Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

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

Elynn Y. Chen, Ruey S. Tsay & Rong Chen

To cite this article: Elynn Y. Chen, Ruey S. Tsay & Rong Chen (2020) Constrained Factor Models for High-Dimensional Matrix-Variate Time Series, Journal of the American Statistical Association, 115:530, 775-793, DOI: 10.1080/01621459.2019.1584899

To link to this article: https://doi.org/10.1080/01621459.2019.1584899

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2020, VOL. 115, NO. 530, 775–793: Theory and Methods
https://doi.org/10.1080/01621459.2019.1584899

Taylor & Francis
Taylor & Francis Group

Check for updates

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

Elynn Y. Chen[a], Ruey S. Tsay[b], and Rong Chen[c]

[a]Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ; [b]Booth School of Business, University of Chicago, Chicago, IL; [c]Department of Statistics, Rutgers University, Piscataway, NJ

ABSTRACT
High-dimensional matrix-variate time series data are becoming widely available in many scientific fields, such as economics, biology, and meteorology. To achieve significant dimension reduction while preserving the intrinsic matrix structure and temporal dynamics in such data, Wang, Liu, and Chen proposed a matrix factor model, that is, shown to be able to provide effective analysis. In this article, we establish a general framework for incorporating domain and prior knowledge in the matrix factor model through linear constraints. The proposed framework is shown to be useful in achieving parsimonious parameterization, facilitating interpretation of the latent matrix factor, and identifying specific factors of interest. Fully utilizing the prior-knowledge-induced constraints results in more efficient and accurate modeling, inference, dimension reduction as well as a clear and better interpretation of the results. 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 finite-sample performance of the proposed method and its associated asymptotic properties. We illustrate 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. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received March 2018
Accepted January 2019

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

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 30 variables spanning over 100 years and over 200 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 indexes. Typical factor analysis of such data either converts the matrix into a vector or modeling the row or column vectors separately (Chamberlain 1983; Chamberlain and Rothschild 1983; Bai 2003; Bai and Ng 2002, 2007; Forni et al. 2000, 2004; Pan and Yao 2008; Lam, Yao, and Bathia 2011; Lam and Yao 2012; Chang, Guo, and Yao 2015). However, the components of matrix-variates are often dependent among rows and columns with certain 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, Liu, and Chen (2019) proposed a matrix factor model that maintains and uses the matrix structure of the data to achieve significant dimension reduction.

In factor analysis of matrix time series and in many other types of high-dimensional data, the problem of factor interpretations is of paramount importance. Furthermore, it is important in many practical applications to obtain specific latent factors related to certain domain theories, and with the aid of these specific factors to predict future values of interest more accurately. For example, financial researchers may be interested in extracting the latent factors of level, slope, and curvatures of the interest-rate yield curve and in predicting future equity prices based on those factors (Diebold, Piazzesi, and Rudebusch 2005; Diebold, Rudebusch, and Aruoba 2006; Rudebusch and Wu 2008; Bansal, Connolly, and Stivers 2014).

In many applications, relevant prior, or domain knowledge is available or data themselves exhibit certain specific structure. Additional covariates may also be 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 build proper constraints based on those measurable variables to effectively obtain the factors of interest.

To address these important issues and practical needs, we extend the matrix factor model of Wang, Liu, and Chen (2019)

CONTACT Rong Chen rongchen@stat.rutgers.edu Department of Statistics, Rutgers University, Piscataway, NJ 08854.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/JASA.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

by imposing natural constraints among the column and row variables to incorporate prior knowledge or to induce specific factors. 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 often encountered by 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 and Friedman (1993), Tibshirani (1996), Liu et al. (2007), Fan and Li (2001), Zou (2006), and Zhang and Lu (2007). It is shown 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 and 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 and 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. The factor models in Tsai and Tsay (2010) and Tsai et al. (2016), following the definition in Bai (2003), Bai and Ng (2002, 2007), Forni et al. (2000, 2004), attempt to separate the common factors that affect the dynamics of most original component series from the idiosyncratic series that at most affect the dynamics of a few original time series. Such a definition is appealing in analyzing economic and financial phenomena. But the fact that idiosyncratic part may exhibit serial correlations poses technical difficulties in both identification and inference. These factor models are only asymptotically identifiable because a rigorous definition of the common factors can only be established when the dimension of time series goes to infinity. In our setting, 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. See Lam, Yao, and Bathia (2011), Lam and Yao (2012), Chang, Guo, and Yao (2015), and Wang, Liu, and Chen (2019) for more detailed comparisons between these two different model definitions.

The rest of the article 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 estimators. Section 5 presents some simulation results, whereas Section 6 contains three applications. Section 7 concludes. All proofs are in the Appendix (supplementary materials).

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 jth column vector and the kth row vector of the matrix $\boldsymbol{A}$ are denoted by $A_{\cdot j}$ and $A_{k\cdot}$, respectively.

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

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

Wang, Liu, and Chen (2019) proposed the following factor model for $\boldsymbol{Y}_t$,

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

where $\boldsymbol{F}_t$ is a $k_1 \times k_2$ latent 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 Equation (1), $(\boldsymbol{\Lambda}, \boldsymbol{\Gamma})$ and $(c\boldsymbol{\Lambda}, \boldsymbol{\Gamma}/c)$ are equivalent if $c \neq 0$.

In Model (1), we assume that $\mathrm{vec}(\boldsymbol{U}_t) \sim \mathrm{WN}(\boldsymbol{0}, \boldsymbol{\Sigma}_e)$ and is independent of the factor process $\mathrm{vec}(\boldsymbol{F}_t)$. That is, $\{\boldsymbol{U}_t\}_{t=1}^T$ 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, Liu, and Chen (2019) 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, Yao, and Bathia (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. The method in Wang, Liu, and Chen (2019) can only identify "common" factors in the sense that those identifiable factors must be of certain strength. Weak factors will be left "erroneously" in the noise in application. 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 the 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

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

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 $k_1 \le m_1 << p_1$ and $k_2 \le m_2 << p_2$. Compared with the matrix factor model in Equation (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_1k_1, smaller than p_1k_1 of the unconstrained model. The number of parameters in the column loading matrix $\boldsymbol{C}$ also decreases from p_2k_2 to m_2k_2. The constraint matrices $\boldsymbol{H}_R$ and $\boldsymbol{H}_C$ are constructed based on prior or domain knowledge of the variables.

2.1. Examples of Constraint Matrices

We first consider discrete covariate-induced constraint matrices, using dummy variables. Continuous covariate may be segmented into regimes. As an illustration, we consider the following toy example of corporate financial matrix-valued time series. Suppose we have eight 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 in Table 1.

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

Further, we consider the constraint incorporating an interaction term between industry and market cap grouping information. Now the ith 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 factor model. Nelson and 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 $y(\tau)$ denotes the set of zero-coupon yields and τ denotes time to maturity.

Diebold and Li (2006) and Diebold, Rudebusch, and Aruoba (2006) interpreted 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 0 (to -1) as maturities increase. And the factor C_t has a loading, that is, -1 at the shortest maturity, increases to an intermediate maturity (equal to 2) and then falls back to -1 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 $\{\boldsymbol{Y}_t\}_{t=1}^T$, 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 $\boldsymbol{h}_1 = (1, 1, 1, 1, 1)'$, $\boldsymbol{h}_2 = (1, 1, 0, -1, -1)'$, 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.

2.2. Multi-Term and Partially Constrained Matrix Factor Models

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 *multi-term matrix factor model* as

$$\boldsymbol{Y}_t = \boldsymbol{H}_{R_1}\boldsymbol{R}_1\boldsymbol{F}_{1t}\boldsymbol{C}_1'\boldsymbol{H}_{C_1}' + \boldsymbol{H}_{R_2}\boldsymbol{R}_2\boldsymbol{F}_{2t}\boldsymbol{C}_2'\boldsymbol{H}_{C_2}' + \boldsymbol{U}_t. \quad (3)$$

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

Table 1. Groups of companies by industry and market capitalization.

		Market Cap	
		C1. Large	C2. Medium
Industry	I1. Tech	Apple, Microsoft	Brocade, FireEye
	I2. Retail	Walmart, Target	JC Penny, Kohl's

	Industry	Market Cap
Apple	I1	C1
Microsoft	I1	C1
Brocade	I1	C2
FireEye	I1	C2
Walmart	I2	C1
Target	I2	C1
JC Penny	I2	C2
Kohl's	I2	C2

Table 2. Illustration of constraint matrices constructed from grouping information by additive model.

$$\boldsymbol{H}_R^{(1)} = \left[\begin{array}{cc} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ \hline 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{array}\right] \qquad \boldsymbol{H}_R^{(2)} = \left[\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ \hline 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{array}\right] \qquad \boldsymbol{H}_R^{(3)} = \left[\begin{array}{cc|cc|c} 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & 1 & -1 \\ \hline 0 & 1 & 1 & 0 & -1 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{array}\right]$$

as natural resource based and manufacture based economies, and the corresponding factors may not interact with each other.

Note that Equation (3) can be rewritten as Equation (2), with $\boldsymbol{H}_R = [\boldsymbol{H}_{R_1} \quad \boldsymbol{H}_{R_2}], \boldsymbol{H}_C = [\boldsymbol{H}_{C_1} \quad \boldsymbol{H}_{C_2}]$,

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

Hence, Equation (3) is a special case of Equation (2) with the strong assumption that the factor matrix is block diagonal. Such a simplification can greatly enhance 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 $\boldsymbol{H}_{C_2}$. An estimation procedure is presented in Remark 3 in Section 3.3. 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

$$\boldsymbol{Y}_t = [\boldsymbol{H}_{R_1}\boldsymbol{R}_1 \quad \boldsymbol{\Lambda}_2] \begin{bmatrix} \boldsymbol{F}_{11,t} & \boldsymbol{F}_{12,t} \\ \boldsymbol{F}_{21,t} & \boldsymbol{F}_{22,t} \end{bmatrix} \begin{bmatrix} \boldsymbol{C}_1'\boldsymbol{H}_{C_1}' \\ \boldsymbol{\Gamma}_2' \end{bmatrix} + \boldsymbol{U}_t, \quad (4)$$

where $\boldsymbol{H}_{R_1}$, $\boldsymbol{R}_1$, $\boldsymbol{H}_{C_1}$, and $\boldsymbol{C}_1$ are defined similarly as those in Equation (3). $\boldsymbol{F}_{ij,t}$'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 $q_1 < p_1, q_2 < p_2$ and $\text{vec}(\boldsymbol{F}_{ij,t})$'s are independent with $\text{vec}(\boldsymbol{U}_t)$. We assume that $\boldsymbol{H}_{R_1}'\boldsymbol{\Lambda}_2 = \boldsymbol{0}$ 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

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

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 (4) can be further simplified as

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

Model (6) is different from the multi-term model of Equation (3) in that the matrix $\boldsymbol{H}_{R_2}$ in Equation (5) is induced from $\boldsymbol{H}_{R_1}$ while the $\boldsymbol{H}_{R_2}$ in Equation (3) 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 Equation (5) is the same as that in the unconstrained model (1). However, as shown by Theorem 1 in Section 4, the rates of convergence are faster than those of the unconstrained matrix factor model. Furthermore, in many applications, inclusion of only two matrix-factor terms is adequate in explaining a high percentage of variability, as exemplified by the three applications in Section 6.

Remark 2. Subpanel structure in multivariate time series is encountered frequently in real applications. For example, macroeconometric data often consist of large panels of time series which can be further divided into smaller but still quite large subpanels or blocks. Built upon Forni et al. (2004), Hallin and Liška (2007), and Hallin and Liška (2011) considered n-dimensional random variable $\boldsymbol{x} = [\boldsymbol{y}' \;\; \boldsymbol{z}']'$ with subpanel vectors $\boldsymbol{y} \in \mathbb{R}^{n_y}$ and $\boldsymbol{z} \in \mathbb{R}^{n_z}$ and proposed a method to identify and estimate joint and block-specific common factors. There are connections between the subpanel structure and the constrained structure considered in this article. Both approaches produce certain block structures in the loading matrix. Consider the vector factor model case. With two subpanels, the model becomes

$$\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{z}\end{bmatrix} = \begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12} & \boldsymbol{0}\\ \boldsymbol{A}_{21} & \boldsymbol{0} & \boldsymbol{A}_{23}\end{bmatrix}\begin{bmatrix}\boldsymbol{f}_1\\ \boldsymbol{f}_2\\ \boldsymbol{f}_3\end{bmatrix} + \begin{bmatrix}\boldsymbol{\epsilon}_y\\ \boldsymbol{\epsilon}_z\end{bmatrix}.$$

Such a model can be constructed under the constraint approach by specifying

$$\boldsymbol{H} = \begin{bmatrix}\boldsymbol{I} & \boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{I}\end{bmatrix} \quad \text{and} \quad \boldsymbol{R} = \begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{A}_{12} & \boldsymbol{0}\\ \boldsymbol{A}_{21} - \boldsymbol{A}_{11} & \boldsymbol{0} & \boldsymbol{A}_{23}\end{bmatrix},$$

where $\boldsymbol{I}$'s and $\boldsymbol{0}$'s are identity and zero matrices of proper dimensions. However, our current estimation procedure is not able to force certain submatrice in $\boldsymbol{R}$ to zero, though the model can be turned into a multi-term factor model as discussed in Section 2.2. On the other hand, the constraint approach is more flexible in introducing various types of structure in the loading matrix as illustrated in Section 2.1.

The benefits of considering partially constrained matrix factor models are 2-folds. Firstly, the model is capable of identifying, from the complement spaces 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 $(p_2 - m_2)$, respectively. This is because the constraint part should have accommodated the main and key common factors. The spirit is similar to the two-step estimation of Lam and 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 model is able to identify the factors corresponding to the pre-specified constraint matrices and their inherit interpretation. That is, $\boldsymbol{F}_{11,t}$ represents the factor matrix with row and column factors affecting the observed matrix-variate time series 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 in Section 2.1, 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 $\boldsymbol{h}_1 = (1, 1, 1, 1, 1)'$, $\boldsymbol{h}_2 = (1, 1, 0, -1, -1)'$, and $\boldsymbol{h}_3 = (-1, 0, 2, 0, -1)$. Then the estimation method in Section 3 is capable of isolating $\boldsymbol{H}_{R_1}\boldsymbol{R}_1\boldsymbol{F}_{11,t}\boldsymbol{C}_1'\boldsymbol{H}_{C_1}'$, hence correctly estimating the loadings and the specified level, slope, and curvature factors in the constrained spaces. As a result, 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). Here, 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 Equation (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 and

Tsay (2010) and Wang, Liu, and Chen (2019). 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. For ease in representation, we assume that the process $\boldsymbol{F}_t$ has mean $\boldsymbol{0}$, and the observation $\boldsymbol{Y}_t$'s are centered and standardized throughout the article.

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 Equation (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 $\boldsymbol{H}_R^{(1)}$ of Table 2, 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 to be a white-noise process, the estimation method in Wang, Liu, and Chen (2019) 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, Liu, and Chen (2019) for details.

To facilitate the estimation, we use the QR decomposition $\boldsymbol{R} = \boldsymbol{Q}_1\boldsymbol{W}_1$ and $\boldsymbol{C} = \boldsymbol{Q}_2\boldsymbol{W}_2$. The estimation of column spaces of $\boldsymbol{R}$ and $\boldsymbol{C}$ is equivalent to the estimation of column spaces of $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$. Thus, Model (7) can be reexpressed 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 $\boldsymbol{Q}_2'\boldsymbol{Q}_2 = \boldsymbol{I}_{m_2}$.

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 $\{\boldsymbol{Z}_t\boldsymbol{Q}_2'\}_{t=1}^T$ and $\{\boldsymbol{X}_t\}_{t=1}^T$, respectively. For $h > 0$, $\boldsymbol{\Omega}_{x,ij}(h)$ defined in Equation (10) does not involve the covariance terms incurred by $\boldsymbol{E}_t$ because of the whiteness condition.

For a fixed $h_0 \geq 1$ satisfying Condition 2 in Appendix A (supplementary materials), define

$$\boldsymbol{M} = \sum_{h=1}^{h_0}\sum_{i=1}^{m_2}\sum_{j=1}^{m_2} \boldsymbol{\Omega}_{x,ij}(h)\boldsymbol{\Omega}_{x,ij}(h)'$$

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

Since $\boldsymbol{M}$ and the matrix sandwiched by $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_1'$ are positive definite matrices, Equation (11) implies that the eigen-space of $\boldsymbol{M}$ is the same as the column space of $\boldsymbol{Q}_1$ if the middle term is full rank (Condition 2 in Appendix A, supplementary materials). Hence, $\mathcal{M}(\boldsymbol{Q}_1)$ can be estimated by the space spanned by the eigenvectors of the sample version of $\boldsymbol{M}$. The normalized eigenvectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{k_1}$ corresponding to the k_1 nonzero eigenvalues of $\boldsymbol{M}$ are uniquely defined up to a sign change. Thus $\boldsymbol{Q}_1$ is uniquely defined by $\boldsymbol{Q}_1 = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{k_1})$ up to a sign change. We estimate $\widehat{\boldsymbol{Q}}_1 = (\widehat{\boldsymbol{q}}_1, \ldots, \widehat{\boldsymbol{q}}_{k_1})$ as a representative of $\mathcal{M}(\boldsymbol{Q}_1)$ or $\mathcal{M}(\boldsymbol{R})$

The estimation procedure is based on the sample version of these quantities. For $h \geq 1$ and a prescribed positive integer h_0, define the sample version of $\boldsymbol{M}$ in Equation (11) as the following,

$$\widehat{\mathbf{M}} = \sum_{h=1}^{h_0}\sum_{i=1}^{m_2}\sum_{j=1}^{m_2} \widehat{\Omega}_{x,ij}(h)\widehat{\Omega}_{x,ij}(h)', \text{ where}$$

$$\widehat{\Omega}_{x,ij}(h) = \frac{1}{T-h}\sum_{t=1}^{T-h} X_{t,\cdot i}X_{t+h,\cdot j}'. \tag{12}$$

Then, $\mathcal{M}(\boldsymbol{Q}_1)$ can be estimated by $\mathcal{M}(\widehat{\boldsymbol{Q}}_1)$, where $\widehat{\boldsymbol{Q}}_1 = (\widehat{\boldsymbol{q}}_1, \ldots, \widehat{\boldsymbol{q}}_{k_1})$ and $\widehat{\boldsymbol{q}}_i$ is an eigenvector of $\widehat{\boldsymbol{M}}$, corresponding to its ith largest eigenvalue. The $\boldsymbol{Q}_2$ is defined similarly for the column loading matrix $\boldsymbol{C}$ and $\mathcal{M}(\widehat{\boldsymbol{Q}}_2)$ and $\widehat{\boldsymbol{Q}}_2$ can be estimated with the same procedure to the transpose of $\boldsymbol{X}_t$. 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 $\widehat{\boldsymbol{U}}_t = \boldsymbol{Y}_t - \boldsymbol{H}_R\widehat{\boldsymbol{Q}}_1\widehat{\boldsymbol{Z}}_t\widehat{\boldsymbol{Q}}_2'\boldsymbol{H}_C'$.

The above estimation procedure assumes that the number of row factors k_1 is known. To determine k_1, Wang, Liu, and Chen (2019) used the eigenvalue ratio-based estimator of Lam and 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 $K = \lceil p_1/2 \rceil$. k_2 can be estimated with the same procedure with the $\widehat{\boldsymbol{M}}$-matrix corresponding to the transpose of $\boldsymbol{X}_t$.

Although the estimation procedure on the transformed series $\boldsymbol{X}_t$ is exactly the same as that of Wang, Liu, and Chen (2019), the asymptotic properties of the estimator are different due to the transformation, as shown in Section 4, and $\boldsymbol{X}_t$ is of lower dimension.

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 and 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 as that in Section 3.1 to obtain $\widehat{\boldsymbol{Q}}_1^*$ and $\widehat{\boldsymbol{Q}}_2^*$ as the representatives of $\mathcal{M}(\widehat{\boldsymbol{R}}^*)$ and $\mathcal{M}(\widehat{\boldsymbol{C}}^*)$. Then the estimators of $\boldsymbol{R}$ and $\boldsymbol{C}$ are $\widehat{\boldsymbol{R}} = \boldsymbol{K}_R^{-1}\widehat{\boldsymbol{Q}}_1^*$ and $\widehat{\boldsymbol{C}} = \boldsymbol{K}_C^{-1}\widehat{\boldsymbol{Q}}_2^*$. 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 $t = 1, 2, \dots, T$,

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

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 3. For multi-term constrained model (3), $\boldsymbol{H}_{R_1}$ and $\boldsymbol{H}_{R_2}$ (or $\boldsymbol{H}_{C_1}$ and $\boldsymbol{H}_{C_2}$) may not necessarily be orthogonal. In this case, we illustrate the estimation procedure for the column loadings. 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 Equation (3) 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

$$\boldsymbol{Y}_t'\boldsymbol{P}_{\boldsymbol{H}_{R_2}^\perp} = \boldsymbol{H}_{C_1}\boldsymbol{C}_1\boldsymbol{F}_{1,t}'\boldsymbol{R}_1'\boldsymbol{H}_{R_1}'\boldsymbol{P}_{\boldsymbol{H}_{R_2}^\perp} + \boldsymbol{U}_t'\boldsymbol{P}_{\boldsymbol{H}_{R_2}^\perp},$$
$$\boldsymbol{Y}_t'\boldsymbol{P}_{\boldsymbol{H}_{R_1}^\perp} = \boldsymbol{H}_{C_2}\boldsymbol{C}_2\boldsymbol{F}_{2,t}'\boldsymbol{R}_2'\boldsymbol{H}_{R_2}'\boldsymbol{P}_{\boldsymbol{H}_{R_1}^\perp} + \boldsymbol{U}_t'\boldsymbol{P}_{\boldsymbol{H}_{R_1}^\perp},$$

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 a white-noise process.

Let $\boldsymbol{M}^{(lk)}$ represent the $\boldsymbol{M}$ matrix defined in Equation (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

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

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 Equation (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 $p_1 \times p_1$ ambient space of row loading matrix 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 they are latent variables.

Remark 4. In situations when the prior or domain knowledge captures most major factors, it is reasonable to assume that m_i grows slower than p_i and the row (column) factor strength (defined in Condition 6 in Section 4) of the main factor $\boldsymbol{F}_{11,t}$ is no weaker than that of the remainder factor $\boldsymbol{F}_{22,t}$. 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 Equation (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 $k = 1, 2$.

4. Theoretical Properties

In this section, we present the convergence rates of the estimators under the setting that p_1, p_2, m_1, m_2, and T all go to infinity while the dimensions k_1, k_2 and the structure of the latent factor are fixed over time. In what follows, let $\|\boldsymbol{A}\|_2$, $\|\boldsymbol{A}\|_F$, and $\|\boldsymbol{A}\|_{\min}$ denote the spectral norm, Frobenius norm, and the smallest nonzero singular value of $\boldsymbol{A}$, respectively. When $\boldsymbol{A}$ is a square matrix, we denote by $tr(\boldsymbol{A})$, $\lambda_{\max}(\boldsymbol{A})$ and $\lambda_{\min}(\boldsymbol{A})$ 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 $b_N = O(a_N)$.

The asymptotic convergence rates are significantly different from those in Wang, Liu, and Chen (2019) 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 1–5) are listed in the Appendix (supplementary materials). They are similar to those in Wang, Liu, and Chen (2019) and are used to derive the limiting behavior of Equation (12) toward 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 Condition 6 implies 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 that 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, Yao, and Bathia (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 at the rate of $\sqrt{p_1^{1-\delta_1}/m_1}$. 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 (see Lemma 1 in Appendix A (supplementary materials), 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-to-noise ratio in the constrained model.

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

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

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

Table 3. Convergence rate of the loading space estimator.

	$m_1m_2 \asymp p_1p_2$	$p_1^{1-\delta_1}p_2^{1-\delta_2} = o(m_1m_2)$	$m_1m_2 = o(p_1^{1-\delta_1}p_2^{1-\delta_2})$
Rate	$p_1^{\delta_1}p_2^{\delta_2}T^{-1/2}$	$m_1m_2p_1^{-1}p_2^{-1}\cdot p_1^{\delta_1}p_2^{\delta_2}T^{-1/2}$	$T^{-1/2}$

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

Remark 5. The convergence rate for the unconstrained model is $p_1^{\delta_1}p_2^{\delta_2}T^{-1/2}$ in Wang, Liu, and Chen (2019). The rates for the constrained model under different relations between m_1m_2 and p_1p_2 are shown in Table 3.

For strong factors with $\delta_1 = \delta_2 = 0$, the convergence rates are the same for the constrained and unconstrained models. However, $m_1p_1^{-1+\delta_1}m_2p_2^{-1+\delta_2}T^{-1/2} = o(1)$ is automatically satisfied since $p_i >> m_i$. Also, the constrained models have much smaller number of parameters, hence potentially have higher efficiency.

For weak factors, the constrained models have better convergence rate in most cases. It depends on the growth rate of the ratio between m_1m_2 and $p_1^{1-\delta_1}p_2^{1-\delta_2}$. The smaller the ratio, the faster the convergence rate. It can be viewed as strength gained due to the constraints. For example, when $m_1 = p_1^{\alpha_1}$ and $m_2 = p_2^{\alpha_2}$, the convergence 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 $\alpha_2 < 1$. It effectively increases the strength from δ_1 and δ_2 to $\delta_1 - (1 - \alpha_1)$ and $\delta_2 - (1 - \alpha_2)$, respectively. Hence, the constraints are particularly useful for weak strength cases.

When $m_1m_2 = O(p_1^{1-\delta_1}p_2^{1-\delta_2})$, we achieve the optimal rate $O_p\left(T^{-1/2}\right)$. Note the unconstrained model can only achieve this rate in the case of strong factor. The constrained model can achieve the optimal rate even in the weak factor case. A special case is when the dimensions of the constrained row and column loading spaces m_1 and m_2 are fixed, the convergence rate is $T^{-1/2}$ regardless of the factor strength condition. Increasing p_1 or p_2 while keeping m_1 and m_2 fixed amounts to increasing the sample points in the constrained spaces. When the constrained spaces are properly specified, the additional information introduced from more sample points will accrue and translate into the transformed signal part in Equation (7), while the transformed noise gets canceled out by averaging. However, the convergence rate is still bounded below by the convergence rate of the estimated covariance matrix. When $m_1m_2 \asymp p_1p_2$, the convergence rates of the constrained and unconstrained models are the same. A special case is when $m_1 = c_1p_1$ and $m_2 = c_2p_2$, that is, the dimensions of the constrained loading spaces increase with p's linearly.

Remark 6. Under some conditions the convergence rates in Theorem 1 may improve significantly. For example, if $\Sigma_u \equiv \mathrm{var}(\mathrm{vec}(\boldsymbol{U}_t))$ is diagonal (i.e., $U_{t,ij}$ and $U_{t,lk}$ are uncorrelated for $(i,j) \neq (l,k)$) and if we have the grouping constraints (i.e., $\boldsymbol{H}_R^{(1)}$ in Section 2.1), then each elements in $\boldsymbol{E}_t$ is a group average. $\mathrm{var}(\boldsymbol{E}_{t,ij})$ is smaller by a factor of $\frac{m_1m_2}{p_1p_2}$ and goes to zero when $\frac{m_1m_2}{p_1p_2} = o(1)$.

Remark 7. If the constraints are correct, it would induce certain intrinsic sparsity in the auto-cross-correlation matrix under the unconstrained model. For such intrinsic sparsity conditions, we may instead use thresholding estimator for large covariance matrix by Bickel and Levina (2008) in Equation (11). This will lead to faster convergence rates. See Section 3.2 of Chang, Guo, and Yao (2018). By explicitly incorporating constraints in the model, the loading matrix is condensed and the sparsity issue becomes less serious.

Remark 8. The factors under our definition contain the classic "common factors" and the serially correlated idiosyncratic components. As shown by the theoretical properties and simulation studies, the constrained matrix factor model helps identify the weak factors. However, the method is still limited in the sense that it can only identify "common" factors of some strength $\delta < 1$. In the case of $\delta = 1$, although the loading spaces can still be consistently estimated with very large T ($pT^{-1/2} = o(1)$), the factor itself cannot be consistently estimated. Therefore, serially correlated idiosyncratic components for which $\delta = 1$ are left "erroneously" in the noise in application. Hopefully, the constraints may improve the effective factor strength.

Theorem 2. Under Conditions 1–6, and if $m_1 p_1^{-1+\delta_1} m_2 p_2^{-1+\delta_2} T^{-1/2} = o(1)$ and the $\boldsymbol{M}$ matrix in Equation (11) has k_1 distinct positive eigenvalues, then the eigenvalues $\{\hat{\lambda}_1, \ldots, \hat{\lambda}_{m_1}\}$ of $\widehat{\boldsymbol{M}}$, sorted in the descending order, satisfy

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

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

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

Theorem 2 shows that the estimators of the nonzero eigenvalues of $\boldsymbol{M}$ converge more slowly than those of the zero eigenvalues. This provides the theoretical support for the ratio-based estimator of the number of factors described in Section 3.1. 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.

The convergence rates for the unconstrained model are $\Delta_{pT}^{\lambda} \equiv p_1^{2-\delta_1} p_2^{2-\delta_2} T^{-1/2}$ for the nonzero eigenvalues and $p_1^{\delta_1} p_2^{\delta_2} T^{-1/2} \cdot \Delta_{pT}^{\lambda}$ for the zero eigenvalues, respectively. See Wang, Liu, and Chen (2019). The rates for the constrained model under different relations between $m_1 m_2$ and $p_1 p_2$ are shown in Table 4.

In the cases of strong factors or weak factors with $m_1 m_2 \asymp p_1 p_2$, our result is the same as that of Wang, Liu, and Chen (2019). In all other cases, 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 $\boldsymbol{Y}_t$, that is, $\boldsymbol{S}_t = \boldsymbol{H}_R \boldsymbol{R} \boldsymbol{F}_t \boldsymbol{C}' \boldsymbol{H}_C' = \boldsymbol{H}_R \boldsymbol{Q}_1 \boldsymbol{Z}_t \boldsymbol{Q}_2' \boldsymbol{H}_C'$. 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 $m_1 p_1^{-1+\delta_1} m_2 p_2^{-1+\delta_2} T^{-1/2} = o(1)$, we have

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

Theorem 3 shows that as long as $m_1 m_2 = o(p_1 p_2)$ we achieve for weak factor cases a faster convergence rate than $O_p\left(p_1^{\delta_1/2} p_2^{\delta_2/2} T^{-1/2} + p_1^{-1/2} p_2^{-1/2}\right)$—the convergence rate of the unconstrained model in Wang, Liu, and Chen (2019). When $m_1 m_2 = O_p(p_1^{1-\delta_1} p_2^{1-\delta_2})$, we get an even better rate. 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 $p_1, p_2 \to \infty$.

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

5. Simulation

In this section, we present some simulation study to illustrate 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 with 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 noises shows similar results.

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, Yao, and Bathia (2011) and Wang, Liu, and Chen (2019), we adopt the discrepancy measure used by Chang, Guo, and Yao (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, Equation (15) is the cosine similarity measure. We report this space distance $\mathcal{D}(\cdot, \cdot)$ as a measurement of the discrepancy between estimated and true loading spaces.

Table 4. Convergence rate of estimators for nonzero and zero eigenvalues of $\boldsymbol{M}$.

λ_j	$m_1m_2 \asymp p_1p_2$	$p_1^{1-\delta_1}p_2^{1-\delta_2} = o(m_1m_2)$	$m_1m_2 = o(p_1^{1-\delta_1}p_2^{1-\delta_2})$
Zero	$p_1^{\delta_1}p_2^{\delta_2}T^{-1/2}\cdot\Delta_{pT}^{\lambda}$	$(\frac{m_1m_2}{p_1p_2})^2p_1^{\delta_1}p_2^{\delta_2}T^{-1/2}\cdot\Delta_{pT}^{\lambda}$	$p_1^{-\delta_1}p_2^{-\delta_2}T^{-1/2}\cdot\Delta_{pT}^{\lambda}$
Nonzero	Δ_{pT}^{λ}	$m_1m_2p_1^{-1}p_2^{-1}\cdot\Delta_{pT}^{\lambda}$	$p_1^{-\delta_1}p_2^{-\delta_2}\cdot\Delta_{pT}^{\lambda}$
Ratio	$p_1^{\delta_1}p_2^{\delta_2}T^{-1/2}$	$m_1m_2p_1^{-1+\delta_1}p_2^{-1+\delta_2}T^{-1/2}$	$T^{-1/2}$

5.1. Case 1. Orthogonal Constraints

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

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

under the following 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_1k_2 independent $AR(1)$ processes with Gaussian white noise $\mathcal{N}(0,1)$ innovations. Specifically, $\text{vec}(\boldsymbol{F}_t) = \boldsymbol{\Phi}_F\,\text{vec}(\boldsymbol{F}_{t-1}) + \boldsymbol{\epsilon}_t$ with $\boldsymbol{\Phi}_F = \text{diag}(-0.5, 0.6, 0.8, -0.4, 0.7, 0.3)$. 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 $\text{cov}(\text{vec}(\boldsymbol{U}_t)) = \boldsymbol{\Gamma}_2 \otimes \boldsymbol{\Gamma}_1$, 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. The 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 Equation (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 $\boldsymbol{h}_{c,1} = [\boldsymbol{1}_7/\sqrt{7}, \boldsymbol{0}_7, \boldsymbol{0}_6]'$, $\boldsymbol{h}_{c,2} = [\boldsymbol{0}_7, \boldsymbol{1}_7/\sqrt{7}, \boldsymbol{0}_6]'$, $\boldsymbol{h}_{c,3} = [\boldsymbol{0}_7, \boldsymbol{0}_7, \boldsymbol{1}_6/\sqrt{6}]'$, where $\boldsymbol{0}_k$ 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 5 shows the performance of estimating the true number of row and column factors separately. 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, that is, $(\delta_1, \delta_2) = (0, 0)$, both constrained and unconstrained models can estimate accurately the number of factors, but the constrained models perform 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, that is, $(\delta_1, \delta_2) = (0.5, 0)$, the unconstrained models fail to estimate the number of row factors k_1, but the constrained models continue to perform well for both k_1 and k_2. 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, that is, $(\delta_1, \delta_2) = (0.5, 0.5)$, both models encounter difficulties in estimating the correct number of row factors k_1 for the sample sizes used. However, the constrained models continue to perform well for the number of column factors k_2. Here, m_1 and m_2 are different and play a role. Since $m_1 > m_2$, k_2 is estimated with higher accuracy, especially in the weak factor case.

Figure 1 shows the boxplots 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, that is, $(\delta_1, \delta_2) = (0, 0)$, and the number of factors is properly estimated, the mean and SD 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, that is, $(\delta_1, \delta_2) = (0.5, 0)$, and the true number of factors is used, the estimation error of constrained models remains small whereas that of the unconstrained models is substantially larger.

Table 6 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 dimension of the constrained column space (m_2) is smaller than that of the constrained row space (m_1). Intuitively, after transformation Equation (7), the ratio of the effective column factor strength $\|\boldsymbol{C}\|_2^2$ and noise level $\|\boldsymbol{E}_t\|_2^2$ is larger than the ratio of the effective row factor strength $\|\boldsymbol{R}\|_2^2$ and noise level $\|\boldsymbol{E}_t\|_2^2$. From the table, we see that (a) the mean of estimation errors decreases, as expected, as the sample size increases and (b) the mean of estimation errors is inversely proportional to the strength of row factors.

Table 5. Relative frequencies of correctly estimating the number of row (column) factors k_1 (k_2) in the case of orthogonal constraints, where p_i are the dimension, T is the sample size, and f_u and f_c denote the results of unconstrained and constrained factor model, respectively. .

k_1				$T = 0.5p_1p_2$		$T = p_1p_2$		$T = 1.5p_1p_2$		$T = 2p_1p_2$	
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	0.264	0.942	0.728	0.996	0.952	1	0.996	1
		20	40	0.734	1	0.998	1	1	1	1	1
		40	20	0.786	0.994	1	1	1	1	1	1
		40	40	1	1	1	1	1	1	1	1
0.5	0	20	20	0	0.202	0	0.508	0	0.734	0	0.926
		20	40	0	0.692	0	0.97	0	0.994	0	1
		40	20	0	0.352	0	0.77	0	0.908	0	0.964
		40	40	0	0.872	0	0.984	0	0.996	0	0.994
0.5	0.5	20	20	0	0.052	0	0.036	0	0.01	0	0.006
		20	40	0	0.052	0	0.022	0	0.008	0	0.004
		40	20	0	0.062	0	0.006	0	0	0	0.002
		40	40	0	0.018	0	0.006	0	0.006	0	0.044
k_2				$T = 0.5p_1p_2$		$T = p_1p_2$		$T = 1.5p_1p_2$		$T = 2p_1p_2$	
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	1	1	1	1	1	1	1	1
		20	40	1	1	1	1	1	1	1	1
		40	20	1	1	1	1	1	1	1	1
		40	40	1	1	1	1	1	1	1	1
0.5	0	20	20	0.988	0.996	1	1	1	1	1	1
		20	40	1	1	1	1	1	1	1	1
		40	20	1	1	1	1	1	1	1	1
		40	40	1	1	1	1	1	1	1	1
0.5	0.5	20	20	0.02	0.762	0.104	0.962	0.272	0.992	0.58	1
		20	40	0	0.956	0.09	0.998	0.472	1	0.85	1
		40	20	0	0.952	0.026	1	0.196	1	0.572	1
		40	40	0	1	0.03	1	0.438	1	0.906	1

NOTE: Table on top is for k_1, while table on the bottom is for k_2.

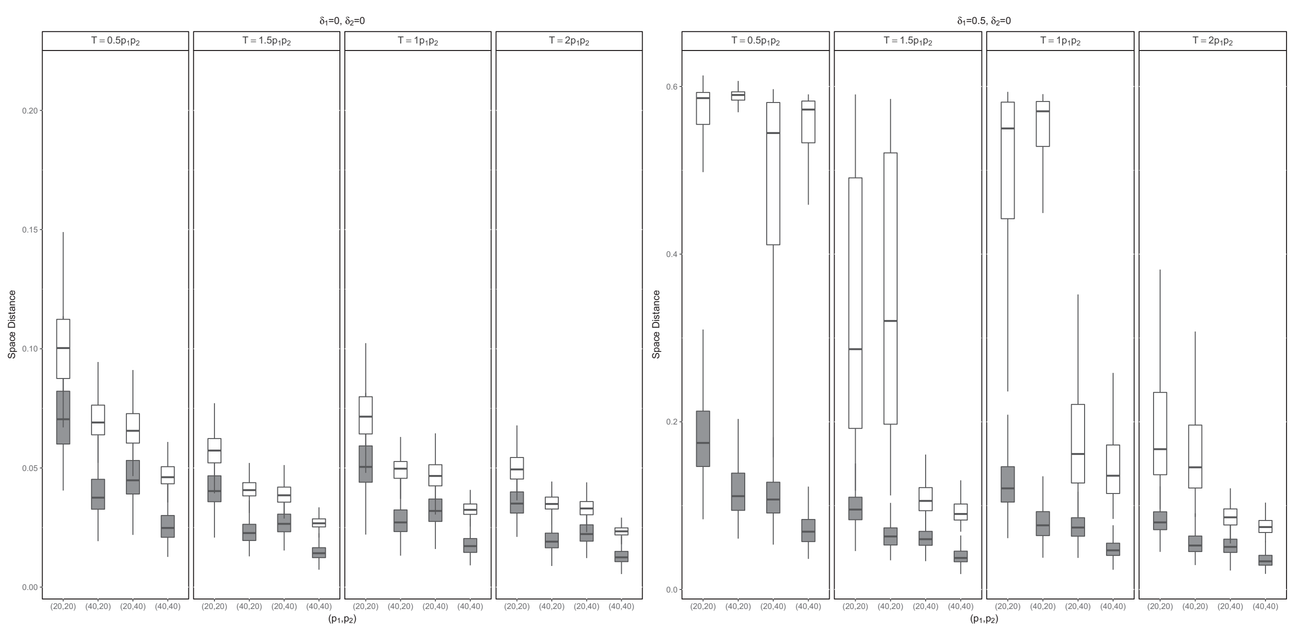


Figure 1. Boxplots of the estimation accuracy measured by $\mathcal{D}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$ of Equation (15) for the case of orthogonal constraints. Gray boxes represent the constrained model. The results are based on 500 iterations. See Table 17 in Appendix C (supplementary materials) for plotted values.

To investigate the performance of estimation under different choices of h_0, which is the number of lags used in Equation (11), we change the underlying generating model of $\mathrm{vec}(\boldsymbol{F}_t)$ to a VAR(2) process without the lag-1 term, $\mathrm{vec}(\boldsymbol{F}_t) = \boldsymbol{\Phi}_F \mathrm{vec}(\boldsymbol{F}_{t-2}) + \boldsymbol{\epsilon}_t$. 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 the prior simulation. Table 7 presents the simulation results. Since $\mathrm{vec}(\boldsymbol{F}_t)$, and hence $\mathrm{vec}(\boldsymbol{Y}_t)$, has zero auto-covariance matrix at lag 1, $\widehat{\boldsymbol{M}}$ under $h_0 = 1$ contains

Table 6. Means and SDs (in parentheses) of the estimation accuracy measured by $\mathcal{D}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$ of Equation (15) for constrained factor models. The case of orthogonal constraints is used.

				$T = 0.5p_1p_2$		$T = p_1p_2$		$T = 1.5p_1p_2$		$T = 2p_1p_2$	
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)

NOTES: The subscripts 1 and 2 denote row and column, respectively. All numbers in the table are 10 times of the true numbers for clear presentation. The results are based on 500 simulations.

Table 7. Performance of estimation under different choices of h_0 when $\text{vec}(\boldsymbol{F}_t) = \boldsymbol{\Phi}_F \text{vec}(\boldsymbol{F}_{t-2}) + \boldsymbol{\epsilon}_t$.

	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)

NOTES: Metrics reported are relative frequencies of correctly estimating k, means and SDs (in parentheses) of the estimation accuracy measured by $\mathcal{D}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$. Means and SDs are multiplied by 10 for ease in presentation. f_u and f_c denote unconstrained and constrained models.

no information on the signal, 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 similar results shows that the choice of h_0 is not critical to the performance of the proposed method as long as it is sufficiently large to describe the pattern of the autocovariance matrices of the data. See Condition 2 in Appendix A (supplementary materials). 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_1q_2 independent AR(1) processes with Gaussian white noise $\mathcal{N}(0, 1)$ innovations, $\text{vec}(\boldsymbol{G}_t) = \boldsymbol{\Phi}_G \text{vec}(\boldsymbol{G}_{t-1}) + \boldsymbol{\epsilon}_t$ 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 $p_i^{-\eta_i/2}\sqrt{p_i/(p_i - m_i)}$ 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 8 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_{con_1} and f_{con_2} denote the same frequency metric for the first matrix factor $\boldsymbol{F}_t$ and

Table 8. Relative frequencies of correctly estimating the number of factors for partially constrained factor models.

						$T = 0.5p_1p_2$			$T = p_1p_2$			$T = 1.5p_1p_2$			$T = 2p_1p_2$		
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

NOTE: The full table including all combinations is presented in Table 18 in Appendix C (supplementary materials).

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 $m_1m_2 = 36$, 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 8 only contains selected combinations of factor strength parameters δ_i's ($i = 1, \ldots, 4$). The results of all combinations of factor strength are given in Table 18 in Appendix C (supplementary materials).

Figures 2 and 3 present boxplots of estimation errors under weak and strong factors based on 500 simulations, respectively. Again, the results show that the constrained approach effectively 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.

6. Applications

In this section, we demonstrate the advantages of 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 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 and Yao (2012) and Wang, Liu, and Chen (2019). For any given (k_1, k_2), the dimension 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.

6.1. Example 1: Multinational Macroeconomic Indices

We apply the constrained and partially constrained factor models to the macroeconomic indexes collected from OECD. The dataset contains 10 quarterly macroeconomic indexes of 14 countries from 1990.Q2 to 2016.Q4 for 107 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 indexes cover four major groups, namely production, consumer price, money market, and international trade. Each original univariate time series is transformed by taking the first or second difference or logarithm to satisfy the mixing condition specified in Condition 4 in the supplementary materials. Detailed descriptions of the dataset and the transformation used are given in Table 15 and 16 of Appendix B (supplementary materials). Figure 4 shows the transformed time series of macroeconomic indicators of multiple countries.

We first fit an unconstrained matrix factor model and obtain estimates of the row loading matrix and the column loading matrix. In the row loading matrix, each row represents a country by its factor loadings, whereas, in the column loading matrix, each row represents a macroeconomic index by its factor loadings. A hierarchical clustering algorithm (Xu and Wunsch 2005; Murtagh and Legendre 2014) is employed to cluster countries and macroeconomic indices based on their representations in the common row and column factor spaces, respectively, under Euclidean distance and *ward.D* criterion. Figure 5 shows the hierarchical clustering results. Based on the clustering result, we construct the row and column constraint matrices. It seems that the row constraint matrix divides countries into six 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 partitions with Norway different from all others due to its rich oil production and other distinct economic characteristics. The column constraint matrix divides macroeconomic indexes into five 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 9 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 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

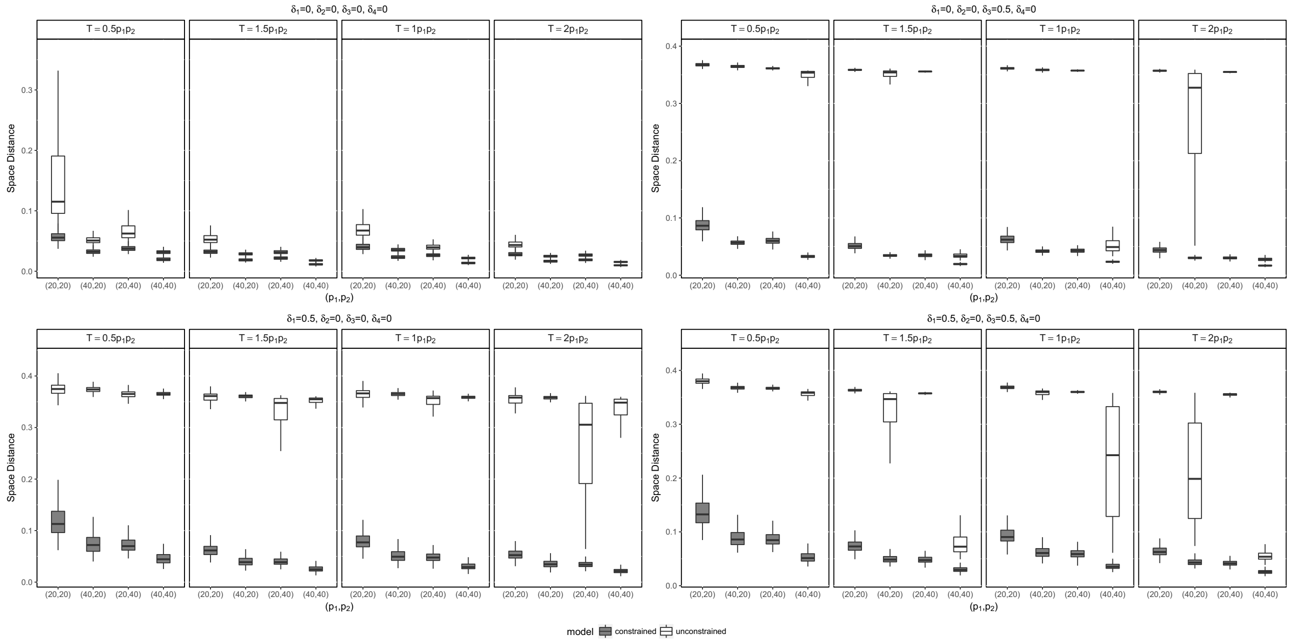


Figure 2. The strong factors case. Boxplots of the estimation accuracy measured by $\mathcal{D}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$ for partially constrained factor models. The gray boxes are for the constrained approach. The results are based on 500 realizations. See Table 19 in Appendix C (supplementary materials) for the plotted values.

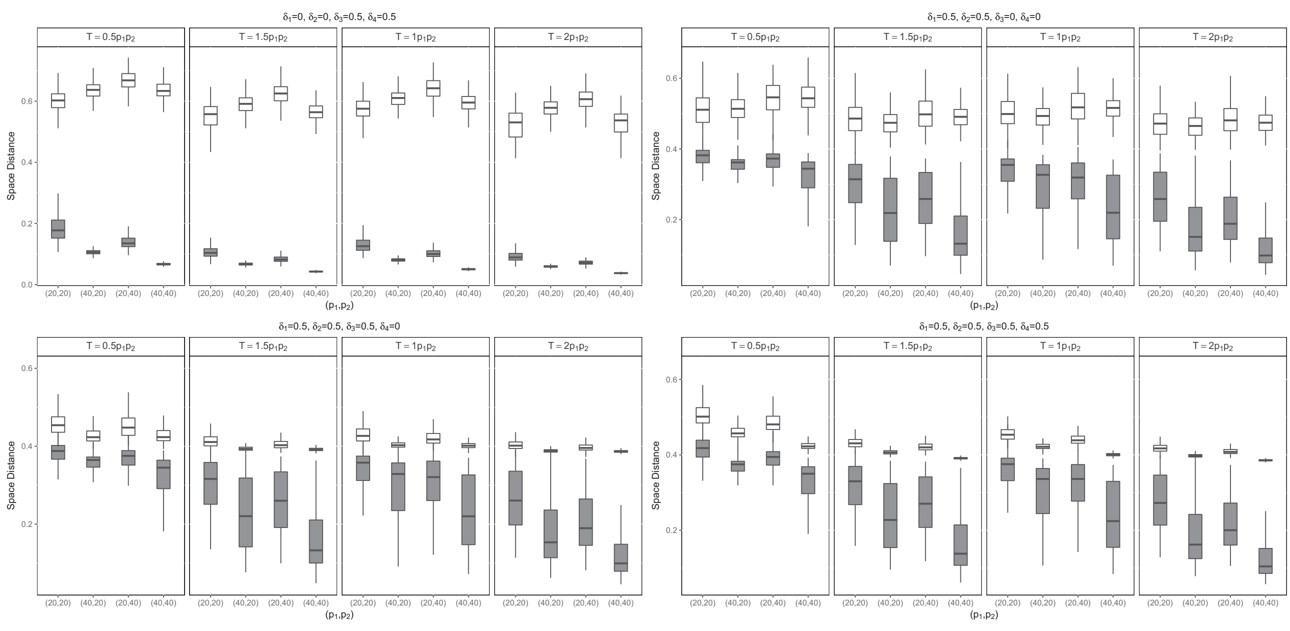


Figure 3. The weak factors case. Boxplots of the estimation accuracy measured by $\mathcal{D}(\widehat{\boldsymbol{Q}}, \boldsymbol{Q})$ for partially constrained factor models. The gray boxes are for the constrained approach. The results are based on 500 realizations. See Table 19 in Appendix C (supplementary materials) for the plotted values.

unconstrained and constrained models, partially validating the constraints while simplifying the analysis.

Table 10 provides the estimates under the same setting as that of Table 9 but without any rotation. From the table, it is seen that except for the first common factor of the row loading matrices there exist some differences 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 indexes, interest rates, productions, and international trade.

Table 11 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. We divide the entire time span into 10 sections and choose each of them as testing data. With time series data, the training data may contain two disconnected

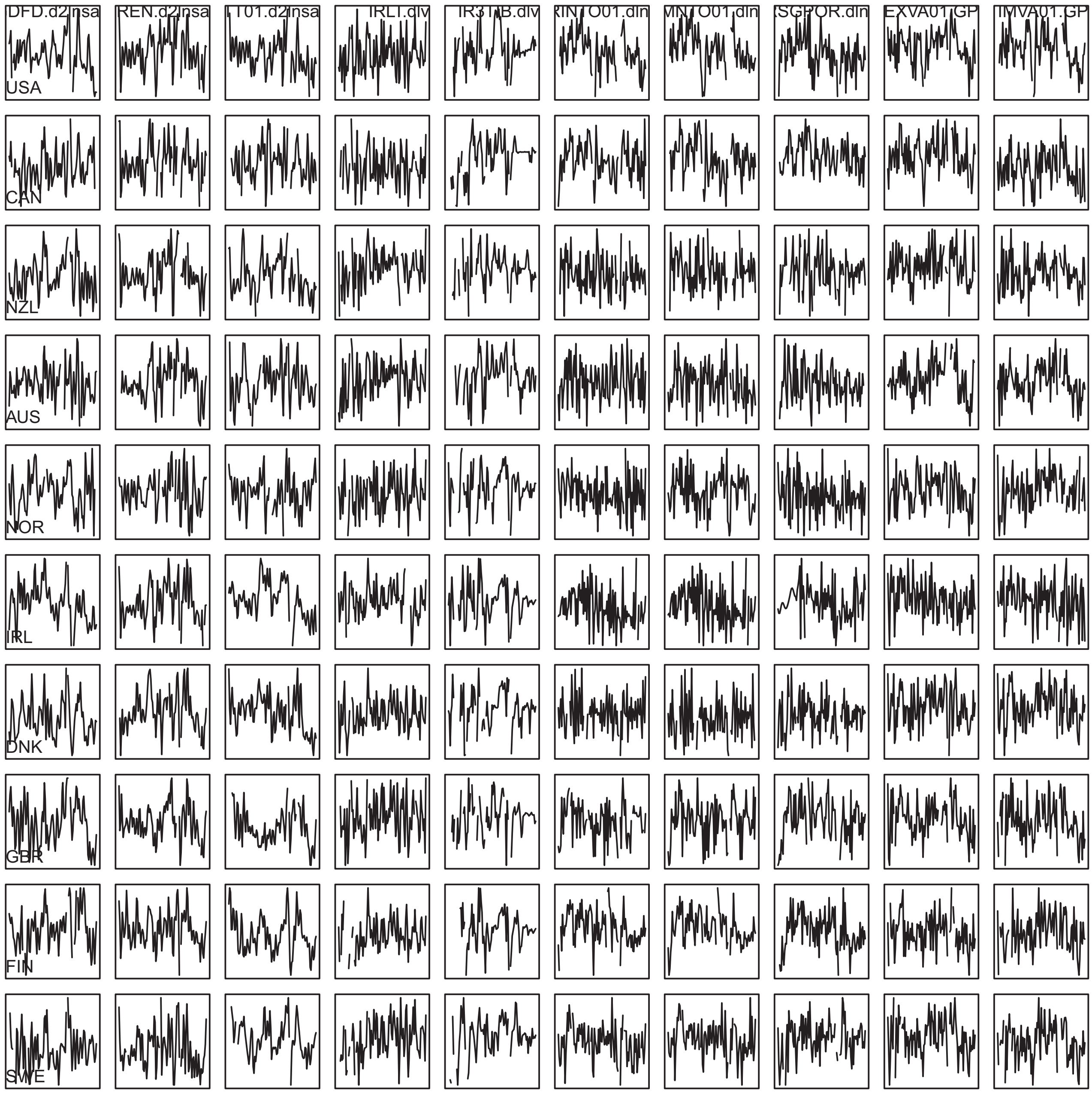


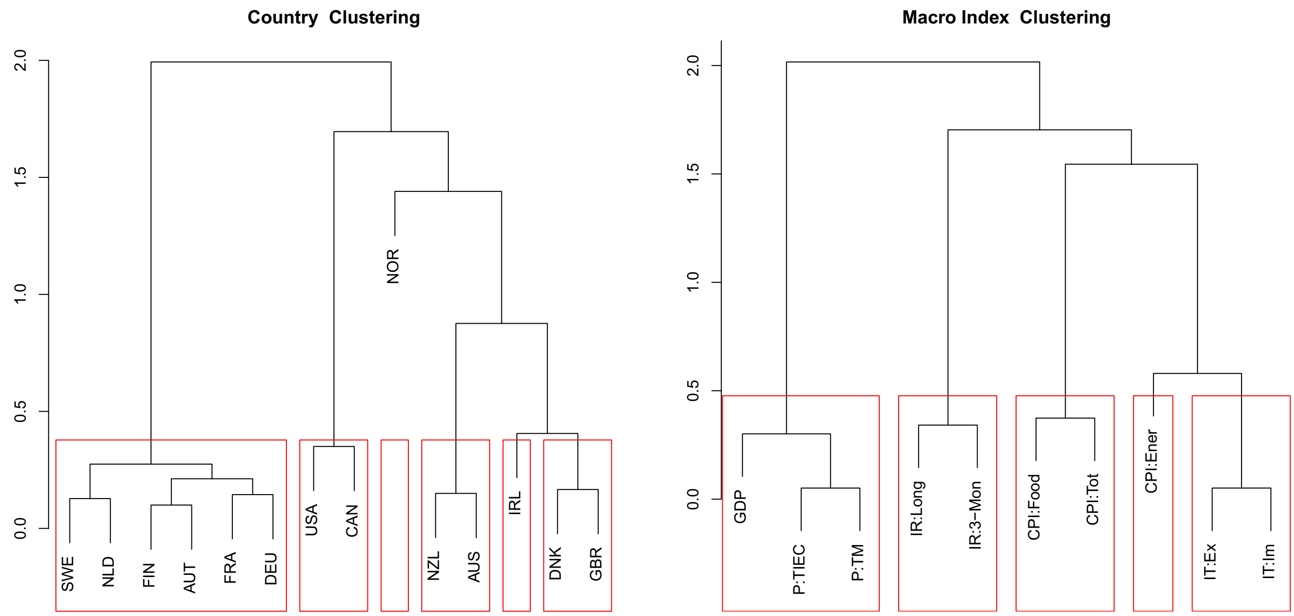
Figure 4. Time series plots of macroeconomic indicators of multiple countries (after data transformation). Only a subset of the countries and indicators are plotted due to the space limit. Country and indicator abbreviation are given in Tables 15 and 16 of Appendix C (supplementary materials).

time spans. We calculate two matrices $\widehat{\boldsymbol{M}}_1$ and $\widehat{\boldsymbol{M}}_2$ according to Equation (12) for the two disconnected time spans separately. The matrix $\widehat{\boldsymbol{M}}$ is redefined as the sum of $\widehat{\boldsymbol{M}}_1$ and $\widehat{\boldsymbol{M}}_2$. Loading matrices and latent dimensions are estimated from this newly defined $\widehat{\boldsymbol{M}}$ with procedures in Section 3. Residual sum of squares (RSS), their ratios to the total sum of squares (RSS/TSS), and the number of parameters are the average 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 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 Financial Measurements

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 D (supplementary materials) contains the descriptions of variables used and their definitions, the 200



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

Figure 5. Macroeconomic series: clustering loading matrices.

Table 9. Estimations of row and column loading matrices (varimax rotated) of constrained and unconstrained matrix factor models for multinational macroeconomic indices.

Model	Loading	Row	USA	CAN	NZL	AUS	NOR	IRL	DNK	GBR	FIN	SWE	FRA	NLD	AUT	DEU
$R_{\text{unc,rot}}$	$\widehat{R}'_{\text{rot}}$	1	7	7	1	1	−1	−2	−1	0	1	0	0	0	0	−1
		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
$R_{\text{con,rot}}$	$\widehat{R}'_{\text{rot}}H'_R$	1	6	6	0	0	0	2	2	2	−1	−1	0	0	0	0
		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:Ener	CPI:Tot	IR:Long	IR:3-Mon	P:TIEC	P:TM	GDP	IT:Ex	IT:Im
$C_{\text{unc,rot}}$	$\widehat{C}'_{\text{rot}}$	1	6	7	3	−1	1	0	0	−1	−1	0
		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
$C_{\text{con,rot}}$	$\widehat{C}'_{\text{rot}}H'_C$	1	7	7	0	0	0	0	0	0	0	0
		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

NOTE: The loadings matrix are multiplied by 10 and rounded to integers for ease in display.

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 3200. Following the convention in eigenanalysis, we standardize the individual series before applying factor analysis. This dataset was used in Wang, Liu, and Chen (2019) for an unconstrained 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 financial, we do not 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. Table 12 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 numbers 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 3200 time series is typ-

Table 10. Estimations of row and column loading matrices of constrained and unconstrained matrix factor models for multinational macroeconomic indices.

Model	Loading	Row	USA	CAN	NZL	AUS	NOR	IRL	DNK	GBR	FIN	SWE	FRA	NLD	AUT	DEU
		1	3	2	2	2	2	2	3	3	3	3	3	3	3	3
R_{unc}	$\widehat{R}'$	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
		1	1	1	2	2	2	3	3	3	4	4	3	3	3	3
R_{con}	$\widehat{R}'H_R'$	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
		1	1	4	2	4	3	3	3	3	4	4
C_{unc}	$\widehat{C}'$	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
		1	6	−2	6	4	4	0	0	0	−2	−2
C_{con}	$\widehat{C}'H_C'$	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

NOTES: No rotation is used. The loadings matrix are multiplied by 10 and rounded to integers for ease in display.

Table 11. Results of 10-fold CV of out-of-sample performance for the multinational macroeconomic indexes.

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

NOTES: The numbers shown are average over the cross-validation, where RSS and TSS stand for residual and total sum of squares, respectively.

Table 12. Summary of 10-fold CV of out-of-sample analysis for the 16 corporate financial measurements for each of 200 companies.

Model	# Factor 1	# Factor 2	RSS	RSS/SST	# parameters
	(4,10)		8140.32	0.869	2064
Full	(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
	(4, 20)		7539.68	0.805	4064
Full	(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
	(5,10)		8012.10	0.855	2080
Full	(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
	(5,20)		7368.63	0.787	7960
Full	(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
	(5,30)		6960.70	0.743	6080
Full	(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

NOTES: The numbers shown are average over the cross-validation and RSS and TSS denote, respectively, the residual and total sum of squares.

ical in high-dimensional time series analysis. The number of parameters involved is usually huge in a unconstrained model. Via the example, we showed that constrained matrix factor models can substantially reduce the number of parameters while keep 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 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-variate 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, Liu, and Chen (2019) carried out a clustering analysis on the ME and BE/ME loading matrices after rotation. Their results suggest

Table 13. Estimates of the loading matrices of constrained and unconstrained matrix factor modes for Fama–French 10 × 10 portfolio returns.

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

NOTES: The loading matrices are varimax rotated and normalized for ease in comparison.

$\boldsymbol{H}_R = [\boldsymbol{h}_{R_1}, \boldsymbol{h}_{R_2}, \boldsymbol{h}_{R_3}]$, where $\boldsymbol{h}_{R_1} = [\mathbf{1}(5)/\sqrt{5}, \mathbf{0}(5)]$, $\boldsymbol{h}_{R_2} = [\mathbf{0}(5), \mathbf{1}(4)/2, 0]$, and $\boldsymbol{h}_{R_3} = [\mathbf{0}(9), 1]$. Therefore, ME factors are classified into three groups of smallest 5 MEs, middle 4 MEs, and the largest ME, respectively. For cases when we need 4 row constraints, we redefine $\boldsymbol{h}_{R_2} = [\mathbf{0}(5), \mathbf{1}(3)/\sqrt{3}, \mathbf{0}(2)]$ and add a fourth column $\boldsymbol{h}_{R_4} = [\mathbf{0}(8), 1, 0]$. For column constraints, $\boldsymbol{H}_C = [\boldsymbol{h}_{C_1}, \boldsymbol{h}_{C_2}, \boldsymbol{h}_{C_3}]$, where $\boldsymbol{h}_{C_1} = [1, \mathbf{0}(9)]$, $\boldsymbol{h}_{C_2} = [0, \mathbf{1}(3)/\sqrt{3}, \mathbf{0}(6)]$, $\boldsymbol{h}_{C_3} = [\mathbf{0}(4), \mathbf{1}(6)]$. Therefore, BE/ME factors are divided into three groups of the smallest BE/MEs, middle 3 BE/MEs, and the 6 largest BE/ME, respectively. For cases when we need 4 column constraints, we redefine $\boldsymbol{h}_{C_3} = [\mathbf{0}(4), \mathbf{1}(4)/2, \mathbf{0}(2)]$ and add a fourth column $\boldsymbol{h}_{C_4} = [\mathbf{0}(8), \mathbf{1}(2)]$.

Table 13 shows the estimates of the loading matrices for the constrained and unconstrained 2 × 2 factor models. The loading matrices are varimax-rotated 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 result demonstrates the power of using constrained factor models.

Table 14 compares the out-of-sample performance of unconstrained and constrained matrix factor models using a 10-fold CV for models with different number of factors constructed similarly to that of Table 11. 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 to be addressed in future research. On the other hand, the constrained model uses a much smaller number of parameters than the unconstrained model.

Table 14. 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.

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

7. Summary and Discussion

This article established a general framework for incorporating domain or prior knowledge induced linear constraints in the matrix factor model. We developed efficient estimation procedures for constrained, multi-term, and partially constrained matrix factor models. 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 datasets, 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.

Under the model setting we adopt, both strong and weak factors exist in the dynamic component. The proposed constrained model incorporates prior information and improves the rates of convergence in the case of weak factors. For the strong factor case, it achieves the same asymptotic rates as those of the unconstrained models. Yet it entails smaller number of parameters and requires weaker assumption on the growth rates of dimensions and sample size. Several interesting topics are open for further researches. First, a natural question is how we know the existence of weak factors in real data. Lam and Yao (2012) used a two-step approach to facilitate the discovery of weak factors that may be masked by strong factors. They ran a second decomposition to the residual from the first step to find the weak factors that may be masked from the strong factor in the first step. A possible method to test the existence of weak factor will be to test the existence of common factors in the second step. Also, data containing subpanels or block structure is a common situation where weak factors arise. Hallin and Liška (2011) developed method to identify and estimate joint and block-specific common factors among different panels. Similar result can be achieved by using constrained vector factor model in Tsai and Tsay (2010). Data containing subpanels can also be cast into matrix observations by putting subpanels as columns. The column spaces of loadings can be divided into subspaces that correspond to the joint and block-specific common factors. However, more sophisticated estimation procedures need to be developed to exclude overlap of the column spaces. The constrained matrix factor model provides building blocks for future research on combining constraints to represent different structures and on devising estimation procedures.

Supplementary Materials

The supplementary materials contain all technical proofs, more information on the data sets used in the real applications, and some extra simulation results. (UASA_A_1584899_SM4831.pdf)

Funding

EYC supported in part by NSF Grants DMS-1803241. RST supported in part by the Booth School of Business, University of Chicago. RC supported in part by NSF Grants DMS-1503409, DMS-1737857, DMS-1803241, and IIS-1741390.

ORCID

Rong Chen http://orcid.org/0000-0003-0793-4546

References

Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171. [775,776]

Bai, J., and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221. [775,776]

——— (2007), "Determining the Number of Primitive Shocks in Factor Models," *Journal of Business & Economic Statistics*, 25, 52–60. [775,776]

Bansal, N., Connolly, R. A., and Stivers, C. (2014), "The Stock-Bond Return Relation, the Term Structure's Slope, and Asset-Class Risk Dynamics," *Journal of Financial and Quantitative Analysis*, 49, 699–724. [775]

Bickel, P. J., and Levina, E. (2008), "Covariance Regularization by Thresholding," *The Annals of Statistics*, 36, 2577–2604. [783]

Chamberlain, G. (1983), "Funds, Factors, and Diversification in Arbitrage Pricing Models," *Econometrica*, 51, 1305–1323. [775]

Chamberlain, G., and Rothschild, M. (1983), "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–1304. [775]

Chang, J., Guo, B., and Yao, Q. (2015), "High Dimensional Stochastic Regression With Latent Factors, Endogeneity and Nonlinearity," *Journal of Econometrics*, 189, 297–312. [775,776,783]

——— (2018), "Principal Component Analysis for Second-Order Stationary Vector Time Series," *Annals of Statistics*, 46, 2094–2124. [783]

Diebold, F. X., and Li, C. (2006), "Forecasting the Term Structure of Government Bond Yields," *Journal of Econometrics*, 130, 337–364. [777]

Diebold, F. X., Piazzesi, M., and Rudebusch, G. (2005), "Modeling Bond Yields in Finance and Macroeconomics," Technical Report, National Bureau of Economic Research. [775]

Diebold, F. X., Rudebusch, G. D., and Aruoba, S. B. (2006), "The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach," *Journal of Econometrics*, 131, 309–338. [775,777]

Fan, J., and Li, R. (2001), "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties," *Journal of the American Statistical Association*, 96, 1348–1360. [776]

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000), "The Generalized Dynamic-Factor Model: Identification and Estimation," *Review of Economics and Statistics*, 82, 540–554. [775,776]

——— (2004), "The Generalized Dynamic Factor Model Consistency and Rates," *Journal of Econometrics*, 119, 231–255. [775,776,779]

Frank, L. E., and Friedman, J. H. (1993), "A Statistical View of Some Chemometrics Regression Tools," *Technometrics*, 35, 109–135. [776]

French, K. R. (2017), "100 Portfolios Formed on Size and Book-to-Markete," available at: *http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_100_port_sz.html*. [791]

Geyer, C. J. (1991), "Constrained Maximum Likelihood Exemplified by Isotonic Convex Logistic Regression," *Journal of the American Statistical Association*, 86, 717–724. [776]

Hallin, M., and Liška, R. (2007), "Determining the Number of Factors in the General Dynamic Factor Model," *Journal of the American Statistical Association*, 102, 603–617. [779]

——— (2011), "Dynamic Factors in the Presence of Blocks," *Journal of Econometrics*, 163, 29–41. [779,793]

Hathaway, R. J. (1985), "A Constrained Formulation of Maximum-Likelihood Estimation for Normal Mixture Distributions," *The Annals of Statistics*, 13, 795–800. [776]

Lam, C., and Yao, Q. (2012), "Factor Modeling for High-Dimensional Time Series: Inference for the Number of Factors," *The Annals of Statistics*, 40, 694–726. [775,776,779,780,787,793]

Lam, C., Yao, Q., and Bathia, N. (2011), "Estimation of Latent Factors for High-Dimensional Time Series," *Biometrika*, 98, 901–918. [775,776,782,783]

Liu, Y., Zhang, H. H., Park, C., and Ahn, J. (2007), "Support Vector Machines With Adaptive LQ Penalty," *Computational Statistics & Data Analysis*, 51, 6380–6394. [776]

Murtagh, F., and Legendre, P. (2014), "Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?," *Journal of Classification*, 31, 274–295. [787]

Nelson, C. R., and Siegel, A. F. (1987), "Parsimonious Modeling of Yield Curves," *Journal of Business*, 60, 473–489. [777]

Pan, J., and Yao, Q. (2008), "Modelling Multiple Time Series via Common Factors," *Biometrika*, 95, 365–379. [775]

Rudebusch, G. D., and Wu, T. (2008), "A Macro-Finance Model of the Term Structure, Monetary Policy and the Economy," *The Economic Journal*, 118, 906–926. [775]

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, Series B, 58, 267–288. [776]

Tsai, H., and Tsay, R. S. (2010), "Constrained Factor Models," *Journal of the American Statistical Association*, 105, 1593–1605. [776,780,793]

Tsai, H., Tsay, R. S., Lin, E. M., and Cheng, C.-W. (2016), "Doubly Constrained Factor Models With Applications," *Statistica Sinica*, 26, 1453–1478. [776]

Wang, D., Liu, X., and Chen, R. (2019), "Factor Models for Matrix-Valued High-Dimensional Time Series," *Journal of Econometrics*, 208, 231–248. [775,776,780,781,782,783,787,790,791]

Xu, R., and Wunsch, D. (2005), "Survey of Clustering Algorithms," *IEEE Transactions on Neural Networks*, 16, 645–678. [787]

Zhang, H. H., and Lu, W. (2007), "Adaptive Lasso for Cox's Proportional Hazards Model," *Biometrika*, 94, 691–703. [776]

Zou, H. (2006), "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101, 1418–1429. [776]