure (12) reduces to

$$\mathcal{D}\{\mathcal{M}(\mathbf{S}_1),\, \mathcal{M}(\mathbf{S}_2)\} = \left\{1 - \frac{\operatorname{tr}(\mathbf{O}_1\mathbf{O}_1'\mathbf{O}_2\mathbf{O}_2')}{q}\right\}^{1/2}, \tag{13}$$

which was used in Chang et al. (2015) and Liu and Chen (2016). Since the number of factors is usually unknown in practice and may be estimated in a nonperfect way, we shall in the current paper use the generalized version in (12) to measure the distance between two linear spaces.

The following regularity conditions are also needed for theoretical properties.

Condition 1. Let $\mathcal{F}_\ell^j$ be the σ-field generated by $\{(\mathbf{x}_{t,1}, \mathbf{x}_{t,2}) : \ell \leqslant t \leqslant j\}$. The latent process $\{\mathbf{x}_{t,1}, \mathbf{x}_{t,2}\}$ is α-mixing with mixing coefficients satisfying

$$\sum_{t=1}^{\infty} \alpha(t)^{1-2/\zeta} < \infty,$$

for some $\zeta > 2$, where $\alpha(t) = \sup_j \sup_{A \in \mathcal{F}_{-\infty}^j, B \in \mathcal{F}_{j+t}^{\infty}} |P(A \cap B) - P(A)P(B)|$.

Condition 2. For any $i = 1, 2$, $j = 1, \ldots, k_i$, and $t = 1, \ldots, n$, $E(|x_{t,i,j}|^{4\zeta}) < \sigma_x^{4\zeta}$, where $x_{t,i,j}$ is the j-th element of $\mathbf{x}_{t,i}$, $\sigma_x > 0$ is a constant, and ζ is given in Condition 1.

Condition 3. $\{\boldsymbol{\varepsilon}_t\}$ is an independent noise process with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_t$ at time t. $\{\boldsymbol{\varepsilon}_t\}$ and $\{\mathbf{x}_{t,1}, \mathbf{x}_{t,2}\}$ are uncorrelated given $\mathcal{F}_{-\infty}^{t-1}$. Each element of $\boldsymbol{\Sigma}_t$ remains bounded by a positive constant σ_ε^2 as p increases to infinity.

Instead of making specific assumptions on the dynamics of the factor process as in Peña and Box (1987) and Forni et al. (2000), here we consider a general setting where the factor process only needs to satisfy the mixing condition with bounded moments (Chang et al., 2015). Compared with the method proposed in Barigozzi et al. (2018) which is designed to detect changes in the second-order structure of the observed data, our approach does not require the factor and noise processes to be 'close' to stationary processes, and we allow heteroskedasticity in $\mathbf{x}_{t,1}$, $\mathbf{x}_{t,2}$ and $\boldsymbol{\varepsilon}_t$ not only through their cross-sectional dimension but also the time dimension; see our simulation results in Section 5.2. Fan et al. (2013) defines the mixing coefficients for a strictly stationary process in a factor model by

$$\alpha(t) = \sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_t^{\infty}} |P(A \cap B) - P(A)P(B)|,$$

and we shall here use its generalized version for the non-stationary setting as defined in Condition 1; see also Chang et al. (2015). Condition 3 assumes that the noise process is serially independent, but may have strong cross-sectional dependence.

Condition 4. For $i = 1, 2$, there exists a constant $\delta_i \in [0, 1]$ such that $\|\mathbf{A}_i\|_2^2 \asymp \|\mathbf{A}_i\|_{\min}^2 \asymp p^{1-\delta_i}$, as p goes to infinity.

Condition 5. $\gamma_0 \in (\eta_1, \eta_2)$. For any $\gamma \in [\eta_1, \eta_2]$, there exists an integer $h_i \in [1, h_0]$ such that $\boldsymbol{\Sigma}_{x,i}(h_i, \gamma)$ is full rank and $\|\boldsymbol{\Sigma}_{x,i}(h_i, \gamma)\|_{\min}$ is uniformly bounded above 0, for $i = 1, 2$.

Condition 6. $\mathbf{M}_i(\gamma)$ admits k_i distinct positive eigenvalues, for $\gamma \in [\eta_1, \eta_2]$, $i = 1, 2$.

Condition 4 defines the factor strength before and after the change point. Condition 5 ensures that $\mathbf{M}_i(\gamma)$ is full rank and contains information from all components in the factor process. Condition 6 assumes that the nonzero eigenvalues of $\mathbf{M}_i(\gamma)$ are distinct from each other. Condition

7 and Condition 8 shown below make two linear spaces before and after the change point are differentiable as n and p go to infinity.

For $0 \leqslant c_1 < c_2 \leqslant 1$, define

$$N(c_1, c_2) = \lfloor c_1 n \rfloor - \lfloor c_2 n \rfloor, \quad \mathbf{x}_t = \sum_{i=1}^{2} \mathbf{x}_{t,i} I_{t,i}(\gamma_0),$$

and three intervals

$$I_1(h) = [0, \gamma_0 - h/n], \quad I_2(h) = (\gamma_0 - h/n, \gamma_0], \quad I_3(h) = (\gamma_0, 1]. \tag{14}$$

For any $0 \leqslant c_1 < c_2 \leqslant 1$ and both c_1 and c_2 are from the same interval, I_1, I_2 or I_3, let

$$\mathbf{\Gamma}_x(h, c_1, c_2) = \frac{\sum_{t=1}^{n} \mathrm{E}[\mathbf{x}_t \mathbf{x}'_{t+h} I_{\{\lfloor c_1 n \rfloor < t \leqslant \lfloor c_2 n \rfloor\}}]}{N(c_1, c_2)}.$$

Condition 7. For any $\gamma \in (\eta_1, \gamma_0)$, there exists an integer $h_1^* \in [1, h_0]$ such that $\mathbf{\Gamma}_x(h_1^*, \gamma, \gamma_0 - h_1^*/n)$ is full rank. For any $\gamma \in (\gamma_0, \eta_2)$, there exists an integer $h_2^* \in [1, h_0]$ such that $\mathbf{\Gamma}_x(h_2^*, \gamma_0, \gamma)$ is full rank. The minimum singular values of these two matrices mentioned are uniformly bounded above $u_0 > 0$.

Condition 8. There exists a positive constant d such that $\mathcal{D}[\mathcal{M}(\mathbf{Q}_1), \mathcal{M}(\mathbf{Q}_2)] > d$ as n and p go to infinity.

Theorem 1 provides the explicit bound for the proposed projection-based change point estimator, from which we can see that the convergence rate depends on an interplay between the dimensionality of the observed time series and the strength of the factor loading.

Theorem 1. *Assume Conditions 1–8. If $p^{\delta_{\max}} n^{-1/2} = o(1)$, then for any $\epsilon > 0$, with true k_1 and k_2 we have*

$$P(\widehat{\gamma} < \gamma_0 - \epsilon) \leqslant \frac{Cp^{\delta_1}}{\epsilon n^{1/2}}, \quad P(\widehat{\gamma} > \gamma_0 + \epsilon) \leqslant \frac{Cp^{\delta_2}}{\epsilon n^{1/2}},$$

as $n, p \to \infty$, where $\delta_{\max} = \max\{\delta_1, \delta_2\}$.

For one-regime factor models where the loading space remains the same over time, $p^{\delta} n^{-1/2} = o(1)$ is a quite standard condition to obtain the consistency for the estimation of the loading space; see for example Lam et al. (2011), Lam and Yao (2012), and Chang et al. (2015). When a change point exists, in order to estimate the loadings spaces consistently, we need to assume that this condition is satisfied in both regimes, namely $p^{\delta_{\max}} n^{-1/2} = o(1)$. If the factors are strong in both regimes with $\delta_1 = \delta_2 = 0$, the condition is reduced to $n^{-1/2} = o(1)$ which is automatically satisfied when $n \to \infty$. On the other hand, if the factors are weak with $\delta_1 = \delta_2 < 0.5$, then the condition can be satisfied even when the dimension p grows as fast as n.

By Theorem 1, the proposed estimator $\widehat{\gamma}$ in (10) for the change-point location is consistent under mild conditions. It also reveals that the estimation performance can depend critically on the strength of factors in both regimes. In particular, if the factors are strong in both regimes ($\delta_1 = \delta_2 = 0$), then the estimation is immune to the curse of dimensionality. On the other hand, if factors are weak in one regime, then the resulting estimator can become less efficient as p increases. When factors have different levels of strengths before and after the break, the probability that the $\widehat{\gamma}$ falls in the weaker regime is larger but the estimation precision in the stronger regime is better. As a result, the overall rate of convergence of $\widehat{\gamma}$ depends on the strength of the weaker regime.

Theorem 2 provides the asymptotic property of the estimated loading spaces when the estimated break date is used.

Theorem 2. *Assume Conditions 1–8. If $p^{\delta_{\max}} n^{-1/2} = o(1)$, then as $n, p \to \infty$, with true k_1 and k_2, we have*

$$\mathcal{D}\{\mathcal{M}[\widehat{\mathbf{Q}}_i(\widehat{\gamma})],\, \mathcal{M}(\mathbf{Q}_i)\} = O_p(p^{\delta_i} n^{-1/2})$$

for $i = 1, 2$.

By Theorem 2, if $\delta_1 = \delta_2 = 0$, the estimator $\mathcal{M}[\widehat{\mathbf{Q}}_i(\widehat{\gamma})]$ converges to $\mathcal{M}(\mathbf{Q}_i)$ at the rate of $n^{-1/2}$, and thus the curse of dimensionality does not exist. If the factors in regime i are weak, however, the convergence rate is slower and the noise process distorts the information on the latent factor; see for example Lam et al. (2011). By Theorem 2, the convergence rate of the associated loading space estimators is the same as that in factors models without breaks (Lam et al., 2011). Compared with the results in Liu and Chen (2020) which used a threshold variable to split the data, the estimation in weak regime does not gain efficiency from data in strong regime because asymptotically there is no interaction between regimes.

In the following, we will show that when the numbers of factors in two regimes are overestimated, our proposed method can estimate the break date and loading spaces as well.

Condition 9. When $\widehat{k}_i > k_i$, there exists a positive constant $\tilde{d}$ such that $\mathcal{D}[\mathcal{M}(\mathbf{Q}_1), \mathcal{M}(\mathbf{Q}_2^*)] > \tilde{d}$ and $\mathcal{D}[\mathcal{M}(\mathbf{Q}_1^*), \mathcal{M}(\mathbf{Q}_2)] > \tilde{d}$, for any $p \times (\widehat{k}_i - k_i)$ matrix $\mathbf{S}_i$ such that $\dim(\mathcal{M}(\mathbf{S}_i) \cap \mathcal{M}(\mathbf{Q}_i)) = 0$ and $i = 1, 2$, where $\mathbf{Q}_i^* = (\mathbf{Q}_i, \mathbf{S}_i)$ is a $p \times \widehat{k}_i$ matrix.

Condition 9 guarantees that the two augmented linear spaces before and after the change point are still differentiable.

Theorem 3. *Assume Conditions 1–9. If* $p^{\delta_{\max}} n^{-1/2} = o(1)$, $\widehat{k}_1 \geqslant k_1$, *and* $\widehat{k}_2 \geqslant k_2$, *then for* $\epsilon > 0$, *we have*

$$P(\widetilde{\gamma} < \gamma_0 - \epsilon) \leqslant \frac{Cp^{\delta_1}}{\epsilon n^{1/2}}, \quad P(\widetilde{\gamma} > \gamma_0 + \epsilon) \leqslant \frac{Cp^{\delta_2}}{\epsilon n^{1/2}},$$

as $n, p \to \infty$.

Theorem 4 will show that the space spanned by the first k_i columns of $\widetilde{\mathbf{Q}}_i(\widetilde{r}, \widehat{k}_i)$ provides an estimate of $\mathcal{M}(\mathbf{Q}_i)$, and it converges as fast as $\mathcal{M}(\widehat{\mathbf{Q}}_i(\widehat{r}))$ in Theorem 2, for $i = 1, 2$. Define $\widetilde{\mathbf{Q}}_i(\widetilde{r})$ which consists of the first k_i column of $\widetilde{\mathbf{Q}}_i(\widetilde{r}, \widehat{k}_i)$,

$$\widetilde{\mathbf{Q}}_i(\widetilde{r}) = (\widehat{\mathbf{q}}_{i,1}(\widetilde{r}), \ldots, \widehat{\mathbf{q}}_{i,k_i}(\widetilde{r})), \text{ for } i = 1, 2.$$

Theorem 4. *Assume Conditions 1–9. If* $p^{\delta_{\max}} n^{-1/2} = o(1)$, $\widehat{k}_1 \geqslant k_1$, *and* $\widehat{k}_2 \geqslant k_2$, *then as* n, $p \to \infty$, *we have*

$$\mathcal{D}\{\mathcal{M}[\widetilde{\mathbf{Q}}_i(\widetilde{\gamma})], \mathcal{M}(\mathbf{Q}_i)\} = O_p(p^{\delta_i} n^{-1/2})$$

for $i = 1, 2$.

By the definition of $\mathcal{D}(\cdot, \cdot)$ in (12), we can simply obtain the follow results.

Corollary 1. *Assume Conditions 1–9. If* $p^{\delta_{\max}} n^{-1/2} = o(1)$, $\widehat{k}_1 \geqslant k_1$, *and* $\widehat{k}_2 \geqslant k_2$, *then as* n, $p \to \infty$, *we have*

$$\mathcal{D}\{\mathcal{M}[\widetilde{\mathbf{Q}}_i(\widetilde{\gamma}, \widehat{k}_i)], \mathcal{M}(\mathbf{Q}_i)\} = O_p(p^{\delta_i} n^{-1/2})$$

for $i = 1, 2$.

It can be seen from Theorem 3 and Corollary 1 that when the numbers of factors are overestimated, our estimators for break date and the loading spaces are still consistent. Their asymptotic properties are the same with those when k_1 and k_2 are correctly estimated.

Proposition 1 shown below, similar to the results in Lam and Yao (2012), proves that the ratios of estimators for nonzero eigenvalues of $\mathbf{M}_i(\eta_i)$ converge at different rates as n and p grow.

Proposition 1. *Assume Conditions 1–8. If* $p^{\delta_{\max}} n^{-1/2} = o(1)$, *then as* n, $p \to \infty$, *we have*

$$\begin{aligned} &\widehat{\lambda}_{i,k+1}(\eta_i)/\widehat{\lambda}_{i,k}(\eta_i) \asymp 1, && \text{for } k = 1, \ldots, k_i - 1, \\ &\widehat{\lambda}_{i,k_i+1}(\eta_i)/\widehat{\lambda}_{i,k_i}(\eta_i) = O_p(p^{2\delta_i} n^{-1}) \xrightarrow{p} 0, \end{aligned}$$

for $i = 1, 2$.

Proposition 1 indicates that the plot of the estimated eigenvalue ratio will drop sharply at $k = k_i$, which provides a partial theoretical underpinning for the estimator of k_i for $i = 1, 2$; see also Lam and Yao (2012). When $k > k_i$, the eigenvalue $\lambda_{i,k}$ is theoretically zero and thus the property of the ratio $\widehat{\lambda}_{i,k+1}(\eta_i)/\widehat{\lambda}_{i,k}(\eta_i)$ can be difficult to investigate. According to Lam and Yao (2012), although the consistency of (11) cannot be confirmed theoretically, the estimator performs well in numerical experiments (Chang et al., 2015; Liu and Chen, 2016; Wang et al., 2019; Liu and Chen, 2020).

4 Multiple Change-Point Detection

We shall in this section consider the situation with multiple change points and extend our results in Sections 2 and 3 to propose an algorithm for estimating the change-point locations. For this, we consider the factor model with multiple change points:

$$\mathbf{y}_t = \begin{cases} \mathbf{A}_1\mathbf{x}_{t,1} + \boldsymbol{\varepsilon}_t, & \text{if } 0 \leqslant t/n \leqslant \gamma_0; \\ \mathbf{A}_2\mathbf{x}_{t,2} + \boldsymbol{\varepsilon}_t, & \text{if } \gamma_0 < t/n \leqslant \gamma_1; \\ \cdots & \\ \mathbf{A}_{m+1}\mathbf{x}_{t,m+1} + \boldsymbol{\varepsilon}_t, & \text{if } \gamma_{m-1} < t/n \leqslant \gamma_m = 1, \end{cases} \tag{15}$$

where $\mathbf{A}_i \in \mathbb{R}^{p\times k_i}$ for $i = 1, \ldots, m+1$, and $\mathcal{D}\{\mathcal{M}(\mathbf{A}_i), \mathcal{M}(\mathbf{A}_{i+1})\} \neq 0$ for $i = 1, \ldots, m$. The model in (15) has m change points, and the case with $m = 0$ relates to the situation with no change point. To detect change points in (15) and estimate their locations, we propose to exploit the effect of a change point on the estimated number of factors. We shall first use the simple example with $m = 1$ to illustrate the idea. In this case, the loading space contains a change point, and if one ignores the change point and calculate

$$\boldsymbol{\Sigma}_y(h) = \frac{1}{n}\sum_{t=1}^{n-h} \mathrm{E}(\mathbf{y}_t\mathbf{y}_{t+h}'), \quad \mathbf{M} = \sum_{h=1}^{h_0} \boldsymbol{\Sigma}_y(h)\boldsymbol{\Sigma}_y(h)', \tag{16}$$

$$\widehat{\boldsymbol{\Sigma}}_y(h) = \frac{1}{n}\sum_{t=1}^{n-h} \mathbf{y}_t\mathbf{y}_{t+h}', \quad \widehat{\mathbf{M}} = \sum_{h=1}^{h_0} \widehat{\boldsymbol{\Sigma}}_y(h)\widehat{\boldsymbol{\Sigma}}_y(h)',$$

then it will lead to an overestimated number of factors. For this, let $\widehat{\lambda}_k$ be the k-th largest eigenvalue of $\widehat{\mathbf{M}}$, and we need the following conditions.

Condition 10. The nonzero eigenvalues of $\mathbf{M}$ are distinct.

Condition 11. Define $\tilde{k} = \dim(\mathcal{M}(\mathbf{Q}_1) \cap \mathcal{M}(\mathbf{Q}_2))$. $\tilde{k} < \min(k_1, k_2)$ is fixed and $\sigma_{\tilde{k}+1}(\mathbf{Q}_1'\mathbf{Q}_2) = \nu$, where ν is a positive constant such that $\nu < 1$ as n and p go to infinity.

Condition 11 is stricter than Condition 8. When $\tilde{k}=0$, as n and p grow to infinity, Condition 11 requires that $\|\mathbf{Q}_1'\mathbf{Q}_2\|_2 < 1$, while Condition 8 indicates that $\|\mathbf{Q}_1'\mathbf{Q}_2\|_F < 1$. Condition 11 ensures that as n and p increases, the non-overlapped subspaces in $\mathcal{M}(\mathbf{Q}_1)$ and $\mathcal{M}(\mathbf{Q}_2)$ are still well apart.

Corollary 2. *Assume Conditions 1–7, 10 and 11. If $\delta_1=\delta_2=\delta_0$ and $p^{\delta_0}n^{-1/2}=o(1)$, then as n, $p\to\infty$ we have*

$$\begin{aligned}\widehat{\lambda}_{k+1}/\widehat{\lambda}_k &\asymp 1, && \text{for } k=1,\ldots,k_1+k_2-\tilde{k}-1,\\ \widehat{\lambda}_{k_1+k_2-\tilde{k}+1}/\widehat{\lambda}_{k_1+k_2-\tilde{k}} &= O_p(p^{2\delta_0}n^{-1}) \xrightarrow{p} 0.\end{aligned}$$

Corollary 2 implies that the ratio of estimated eigenvalues will drop sharply at $k=k_1+k_2-\tilde{k}$, if we combine data from two regimes when the factor strength level across regimes remains the same. Motivated by this observation and the methods proposed in Ma and Su (2018) and Wu (2021), we divide the whole time span into subintervals to monitor the change of the number of factors, and thus detect the structural breaks in the factor model. To be more specific, let J be a prescribed integer satisfying $n\gg J\gg m$. After dividing $[0,1]$ into J equally-spaced intervals $S_j=[\frac{j-1}{J},\frac{j}{J})$ for $j=1,\ldots,J-1$ and $S_J=[\frac{J-1}{J},1]$, we further assume that there is no break in $[0,\frac{1}{2J})\cup[\frac{2J-1}{2J},1]$ and the distance of any two change points is greater than $2/J$. We fit the data in each subinterval with a factor model without structural breaks and estimate the number of factors in each subinterval. In particular, let

$$\widehat{\mathbf{\Sigma}}_{y,j}^J(h)=\frac{1}{n}\sum_{t=\lfloor\frac{n(j-1)}{J}\rfloor+1}^{\lfloor\frac{nj}{J}\rfloor}\mathbf{y}_t\mathbf{y}_{t+h}',\quad \widehat{\mathbf{M}}_j^J=\sum_{h=1}^{h_0}\widehat{\mathbf{\Sigma}}_{y,j}^J(h)\widehat{\mathbf{\Sigma}}_{y,j}^J(h)',\tag{17}$$

for $j=1,\ldots,J$, then the number of factors in the j-th subinterval can be estimated by

$$\widehat{k}_j^J=\arg\min_{1\leqslant k\leqslant R}\frac{\widehat{\lambda}_{j+1,k}^J}{\widehat{\lambda}_{j,k}^J},\tag{18}$$

where $\widehat{\lambda}_{j,k}^J$ is the k-th largest eigenvalue of $\widehat{\mathbf{M}}_j^J$. When we track the possible changes of the number of factors in these subintervals, three situations that can happen to $\widehat{k}_j^J$ and S_j need to be considered:

(i) when $\widehat{k}_j^J\neq\widehat{k}_{j-1}^J$, $\widehat{k}_{j-1}^J=\widehat{k}_{j-2}^J$, and $\widehat{k}_j^J\neq\widehat{k}_{j+1}^J$, the break happens in the interior of the interval S_j;

(ii) when $\widehat{k}_j^J=\widehat{k}_{j-1}^J$, no break happens in the interior of S_j; or the break happens near the left end of S_j, and the number of factors remains the same after the break;

(iii) when $\widehat{k}_j^J \neq \widehat{k}_{j-1}^J$, $\widehat{k}_{j-1}^J = \widehat{k}_{j-2}^J$, and $\widehat{k}_j^J = \widehat{k}_{j+1}^J$, the break happens near the left end of S_j, and the number of factors changes as well after the break.

The case when $\widehat{k}_j^J \neq \widehat{k}_{j-1}^J$, $\widehat{k}_{j-1}^J \neq \widehat{k}_{j-2}^J$ and $\widehat{k}_j^J = \widehat{k}_{j+1}^J$ is not included here because it shall be considered when discussing $\widehat{k}_{j-1}^J$ and indicates that the change point may happen in the interior of S_{j-1}. The case when $\widehat{k}_j^J \neq \widehat{k}_{j-1}^J$, $\widehat{k}_{j-1}^J \neq \widehat{k}_{j-2}^J$ and $\widehat{k}_j^J \neq \widehat{k}_{j+1}^J$, is not discussed above because there is at most one change point in two consecutive subintervals under our assumptions. For real data analysis, if it happens or the estimated number of factors varies frequently, it may indicate that a larger J should be considered.

For case (i), the complement loading spaces before the break and after the break in (9) can be estimated by with data in the intervals S_{j-1} and S_{j+1} respectively, and then the estimate of the location of the break is obtained by the method in Section 2.2. For cases (ii), to detect the existence of a change point near or in the ends of S_j, we re-divide the interval $[0,1]$ into $J+1$ subintervals $S_1^* = [0, \frac{1}{2J})$, $S_j^* = [\frac{2j+1}{2J}, \frac{2j+3}{2J})$ for $j = 2, \ldots, J-1$ and $S_{J+1}^* = [\frac{2J-1}{2J}, 1]$, and then estimate the number of factors in these subintervals, denoted by $\widehat{k}_1^{J*}, \ldots, \widehat{k}_{J+1}^{J*}$. Note that the midpoints in subintervals $S_1, \ldots, S_J$ consist of the endpoints of subintervals $S_2^*, \ldots, S_{J-1}^*$. There are two situations that can happen to $\widehat{k}_j^{J*}$ in case (ii):

(a) when $\widehat{k}_j^{J*} = \widehat{k}_j^J$, there is no break in the end of S_j;

(b) when $\widehat{k}_j^{J*} \neq \widehat{k}_j^J$, the break happens near the left end of S_j.

For case(b) and case (iii) discussed above, we can estimate the complement loading space before the break with data in the interval S_{j-1}^* and that after the break with data in the interval S_{j+1}^*, and then construct the sample objective function in (9) to estimate the location of the change point.

5 Empirical Illustration

We shall here conduct a Monte Carlo simulation study to examine the finite sample performance of the proposed change-point estimation procedure and multiple change-point detection algorithm. Throughout the simulation, we set $h_0 = 1$ for simplicity, and results for estimation performance of proposed estimators and comparisons among different methods are presented in Sections 5.1 and 5.2 respectively.

5.1 Estimation performance for change-point location and loading space

We first examine the estimation performance for change-point location and loading spaces discussed in Section 2.2 and 2.3. For this, we generate the noise process as a Gaussian process whose covariance matrix has 1 on the diagonal and 0.5 in all off-diagonal entries. Let $k_1 = k_2 = 3$ and the factor process is simulated from 3 independent autoregressive(AR) models of order 1 with AR coefficients 0.9, -0.7, and 0.8 and with innovation standard deviation 2. We generate entries of the loading matrix $\mathbf{A}_i$ as independent sample from the uniform distribution on $[-p^{-\delta_i/2}, p^{-\delta_i/2}]$, then δ_i characterizes the factor strength of $\mathbf{A}_i$. Let the change-point location $\gamma_0 = 0.5$, and we consider four different scenarios on the factor strength, namely SS ($\delta_1 = \delta_2 = 0$) in which strong factors are used both before and after the change point, SW ($\delta_1 = 0$ and $\delta_2 = 0.25$) in which strong factors are used before the change point and weak factors after, WS ($\delta_1 = 0.25$ and $\delta_2 = 0$) in which weak factors are used before the change point and strong factors after, and WW ($\delta_1 = \delta_2 = 0.25$) in which weak factors are used both before and after the change point. Let $\eta_1 = 0.1$ and $\eta_2 = 0.9$. For each setting, we generate 1000 realizations, and examine if the procedure proposed in Section 2.2 can successfully identify the change-point location and the associated loading spaces before and after the change point.

We first consider the case when the numbers of factors are known, and Figure 1 provides the histograms of the proposed change-point location estimator $\widehat{\gamma}$ for different settings when $n = 1000$. It can be seen from Figure 1 that, if the factor strength is weak in at least one regime, before or after the change point, then the estimation efficiency in that regime suffers from the increase in dimension. In contrast, the estimation efficiency in the strong regime does not seem to be affected by the curse of dimensionality. This is in line with the results in Theorem 1; see also the discussions thereafter. In addition, it can be seen from the middle panels in Figure 1 that, when the factor strengths before and after the change point are different, namely settings SW and WS, the estimation bias, though asymptotically negligible, is more likely to be toward the regime with weaker factors. In particular, when the factor strength after the change point is weaker as in the SW setting, then it is more likely to overestimate γ_0. On the other hand, if the factor strength before the change point is weaker as in the WS setting, then it is more likely to underestimate γ_0. We also provide in Table 1 a summary of the estimation error $|\widehat{\gamma} - \gamma_0|$ when $n = 400, 1000$. The estimation errors for the loading spaces are summarized in Table 2, from which we can see that the estimation procedure proposed in Section 2.2 performs reasonably well under all the considered settings.

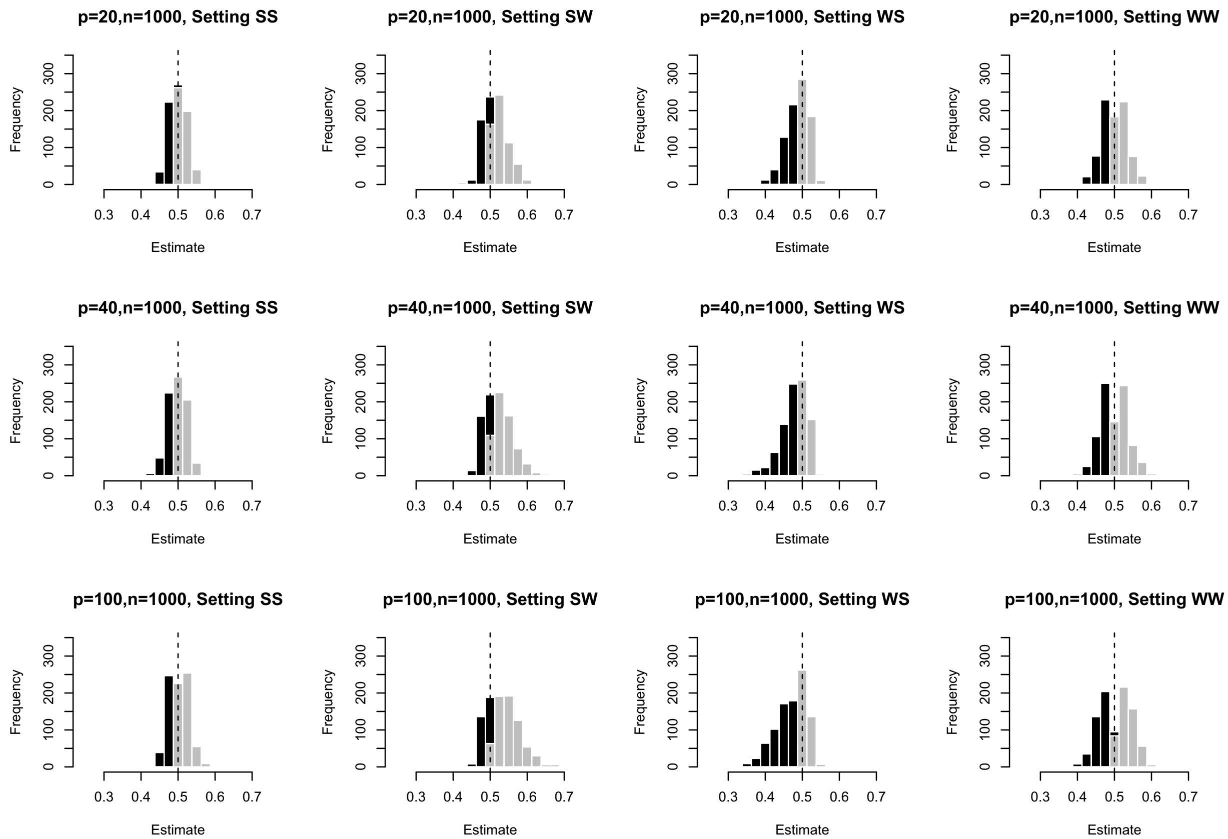

Figure 1: Histograms of estimated change-point location under different settings when $n = 1000$ and k_1 and k_2 are known. The dashed line shows the true change-point location $\gamma_0 = 0.5$, black bars show the frequencies of underestimation, and grey bars show the frequencies of overestimation.

Table 1: Average estimation error $|\widehat{\gamma} - \gamma_0|$ when k_1 and k_2 are known in Section 5.1

n	$n = 400$			$n = 1000$		
p	20	40	100	20	40	100
SS	0.035	0.039	0.040	0.015	0.018	0.018
SW	0.051	0.066	0.083	0.023	0.029	0.039
WS	0.054	0.060	0.081	0.023	0.027	0.038
WW	0.053	0.060	0.071	0.024	0.028	0.034

For the case where the number of factors are unknown, Tables 3 and 4 provide the average

Table 2: Average estimation error $\mathcal{D}\{\mathcal{M}[\widehat{\mathbf{Q}}_i(\widehat{r})], \mathcal{M}(\mathbf{Q}_i)\}$ when k_1 and k_2 are known in Section 5.1

	n	$n = 400$			$n = 1000$		
	p	20	40	100	20	40	100
Setting SS	$\delta_1 = 0$	0.059	0.055	0.053	0.033	0.032	0.031
	$\delta_2 = 0$	0.059	0.057	0.055	0.033	0.032	0.031
Setting SW	$\delta_1 = 0$	0.058	0.056	0.053	0.033	0.032	0.031
	$\delta_2 = 0.25$	0.095	0.099	0.113	0.049	0.053	0.058
Setting WS	$\delta_1 = 0.25$	0.093	0.094	0.110	0.048	0.052	0.058
	$\delta_2 = 0$	0.060	0.056	0.053	0.034	0.032	0.032
Setting WW	$\delta_1 = 0.25$	0.089	0.095	0.103	0.049	0.052	0.057
	$\delta_2 = 0.25$	0.093	0.094	0.105	0.050	0.052	0.057

estimation errors for loading spaces and threshold value when $\widehat{k}_1 = \widehat{k}_2 = 4$. It can be seen that their patterns are similar to those in Table 1 and Table 2, and the proposed method works reasonably well when k_1 and k_2 are overestimated.

Table 3: Average estimation error $|\widehat{\gamma} - \gamma_0|$ when the numbers of factors are unknown and overestimated as 4 in Section 5.1

n	$n = 400$			$n = 1000$		
p	20	40	100	20	40	100
SS	0.029	0.030	0.034	0.012	0.013	0.016
SW	0.041	0.051	0.074	0.019	0.024	0.034
WS	0.041	0.050	0.069	0.018	0.023	0.030
WW	0.042	0.047	0.059	0.019	0.022	0.028

5.2 Performance of multiple change-point detection

In this subsection we investigate the performance of our method when the number of change points is unknown, and compare it with the method introduced in Cho and Fryzlewicz (2015), which is designed to detect multiple change points for high-dimensional time series without a factor structure and requires the observed process to be stationary within each regime.

Table 4: Average estimation error $\mathcal{D}\{\mathcal{M}[\widehat{\mathbf{Q}}_i(\hat{r})], \mathcal{M}(\mathbf{Q}_i)\}$ when the numbers of factors are unknown and overestimated as 4 in Section 5.1

	n	$n=400$			$n=1000$		
	p	20	40	100	20	40	100
Setting SS	$\delta_1 = 0$	0.045	0.045	0.045	0.027	0.027	0.028
	$\delta_2 = 0$	0.045	0.046	0.046	0.027	0.028	0.027
Setting SW	$\delta_1 = 0$	0.047	0.048	0.048	0.028	0.028	0.029
	$\delta_2 = 0.25$	0.073	0.085	0.085	0.040	0.043	0.048
Setting WS	$\delta_1 = 0.25$	0.073	0.084	0.084	0.038	0.043	0.048
	$\delta_2 = 0$	0.047	0.047	0.047	0.027	0.028	0.028
Setting WW	$\delta_1 = 0.25$	0.074	0.083	0.083	0.040	0.044	0.050
	$\delta_2 = 0.25$	0.073	0.083	0.083	0.039	0.044	0.049

Factors are assumed to be strong in all regimes, and the entries of the loading matrices in all settings are generated as independent sample from the uniform distribution on $[-1, 1]$.

Three settings are considered.

1. One single change point at 0.5. $k_1 = 1$ and $k_2 = 2$. The factor process is stationary, follows an AR(1) model with AR coefficient 0.9 before the break, and consists of two independent AR processes with AR coefficients 0.9 and -0.8 after the break. The noise process is stationary and Gaussian whose covariance matrix has 1 on the diagonal and 0.1 in all the the off-diagonal entries.

2. Two change points at 0.33 and 0.6. $k_1 = k_2 = k_3$. The factor process is non-stationary

$$x_t = -0.1t/n + 0.9x_{t-1} + e_t, \text{ for } t = 1, \dots, n,$$

where $e_t \sim N(0, 3)$, and $\{\boldsymbol{\varepsilon}_t\}$ is an independent Gaussian process, whose covariance matrix at time t has $0.9 + 0.5\sin(2\pi t/n)$ on the diagonal and 0.1 in all the off-diagonal entries.

3. No change points. There is only one factor which is an AR(1) process with AR coefficient 0.9. $\{\boldsymbol{\varepsilon}_t\}$ is an independent Gaussian process, whose covariance matrix at time t has $2 - 4t/n + 4t^2/n^2$ on the diagonal and 0.2 in all the off-diagonal entries.

Set $p = 50, 100, 200$ and $n = 500, 1000$. When $n = 500$, $J = 10$. When $n = 1000$, $J = 15$. We run 1000 replications. Following Ma and Su (2018), we evaluate the performance of the multiple change-point procedure with the relative frequency of correct estimation of the number of breaks shown in Table 5, and conditional on the correct estimation of m, the accuracy of change-point estimation, which is measured by Hausdorff distance of the estimated and true locations of change points. Let $\mathcal{D}(A,B) = \sup_{b\in B}\inf_{a\in A}|a-b|$ for any two sets A and B, then the Hausdorff distance between A and B is defined as $\max\{\mathcal{D}(A,B),\mathcal{D}(B,A)\}$. Table 6 shows the mean and standard deviation of the Hausdorff distance of estimates and true parameters, and Figure 2 plots the histogram of the estimated locations of change points conditional on the correct estimation of m.

Table 5: Relative frequency of correct detection of the number of breaks in Section 5.2

		Setting 1		Setting 2		Setting 3	
n	p	CF	Our method	CF	Our method	CF	Our method
500	50	0.970	0.962	0.617	0.840	0.779	0.994
	100	0.962	0.947	0.823	0.866	0.590	0.991
	200	0.956	0.966	0.890	0.866	0.304	0.993
1000	50	0.961	0.974	0.693	0.915	0.199	0.998
	100	0.938	0.976	0.645	0.934	0.039	1.000
	200	0.914	0.976	0.559	0.922	0.004	0.997

Note: 'CF' denotes the change-point detection algorithm proposed by Cho and Fryzlewicz (2015).

From Table 5, Table 6, and Figure 2 we can see that under setting 1, when sample size is 500, the method by Cho and Fryzlewicz (2015) identifies the breaks slightly more frequently than ours but estimates the break locations less accurate than ours; when sample size increases to 1000, our method performs better in both break date identification and estimation. When analyzing data which is not stationary between change points under setting 2 and setting 3, our method successfully detects all the breaks with a much higher frequency and estimate the break locations much more precise than the algorithm introduced in Cho and Fryzlewicz (2015).

6 Real Data Analysis

We applied our method to the Stock-Watson data (Stock and Watson, 1998, 2005), containing 132 U.S. monthly economic indicators from March 1960 to December 2003, with $n = 526$ and $p = 132$. The data include real output and income, employment, real retail, manufacturing and

Table 6: Mean and standard deviation (in the parentheses) of Hausdorff distance between estimated and true change-point locations conditional on the correct estimation of m in Section 5.2

		Setting 1		Setting 2	
n	p	CF	Our method	CF	Our method
500	50	1.556(2.828)	0.921(0.921)	4.461(3.839)	0.810(1.042)
	100	1.557(2.791)	0.869(0.831)	4.132(3.777)	0.822(0.994)
	200	1.180(2.101)	0.995(0.959)	3.692(3.440)	0.910(1.031)
1000	50	0.685(1.236)	0.469(0.478)	3.662(3.060)	0.455(0.597)
	100	0.646(1.170)	0.521(0.514)	2.634(2.253)	0.527(0.641)
	200	0.633(1.043)	0.521(0.514)	2.277(1.964)	0.565(0.602)

Note: For ease of presentation, all values in this table are multiplied by 100.

trade sales, consumption, interest rates, price index and other economic indicators. Stock and Watson (2005) provided more detailed information about this data set and transformations needed before analysis.

Set $h_0 = 1$ and $J = 12$. Figure 3 plots the estimated number of factors in the subintervals, and it indicates that there might be two change points. One happens in S^*_{10} and the number of factors remains at 1 after the break, and the other one happens in S^*_{12} and the number of factors increases to 2 after the change point. Using methods described in Section 4, we obtain the estimates for two change points, 0.735 and 0.920, and $\widehat{k}_1 = \widehat{k}_2 = 1$ and $\widehat{k}_3 = 2$.

It implies that the dynamics of economic indicators experienced permanent structural changes around May 1992 possibly due to the economic downturn in early 1990s and around June 2000 because of the dot-com bubble.

7 Conclusion

Although factor models have been frequently used in the study of high-dimensional time series, existing results were mostly developed under the framework of Chamberlain and Rothschild (1983) and Bai (2003). Such a factor modeling framework, however, typically requires the noise process to be idiosyncratic and as a result does not allow the existence of strong cross-sectional dependence.

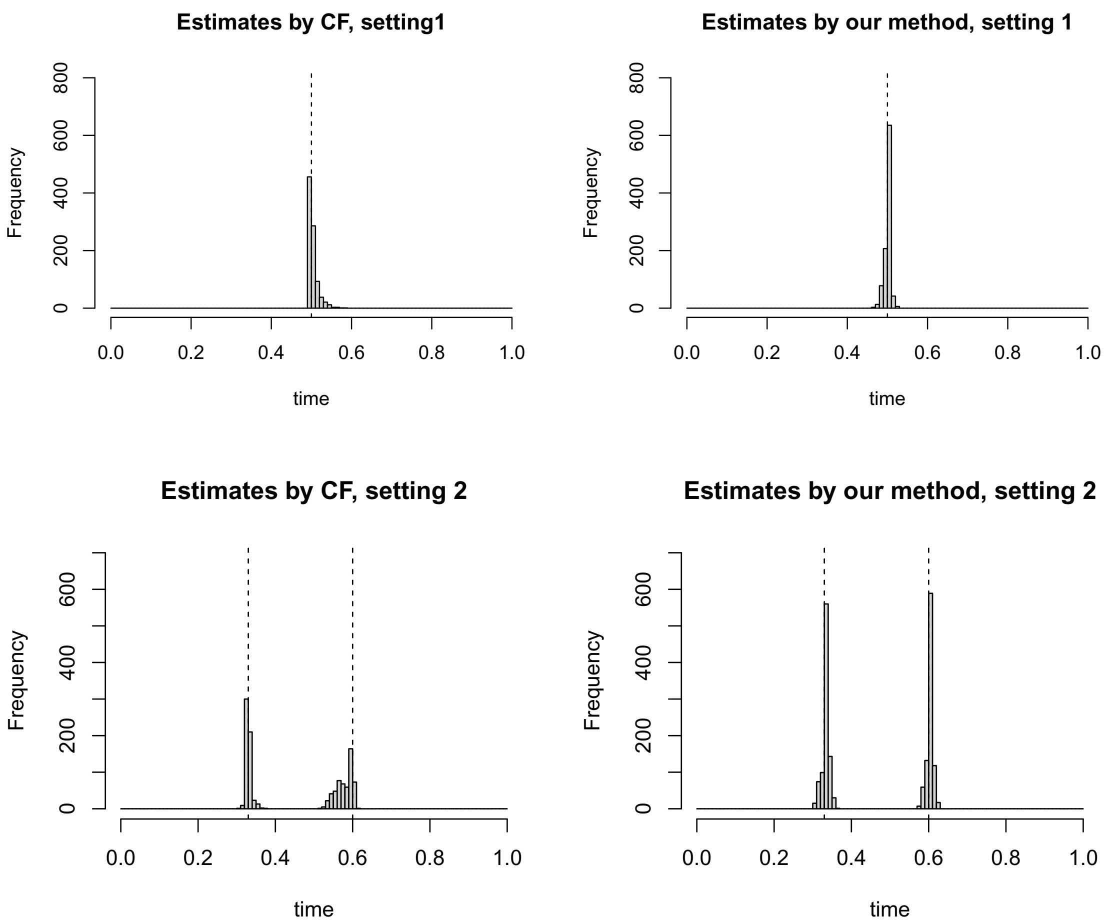

Figure 2: Histogram of estimated locations of change points by the method in Cho and Fryzlewicz (2015) and ours conditional on the correct estimation of m under setting 1 and setting 2 when $n = 1000$ and $p = 200$ in Section 5.2. Dashed lines show the true locations of change points.

In addition, it may suffer from certain identifiability issues as discussed in Lam et al. (2011). To address these, Lam et al. (2011) in their influential paper proposed a new framework for factor analysis of high-dimensional time series. The major goal of the current paper is to consider the recently proposed factor model of Lam et al. (2011) in the change-point setting, and develop consistent estimators for the change-point locations and the associated factor loading spaces. Asymptotic properties of the proposed estimators have been carefully studied in Section 3, from which we can see that the convergence rates depend on an interplay between the dimension of the observed time series and the strength of the factor loading. Furthermore, we show that the proposed estimators are still consistent when the numbers of factors are overestimated. The

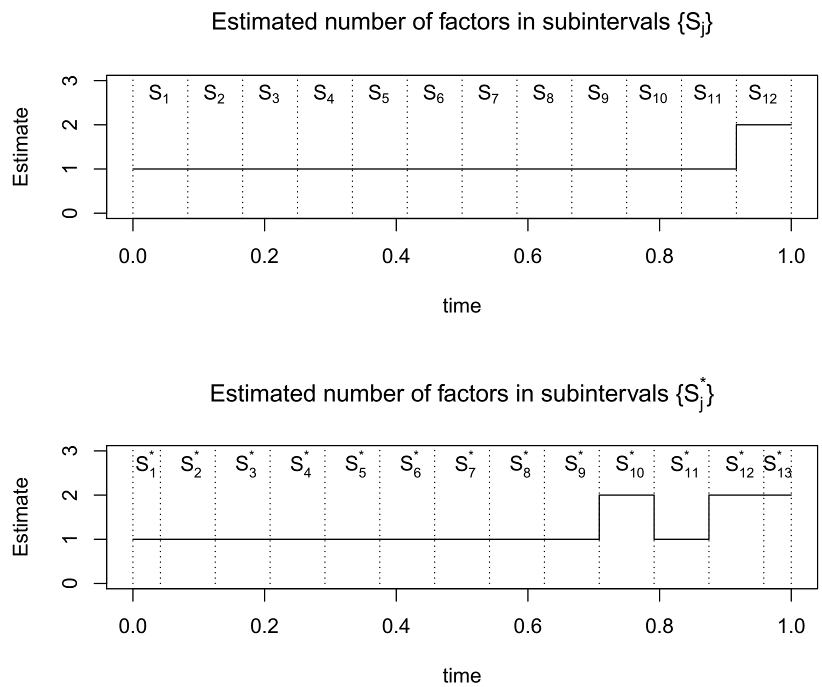

Figure 3: Estimated numbers of factors in subintervals $\{S_j\}$ and $\{S_j^*\}$ for real data analysis in Section 6.

algorithm for multiple change-point detection is also proposed and discussed. Compared with existing results on change point estimation of factor models for high-dimensional time series, a distinguished feature of the current paper is the allowance of strong cross-sectional dependence. Another advantage of the proposed algorithm is that we exclusively focus on the changes of the factor loading space and can handle the situation when the factor or noise process is non-stationary over the sampling period while most existing multiple change-point detection approaches for high-dimensional time series require the observed process to be piecewise stationary or 'close' to a piecewise stationary process. In particular, it can be seen from the simulation results in Section 5.2 that our algorithm performs well in both change-point identification and estimation when the dynamics of factor and noise processes vary along with time.

Acknowledgements

The authors wish to thank the associate editor and anonymous referee for their constructive and insightful comments, which helped us substantially improve the paper.

Appendix: Proofs and Lemmas

In this section, we mainly focus on the mathematical proofs for before the break and when $\epsilon > 0$. The results for after the break or $\epsilon < 0$ are included, but most of proofs are omitted since they are quite similar. For any fixed $\epsilon \neq 0$, there exists a positive integer N such that when $n \geqslant N$, $|\epsilon| > (h+1)/n$, therefore, for Lemmas 2-7, we only consider when $|\epsilon| > (|h|+1)/n$. In addition, the model is not distinguishable for all values in $[k/n, (k+1)/n)$ as the break point where $k \in \mathbb{Z}^+$, so for simplicity we treat ϵn as an integer in the proofs. We use Cs to denote generic uniformly positive constants which only depend on the parameters.

Lemma 1. *For $0 \leqslant c_1 < c_2 \leqslant 1$, and c_1 and c_2 are from the same one of the three intervals, I_1, I_2 or I_3 defined in (14), let*

$$\widehat{\mathbf{\Gamma}}_x(h, c_1, c_2) = \frac{\sum_{t=1}^n \mathbf{x}_t \mathbf{x}'_{t+h} I_{\{\lfloor c_1 n\rfloor < t \leqslant \lfloor c_2 n \rfloor\}}}{N(c_1, c_2)}.$$

Under Conditions 1 and 2, for any $h \in [1, h_0]$, it holds that

$$\|\mathbf{\Gamma}_x(h, c_1, c_2)\|_2^2 \leqslant k_{\max}^2 \sigma_x^4,$$

$$\mathrm{E}\left(\|\widehat{\mathbf{\Gamma}}_x(h, c_1, c_2) - \mathbf{\Gamma}_x(h, c_1, c_2)\|_2^2\right) \leqslant \frac{(3h + 8\alpha) k_{\max}^2 \sigma_x^4}{N(c_1, c_2)},$$

where $\alpha = \sum_{t=1}^{\infty} \alpha(t)^{1-2/\zeta}$, and $k_{\max} = \max\{k_1, k_2\}$.

Proof: Let $a_{q,\ell}$ and $\widehat{a}_{q,\ell}$ be the (q,ℓ)-th entry in $\mathbf{\Gamma}_x(h, c_1, c_2)$ and $\widehat{\mathbf{\Gamma}}_x(h, c_1, c_2)$ respectively. By Condition 2 and Jensen's inequality we know that $\mathrm{E}(x_{t,i,j}^2) < \sigma_x^2$ and $\mathrm{E}(x_{t,i,j}^4) < \sigma_x^4$, for $i = 1, 2$, $j = 1, \ldots, k_i$, and $t = 1, \ldots, n$. Let $x_{t,q}$ be the q-th entry in $\mathbf{x}_t$. We have $\mathrm{E}(x_{t,j}^2) < \sigma_x^2$ and $\mathrm{E}(x_{t,j}^4) < \sigma_x^4$, for $j = 1, \ldots, k_1$ when $t \leqslant \gamma_0 n$, for $j = 1, \ldots, k_2$ when $t > \gamma_0 n$. By Cauchy-Schwarts inequality,

$$|a_{q,\ell}|^2 = \left| \frac{1}{N(c_1, c_2)} \sum_{t=\lfloor c_1 n \rfloor + 1}^{\lfloor c_2 n \rfloor} \mathrm{E}(x_{t,q} x_{t+h,\ell}) \right|^2 \leqslant \left| \frac{1}{N(c_1, c_2)} \sum_{t=\lfloor c_1 n \rfloor + 1}^{\lfloor c_2 n \rfloor} \sqrt{\mathrm{E}(x_{t,q}^2) \mathrm{E}(x_{t+h,\ell}^2)} \right|^2 = \sigma_x^4.$$

It follows that $\|\mathbf{\Gamma}_x(h, c_1, c_2)\|_2^2 \leqslant \|\mathbf{\Gamma}_x(h, c_1, c_2)\|_F^2 \leqslant k_{\max}^2 \sigma_x^4$.

By Proposition 2.5 in Fan and Yao (2003), we have

$$
\begin{aligned}
\mathrm{E}(\widehat{a}_{q,\ell}-a_{q,\ell})^2 &= \frac{1}{N(c_1,c_2)^2}\mathrm{E}\left|\sum_{t=1}^{n}[x_{t,q}x_{t+h,\ell}-E(x_{t,q}x_{t+h,\ell})]I_{\{\lfloor c_1 n\rfloor<t\leqslant\lfloor c_2 n\rfloor\}}\right|^2\\
&= \frac{1}{(N(c_1,c_2)^2}\sum_{\substack{|t_1-t_2|\leqslant h\\ \lfloor c_1 n\rfloor<t_1,t_2\leqslant\lfloor c_2 n\rfloor}}\mathrm{E}[x_{t_1,q}x_{t_1+h,\ell}-E(x_{t_1,q}x_{t_1+h,\ell})][x_{t_2,q}x_{t_2+h,\ell}-E(x_{t_2,q}x_{t_2+h,\ell})]\\
&\quad+\frac{1}{N(c_1,c_2)^2}\sum_{\substack{|t_1-t_2|>h\\ \lfloor c_1 n\rfloor<t_1,t_2\leqslant\lfloor c_2 n\rfloor}}\mathrm{E}[x_{t_1,q}x_{t_1+h,\ell}-E(x_{t_1,q}x_{t_1+h,\ell})][x_{t_2,q}x_{t_2+h,\ell}-E(x_{t_2,q}x_{t_2+h,\ell})]\\
&\leqslant \frac{[(2h+1)N(c_1,c_2)-h^2-h]\sigma_x^4}{N(c_1,c_2)^2}+\frac{(N(c_1,c_2)-h)\sigma_x^4}{N(c_1,c_2)^2}\sum_{u=1}^{N(c_1,c_2)-2h-1}\alpha(u)^{1-2/\zeta}\\
&\leqslant \frac{3hN(c_1,c_2)\sigma_x^4}{N(c_1,c_2)^2}+\frac{(N(c_1,c_2)-h)\alpha\sigma_x^4}{N(c_1,c_2)^2}<\frac{(3h+8\alpha)\sigma_x^4}{N(c_1,c_2)}.
\end{aligned}
$$

Thus, $\mathrm{E}\|\widehat{\boldsymbol{\Gamma}}_x(h,c_1,c_2)-\boldsymbol{\Gamma}_x(h,c_1,c_2)\|_2^2\leqslant\mathrm{E}\|\widehat{\boldsymbol{\Gamma}}_x(h,c_1,c_2)-\boldsymbol{\Gamma}_x(h,c_1,c_2)\|_F^2\leqslant(3h+8\alpha)k_{\max}^2\sigma_x^4/N(c_1,c_2)$.
■

Lemma 2. *Under Conditions 1-4 and 6, for $\epsilon\in(-\gamma_0,1-\gamma_0)$ and $|\epsilon|>(h+1)/n$, it holds that*

$$
\mathrm{E}\left(\|\widehat{\boldsymbol{\Sigma}}_{y,i}(h,\gamma_0+\epsilon)-\boldsymbol{\Sigma}_{y,i}(h,\gamma_0+\epsilon)\|_2^2\right)\leqslant 144(3h+8\alpha)a_1^4k_{\max}^2\nu^4p^2n^{-1},
$$

where $\nu=\max\{\sigma_x,\sigma_\varepsilon,1\}$, and $a_1>1$ satisfies $\|\mathbf{A}_i\|_2\leqslant a_1p^{1/2-\delta_i/2}$, for i=1,2.

Proof: When $\epsilon>0$,

$$
\begin{aligned}
&\widehat{\boldsymbol{\Sigma}}_{y,1}(h,\gamma_0+\epsilon)-\boldsymbol{\Sigma}_{y,1}(h,\gamma_0+\epsilon)\\
&= \Big[(\gamma_0-\frac{h}{n})\mathbf{A}_1\left(\widehat{\boldsymbol{\Gamma}}_x(h,0,\gamma_0-\frac{h}{n})-\boldsymbol{\Gamma}_x(h,0,\gamma_0-\frac{h}{n})\right)\mathbf{A}_1'\\
&\quad+\frac{h}{n}\mathbf{A}_1\left(\widehat{\boldsymbol{\Gamma}}_x(h,\gamma_0-\frac{h}{n},\gamma_0)-\boldsymbol{\Gamma}_x(h,\gamma_0-\frac{h}{n},\gamma_0)\right)\mathbf{A}_2'\\
&\quad+(\epsilon-\frac{h}{n})\mathbf{A}_2\left(\widehat{\boldsymbol{\Gamma}}_x(h,\gamma_0,\gamma_0+\epsilon-\frac{h}{n})-\boldsymbol{\Gamma}_x(h,\gamma_0,\gamma_0+\epsilon-\frac{h}{n})\right)\mathbf{A}_2'\Big]\\
&\quad+\frac{1}{n}\sum_{t=1}^{r_0-h}(\mathbf{A}_1\mathbf{x}_{t,1}\boldsymbol{\varepsilon}_{t+h}'+\boldsymbol{\varepsilon}_t\mathbf{x}_{t+h,1}'\mathbf{A}_1'+\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t+h}')+\frac{1}{n}\sum_{t=r_0-h+1}^{r_0}(\mathbf{A}_1\mathbf{x}_{t,1}\boldsymbol{\varepsilon}_{t+h}'+\boldsymbol{\varepsilon}_t\mathbf{x}_{t+h,2}'\mathbf{A}_2'+\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t+h}')\\
&\quad+\frac{1}{n}\sum_{t=r_0+1}^{r_0+\lfloor\epsilon n\rfloor-h}(\mathbf{A}_2\mathbf{x}_{t,2}\boldsymbol{\varepsilon}_{t+h}'+\boldsymbol{\varepsilon}_t\mathbf{x}_{t+h,2}'\mathbf{A}_2'+\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t+h}')\\
&= I_1+I_2+I_3+I_4.
\end{aligned}
$$

Condition 4 implies that there exists a positive constant a_0 such that $\|\mathbf{A}_i\|_2\leqslant a_0p^{1/2-\delta_i/2}$ for

$i=1,2$. Let $a_1=\max\{a_0,1\}$. By Lemma 1 and Condition 3, we have

$$\begin{aligned}
\mathrm{E}\|I_1\|_2^2 &\leqslant 3(\gamma_0-\frac{h}{n})^2\|\mathbf{A}_1\|_2^4\cdot \mathrm{E}\left(\|\widehat{\boldsymbol{\Gamma}}_x(h,0,\gamma_0-\frac{h}{n})-\boldsymbol{\Gamma}_x(h,0,\gamma_0-\frac{h}{n})\|_2^2\right)\\
&\quad+\frac{3h^2}{n^2}\|\mathbf{A}_1\|_2^2\cdot \mathrm{E}\left(\|\widehat{\boldsymbol{\Gamma}}_x(h,\gamma_0-\frac{h}{n},\gamma_0)-\boldsymbol{\Gamma}_x(h,\gamma_0-\frac{h}{n},\gamma_0)\|_2^2\right)\cdot\|\mathbf{A}_2\|_2^2\\
&\quad+3(\epsilon-\frac{h}{n})^2\|\mathbf{A}_2\|_2^4\cdot \mathrm{E}\left(\|\widehat{\boldsymbol{\Gamma}}_x(h,\gamma_0,\gamma_0+\epsilon-\frac{h}{n})-\boldsymbol{\Gamma}_x(h,\gamma_0,\gamma_0+\epsilon-\frac{h}{n})\|_2^2\right)\\
&\leqslant \frac{3(3h+8\alpha)k_{\max}^2\sigma_x^4}{n^2}\left((\gamma_0-\frac{h}{n})a_1^4p^{2-2\delta_1}+ha_1^4p^{2-\delta_1-\delta_2}n^{-1}+(\epsilon-\frac{h}{n})a_1^4p^{2-2\delta_2}\right)\\
&\leqslant \frac{3(3h+8\alpha)a_1^4k_{\max}^2\sigma_x^4}{n}\left(\gamma_0p^{2-2\delta_1}+hp^{2-\delta_1-\delta_2}n^{-1}+\epsilon p^{2-2\delta_2}\right).
\end{aligned}$$

Since $\mathbf{x}_t$ and $\boldsymbol{\varepsilon}_t$ are independent,

$$\begin{aligned}
&\mathrm{E}\left\|\frac{1}{n}\sum_{t=1}^{r_0-h}\mathbf{A}_1\mathbf{x}_{t,1}\boldsymbol{\varepsilon}_{t+h}'\right\|_2^2\leqslant\|\mathbf{A}_1\|_2^2\cdot\mathrm{E}\left\|\frac{1}{n}\sum_{t=1}^{r_0-h}\mathbf{x}_{t,1}\boldsymbol{\varepsilon}_{t+h}'\right\|_F^2\\
\leqslant\;& \frac{a_1^2p^{1-\delta_1}}{n^2}\sum_{i=1}^{k_i}\sum_{j=1}^{p}\mathrm{E}\left(\sum_{t=1}^{r_0-h}x_{t,1,i}\varepsilon_{t+h,j}\right)^2\leqslant\frac{a_1^2p^{1-\delta_1}}{n^2}\sum_{i=1}^{k_i}\sum_{j=1}^{p}\mathrm{E}\left(\sum_{t=1}^{r_0-h}x_{t,1,i}^2\varepsilon_{t+h,j}^2\right)\\
\leqslant\;& \frac{\gamma_0a_1^2k_{\max}\sigma_x^2\sigma_\varepsilon^2p^{2-\delta_1}}{n},
\end{aligned}\tag{19}$$

and

$$\mathrm{E}\left\|\frac{1}{n}\sum_{t=1}^{r_0-h}\boldsymbol{\varepsilon}_t\mathbf{x}_{t+h,1}'\mathbf{A}_1'\right\|_2^2\leqslant\frac{\gamma_0a_1^2k_{\max}\sigma_x^2\sigma_\varepsilon^2p^{2-\delta_1}}{n},\tag{20}$$

where $\varepsilon_{t,j}$ is the j-th entry in $\boldsymbol{\varepsilon}_t$. On the other hand,

$$\mathrm{E}\left\|\frac{1}{n}\sum_{t=1}^{r_0-h}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t+h}'\right\|_2^2\leqslant\mathrm{E}\left\|\frac{1}{n}\sum_{t=1}^{r_0-h}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t+h}'\right\|_F^2\leqslant\frac{1}{n^2}\sum_{t=1}^{r_0-h}\sum_{i=1}^{p}\sum_{j=1}^{p}\mathrm{E}(\varepsilon_{t,i}^2\varepsilon_{t+h,j}^2)\leqslant\frac{\gamma_0\sigma_\varepsilon^4p^2}{n}.$$

Together with (19) and (20) we have

$$E\|I_2\|_2^2\leqslant\frac{3\gamma_0a_1^2k_{\max}\sigma_x^2\sigma_\varepsilon^2p^{1-\delta_1}}{n}+\frac{3\gamma_0a_1^2k_{\max}\sigma_x^2\sigma_\varepsilon^2p^{1-\delta_1}}{n}+\frac{3\gamma_0\sigma_\varepsilon^4p^2}{n}\leqslant\frac{9\gamma_0a_1^2k_{\max}\nu^4p^2}{n},$$

where $\nu=\max\{\sigma_x,\sigma_\varepsilon,1\}$. Similarly, we can show that

$$\mathrm{E}\|I_3\|_2^2\leqslant\frac{9a_1^2hk_{\max}\nu^4p^2}{n^2},\quad \mathrm{E}\|I_4\|_2\leqslant\frac{9\epsilon a_1^2k_{\max}\nu^4p^2}{n}.$$

Hence,

$$\begin{aligned}
&\mathrm{E}\left\|\widehat{\boldsymbol{\Sigma}}_{y,1}(h,\gamma_0+\epsilon)-\boldsymbol{\Sigma}_{y,1}(h,\gamma_0+\epsilon)\right\|_2^2\leqslant\mathrm{E}(\|I_1\|_2+\|I_2\|_2+\|I_3\|+\|I_4\|_2)^2\\
\leqslant\;& 4\mathrm{E}\|I_1\|_2^2+4\mathrm{E}\|I_2\|_2^2+4\mathrm{E}\|I_3\|_2^2+4\mathrm{E}\|I_4\|_2^2\\
\leqslant\;& 12(3h+8\alpha)(\gamma_0+\frac{h}{n}+\epsilon)a_1^4k_{\max}^2\sigma_x^4p^2n^{-1}+36(\gamma_0+\frac{h}{n}+\epsilon)a_1^2k_{\max}\nu^4p^2n^{-1}\\
\leqslant\;& 48(3h+8\alpha)(\gamma_0+\frac{h}{n}+\epsilon)a_1^4k_{\max}^2\nu^4p^2n^{-1}\leqslant144(3h+8\alpha)a_1^4k_{\max}^2\nu^4p^2n^{-1}.
\end{aligned}$$

When $\epsilon < 0$, it can be proven in a similar fashion. ∎

Lemma 3. *Under Conditions 1-4 and 6, for $\epsilon \in (-\gamma_0, 1-\gamma_0)$ and $|\epsilon| > (h+1)/n$, it holds that*

$$\|\mathbf{\Sigma}_{y,1}(h,\gamma_0+\epsilon)\|_2 \leqslant \begin{cases} \gamma_0 a_1^2 k_{\max}\sigma_x^2 p^{1-\delta_1}, & \epsilon \in (-\gamma_0, \frac{-(h+1)}{n}); \\ a_1^2 k_{\max}\sigma_x^2(\gamma_0 p^{1-\delta_1} + hp^{1-\delta_1/2-\delta_2/2}n^{-1} + \epsilon p^{1-\delta_2}), & \epsilon \in (\frac{h+1}{n}, 1-\gamma_0), \end{cases}$$

and

$$\|\mathbf{\Sigma}_{y,2}(h,\gamma_0+\epsilon)\|_2 \leqslant \begin{cases} a_1^2 k_{\max}\sigma_x^2[-\epsilon p^{1-\delta_1} + hp^{1-\delta_1/2-\delta_2/2}n^{-1} + (1-\gamma_0)p^{1-\delta_2}], & \epsilon \in (-\gamma_0, \frac{-(h+1)}{n}); \\ (1-\gamma_0)a_1^2 k_{\max}\sigma_x^2 p^{1-\delta_2}, & \epsilon \in (\frac{h+1}{n}, 1-\gamma_0). \end{cases}$$

Proof: By the definition of $\mathbf{\Sigma}_{y,i}(h,\gamma)$ and Lemma 1, when $\epsilon > (h+1)/n$, we have

$$\begin{aligned}
\left\|\mathbf{\Sigma}_{y,1}(h,\gamma_0+\epsilon)\right\|_2 &= \frac{1}{n^2}\left\| \sum_{t=1}^{r_0-h} E(\mathbf{y}_t\mathbf{y}_{t+h}') + \sum_{t=r_0-h+1}^{r_0} E(\mathbf{y}_t\mathbf{y}_{t+h}') + \sum_{t=r_0+1}^{r_0+\lfloor\epsilon n\rfloor-h} E(\mathbf{y}_t\mathbf{y}_{t+h}')\right\|_2 \\
&\leqslant \quad (\gamma_0-\frac{h}{n})\|\mathbf{A}_1\|_2^2\cdot\left\|\mathbf{\Gamma}_x(h,0,\gamma_0-\frac{h}{n})\right\|_2 + \frac{h}{n}\|\mathbf{A}_1\|_2\cdot\|\mathbf{A}_2\|_2\cdot\left\|\mathbf{\Gamma}_x(h,\gamma_0-\frac{h}{n},\gamma_0)\right\|_2 \\
&\quad +(\epsilon-\frac{h}{n})\|\mathbf{A}_2\|_2^2\cdot\left\|\mathbf{\Gamma}_x(h,\gamma_0,\gamma_0+\epsilon-\frac{h}{n})\right\|_2 \\
&\leqslant \quad \gamma_0 a_1^2 k_{\max}\sigma_x^2 p^{1-\delta_1} + a_1^2 h k_{\max}\sigma_x^2 p^{1-\delta_1/2-\delta_2/2}n^{-1} + \epsilon a_1^2 k_{\max}\sigma_x^2 p^{1-\delta_2}.
\end{aligned}$$

∎

Lemma 4. *Under Conditions 1-4 and 6, for $\epsilon \in (-\gamma_0, 1-\gamma_0)$and $|\epsilon| > (h+1)/n$, it holds that*

$$\|\mathbf{B}_1'\mathbf{\Sigma}_{y,1}(h,\gamma_0+\epsilon)\|_2 \begin{cases} = 0, & \epsilon \in (-\gamma_0, \frac{-(h+1)}{n}); \\ \leqslant \epsilon a_1^2 k_{\max}\sigma_x^2 p^{1-\delta_2}, & \epsilon \in (\frac{h+1}{n}, 1-\gamma_0), \end{cases}$$

and

$$\|\mathbf{B}_2'\mathbf{\Sigma}_{y,2}(h,\gamma_0+\epsilon)\|_2 \begin{cases} \leqslant a_1^2 k_{\max}\sigma_x^2(-\epsilon p^{1-\delta_1} + hp^{1-\delta_1/2-\delta_2/2}n^{-1}). & \epsilon \in (-\gamma_0, \frac{-(h+1)}{n}); \\ = 0, & \epsilon \in (\frac{h+1}{n}, 1-\gamma_0). \end{cases}$$

Proof: When $\epsilon > (h+1)/n$, by Lemma 1

$$\begin{aligned}
\left\|\mathbf{B}_1'\mathbf{\Sigma}_{y,1}(h,\gamma_0+\epsilon)\right\|_2 &= \frac{1}{n}\left\|\sum_{t=1}^{\lfloor\gamma_0 n+\epsilon n\rfloor-h} \mathbf{B}_1'\mathrm{E}\left(\mathbf{y}_t\mathbf{y}_{t+h}'\right)\right\|_2 \\
&= \quad \frac{1}{n}\left\|\sum_{t=1}^{r_0}\mathbf{B}_1'\mathbf{A}_1\mathrm{E}\left(\mathbf{x}_{t,1}\mathbf{y}_{t+h}'\right) + \sum_{t=r_0+1}^{r_0+\lfloor\epsilon n\rfloor-h}\mathbf{B}_1'\mathbf{A}_2\mathrm{E}(\mathbf{x}_{t,2}\mathbf{x}_{t+h,2}')\mathbf{A}_2'\right\|_2 \\
&\leqslant \quad \frac{N(\gamma_0,\gamma_0+\epsilon-\frac{h}{n})}{n}\|\mathbf{B}_1\|_2\cdot\|\mathbf{A}_2\mathbf{\Gamma}_x(h,\gamma_0,\gamma_0+\epsilon-\frac{h}{n})\mathbf{A}_2'\|_2 \leqslant \epsilon a_1^2 k_{\max}\sigma_x^2 p^{1-\delta_2}.
\end{aligned}$$

∎

Lemma 5. *Under Conditions 4 and 7, it holds that*

$$\|\mathbf{B}_1'\mathbf{A}_2\|_2^2 \geqslant a_2^2 d^2 \tau p^{1-\delta_2}, \quad \|\mathbf{B}_2'\mathbf{A}_1\|_2^2 \geqslant a_2^2 d^2 \tau p^{1-\delta_1},$$

where a_2 is a positive constant such that $a_2 p^{1/2-\delta_i/2} \leqslant \|\mathbf{A}_i\|_{\min}$ for $i = 1, 2$, and $\tau = \min\{k_2/k_1, k_1/k_2\}$.

Proof: Note that

$$\begin{aligned}\operatorname{tr}\left[\mathbf{Q}_2'\begin{pmatrix}\mathbf{Q}_1 & \mathbf{B}_1\end{pmatrix}\begin{pmatrix}\mathbf{Q}_1'\\ \mathbf{B}_1'\end{pmatrix}\mathbf{Q}_2\right] &= \operatorname{tr}(\mathbf{Q}_2'\mathbf{Q}_1\mathbf{Q}_1'\mathbf{Q}_2) + \operatorname{tr}(\mathbf{Q}_2'\mathbf{B}_1\mathbf{B}_1'\mathbf{Q}_2)\\ &= k_{\min}\{1 - [\mathcal{D}(\mathcal{M}(\mathbf{Q}_1), \mathcal{M}(\mathbf{Q}_2))]^2\} + \operatorname{tr}(\mathbf{Q}_2'\mathbf{B}_1\mathbf{B}_1'\mathbf{Q}_2),\end{aligned}$$

where $k_{\min} = \min\{k_1, k_2\}$. On the other hand,

$$\operatorname{tr}\left[\mathbf{Q}_2'\begin{pmatrix}\mathbf{Q}_1 & \mathbf{B}_1\end{pmatrix}\begin{pmatrix}\mathbf{Q}_1\\ \mathbf{B}_1\end{pmatrix}\mathbf{Q}_2\right] = \operatorname{tr}(\mathbf{Q}_2'\mathbf{Q}_2) = k_2.$$

Hence, $\operatorname{tr}(\mathbf{Q}_2'\mathbf{B}_1\mathbf{B}_1'\mathbf{Q}_2) = k_{\min}[\mathcal{D}(\mathcal{M}(\mathbf{Q}_1), \mathcal{M}(\mathbf{Q}_2))]^2 + k_2 - k_{\min}$. Then we have $\|\mathbf{B}_1'\mathbf{Q}_2\|_2^2 \geqslant \operatorname{tr}(\mathbf{Q}_2'\mathbf{B}_1\mathbf{B}_1'\mathbf{Q}_2)/k_2 \geqslant \tau d^2$. Condition 4 implies that there exists a positive constant a_2 such that $\|\mathbf{A}_2\|_{\min} \geqslant a_2 p^{1/2-\delta_i/2}$. It follows

$$\|\mathbf{B}_1'\mathbf{A}_2\| \geqslant a_2 d\tau p^{1/2-\delta_2/2}.$$

The other inequality can shown in a similar way. ∎

Lemma 6. *Under Conditions 1-8, we have $G(\gamma_0) = 0$ and for $\epsilon \in (\eta_1 - \gamma_0, \eta_2 - \gamma_0)$ and $|\epsilon| > (h+1)/n$,*

$$G(\gamma_0 + \epsilon) \geqslant \begin{cases} a_2^2 d^2 \tau (a_2^2 u_0^2 \epsilon^2 p^{2-2\delta_1}/2 + a_2^2 h_1^* u_0^2 \epsilon p^{2-2\delta_1} n^{-1} - a_1^2 h_1^{*2} k_{\max} \sigma_x^2 p^{2-\delta_1-\delta_2} n^{-2}), & \epsilon \in (\eta_1 - \gamma_0, \frac{-(h+1)}{n});\\ a_2^4 d^2 \tau u_0^2 (\epsilon^2 p^{2-2\delta_2} - 2h_2^* p^{2-2\delta_2} n^{-1}), & \epsilon \in (\frac{h+1}{n}, \eta_2 - \gamma_0).\end{cases}$$

Proof: Under Condition 8, by Lemmas 1 and 5, and Theorem 6 in Merikoski and Kumar (2004), when $\epsilon < -(h+1)/n$,

$$\begin{aligned}
&G(\gamma_0 + \epsilon) \geqslant \|\mathbf{B}_2'\mathbf{M}_2(\gamma_0 + \epsilon)\mathbf{B}_2\|_2 \geqslant \|\mathbf{B}_2'\mathbf{\Sigma}_{y,2}(h, \gamma_0 + \epsilon)\|_2^2\\
&\geqslant \|\mathbf{B}_2'\mathbf{A}_1\|_2^2 \left\|\frac{N(\gamma_0 + \epsilon, \gamma_0 - \frac{h_1^*}{n})}{n}\mathbf{\Gamma}_x(h_1^*, \gamma_0 + \epsilon, \gamma_0 - \frac{h_1^*}{n})\mathbf{A}_1' + \frac{N(\gamma_0 - \frac{h_1^*}{n}, \gamma_0)}{n}\mathbf{\Gamma}_x(h_1^*; \gamma_0 - \frac{h}{n}, \gamma_0)\mathbf{A}_2'\right\|_{\min}^2\\
&\geqslant \|\mathbf{B}_2'\mathbf{A}_1\|_2^2 \left[-(\epsilon + h_1^*/n)\|\mathbf{\Gamma}_x(h_1^*, \gamma_0 + \epsilon, \gamma_0 - \frac{h_1^*}{n})\mathbf{A}_1'\|_{\min} - h\|\mathbf{\Gamma}_x(h_1, \gamma_0 - \frac{h_1^*}{n}, \gamma_0)\mathbf{A}_1'\|_2\right]^2\\
&\geqslant \|\mathbf{B}_2'\mathbf{A}_1\|_2^2 \left[\frac{(\epsilon + h_1^*/n)^2}{2}\|\mathbf{\Gamma}_x(h_1^*, \gamma_0 + \epsilon, \gamma_0 - \frac{h_1^*}{n})\mathbf{A}_1'\|_{\min}^2 - h^2\|\mathbf{\Gamma}_x(h_1, \gamma_0 - \frac{h_1^*}{n}, \gamma_0)\mathbf{A}_1'\|_2^2\right]\\
&\geqslant a_2^2 d^2 \tau p^{1-\delta_1}[(\epsilon + h_1^*/n)^2 a_2^2 u_0^2 p^{1-\delta_1}/2 - a_1^2 h_1^{*2} k_{\max} \sigma_x^2 p^{1-\delta_2} n^{-2}]\\
&= a_2^2 d^2 \tau (a_2^2 u_0^2 \epsilon^2 p^{2-2\delta_1}/2 + a_2^2 u_0^2 h_1^* \epsilon p^{2-2\delta_1} n^{-1} - a_1^2 h_1^{*2} k_{\max} \sigma_x^2 p^{2-\delta_1-\delta_2} n^{-2});
\end{aligned}$$

when $\epsilon > (h+1)/n$,

$$\begin{aligned}
G(\gamma_0+\epsilon) &\geqslant \|\mathbf{B}_1'\mathbf{M}_1(\gamma_0+\epsilon)\mathbf{B}_1\|_2 \geqslant \|\mathbf{B}_1'\boldsymbol{\Sigma}_{y,1}(h,\gamma_0+\epsilon)\|_2^2 \\
&\geqslant (\epsilon-h/n)^2\|\mathbf{B}_1'\mathbf{A}_2\|_2^2\cdot\|\boldsymbol{\Gamma}_x(h_2^*,\gamma_0,\gamma_0+\epsilon-\frac{h_2^*}{n})\mathbf{A}_2'\|_{\min}^2 \\
&= a_2^4 d^2\tau u_0^2(\epsilon^2 p^{2-2\delta_2}-2h_2^* p^{2-2\delta_2}n^{-1}).
\end{aligned}$$

From Lemma 4, we have

$$G(\gamma_0)=0.$$

■

Lemma 7. *Under Conditions 1-6, if $p^{\delta_{\max}}n^{-1/2}=o(1)$, with true k_1 and k_2, as $n,p,\to\infty$, we have*

$$\mathrm{E}\|\widehat{\mathbf{B}}_i(\eta_i)-\mathbf{B}_i(\eta_i)\|_2^2 \leqslant Cp^{2\delta_i}n^{-1},\ \textit{for } i=1,2.$$

Proof: Let $Y_t = x_{t,i,q}x_{t+h,i,\ell}-\mathrm{E}(x_{t,i,q}x_{t+h,i,\ell})$. Condition 2 indicates that there exists a positive constant σ_y such that $\mathrm{E}(|Y_t^{2\zeta}|)<\sigma_y^{2\zeta}$. For any $0\leqslant c_1<c_2\leqslant 1$, by Lemma 6 in Liu and Chen (2020), we have

$$\frac{1}{N(c_1,c_2)^4}\mathrm{E}\left(\sum_{t=\lfloor c_1 n\rfloor+1}^{\lfloor c_2 n\rfloor} Y_t^4\right)\leqslant\frac{(47h+48\alpha+192\alpha^2)\sigma_y^4}{N(c_1,c_2)}.$$

It follows

$$\begin{aligned}
&\mathrm{E}\|\widehat{\boldsymbol{\Gamma}}_x(h,c_1,c_2)-\boldsymbol{\Gamma}_x(h,c_1,c_2)\|_2^4 \\
&\leqslant \mathrm{E}\|\widehat{\boldsymbol{\Gamma}}_x(h,c_1,c_2)-\boldsymbol{\Gamma}_x(h,c_1,c_2)\|_F^4 \leqslant \frac{(47h+48\alpha+192\alpha^2)k_{\max}^2\sigma_y^4}{N(c_1,c_2)}.
\end{aligned}$$

Thus we have

$$\begin{aligned}
&\mathrm{E}\|\widehat{\boldsymbol{\Sigma}}_{y,1}(h,\gamma_1)-\boldsymbol{\Sigma}_{y,1}(h,\gamma_1)\|_2^4 \\
\leqslant\ & \frac{16N(0,\gamma_1-\frac{h}{n})^4}{n^4}\|\mathbf{A}_1\|_2^8\cdot\mathrm{E}\|\widehat{\boldsymbol{\Gamma}}_x(h,0,\gamma_1-\frac{h}{n})-\boldsymbol{\Gamma}_x(h,0,\gamma_1-\frac{h}{n})\|_2^4 \\
&+\frac{16}{n^4}\|\mathbf{A}_1\|_2^4\cdot\mathrm{E}\left(\left\|\sum_{t=1}^{\gamma_1 n-h}\mathbf{x}_{t,1}\boldsymbol{\varepsilon}_{t+h}'\right\|_2^4\right)+\frac{16}{n^4}\|\mathbf{A}_1\|_2^4\cdot\mathrm{E}\left(\left\|\sum_{t=1}^{\gamma_1 n-h}\boldsymbol{\varepsilon}_t\mathbf{x}_{t+h,1}'\right\|_2^4\right)+\frac{16}{n^4}\mathrm{E}\left(\left\|\sum_{t=1}^{\gamma_1 n-h}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t+h}'\right\|_2^4\right) \\
\leqslant\ & \frac{C_1p^{4-4\delta_1}}{n}+\frac{16C_2p^{2-2\delta_1}}{n^4}\mathrm{E}\left(\sum_{t=1}^{\gamma_1 n-h}\sum_{q=1}^{k_1}\sum_{v=1}^{p}x_{t,1,q}^2\epsilon_{t+h,v}^2\right)^2+\frac{16C_2p^{2-2\delta_1}}{n^4}\mathrm{E}\left(\sum_{t=1}^{\gamma_1 n-h}\sum_{q=1}^{k_1}\sum_{v=1}^{p}\epsilon_{t,q}^2x_{t+h,1,v}^2\right)^2 \\
&+\frac{16C_3}{n^4}\mathrm{E}\left(\sum_{t=1}^{\gamma_1 n-h}\sum_{q=1}^{p}\sum_{v=1}^{p}\epsilon_{t,q}^2\epsilon_{t+h,v}^2\right)^2 \\
\leqslant\ & \frac{C_1p^{4-4\delta_1}}{n}+\frac{C_2p^{2-2\delta_1}}{n^2},
\end{aligned}$$

where C_1,C_2 and C_3 are positive constants and depend only on the parameters.

Hence, with Lemmas 2 and 3

$$
\begin{aligned}
& \mathrm{E}\|\widehat{\mathbf{M}}_1(\gamma_1)-\mathbf{M}_1(\gamma_1)\|_2^2 \\
\leqslant\ & h_0 \sum_{h=1}^{h_0} \mathrm{E}\|\widehat{\mathbf{\Sigma}}_{y,1}(h,\gamma_1)\widehat{\mathbf{\Sigma}}_{y,1}(h,\gamma_1)'-\mathbf{\Sigma}_{y,1}(h,\gamma_1)\mathbf{\Sigma}_{y,1}(h,\gamma_1)'\|_2^2 \\
\leqslant\ & 2h_0 \sum_{h=1}^{h_0} \Big[\mathrm{E}\|\widehat{\mathbf{\Sigma}}_{y,1}(h,\gamma_1)-\mathbf{\Sigma}_{y,1}(h,\gamma_1)\|_2^4+\|\mathbf{\Sigma}_{y,1}(h,\gamma_1)\|_2^2\cdot\mathrm{E}\|\widehat{\mathbf{\Sigma}}_{y,1}(h,\gamma_1)-\mathbf{\Sigma}_{y,1}(h,\gamma_1)\|_2^2\Big] \\
\leqslant\ & Cp^{4-2\delta_1}n^{-1}.
\end{aligned}
$$

Following the proof of Theorem 1 in Lam et al. (2011), we can reach the conclusion. ∎

Lemma 8. *Under Conditions 1-8, for $\epsilon\in[-\gamma_0,1-\gamma_0]$, it holds that*

$$
\begin{aligned}
& \mathrm{E}|\widehat{G}(h,\gamma_0+\epsilon)-G(h,\gamma_0+\epsilon)| \\
\leqslant\ & \begin{cases}
C_1p^2n^{-1}+C_2\epsilon p^{2-\delta_1}n^{-1/2}+C_3\epsilon^2p^{2-2\delta_1+\delta_2}n^{-1/2}, & \epsilon\in(-\gamma_0,-\frac{2}{n}); \\
C_1p^2n^{-1}, & \epsilon=0; \\
C_1p^2n^{-1}+C_2\epsilon p^{2-\delta_2}n^{-1/2}+C_3\epsilon^2p^{2+\delta_1-2\delta_2}n^{-1/2}, & \epsilon\in(\frac{2}{n},1-\gamma_0).
\end{cases}
\end{aligned}
$$

Proof: By the definition of $\mathbf{M}_i(\eta_i)$, we can see that $\mathcal{M}(\mathbf{B}_i)=\mathcal{M}(\mathbf{B}_i(\eta_i))$. It implies that there exists an orthogonal $(p-k_i)\times(p-k_i)$ matrix $\mathbf{R}_i$ such that $\mathbf{B}_i=\mathbf{B}_i(\eta_i)\mathbf{R}_i$.

$$
G(\gamma)=\sum_{i=1}^2\|\mathbf{R}_i'\mathbf{B}_i(\eta_i)'\mathbf{M}_i(\gamma)\mathbf{B}_i(\eta_i)\mathbf{R}_i\|_2=\sum_{i=1}^2\|\mathbf{B}_i(\eta_i)'\mathbf{M}_i(\gamma)'\mathbf{B}_i(\eta_i)\|_2.
$$

By the definition of $\widehat{G}(\gamma)$ we have,

$$
\begin{aligned}
& |\widehat{G}(\gamma)-G(\gamma)| \\
\leqslant\ & \sum_{i=1}^2\sum_{h=1}^{h_0}\Big\|\widehat{\mathbf{B}}_i(\eta_i)'\widehat{\mathbf{\Sigma}}_{y,i}(h,\gamma)\widehat{\mathbf{\Sigma}}_{y,i}(h,\gamma)'\widehat{\mathbf{B}}_i-\mathbf{B}_i(\eta_i)'\mathbf{\Sigma}_{y,i}(h,\gamma)\mathbf{\Sigma}_{y,i}(h,\gamma)'\mathbf{B}_i(\eta_i)\Big\|_2 \\
\leqslant\ & \sum_{i=1}^2\sum_{h=1}^{h_0}\Big[\big\|\widehat{\mathbf{B}}_i(\eta_i)'\widehat{\mathbf{\Sigma}}_{y,i}(h,\gamma)-\mathbf{B}_i(\eta_i)'\mathbf{\Sigma}_{y,i}(h,\gamma)\big\|_2^2 \\
& +2\big\|\mathbf{B}_i(\eta_i)'\mathbf{\Sigma}_{y,i}(h,\gamma)\big\|_2\cdot\big\|\widehat{\mathbf{B}}_i(\eta_i)'\widehat{\mathbf{\Sigma}}_{y,i}(h,\gamma)-\mathbf{B}_i(\eta_i)'\mathbf{\Sigma}_{y,i}(h,\gamma)\big\|_2\Big] \\
\leqslant\ & \sum_{i=1}^2\sum_{h=1}^{h_0}\Bigg[\Big(\|\widehat{\mathbf{B}}_i(\eta_i)\|_2\cdot\big\|\widehat{\mathbf{\Sigma}}_{y,i}(h,\gamma)-\mathbf{\Sigma}_{y,i}(h,\gamma)\big\|_2+\big\|\widehat{\mathbf{B}}_i(\eta_i)-\mathbf{B}_i(\eta_i)\big\|_2\cdot\|\mathbf{\Sigma}_{y,i}(h,\gamma)\|_2\Big)^2 \\
& +2\|\mathbf{B}_i(\eta_i)'\mathbf{\Sigma}_{y,i}(h,\gamma)\|_2\left(\|\widehat{\mathbf{B}}_i(\eta_i)\|_2\|\widehat{\mathbf{\Sigma}}_{y,i}(h,\gamma)-\mathbf{\Sigma}_{y,i}(h,\gamma)\|_2+\|\widehat{\mathbf{B}}_i(\eta_i)-\mathbf{B}_i(\eta_i)\|_2\|\mathbf{\Sigma}_{y,i}(h,\gamma)\|_2\right)\Bigg] \\
=\ & \sum_{i=1}^2\sum_{h=1}^{h_0}L_{i,1}(h,\gamma)+L_{i,2}(h,\gamma). \qquad (21)
\end{aligned}
$$

By Lemmas 2-4 and 7,

$$\mathrm{E}(L_{1,1}(h,\gamma_0+\epsilon)) \leqslant \begin{cases} C_1 p^2 n^{-1}, & \epsilon \in (-\gamma_0, -\frac{h+1}{n}); \\ C_1 p^2 n^{-1} + C_3 \epsilon^2 p^{2+2\delta_1-2\delta_2} n^{-1}, & \epsilon \in (\frac{h+1}{n}, 1-\gamma_0), \end{cases}$$

$$\mathrm{E}(L_{1,2}(h,\gamma_0+\epsilon)) \begin{cases} = 0, & \epsilon \in (-\gamma_0, -\frac{h+1}{n}); \\ \leqslant C_1 \epsilon p^{2-\delta_2} n^{-1/2} + C_3 \epsilon^2 p^{2+\delta_1-2\delta_2} n^{-1/2}, & \epsilon \in (\frac{h+1}{n}, 1-\gamma_0), \end{cases}$$

$$\mathrm{E}(L_{2,1}(h,\gamma_0+\epsilon)) \leqslant \begin{cases} C_1 p^2 n^{-1} + C_3 \epsilon^2 p^{2-2\delta_1+2\delta_2} n^{-1}, & \epsilon \in (-\gamma_0, -\frac{h+1}{n}); \\ C_1 p^2 n^{-1}, & \epsilon \in (\frac{h+1}{n}, 1-\gamma_0), \end{cases}$$

$$\mathrm{E}(L_{2,2}(h,\gamma_0+\epsilon)) \begin{cases} \leqslant C_1 p^{2-\delta_1/2-\delta_2/2} n^{-3/2} + C_2 \epsilon p^{2-\delta_1} n^{-1/2} + C_3 \epsilon^2 p^{2-2\delta_1+\delta_2} n^{-1/2}, & \epsilon \in (-\gamma_0, -\frac{h+1}{n}); \\ = 0, & \epsilon \in (\frac{h+1}{n}, 1-\gamma_0). \end{cases}$$

From (21), it follows,

$$\begin{aligned} &\mathrm{E}|\widehat{G}(\gamma_0+\epsilon) - G(\gamma_0+\epsilon)| \\ &\leqslant \begin{cases} C_1 p^2 n^{-1} + C_2 \epsilon p^{2-\delta_1} n^{-1/2} + C_3 \epsilon^2 p^{2-2\delta_1+\delta_2} n^{-1/2}, & \epsilon \in (-\gamma_0, -\frac{2}{n}); \\ C_1 p^2 n^{-1}, & \epsilon = 0; \\ C_1 p^2 n^{-1} + C_2 \epsilon p^{2-\delta_2} n^{-1/2} + C_3 \epsilon^2 p^{2+\delta_1-2\delta_2} n^{-1/2}, & \epsilon \in (\frac{2}{n}, 1-\gamma_0). \end{cases} \end{aligned}$$

■

Proof of Theorem 1. Since $G(r) \geqslant 0$ and $G(r_0) = 0$, for any fixed $\epsilon > (h+1)/n$, it follows that

$$\begin{aligned} P(\widehat{r} - r_0 > \epsilon) &= P[\widehat{G}(r_0) > \widehat{G}(\widehat{r}), \widehat{r} > r_0 + \epsilon] \\ &= P\Big[\widehat{G}(r_0) - G(r_0) > \widehat{G}(\widehat{r}) - G(\widehat{r}) + G(\widehat{r}), \widehat{r} > r_0 + \epsilon\Big] \\ &= P\Big[\widehat{G}(r_0) - G(r_0) + G(\widehat{r}) - \widehat{G}(\widehat{r}) + \frac{3}{4} a_2^4 d^2 \tau u_0^2 \epsilon^2 p^{2-2\delta_2} - G(\widehat{r}) > \frac{3}{4} a_2^4 d^2 \tau u_0^2 \epsilon^2 p^{2-2\delta_2}, \widehat{r} > r_0 + \epsilon\Big] \\ &\leqslant P\Big[\big|\widehat{G}(r_0) - G(r_0)\big| > +\frac{1}{4} a_2^4 d^2 \tau u_0^2 \epsilon^2 p^{2-2\delta_2}\Big] + P\Big[\big|\widehat{G}(\widehat{r}) - G(\widehat{r})\big| > \frac{1}{4} a_2^4 d^2 \tau u_0^2 \epsilon^2 p^{2-2\delta_2}, \widehat{r} > r_0 + \epsilon\Big] \\ &\quad + P\Big[\frac{3}{4} a_2^4 d^2 \tau u_0^2 \epsilon^2 p^{2-2\delta_2} - G(\widehat{r}) > \frac{1}{4} a_2^4 d^2 \tau u_0^2 \epsilon^2 p^{2-2\delta_2}, \widehat{r} > r_0 + \epsilon\Big] \\ &= P\Big[\big|\widehat{G}(r_0) - G(r_0)\big| > \frac{1}{4} a_2^4 d^2 \tau u_0^2 \epsilon^2 p^{2-2\delta_2}\Big] + P\Big[\big|\widehat{G}(\widehat{r}) - G(\widehat{r})\big| > \frac{1}{4} a_2^4 d^2 \tau u_0^2 \epsilon^2 p^{2-2\delta_2}, \widehat{r} > r_0 + \epsilon\Big] \\ &\quad + P\Big[G(\widehat{r}) < \frac{1}{2} a_2^4 d^2 \tau u_0^2 \epsilon^2 p^{2-2\delta_2}, \widehat{r} > r_0 + \epsilon\Big] \\ &= I_1 + I_2 + I_3. \end{aligned}$$

By Lemma 6, Lemma 8, and Chebyshev's inequality, if $p^{\max}n^{-1/2} = o(1)$ and n is large enough, when $\widehat{r} > r_0 + \epsilon$, we have

$$I_1 < C_1 p^{2\delta_2} n^{-1}, \quad I_2 < \frac{C_2 p^{\delta_2} n^{-1/2}}{\epsilon}, \quad I_3 = 0.$$

Hence, there exists a constant C such that

$$P(\widehat{r} > r_0 + \epsilon) \leqslant \frac{C p^{\delta_2} n^{-1/2}}{\epsilon}, \text{ for } \epsilon > 0.$$

■

Proof of Theorem 2. When $\widehat{\gamma} > \gamma_0$, from Theorem 1 in Lam et al. (2011), it follows

$$\mathcal{D}\{\mathcal{M}[\widehat{\mathbf{Q}}_2(\widehat{\gamma})], \mathcal{M}(\mathbf{Q}_2)\} = O_p(p^{\delta_2} n^{-1/2}), \text{ as } n, p \to \infty.$$

Now we start to investigate the asymptotic properties of $\mathcal{M}[\widehat{\mathbf{Q}}_1(\widehat{\gamma})]$ when $\widehat{\gamma} > \gamma_0$.

For $\epsilon > 0$, Lemmas 2-4 imply that

$$\begin{aligned}
&\|\widehat{\mathbf{M}}_1(\gamma_0+\epsilon) - \mathbf{M}_1(\gamma_0+\epsilon)\|_2 \\
&\leqslant \sum_{h=1}^{h_0} \Big(\|\widehat{\boldsymbol{\Sigma}}_{y,1}(h,\gamma_0+\epsilon) - \boldsymbol{\Sigma}_{y,1}(h,\gamma_0+\epsilon)\|_2^2 + 2\|\boldsymbol{\Sigma}_{y,1}(h,\gamma_0+\epsilon)\|_2 \cdot \|\widehat{\boldsymbol{\Sigma}}_{y,1}(h,\gamma_0+\epsilon) - \boldsymbol{\Sigma}_{y,1}(h,\gamma_0+\epsilon)\|_2 \Big) \\
&= O_p(p^2 n^{-1}) + O_p(p^{2-\delta_1} n^{-1/2}) + O_p(\epsilon p^{2-\delta_2} n^{-1/2}) \\
&= O_p(p^{2-\delta_1} n^{-1/2}) + O_p(\epsilon p^{2-\delta_2} n^{-1/2}). \qquad (22)
\end{aligned}$$

Under Conditions 2 and 4, it follows from Lemma 1

$$\begin{aligned}
&\|\boldsymbol{\Sigma}_{y,1}(h, r_0+\epsilon) - \boldsymbol{\Sigma}_{y,1}(h, r_0)\|_2 = \frac{1}{n}\Big\| \sum_{t=r_0-h+1}^{\gamma_0+\lfloor \epsilon n\rfloor - h} \mathrm{E}(\mathbf{y}_t \mathbf{y}_{t+h}') \Big\|_2 \\
&= \Big\| \frac{N(\gamma_0 - h/n, \gamma_0)}{n} \mathbf{A}_1 \boldsymbol{\Gamma}_x(h, \gamma_0 - h/n, \gamma_0)\mathbf{A}_2' + \frac{N(\gamma_0, \gamma_0+\epsilon-h/n)}{n} \mathbf{A}_2 \boldsymbol{\Gamma}_x(h, \gamma, \gamma_0+\epsilon-h/n)\mathbf{A}_2' \Big\|_2 \\
&= O(\epsilon p^{1-\delta_2}) + O(p^{1-\delta_2/2-\delta_{\min}/2} n^{-1}).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\|\mathbf{M}_1(r_0+\epsilon) - \mathbf{M}_1\|_2 \\
&\leqslant \sum_{h=1}^{h_0} \big\| \boldsymbol{\Sigma}_{y,1}(h, r_0+\epsilon)\boldsymbol{\Sigma}_{y,1}(h, r_0+\epsilon)' - \boldsymbol{\Sigma}_{y,1}(h, r_0)\boldsymbol{\Sigma}_{y,1}(h, r_0)' \big\|_2 \\
&\leqslant \sum_{h=1}^{h_0} \big(\|\boldsymbol{\Sigma}_{y,1}(h, r_0+\epsilon) - \boldsymbol{\Sigma}_{y,1}(h, r_0)\|_2^2 + 2\|\boldsymbol{\Sigma}_{y,1}(h, r_0)\|_2 \cdot \|\boldsymbol{\Sigma}_{y,1}(h, r_0+\epsilon) - \boldsymbol{\Sigma}_{y,1}(h, r_0)\|_2 \big) \\
&= O(\epsilon^2 p^{2-2\delta_2}) + O(p^{2-\delta_2-\delta_{\min}} n^{-2}) + O(\epsilon p^{2-\delta_1-\delta_2}) + O(p^{2-\delta_1-\delta_2/2-\delta_{\min}/2} n^{-1}).
\end{aligned}$$

If $p^{\delta_{\max}} n^{-1/2} = o(1)$, together with (22), we have

$$\begin{aligned}
& \|\widehat{\mathbf{M}}_1(r_0+\epsilon) - \mathbf{M}_1\|_2 \\
\leqslant \; & \|\widehat{\mathbf{M}}_1(r_0+\epsilon) - \mathbf{M}_1(r_0+\epsilon)\|_2 + \|\mathbf{M}_1(r_0+\epsilon) - \mathbf{M}_1\|_2 \\
= \; & O_p(p^{2-\delta_1} n^{-1/2}) + O(\epsilon p^{2-\delta_1-\delta_2}) + O(\epsilon^2 p^{2-2\delta_2}).
\end{aligned}$$

Theorem 1 tells us if $\widehat{r} > r_0$, $|\widehat{r} - r_0| = O_p(p^{\delta_2} n^{-1/2})$. Therefore,

$$\|\widehat{\mathbf{M}}_1(\widehat{r}) - \mathbf{M}_1\|_2 = O_p(p^{2-\delta_1} n^{-1/2}).$$

Under Condition 5, by Theorem 9 in Merikoski and Kumar (2004), we can see that

$$\|\mathbf{M}_1\|_{\min} = \|\boldsymbol{\Sigma}_{y,1}(h,\gamma_0)\|_{\min}^2 \geqslant \|\mathbf{A}_1\|_2^2 \, \|\boldsymbol{\Sigma}_{x,1}(h,\gamma_0)\|_{\min}^2 \, \|\mathbf{A}_1\|_2^2 = O(p^{2-2\delta_1}).$$

Following the proof of Theorem 2 in Liu and Chen (2016), we have

$$\mathcal{D}\{\mathcal{M}[\widehat{\mathbf{Q}}_1(\widehat{\gamma})], \mathcal{M}(\mathbf{Q}_1)\} = O_p(p^{\delta_1} n^{-1/2}),$$

as $n, p \to \infty$, when $\widehat{r} > r_0$.

The conclusions for $\widehat{r} < r_0$ can be proven in a similar way. ∎

Lemma 9. *Let $\mathbf{B}_i^*$ be a $p \times (p - \widehat{k}_i)$ orthogonal matrix such that $\mathcal{M}(\mathbf{B}_i^*) \in \mathcal{M}(\mathbf{B}_i)$ for $i = 1, 2$. Under Conditions 1-4 and 9, for any $\mathbf{B}_i^*$ and $\epsilon \in (-\gamma_0, 1-\gamma_0)$ and $|\epsilon| > (h+1)/n$,*

$$\|\mathbf{B}_1^{*'} \boldsymbol{\Sigma}_{y,1}(h, \gamma_0+\epsilon)\|_2 \begin{cases} = 0, & \epsilon \in (-\gamma_0, -\frac{h+1}{n}), \\ \leqslant O(\epsilon p^{1-\delta_2}), & \epsilon \in (\frac{h+1}{n}, 1), \end{cases}$$

$$\|\mathbf{B}_2^{*'} \boldsymbol{\Sigma}_{y,2}(h, \gamma_0+\epsilon)\|_2 \begin{cases} \leqslant O(\epsilon p^{1-\delta_1}) + O(p^{1-\delta_1/2-\delta_2/2} n^{-1}), & \epsilon \in (-\gamma_0, -\frac{h+1}{n}), \\ = 0, & \epsilon \in (\frac{h+1}{n}, 1). \end{cases}$$

Proof. Note that for $\mathbf{B}_i^*$ such that $\mathbf{B}_i^{*'} \mathbf{A}_i = \mathbf{0}$, following the proof of Lemma 4, we can reach the conclusion. ∎

Proof of Theorem 3: Under Conditions 1-9, if $p^{\delta_{\max}} n^{-1/2} = o(1)$, similar to the proof of Theorem 1, we obtain that

$$\left\|\widehat{\mathbf{B}}_i(\eta_i) - \mathbf{B}_i(\eta_i)\right\|_2 = O_p(p^{\delta_i} n^{-1/2}), \text{ for } i = 1, 2.$$

Since

$$\widehat{\mathbf{B}}_i(\eta_i) = \Big(\; \widehat{\mathbf{q}}_{i,k_i+1}(\eta_i), \quad \ldots, \quad \widehat{\mathbf{q}}_{i,\widehat{k}_i}(\eta_i), \quad \widehat{\mathbf{B}}_{i,\widehat{k}_i}(\eta_i), \; \Big)$$

we have

$$\big\|\widehat{\mathbf{B}}_{i,\widehat{k}_i}(\eta_i)-\mathbf{B}_{i,\widehat{k}_i}(\eta_i)\big\|_2 \leqslant \big\|\widehat{\mathbf{B}}_i(\eta_i)-\mathbf{B}_i(\eta_i)\big\|_2 = O_p(p^{\delta_i}n^{-1/2}),\ \text{for } i=1,2.$$

With Lemma 7, similar to the proof of Theorem 1, we can complete the proof.

Proof of Theorem 4: Similar to proof of Theorem 2, we can obtain the results.

Proof of Corollary 1: Since $\widehat{k}_1 \geqslant k_1$ and $\widehat{k}_2 \geqslant k_2$, by the definition of $\mathcal{D}(\cdot,\cdot)$ and Theorem 4, the conclusion follows.

Proof of Proposition 1: Similar to proof of Corollary 1 in Lam and Yao (2012).

Proof of Corollary 2: We consider the case when $\tilde{k} > 0$ first.

An alternative way to denote $\mathcal{M}(\mathbf{Q}_1)$ and $\mathcal{M}(\mathbf{Q}_2)$ is introduced. Let $\bar{\mathbf{q}}_1,\ldots,\bar{\mathbf{q}}_{\tilde{k}}$ be an orthonormal basis of $\mathcal{M}(\mathbf{Q}_1)\cap\mathcal{M}(\mathbf{Q}_2)$. Define

$$\bar{\mathbf{Q}}_1 = (\bar{\mathbf{q}}_1,\ldots,\bar{\mathbf{q}}_{\tilde{k}},\bar{\mathbf{q}}_{1,\tilde{k}+1},\ldots,\bar{\mathbf{q}}_{1,k_1}), \quad \bar{\mathbf{Q}}_2 = (\bar{\mathbf{q}}_1,\ldots,\bar{\mathbf{q}}_{\tilde{k}},\bar{\mathbf{q}}_{2,\tilde{k}+1},\ldots,\mathbf{q}_{2,k_2}),$$

as orthonormal base of $\mathcal{M}(\mathbf{Q}_1)$ and $\mathcal{M}(\mathbf{Q}_2)$, respectively. It implies that there exists an orthonormal $k_i\times k_i$ matrix $\mathbf{V}_i$ such that $\mathbf{Q}_i = \bar{\mathbf{Q}}_i\mathbf{V}_i$ for $i=1,2$.

Define $\bar{\mathbf{Q}}^0 = (\bar{\mathbf{q}}_1,\ldots,\bar{\mathbf{q}}_{\tilde{k}})$, $\bar{\mathbf{Q}}_1^1 = (\bar{\mathbf{q}}_{\tilde{k}+1},\ldots,\bar{\mathbf{q}}_{k_1})$, and $\bar{\mathbf{Q}}_2^1 = (\bar{\mathbf{q}}_{\tilde{k}+1},\ldots,\bar{\mathbf{q}}_{k_2})$. Note that $\bar{\mathbf{Q}}^{0\prime}\bar{\mathbf{Q}}_1^1 = \mathbf{0}$ and $\bar{\mathbf{Q}}^{0\prime}\bar{\mathbf{Q}}_2^1 = \mathbf{0}$, but $\bar{\mathbf{Q}}_1^{1\prime}\bar{\mathbf{Q}}_2^1 = \mathbf{0}$ may not be true.

By Theorem 9 in Merikoski and Kumar (2004), it follows,

$$\begin{aligned}
&\sigma_{\tilde{k}+1}(\mathbf{Q}_1'\mathbf{Q}_2) = \sigma_{\tilde{k}+1}(\mathbf{V}_1'\bar{\mathbf{Q}}_1'\bar{\mathbf{Q}}_2\mathbf{V}_2) = \sigma_{\tilde{k}+1}(\bar{\mathbf{Q}}_1'\bar{\mathbf{Q}}_2)\\
=&\sigma_{\tilde{k}+1}\left(\begin{pmatrix}\bar{\mathbf{Q}}^{0\prime}\\ \bar{\mathbf{Q}}_1^{1\prime}\end{pmatrix}\begin{pmatrix}\bar{\mathbf{Q}}^0 & \bar{\mathbf{Q}}_2^1\end{pmatrix}\right) = \sigma_{\tilde{k}+1}\left(\begin{pmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \bar{\mathbf{Q}}_1^{1\prime}\bar{\mathbf{Q}}_2^1\end{pmatrix}\right) = \|\bar{\mathbf{Q}}_1^{1\prime}\bar{\mathbf{Q}}_2^1\|_2,
\end{aligned}$$

and together with Condition 11 we can show that $\|\bar{\mathbf{Q}}_1^{1\prime}\bar{\mathbf{Q}}_2^1\|_2 = \nu < 1$.

By the definition of $\mathbf{M}$ in (16), we have

$$\begin{aligned}
&\mathbf{M}\\
=&\left[\sum_{h=1}^{h_0}\big(\gamma_0\mathbf{A}_1\boldsymbol{\Sigma}_{x,1}(h,\gamma_0)\mathbf{A}_1' + (1-\gamma_0)\mathbf{A}_2\boldsymbol{\Sigma}_{x,2}(h,\gamma_0)\mathbf{A}_2'\big)\right] \qquad (23)\\
&\cdot\left[\sum_{h=1}^{h_0}\big(\gamma_0\mathbf{A}_1\boldsymbol{\Sigma}_{x,1}(h,\gamma_0)\mathbf{A}_1' + (1-\gamma_0)\mathbf{A}_2\boldsymbol{\Sigma}_{x,2}(h,\gamma_0)\mathbf{A}_2'\big)\right]' + o(1/n). \qquad (24)
\end{aligned}$$

The reason that $o(1/n)$ exists in the above equation is that the observations at time $\lfloor\gamma_0 n\rfloor - h,\ldots,\lfloor\gamma_0 n\rfloor$ are not counted when calculating $\boldsymbol{\Sigma}_{x,1}(h,\gamma_0)$ and $\boldsymbol{\Sigma}_{x,2}(h,\gamma_0)$ in (5) and $\boldsymbol{\Sigma}_y(h)$ has h more terms than $\boldsymbol{\Sigma}_{y,1}(h,\gamma_0)+\boldsymbol{\Sigma}_{y,2}(h,\gamma_0)$.

Condition 4 tells us that there exist two $k_i \times k_i$ non-singular matrices $\mathbf{\Gamma}_i$ such that $\mathbf{A}_i = \bar{\mathbf{Q}}_i \mathbf{\Gamma}_i$ and $\|\mathbf{\Gamma}_i\|_2 \asymp \|\mathbf{\Gamma}_i\|_{\min} \asymp p^{1/2-\delta_0/2}$ for $i = 1, 2$. Let $\mathbf{\Sigma}_1 = \gamma_0 \sum_{h=1}^{h_0} \mathbf{\Gamma}_1 \mathbf{\Sigma}_{x,1}(h, \gamma_0) \mathbf{\Gamma}_1'$ and $\mathbf{\Sigma}_2 = (1-\gamma_0) \sum_{h=1}^{h_0} \mathbf{\Gamma}_2 \mathbf{\Sigma}_{x,2}(h, \gamma_0) \mathbf{\Gamma}_2'$. With Condition 5, we can show that $\|\mathbf{\Sigma}_i\|_2 \asymp \|\mathbf{\Sigma}_i\|_{\min} \asymp O_p(p^{1-\delta_0})$ for $i = 1, 2$. By the definition of singular value, we have

$$
\begin{aligned}
&\left[\sigma_{k_1+k_2-\tilde{k}}(\bar{\mathbf{Q}}_1 \mathbf{\Sigma}_1 \bar{\mathbf{Q}}_1' + \bar{\mathbf{Q}}_2 \mathbf{\Sigma}_2 \bar{\mathbf{Q}}_2')\right]^2 \\
&= \max_{\substack{S \in \mathbb{R}^p \\ \dim(S) = k_1+k_2-\tilde{k}}} \min_{\substack{\|\mathbf{u}\|=1 \\ \mathbf{u} \in S}} \|(\bar{\mathbf{Q}}_1 \mathbf{\Sigma}_1 \bar{\mathbf{Q}}_1' + \bar{\mathbf{Q}}_2 \mathbf{\Sigma}_2 \bar{\mathbf{Q}}_2')\mathbf{u}\|_2^2 \\
&\geqslant \min_{\substack{\|\mathbf{u}\|=1 \\ \mathbf{u} \in \mathcal{M}(\mathbf{Q}_1) \cup \mathcal{M}(\mathbf{Q}_2)}} \|(\bar{\mathbf{Q}}_1 \mathbf{\Sigma}_1 \bar{\mathbf{Q}}_1' + \bar{\mathbf{Q}}_2 \mathbf{\Sigma}_2 \bar{\mathbf{Q}}_2')\mathbf{u}\|_2^2. \qquad (25)
\end{aligned}
$$

We write $\mathbf{u} = \sum_{i=1}^{\tilde{k}} a_i \bar{\mathbf{q}}_i + \sum_{i=\tilde{k}+1}^{k_1} b_i \bar{\mathbf{q}}_{1,i} + \sum_{i=\tilde{k}+1}^{k_2} c_i \bar{\mathbf{q}}_{2,i}$ for any $\mathbf{u} \in \mathcal{M}(\mathbf{Q}_1) \cup \mathcal{M}(\mathbf{Q}_2)$, and $\mathbf{a} = (a_1, \ldots, a_{\tilde{k}})'$, $\mathbf{b} = (b_{\tilde{k}+1}, \ldots, b_{k_1})'$, and $\mathbf{c} = (c_{\tilde{k}+1}, \ldots, c_{k_2})'$. We have

$$
\bar{\mathbf{Q}}_1'\mathbf{u} = \begin{pmatrix} \bar{\mathbf{Q}}^{0\prime} \\ \bar{\mathbf{Q}}_1^{1\prime} \end{pmatrix} \begin{pmatrix} \bar{\mathbf{Q}}^0 & \bar{\mathbf{Q}}_1^1 & \bar{\mathbf{Q}}_2^1 \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \end{pmatrix} = \begin{pmatrix} \mathbf{a} \\ \mathbf{b} + \bar{\mathbf{Q}}_1^{1\prime} \bar{\mathbf{Q}}_2^1 \mathbf{c} \end{pmatrix},
$$

and

$$
\bar{\mathbf{Q}}_2'\mathbf{u} = \begin{pmatrix} \bar{\mathbf{Q}}^{0\prime} \\ \bar{\mathbf{Q}}_2^{1\prime} \end{pmatrix} \begin{pmatrix} \bar{\mathbf{Q}}^0 & \bar{\mathbf{Q}}_1^1 & \bar{\mathbf{Q}}_2^1 \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \end{pmatrix} = \begin{pmatrix} \mathbf{a} \\ \bar{\mathbf{Q}}_2^{1\prime} \bar{\mathbf{Q}}_1^1 \mathbf{b} + \mathbf{c} \end{pmatrix}.
$$

If $\|\mathbf{b}\|_2 \geqslant \|\mathbf{c}\|_2$, since $\|\mathbf{u}\|_2 = 1$ and $\|\bar{\mathbf{Q}}_1^{1\prime} \bar{\mathbf{Q}}_2^1\|_2 = \nu < 1$, we have

$$
\begin{aligned}
\|\bar{\mathbf{Q}}_1'\mathbf{u}\|_2^2 &= \|\mathbf{a}\|_2^2 + \|\mathbf{b} + \bar{\mathbf{Q}}_1^{1\prime} \bar{\mathbf{Q}}_2^1 \mathbf{c}\|_2^2 \geqslant \|\mathbf{a}\|_2^2 + (\|\mathbf{b}\|_2 - \nu\|\mathbf{c}\|_2)^2 \\
&\geqslant \|\mathbf{a}\|_2^2 + (1-\nu)^2\|\mathbf{b}\|_2^2 \geqslant \|\mathbf{a}\|_2^2 + \tfrac{(1-\nu)^2}{2}\|\mathbf{b}\|_2^2 + \tfrac{(1-\nu)^2}{2}\|\mathbf{c}\|_2^2 \geqslant \tfrac{(1-\nu)^2}{2}. \qquad (26)
\end{aligned}
$$

Let $\mathbf{\Sigma}_1 \bar{\mathbf{Q}}_1'\mathbf{u} = \mathbf{v}_1$ and $\mathbf{\Sigma}_2 \bar{\mathbf{Q}}_2'\mathbf{u} = \mathbf{v}_2$. By Theorem 9 in Merikoski and Kumar (2004) and (26), we can show that $\|\mathbf{v}_1\|_2 \geqslant Cp^{1-\delta_0}$. It follows

$$
\begin{aligned}
&\|(\bar{\mathbf{Q}}_1 \mathbf{\Sigma}_1 \bar{\mathbf{Q}}_1' + \bar{\mathbf{Q}}_2 \mathbf{\Sigma}_2 \bar{\mathbf{Q}}_2')\mathbf{u}\|_2^2 \\
&= \|\bar{\mathbf{Q}}_1 \mathbf{v}_1 + \bar{\mathbf{Q}}_2^1 \mathbf{v}_2\|_2^2 = \mathbf{v}_1' \bar{\mathbf{Q}}_1' \bar{\mathbf{Q}}_1 \mathbf{v}_1 + 2\mathbf{v}_1' \bar{\mathbf{Q}}_1' \bar{\mathbf{Q}}_2 \mathbf{v}_2 + \mathbf{v}_2' \bar{\mathbf{Q}}_2' \bar{\mathbf{Q}}_2 \mathbf{v}_2 \\
&\geqslant \|\mathbf{v}_1\|_2^2 - 2\nu\|\mathbf{v}_1\|_2 \cdot \|\mathbf{v}_2\|_2 + \|\mathbf{v}_2\|_2^2 = (1-\nu^2)\|\mathbf{v}_1\|_2^2 + (\nu\|\mathbf{v}_1\|_2 - \|\mathbf{v}_2\|_2)^2 \geqslant Cp^{2-2\delta_0}. \qquad (27)
\end{aligned}
$$

If $\|\mathbf{b}\|_2 < \|\mathbf{c}\|_2$, then $\|\mathbf{v}_2\|_2 \geqslant Cp^{1/2-\delta_0/2}$. We can also obtain (27).

When $\tilde{k} = 0$, the conclusions can be reached in a similar fashion.

Together with (24), (25) and (27), we have

$$\lambda_{k_1+k_2-\tilde{k}}(\mathbf{M}) \geqslant Cp^{2-2\delta_0}.$$

Following the proof of Corollary 1 in Lam and Yao (2012), we can complete the proof. ∎