Constrained Factor Models for High-Dimensional Matrix-Variate Time Series

Yi Chen *
Department of Statistics and Biostatistics, Rutgers University
Ruey S. Tsay
Booth School of Business, University of Chicago
Rong Chen
Department of Statistics and Biostatistics, Rutgers University

November 12, 2017

Abstract

In many scientific fields, including economics, biology, and meteorology, high dimensional matrix-variate data are routinely collected over time. To incorporate the structural interrelations between columns and rows and to achieve significant dimension reduction when dealing with high-dimensional matrix-variate time series, Wang et al. (2017) proposed a matrix factor model that is shown to be effective in analyzing such data. In this paper, we establish a general framework for incorporating domain or prior knowledge induced linear constraints in the matrix-variate factor model. The constraints can be used to achieve parsimony in parameterization, to facilitate interpretation of the latent matrix factor, and to target specific factors of interest based on domain theories. Fully utilizing the constraints results in more efficient and accurate modeling, inference, dimension reduction as well as a clear and better interpretation of the results. In this paper, constrained, multi-term, and partially constrained factor models for matrix-variate time series are developed, with efficient estimation procedures and their asymptotic properties. We show that the convergence rates of the constrained factor loading matrices are much faster than those of the conventional matrix factor analysis under many situations. Simulation studies are carried out to demonstrate performance of the proposed method and the associated asymptotic properties. We demonstrate the proposed model with three applications, where the constrained matrix factor models outperform their unconstrained counterparts in the power of variance explanation under the out-of-sample 10-fold cross-validation setting.

Keywords: Constrained eigen-analysis; Convergence in L2-norm; Dimension reduction; Factor model, Matrix-variate time series.

*The authors gratefully acknowledge *please remember to list all relevant funding sources in the unblinded version*

1 Introduction

High-dimensional matrix-variate time series have been widely observed nowadays in a variety of scientific fields including economics, meteorology, and ecology. For example, the World Bank and the International Monetary Fund collect and publish macroeconomic data of more than thirty variables spanning over one hundred years and over two hundred countries covering a variety of demographic, social, political, and economic topics. These data neatly form a matrix-variate time series with rows representing the countries and columns representing various macroeconomic indices. Typical factor analysis of such data either converts the matrix into a vector or modeling the row or column vectors separately (See Chamberlain (1983), Chamberlain & Rothschild (1983), Bai (2003), Bai & Ng (2002), Bai & Ng (2007), Forni et al. (2000), Forni et al. (2004), Pan & Yao (2008), Lam et al. (2011), and Lam & Yao (2012)). However, the components of matrix-variates are dependent among rows and columns with a well-defined structure. Vectorizing a matrix-valued response, or modeling the row or column vectors separately may overlook some intrinsic dependency and fail to capture the matrix structure. Wang et al. (2017) propose a matrix factor model that maintains and utilizes the matrix structure of the data to achieve significant dimension reduction.

However, in factor analysis of matrix time series and many other types of high-dimensional data, the problem of factor interpretations is of paramount importance. Even more important, in many practical applications, is the problem of obtaining specific latent factors related to certain domain theories, and with the aid of these specific factors further predicting future values of interest. For example, financial researchers may be interested in extracting the latent factors of level, slope, and curvatures of the interest-rate yield curve and predicting future equity prices based on those factors (Diebold et al. (2005), Diebold et al. (2006), Rudebusch & Wu (2008), and Bansal et al. (2014)).

In many applications, relevant prior or domain knowledge is available or data themselves exhibit certain specific structure. Additional covariates may also have been measured. For example, in business and economic forecasting, sector or group information of variables under study is often available. Such *a-priori* information can be incorporated to improve the accuracy and inference of the analysis and to produce more parsimonious and interpretable factors. In other cases, the existing domain knowledge may intrigue researchers' interest in some specific factors. The theories and prior experience may provide guidance for specifying the measurable variables related to the specific factors of interest. It is then desirable to constrain the dimension of the factor representation in order to obtain effectively an adequate representation of the collected variables.

To address these important issues and practical needs, we extend the matrix factor model of Wang et al. (2017) to incorporate natural constraints among the column and row variables. Incorporating a-priori information in parameter estimation has been widely used in statistical analysis, such as the constrained maximum likelihood estimation, constrained least squares, and penalized least squares. Constrained maximum likelihood estimation with the parameter space defined by linear or smooth nonlinear constraints have been explored in the literature. Hathaway (1985) applies the constrained maximum likelihood estimation to the problem of mixture normal distributions and shows that the constrained estimation avoids the problems of singularities and spurious maximizers facing an unconstrained estimation. Geyer (1991) proposes a general approach applicable to many models specified by constraints on the parameter space and illustrates his approach with a constrained logistic regression of the incidence of Down's syndrome on maternal age. Penalty methods have also been customarily used to enforce constraints in statistical models including generalized linear models, generalized estimating equations, proportional hazards models, and M-estimators. See, for example, Frank & Friedman (1993), Tibshirani (1996), Liu et al. (2007), Fan & Li (2001), Zou (2006), and Zhang & Lu (2007). The results of these articles show that including the soft constraints as penalizing term enhances the prediction accuracy and improves the interpretation of the resulting statistical model.

For factor models of time series, Tsai & Tsay (2010) and Tsai et al. (2016) impose constraints, constructed by some empirical procedures, that incorporate the inherent data structure, to both the classical and approximate factor models. Their results show that the constraints are useful tools to obtain parsimonious econometric models for forecasting, to simplify the interpretations of common factors, and to reduce the dimension. Motivated by similar concerns, we consider constrained, multi-term, and partially constrained factor models for high-dimensional matrix-variate time series. Our methods differs from Tsai & Tsay (2010) in several aspects. First, we deal with matrix factor model and thus have the flexibility to impose row and column constraints. The interaction between the row and column constraints are explored. Second, we adopt a different set of assumptions for factor model defined in Lam et al. (2011) and Lam & Yao (2012). The matrix-variate time series is decomposed into two parts: a dynamic part driven by a lower-dimensional factor time series and a static part consisting of matrix white noises. Since the white-noise series exhibits no dynamic correlations, the decomposition is unique in the sense that both the dimension of the factor process and the factor loading space are identifiable for a given finite sample size.

The rest of the paper is organized as follows. Section 2 introduces the constrained, multi-term, and partially constrained matrix-variate factor models. Section 3 presents estimation procedures for constrained and partially constrained factor models with different constraints. Section 4 investigates theoretical properties of the estimates. Section 5 presents some simulation results whereas Section 6 contains three applications. Section 7 concludes. All proofs are in the Appendix.

2 The Constrained Matrix Factor Model

For consistency in notation, we adopt the following conventions. A bold capital letter $\boldsymbol{A}$ represents a matrix, a bold lower letter $\boldsymbol{a}$ represents a column vector, and a lower letter a represents a scalar. The j-th column vector and the k-th row vector of the matrix $\boldsymbol{A}$ are denoted by $A_{\cdot j}$ and $A_{k\cdot}$, respectively.

Let $\{\boldsymbol{Y}_t\}_{t=1,\dots,T}$ be a matrix-variate time series, where $\boldsymbol{Y}_t$ is a $p_1 \times p_2$ matrix, that is

$$\boldsymbol{Y}_t = (Y_{\cdot 1,t}, \cdots, Y_{\cdot p_2,t}) = \begin{pmatrix} Y'_{1\cdot,t} \\ \vdots \\ Y'_{p_1\cdot,t} \end{pmatrix} = \begin{pmatrix} y_{11,t} & \cdots & y_{1p_2,t} \\ \vdots & \ddots & \vdots \\ y_{p_1 1,t} & \cdots & y_{p_1 p_2,t} \end{pmatrix}.$$

Wang et al. (2017) propose a factor model for $\boldsymbol{Y}_t$ as

$$\boldsymbol{Y}_t = \boldsymbol{\Lambda} \boldsymbol{F}_t \boldsymbol{\Gamma}' + \boldsymbol{U}_t, \qquad t = 1, 2, \dots, T, \tag{1}$$

where $\boldsymbol{F}_t$ is a $k_1 \times k_2$ unobserved matrix-variate time series of common fundamental factors, $\boldsymbol{\Lambda}$ is a $p_1 \times k_1$ row loading matrix, $\boldsymbol{\Gamma}$ is a $p_2 \times k_2$ column loading matrix, and $\boldsymbol{U}_t$ is a $p_1 \times p_2$ matrix of random errors.

In Model (1), we assume that $vec(\boldsymbol{U}_t) \sim WN(\boldsymbol{0}, \boldsymbol{\Sigma}_e)$ and is independent of the factor process $vec(\boldsymbol{F}_t)$. That is, $\{\boldsymbol{U}_t\}_{t=1,\dots,T}$ is a white noise matrix-variate time series and the common fundamental factors $\boldsymbol{F}_t$ drive all dynamics and co-movement of $\boldsymbol{Y}_t$. $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ reflect the importance of common factors and their interactions. Wang et al. (2017) provide several interpretations of the loading matrices $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$. Essentially, $\boldsymbol{\Lambda}$ ($\boldsymbol{\Gamma}$) can be viewed as the row (column) loading matrix that reflects how each row (column) in $\boldsymbol{Y}_t$ depends on the factor matrix $\boldsymbol{F}_t$. The interaction between the row and column is introduced through the multiplication of these terms.

The definition of common factors in Model (1) is similar to that of Lam et al. (2011). This decomposition facilitates model identification in finite samples and simplifies the procedure of model identification and statistical inference. However, under the definition, both the "common factors" defined in the traditional factor models and the serially correlated idiosyncratic components will be identified as factors. This poses challenges to the interpretation of the estimated factors, which are usually of special interest in many applications.

Moreover, when the dimensions p_1 and p_2 are sufficiently large, interpretation of the estimated common factors $\widehat{\boldsymbol{F}}_t$ becomes difficult because of the uncertainty and dependence involved in the estimates of the loading matrices $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$.

To mitigate th aforementioned difficulties and, more importantly, to incorporate natural and known constraints among the column and row variables, we consider the following constrained and partially constrained matrix factor models.

A *constrained matrix factor model* can be written as

$$\boldsymbol{Y}_t = \boldsymbol{H}_R \boldsymbol{R} \boldsymbol{F}_t \boldsymbol{C}' \boldsymbol{H}_C' + \boldsymbol{U}_t, \tag{2}$$

where $\boldsymbol{H}_R$ and $\boldsymbol{H}_C$ are pre-specified full column-rank $p_1 \times m_1$ and $p_2 \times m_2$ constraint matrices, respectively, and $\boldsymbol{R}$ and $\boldsymbol{C}$ are $m_1 \times k_1$ row loading matrix and $m_2 \times k_2$ column loading matrix, respectively. For meaningful constraints, we assume $k_1 \leq m_1 << p_1$ and $k_2 \leq m_2 << p_2$. Compared with the matrix factor model in (1), we set $\boldsymbol{\Lambda} = \boldsymbol{H}_R \boldsymbol{R}$ and $\boldsymbol{\Gamma} = \boldsymbol{H}_C \boldsymbol{C}$ with $\boldsymbol{H}_R$ and $\boldsymbol{H}_C$ given. The number of parameters in the left loading matrix $\boldsymbol{R}$ is $m_1 k_1$, smaller than $p_1 k_1$ of the unconstrained model. The number of parameters in the column loading matrix $\boldsymbol{C}$ also decreases from $p_2 k_2$ to $m_2 k_2$. The constraint matrices $\boldsymbol{H}_R$ and $\boldsymbol{H}_C$ can be constructed based on prior or domain knowledge of the variables. For example, if $\boldsymbol{H}_R$ consists of orthogonal binary vectors, it represents a classification or grouping of the rows of the observed matrix.

Consider a simplified model with only row constraints $\boldsymbol{Y}_t = \boldsymbol{H}_R \boldsymbol{R} \boldsymbol{F}_t \boldsymbol{C}' + \boldsymbol{U}_t$. If

$$\boldsymbol{H}_R = \begin{bmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \ldots & 0 & 1 & \ldots & 1 \end{bmatrix}', \tag{3}$$

we are effectively imposing the constraint that there are two groups of row variables (say countries) in which the 'row' behavior of each variable in a group is the same. Specifically, the model becomes

$$\boldsymbol{Y}_t^{(1)} = \boldsymbol{R}_1 \boldsymbol{F}_t \boldsymbol{C}' + \boldsymbol{U}_t^{(1)} \quad \text{and} \quad \boldsymbol{Y}_t^{(2)} = \boldsymbol{R}_2 \boldsymbol{F}_t \boldsymbol{C}' + \boldsymbol{U}_t^{(2)}$$

where $\boldsymbol{Y}_t^{(1)}$ consists of the first $p_1^{(1)}$ rows of $\boldsymbol{Y}_t$ – all the countries in the first group, and $\boldsymbol{Y}_t^{(2)}$ consists of the rest of the rows in the second group. In this case, $\boldsymbol{R}_1$ is a $1 \times k_1$ row vector that is common to all rows in the first group $\boldsymbol{Y}_t^{(1)}$. Comparing to the general matrix factor model (2), the constrained model imposes the constraint that the loading matrix $\boldsymbol{\Lambda}$ have the form $\boldsymbol{\Lambda} = [\boldsymbol{R}_1' \cdots \boldsymbol{R}_1' \boldsymbol{R}_2' \cdots \boldsymbol{R}_2']'$. The countries within the same group have the same row loadings. Note that the two groups still share the same factor matrix $\boldsymbol{F}_t$ and the same column loading matrix $\boldsymbol{C}$. The two groups related to the global common factor $\boldsymbol{F}_t$ differently. The smaller loading matrix $\boldsymbol{R}$ of dimension $2 \times m_1$, instead of the unconstrained $p_1 \times m_1$ loading matrix, provides a much simpler interpretation. More complicated constraints can be used. See Appendix A for an illustration of some constraint matrices.

If there are two "distinct" sets of constraints and the factors corresponding to these two sets do not interact, Model (2) can be extended to a *multiple-term matrix factor model* as

$$\boldsymbol{Y}_t = \boldsymbol{H}_{R_1}\boldsymbol{R}_1\boldsymbol{F}_{1t}\boldsymbol{C}_1'\boldsymbol{H}_{C_1}' + \boldsymbol{H}_{R_2}\boldsymbol{R}_2\boldsymbol{F}_{2t}\boldsymbol{C}_2'\boldsymbol{H}_{C_2}' + \boldsymbol{U}_t. \tag{4}$$

For example, countries can be grouped according to their geographic locations, such as European and Asian countries, and also grouped according to their economic characteristics, such as natural resource based and manufacture based economies, and the corresponding factors may not interact with each other.

Note that (4) can be rewritten as (2), with $\boldsymbol{H}_R = \begin{bmatrix}\boldsymbol{H}_{R_1}, & \boldsymbol{H}_{R_2}\end{bmatrix}$, $\boldsymbol{H}_C = \begin{bmatrix}\boldsymbol{H}_{C_1}, & \boldsymbol{H}_{C_2}\end{bmatrix}$,

$$\boldsymbol{R} = \begin{bmatrix}\boldsymbol{R}_1 & 0 \\ 0 & \boldsymbol{R}_2\end{bmatrix}, \boldsymbol{C} = \begin{bmatrix}\boldsymbol{C}_1 & 0 \\ 0 & \boldsymbol{C}_2\end{bmatrix}, \text{ and } \boldsymbol{F}_t = \begin{bmatrix}\boldsymbol{F}_{1t} & 0 \\ 0 & \boldsymbol{F}_{2t}\end{bmatrix}.$$

Hence (4) is a special case of (2) with the strong assumption that the factor matrix is block diagonal. Such a simplification can greatly increase the interpretation of the model.

Remark 1. The pre-specified constraint matrices $\boldsymbol{H}_{R_1}$ and $\boldsymbol{H}_{R_2}$ do not have to be orthogonal. Neither does the pair $\boldsymbol{H}_{C_1}$ and $\boldsymbol{H}_{C_2}$. This is so because of the assumption of low dimensionality of the latent matrix factors. Estimation procedures of Section 3.3 are able to identify the loading matrices and the latent matrix factors if the transformed observations still contain adequate information on the latent matrix factors. The rates of convergence will change as a result of information loss from the estimation procedure to deal with the nonorthogonality of $\boldsymbol{H}_{R_1}$ and $\boldsymbol{H}_{R_2}$. Since we can always transform non-orthogonal constraint matrices to some orthogonal constraint matrices, we shall focus on the case when $\boldsymbol{H}_{R_1}$ and $\boldsymbol{H}_{R_2}$ (or $\boldsymbol{H}_{C_1}$ and $\boldsymbol{H}_{C_2}$) are orthogonal.

In many applications, prior or domain knowledge may not be sufficiently comprehensive or may only provide a partial specification of the constraint matrices. In the above example, it is possible that the countries within a group react to one set of factors the same way, but differently to another set of factors. In such cases, a partially constrained factor model would be more appropriate. Specifically, a *partially constrained matrix factor model* can be written as

$$\boldsymbol{Y}_t = \begin{bmatrix}\boldsymbol{H}_{R_1}\boldsymbol{R}_1 & \boldsymbol{\Lambda}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{F}_{11,t} & \boldsymbol{F}_{12,t} \\ \boldsymbol{F}_{21,t} & \boldsymbol{F}_{22,t}\end{bmatrix}\begin{bmatrix}\boldsymbol{C}_1'\boldsymbol{H}_{C_1}' \\ \boldsymbol{\Gamma}_2'\end{bmatrix} + \boldsymbol{U}_t,$$

where $\boldsymbol{H}_{R_1}$, $\boldsymbol{R}_1$, $\boldsymbol{H}_{C_1}$ and $\boldsymbol{C}_1$ are defined similarly as those in (4). $\boldsymbol{F}_{ij,t}$'s are common matrix factors corresponding to the interactions of the row and column loading space spanned by the columns of $\boldsymbol{H}_R$ and $\boldsymbol{H}_C$ and their complements, $\boldsymbol{\Lambda}_2$ is $p_1 \times q_1$ row loading matrix and $\boldsymbol{\Gamma}_2$ is a $p_2 \times q_2$ column loading matrix. Again, we have $q_1 < p_1$ and $q_2 < p_2$. We further assume that $vec(\boldsymbol{F}_{ij,t})$'s are independent with $vec(\boldsymbol{U}_t)$. $\boldsymbol{H}'_{R_1}\boldsymbol{\Lambda}_2 = \boldsymbol{0}$ and $\boldsymbol{H}'_{C_1}\boldsymbol{\Gamma}_2 = \boldsymbol{0}$, because all the row loadings that are in the space of $\boldsymbol{H}_{R_1}$ and all the column loadings that are in the space of $\boldsymbol{H}_{C_1}$ could be absorbed into the first parts of loading matrices. Thus, we could explicitly rewrite the model as

$$\boldsymbol{Y}_t = \begin{bmatrix} \boldsymbol{H}_{R_1}\boldsymbol{R}_1 & \boldsymbol{H}_{R_2}\boldsymbol{R}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{F}_{11,t} & \boldsymbol{F}_{12,t} \\ \boldsymbol{F}_{21,t} & \boldsymbol{F}_{22,t} \end{bmatrix} \begin{bmatrix} \boldsymbol{C}'_1\boldsymbol{H}'_{C_1} \\ \boldsymbol{C}'_2\boldsymbol{H}'_{C_2} \end{bmatrix} + \boldsymbol{U}_t, \tag{5}$$

where $\boldsymbol{H}_{R_2}$ is a $p_1 \times (p_1 - m_1)$ constraint matrix satisfying $\boldsymbol{H}'_{R_1}\boldsymbol{H}_{R_2} = \boldsymbol{0}$, $\boldsymbol{H}_{C_2}$ is a $p_2 \times (p_2 - m_2)$ constraint matrix satisfying $\boldsymbol{H}'_{C_1}\boldsymbol{H}_{C_2} = \boldsymbol{0}$, $\boldsymbol{R}_2$ is $(p_1 - m_1) \times q_1$ row loading matrix, and $\boldsymbol{C}_2$ is a $(p_2 - m_2) \times q_2$ column loading matrix.

In the special case when $\boldsymbol{F}_{21,t} = \boldsymbol{0}$ and $\boldsymbol{F}_{12,t} = \boldsymbol{0}$, model (5) can be further simplified as

$$\boldsymbol{Y}_t = \boldsymbol{H}_{R_1}\boldsymbol{R}_1\boldsymbol{F}_{11,t}\boldsymbol{C}'_1\boldsymbol{H}'_{C_1} + \boldsymbol{H}_{R_2}\boldsymbol{R}_2\boldsymbol{F}_{22,t}\boldsymbol{C}'_2\boldsymbol{H}'_{C_2} + \boldsymbol{U}_t. \tag{6}$$

Model (6) is different from the multi-term model of (4) in that the matrix $\boldsymbol{H}_{R_2}$ in (5) is induced from $\boldsymbol{H}_{R_1}$ while $\boldsymbol{H}_{R_2}$ in (4) is an informative constraint, with a lower dimension.

In the special case when $\boldsymbol{H}_{C_1} = \boldsymbol{I}_{p_1}$ (there is no column constraint), model (5) becomes

$$\boldsymbol{Y}_t = \begin{bmatrix} \boldsymbol{H}_{R_1}\boldsymbol{R}_1 & \boldsymbol{H}_{R_2}\boldsymbol{R}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{F}_{1,t} \\ \boldsymbol{F}_{2,t} \end{bmatrix} \boldsymbol{C}' + \boldsymbol{U}_t,$$

where $\boldsymbol{F}_{1,t} = [\boldsymbol{F}_{11,t}, \boldsymbol{F}_{12,t}]$ and $\boldsymbol{F}_{2,t} = [\boldsymbol{F}_{21,t}, \boldsymbol{F}_{22,t}]$. The left loading matrix still spans the entire p_1 dimensional space, but the first part of loading matrix $\boldsymbol{R}_1$ has a clearer interpretation.

The partially constrained matrix factor model (5) incorporates partial information $\boldsymbol{H}_{R_1}$ and $\boldsymbol{H}_{C_1}$ in the unconstrained model (1) without ignoring the possible remainders. If we include all four matrix factors in the four subspaces divided by the interactions of $\boldsymbol{H}_{R_1}$ and $\boldsymbol{H}_{C_1}$ and their complements, the number of parameters in (5) is the same as that in the unconstrained model (1). However, as shown by the theorems in Section 4, the rates of convergence are much faster than those of the unconstrained matrix factor model. Furthermore, in most applications, inclusion of only two matrix-factor terms is adequate in explaining high percentage of variability, as exemplified by the three applications in Section 6.

The benefits of partially constrained matrix factor models are two-folds. Firstly, it is capable of picking up, from the complement space of $\boldsymbol{H}_R$ and $\boldsymbol{H}_C$, the factors that are unknown to researchers. In this case, the dimensions of $\boldsymbol{F}_{22,t}$ are typically much smaller than those of $\boldsymbol{F}_{11,t}$ even though the loading matrices $\boldsymbol{R}_2$ and $\boldsymbol{C}_2$ still have large numbers of rows $(p_1 - m_1)$ and columns $(p_2 - m_2)$, respectively, since the constraint part should have accommodated the main and key common factors. The spirit is similar to the two-step estimation of Lam & Yao (2012) in which one fits a second-stage factor model to the residuals obtained by subtracting the common part of the first-stage factor model.

The second benefit is that the partially constrained matrix factor model is able to identify matrix factors whose dimensions are completely explained by the pre-specified constraint matrices. Specifically, $\boldsymbol{F}_{11,t}$ represents the factor matrix with row and column factors affecting the observed matrix-variate in the way as specified by the constraints $\boldsymbol{H}_R$ and $\boldsymbol{H}_C$ completely. Consider the multinational macroeconomic index example. If $\boldsymbol{H}_R$ is built from the country classification information, how the rows in $\boldsymbol{F}_{11,t}$ affect the observations can be completely explained by the country groups instead of individual countries and the row factors in $\boldsymbol{F}_{11,t}$ have a clearer interpretation related to the classification. In many practical applications, researchers are interested in obtaining specific latent factors related to some domain theories and use these specific factors to predict future values of interest as guided by domain theories. For example, in the yield curve example of Appendix A, economic theory implies that the level, slope, and curvature factors affect the observations in the way specified by, for example, $\boldsymbol{H}_R = [\boldsymbol{h}_1, \boldsymbol{h}_2, \boldsymbol{h}_3]$, where $\boldsymbol{h}_1 = (1, 1, 1, 1, 1)'$, $bh_2 = (1, 1, 0, -1, -1)'$, and $\boldsymbol{h}_3 = (-1, 0, 2, 0, -1)$. Then the estimation method in Section 3 is capable of isolating $\boldsymbol{H}_{R_1}\boldsymbol{R}_1\boldsymbol{F}_{11,t}\boldsymbol{C}_1'\boldsymbol{H}_{C_1}'$ and correctly estimating the loadings and the specified level, slope, and curvature factors in the constrained spaces. Thus, the constrained factor model can serve as a method to identify and isolate specific factors suggested by domain theories or prior knowledge.

3 Estimation Procedure

Similar to all factor models, identification issue exits in the constrained matrix-variate factor model (2). Let $\boldsymbol{O}_1$ and $\boldsymbol{O}_2$ be two invertible matrices of size $k_1 \times k_1$ and $k_2 \times k_2$. Then the triples $(\boldsymbol{R}, \boldsymbol{F}_t, \boldsymbol{C})$ and $(\boldsymbol{R}\boldsymbol{O}_1, \boldsymbol{O}_1^{-1}\boldsymbol{F}_t\boldsymbol{O}_2^{-1}, \boldsymbol{O}_2\boldsymbol{C})$ are equivalent under Model (2). We may assume that the columns of $\boldsymbol{R}$ and $\boldsymbol{C}$ are orthonormal, that is, $\boldsymbol{R}'\boldsymbol{R} = \boldsymbol{I}_{k_1}$ and $\boldsymbol{C}'\boldsymbol{C} = \boldsymbol{I}_{k_2}$, where $\boldsymbol{I}_d$ denotes the $d \times d$ identity matrix. Even with these constraints, $\boldsymbol{R}$, $\boldsymbol{F}_t$ and $\boldsymbol{C}$ are not uniquely determined in (2), as aforementioned replacement is still valid for any orthonormal $\boldsymbol{O}$. However, the column spaces of the loading matrices $\boldsymbol{R}$ and $\boldsymbol{C}$ are uniquely determined. Hence, in the following sections, we focus on the estimation of the column spaces of $\boldsymbol{R}$ and $\boldsymbol{C}$. We denote the row and column factor loading spaces by $\mathcal{M}(\boldsymbol{R})$ and $\mathcal{M}(\boldsymbol{C})$, respectively. For simplicity, we suppress the matrix column space notation and use the matrix notation directly.

3.1 Orthogonal Constraints

We start with the estimation of the constrained matrix-variate factor model (2). The approach follows the ideas of Tsai & Tsay (2010) and Wang et al. (2017). In what follows, we illustrate the estimation procedure for the column space of $\boldsymbol{R}$. The column space of $\boldsymbol{C}$ can be obtained similarly from the transpose of $\boldsymbol{Y}_t$'s.

Suppose we have orthogonal constraints $\boldsymbol{H}_R'\boldsymbol{H}_R = \boldsymbol{I}_{m_1}$ and $\boldsymbol{H}_C'\boldsymbol{H}_C = \boldsymbol{I}_{m_2}$. Define the transformation $\boldsymbol{X}_t = \boldsymbol{H}_R'\boldsymbol{Y}_t\boldsymbol{H}_C$. It follows from (2) that

$$\boldsymbol{X}_t = \boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}' + \boldsymbol{E}_t, \qquad t = 1, 2, \ldots, T, \tag{7}$$

where $\boldsymbol{E}_t = \boldsymbol{H}_R'\boldsymbol{U}_t\boldsymbol{H}_C$.

This transformation projects the observed matrix time series into the constrained space. For example, if $\boldsymbol{H}_R$ is the orthonormal matrix corresponding to the group constraint in (3), then $\boldsymbol{H}_R'\boldsymbol{Y}_t$ is a $2 \times p_2$ matrix, with the first row being the normalized average of the rows of $\boldsymbol{Y}_t$ in the first group and the second row being that in the second group. Such an operation conveniently incorporates the constraints while reduces the dimension of data matrix from $p_1 \times p_2$ to $m_1 \times m_2$, making the analysis more efficient.

Since $\boldsymbol{E}_t$ remains a white noise process, the estimation method in Wang et al. (2017) directly applies to the transformed $m_1 \times m_2$ matrix time series $\boldsymbol{X}_t$ in model (7). For completeness, we outline briefly the procedure. See Wang et al. (2017) for details.

To facilitate the estimation, we use the QR decomposition $\boldsymbol{R} = \boldsymbol{Q}_1\boldsymbol{W}_1$ and $\boldsymbol{C} = \boldsymbol{Q}_2\boldsymbol{W}_2$ to normalize the loading matrices, so that model (7) can be re-expressed as

$$\boldsymbol{X}_t = \boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}' + \boldsymbol{E}_t = \boldsymbol{Q}_1\boldsymbol{Z}_t\boldsymbol{Q}_2' + \boldsymbol{E}_t, \qquad t = 1, 2, \ldots, T, \tag{8}$$

where $\boldsymbol{Z}_t = \boldsymbol{W}_1\boldsymbol{F}_t\boldsymbol{W}_2'$, $\boldsymbol{Q}_1'\boldsymbol{Q}_1 = \boldsymbol{I}_{m_1}$ and $\boldsymbol{Q}_2'\boldsymbol{Q}_2 = \boldsymbol{I}_{m_2}$.

We assume that both $\boldsymbol{F}_t$ and $\boldsymbol{E}_t$ are zero mean and thus $E(X_{t,\cdot j}) = 0$. Let h be a positive integer. For $i, j = 1, 2, \ldots, m_2$, define

$$\boldsymbol{\Omega}_{zq,ij}(h) = \frac{1}{T-h}\sum_{t=1}^{T-h} Cov(\boldsymbol{Z}_t Q_{2,i\cdot}, \boldsymbol{Z}_{t+h} Q_{2,j\cdot}), \text{ and} \tag{9}$$

$$\boldsymbol{\Omega}_{x,ij}(h) = \frac{1}{T-h}\sum_{t=1}^{T-h} Cov(X_{t,\cdot i}, X_{t+h,\cdot j}), \tag{10}$$

which can be interpreted as the auto-cross-covariance matrices at lag h between column i and column j of $\{\boldsymbol{Z}_t\boldsymbol{Q}_2'\}_{t=1,\cdots,T}$ and $\{\boldsymbol{X}_t\}_{t=1,\cdots,T}$, respectively. For $h > 0$, both terms do not involve $\boldsymbol{E}_t$ due to the whiteness condition.

For a fixed $h_0 \geq 1$ satisfying Condition 2 in Section 4, define

$$\boldsymbol{M} = \sum_{h=1}^{h_0} \sum_{i=1}^{m_2} \sum_{j=1}^{m_2} \boldsymbol{\Omega}_{x,ij}(h) \boldsymbol{\Omega}_{x,ij}(h)' = \boldsymbol{Q}_1 \left\{ \sum_{h=1}^{h_0} \sum_{i=1}^{m_2} \sum_{j=1}^{m_2} \boldsymbol{\Omega}_{zq,ij}(h) \boldsymbol{\Omega}_{zq,ij}(h)' \right\} \boldsymbol{Q}_1'. \tag{11}$$

Under Condition 2 the column space of $\boldsymbol{M}$ is the same as that of $\boldsymbol{Q}_1$, and the columns of the factor loading matrix $\boldsymbol{Q}_1$ can be obtained as the k_1 orthogonal eigenvectors of the matrix $\boldsymbol{M}$ corresponding to its k_1 non-zero eigenvalues arranged in the descending order. Suppose we have centered the transformed observations $\{\mathbf{X}_t\}_{t=1,\ldots,T}$, then for $h \geq 1$ and a prescribed positive integer h_0, define the sample version of $\boldsymbol{M}$ in (11) as the following

$$\widehat{\mathbf{M}} = \sum_{h=1}^{h_0} \sum_{i=1}^{m_2} \sum_{j=1}^{m_2} \widehat{\boldsymbol{\Omega}}_{x,ij}(h) \widehat{\boldsymbol{\Omega}}_{x,ij}(h)', \text{ where } \quad \widehat{\boldsymbol{\Omega}}_{x,ij}(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} X_{t,\cdot i} X'_{t+h,\cdot j}. \tag{12}$$

A natural estimator for $\boldsymbol{Q}_1$ is $\widehat{\boldsymbol{Q}}_1 = \left\{ \widehat{\boldsymbol{q}}_1, \cdots, \widehat{\boldsymbol{q}}_{k_1} \right\}$, where $\widehat{\boldsymbol{q}}_i$ is an eigenvector of $\widehat{\boldsymbol{M}}$, corresponding to its i-th largest eigenvalue. Consequently, we estimate the normalized factors and residuals, respectively, by $\widehat{\boldsymbol{Z}}_t = \widehat{\boldsymbol{Q}}_1' \boldsymbol{X}_t \widehat{\boldsymbol{Q}}_2$ and $\widehat{\boldsymbol{U}}_t = \boldsymbol{Y}_t - \widehat{\boldsymbol{H}}_R \widehat{\boldsymbol{Q}}_1 \widehat{\boldsymbol{Z}}_t \widehat{\boldsymbol{Q}}_2' \widehat{\boldsymbol{H}'}_C$.

The above estimation procedure assumes that the number of row factors k_1 is known. To determine k_1, Wang et al. (2017) used the eigenvalue ratio-based estimator of Lam & Yao (2012). Let $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq \widehat{\lambda}_{m_1} \geq 0$ be the ordered eigenvalues of $\widehat{\boldsymbol{M}}$. The ratio-based estimator for k_1 is defined as

$$\widehat{k}_1 = \arg \min_{1 \leq j \leq K} \frac{\widehat{\lambda}_{j+1}}{\widehat{\lambda}_j},$$

where $k_1 \leq K \leq p_1$ is an integer. In practice we may take $K = p_1/2$.

Although the estimation procedure on the transformed series $\boldsymbol{X}_t$ is exactly the same as that of Wang et al. (2017), the asymptotic properties of the estimator is different due to the transformation, as shown in Section 4.

3.2 Nonorthogonal Constraints

If the constraint matrix $\boldsymbol{H}_R$ (or $\boldsymbol{H}_C$) is not orthogonal, we can perform column orthogonalization and standardization, similar to that in Tsai & Tsay (2010). Specifically, we obtain

$$\boldsymbol{H}_R = \boldsymbol{\Theta}_R \boldsymbol{K}_R,$$

where $\boldsymbol{\Theta}_R$ is an orthonormal matrix and $\boldsymbol{K}_R$ is a $m_1 \times m_1$ upper triangular matrix with nonzero diagonal elements. $\boldsymbol{H}_C = \boldsymbol{\Theta}_C \boldsymbol{K}_C$ can be obtained in the same way.

Letting $\boldsymbol{X}_t = \boldsymbol{\Theta}_R' \boldsymbol{Y}_t \boldsymbol{\Theta}_C$, $\boldsymbol{R}^* = \boldsymbol{K}_R \boldsymbol{R}$ and $\boldsymbol{C}^* = \boldsymbol{K}_C \boldsymbol{C}$, we have

$$\boldsymbol{X}_t = \boldsymbol{R}^* \boldsymbol{F}_t \boldsymbol{C}^{*'} + \boldsymbol{E}_t, \qquad t = 1, 2, \ldots, T, \tag{13}$$

where $\boldsymbol{E}_t = \boldsymbol{\Theta}_R' \boldsymbol{U}_t \boldsymbol{\Theta}_C$. Since $\boldsymbol{E}_t$ remains a white noise process, we apply the same estimation method in Section 3.1 to obtain the estimates $\widehat{\boldsymbol{R}}^*$ and $\widehat{\boldsymbol{C}}^*$. The estimates of $\boldsymbol{R}$ and $\boldsymbol{C}$ are $\widehat{\boldsymbol{R}} = \boldsymbol{K}_R^{-1}\widehat{\boldsymbol{R}}^*$ and $\widehat{\boldsymbol{C}} = \boldsymbol{K}_C^{-1}\widehat{\boldsymbol{C}}^*$. Note that $\boldsymbol{K}_R$ and $\boldsymbol{K}_C$ are invertible lower triangular matrices.

3.3 Multi-term Constrained Matrix Factor Model

Without loss of generality, we assume that both row and column constraint matrices are orthogonal matrices. If $\boldsymbol{H}_{R_1}$ and $\boldsymbol{H}_{R_2}$ (or $\boldsymbol{H}_{C_1}$ and $\boldsymbol{H}_{C_2}$) are orthogonal, we obtain, for $t = 1, 2, \ldots, T$,

$$\begin{aligned} \boldsymbol{H}_{R_1}'\boldsymbol{Y}_t\boldsymbol{H}_{C_1} &= \boldsymbol{R}_1\boldsymbol{F}_{1,t}\boldsymbol{C}_1' + \boldsymbol{H}_{R_1}'\boldsymbol{U}_t\boldsymbol{H}_{C_1}, \\ \boldsymbol{H}_{R_2}'\boldsymbol{Y}_t\boldsymbol{H}_{C_2} &= \boldsymbol{R}_2\boldsymbol{F}_{2,t}\boldsymbol{C}_2' + \boldsymbol{H}_{R_2}'\boldsymbol{U}_t\boldsymbol{H}_{C_2}, \end{aligned}$$

where $\boldsymbol{H}_{R_1}'\boldsymbol{U}_t\boldsymbol{H}_{C_1}$ and $\boldsymbol{H}_{R_2}'\boldsymbol{U}_t\boldsymbol{H}_{C_2}$ are white noises. The estimators of $\widehat{\boldsymbol{R}}_1$, $\widehat{\boldsymbol{C}}_1$, $\widehat{\boldsymbol{F}}_{1,t}$, $\widehat{\boldsymbol{R}}_2$, $\widehat{\boldsymbol{C}}_2$ and $\widehat{\boldsymbol{F}}_{2,t}$ can be obtained by applying the estimation procedure described in Section 3.1 to $\boldsymbol{H}_{R_1}'\boldsymbol{Y}_t\boldsymbol{H}_{C_1}$ and $\boldsymbol{H}_{R_2}'\boldsymbol{Y}_t\boldsymbol{H}_{C_2}$, respectively.

Remark 2. For multi-term constrained model (4), $\boldsymbol{H}_{R_1}$ and $\boldsymbol{H}_{R_2}$ (or $\boldsymbol{H}_{C_1}$ and $\boldsymbol{H}_{C_2}$) may not necessarily be orthogonal. Under this situation, we illustrate the estimation procedure for the column loadings, while the row loading estimators for $\widehat{\boldsymbol{R}}_1$ and $\widehat{\boldsymbol{R}}_2$ can be obtained from the same procedure applied to the transpose of $\boldsymbol{Y}_t$. Define projection matrices $\boldsymbol{P}_{\boldsymbol{H}_{R_1}^\perp} = \boldsymbol{I} - \boldsymbol{H}_{R_1}\boldsymbol{H}_{R_1}'$ and $\boldsymbol{P}_{\boldsymbol{H}_{R_2}^\perp} = \boldsymbol{I} - \boldsymbol{H}_{R_2}\boldsymbol{H}_{R_2}'$, which represent the projections onto the spaces perpendicular to the column spaces of $\boldsymbol{H}_{R_1}$ and $\boldsymbol{H}_{R_2}$, respectively. Left multiplying equations (4) by $\boldsymbol{P}_{\boldsymbol{H}_{R_2}^\perp}$ and $\boldsymbol{P}_{\boldsymbol{H}_{R_1}^\perp}$, respectively, and taking transpose of the resulting matrices, we have $\boldsymbol{Y}_t'\boldsymbol{P}_{\boldsymbol{H}_{R_2}^\perp} = \boldsymbol{H}_{C_1}\boldsymbol{C}_1\boldsymbol{F}_{1,t}'\boldsymbol{R}_1'\boldsymbol{H}_{R_1}'\boldsymbol{P}_{\boldsymbol{H}_{R_2}^\perp} + \boldsymbol{U}_t'\boldsymbol{P}_{\boldsymbol{H}_{R_2}^\perp}$ and $\boldsymbol{Y}_t'\boldsymbol{P}_{\boldsymbol{H}_{R_1}^\perp} = \boldsymbol{H}_{C_2}\boldsymbol{C}_2\boldsymbol{F}_{2,t}'\boldsymbol{R}_2'\boldsymbol{H}_{R_2}'\boldsymbol{P}_{\boldsymbol{H}_{R_1}^\perp} + \boldsymbol{U}_t'\boldsymbol{P}_{\boldsymbol{H}_{R_1}^\perp}$, where $\boldsymbol{P}_{\boldsymbol{H}_{R_2}^\perp}\boldsymbol{U}_t$ and $\boldsymbol{P}_{\boldsymbol{H}_{R_1}^\perp}\boldsymbol{U}_t$ are white noises. The column loading estimators $\widehat{\boldsymbol{C}}_1$ and $\widehat{\boldsymbol{C}}_2$ can be obtained by applying the procedure described in Section 3.1 to $\boldsymbol{H}_{C_1}'\boldsymbol{Y}_t'\boldsymbol{P}_{\boldsymbol{H}_{R_2}^\perp}$ and $\boldsymbol{H}_{C_2}'\boldsymbol{Y}_t'\boldsymbol{P}_{\boldsymbol{H}_{R_1}^\perp}$, respectively. Note that the $p_1 \times m_1$ matrix $\boldsymbol{P}_{\boldsymbol{H}_{R_2}^\perp}\boldsymbol{H}_{R_1}$ is no longer full rank or orthonormal. However, the row and column loading spaces and latent factors can be fully recovered if the dimension of the reduced constrained loading spaces still larger than the dimensions of the latent factor spaces. However, the rates of convergence will change. For example, the rate of convergence of $\widehat{\boldsymbol{C}}_1$ will depend on $\|\boldsymbol{P}_{\boldsymbol{H}_{R_2}^\perp}\boldsymbol{H}_{R_1}\boldsymbol{R}_1\|_2^2$ instead of $\|\boldsymbol{H}_{R_1}\boldsymbol{R}_1\|_2^2$.

3.4 Partially Constrained Matrix Factor Model

For the partially constrained matrix factor model (5), we assume that $\boldsymbol{H}_{R_1}'\boldsymbol{H}_{R_2} = \boldsymbol{0}$ and $\boldsymbol{H}_{C_1}'\boldsymbol{H}_{C_2} = \boldsymbol{0}$. Define the transformation $\boldsymbol{X}_t^{(lk)} = \boldsymbol{H}_{R_l}'\boldsymbol{Y}_t\boldsymbol{H}_{C_k}$ for $l, k = 1, 2$. Then the transformed data follow the structure,

$$\boldsymbol{X}_t^{(lk)} = \boldsymbol{R}_l \boldsymbol{F}_{lk,t} \boldsymbol{C}_k' + \boldsymbol{E}_t^{(lk)}, \quad l,k = 1,2,$$

where $\boldsymbol{E}_t^{(lk)} = \boldsymbol{H'}_{R_l} \boldsymbol{U}_t \boldsymbol{H}_{C_k}$ remains white noise processes.

Let $\boldsymbol{M}^{(lk)}$ represent the $\boldsymbol{M}$ matrix defined in (11) for each $\boldsymbol{X}_t^{(lk)}$, $l,k = 1,2$. Define $\boldsymbol{M}^{(l\cdot)} = \sum_{k=1}^{2} \boldsymbol{M}^{(lk)}$ for $l = 1,2$, then

$$\boldsymbol{M}^{(l\cdot)} = \boldsymbol{Q}_1^{(l)} \left\{ \sum_{k=1}^{2} \sum_{h=1}^{h_0} \sum_{i=1}^{m_2} \sum_{j=1}^{m_2} \boldsymbol{\Omega}_{zq,ij}^{(lk)}(h) \boldsymbol{\Omega}_{zq,ij}^{(lk)}(h)' \right\} \boldsymbol{Q}_1^{(l)\prime}, \quad l = 1,2, \tag{14}$$

has the same column space as that of $\boldsymbol{R}_l$, for $l = 1,2$, respectively.

The estimators of $\widehat{\boldsymbol{R}}_l$, $l = 1,2$, can be obtained by applying eigen-decomposition on the sample version of $\boldsymbol{M}^{(l\cdot)}$ defined similarly to (12). $\boldsymbol{C}_k$, $k = 1,2$, can be obtained by using the same procedure on the transposes of $\boldsymbol{X}_t^{(lk)}$ for $l,k = 1,2$. In the special case of model (6) if $\boldsymbol{F}_{21,t} = \boldsymbol{0}$ and $\boldsymbol{F}_{12,t} = \boldsymbol{0}$, the above estimation is essentially the same procedure as those described in Section 3.1 applying to $\boldsymbol{X}_t^{(ll)}$ for $l = 1,2$.

This procedure effectively projects the observed matrix time series $\boldsymbol{Y}_t$ into four orthogonal subspaces, based on the constraints obtained from the domain knowledge or some empirical procedure. Because $\boldsymbol{X}_t^{(lk)}, l,k = 1,2$ are orthogonal, they can be analyzed separately. In our setting, we divide a $p_1 \times p_1$ row loading matrix space into two orthogonal $p_1 \times m_1$ and $p_1 \times (p_1 - m_1)$ subspaces. The estimation procedure for the partially constrained model ensures the structural requirement that $\boldsymbol{X}_t^{(l1)}$ and $\boldsymbol{X}_t^{(l2)}$ share the same row loading matrix for the same l without sacrificing the dimension reduction benefit from column space division. More generally, we could divide the space of loading matrix into more than two parts to accommodate each application. Under this partially constrained model, the orthogonality assumption between $\boldsymbol{F}_{lk,t}, l,k = 1,2$ is not important as all are latent variables.

Remark 3. In situations when the prior or domain knowledge captures most major factors, we know further that m_i grows slower than p_i and the row (column) factor strength of $\boldsymbol{F}_{11,t}$ is no weaker than that of $\boldsymbol{F}_{22,t}$. Improved estimators of $\hat{\boldsymbol{R}}_l$, $l = 1,2$, can be obtained by applying eigen-decomposition on the sample version of $\boldsymbol{M}^{(l1)}$ defined similarly to (12). Improved estimators of $\hat{\boldsymbol{C}}_k$, $k = 1,2$, can be obtained by using the same procedure on the transposes of $\boldsymbol{X}_t^{(1k)}$ for $k = 1,2$. Here, the estimation procedure discards the noisy part in (14) and results in improved estimators.

4 Theoretical Properties

In this section, we present the rates of convergence for the estimators under the setting that p_1, p_2, m_1, m_2 and T all go to infinity while k_1 and k_2 are fixed and the factor structure does not change over time. Note that the factor decomposition (2) is practically useful only when $k_1 << p_1$ or $k_2 << p_2$.

The asymptotic convergence rates are significantly different from those in Wang et al. (2017) due to the constraints. The results reveal more clearly the impact of the constraints on signals and noises and the interaction between them. We only consider the case of the orthogonal constrained model (2). Asymptotic properties of nonorthogonal, multi-term, and partially constrained matrix factor model are trivial extensions.

Several regularity conditions (Conditions 1 to 5) are listed in the Appendix. They are similar to those in Wang et al. (2017) and are used to derive the limiting behavior of (12) towards its population version. The following condition requires some discussion.

Condition 6. Factor Strength. There exist constants δ_1 and δ_2 in $[0, 1]$ such that $\|\boldsymbol{H}_R\boldsymbol{R}\|_2^2 \asymp p_1^{1-\delta_1} \asymp \|\boldsymbol{H}_R\boldsymbol{R}\|_{min}^2$ and $\|\boldsymbol{H}_C\boldsymbol{C}\|_2^2 \asymp p_2^{1-\delta_2} \asymp \|\boldsymbol{H}_C\boldsymbol{C}\|_{min}^2$.

Since only $\boldsymbol{Y}_t$ is observed in model (2), how well we can recover the factor $\boldsymbol{F}_t$ from $\boldsymbol{Y}_t$ depends on the 'factor strength' reflected by the coefficients in the row and column factor loading matrices $\boldsymbol{H}_R\boldsymbol{R}$ and $\boldsymbol{H}_C\boldsymbol{C}$. For example, in the case of $\boldsymbol{H}_R\boldsymbol{R} = \boldsymbol{0}$ or $\boldsymbol{H}_C\boldsymbol{C} = \boldsymbol{0}$, $\boldsymbol{Y}_t$ carries no information on $\boldsymbol{F}_t$. In the following, we assume $\|\boldsymbol{F}_t\|$ does not change as p_1, p_2, m_1, and m_2 change.

The rates δ_1 and δ_2 in Condition 6 are called the strength for the row factors and the column factors, respectively. If $\delta_1 = 0$, the corresponding row factors are called strong factors because the case includes situation where each element of the row loading matrix is $O(1)$, implying that the factors have impacts on the majority of p_1 vector time series. The amount of information that observed process $\boldsymbol{Y}_t$ carries about the strong factors increases at the same rate as the number of observations or the amount of noise increases. If $\delta_1 > 0$, the row factors are weak, which means the information contained in $\boldsymbol{Y}_t$ about the factors grows more slowly than the noises introduced as p_1 increases. The smaller the $\delta' s$, the stronger the factors. In the strong factor case, the loading matrix is dense. See Lam et al. (2011) for further discussions.

If we restrict $\boldsymbol{H}_R$ to be orthonormal, $||\boldsymbol{H}_R\boldsymbol{R}||_2^2 = ||\boldsymbol{R}||_2^2 \asymp p_1^{1-\delta_1}$ and there is an interplay between $\boldsymbol{H}_R$ and $\boldsymbol{R}$ as p_1 increases. In order for $\boldsymbol{H}_R$ to remain orthonormal, when p_1 increases, each element of $\boldsymbol{H}_R$ decreases at the rate of $p_1^{-1/2}$. At the same time, each element of $\boldsymbol{R}$ on average increases $\sqrt{p_1^{1-\delta_1}/m_1}$. The column factor loading $||\boldsymbol{H}_C\boldsymbol{C}||_2^2$ behaves in the same way. As p_1 and p_2 increase, each element of the transformed error $\boldsymbol{E}_t$ remains a growth rate of 1 under Condition 3, but the dimension of $\boldsymbol{E}_t$ is $m_1 \times m_2$ which grows at a slower rate than $p_1 \times p_2$. The factor strength is defined in terms of the observed dimension p_1 and p_2 and the overall loading matrices $\boldsymbol{H}_R\boldsymbol{R}$ and $\boldsymbol{H}_C\boldsymbol{C}$, but clearly how m_1 and m_2 increase with p_1, p_2 is also important because it controls the signal-noise ratio in the constrained model. For example, if $m_i/p_i = c_i < 1$, $i = 1, 2$, that is, the number of members in each group is fixed, then $||\boldsymbol{R}||_2^2||\boldsymbol{C}||_2^2 \asymp m_1^{1-\delta_1}m_2^{1-\delta_2}/c_1^{1-\delta_1}c_2^{1-\delta_2}$, compared to $||\boldsymbol{E}_t||_2^2 \asymp m_1m_2$. If $m_i = p_i^{\alpha_i}$, $\alpha_i < 1$, $i = 1, 2$, then $||\boldsymbol{R}||_2^2||\boldsymbol{C}||_2^2 \asymp m_1^{(1-\delta_1)/\alpha_1}m_2^{(1-\delta_2)/\alpha_2}$ compared to $||\boldsymbol{E}_t||_2^2 \asymp m_1m_2$. Since $c_i < 1$ and $\alpha_i < 1$, the signal-noise ratio is larger than $m_1^{-\delta_1}m_2^{-\delta_2}$, which is the signal-noise ratio of a unconstrained matrix factor model when $p_1 = m_1$ and $p_2 = m_2$.

We have the following theorems for the constrained matrix factor model. Asymptotic properties for the multi-term and the partially constrained matrix factor models are similar and can be derived easily.

Theorem 1. *Under Conditions 1-6 and* $m_1p_1^{-1+\delta_1}m_2p_2^{-1+\delta_2}T^{-1/2} = o_p(1)$, *as* m_1, p_1, m_2, p_2, *and* T *go to* ∞, *it holds that*

$$\begin{aligned} \|\widehat{\boldsymbol{Q}}_1 - \boldsymbol{Q}_1\|_2 &= O_p\left(\max\left(T^{-1/2},\ \frac{m_1}{p_1^{1-\delta_1}}\frac{m_2}{p_2^{1-\delta_2}}T^{-1/2}\right)\right), \\ \|\widehat{\boldsymbol{Q}}_2 - \boldsymbol{Q}_2\|_2 &= O_p\left(\max\left(T^{-1/2},\ \frac{m_1}{p_1^{1-\delta_1}}\frac{m_2}{p_2^{1-\delta_2}}T^{-1/2}\right)\right). \end{aligned}$$

Remark 4. In the unconstrained case, Wang et al. (2017) obtained the rate of $p_1^{\delta_1}p_2^{\delta_2}T^{-1/2}$. The difference between the two models is of the order of $\frac{m_1m_2}{p_1p_2}$.

The rate of convergence in Theorem 1 depends on the relative growth rate of m_1m_2 and $p_1^{1-\delta_1}p_2^{1-\delta_2}$. We achieve $\|\widehat{\boldsymbol{Q}}_i - \boldsymbol{Q}_i\|_2 = O_p\left(T^{-1/2}\right)$ not only in the strong factor case but also in the weak factor case (i.e. $\delta > 0$) so long as m_1m_2 increases at a slower rate than or equal to that of $p_1^{1-\delta_1}p_2^{1-\delta_2}$, e.g. $m_1m_2 \sim O_p(p_1^{1-\delta_1}p_2^{1-\delta_2})$. The improvement on the convergence rate for the weak factor case results from the fact that the constraints effectively reduce the dimension of the data matrix from $p_1 \times p_2$ to $m_1 \times m_2$ whereas the error magnitude remains the same.

For fixed dimensions of the constrained row and column loading spaces m_1 and m_2, we have more and more observations in each group and the convergence rate is $p_1^{\delta_1-1}p_2^{\delta_2-1}T^{-1/2}$. Here, increases in p_1 or p_2 improve the convergence rate. This is because, if constraints are properly specified, the additional information introduced from increasing p_1 or p_2 will accrue and translate into the transformed signal part in (7), while the transformed noise part does not accumulate. In a sense, constraints allow signals to pass through while filtering out the noises.

If $m_1 = p_1^{\alpha_1}$ and $m_2 = p_2^{\alpha_2}$, the rate is $p_1^{\delta_1+\alpha_1-1} p_2^{\delta_2+\alpha_2-1} T^{-1/2}$, and we achieve a better rate than that of the unconstrained case if $\alpha_1 < 1$ or $\alpha_2 < 1$. If $m_1 = c_1 p_1$ and $m_2 = c_2 p_2$, that is, the dimensions of the constrained loading spaces increase with p's linearly, then the rate is the same as that of the unconstrained model.

Remark 5. The strengths of row factors and column factors δ_1 and δ_2 determine the convergence rate jointly. An increase in the strength of row factors is able to improve the estimation of the column factors loading space and vice versa.

Theorem 2. *Under Conditions 1-6, and if $m_1 p_1^{-1+\delta_1} m_2 p_2^{-1+\delta_2} T^{-1/2} = o_p(1)$ and the $\boldsymbol{M}$ matrix has k_1 distinct positive eigenvalues, then the eigenvalues $\{\hat{\lambda}_1, \ldots, \hat{\lambda}_{m_1}\}$ of $\widehat{\boldsymbol{M}}$, sorted by the descending order, satisfy*

$$\begin{aligned} |\hat{\lambda}_j - \lambda_j| &= O_p\left(\max\left(p_1^{2-2\delta_1} p_2^{2-2\delta_2},\ m_1 p_1^{1-\delta_1} m_2 p_2^{1-\delta_2}\right) \cdot T^{-1/2}\right), \quad for \quad j = 1, 2, \ldots, k_1, \\ |\hat{\lambda}_j| &= O_p\left(\max\left(p_1^{2-2\delta_1} p_2^{2-2\delta_2},\ m_1^2 m_2^2\right) \cdot T^{-1}\right), \quad for \quad j = k_1+1, \ldots, m_1, \end{aligned}$$

where $\lambda_1 > \lambda_2 > \cdots > \lambda_{m_1}$ are the eigenvalues of $\boldsymbol{M}$.

Theorem 2 shows that the estimators for the nonzero eigenvalues of $\boldsymbol{M}$ converge more slowly than those for the zero eigenvalues. This provides the theoretical support for the ratio-based estimator of the number of factors described in Section 3.1, similar to that in Lam et al. (2011). The assumption that $\boldsymbol{M}$ has k_1 distinct positive eigenvalues is not essential, yet it substantially simplifies the presentation and the proof of the convergence properties.

In the cases of strong factors or wake factors with $m_1 m_2 \sim o_p(p_1^{1-\delta_1} p_2^{1-\delta_2})$, our result is the same as that of Wang et al. (2017). When the factors are weak and $p_1^{1-\delta_1} p_2^{1-\delta_2} \sim O_p(m_1 m_2)$, the gap between the convergence rates of nonzero and zero eigenvalues of $\boldsymbol{M}$ is larger in the constrained case.

Let $\boldsymbol{S}_t$ be the dynamic signal part of $\boldsymbol{Y}_t$, i.e. $\boldsymbol{S}_t = \boldsymbol{H}_R \boldsymbol{R} \boldsymbol{F}_t \boldsymbol{C}' \boldsymbol{H}_C' = \boldsymbol{H}_R \boldsymbol{Q}_1 \boldsymbol{Z}_t \boldsymbol{Q}_2' \boldsymbol{H}_C'$. From the discussion in Section 3.1, $\boldsymbol{S}_t$ can be estimated by

$$\widehat{\boldsymbol{S}}_t = \boldsymbol{H}_R \widehat{\boldsymbol{Q}}_1 \widehat{\boldsymbol{Z}}_t \widehat{\boldsymbol{Q}}_2' \boldsymbol{H}_C'.$$

Some theoretical properties of $\widehat{\boldsymbol{S}}_t$ are given below:

Theorem 3. *Under Conditions 1-6 and $m_1 p_1^{-1+\delta_1} m_2 p_2^{-1+\delta_2} T^{-1/2} = o_p(1)$, we have*

$$\frac{1}{\sqrt{p_1 p_2}} \|\widehat{\boldsymbol{S}}_t - \boldsymbol{S}_t\|_2 = O_p\left(\max\left(p_1^{-\delta_1/2} p_2^{-\delta_2/2}, \quad m_1 p_1^{-1+\delta_1/2} m_2 p_2^{-1+\delta_2/2}\right) \cdot \frac{1}{\sqrt{T}} + \frac{1}{\sqrt{p_1 p_2}}\right).$$

When $m_1 m_2 \sim o_p(p_1^{1-\delta_1} p_2^{1-\delta_2})$, the rate in Theorem 3 becomes $\frac{1}{\sqrt{p_1 p_2}} \|\widehat{\boldsymbol{S}}_t - \boldsymbol{S}_t\|_2 = O_p\left(p_1^{-\delta_1/2} p_2^{-\delta_2/2} T^{-1/2} + p_1^{-1/2} p_2^{-1/2}\right)$ and increases in p_1 and p_2 improve the convergence rate. In other cases, we get $\frac{1}{\sqrt{p_1 p_2}} \|\widehat{\boldsymbol{S}}_t - \boldsymbol{S}_t\|_2 = O_p\left(m_1 p_1^{-1+\delta_1/2} m_2 p_2^{-1+\delta_2/2} T^{-1/2} + p_1^{-1/2} p_2^{-1/2}\right)$. And so long as $m_1 m_2$ increases slower than $p_1 p_2$ does, we get a faster convergence rate than that of the unconstrained model in Wang et al. (2017). Note that the estimation of the loading spaces are consistent with fixed p_1 and p_2 in Theorem 1. But the consistency of the signal estimate requires $p_1, p_2 \to \infty$.

As noted in Section 3, the row and column factor loading matrices $\boldsymbol{\Lambda} = \boldsymbol{H}_R \boldsymbol{R}$ and $\boldsymbol{\Gamma} = \boldsymbol{H}_C \boldsymbol{C}$ are only identifiable up to a linear space spanned by its columns. Following Lam et al. (2011) and Wang et al. (2017), we adopt the discrepancy measure used by Chang et al. (2015): for two orthogonal matrices $\boldsymbol{O}_1$ and $\boldsymbol{O}_2$ of size $p \times q_1$ and $p \times q_2$, then the difference between the two linear spaces $\mathcal{M}(\boldsymbol{O}_1)$ and $\mathcal{M}(\boldsymbol{O}_2)$ is measured by

$$\mathcal{D}(\mathcal{M}(\boldsymbol{O}_1), \mathcal{M}(\boldsymbol{O}_2)) = \left(1 - \frac{1}{\max(q_1, q_2)} tr\left(\boldsymbol{O}_1 \boldsymbol{O}_1' \boldsymbol{O}_2 \boldsymbol{O}_2'\right)\right)^{1/2}. \tag{15}$$

Clearly, $\mathcal{D}(\mathcal{M}(\boldsymbol{O}_1), \mathcal{M}(\boldsymbol{O}_2))$ assumes values in [0,1]. It equals to 0 if and only if $\mathcal{M}(\boldsymbol{O}_1) = \mathcal{M}(\boldsymbol{O}_2)$ and equals to 1 if and only if $\mathcal{M}(\boldsymbol{O}_1) \perp \mathcal{M}(\boldsymbol{O}_2)$. If $\boldsymbol{O}_1$ and $\boldsymbol{O}_2$ are vectors, (15) is the cosine similarity measure. The following Theorem 4 shows that the error in estimating loading spaces goes to zero as p_1, p_2 and T go to infinity and the convergence rate is of the same order as that for estimated $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$.

Theorem 4. *Under Conditions 1-6 and if* $m_1 p_1^{-1+\delta_1} m_2 p_2^{-1+\delta_2} T^{-1/2} = o_p(1)$*, then*

$$\begin{aligned} \mathcal{D}(\mathcal{M}(\widehat{\boldsymbol{\Lambda}}), \mathcal{M}(\boldsymbol{\Lambda})) &= \mathcal{D}(\mathcal{M}(\widehat{\boldsymbol{\Gamma}}), \mathcal{M}(\boldsymbol{\Gamma})) \\ &= O_p\left(\max\left(T^{-1/2}, \frac{m_1}{p_1^{1-\delta_1}} \frac{m_2}{p_2^{1-\delta_2}} T^{-1/2}\right)\right). \end{aligned}$$

Asymptotic theories for estimators of nonorthogonal, multi-term constrained factors model are trivial extensions of the above properties for the orthogonal constrained model.

5 Simulation

In this section, we use simulation to study the performance of the estimation methods of Section 3 in finite samples. We also compare the results with those of unconstrained models. We employ data generating models under orthogonal full and partial constraints, respectively. In the simulation, we use the Student-t distribution with 5 degrees of freedom to generate the entries in the disturbances $\boldsymbol{U}_t$. Using Gaussian noise shows similar results.

5.1 Case 1. Orthogonal Constraints

In this case, the observed data $\boldsymbol{Y}_t$'s are generated according to Model (2),

$$\boldsymbol{Y}_t = \boldsymbol{H}_R \boldsymbol{R} \boldsymbol{F}_t \boldsymbol{C}' \boldsymbol{H}_C' + \boldsymbol{U}_t, \qquad t = 1, \ldots, T,$$

under the following simulation design.

The latent factor process $\boldsymbol{F}_t$ is of dimension $k_1 \times k_2 = 3 \times 2$. The entries of $\boldsymbol{F}_t$ follow $k_1 k_2$ independent $AR(1)$ processes with Gaussian white noise $\mathcal{N}(0,1)$ innovations. Specifically, $vec(\boldsymbol{F}_t) = \boldsymbol{\Phi}_F\, vec(\boldsymbol{F}_{t-1}) + \boldsymbol{\epsilon}_t$ with $\boldsymbol{\Phi}_F = diag(-0.5, 0.6, 0.8, -0.4, 0.7, 0.3)$. The dimensions of the constrained row and column loading spaces are $m_1 = 12$ and $m_2 = 3$, respectively. Hence, $\boldsymbol{R}$ is 12×3 and $\boldsymbol{C}$ is 3×2. The entries of $\boldsymbol{R}$ and $\boldsymbol{C}$ are independently sampled from the uniform distribution $U(-p_i^{-\delta_i/2}\sqrt{m_i/p_i},\ p_i^{-\delta_i/2}\sqrt{m_i/p_i})$ for $i = 1, 2$, respectively, so that the condition on the factor strength is satisfied. The disturbance $\boldsymbol{U}_t = \boldsymbol{\Psi}^{1/2}\boldsymbol{\Xi}_t$ is a white noise process, where the elements of $\boldsymbol{\Xi}_t$ are independent random variables of Student-t distribution with five degrees of freedom and the matrix $\boldsymbol{\Psi}^{1/2}$ is chosen so that $\boldsymbol{U}_t$ has a Kronecker product covariance structure $cov(vec(\boldsymbol{U}_t)) = \boldsymbol{\Gamma}_2 \otimes \boldsymbol{\Gamma}_1$, where $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ are of size $p_1 \times p_1$ and $p_2 \times p_2$ respectively. For $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$, the diagonal elements are 1 and the off-diagonal elements are 0.2.

The effects of factor strength are investigated by varying factor strength parameter (δ_1, δ_2) among $(0,0)$, $(0.5, 0)$, $(0.5, 0.5)$. For each pair of δ_i's, the dimensions (p_1, p_2) are chosen to be $(20, 20)$, $(20, 40)$, $(40, 20)$ and $(40, 40)$. The sample sizes T are $0.5p_1p_2$, p_1p_2, $1.5p_1p_2$ and $2p_1p_2$. For each combination of the parameters, we use 500 realizations. And we use $h_0 = 1$ for all simulations. Estimation error of $\mathcal{M}(\widehat{\boldsymbol{Q}}_i)$ is defined as $\mathcal{D}(\widehat{\boldsymbol{Q}}_i, \boldsymbol{Q}_i)$, where the distance $\mathcal{D}$ is defined in (15).

The row constraint matrix $\boldsymbol{H}_R$ is a $p_1 \times 12$ orthogonal matrix. For $p_1 = 20$, $\boldsymbol{H}_R$ is assumed to be a block diagonal matrix $\boldsymbol{I}_4 \otimes \boldsymbol{D}$, where $\boldsymbol{I}_k$ is the identify matrix of dimension k and $\boldsymbol{D} = [\boldsymbol{d}_1, \boldsymbol{d}_2, \boldsymbol{d}_3]$ is a 5×3 matrix with $\boldsymbol{d}_1' = (1,1,1,1,1)/\sqrt{5}$, $\boldsymbol{d}_2' = (-1,-1,0,1,1)/2$, $\boldsymbol{d}_3' = (-1,0,2,0,-1)/\sqrt{6}$. These three $\boldsymbol{d}_j$ vectors can be viewed as the level, slope and curvature, respectively, of a group of five variables. Therefore, the 20 rows are divided into 4 groups of size 5. When we increase p_1 to 40 while keeping $m_1 = 12$ fixed, we double the length of each vector in the columns of $\boldsymbol{D}$, using $\boldsymbol{d}_1' = (1,1,1,1,1,1,1,1,1,1)/\sqrt{10}$, $\boldsymbol{d}_2' = (-1,-1,-1,-1,0,0,1,1,1,1)/\sqrt{8}$ and $\boldsymbol{d}_3' = (-1,-1,0,0,2,2,0,0,-1,-1)/\sqrt{12}$.

The column constraint matrix $\boldsymbol{H}_C$ is a $p_2 \times 3$ orthogonal matrix. For $p_2 = 20$, the three columns of $\boldsymbol{H}_C$ are generated as $\boldsymbol{h}_{c,1} = [\mathbf{1}_7/\sqrt{7}, \mathbf{0}_7, \mathbf{0}_6]'$, $\boldsymbol{h}_{c,2} = [\mathbf{0}_7, \mathbf{1}_7/\sqrt{7}, \mathbf{0}_6]'$, $\boldsymbol{h}_{c,3} = [\mathbf{0}_7, \mathbf{0}_7, \mathbf{1}_6/\sqrt{6}]'$, where $\mathbf{0}_k$ denotes a k-dimensional zero row vector. The constraints represent a 3-group classification. The 20 columns are divided into 3 groups of size 7, 7 and 6 respectively. In increasing p_2 to 40 while keeping $m_2 = 3$ fixed, we double the length of each vector in the columns defined above.

Table 1 shows the performance of estimating the true number of factors. We compare the total number of estimated factors $\widehat{k} = \widehat{k}_1\widehat{k}_2$ with the true value $k = k_1k_2 = 6$. The subscripts c and u denote results from the constrained model (2) and unconstrained model (1), respectively. f_c and f_u denote the relative frequency of correctly estimating the true number of factors k. From the table, we make the following observations. First, when the row and column factors are strong, i.e. $(\delta_1, \delta_2) = (0, 0)$, both constrained and unconstrained models can estimate accurately the number of factors, but the constrained models fare better when the sample size is small. Second, if the strength of the row factors is weak, but the strength of the column factors is strong, i.e. $(\delta_1, \delta_2) = (0.5, 0)$, the unconstrained models fail to estimate the number of factors, but the constrained models continue to perform well. Furthermore, as expected, the performance of the constrained models improves with the sample size. Finally, if the strength of the row and columns factors is weak, i.e. $(\delta_1, \delta_2) = (0.5, 0.5)$, both models encounter difficulties in estimating the correct number of factors for the sample sizes used. This is not surprising as weak signals are hard to detect in general.

				$T = 0.5\,p_1\,p_2$		$T = p_1\,p_2$		$T = 1.5\,p_1\,p_2$		$T = 2\,p_1\,p_2$	
δ_1	δ_2	p_1	p_2	f_u	f_c	f_u	f_c	f_u	f_c	f_u	f_c
0	0	20	20	0.29	0.95	0.77	1	0.95	1	0.99	1
		20	40	0.77	1	0.99	1	1	1	1	1
		40	20	0.81	1	1	1	1	1	1	1
		40	40	1	1	1	1	1	1	1	1
0.5	0	20	20	0	0.2	0	0.49	0	0.78	0	0.92
		20	40	0	0.68	0	0.96	0	0.99	0	1
		40	20	0	0.37	0	0.78	0	0.92	0	0.97
		40	40	0	0.86	0	0.98	0	0.99	0	1
0.5	0.5	20	20	0	0.05	0	0.02	0	0.02	0	0.01
		20	40	0	0.03	0	0.02	0	0.01	0	0
		40	20	0	0.05	0	0.01	0	0	0	0.01
		40	40	0	0.05	0	0	0	0.01	0	0.04

Table 1: Relative frequencies of correctly estimating the number of factors k in the case of orthogonal constraints.

Figure 1 shows the box-plots of the estimation errors in estimating the loading spaces of $\boldsymbol{Q} = \boldsymbol{Q}_2 \otimes \boldsymbol{Q}_1$ using the correct number of factors. The gray boxes are for the constrained models. From the plots, it is seen that when both row and column factors are strong, i.e.

$(\delta_1, \delta_2) = (0, 0)$, and the number of factors is properly estimated, the mean and standard deviation of the estimation errors $\mathcal{D}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$ are small for both models, but the constrained model has a smaller mean estimation error. When row factors are weak, i.e. $(\delta_1, \delta_2) = (0.5, 0)$, and the number of factors is given, the estimation error of constrained models remains small whereas that of the unconstrained models is substantially larger.

Table 2 shows the mean and standard deviations of the estimation errors $\mathcal{D}(\widehat{\boldsymbol{Q}}_i, \boldsymbol{Q}_i)$ for row ($i = 1$) and column ($i = 2$) loading spaces separately for the constrained model (2). Column loading spaces are estimated with higher accuracy because the number of column constraints ($p_1 - m_1$) is larger than the number of row constraints ($p_2 - m_2$). From the table, we see that (a), as expected, the mean of estimation errors decreases as the sample size increases and (b) the mean of estimation errors is inversely proportional to the strength of row factors.

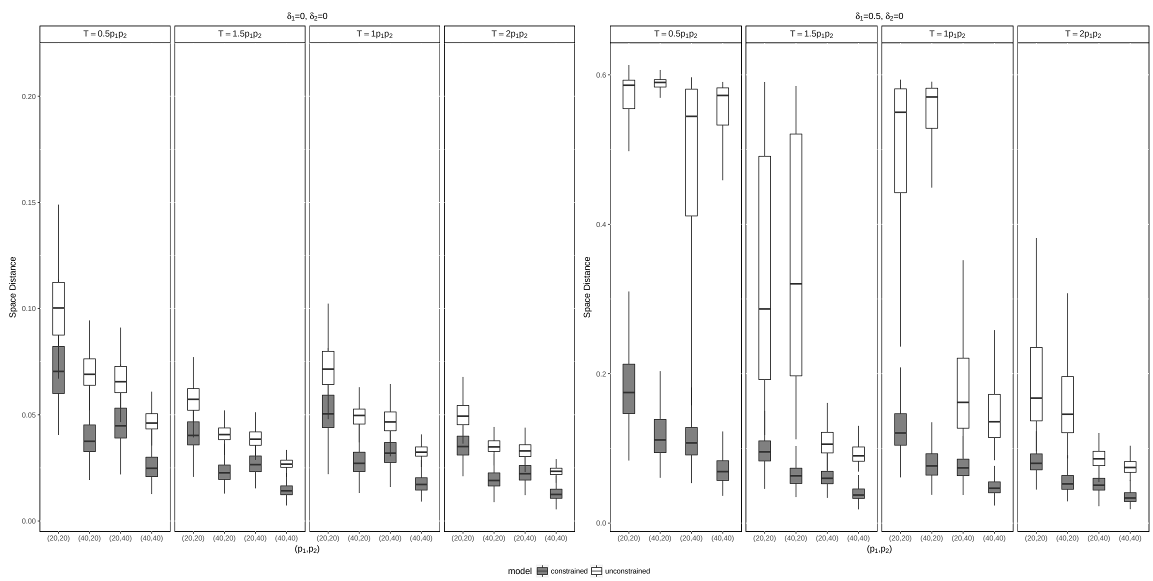


Figure 1: Box-plots of the estimation accuracy measured by $\mathcal{D}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$ for the case of orthogonal constraints. Gray boxes represent the constrained model. The results are based on 500 iterations. See Table 14 in Appendix D for plotted values.

To investigate the performance of estimation under different choices of h_0, which is the number of lags used in (11), we change the underlying generating model of $vec(\boldsymbol{F}_t)$ to a VAR(2) process without the lag-1 term, $vec(\boldsymbol{F}_t) = \boldsymbol{\Phi}_F vec(\boldsymbol{F}_{t-2}) + \boldsymbol{\epsilon}_t$. Here we only consider the strong factor setting with $\delta_1 = \delta_2 = 0$ and use the sample size $T = 2p_1p_2$ for each combination of p_1 and p_2. All the other parameters are the same as those in Section 5.1. Table 3 presents the simulation results. Since $vec(\boldsymbol{F}_t)$, and hence $vec(\boldsymbol{Y}_t)$, has zero auto-covariance matrix at lag 1, $\widehat{\boldsymbol{M}}$ under $h_0 = 1$ contains no information on the signal,

				$T = 0.5\,p_1\,p_2$		$T = p_1\,p_2$		$T = 1.5\,p_1\,p_2$		$T = 2\,p_1\,p_2$	
δ_1	δ_2	p_1	p_2	$\mathcal{D}(\widehat{Q}_1, Q_1)$	$\mathcal{D}(\widehat{Q}_2, Q_2)$	$\mathcal{D}(\widehat{Q}_1, Q_1)$	$\mathcal{D}(\widehat{Q}_2, Q_2)$	$\mathcal{D}(\widehat{Q}_1, Q_1)$	$\mathcal{D}(\widehat{Q}_2, Q_2)$	$\mathcal{D}(\widehat{Q}_1, Q_1)$	$\mathcal{D}(\widehat{Q}_2, Q_2)$
0	0	20	20	0.71(0.18)	0.13(0.07)	0.51(0.13)	0.09(0.05)	0.41(0.09)	0.07(0.04)	0.35(0.07)	0.06(0.03)
		20	40	0.46(0.11)	0.08(0.04)	0.32(0.07)	0.05(0.03)	0.27(0.06)	0.04(0.02)	0.23(0.05)	0.04(0.02)
		40	20	0.40(0.12)	0.07(0.04)	0.28(0.07)	0.05(0.03)	0.23(0.06)	0.04(0.02)	0.19(0.05)	0.04(0.02)
		40	40	0.26(0.07)	0.04(0.02)	0.18(0.04)	0.03(0.02)	0.14(0.04)	0.03(0.01)	0.13(0.03)	0.02(0.01)
0.5	0	20	20	1.84(0.75)	0.5(0.23)	1.23(0.35)	0.30(0.15)	0.95(0.23)	0.22(0.11)	0.81(0.18)	0.17(0.09)
		20	40	1.08(0.30)	0.26(0.13)	0.74(0.18)	0.15(0.08)	0.61(0.14)	0.12(0.06)	0.52(0.12)	0.10(0.05)
		40	20	1.18(0.45)	0.28(0.15)	0.78(0.23)	0.17(0.09)	0.64(0.18)	0.13(0.07)	0.54(0.14)	0.11(0.06)
		40	40	0.71(0.21)	0.14(0.08)	0.48(0.13)	0.09(0.05)	0.39(0.1)	0.07(0.04)	0.35(0.09)	0.06(0.03)
0.5	0.5	20	20	5.84(0.62)	2.04(0.53)	5.35(0.75)	1.63(0.42)	4.68(1.17)	1.33(0.34)	4.20(1.31)	1.13(0.32)
		20	40	5.62(0.68)	1.98(0.40)	4.75(1.13)	1.47(0.30)	3.96(1.33)	1.18(0.27)	3.32(1.35)	0.97(0.24)
		40	20	5.53(0.61)	1.52(0.50)	4.68(1.25)	1.00(0.37)	3.64(1.46)	0.76(0.30)	2.87(1.42)	0.61(0.25)
		40	40	5.01(1.01)	1.32(0.38)	3.64(1.47)	0.84(0.29)	2.62(1.46)	0.61(0.20)	1.98(1.14)	0.49(0.19)

Table 2: Means and standard deviations (in parentheses) of the estimation accuracy measured by $\mathcal{D}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$ for constrained factor models. The case of orthogonal constraints is used. All numbers in the table are 10 times the true numbers for clear presentation. The results are based on 500 simulations.

and, as expected, both the constrained and unconstrained models fail to correctly estimate the number of factors and the loading space. On the other hand, both models are able to correctly estimate the number of factors when $h_0 > 1$ with the constrained model faring better. The fact that $h_0 = 2, 3, 4$ give very similar results shows that the choice of h_0 does not affect the performance much so long as at least one non-zero auto-covariance matrix is included in the calculation. In practice, one can select h_0 by examining the sample cross-correlation matrices of $\boldsymbol{Y}_t$.

5.2 Case 2. Partial Orthogonal Constraints

In this case, the observed data $\boldsymbol{Y}_t$'s are generated using Model (5),

$$\boldsymbol{Y}_t = \boldsymbol{H}_R \boldsymbol{R}_1 \boldsymbol{F}_t \boldsymbol{C}_1' \boldsymbol{H}_C' + \boldsymbol{L}_R \boldsymbol{R}_2 \boldsymbol{G}_t \boldsymbol{C}_2' \boldsymbol{L}_C' + \boldsymbol{U}_t, \qquad t = 1, \ldots, T.$$

Parameter settings of the first part $\boldsymbol{H}_R \boldsymbol{R}_1 \boldsymbol{F}_t \boldsymbol{C}_1' \boldsymbol{H}_C'$ are the same as those in Case 1. The latent factor process $\boldsymbol{G}_t$ is of dimension $q_1 \times q_2 = 5 \times 4$. The entries of $\boldsymbol{G}_t$ follow $q_1 q_2$ independent $AR(1)$ processes with Gaussian white noise $\mathcal{N}(0, 1)$ innovations, $vec(\boldsymbol{G}_t) = \boldsymbol{\Phi}_G\, vec(\boldsymbol{G}_{t-1}) + \boldsymbol{\epsilon}_t$ with $\boldsymbol{\Phi}_G$ being a diagonal matrix with entries $(-0.7, 0.5, -0.2, 0.9, 0.1, 0.4, 0.6, -0.5, 0.7, 0.7, -0.4, 0.4, 0.4, -0.6, -0.6, 0.6, -0.5, -0.3, 0.2, -0.4)$. The row loading matrix $\boldsymbol{L}_R \boldsymbol{R}_2$ is a 20×5 orthogonal matrix, satisfying $\boldsymbol{H}_R' \boldsymbol{L}_R = \boldsymbol{0}$. The column loading matrix $\boldsymbol{L}_C \boldsymbol{C}_2$ is a 20×4 orthogonal matrix, satisfying $\boldsymbol{H}_C' \boldsymbol{L}_C = \boldsymbol{0}$. The entries of $\boldsymbol{R}_2$ and $\boldsymbol{C}_2$ are random draws from the uniform distribution between $-p_i^{-\eta_i/2}\sqrt{p_i/(p_i - m_i)}$ and

	p_1	p_2	$h_0 = 1$	$h_0 = 2$	$h_0 = 3$	$h_0 = 4$
f_c	20	20	0.12	1.00	1.00	1.00
	20	40	0.16	1.00	1.00	1.00
	40	20	0.12	1.00	1.00	1.00
	40	40	0.22	1.00	1.00	1.00
f_u	20	20	0.00	0.89	0.58	0.43
	20	40	0.00	1.00	1.00	0.95
	40	20	0.00	1.00	1.00	0.97
	40	40	0.00	1.00	1.00	1.00
$\mathcal{D}_c(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$	20	20	2.83(1.13)	0.36(0.07)	0.37(0.07)	0.38(0.08)
	20	40	2.69(1.15)	0.23(0.05)	0.23(0.05)	0.24(0.05)
	40	20	2.54(1.21)	0.20(0.05)	0.20(0.05)	0.21(0.06)
	40	40	2.31(1.17)	0.13(0.03)	0.13(0.03)	0.14(0.04)
$\mathcal{D}_u(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$	20	20	4.37(1.29)	0.51(0.07)	0.53(0.07)	0.53(0.08)
	20	40	4.30(1.30)	0.34(0.04)	0.35(0.04)	0.35(0.04)
	40	20	4.36(1.31)	0.36(0.04)	0.37(0.04)	0.37(0.05)
	40	40	4.34(1.34)	0.24(0.02)	0.24(0.03)	0.25(0.03)

Table 3: Performance of estimation under different choices of h_0 when $vec(\boldsymbol{F}_t) = \boldsymbol{\Phi}_{\boldsymbol{F}} vec(\boldsymbol{F}_{t-2}) + \boldsymbol{\epsilon}_t$. Metrics reported are relative frequencies of correctly estimating k, means and standard deviations (in parentheses) of the estimation accuracy measured by $\mathcal{D}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$. Means and standard deviations are multiplied by 10 for ease in presentation.

$p_i^{-\eta_i/2}\sqrt{p_i/(p_i - m_i)}$ for $i = 1, 2$, respectively, so that the conditions on factor strength are satisfied. Factor strength is controlled by the δ_i's.

Model (5) could be written in the following form:

$$\boldsymbol{Y}_t = (\boldsymbol{H}_R \boldsymbol{R}_1 \ \boldsymbol{L}_R \boldsymbol{R}_2) \begin{pmatrix} \boldsymbol{F}_t & 0 \\ 0 & \boldsymbol{G}_t \end{pmatrix} \begin{pmatrix} \boldsymbol{C}_1' \boldsymbol{H}_C' \\ \boldsymbol{C}_2' \boldsymbol{L}_C' \end{pmatrix} + \boldsymbol{U}_t, \qquad t = 1, \ldots, T.$$

In this form, the true number of factors is $k_0 = (k_1 + r_1)(k_2 + r_2)$ and the true loading matrix is $(\boldsymbol{H}_C \boldsymbol{C}_1 \ \boldsymbol{L}_C \boldsymbol{C}_2) \otimes (\boldsymbol{H}_R \boldsymbol{R}_1 \ \boldsymbol{L}_R \boldsymbol{R}_2)$. Table 4 shows the frequency of correctly estimating k_0 based on 500 iterations. In the table, f_u denotes the frequency of correctly estimating k_0 for unconstrained model. f_1 and f_2 denote the same frequency metric for the first matrix factor $\boldsymbol{F}_t$ and second matrix factor $\boldsymbol{G}_t$ of the constrained model. The number of factors in $\boldsymbol{F}_t$ is estimated with a higher accuracy because the dimension of constrained loading space for $\boldsymbol{F}_t$ is $m_1 m_2 = 36$, which is smaller than that for $\boldsymbol{G}_t$, $(p_1 - m_1)(p_2 - m_2) = 136$. The result again confirms the theoretical results in Section 4. Note that Table 4 only contains selected combinations of factor strength parameters δ_i's ($i = 1, \dots 4$). The results of all combinations of factor strength are given in Table 15 in Appendix D.

Figure 2 and Figure 3 present box-plots of estimation errors under weak and strong factors from 500 simulations, respectively. Again, the results show that the constrained approach efficiently improves the estimation accuracy. The performance of constrained model is good even in the case of weak factors. Moreover, with stronger signals and larger sample sizes, both approaches increase their estimation accuracy.

						$T = 0.5 * p_1 * p_2$			$T = p_1 * p_2$			$T = 1.5 * p_1 * p_2$			$T = 2 * p_1 * p_2$		
δ_1	δ_2	δ_3	δ_4	p_1	p_2	f_u	f_{con_1}	f_{con_2}	f_u	f_{con_1}	f_{con_2}	f_u	f_{con_1}	f_{con_2}	f_u	f_{con_1}	f_{con_2}
0	0	0	0	20	20	0	0.94	0	0	1.00	0	0	1.00	0	0.01	1.00	0
				20	40	0	1.00	0	0	1.00	0	0.03	1.00	0	0.19	1.00	0
				40	20	0.15	0.99	1.00	0.81	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00
				40	40	0.71	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0	0	0.5	0	20	20	0	0.94	0	0	1.00	0	0	1.00	0	0	1.00	0
				20	40	0	1.00	0	0	1.00	0	0	1.00	0	0	1.00	0
				40	20	0	0.99	0.54	0	1.00	0.84	0	1.00	0.97	0	1.00	1.00
				40	40	0	1.00	0.98	0	1.00	1.00	0	1.00	1.00	0	1.00	1.00
0.5	0.5	0.5	0.5	20	20	0	0.07	0	0	0.04	0	0	0.01	0	0	0.01	0
				20	40	0	0.07	0	0	0.02	0	0	0.01	0	0	0.01	0
				40	20	0	0.06	0	0	0.01	0	0	0	0	0	0	0
				40	40	0	0.06	0	0	0	0	0	0	0	0	0.03	0

Table 4: Relative frequencies of correctly estimating the number of factors for partially constrained factor models. Full tables including all combinations are presented in Table 15 in Appendix D.

6 Applications

In this section, we demonstrate the advantages of using constrained matrix-variate factor models with three applications. In practice, the number of common factors (k_1, k_2) and the dimensions of constrained row and column loading spaces (m_1, m_2) must be pre-specified in order to determine an appropriate constrained factor model. The numbers of factors (k_1, k_2) can be determined by any existing methods, such as those in Lam & Yao (2012) and Wang et al. (2017). For any given (k_1, k_2), the the dimensions of constrained row and column loading spaces (m_1,m_2) can be determined by either (a) prior or substantive knowledge or (b) an empirical procedure. The results show that even simple grouping information can substantially increase the accuracy in estimation.

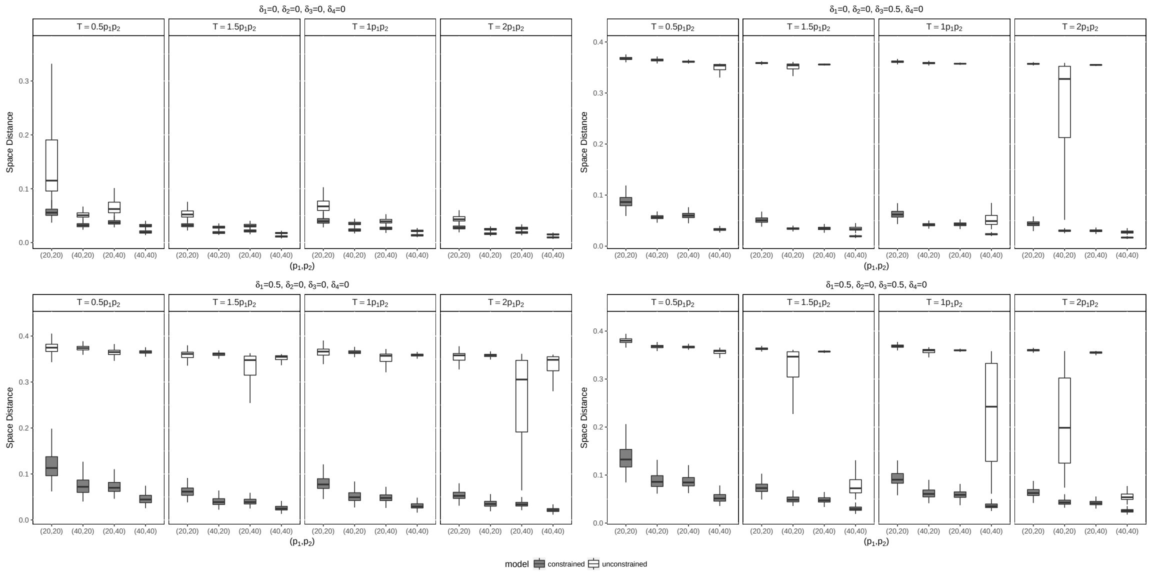


Figure 2: The strong factors case. Box-plots of the estimation accuracy measured by $\mathcal{D}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$ for partially constrained factor models. The gray boxes are for the constrained approach. The results are based on 500 realizations. See Table 16 in Appendix D for the plotted values.

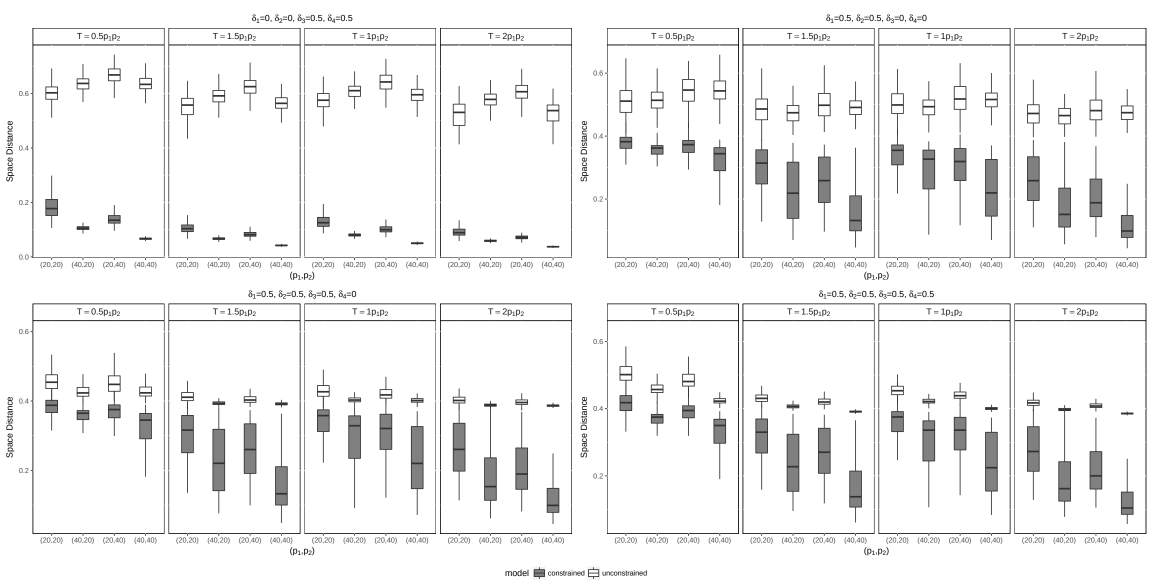


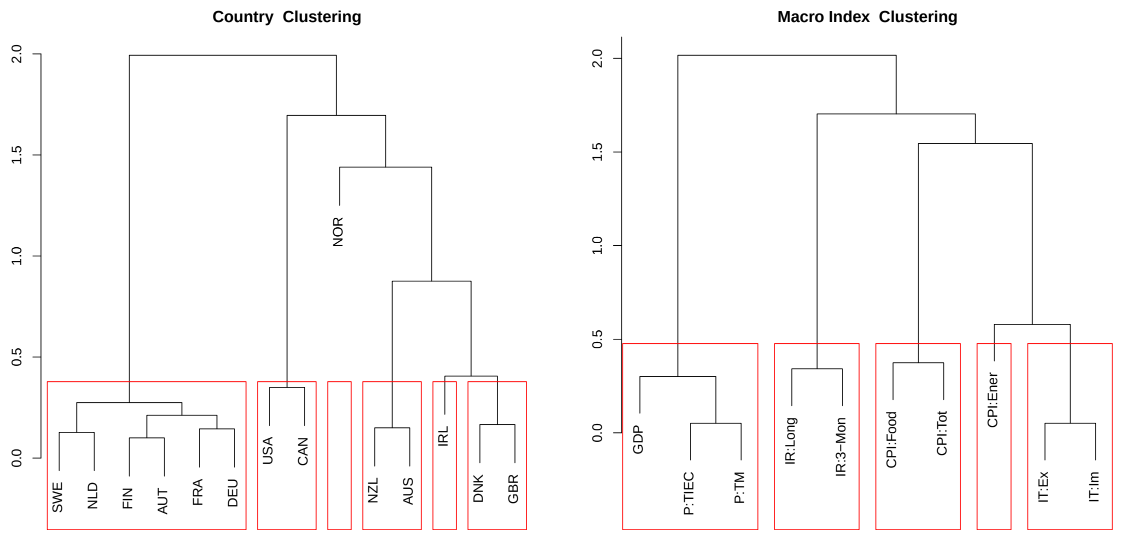
Figure 3: The weak factors case. Box-plots of the estimation accuracy measured by $\mathcal{D}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$ for partially constrained factor models. The gray boxes are for the constrained approach. The results are based on 500 realizations. See Table 16 in Appendix D for the plotted values.

6.1 Example 1: Multinational Macroeconomic Indices

We apply the constrained and partially constrained factor models to the macroeconomic indices dataset collected from OECD. The dataset contains 10 quarterly macroeconomic indices of 14 countries from 1990.Q2 to 2016.Q4 for107 quarters. Thus, we have $T = 107$ and $p_1 \times p_2 = 14 \times 10$ matrix-valued time series. The countries include developed economies from North American, European, and Oceania. The indices cover four major groups, namely production, consumer price, money market, and international trade. Each raw time series is transformed by taking the first or second difference or logarithm to satisfy the mixing condition in Condition 4. Countries, detailed descriptions of the dataset, and transformation procedures are given in Tables 12 and 13 of Appendix C.

We first fit an unconstrained matrix factor model which generates estimators of the row loading matrix and the column loading matrix. In the row loading matrix, each row represents a country by its factor loadings for all common row factors, whereas, in the column loading matrix, each row represents a macroeconomic index by its factor loadings for all common column factors. A hierarchical clustering algorithm is employed to cluster countries and macroeconomic indices based on their representations in the common row and column factor spaces, respectively. Figure 4 shows the hierarchical clustering results. Therefor, we construct the row and column constraint matrices based on the clustering results. It seems that the row constraint matrix divides countries into 6 groups: (i) United States and Canada; (ii) New Zealand and Australia; (iii) Norway; (iv) Ireland, Denmark, and United Kingdom; (v) Finland and Sweden; (vi) France, Netherlands, Austria, and Germany. The grouping more or less follows geographical partition with Norway different from all others due to its rich oil production and other distinct economic characteristics. The column constraint matrix divide macroeconomic indices into 5 categories: (i) GDP, production of total industry excluding construction, and production of total manufacturing ; (ii) long-term government bond yields and 3-month interbank rates and yields; (iii) total CPI and CPI of Food; (iv) CPI of Energy; (v) total exports value and total imports value in goods. Again, the grouping agrees with common economic knowledge.

Table 5 shows estimates of the row and column loading matrices for constrained and unconstrained 4×4 factor models. The loading matrices are normalized so that the norm of each column is one. They are also varimax-rotated to reveal a clear structure. The values shown are the rounded values of the estimates multiplied by 10 for ease in display. From the table, both the row and column loading matrices exhibit similar patterns between unconstrained and constrained models, partially validating the constraints while simplifying the analysis.



(a) Country Loading Clustering (b) Macroeconomic Index Loading Clustering

Figure 4: Macroeconomic series: Clustering loading matrices

Table 6 provides the estimates under the same setting as that of Table 5 but without any rotation. From the table, it is seen that except for the first common factors of the row loading matrices there exist some difference in the estimated loading matrices between unconstrained and constrained factor models. The results of constrained models convey more clearly the following observations. Consider the row factors. The first row common factor represents the status of global economy as it is a weighted average of all the countries under study. The remaining three row common factors mark certain differences between country groups. For the column factors, the first column common factor is dominated by the price index and interest rates; The second column common factor is mainly the production and international trade; The remaining two column common factors represent interaction between price indices, interest rates, productions, and international trade.

Table 7 compares the out-of-sample performance of unconstrained, constrained, and partially constrained factor models using a 10-fold cross validation (CV) for models with different number of factors. Residual sum of squares (RSS), their ratios to the total sum of squares (RSS/TSS), and the number of parameters are means of the 10-fold CV. Clearly, the constrained factor model uses far fewer parameters in the loading matrices yet achieves slightly better results than the unconstrained model. Using the same number of parameters, the partially constrained model is able to reduce markedly the RSS over the unconstrained model.

In this particular application, the constrained matrix factor model with the specified

Model	Loading	Row	USA	CAN	NZL	AUS	NOR	IRL	DNK	GBR	FIN	SWE	FRA	NLD	AUT	DEU
		1	7	7	1	1	-1	-2	-1	0	1	0	0	0	0	-1
$R_{unc,rot}$	$\widehat{R}'_{rot}$	2	0	1	-2	-1	1	1	1	2	4	3	4	4	4	4
		3	2	-1	5	5	1	5	3	2	-1	1	1	0	0	0
		4	-1	1	1	2	9	-3	0	0	0	1	-1	1	0	0
		1	6	6	0	0	0	2	2	2	-1	-1	0	0	0	0
$R_{con,rot}$	$\widehat{R}'_{rot}H'_R$	2	-1	-1	0	0	0	3	3	3	4	4	3	3	3	3
		3	0	0	7	7	0	1	1	1	1	1	-1	-1	-1	-1
		4	0	0	0	0	10	0	0	0	1	1	0	0	0	0

Model	Loading	Row	CPI:Food	CPI:Tot	CPI:Ener	IR:Long	IR:3-Mon	P:TIEC	P:TM	GDP	IT:Ex	IT:Im
		1	6	7	3	-1	1	0	0	-1	-1	0
$C_{unc,rot}$	$\widehat{C}'_{rot}$	2	-2	1	4	1	-1	0	0	0	6	6
		3	0	0	1	8	6	-1	0	1	0	0
		4	1	-1	0	0	0	6	6	5	0	0
		1	7	7	0	0	0	0	0	0	0	0
$C_{con,rot}$	$\widehat{C}'_{rot}H'_C$	2	0	0	6	0	0	0	0	0	6	6
		3	0	0	0	7	7	0	0	0	0	0
		4	0	0	-2	0	0	6	6	6	1	1

Table 5: Estimations of row and column loading matrices (varimax rotated) of constrained and unconstrained matrix factor models for multinational macroeconomic indices. The loadings matrix are multiplied by 10 and roundedto integers for ease in display.

constraint matrices seems appropriate and plausible. If incorrect structures (constraint matrices) are imposed on the model, then the constrained model may become inappropriate. As we can see from the next example, a single orthogonal constraint actually hurts the performance. In cases like this, we need a second or a third constraint to achieve satisfactory performance. Nevertheless, the results from the constrained model are better than those from the unconstrained model.

6.2 Example 2: Company Financials

In this application, we investigate the constrained matrix-variate factor models for the time series of 16 quarterly financial measurements of 200 companies from 2006.Q1 to 2015.Q4 for 40 observations. Appendix E contains the descriptions of variables used along with their definitions, the 200 companies and their corresponding industry group and sector information. Data are arranged in matrix-variate time series format. At each t, we observe a 16×200 matrix, whose rows represent financial variables and columns represent companies. Thus we have $T = 40$, $p_1 = 16$ and $p_2 = 200$. The total number of time series is $3,200$. Following the convention in eigenanalysis, we standardize the individual series before applying factor analysis. This data set was used in Wang et al. (2017) for an unconstrained

Model	Loading	Row	USA	CAN	NZL	AUS	NOR	IRL	DNK	GBR	FIN	SWE	FRA	NLD	AUT	DEU
R_{unc}	$\widehat{R}'$	1	3	2	2	2	2	2	3	3	3	3	3	3	3	3
		2	4	2	5	5	1	0	1	0	-3	-1	-2	-2	-2	-3
		3	3	6	-2	-2	4	-5	-3	-1	1	0	-1	1	0	0
		4	-4	-3	0	2	8	-1	1	0	-1	1	0	1	0	0
R_{con}	$\widehat{R}'H_R'$	1	1	1	2	2	2	3	3	3	4	4	3	3	3	3
		2	5	5	3	3	4	0	0	0	-2	-2	-2	-2	-2	-2
		3	-1	-1	5	5	-6	0	0	0	0	0	-1	-1	-1	-1
		4	-4	-4	3	3	6	-2	-2	-2	1	1	-1	-1	-1	-1
Model	Loading	Row	CPI:Food	CPI:Ener	CPI:Tot	IR:Long	IR:3-Mon	P:TIEC	P:TM	GDP	IT:Ex	IT:Im				
C_{unc}	$\widehat{C}'$	1	1	4	2	4	3	3	3	3	4	4				
		2	5	3	6	-1	1	-3	-4	-4	0	0				
		3	5	-1	2	-1	1	4	4	3	-4	-4				
		4	0	-1	-2	7	5	-2	-2	0	-3	-3				
C_{con}	$\widehat{C}'H_C'$	1	6	-2	6	4	4	0	0	0	-2	-2				
		2	0	0	0	3	3	5	5	5	3	3				
		3	-3	3	-3	5	5	-3	-3	-3	1	1				
		4	3	5	3	-1	-1	-2	-2	-2	5	5				

Table 6: Estimations of row and column loading matrices of constrained and unconstrained matrix factor models for multinational macroeconomic indices. No rotation is used. The loadings matrix are multiplied by 10 and rounded to integers for ease in display.

matrix factor model.

The column constraint matrix $\boldsymbol{H}_C$ is constructed based on the industrial classification of Bloomberg. The 200 companies are classified into 51 industrial groups, such as biotechnology, oil & gas, computer, among others. Thus the dimension of $\boldsymbol{H}_C$ is 200×51. Since we do not have adequate prior knowledge on corporate financials, we do no impose any constraint on the row loading matrix. Thus, in this application, we use $\boldsymbol{H}_R = \boldsymbol{I}_{16}$.

We apply the unconstrained model (1), the orthogonal constrained model (7), and the partial constrained model (5) to the data set. Table 8 shows the average residual sum of squares (RSS) and their ratios to the total sum of squares (TSS) from a 10-fold CV for models with different number of factors. Again, it is clear, from the table, that the constrained matrix factor models use fewer number of parameters in loading matrices and achieve similar results. If we use the same number of parameters in the loading matrices, variances explained by the constrained matrix factor models are much larger than those of the unconstrained ones, indicating the impact of over-parameterization. This application with $3,200$ tmie series is typical in high-dimensional time series. The number of parameters involved is usually huge in a unconstrained model. Via the example, we showed that constrained matrix factor models can largely reduce the number of parameters while keeping

Model	# Factor 1	# Factor 2	RSS	RSS/TSS	# Parameters
Full	(6,5)		570.50	0.449	134
Constrained	(6,5)		560.31	0.442	61
Partial	(6,5)	(6,5)	454.41	0.358	134
Full	(5,5)		613.26	0.482	120
Constrained	(5,5)		604.63	0.477	55
Partial	(5,5)	(5,5)	516.27	0.407	120
Full	(4,5)		658.15	0.517	106
Constrained	(4,5)		649.85	0.512	49
Partial	(4,5)	(4,5)	576.94	0.454	106
Full	(4,4)		729.46	0.573	96
Constrained	(4,4)		721.96	0.568	44
Partial	(4,4)	(4,4)	657.13	0.517	96
Full	(3,4)		787.80	0.620	82
Constrained	(3,4)		768.64	0.605	38
Partial	(3,4)	(3,4)	719.46	0.567	82
Full	(3,3)		868.43	0.684	72
Constrained	(3,3)		852.76	0.671	33
Partial	(3,3)	(3,3)	813.16	0.640	72

Table 7: Results of 10-fold CV of out-of-sample performance for the multinational macroeconomic indices. The numbers shown are average over the cross validation, where RSS and TSS stand for residual and total sum of squares, respectively.

the same explanation power.

6.3 Example 3: Fama-French 10 by 10 Series

Finally, we investigate constrained matrix-variate factor models for the monthly market-adjusted return series of Fama-French 10×10 portfolios from January 1964 to December 2015 for total 624 months and overall $62,400$ observations. The portfolios are the intersections of 10 portfolios formed by size (market equity, ME) and 10 portfolios formed by the ratio of book equity to market equity (BE/ME). Thus, we have $T = 624$ and $p_1 \times p_2 = 10 \times 10$ matrix time series. The series are constructed by subtracting the monthly excess market returns from each of the original portfolio returns obtained from French (2017), so they are free of the market impact.

Using an unconstrained matrix factor model, Wang et al. (2017) carried out a clustering analysis on the ME and BE/ME loading matrices after rotation. Their results suggest $\boldsymbol{H}_R = [\boldsymbol{hr}_1, \boldsymbol{hr}_2, \boldsymbol{hr}_3]$, where $\boldsymbol{hr}_1 = [\mathbf{1}(5)/\sqrt{5}, \mathbf{0}(5)]$, $\boldsymbol{hr}_2 = [\mathbf{0}(5), \mathbf{1}(4)/2, 0]$, and $\boldsymbol{hr}_3 = [\mathbf{0}(9), 1]$. Therefore, ME factors are classified into three groups of smallest 5 ME's, middle 4 ME's, and the largest ME, respectively. For cases when we need 4 row constraints,

Model	# Factor 1	# Factor 2	RSS	RSS/SST	# parameters
Full	(4,10)		8140.32	0.869	2064
	(4,12)		7990.04	0.853	2464
	(4,19)		7587.11	0.810	3864
Constrained	(4,10)		8062.63	0.861	574
Partial	(4,10)	(4,2)	7969.83	0.851	936
	(4,10)	(4,9)	7623.25	0.814	1979
Full	(4, 20)		7539.68	0.805	4064
	(4, 27)		7261.49	0.775	5464
	(4, 39)		6872.18	0.734	7864
Constrained	(4, 20)		7646.70	0.816	1084
Partial	(4, 20)	(4,7)	7292.06	0.779	2191
	(4, 20)	(4,19)	6815.96	0.728	3979
Full	(5,10)		8012.10	0.855	2080
	(5,12)		7849.34	0.838	2480
	(5,19)		7420.04	0.792	3880
Constrained	(5,10)		7942.95	0.848	590
Partial	(5,10)	(5,2)	7849.40	0.838	968
	(5,10)	(5,9)	7472.10	0.798	2011
Full	(5,20)		7368.63	0.787	7960
	(5,23)		7250.73	0.774	4680
	(5,39)		6641.13	0.709	7880
Constrained	(5,20)		7489.20	0.800	1100
Partial	(5,20)	(5,3)	7357.80	0.786	1627
	(5,20)	(5,19)	6595.03	0.704	4011
Full	(5,30)		6960.70	0.743	6080
	(5,34)		6813.93	0.727	6880
	(5,59)		5988.15	0.639	11880
Constrained	(5,30)		7184.53	0.767	1610
Partial	(5,30)	(5,4)	6997.21	0.747	2286
	(5,30)	(5,29)	5936.64	0.634	6011

Table 8: Summary of 10-fold CV of out-of-sample analysis for the corporate financials of 16 series for each of 200 companies. The numbers shown are average over the cross validation and RSS and TSS denote, respectively, the residual and total sum of squares.

we redefine $\boldsymbol{hr}_2 = [\mathbf{0}(5), \mathbf{1}(3)/\sqrt{3}, \mathbf{0}(2)]$ and add a fourth column $\boldsymbol{hr}_4 = [\mathbf{0}(8), 1, 0]$. For column constraints, $\boldsymbol{H}_C = [\boldsymbol{hc}_1, \boldsymbol{hc}_2, \boldsymbol{hc}_3]$, where $\boldsymbol{hc}_1 = [1, \mathbf{0}(9)]$, $\boldsymbol{hc}_2 = [0, \mathbf{1}(3)/\sqrt{3}, \mathbf{0}(6)]$, $\boldsymbol{hc}_3 = [\mathbf{0}(4), \mathbf{1}(6)]$. Therefore, BE/ME factors are divided into three groups of the smallest BE/ME's, middle 3 BE/ME's, and the 6 largest BE/ME, respectively. For cases when we need 4 column constraints, we redefine $\boldsymbol{hc}_3 = [\mathbf{0}(4), \mathbf{1}(4)/2, \mathbf{0}(2)]$ and add a fourth column $\boldsymbol{hc}_4 = [\mathbf{0}(8), \mathbf{1}(2)]$.

Table 9 shows the estimates of the loading matrices for the constrained and unconstrained 2 × 2-factor models. The loading matrices are VARIMAX roated for ease in interpretation and normalized so that the norm of each column is one. From the table, the loading matrices exhibit similar patterns, but those of the constrained model convey the following observations more clearly. Consider the row factors, the first factor represents the difference between the average of the 5 smallest ME group and the weighted average of the remaining portfolio whereas the second factor is mainly the average of the medium 4 ME portfolios. For the column loading matrix, the first factor is a weighted average of the smallest BE/ME portfolio and the middle three portfolios. The second factor marks the difference between the smallest BE/ME portfolio from a weighted average of the two remaining groups. Finally, it is interesting to see that the constrained model uses only 16 parameters, yet it can reveal information similar to the unconstrained model that employs 40 parameters. This latter result demonstrates the power of using constrained factor models.

Model	Loading	Column	Rotated Estimated Loadings									
R_u	$\widehat{R}'$	1	0.43	0.46	0.44	0.43	0.33	0.16	0.05	-0.02	-0.20	-0.23
		2	-0.01	-0.01	-0.05	0.09	0.18	0.39	0.39	0.62	0.51	0.16
	$\widehat{R}'H_R'$	1	0.44	0.44	0.44	0.44	0.44	-0.04	-0.04	-0.04	-0.04	-0.15
		2	0.04	0.04	0.04	0.04	0.04	0.50	0.50	0.50	0.50	0.06
C_u	$\widehat{C}'$	1	0.70	0.48	0.37	0.30	0.14	0.07	0.05	-0.05	-0.09	0.15
		2	0.29	-0.07	-0.10	-0.23	-0.30	-0.32	-0.34	-0.44	-0.48	-0.34
	$\widehat{C}'H_C'$	1	0.78	0.36	0.36	0.36	0	0	0	0	0	0
		2	0.24	-0.18	-0.18	-0.18	-0.37	-0.37	-0.37	-0.37	-0.37	-0.37

Table 9: Estimates of the loading matrices of constrained and unconstrained matrix factor modes for Fama-French 10×10 portfolio returns. The loading matrices are varimax rotated and normalized for ease in comparison.

Table 10 compares the out-of-sample performance of unconstrained and constrained matrix factor models using a 10-fold cross-validation (CV) for models with different number of factors. The table contains the average of the residual sum of squares (RSS), their ratios to the total sum of squares (RSS/TSS), and the number of parameters used in the 10-fold CV study. In this case, the prediction RSS of the constrained model is slightly larger than that of the unconstrained one with the same number of factors, which may results from the misspecification of the constrained matrices. Testing the adequacy of the constrained matrix is an important research topic that will be addressed in future research. On the other hand, the constrained model still has much less number of parameters than the unconstrained model.

Model	# Factor 1	# Factor 2	RSS	RSS/SST	# Parameters
Full	(3,3)		3064.40	0.500	60
	(3,4)		2905.79	0.474	70
	(3,6)		2644.59	0.431	90
Constrained	(3,3)		3115.16	0.508	24
Partial	(3,3)	(3,3)	2819.06	0.460	60
	(3,3)	(1,1)	3079.79	0.502	36
Full	(3,2)		3316.55	0.541	50
	(3,4)		2905.79	0.474	70
Constrained	(3,2)		3361.03	0.548	18
Partial	(3,2)	(3,2)	3169.79	0.517	50
	(3,2)	(1,1)	3323.25	0.542	31
Full	(2,3)		3269.50	0.533	50
	(2,4)		3152.63	0.514	60
	(2,6)		2976.18	0.431	90
Constrained	(2,3)		3372.79	0.550	18
Partial	(2,3)	(2,3)	3154.36	0.514	50
	(2,3)	(1,2)	3296.73	0.538	37
Full	(2,2)		3473.32	0.567	40
	(2,3)		3269.50	0.533	50
	(2,4)		3152.63	0.514	60
Constrained	(2,2)		3535.56	0.577	16
Partial	(2,2)	(2,2)	3415.25	0.557	40
	(2,2)	(2,1)	3486.15	0.569	33

Table 10: Performance of out-of-sample 10-fold CV of constrained and unconstrained factor models using Fama-French 10×10 portfolio return series, where RSS and RSS/TSS denote, respectively, the residual and total sum of squares.

7 Summary

This paper established a general framework for incorporating domain or prior knowledge induced linear constraints in the matrix factor model. We developed efficient estimation procedures for multi-term and partially constrained matrix factor models as well as the constrained model. Constraints can be used to achieve parsimony in parameterization, to facilitate factor interpretation, and to target specific factors indicated by the domain theories. We derived asymptotic theorems justifying the benefits of imposing constraints. Simulation results confirmed the advantages of constrained matrix factor model over the unconstrained one in finite samples. Finally, we illustrated the applications of constrained matrix factor models with three real data sets, where the constrained factor models outperform their unconstrained counterparts in explaining the variabilities of the data using out-of-sample 10-fold cross validation and in factor interpretation.

References

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

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

Bai, J. & Ng, S. (2007), 'Determining the number of primitive shocks in factor models', *Journal of Business & Economic Statistics* **25**(1), 52–60.

Bansal, N., Connolly, R. A. & Stivers, C. (2014), 'The stock-bond return relation, the term structures slope, and asset-class risk dynamics', *Journal of Financial and Quantitative Analysis* **49**(3), 699–724.

Chamberlain, G. (1983), 'Funds, Factors, and Diversification in Arbitrage Pricing Models', *Econometrica* **51**(5), 1305–23.

Chamberlain, G. & Rothschild, M. (1983), 'Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets', *Econometrica* **51**(5), 1281–304.

Chang, J., Guo, B. & Yao, Q. (2015), 'High dimensional stochastic regression with latent factors, endogeneity and nonlinearity', *Journal of Econometrics* **189**(2), 297–312.

Diebold, F. X. & Li, C. (2006), 'Forecasting the term structure of government bond yields', *Journal of econometrics* **130**(2), 337–364.

Diebold, F. X., Piazzesi, M. & Rudebusch, G. (2005), Modeling bond yields in finance and macroeconomics, Technical report, National Bureau of Economic Research.

Diebold, F. X., Rudebusch, G. D. & Aruoba, S. B. (2006), 'The macroeconomy and the yield curve: a dynamic latent factor approach', *Journal of econometrics* **131**(1), 309–338.

Fan, J. & Li, R. (2001), 'Variable selection via nonconcave penalized likelihood and its oracle properties', *Journal of the American statistical Association* **96**(456), 1348–1360.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2000), 'The generalized dynamic-factor model: Identification and estimation', *Review of Economics and statistics* **82**(4), 540–554.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2004), 'The generalized dynamic factor model consistency and rates', *Journal of Econometrics* **119**(2), 231–255.

Frank, L. E. & Friedman, J. H. (1993), 'A statistical view of some chemometrics regression tools', *Technometrics* **35**(2), 109–135.

French, K. R. (2017), '100 Portfolios Formed on Size and Book-to-Markete', `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_100_port_sz.html`. [Online; Accessed 01-Jan-2017].

Geyer, C. J. (1991), 'Constrained maximum likelihood exemplified by isotonic convex logistic regression', *Journal of the American Statistical Association* **86**(415), 717–724.

Hathaway, R. J. (1985), 'A constrained formulation of maximum-likelihood estimation for normal mixture distributions', *The Annals of Statistics* pp. 795–800.

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

Lam, C., Yao, Q. & Bathia, N. (2011), 'Estimation of latent factors for high-dimensional time series', *Biometrika* **98**(4), 901–18.

Liu, X. & Chen, R. (2016), 'Regime-switching factor models for high-dimensional time series', *Statistica Sinica* **26**, 1427–1451.

Liu, Y., Zhang, H. H., Park, C. & Ahn, J. (2007), 'Support vector machines with adaptive lq penalty', *Computational Statistics & Data Analysis* **51**(12), 6380–6394.

Nelson, C. R. & Siegel, A. F. (1987), 'Parsimonious modeling of yield curves', *Journal of business* pp. 473–489.

Pan, J. & Yao, Q. (2008), 'Modelling multiple time series via common factors', *Biometrika* pp. 365–379.

Rudebusch, G. D. & Wu, T. (2008), 'A macro-finance model of the term structure, monetary policy and the economy', *The Economic Journal* **118**(530), 906–926.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 267–288.

Tsai, H. & Tsay, R. S. (2010), 'Constrained factor models', *Journal of the American Statistical Association* **105**(492), 1593–1605.

Tsai, H., Tsay, R. S., Lin, E. M. & Cheng, C.-W. (2016), 'Doubly constrained factor models with applications', *Statistica Sinica* **26**, 1453–1478.

Wang, D., Liu, X. & Chen, R. (2017), Factor models for matrix-valued high-dimensional time series, Technical report.

Zhang, H. H. & Lu, W. (2007), 'Adaptive lasso for cox's proportional hazards model', *Biometrika* **94**(3), 691–703.

Zou, H. (2006), 'The adaptive lasso and its oracle properties', *Journal of the American statistical association* **101**(476), 1418–1429.

Appendix A Constraint Matrices

We first consider **covariate-induced constraint matrices**, using only dummy variables. Constrained matrix factor model based on continuous variables is out of the scope of this article. As an illustration we consider a tiny data set of corporate financial matrix-valued time series and constraint matrices for companies (rows). Suppose we have 8 companies, which can be grouped according to their industrial classification (Tech and Retail) and also their market capitalization (Large and Medium). The two groups form 2×2 combinations as shown below,

		Market Cap	
		1. Large	2. Medium
Industry	1. Tech	Apple, Microsoft	Brocade, FireEye
	2. Retail	Walmart, Target	JC Penny, Kohl's

	Industry	Market Cap
Apple	1	1
Microsoft	1	1
Brocade	1	2
FireEye	1	2
Walmart	2	1
Target	2	1
JC Penny	2	2
Kohl's	2	2

Constraint matrix $\boldsymbol{H}_R^{(1)}$ in Table 11 utilizes only industrial classification. To combine both industrial classification and market cap information, we first consider an additive model constraint on the $8 \times k_1$ ($k_1 \leq 3$) loading matrix $\boldsymbol{\Lambda}$ in model (1). The additive model constraint means that the i-th row of $\boldsymbol{\Lambda}$, that is, the loadings of k_1 row factors on the i-th variable, must have the form $\boldsymbol{\lambda}_{i\cdot} = \boldsymbol{u}_{j\cdot} + \boldsymbol{v}_{l\cdot}$, where the i-th variable falls in group $(Industry_j, MarketCap_l)$, k_1-dimensional vectors $\boldsymbol{u}_{j\cdot}$ and $\boldsymbol{v}_{l\cdot}$ are the loadings of k_1 row factors on the j-th market cap group and l-th industrial group, respectively. The most obvious way to express the additive model constraint is to use row constraints $\boldsymbol{H}_R^{(2)}$ in Table 11. Then, in the constrained matrix factor model (2), $\boldsymbol{H}_R = \boldsymbol{H}_R^{(2)}$ and $\boldsymbol{R} = (\boldsymbol{u}_{1\cdot}, \boldsymbol{u}_{2\cdot}, \boldsymbol{v}_{1\cdot}, \boldsymbol{v}_{2\cdot})'$.

$$
\boldsymbol{H}_R^{(1)} = \begin{array}{|cc|}
\hline
1 & \\
1 & \\
1 & \\
1 & \\
\hline
 & 1 \\
 & 1 \\
 & 1 \\
 & 1 \\
\hline
\end{array}
\qquad
\boldsymbol{H}_R^{(2)} = \begin{array}{|cc|cc|}
\hline
1 & & 1 & \\
1 & & 1 & \\
1 & & & 1 \\
1 & & & 1 \\
\hline
 & 1 & 1 & \\
 & 1 & 1 & \\
 & 1 & & 1 \\
 & 1 & & 1 \\
\hline
\end{array}
\qquad
\boldsymbol{H}_R^{(3)} = \begin{array}{|cc|cc|c|}
\hline
1 & & 1 & & 1 \\
1 & & 1 & & 1 \\
1 & & & 1 & -1 \\
1 & & & 1 & -1 \\
\hline
 & 1 & 1 & & -1 \\
 & 1 & 1 & & -1 \\
 & 1 & & 1 & 1 \\
 & 1 & & 1 & 1 \\
\hline
\end{array}
$$

Table 11: Illustration of constraint matrices constructed from grouping information by additive model.

Further, we consider the constraint incorporating an interaction term between industry and market cap grouping information. Now the i-th row of $\boldsymbol{\Lambda}$ has the form $\boldsymbol{\lambda}_{i\cdot} = \boldsymbol{u}_{j\cdot} + \boldsymbol{v}_{l\cdot} + \alpha_{j,l}\boldsymbol{w}$, where $\boldsymbol{w}$ is the k_1-dimensional interaction vector containing loadings of k_1 row factors and α_{ij} is the interaction term determined by $\boldsymbol{u}_{j\cdot}$ and $\boldsymbol{v}_{l\cdot}$ jointly. For example,

$$
\alpha_{j,l} = \begin{cases} 1 & if \quad j = l = 1 \text{ or } 2, \\ -1 & if \quad j = 1, l = 2 \text{ or vice versa.} \end{cases}
$$

In this case, for the constrained matrix factor model (2), $\boldsymbol{H}_R = \boldsymbol{H}_R^{(3)}$ and $\boldsymbol{R} = (\boldsymbol{u}_{1\cdot}, \boldsymbol{u}_{2\cdot}, \boldsymbol{v}_{1\cdot}, \boldsymbol{v}_{2\cdot}, \boldsymbol{w})'$. Note that $\boldsymbol{H}_R^{(2)}$ and $\boldsymbol{H}_R^{(3)}$ here are not full column rank and can be reduced to a full column rank matrix satisfying the requirement in Section 3. But the presentations of $\boldsymbol{H}_R^{(2)}$ and $\boldsymbol{H}_R^{(3)}$ are sufficient to illustrate the ideas of constructing complex constraint matrices.

To illustrate a **theory-induced constraint matrix**, we consider the yield curve latent factors model. Nelson & Siegel (1987) propose the Nelson-Siegel representation of the yield curve using a variation of the three-component exponential approximation to the cross-section of yields at any moment in time,

$$
y(\tau) = \beta_1 + \beta_2 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_3 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right),
$$

where $\boldsymbol{y}(\tau)$ denotes the set of zero-coupon yields and τ denotes time to maturity.

Diebold & Li (2006) and Diebold et al. (2006) interpret the Nelson-Siegel representation as a dynamic latent factor model where β_1, β_2, and β_3 are time-varying latent factors that capture the level (L), slope (S), and curvature (C) of the yield curve at each period t, while the terms that multiply the factors are respective factor loadings, that is

$$
y(\tau) = L_t + S_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + C_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right).
$$

The factor L_t may be interpreted as the overall level of the yield curve since its loading is equal for all maturities. The factor S_t, representing the slope of the yield curve, has a maximum loading (equal to 1) at the shortest maturity and then monotonically decays through zero as maturities increase. And the factor C_t has a loading that is 0 at the shortest maturity, increases to an intermediate maturity and then falls back to 0 as maturities increase. Hence, S_t and C_t capture the short-end and medium-term latent components of the yield curve. The coefficient λ controls the rate of decay of the loading of C_t and the maturity where S_t has maximum loading.

Multinational yield curve can be represented as a matrix time series $\{\boldsymbol{Y}_t\}_{t=1,\ldots,T}$, where rows of $\boldsymbol{Y}_t$ represent time to maturity and columns of $\boldsymbol{Y}_t$ denotes countries. To capture the characteristics of loading matrix specific to the level, slope, and curvature factors, we could set row loading constraint matrix to, for example, $\boldsymbol{H}_R = [\boldsymbol{h}_1, \boldsymbol{h}_2, \boldsymbol{h}_3]$, where $\boldsymbol{h}_1 = (1,1,1,1,1)'$, $bh_2 = (1,1,0,-1,-1)'$ and $\boldsymbol{h}_3 = (-1,0,2,0,-1)$. In Section 5, we try to mimic multinational yield curve and generate our samples from this type of constraints.

Appendix B Proofs

In what follows, let $\|\boldsymbol{A}\|_1$, $\|\boldsymbol{A}\|_2$ and $\|\boldsymbol{A}\|_F$ denote the L_1, spectral, and Frobenius norms of the matrix $\boldsymbol{A}$, respectively. They are defined as $\|\boldsymbol{A}\|_1 = \max_i \sum_j |\boldsymbol{A}_{ij}|$, $\|\boldsymbol{A}\|_2 = \sqrt{\lambda_{max}(\boldsymbol{A}'\boldsymbol{A})}$, and $\|\boldsymbol{A}\|_F = \sqrt{tr(\boldsymbol{A}'\boldsymbol{A})}$. Note that for a vector $\boldsymbol{a}$, both $\|\boldsymbol{a}\|_2$ and $\|\boldsymbol{a}\|_F$ are equal to the Euclidean norm. $\|\boldsymbol{A}\|_{min}$ denotes the positive square root of the minimal eigenvalue of $\boldsymbol{A}'\boldsymbol{A}$ or $\boldsymbol{A}\boldsymbol{A}'$, whichever is a smaller matrix. When $\boldsymbol{A}$ is a square matrix, we denote by $tr(\boldsymbol{A})$, $\lambda_{max}(\boldsymbol{A})$ and $\lambda_{max}(\boldsymbol{A})$ the trace, maximum and minimum eigenvalues of the matrix $\boldsymbol{A}$, respectively. For two sequences a_N and b_N, we write $a_N \asymp b_N$ if $a_N = O(b_N)$ and $b_N = O(a_N)$.

We define the following notation. For $h \geq 0$, let $\boldsymbol{\Sigma}_{f,u}(h) = Cov(vec(\boldsymbol{F}_t), vec(\boldsymbol{U}_{t+h}))$,

$$\tilde{\boldsymbol{\Sigma}}_{f,u}(h) = \frac{1}{T-h}\sum_{t=1}^{T-h} vec(\boldsymbol{F}_t)vec(\boldsymbol{U}_{t+h})', \quad \text{and} \quad \tilde{\boldsymbol{\Sigma}}_y(h) = \frac{1}{T-h}\sum_{t=1}^{T-h} vec(\boldsymbol{Y}_t)vec(\boldsymbol{Y}_{t+h})'.$$

The auto-covariance matrices of $\boldsymbol{\Sigma}_{u,f}(h)$, $\boldsymbol{\Sigma}_f(h)$, $\boldsymbol{\Sigma}_u(h)$ and their sample versions are defined in a similar manner. The following regularity and factor strength conditions are needed.

Condition 1. No linear combination of the components of $\boldsymbol{F}_t$ is white noise.

Condition 2. There exists at least one $h = 1, \ldots, h_0$, where $h_0 \geq 1$ is a positive integer, such that $\sum_{i=1}^{m_2}\sum_{j=1}^{m_2} \boldsymbol{\Omega}_{zq,ij}(h)\boldsymbol{\Omega}_{zq,ij}(h)'$ in equation (9) is of full rank.

Condition 1 is natural, as all the white noise linear combinations of $\boldsymbol{F}_t$ should be absorbed into $\boldsymbol{U}_t$, which ensures that there exists at least one $h \geq 1$ for which $\boldsymbol{\Omega}_{zq,ij}(h)$ is full-ranked. Condition 2 further ensures that $\boldsymbol{M}$ has k_1 positive eigenvalues.

Condition 3. For $h \geq 0$, the maximum eigenvalue of $\boldsymbol{\Sigma}_{f,u}(h)$ and $\boldsymbol{\Sigma}_u$ remains bounded as T, p_1 and p_2 increase to infinity.

In model (2), $\boldsymbol{H}_R \boldsymbol{R} \boldsymbol{F}_t \boldsymbol{C}' \boldsymbol{H}_C'$ can be viewed as the signal part of the observation $\boldsymbol{Y}_t$, and $\boldsymbol{U}_t$ as the noise. Condition 3 requires two things. First, each element of $\boldsymbol{\Sigma}_u$ remains bounded as p_1 and p_2 increase to infinity. Thus each noise component does not goes to infinity so that the signals are not obscured by the noises. Second, as dimensions increase, the covariance matrix of noises does not have information concentrated in a few directions. Thus the noise part does not contain any useful information. This is reasonable since all the common components should be absorbed in the signal.

Condition 4. The vector-valued process $vec(\boldsymbol{F}_t)$ is α-mixing. For some $\gamma > 2$, the mixing coefficients satisfy the condition that

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

where $\alpha(h) = \sup_{\tau} \sup_{A \in \mathcal{F}_{-\infty}^{\tau}, B \in \mathcal{F}_{\tau+h}^{\infty}} |P(A \cap B) - P(A)P(B)|$ and $\mathcal{F}_{\tau}^{s}$ is the σ-field generated by $\{vec(\boldsymbol{F}_t) : \tau \leq t \leq s\}$.

Condition 5. Let $F_{t,ij}$ be the ij-th entry of $\boldsymbol{F}_t$. Then, $E(|F_{t,ij}|^{2\gamma}) \leq C$ for any $i = 1, \ldots, k_1$, $j = 1, \ldots, k_2$ and $t = 1, \ldots, T$, where C is a positive constant and γ is given in Condition 4. In addition, there exists an integer h satisfying $1 \leq h \leq h_0$ such that $\boldsymbol{\Sigma}_f(h)$ is of rank $k = \max(k_1, k_2)$ and $\|\boldsymbol{\Sigma}_f(h)\|_2 \asymp O(1) \asymp \sigma_k(\boldsymbol{\Sigma}_f(h))$. For $i = 1, \ldots, k_1$ and $j = 1, \ldots, k_2$, $\frac{1}{T-h} \sum_{t=1}^{T-h} Cov(F_{t,i\cdot}, F_{t+h,i\cdot}) \neq \boldsymbol{0}$ and $\frac{1}{T-h} \sum_{t=1}^{T-h} Cov(F_{t,\cdot j}, F_{t+h,\cdot j}) \neq \boldsymbol{0}$.

Condition 4 and Condition 5 specify that the latent process $\{\boldsymbol{F}_t\}_{t=1,\ldots,T}$ only needs to satisfy the mixing condition specified in Condition 4 instead of the stationary condition. And we make use of the auto-covariance structure of the latent process $\{\boldsymbol{F}_t\}_{t=1,\ldots,T}$ without assuming any specific model. These two features make our estimation procedure more attractive and general than the standard principal component analysis.

We focus on the case of orthogonal constraints. Results for the non-orthogonal case and the partially-constrained case are similar.

The constrained factor model is $\boldsymbol{Y}_t = \boldsymbol{H}_R \boldsymbol{R} \boldsymbol{F}_t \boldsymbol{C}' \boldsymbol{H}_C' + \boldsymbol{U}_t$. Suppose we have orthogonal constraints, that is $\boldsymbol{H}_R' \boldsymbol{H}_R = \boldsymbol{I}_{m_1}$ and $\boldsymbol{H}_C' \boldsymbol{H}_C = \boldsymbol{I}_{m_2}$, then the transformed $m_1 \times m_2$ data $\boldsymbol{X}_t = \boldsymbol{H}_R' \boldsymbol{X}_t \boldsymbol{H}_C = \boldsymbol{R} \boldsymbol{F}_t \boldsymbol{C}' + \boldsymbol{E}_t$, where $\boldsymbol{E}_t = \boldsymbol{H}_R' \boldsymbol{U}_t \boldsymbol{H}_C$ and $\boldsymbol{E}_t$ is still white noise process.

Lemma 1. *Under Condition 3, each element of* $\boldsymbol{\Sigma}_e = Cov(vec(\boldsymbol{E}))$ *is uniformly bounded as* p_1 *and* p_2 *increase to infinity.*

Proof.

$$\begin{aligned}\boldsymbol{\Sigma}_e &= Cov(vec(\boldsymbol{H}'_R \boldsymbol{U}_t \boldsymbol{H}_C)) \\ &= Cov((\boldsymbol{H}'_R \otimes \boldsymbol{H}'_C) \cdot vec(\boldsymbol{U}_t)) \\ &= (\boldsymbol{H}_R \otimes \boldsymbol{H}_C)' \cdot \boldsymbol{\Sigma}_u \cdot (\boldsymbol{H}_R \otimes \boldsymbol{H}_C).\end{aligned}$$

Let $\boldsymbol{A} = \boldsymbol{H}_R \otimes \boldsymbol{H}_C$. Since $\boldsymbol{H}_R$ and $\boldsymbol{H}_C$ are $p_1 \times m_1$ and $p_2 \times m_2$ orthogonal matrices respectively, $\boldsymbol{A}$ is a $p_1 p_2 \times m_1 m_2$ orthogonal matrix.

Let e_i be the i-th element of $vec(\boldsymbol{E}_t)$, $A_{\cdot i}$ be the i-th column vector of $\boldsymbol{A}$ for $i = 1, \ldots, m_1 m_2$, then the diagonal elements of $\boldsymbol{\Sigma}_e$ are

$$Var(e_i) = A'_{\cdot i} \boldsymbol{\Sigma}_u A_{\cdot i} \leq \ \lambda_{\max}(\boldsymbol{\Sigma}_u) \text{ for } i = 1, \ldots, m_1 m_2.$$

Condition 3 assumes $\lambda_{\max}(\boldsymbol{\Sigma}_u) \sim O(1)$, hence $Var(e) \sim O(1)$ for $i = 1, \ldots, m_1 m_2$.

And off-diagonal elements of $\boldsymbol{\Sigma}_e$ are

$$Cov(e_i, e_j) \leq Var(e_i)^{\frac{1}{2}} Var(e_j)^{\frac{1}{2}} \sim O(1) \text{ for } i \neq j,\ i, j = 1, \ldots, m_1 m_2.$$

Thus, each element of $\boldsymbol{\Sigma}_e$ remains bounded if the maximum eigenvalue of $\boldsymbol{\Sigma}_e = Cov(vec(\boldsymbol{E}))$ is bounded as p_1 and p_2 increase to infinity. □

Lemma 2. *Under the assumption that* $\boldsymbol{H}_R$ *and* $\boldsymbol{H}_C$ *are orthogonal. Condition 6 also ensures that* $\|\boldsymbol{R}\|_2^2 \asymp p_1^{1-\delta_1} \asymp \|\boldsymbol{R}\|_{min}^2$ *and* $\|\boldsymbol{C}\|_2^2 \asymp p_2^{1-\delta_1} \asymp \|\boldsymbol{C}\|_{min}^2$.

Proof. For any orthogonal matrix $\boldsymbol{H}$, we have $\|\boldsymbol{H}\boldsymbol{R}\|_2^2 = \|\boldsymbol{R}\|_2^2$ and $\|\boldsymbol{H}\boldsymbol{R}\|_{min}^2 = \|\boldsymbol{R}\|_{min}^2$. And the results follow. □

In the following proofs, we work with the transformed model (7), as in $\boldsymbol{X}_t = \boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}' + \boldsymbol{E}_t$ where $\boldsymbol{X}_t$ and $\boldsymbol{E}_t$ are $m_1 \times m_2$ matrices, $\boldsymbol{F}_t$ is $k_1 \times k_2$ matrix, $\boldsymbol{R}$ is the $m_1 \times k_1$ row loading matrix, and $\boldsymbol{C}$ is the $m_2 \times k_2$ column loading matrix for the transformed model.

We start by defining some quantities used in the proofs. Write

$$\begin{aligned}
\boldsymbol{\Omega}_{s,ij}(h) &= \frac{1}{T-h}\sum_{t=1}^{T-h} Cov(\boldsymbol{R}\boldsymbol{F}_t C_{i\cdot}, \boldsymbol{R}\boldsymbol{F}_{t+h} C_{j\cdot}),\\
\boldsymbol{\Omega}_{fc,ij}(h) &= \frac{1}{T-h}\sum_{t=1}^{T-h} Cov(\boldsymbol{F}_t C_{i\cdot}, \boldsymbol{F}_{t+h} C_{j\cdot}),\\
\hat{\boldsymbol{\Omega}}_{s,ij}(h) &= \frac{1}{T-h}\sum_{t=1}^{T-h} \boldsymbol{R}\boldsymbol{F}_t C_{i\cdot} C'_{j\cdot} \boldsymbol{F}'_{t+h} \boldsymbol{R}',\\
\hat{\boldsymbol{\Omega}}_{se,ij}(h) &= \frac{1}{T-h}\sum_{t=1}^{T-h} \boldsymbol{R}\boldsymbol{F}_t C_{i\cdot} E'_{t+h,\cdot j},\\
\hat{\boldsymbol{\Omega}}_{es,ij}(h) &= \frac{1}{T-h}\sum_{t=1}^{T-h} E_{t,\cdot j} C'_{i\cdot} \boldsymbol{F}'_{t+h} \boldsymbol{R}',\\
\hat{\boldsymbol{\Omega}}_{e,ij}(h) &= \frac{1}{T-h}\sum_{t=1}^{T-h} E_{t,\cdot j} E'_{t+h,\cdot j},\\
\hat{\boldsymbol{\Omega}}_{fc,ij}(h) &= \frac{1}{T-h}\sum_{t=1}^{T-h} \boldsymbol{F}_t C_{i\cdot} C'_{j\cdot} \boldsymbol{F}'_{t+h}.
\end{aligned}$$

The following Lemma 3 from Wang et al. (2017) establishes the entry-wise convergence rate of the covariance matrix estimation of the vectorized latent factor process $vec(\boldsymbol{F}_t)$.

Lemma 3. *Let $F_{t,ij}$ denote the ij-th entry of $\boldsymbol{F}_t$. Under Condition 4 and Condition 5, for any $i,k = 1,\dots,k_1$ and $j,l = 1,\cdots,k_2$, we have*

$$\left| \frac{1}{T-h}\sum_{t=1}^{T-h} \left(F_{t,ij}F_{t+h,kl} - Cov(F_{t,ij}F_{t+h,kl})\right) \right| = O_p(T^{-1/2}). \tag{16}$$

Under the matrix-variate factor Model (7), the $\boldsymbol{R}\boldsymbol{F}_t\boldsymbol{C}'$ is the signal and $\boldsymbol{E}_t$ is the noise.

Lemma 4. *Under Conditions 1-6, it holds that*

$$\begin{aligned}
\sum_{i=1}^{m_2}\sum_{j=1}^{m_2} \|\widehat{\boldsymbol{\Omega}}_{s,ij}(h) - \boldsymbol{\Omega}_{s,ij}(h)\|_2^2 &= O_p(p_1^{2-2\delta_1} p_2^{2-2\delta_2} T^{-1}),, \quad (17)\\
\sum_{i=1}^{m_2}\sum_{j=1}^{m_2} \|\widehat{\boldsymbol{\Omega}}_{se,ij}(h) - \boldsymbol{\Omega}_{se,ij}(h)\|_2^2 &= O_p(m_1 p_1^{1-\delta_1} m_2 p_2^{1-\delta_2} T^{-1}),, \quad (18)\\
\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}, \|\widehat{\boldsymbol{\Omega}}_{es,ij}(h) - \boldsymbol{\Omega}_{es,ij}(h)\|_2^2 &= O_p(m_1 p_1^{1-\delta_1} m_2 p_2^{1-\delta_2} T^{-1}),, \quad (19)\\
\sum_{i=1}^{m_2}\sum_{j=1}^{m_2} \|\widehat{\boldsymbol{\Omega}}_{e,ij}(h) - \boldsymbol{\Omega}_{e,ij}(h)\|_2^2 &= O_p(m_1^2 m_2^2 T^{-1}). \quad (20)
\end{aligned}$$

Proof. To prove the convergence rate of $\widehat{\mathbf{\Omega}}_{s,ij}(h)$ in (17), we first establish the convergence rate of estimating $\mathbf{\Omega}_{fc,ij}(h) = \frac{1}{T-h}\sum_{t=1}^{T-h} Cov(\boldsymbol{F}_t C_{i\cdot}, \boldsymbol{F}_{t+h} C_{j\cdot})$.

$$
\begin{aligned}
&\|\widehat{\mathbf{\Omega}}_{fc,ij}(h) - \mathbf{\Omega}_{fc,ij}(h)\|_2^2 \leq \|\widehat{\mathbf{\Omega}}_{fc,ij}(h) - \mathbf{\Omega}_{fc,ij}(h)\|_F^2 \\
&= \left\| \frac{1}{T-h} \sum_{t=1}^{T-h} (\boldsymbol{F}_{t+h} \otimes \boldsymbol{F}_t - E(\boldsymbol{F}_{t+h} \otimes \boldsymbol{F}_t)) \cdot vec(C_{i\cdot} C'_{j\cdot}) \right\|_2^2 \\
&\leq \left\| \frac{1}{T-h} \sum_{t=1}^{T-h} (\boldsymbol{F}_{t+h} \otimes \boldsymbol{F}_t - E(\boldsymbol{F}_{t+h} \otimes \boldsymbol{F}_t)) \right\|_F^2 \cdot \|C_{i\cdot}\|_2^2 \cdot \|C_{j\cdot}\|_2^2. \qquad (21)
\end{aligned}
$$

Hence, we have

$$
\begin{aligned}
&\sum_{i=1}^{m_2} \sum_{j=1}^{m_2} \|\widehat{\mathbf{\Omega}}_{s,ij}(h) - \mathbf{\Omega}_{s,ij}(h)\|_2^2 \\
&= \sum_{i=1}^{m_2} \sum_{j=1}^{m_2} \|\boldsymbol{R} \cdot (\widehat{\mathbf{\Omega}}_{fc,ij}(h) - \mathbf{\Omega}_{fc,ij}(h)) \cdot \boldsymbol{R}'\|_2^2 \\
&\leq \|\boldsymbol{R}\|_2^4 \cdot \left\| \frac{1}{T-h} \sum_{t=1}^{T-h} (\boldsymbol{F}_{t+h} \otimes \boldsymbol{F}_t - E(\boldsymbol{F}_{t+h} \otimes \boldsymbol{F}_t)) \right\|_F^2 \cdot \left(\sum_{i=1}^{m_2} \|C_{i\cdot}\|_2^2 \right)^2 \\
&= \|\boldsymbol{R}\|_2^4 \cdot \left\| \frac{1}{T-h} \sum_{t=1}^{T-h} (\boldsymbol{F}_{t+h} \otimes \boldsymbol{F}_t - E(\boldsymbol{F}_{t+h} \otimes \boldsymbol{F}_t)) \right\|_F^2 \cdot \|\boldsymbol{C}\|_F^4 \\
&\leq k_2^2 \, \|\boldsymbol{R}\|_2^4 \cdot \left\| \frac{1}{T-h} \sum_{t=1}^{T-h} (\boldsymbol{F}_{t+h} \otimes \boldsymbol{F}_t - E(\boldsymbol{F}_{t+h} \otimes \boldsymbol{F}_t)) \right\|_F^2 \cdot \|\boldsymbol{C}\|_2^4 \\
&= O_p(p_1^{2-2\delta_1} p_2^{2-2\delta_2} T^{-1}).
\end{aligned}
$$

The first inequality comes from (21) and the last inequality follows from Condition 6 and Lemma 1.

To prove the convergence rate of covariance between signal at t and noise at $t+h$ in (18), we first establish the convergence rate of covariance between $\boldsymbol{F}_t C_{i\cdot}$ and $E_{t+h,\cdot j}$.

$$
\begin{aligned}
&\left\| \frac{1}{T-h} \sum_{t=1}^{T-h} \boldsymbol{F}_t C_{i\cdot} E'_{t+h,\cdot j} - \frac{1}{T-h} \sum_{t=1}^{T-h} E\left(\boldsymbol{F}_t C_{i\cdot} E'_{t+h,\cdot j}\right) \right\|_2^2 \\
&\leq \left\| \frac{1}{T-h} \sum_{t=1}^{T-h} vec(\boldsymbol{F}_t C_{i\cdot} E'_{t+h,\cdot j}) - \frac{1}{T-h} \sum_{t=1}^{T-h} E\left(vec(\boldsymbol{F}_t C_{i\cdot} E'_{t+h,\cdot j})\right) \right\|_2^2 \\
&\leq \| \frac{1}{T-h} \sum_{t=1}^{T-h} \left(E_{t+h,\cdot j} \otimes \boldsymbol{F}_t - E\left(E_{t+h,\cdot j} \otimes \boldsymbol{F}_t\right)\right) \cdot vec(C_{i\cdot}) \|_2^2 \\
&\leq \| \frac{1}{T-h} \sum_{t=1}^{T-h} \left(E_{t+h,\cdot j} \otimes \boldsymbol{F}_t - E\left(E_{t+h,\cdot j} \otimes \boldsymbol{F}_t\right)\right) \|_2^2 \cdot \|C_{i\cdot}\|_2^2.
\end{aligned}
$$

Hence, we have

$$
\begin{aligned}
&\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\|\widehat{\boldsymbol{\Omega}}_{se,ij}(h)-\boldsymbol{\Omega}_{se,ij}(h)\|_2^2\\
&=\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\left\|\frac{1}{T-h}\sum_{t=1}^{T-h}\boldsymbol{R}\boldsymbol{F}_tC_{i\cdot}E'_{t+h,\cdot j}-\frac{1}{T-h}\sum_{t=1}^{T-h}E\left(\boldsymbol{R}\boldsymbol{F}_tC_{i\cdot}E'_{t+h,\cdot j}\right)\right\|_2^2\\
&\le\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\|\boldsymbol{R}\|_2^2\cdot\left\|\frac{1}{T-h}\sum_{t=1}^{T-h}\boldsymbol{F}_tC_{i\cdot}E'_{t+h,\cdot j}-\frac{1}{T-h}\sum_{t=1}^{T-h}E\left(\boldsymbol{F}_tC_{i\cdot}E'_{t+h,\cdot j}\right)\right\|_2^2\\
&\le\|\boldsymbol{R}\|_2^2\cdot\sum_{i=1}^{m_2}\|\frac{1}{T-h}\sum_{t=1}^{T-h}\left(E_{t+h,\cdot j}\otimes\boldsymbol{F}_t-E\left(E_{t+h,\cdot j}\otimes\boldsymbol{F}_t\right)\right)\|_2^2\cdot\sum_{j=1}^{m_2}\|C_{i\cdot}\|_2^2\\
&\le k_2\|\boldsymbol{R}\|_2^2\cdot\sum_{i=1}^{m_2}\|\frac{1}{T-h}\sum_{t=1}^{T-h}\left(E_{t+h,\cdot j}\otimes\boldsymbol{F}_t-E\left(E_{t+h,\cdot j}\otimes\boldsymbol{F}_t\right)\right)\|_2^2\cdot\sum_{j=1}^{m_2}\|\boldsymbol{C}\|_2^2\\
&=O_p(m_1p_1^{1-\delta_1}m_2p_2^{1-\delta_2}T^{-1}).
\end{aligned}
$$

To prove the convergence rate of covariance between noise at t and signal at $t+h$ in (19), we use similar arguments and get

$$
\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\|\widehat{\boldsymbol{\Omega}}_{es,ij}(h)-\boldsymbol{\Omega}_{es,ij}(h)\|_2^2=O_p(m_1p_1^{1-\delta_1}m_2p_2^{1-\delta_2}T^{-1/2}).
$$

.

And the convergence rate of $\widehat{\boldsymbol{\Omega}}_{e,ij}(h)$ in (20) is given by

$$
\begin{aligned}
&\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\|\widehat{\boldsymbol{\Omega}}_{e,ij}(h)-\boldsymbol{\Omega}_{e,ij}(h)\|_2^2\\
&=\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\left\|\frac{1}{T-h}\sum_{t=1}^{T-h}E_{t,\cdot i}E'_{t+h,\cdot j}\right\|\\
&=O_p(m_1^2m_2^2T^{-1}).
\end{aligned}
$$

□

With the four rates established in Lemma 4, we can study the rate of convergence for the transformed observed covariance matrix $\widehat{\boldsymbol{\Omega}}_{x,ij}(h)$.

Lemma 5. *Under Conditions 1-6, it holds that*

$$
\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\|\widehat{\boldsymbol{\Omega}}_{x,ij}(h)-\boldsymbol{\Omega}_{x,ij}(h)\|_2^2=O_p\left(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2},\ m_1^2m_2^2)\cdot T^{-1}\right). \qquad (22)
$$

Proof. By definition of $\widehat{\mathbf{\Omega}}_{x,ij}(h)$ in Section 3, we can decompose $\widehat{\mathbf{\Omega}}_{x,ij}(h)$ into the following four parts,

$$\begin{aligned}\widehat{\mathbf{\Omega}}_{x,ij}(h) &= \frac{1}{T-h}\sum_{t=1}^{T-h} X_{t,\cdot i}X'_{t+h,\cdot j}\\ &= \frac{1}{T-h}\sum_{t=1}^{T-h}(\boldsymbol{RF}_t C_{i\cdot} + E_{t,i\cdot})(\boldsymbol{RF}_t C_{i\cdot} + E_{t+h,j\cdot})'\\ &= \widehat{\mathbf{\Omega}}_{s,ij}(h) + \widehat{\mathbf{\Omega}}_{se,ij}(h) + \widehat{\mathbf{\Omega}}_{es,ij}(h) + \widehat{\mathbf{\Omega}}_{e,ij}(h).\end{aligned}$$

Thus from Lemma 4, we have

$$\begin{aligned}&\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\|\widehat{\mathbf{\Omega}}_{x,ij}(h) - \mathbf{\Omega}_{x,ij}(h)\|_2^2\\ &\leq 4\ \sum_{i=1}^{m_2}\sum_{j=1}^{m_2}(\|\widehat{\mathbf{\Omega}}_{s,ij}(h) - \mathbf{\Omega}_{s,ij}(h)\|_2^2 + \|\widehat{\mathbf{\Omega}}_{se,ij}(h) - \mathbf{\Omega}_{se,ij}(h)\|_2^2\\ &\qquad + \|\widehat{\mathbf{\Omega}}_{es,ij}(h) - \mathbf{\Omega}_{es,ij}(h)\|_2^2 + \|\widehat{\mathbf{\Omega}}_{e,ij}(h) - \mathbf{\Omega}_{e,ij}(h))\|_2^2\\ &= O_p\left(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2},\ m_1^2 m_2^2)\cdot T^{-1}\right).\end{aligned}$$

□

Lemma 6. *Under Conditions 1-6 and* $m_1 p_1^{-1+\delta_1} m_2 p_2^{-1+\delta_2} T^{-1/2} = o_p(1)$*, it holds that*

$$\|\widehat{\boldsymbol{M}} - \boldsymbol{M}\|_2 = O_p\left(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2},\ m_1 p_1^{1-\delta_1} m_2 p_2^{1-\delta_2})\cdot T^{-1/2}\right). \tag{23}$$

Proof. By definitions of $\boldsymbol{M}$ in (11) and its sample version $\widehat{\boldsymbol{M}}$, we have

$$\begin{aligned}\|\widehat{\boldsymbol{M}} - \boldsymbol{M}\|_2 &= \left\|\sum_{h=1}^{h_0}\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}(\widehat{\mathbf{\Omega}}_{x,ij}(h)\widehat{\mathbf{\Omega}}'_{x,ij}(h) - \mathbf{\Omega}_{x,ij}(h)\mathbf{\Omega}'_{x,ij}(h))\right\|_2\\ &\leq \sum_{h=1}^{h_0}\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\left(\big\|(\widehat{\mathbf{\Omega}}_{x,ij}(h) - \mathbf{\Omega}_{x,ij}(h))(\widehat{\mathbf{\Omega}}_{x,ij}(h) - \mathbf{\Omega}_{x,ij}(h))'\big\|_2 + 2\big\|\mathbf{\Omega}_{x,ij}(h)\big\|_2\big\|\widehat{\mathbf{\Omega}}_{x,ij}(h) - \mathbf{\Omega}_{x,ij}(h)\big\|_2\right)\\ &= \sum_{h=1}^{h_0}\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\big\|\widehat{\mathbf{\Omega}}_{x,ij}(h) - \mathbf{\Omega}_{x,ij}(h)\big\|_2^2 + 2\sum_{h=1}^{h_0}\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\big\|\mathbf{\Omega}_{x,ij}(h)\big\|_2\big\|\widehat{\mathbf{\Omega}}_{x,ij}(h) - \mathbf{\Omega}_{x,ij}(h)\big\|_2.\end{aligned}$$

Now we investigate each item in the above formula.

$$
\begin{aligned}
&\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\|\boldsymbol{\Omega}_{x,ij}(h)\|_2^2 = \sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\|\boldsymbol{R}\boldsymbol{\Omega}_{fc,ij}(h)\boldsymbol{R}'\|_2^2 \\
&\leq \sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\|\boldsymbol{R}\|_2^4\cdot\|\boldsymbol{\Omega}_{fc,ij}(h)\|_2^2 \\
&\leq \sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\|\boldsymbol{R}\|_2^4\cdot\left\|\frac{1}{T-h}\sum_{t=1}^{T-h}E(\boldsymbol{F}_tC_{i\cdot}C_{j\cdot}'\boldsymbol{F}_{t+h}')\right\|_2^2 \\
&\leq \sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\|\boldsymbol{R}\|_2^4\cdot\left\|\frac{1}{T-h}\sum_{t=1}^{T-h}E\left(vec(\boldsymbol{F}_tC_{i\cdot}C_{j\cdot}'\boldsymbol{F}_{t+h}')\right)\right\|_2^2 \\
&= \sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\|\boldsymbol{R}\|_2^4\cdot\left\|\frac{1}{T-h}\sum_{t=1}^{T-h}E\left(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t\right)\cdot vec(C_{i\cdot}C_{j\cdot}')\right\|_2^2 \\
&\leq \sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\|\boldsymbol{R}\|_2^4\cdot\left\|\frac{1}{T-h}\sum_{t=1}^{T-h}E\left(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t\right)\right\|_2^2\cdot\|vec(C_{i\cdot}C_{j\cdot}')\|_2^2 \\
&\leq \sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\|\boldsymbol{R}\|_2^4\cdot\left\|\frac{1}{T-h}\sum_{t=1}^{T-h}E\left(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t\right)\right\|_2^2\cdot\|C_{i\cdot}C_{j\cdot}'\|_F^2 \\
&= \sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\|\boldsymbol{R}\|_2^4\cdot\left\|\frac{1}{T-h}\sum_{t=1}^{T-h}E\left(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t\right)\right\|_2^2\cdot\|C_{i\cdot}\|_2^2\|C_{j\cdot}'\|_2^2 \\
&= \|\boldsymbol{R}\|_2^4\cdot\left\|\frac{1}{T-h}\sum_{t=1}^{T-h}E\left(\boldsymbol{F}_{t+h}\otimes\boldsymbol{F}_t\right)\right\|_2^2\cdot\left(\sum_{i=1}^{m_2}\|C_{i\cdot}\|_2^2\right)^2 \\
&= O_p(p_1^{2-2\delta_1}p_2^{2-2\delta_2}).
\end{aligned}
$$

From Lemma 5, we have $\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\|\widehat{\boldsymbol{\Omega}}_{x,ij}(h)-\boldsymbol{\Omega}_{x,ij}(h)\|_2^2 = O_p\left(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2},\ m_1^2m_2^2)\cdot T^{-1}\right)$, then

$$
\begin{aligned}
&\left(\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\big\|\boldsymbol{\Omega}_{x,ij}(h)\big\|_2\big\|\widehat{\boldsymbol{\Omega}}_{x,ij}(h)-\boldsymbol{\Omega}_{x,ij}(h)\big\|_2\right)^2 \\
&\leq \sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\big\|\boldsymbol{\Omega}_{x,ij}(h)\big\|_2^2\cdot\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\big\|\widehat{\boldsymbol{\Omega}}_{x,ij}(h)-\boldsymbol{\Omega}_{x,ij}(h)\big\|_2^2 \\
&= O_p(p_1^{2-2\delta_1}p_2^{2-2\delta_2})\cdot O_p(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2},\ m_1^2m_2^2)\cdot T^{-1} \\
&= O_p(\max(p_1^{4-4\delta_1}p_2^{4-4\delta_2},\ m_1^2p_1^{2-2\delta_1}m_2^2p_2^{2-2\delta_2})\cdot T^{-1}.
\end{aligned}
$$

Thus, from the above results, Lemma 5 and the condition that $m_1p_1^{-1+\delta_1}m_2p_2^{-1+\delta_2}T^{-1/2} =$

$o_p(1)$, we have

$$\|\widehat{\boldsymbol{M}} - \boldsymbol{M}\|_2 = O_p\left(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2},\ m_1^2m_2^2)\cdot T^{-1}\right) + O_p\left(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2},\ m_1p_1^{1-\delta_1}m_2p_2^{1-\delta_2}\right)\cdot T^{-1/2}$$
$$= O_p\left(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2},\ m_1p_1^{1-\delta_1}m_2p_2^{1-\delta_2})\cdot T^{-1/2}\right).$$

□

Similar to the proof of Lemma 5 in Wang et al. (2017), we have

Lemma 7. *Under Condition 3 and Condition 5, we have*

$$\lambda_i(\boldsymbol{M}) \asymp p_1^{2-2\delta_1}p_2^{2-2\delta_2}, \qquad i = 1, 2, \ldots, k_1,$$

where $\lambda_i(\boldsymbol{M})$ denotes the i-th largest singular value of $\boldsymbol{M}$.

Proof of Theorem 1

Proof. By Lemma 3-7, and Lemma 3 in Lam et al. (2011), we have

$$\|\widehat{\boldsymbol{Q}}_1 - \boldsymbol{Q}_1\|_2 \le \frac{8}{\lambda_{min}(\boldsymbol{M})}\|\widehat{\boldsymbol{M}} - \boldsymbol{M}\|_2 = O_p\left(\max\left(T^{-1/2},\ \frac{m_1}{p_1^{1-\delta_1}}\frac{m_2}{p_2^{1-\delta_2}}T^{-1/2}\right)\right).$$

Proof for $\|\widehat{\boldsymbol{Q}}_2 - \boldsymbol{Q}_2\|_2$ is similar. □

Proof of Theorem 2

Proof. The proof is similar to that of Theorem 1 of Lam & Yao (2012). Let λ_j and $\boldsymbol{q}_j$ be the j-th largest eigenvalue and eigenvector of $\boldsymbol{M}$, respectively. The corresponding sample versions are denoted by $\widehat{\lambda}_j$ and $\widehat{\boldsymbol{q}}_j$ for the matrix $\widehat{\boldsymbol{M}}$. Let $\boldsymbol{Q}_1 = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{k_1})$, $\boldsymbol{B}_1 = (\boldsymbol{q}_{k_1+1}, \ldots, \boldsymbol{q}_{m_1})$, $\widehat{\boldsymbol{Q}}_1 = (\hat{\boldsymbol{q}}_1, \ldots, \hat{\boldsymbol{q}}_{k_1})$ and $\widehat{\boldsymbol{B}}_1 = (\hat{\boldsymbol{q}}_{k_1+1}, \ldots, \hat{\boldsymbol{q}}_{m_1})$.

Eigenvalues λ_j, $j = 1, \ldots, k_1$

For $j = 1, \ldots, k_1$, we have

$$\widehat{\lambda}_j - \lambda_j = \widehat{\boldsymbol{q}}_j'\widehat{\boldsymbol{M}}\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j'\boldsymbol{M}\boldsymbol{q}_j = I_1 + I_2 + I_3 + I_4 + I_5,$$

where

$$I_1 = (\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j)'(\widehat{\boldsymbol{M}} - \boldsymbol{M})\widehat{\boldsymbol{q}}_j,\ I_2 = (\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j)'\boldsymbol{M}(\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j), \tag{24}$$
$$I_3 = (\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j)'\boldsymbol{M}\boldsymbol{q}_j,\ I_4 = \boldsymbol{q}_j'(\widehat{\boldsymbol{M}} - \boldsymbol{M})\boldsymbol{q}_j,\ I_5 = \boldsymbol{q}_j'(\widehat{\boldsymbol{M}} - \boldsymbol{M})(\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j). \tag{25}$$

We have, from Theorem 1,

$$\|\widehat{\boldsymbol{q}}_j' - \boldsymbol{q}_j\|_2 \le \|\widehat{\boldsymbol{Q}}_1 - \boldsymbol{Q}_1\|_2 = O_p\left(\max\left(T^{-1/2},\ \frac{m_1}{p_1^{1-\delta_1}}\frac{m_2}{p_2^{1-\delta_2}}T^{-1/2}\right)\right),\ \textit{for } j = 1, \ldots, k_1.$$

And by Lemma 6, $\|\widehat{\boldsymbol{M}}-\boldsymbol{M}\|_2 = O_p\left(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2},\ m_1p_1^{1-\delta_1}m_2p_2^{1-\delta_2}\right)\cdot T^{-1/2}$.

Also from Lemma 7, we have $\|\boldsymbol{M}\|_2 = O_p(p_1^{2-2\delta_1}p_2^{2-2\delta_2})$.

Then,

$$
\begin{aligned}
&\|I_1\|_2 = \|(\widehat{\boldsymbol{q}}_j-\boldsymbol{q}_j)'(\widehat{\boldsymbol{M}}-\boldsymbol{M})\widehat{\boldsymbol{q}}_j\|_2 \le \|\widehat{\boldsymbol{q}}_j-\boldsymbol{q}_j\|_2\cdot\|\widehat{\boldsymbol{M}}-\boldsymbol{M}\|_2\cdot\|\widehat{\boldsymbol{q}}_j\|_2 = O_p\left(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2}, m_1^2m_2^2)\cdot T^{-1}\right)\\
&\|I_2\|_2 = \|(\widehat{\boldsymbol{q}}_j-\boldsymbol{q}_j)'\boldsymbol{M}(\widehat{\boldsymbol{q}}_j-\boldsymbol{q}_j)\|_2 \le \|\widehat{\boldsymbol{q}}_j-\boldsymbol{q}_j\|_2^2\cdot\|\boldsymbol{M}\|_2 = O_p\left(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2}, m_1^2m_2^2)\cdot T^{-1}\right)\\
&\|I_3\|_2 = \|\widehat{\boldsymbol{q}}_j-\boldsymbol{q}_j)'\boldsymbol{M}\boldsymbol{q}_j\|_2 \le \|\widehat{\boldsymbol{q}}_j'-\boldsymbol{q}_j\|_2\cdot\|\boldsymbol{M}\|_2\cdot\|\boldsymbol{q}_j\|_2 = O_p\left(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2}, m_1p_1^{1-\delta_1}m_2p_2^{1-\delta_2})\cdot T^{-1/2}\right)\\
&\|I_4\|_2 = \|\boldsymbol{q}_j'(\widehat{\boldsymbol{M}}-\boldsymbol{M})\boldsymbol{q}_j\|_2 \le \|\boldsymbol{q}_j\|_2\|\widehat{\boldsymbol{M}}-\boldsymbol{M}\|_2\|\boldsymbol{q}_j\|_2 = O_p\left(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2},\ m_1p_1^{1-\delta_1}m_2p_2^{1-\delta_2})\cdot T^{-1/2}\right)\\
&\|I_5\|_2 = \|\boldsymbol{q}_j'(\widehat{\boldsymbol{M}}-\boldsymbol{M})(\widehat{\boldsymbol{q}}_j-\boldsymbol{q}_j)\|_2 \le \|\boldsymbol{q}_j\|_2\|\widehat{\boldsymbol{M}}-\boldsymbol{M}\|_2\|\widehat{\boldsymbol{q}}_j'-\boldsymbol{q}_j\|_2\\
&\qquad = O_p\left(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2},\ m_1p_1^{1-\delta_1}m_2p_2^{1-\delta_2})\cdot T^{-1/2}\right).
\end{aligned}
$$

Thus, under the condition that $m_1p_1^{-1+\delta_1}m_2p_2^{-1+\delta_2}T^{-1/2} = o_p(1)$, we have

$$
|\hat{\lambda}_j-\lambda_j| = O_p\left(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2},\ m_1p_1^{1-\delta_1}m_2p_2^{1-\delta_2})\cdot T^{-1/2}\right), \qquad for \qquad j=1,\dots,k_1.
$$

Eigenvalues λ_j, $j=k_1+1,\dots,p_1$

Similar to proof of Theorem 1 with Lemma 3 in Lam et al. (2011), we have

$$
\|\widehat{\boldsymbol{B}}_1-\boldsymbol{B}_1\|_2 = O_p\left(\max\left(T^{-1/2},\ \frac{m_1}{p_1^{1-\delta_1}}\frac{m_2}{p_2^{1-\delta_2}}T^{-1/2}\right)\right).
$$

And hence

$$
\|\widehat{\boldsymbol{q}}_j'-\boldsymbol{q}_j\|_2 \le \|\widehat{\boldsymbol{Q}}_1-\boldsymbol{Q}_1\|_2 = O_p\left(\max\left(T^{-1/2},\ \frac{m_1}{p_1^{1-\delta_1}}\frac{m_2}{p_2^{1-\delta_2}}T^{-1/2}\right)\right),\ for\ j=k_1+1,\dots,p_1.
$$

Define $\widetilde{\boldsymbol{M}} = \sum_{h=1}^{h_0}\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\widehat{\boldsymbol{\Omega}}_{i,j}(h)\boldsymbol{\Omega}_{i,j}'(h)$, then

$$
\begin{aligned}
\|\widetilde{\boldsymbol{M}}-\boldsymbol{M}\| &= \left\|\sum_{h=1}^{h_0}\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\left(\widehat{\boldsymbol{\Omega}}_{ij}(h)\boldsymbol{\Omega}_{ij}'(h)-\boldsymbol{\Omega}_{ij}(h)\boldsymbol{\Omega}_{ij}'(h)\right)\right\|_2\\
&\le \sum_{h=1}^{h_0}\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\left\|\left(\widehat{\boldsymbol{\Omega}}_{ij}(h)-\boldsymbol{\Omega}_{ij}(h)\right)\right\|_2\left\|\boldsymbol{\Omega}_{ij}'(h)\right\|_2\\
&= O_p(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2},\ m_1p_1^{1-\delta_1}m_2p_2^{1-\delta_2})\cdot T^{-1/2},\ \text{from Lemma 6.}
\end{aligned}
$$

For $j=k_1+1,\dots,p_1$, since $\lambda_j=0$ we have

$$
\widehat{\lambda}_j = \widehat{\boldsymbol{q}}_j'\widehat{\boldsymbol{M}}\widehat{\boldsymbol{q}}_j = K_1+K_2+K_3,
$$

where $K_1 = \widehat{\boldsymbol{q}}_j'(\widehat{\boldsymbol{M}}-\widetilde{\boldsymbol{M}}-\widetilde{\boldsymbol{M}}'+\boldsymbol{M})\widehat{\boldsymbol{q}}_j$, $K_2 = 2\widehat{\boldsymbol{q}}_j'(\widetilde{\boldsymbol{M}}-\boldsymbol{M})(\widehat{\boldsymbol{q}}_j-\boldsymbol{q}_j)$ and $K_3 = (\widehat{\boldsymbol{q}}_j-\boldsymbol{q}_j)'\boldsymbol{M}(\widehat{\boldsymbol{q}}_j-\boldsymbol{q}_j)$.

Then,

$$\|K_1\|_2 = \left\|\widehat{\boldsymbol{q}}_j' \sum_{h=1}^{h_0}\sum_{i=1}^{m_2}\sum_{j=1}^{m_2} \left(\widehat{\boldsymbol{\Omega}}_{ij}(h)\widehat{\boldsymbol{\Omega}}_{ij}'(h) - \widehat{\boldsymbol{\Omega}}_{ij}(h)\boldsymbol{\Omega}_{ij}'(h) - \boldsymbol{\Omega}_{ij}(h)\widehat{\boldsymbol{\Omega}}_{ij}'(h) + \boldsymbol{\Omega}_{ij}(h)\boldsymbol{\Omega}_{ij}'(h)\right)\widehat{\boldsymbol{q}}_j\right\|_2$$

$$= \left\|\widehat{\boldsymbol{q}}_j' \sum_{h=1}^{h_0}\sum_{i=1}^{m_2}\sum_{j=1}^{m_2} \left(\widehat{\boldsymbol{\Omega}}_{ij}(h) - \boldsymbol{\Omega}_{ij}(h)\right)\left(\widehat{\boldsymbol{\Omega}}_{ij}(h) - \boldsymbol{\Omega}_{ij}(h)\right)'\widehat{\boldsymbol{q}}_j\right\|_2$$

$$\le \sum_{h=1}^{h_0}\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}\left\|\left(\widehat{\boldsymbol{\Omega}}_{ij}(h) - \boldsymbol{\Omega}_{ij}(h)\right)\right\|_2^2 = O_p(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2},\ m_1^2m_2^2)\cdot T^{-1})$$

$$\|K_2\|_2 = \left\|2\widehat{\boldsymbol{q}}_j'\cdot\left(\widetilde{\boldsymbol{M}} - \boldsymbol{M}\right)\cdot(\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j)\right\|_2 \le 2\left\|\widetilde{\boldsymbol{M}} - \boldsymbol{M}\right\|_2\cdot\|\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j\|_2 = O_p\left(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2}, m_1^2m_2^2)\cdot T^{-1}\right)$$

$$\|K_3\|_2 = \left\|(\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j)'\boldsymbol{M}(\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j)\right\|_2 \le \left\|(\widehat{\boldsymbol{q}}_j - \boldsymbol{q}_j)\right\|_2^2\|\boldsymbol{M}\|_2 = O_p\left(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2}, m_1^2m_2^2)\cdot T^{-1}\right).$$

Thus, we have

$$|\hat{\lambda}_j| = O_p\left(\max(p_1^{2-2\delta_1}p_2^{2-2\delta_2}, m_1^2m_2^2)\cdot T^{-1}\right), \qquad for \qquad j = 1,\ldots,k_1.$$

□

Proof of Theorem 3

Proof. $\boldsymbol{S}_t$ is the dynamic signal part of $\boldsymbol{X}_t$, i.e. $\boldsymbol{S}_t = \boldsymbol{H}_R\boldsymbol{Q}_1\boldsymbol{Z}_t\boldsymbol{Q}_2'\boldsymbol{H}_C'$. And its estimator is $\widehat{\boldsymbol{S}}_t = \boldsymbol{H}_R\widehat{\boldsymbol{Q}}_1\widehat{\boldsymbol{Q}}_1'\boldsymbol{X}_t\widehat{\boldsymbol{Q}}_2\widehat{\boldsymbol{Q}}_2'\boldsymbol{H}_C'$. We have

$$\begin{aligned}
\widehat{\boldsymbol{S}}_t - \boldsymbol{S}_t &= \boldsymbol{H}_R\left(\widehat{\boldsymbol{Q}}_1\widehat{\boldsymbol{Q}}_1'\boldsymbol{X}_t\widehat{\boldsymbol{Q}}_2\widehat{\boldsymbol{Q}}_2' - \boldsymbol{Q}_1\boldsymbol{Z}_t\boldsymbol{Q}_2'\right)\boldsymbol{H}_C' = \boldsymbol{H}_R\left(\widehat{\boldsymbol{Q}}_1\widehat{\boldsymbol{Q}}_1'\left(\boldsymbol{Q}_1\boldsymbol{Z}_t\boldsymbol{Q}_2' + \boldsymbol{E}_t\right)\widehat{\boldsymbol{Q}}_2\widehat{\boldsymbol{Q}}_2' - \boldsymbol{Q}_1\boldsymbol{Z}_t\boldsymbol{Q}_2'\right)\boldsymbol{H}_C' \\
&= \boldsymbol{H}_R\left(\widehat{\boldsymbol{Q}}_1\widehat{\boldsymbol{Q}}_1'\boldsymbol{Q}_1\boldsymbol{Z}_t\boldsymbol{Q}_2'(\widehat{\boldsymbol{Q}}_2\widehat{\boldsymbol{Q}}_2' - \boldsymbol{Q}_2\boldsymbol{Q}_2') + (\widehat{\boldsymbol{Q}}_1\widehat{\boldsymbol{Q}}_1' - \boldsymbol{Q}_1\boldsymbol{Q}_1')\boldsymbol{Q}_1\boldsymbol{Z}_t\boldsymbol{Q}_2' + \widehat{\boldsymbol{Q}}_1\widehat{\boldsymbol{Q}}_1'\boldsymbol{E}_t\widehat{\boldsymbol{Q}}_2\widehat{\boldsymbol{Q}}_2'\right)\boldsymbol{H}_C' \\
&= I_1 + I_2 + I_3.
\end{aligned}$$

Since $\boldsymbol{H}_R$ and $\boldsymbol{H}_C$ are orthogonal matrices, we have

$$\begin{aligned}
\|I_1\|_2^2 &= \left\|\boldsymbol{H}_R\widehat{\boldsymbol{Q}}_1\widehat{\boldsymbol{Q}}_1'\boldsymbol{Q}_1\boldsymbol{Z}_t\boldsymbol{Q}_2'(\widehat{\boldsymbol{Q}}_2\widehat{\boldsymbol{Q}}_2' - \boldsymbol{Q}_2\boldsymbol{Q}_2')\boldsymbol{H}_C'\right\|_2^2 \\
&\le \|\boldsymbol{Z}_t\|_2^2\left\|(\widehat{\boldsymbol{Q}}_2 - \boldsymbol{Q}_2)\widehat{\boldsymbol{Q}}_2' + \boldsymbol{Q}_2(\widehat{\boldsymbol{Q}}_2 - \boldsymbol{Q}_2)'\right\|_2^2 \\
&\le 2\|\boldsymbol{Z}_t\|_2^2\|\widehat{\boldsymbol{Q}}_2 - \boldsymbol{Q}_2\|_2^2
\end{aligned}$$

Thus by Theorem 1, we have

$$\begin{aligned}
\|I_1\| &= O_p\left(p_1^{1/2-\delta_1/2}p_2^{1/2-\delta_2/2}\right)\cdot O_p\left(\max\left(T^{-1/2},\ \frac{m_1}{p_1^{1-\delta_1}}\frac{m_2}{p_2^{1-\delta_2}}T^{-1/2}\right)\right) \\
&= O_p\left(\max\left(p_1^{1/2-\delta_1/2}p_2^{1/2-\delta_2/2},\ m_1p_1^{-1/2+\delta_1/2}m_2p_2^{-1/2+\delta_2/2}\right)\cdot T^{-1/2}\right).
\end{aligned}$$

Similarity, we have

$$\begin{aligned}\|I_2\|_2 &= \left\|(\widehat{\boldsymbol{Q}}_1\widehat{\boldsymbol{Q}}_1' - \boldsymbol{Q}_1\boldsymbol{Q}_1')\boldsymbol{Q}_1\boldsymbol{Z}_t\boldsymbol{Q}_2'\right\|_2 \le 2\|\boldsymbol{Z}_t\|_2\|\widehat{\boldsymbol{Q}}_1 - \boldsymbol{Q}_1\|_2 \\ &= O_p\left(\max\left(p_1^{1/2-\delta_1/2}p_2^{1/2-\delta_2/2},\ m_1 p_1^{-1/2+\delta_1/2} m_2 p_2^{-1/2+\delta_2/2}\right)\cdot T^{-1/2}\right),\end{aligned}$$

and

$$\|I_3\|_2 = \left\|\widehat{\boldsymbol{Q}}_1\widehat{\boldsymbol{Q}}_1'\boldsymbol{E}_t\widehat{\boldsymbol{Q}}_2\widehat{\boldsymbol{Q}}_2'\right\|_2 \le \left\|\widehat{\boldsymbol{Q}}_1'\boldsymbol{E}_t\widehat{\boldsymbol{Q}}_2\right\|_2 \le \|(\widehat{\boldsymbol{Q}}_2'\otimes\widehat{\boldsymbol{Q}}_1')vec(\boldsymbol{E}_t)\|_2 \le k_1k_2\|\boldsymbol{\Sigma}_e\|_2 = O_p(1).$$

Thus,

$$\|\widehat{\boldsymbol{S}}_t - \boldsymbol{S}_t\|_2 = O_p\left(\max\left(p_1^{1/2-\delta_1/2}p_2^{1/2-\delta_2/2},\ m_1 p_1^{-1/2+\delta_1/2} m_2 p_2^{-1/2+\delta_2/2}\right)\cdot T^{-1/2} + 1\right)$$

□

Proof of Theorem 4

Proof.

$$\mathcal{D}(\widehat{\boldsymbol{Q}}_i, \boldsymbol{Q}_i) = \left(1 - \frac{1}{k_i}Tr\left(\widehat{\boldsymbol{Q}}_i\widehat{\boldsymbol{Q}}_i'\boldsymbol{Q}_i\boldsymbol{Q}_i'\right)\right)^{-1/2}, \quad for \quad i = 1, 2.$$

From Liu & Chen (2016), $\mathcal{D}(\widehat{\boldsymbol{Q}}_i, \boldsymbol{Q}_i) = O_p\left(\|\widehat{\boldsymbol{Q}}_i, \boldsymbol{Q}_i\|_2\right) = O_p\left(\max\left(T^{-1/2},\ \frac{m_1}{p_1^{1-\delta_1}}\frac{m_2}{p_2^{1-\delta_2}}T^{-1/2}\right)\right)$ for $i = 1, 2$. Since $\mathcal{D}(\widehat{\boldsymbol{\Lambda}}, \boldsymbol{\Lambda}) = \mathcal{D}(\widehat{\boldsymbol{Q}}_1, \boldsymbol{Q}_1)$ and $\mathcal{D}(\widehat{\boldsymbol{\Gamma}}, \boldsymbol{\Gamma}) = \mathcal{D}(\widehat{\boldsymbol{Q}}_2, \boldsymbol{Q}_2)$, the result follows. □

Appendix C Multinational Macroeconomic Indices Dataset

Table 12 lists the short name of each series, its mnemonic (the series label used in the OECD database), the transformation applied to the series, and a brief data description. All series are from the OECD Database. In the transformation column, Δ denote the first difference, $\Delta\ln$ denote the first difference of the logarithm. GP denotes the measure of growth rate last period.

Short name	Mnemonic	Tran	description
CPI: Food	CPGDFD	$\Delta^2 \ln$	Consumer Price Index: Food, seasonally adjusted
CPI: Ener	CPGREN	$\Delta^2 \ln$	Consumer Price Index: Energy, seasonally adjusted
CPI: Tot	CPALTT01	$\Delta^2 \ln$	Consumer Price Index: Total, seasonally adjusted
IR: Long	IRLT	Δ lv	Interest Rates: Long-term gov bond yields
IR: 3-Mon	IR3TIB	Δ lv	Interest Rates: 3-month Interbank rates and yields
P: TIEC	PRINTO01	$\Delta \ln$	Production: Total industry excl construction
P: TM	PRMNTO01	$\Delta \ln$	Production: Total manufacturing
GDP	LQRSGPOR	$\Delta \ln$	GDP: Original (Index 2010 = 1.00, seasonally adjusted)
IT: Ex	XTEXVA01	$\Delta \ln$	International Trade: Total Exports Value (goods)
IT: Im	XTIMVA01	$\Delta \ln$	International Trade: Total Imports Value (goods)

Table 12: Data transformations, and variable definitions

Country	ISO ALPHA-3 Code	Country	ISO ALPHA-3 Code
United States of America	USA	United Kingdom	GBR
Canada	CAN	Finland	FIN
New Zealand	NZL	Sweden	SWE
Australia	AUS	France	FRA
Norway	NOR	Netherlands	NLD
Ireland	IRL	Austria	AUT
Denmark	DNK	Germany	DEU

Table 13: Countries and ISO Alpha-3 Codes in Macroeconomic Indices Application

Appendix D Tables of Simulation Results

				$T = 0.5\,p_1\,p_2$		$T = p_1\,p_2$		$T = 1.5\,p_1\,p_2$		$T = 2\,p_1\,p_2$	
δ_1	δ_2	p_1	p_2	$\mathcal{D}_u(\widehat{Q}, Q)$	$\mathcal{D}_c(\widehat{Q}, Q)$	$\mathcal{D}_u(\widehat{Q}, Q)$	$\mathcal{D}_c(\widehat{Q}, Q)$	$\mathcal{D}_u(\widehat{Q}, Q)$	$\mathcal{D}_c(\widehat{Q}, Q)$	$\mathcal{D}_u(\widehat{Q}, Q)$	$\mathcal{D}_c(\widehat{Q}, Q)$
0	0	20	20	1.02(0.2)	0.73(0.18)	0.73(0.12)	0.52(0.13)	0.58(0.08)	0.42(0.09)	0.5(0.07)	0.36(0.07)
		20	40	0.67(0.1)	0.47(0.11)	0.47(0.06)	0.33(0.07)	0.39(0.05)	0.27(0.06)	0.33(0.04)	0.23(0.05)
		40	20	0.71(0.1)	0.41(0.12)	0.5(0.06)	0.28(0.07)	0.41(0.05)	0.24(0.06)	0.35(0.04)	0.2(0.05)
		40	40	0.47(0.06)	0.26(0.07)	0.33(0.03)	0.18(0.04)	0.27(0.03)	0.15(0.04)	0.24(0.02)	0.13(0.03)
0.5	0	20	20	5.64(0.5)	1.92(0.74)	4.94(1.17)	1.27(0.34)	3.34(1.56)	0.98(0.22)	2.09(1.11)	0.83(0.18)
		20	40	4.86(1.19)	1.12(0.3)	1.95(1)	0.76(0.18)	1.12(0.28)	0.62(0.14)	0.89(0.17)	0.53(0.12)
		40	20	5.82(0.26)	1.23(0.44)	5.33(0.87)	0.8(0.22)	3.46(1.6)	0.66(0.18)	1.73(0.81)	0.55(0.14)
		40	40	5.37(0.81)	0.73(0.21)	1.56(0.67)	0.49(0.13)	0.96(0.2)	0.4(0.1)	0.77(0.12)	0.36(0.09)
0.5	0.5	20	20	6.81(0.34)	6.08(0.6)	6.46(0.17)	5.54(0.73)	6.32(0.13)	4.84(1.11)	6.24(0.1)	4.34(1.26)
		20	40	6.67(0.3)	5.86(0.66)	6.39(0.15)	4.93(1.08)	6.26(0.08)	4.12(1.28)	6.2(0.05)	3.47(1.3)
		40	20	6.71(0.28)	5.69(0.61)	6.4(0.13)	4.78(1.23)	6.27(0.07)	3.73(1.43)	6.2(0.05)	2.94(1.4)
		40	40	6.62(0.28)	5.15(0.98)	6.32(0.08)	3.74(1.44)	6.23(0.05)	2.7(1.43)	6.17(0.03)	2.05(1.12)

Table 14: Orthogonal constraints case. Means and standard deviations (in parentheses) of the estimation accuracy measured by $\mathcal{D}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$. $\boldsymbol{D}_u$ for the unconstrained model 1. $\boldsymbol{D}_c$ for the constrained model 2. All numbers in the table are 10 times the true numbers for clear presentation. The results are based on 500 iterations.

						$T = 0.5 * p_1 * p_2$			$T = p_1 * p_2$			$T = 1.5 * p_1 * p_2$			$T = 2 * p_1 * p_2$		
δ_1	δ_2	δ_3	δ_4	p_1	p_2	f_u	f_{con_1}	f_{con_2}	f_u	f_{con_1}	f_{con_2}	f_u	f_{con_1}	f_{con_2}	f_u	f_{con_1}	f_{con_2}
0	0	0	0	20	20	0	0.94	0	0	1.00	0	0	1.00	0	0.01	1.00	0
				20	40	0	1.00	0	0	1.00	0	0.03	1.00	0	0.19	1.00	0
				40	20	0.15	0.99	1.00	0.81	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00
				40	40	0.71	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0	0	0.5	0	20	20	0	0.94	0	0	1.00	0	0	1.00	0	0	1.00	0
				20	40	0	1.00	0	0	1.00	0	0	1.00	0	0	1.00	0
				40	20	0	0.99	0.54	0	1.00	0.84	0	1.00	0.97	0	1.00	1.00
				40	40	0	1.00	0.98	0	1.00	1.00	0	1.00	1.00	0	1.00	1.00
0	0	0.5	0.5	20	20	0	0.94	0	0	1.00	0	0	1.00	0	0	1.00	0
				20	40	0	1.00	0	0	1.00	0	0	1.00	0	0	1.00	0
				40	20	0	0.99	0	0	1.00	0	0	1.00	0	0	1.00	0
				40	40	0	1.00	0	0	1.00	0	0	1.00	0	0	1.00	0
0.5	0	0	0	20	20	0	0.21	0	0	0.53	0	0	0.79	0	0	0.92	0
				20	40	0	0.67	0	0	0.97	0	0	1.00	0	0	1.00	0
				40	20	0	0.34	1.00	0	0.79	1.00	0	0.92	1.00	0	0.95	1.00
				40	40	0	0.87	1.00	0	0.97	1.00	0	0.99	1.00	0	0.99	1.00
0.5	0	0.5	0	20	20	0	0.21	0	0	0.53	0	0	0.79	0	0	0.92	0
				20	40	0	0.67	0	0	0.97	0	0	1.00	0	0	1.00	0
				40	20	0	0.34	0.54	0	0.79	0.84	0	0.92	0.97	0	0.95	1.00
				40	40	0	0.87	0.98	0	0.97	1.00	0	0.99	1.00	0	0.99	1.00
0.5	0	0.5	0.5	20	20	0	0.21	0	0	0.53	0	0	0.79	0	0	0.92	0
				20	40	0	0.67	0	0	0.97	0	0	1.00	0	0	1.00	0
				40	20	0	0.34	0	0	0.79	0	0	0.92	0	0	0.95	0
				40	40	0	0.87	0	0	0.97	0	0	0.99	0	0	0.99	0
0.5	0.5	0	0	20	20	0	0.07	0	0	0.04	0	0	0.01	0	0	0.01	0
				20	40	0	0.07	0	0	0.02	0	0	0.01	0	0	0.01	0
				40	20	0	0.06	1.00	0	0.01	1.00	0	0	1.00	0	0	1.00
				40	40	0	0.06	1.00	0	0	1.00	0	0	1.00	0	0.03	1.00
0.5	0.5	0.5	0	20	20	0	0.07	0	0	0.04	0	0	0.01	0	0	0.01	0
				20	40	0	0.07	0	0	0.02	0	0	0.01	0	0	0.01	0
				40	20	0	0.06	0.54	0	0.01	0.84	0	0	0.97	0	0	1.00
				40	40	0	0.06	0.98	0	0	1.00	0	0	1.00	0	0.03	1.00
0.5	0.5	0.5	0.5	20	20	0	0.07	0	0	0.04	0	0	0.01	0	0	0.01	0
				20	40	0	0.07	0	0	0.02	0	0	0.01	0	0	0.01	0
				40	20	0	0.06	0	0	0.01	0	0	0	0	0	0	0
				40	40	0	0.06	0	0	0	0	0	0	0	0	0.03	0

Table 15: Relative frequency of correctly estimating k_1

						$T=0.5*p_1*p_2$		$T=p_1*p_2$		$T=1.5*p_1*p_2$		$T=2*p_1*p_2$	
δ_1	δ_2	δ_3	δ_4	p_1	p_2	$\mathcal{D}_u(\widehat{Q},Q)$	$\mathcal{D}_c(\widehat{Q},Q)$	$\mathcal{D}_u(\widehat{Q},Q)$	$\mathcal{D}_c(\widehat{Q},Q)$	$\mathcal{D}_u(\widehat{Q},Q)$	$\mathcal{D}_c(\widehat{Q},Q)$	$\mathcal{D}_u(\widehat{Q},Q)$	$\mathcal{D}_c(\widehat{Q},Q)$
0	0	0	0	20	20	1.56(0.87)	0.57(0.1)	0.71(0.16)	0.41(0.06)	0.54(0.09)	0.33(0.04)	0.45(0.07)	0.28(0.04)
				20	40	0.71(0.33)	0.38(0.05)	0.4(0.06)	0.27(0.03)	0.32(0.04)	0.22(0.03)	0.27(0.03)	0.19(0.02)
				40	20	0.52(0.07)	0.33(0.05)	0.36(0.04)	0.24(0.03)	0.29(0.03)	0.19(0.03)	0.25(0.02)	0.17(0.02)
				40	40	0.32(0.04)	0.2(0.04)	0.22(0.02)	0.14(0.02)	0.18(0.02)	0.12(0.02)	0.15(0.01)	0.1(0.02)
0	0	0.5	0	20	20	3.68(0.04)	0.88(0.13)	3.61(0.02)	0.63(0.08)	3.59(0.02)	0.51(0.07)	3.57(0.02)	0.44(0.06)
				20	40	3.61(0.02)	0.61(0.06)	3.57(0.01)	0.43(0.04)	3.56(0.01)	0.35(0.03)	3.55(0.02)	0.3(0.03)
				40	20	3.65(0.04)	0.57(0.05)	3.58(0.05)	0.42(0.03)	3.43(0.36)	0.35(0.02)	2.78(0.94)	0.3(0.02)
				40	40	3.36(0.51)	0.33(0.03)	0.59(0.36)	0.24(0.02)	0.35(0.06)	0.2(0.02)	0.28(0.03)	0.17(0.01)
0	0	0.5	0.5	20	20	5.99(0.36)	1.88(0.51)	5.73(0.38)	1.32(0.29)	5.49(0.45)	1.06(0.19)	5.24(0.49)	0.92(0.17)
				20	40	6.67(0.32)	1.42(0.3)	6.42(0.35)	1.02(0.15)	6.24(0.34)	0.83(0.11)	6.06(0.33)	0.72(0.09)
				40	20	6.37(0.29)	1.06(0.09)	6.09(0.28)	0.8(0.06)	5.89(0.31)	0.67(0.04)	5.77(0.29)	0.59(0.04)
				40	40	6.37(0.3)	0.67(0.04)	5.95(0.29)	0.5(0.03)	5.62(0.34)	0.42(0.02)	5.26(0.46)	0.37(0.02)
0.5	0	0	0	20	20	3.72(0.19)	1.22(0.38)	3.61(0.21)	0.8(0.17)	3.55(0.21)	0.63(0.13)	3.47(0.32)	0.55(0.11)
				20	40	3.61(0.17)	0.73(0.17)	3.45(0.33)	0.49(0.1)	3.2(0.59)	0.4(0.08)	2.66(0.9)	0.35(0.06)
				40	20	3.73(0.09)	0.78(0.27)	3.64(0.06)	0.52(0.13)	3.59(0.07)	0.41(0.11)	3.56(0.09)	0.36(0.08)
				40	40	3.65(0.05)	0.46(0.13)	3.57(0.07)	0.31(0.07)	3.49(0.21)	0.26(0.06)	3.29(0.48)	0.22(0.05)
0.5	0	0.5	0	20	20	3.81(0.07)	1.4(0.34)	3.69(0.04)	0.94(0.16)	3.63(0.03)	0.75(0.12)	3.6(0.04)	0.64(0.11)
				20	40	3.67(0.03)	0.87(0.15)	3.6(0.01)	0.6(0.08)	3.57(0.02)	0.49(0.07)	3.54(0.08)	0.42(0.06)
				40	20	3.66(0.09)	0.91(0.24)	3.56(0.13)	0.63(0.11)	3.19(0.58)	0.5(0.09)	2.14(0.92)	0.44(0.07)
				40	40	3.53(0.18)	0.54(0.11)	2.3(1.01)	0.37(0.06)	0.82(0.34)	0.31(0.06)	0.57(0.11)	0.26(0.05)
0.5	0	0.5	0.5	20	20	4.91(0.48)	2.19(0.51)	4.5(0.48)	1.5(0.28)	4.22(0.4)	1.2(0.18)	3.99(0.27)	1.04(0.17)
				20	40	5.69(0.25)	1.56(0.3)	5.45(0.24)	1.11(0.14)	5.23(0.35)	0.9(0.11)	4.85(0.54)	0.78(0.09)
				40	20	5.32(0.29)	1.29(0.2)	5.21(0.28)	0.93(0.09)	4.99(0.44)	0.77(0.07)	4.67(0.56)	0.68(0.06)
				40	40	5.3(0.15)	0.79(0.09)	4.8(0.55)	0.58(0.05)	3.81(0.33)	0.49(0.04)	3.63(0.03)	0.43(0.03)
0.5	0.5	0	0	20	20	5.13(0.47)	3.76(0.4)	5.05(0.46)	3.36(0.5)	4.88(0.44)	2.97(0.68)	4.73(0.38)	2.59(0.76)
				20	40	5.44(0.46)	3.63(0.39)	5.2(0.48)	3.05(0.65)	5.01(0.45)	2.57(0.78)	4.86(0.44)	2.1(0.8)
				40	20	5.17(0.4)	3.49(0.39)	4.91(0.33)	2.93(0.77)	4.75(0.33)	2.26(0.93)	4.64(0.3)	1.82(0.89)
				40	40	5.46(0.41)	3.19(0.6)	5.17(0.36)	2.31(0.92)	4.91(0.31)	1.66(0.89)	4.75(0.29)	1.28(0.77)
0.5	0.5	0.5	0	20	20	4.59(0.31)	3.82(0.4)	4.33(0.27)	3.39(0.5)	4.15(0.21)	3(0.67)	4.05(0.16)	2.62(0.75)
				20	40	4.54(0.34)	3.66(0.39)	4.24(0.25)	3.06(0.64)	4.07(0.18)	2.59(0.78)	3.99(0.15)	2.11(0.79)
				40	20	4.3(0.23)	3.52(0.39)	4.05(0.11)	2.95(0.76)	3.94(0.06)	2.29(0.92)	3.88(0.05)	1.84(0.88)
				40	40	4.3(0.21)	3.2(0.59)	4.03(0.1)	2.32(0.92)	3.92(0.05)	1.67(0.88)	3.87(0.04)	1.29(0.77)
0.5	0.5	0.5	0.5	20	20	5.05(0.28)	4.17(0.43)	4.57(0.22)	3.59(0.48)	4.33(0.17)	3.15(0.63)	4.19(0.13)	2.75(0.72)
				20	40	4.87(0.29)	3.88(0.39)	4.42(0.18)	3.2(0.61)	4.22(0.13)	2.71(0.74)	4.1(0.1)	2.22(0.75)
				40	20	4.61(0.19)	3.63(0.37)	4.23(0.11)	3.03(0.73)	4.07(0.06)	2.37(0.88)	3.98(0.06)	1.93(0.85)
				40	40	4.25(0.13)	3.25(0.58)	4.01(0.05)	2.37(0.9)	3.91(0.03)	1.72(0.86)	3.86(0.02)	1.34(0.75)

Table 16: Means and standard deviations (in parentheses) of the estimation accuracy measured by $\mathcal{D}(\widehat{\boldsymbol{Q}},\boldsymbol{Q})$. For ease of presentation, all numbers in this table are the true numbers multiplied by 10.

Appendix E Corporate Financial Data Information

Short Name	Variable Name	Calculation
Profit.M	Profit Margin	Net Income / Revenue
Oper.M	Operating Margin	Operating Income / Revenue
EPS	Diluted Earing per share	from report
Gross.Margin	Gross Margin	Gross Proitt / Revenue
ROE	Return on equity	Net Income / Shareholders Equity
ROA	Return on assets	Net Income / Total Assets
Revenue.PS	Revenue Per Share	Revenue / Shares Outstanding
LiabilityE.R	Liability/Equity Ratio	Total Liabilities / Shareholders Equity
AssetE.R	Asset/Equity Ratio	Total Assets / Shareholders Equity
Earnings.R	Basic Earnings Power Ratio	EBIT / Total Assets
Payout.R	Payout Ratio	Dividend Per Share / EPS Basic
Cash.PS	Cash Per Share	Cash and other / Shares Outstanding
Revenue.G.Q	Revenue Growth over last Quarter	Revenue / Revenue Last Quarter - 1
Revenue.G.Y	Revenue Growth over same Quarter Last Year	Revenue / Revenue Last Year - 1
Profit.G.Q	Profit Growth over last Quarter	Profit / Profit Last Quarter - 1
Profit.G.Y	Profit Growth over same Quarter last Year	Profit / Profit Last Quarter - 1

Table 17: Variables in coporate financial data

TICKER	INDUSTRY_GROUP	INDUSTRY_SECTOR	TICKER	INDUSTRY_GROUP	INDUSTRY_SECTOR
AAPL	Computers	Technology	KO	Beverages	Consumer Non-cyclical
ABT	Healthcare-Products	Consumer Non-cyclical	KSU	Transportation	Industrial
ADM	Agriculture	Consumer Non-cyclical	LEG	Home Furnishings	Consumer Cyclical
ADP	Commercial Services	Consumer Non-cyclical	LH	Healthcare-Services	Consumer Non-cyclical
AEP	Electric	Utilities	LLTC	Semiconductors	Technology
AES	Electric	Utilities	LLY	Pharmaceuticals	Consumer Non-cyclical
AET	Healthcare-Services	Consumer Non-cyclical	LM	Diversified Finan Serv	Financial
AME	Electrical Compo&Equip	Industrial	LRCX	Semiconductors	Technology
AMGN	Biotechnology	Consumer Non-cyclical	MAS	Building Materials	Industrial
APA	Oil&Gas	Energy	MAT	Toys/Games/Hobbies	Consumer Cyclical
APC	Oil&Gas	Energy	MHFI	Commercial Services	Consumer Non-cyclical
APD	Chemicals	Basic Materials	MMC	Insurance	Financial
APH	Electronics	Industrial	MMM	Miscellaneous Manufactur	Industrial
ARG	Chemicals	Basic Materials	MO	Agriculture	Consumer Non-cyclical
AVY	Household Products/Wares	Consumer Non-cyclical	MOS	Chemicals	Basic Materials
BA	Aerospace/Defense	Industrial	MRK	Pharmaceuticals	Consumer Non-cyclical
BAX	Healthcare-Products	Consumer Non-cyclical	MRO	Oil&Gas	Energy
BCR	Healthcare-Products	Consumer Non-cyclical	MSFT	Software	Technology
BDX	Healthcare-Products	Consumer Non-cyclical	MSI	Telecommunications	Communications
BEN	Diversified Finan Serv	Financial	MUR	Oil&Gas	Energy
BHI	Oil&Gas Services	Energy	MYL	Pharmaceuticals	Consumer Non-cyclical
BLL	Packaging&Containers	Industrial	NBL	Oil&Gas	Energy
BMY	Pharmaceuticals	Consumer Non-cyclical	NEE	Electric	Utilities

CA	Software	Technology	NEM	Mining	Basic Materials
CAH	Pharmaceuticals	Consumer Non-cyclical	NI	Gas	Utilities
CAT	Machinery-Constr&Mining	Industrial	NOC	Aerospace/Defense	Industrial
CCE	Beverages	Consumer Non-cyclical	NSC	Transportation	Industrial
CHD	Household Products/Wares	Consumer Non-cyclical	NUE	Iron/Steel	Basic Materials
CL	Cosmetics/Personal Care	Consumer Non-cyclical	NWL	Housewares	Consumer Cyclical
CLX	Household Products/Wares	Consumer Non-cyclical	OKE	Pipelines	Energy
CMCSA	Media	Communications	OMC	Advertising	Communications
CMI	Machinery-Diversified	Industrial	OXY	Oil&Gas	Energy
CMS	Electric	Utilities	PBI	Office/Business Equip	Technology
CNP	Gas	Utilities	PCAR	Auto Manufacturers	Consumer Cyclical
COG	Oil&Gas	Energy	PCG	Electric	Utilities
COP	Oil&Gas	Energy	PEG	Electric	Utilities
CSX	Transportation	Industrial	PEP	Beverages	Consumer Non-cyclical
CTL	Telecommunications	Communications	PFE	Pharmaceuticals	Consumer Non-cyclical
CVC	Media	Communications	PG	Cosmetics/Personal Care	Consumer Non-cyclical
CVS	Retail	Consumer Cyclical	PH	Miscellaneous Manufactur	Industrial
D	Electric	Utilities	PHM	Home Builders	Consumer Cyclical
DD	Chemicals	Basic Materials	PKI	Electronics	Industrial
DHR	Healthcare-Products	Consumer Non-cyclical	PNR	Miscellaneous Manufactur	Industrial
DIS	Media	Communications	PNW	Electric	Utilities
DOV	Miscellaneous Manufactur	Industrial	PPG	Chemicals	Basic Materials
DOW	Chemicals	Basic Materials	PPL	Electric	Utilities
DTE	Electric	Utilities	PSA	REITS	Financial
DUK	Electric	Utilities	PX	Chemicals	Basic Materials
DVN	Oil&Gas	Energy	QCOM	Semiconductors	Technology
ECL	Commercial Services	Consumer Non-cyclical	R	Transportation	Industrial
ED	Electric	Utilities	RCL	Leisure Time	Consumer Cyclical
EFX	Commercial Services	Consumer Non-cyclical	RHI	Commercial Services	Consumer Non-cyclical
EIX	Electric	Utilities	ROK	Machinery-Diversified	Industrial
EMC	Computers	Technology	ROP	Machinery-Diversified	Industrial
EMR	Electrical Compo&Equip	Industrial	RRC	Oil&Gas	Energy
EQT	Oil&Gas	Energy	SBUX	Retail	Consumer Cyclical
ES	Electric	Utilities	SCG	Electric	Utilities
ESV	Oil&Gas	Energy	SHW	Chemicals	Basic Materials
ETN	Miscellaneous Manufactur	Industrial	SLB	Oil&Gas Services	Energy
ETR	Electric	Utilities	SNA	Hand/Machine Tools	Industrial
EXC	Electric	Utilities	SO	Electric	Utilities
EXPD	Transportation	Industrial	SRE	Gas	Utilities
F	Auto Manufacturers	Consumer Cyclical	STJ	Healthcare-Products	Consumer Non-cyclical
FAST	Distribution/Wholesale	Consumer Cyclical	SWK	Hand/Machine Tools	Industrial
FCX	Mining	Basic Materials	SWKS	Semiconductors	Technology
FLS	Machinery-Diversified	Industrial	SWN	Oil&Gas	Energy
FMC	Chemicals	Basic Materials	SYK	Healthcare-Products	Consumer Non-cyclical
FOX	Media	Communications	SYMC	Internet	Communications
FRT	REITS	Financial	SYY	Food	Consumer Non-cyclical
GAS	Gas	Utilities	T	Telecommunications	Communications
GD	Aerospace/Defense	Industrial	TAP	Beverages	Consumer Non-cyclical
GE	Miscellaneous Manufactur	Industrial	TE	Electric	Utilities
GILD	Biotechnology	Consumer Non-cyclical	TGNA	Media	Communications
GLW	Electronics	Industrial	THC	Healthcare-Services	Consumer Non-cyclical

GPC	Retail	Consumer Cyclical	TMO	Healthcare-Products	Consumer Non-cyclical
GT	Auto Parts&Equipment	Consumer Cyclical	TROW	Diversified Finan Serv	Financial
GWW	Distribution/Wholesale	Consumer Cyclical	TSO	Oil&Gas	Energy
HAR	Home Furnishings	Consumer Cyclical	TSS	Commercial Services	Consumer Non-cyclical
HAS	Toys/Games/Hobbies	Consumer Cyclical	TWX	Media	Communications
HCN	REITS	Financial	TXN	Semiconductors	Technology
HCP	REITS	Financial	TXT	Miscellaneous Manufactur	Industrial
HES	Oil&Gas	Energy	TYC	Building Materials	Industrial
HOG	Leisure Time	Consumer Cyclical	UHS	Healthcare-Services	Consumer Non-cyclical
HON	Electronics	Industrial	UNH	Healthcare-Services	Consumer Non-cyclical
HP	Oil&Gas	Energy	UNP	Transportation	Industrial
HRS	Aerospace/Defense	Industrial	UTX	Aerospace/Defense	Industrial
HSY	Food	Consumer Non-cyclical	VFC	Apparel	Consumer Cyclical
IBM	Computers	Technology	VLO	Oil&Gas	Energy
IFF	Chemicals	Basic Materials	VMC	Building Materials	Industrial
INTC	Semiconductors	Technology	VZ	Telecommunications	Communications
IP	Forest Products&Paper	Basic Materials	WDC	Computers	Technology
IPG	Advertising	Communications	WEC	Electric	Utilities
IR	Miscellaneous Manufactur	Industrial	WHR	Home Furnishings	Consumer Cyclical
JBHT	Transportation	Industrial	WM	Environmental Control	Industrial
JCI	Building Materials	Industrial	WMB	Pipelines	Energy
JNJ	Pharmaceuticals	Consumer Non-cyclical	WY	REITS	Financial
K	Food	Consumer Non-cyclical	XEL	Electric	Utilities
KLAC	Semiconductors	Technology	XLNX	Semiconductors	Technology
KMB	Household Products/Wares	Consumer Non-cyclical	XOM	Oil&Gas	Energy
			XRAY	Healthcare-Products	Consumer Non-cyclical
			XRX	Office/Business Equip	Technology

Table 18: Bloomberg's industry and section information of 200 companies.