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

Elynn Y. Chen †, Ruey S. Tsay †, and Rong Chen ‡

1Department of Operations Research and Financial Engineering, Princeton University
2Booth School of Business, University of Chicago
3Department of Statistics and Biostatistics, Rutgers University

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 et al. (2018) proposed a matrix factor model that is shown to be able to provide effective analysis. In this paper, 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.

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 thirty variables spanning over one hundred years and over two hundred countries covering a variety of demographic, social, political, and economic topics. These data neatly form a matrix-variate time series with rows representing the countries and columns representing various macroeconomic 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 & Rothschild 1983, Bai 2003, Bai & Ng 2002, 2007, Forni et al. 2000, 2004, Pan & Yao 2008, Lam et al. 2011, Lam & Yao 2012, Chang et al. 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 et al. (2018) propose a matrix factor model that maintains and utilizes 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 et al. 2005, 2006, Rudebusch & Wu 2008, Bansal et al. 2014).

In many applications, relevant prior or domain knowledge is available or data themselves exhibit certain specific structure. Additional covariates may also 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 in order to effectively obtain the factors of interest.

To address these important issues and practical needs, we extend the matrix factor model of Wang et al. (2018) 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 & Friedman (1993), Tibshirani (1996), Liu et al. (2007), Fan & Li (2001), Zou (2006), and Zhang & 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 & Tsay (2010) and Tsai et al. (2016) impose constraints, constructed by some empirical procedures, that incorporate the inherent data structure, to both the classical and approximate factor models. Their results show that the constraints are useful tools to obtain parsimonious econometric models for forecasting, to simplify the interpretations of common factors, and to reduce the dimension. Motivated by similar concerns, we consider constrained, multi-term, and partially constrained factor models for high-dimensional matrix-variate time series. Our methods differs from Tsai & Tsay (2010) in several aspects. First, we deal with matrix factor model and thus have the flexibility to impose row and column constraints. The interaction between the row and column constraints are explored. Second, we adopt a different set of assumptions for factor model. The factor models in Tsai & Tsay (2010) and Tsai et al. (2016), following the definition in Bai (2003), Bai & 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 et al. (2011), Lam & Yao (2012), Chang et al. (2015), Wang et al. (2018) for more detailed comparisons between these two different model definitions.

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

2 The Constrained Matrix Factor Model

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

Let \( \{Y_t\}_{t=1}^T \) be a matrix-variate time series, where \( Y_t \) is a \( p_1 \times p_2 \) matrix, that is

\[
Y_t = (Y_{1,t}, \ldots, Y_{p_2,t}) = \begin{pmatrix}
Y_{11,t} & \cdots & y_{1p_2,t} \\
\vdots & \ddots & \vdots \\
y_{p_11,t} & \cdots & y_{p_1p_2,t}
\end{pmatrix}.
\]

Wang et al. (2018) propose the following factor model for \( Y_t \),

\[
Y_t = \Lambda F_t \Gamma' + U_t, \quad t = 1, 2, \ldots, T,
\]

where \( F_t \) is a \( k_1 \times k_2 \) latent matrix-variate time series of common fundamental factors, \( \Lambda \) is a \( p_1 \times k_1 \) row loading matrix, \( \Gamma \) is a \( p_2 \times k_2 \) column loading matrix, and \( U_t \) is a \( p_1 \times p_2 \) matrix of random errors. In Equation (1), \( (\Lambda, \Gamma) \) and \( (c\Lambda, \Gamma/c) \) are equivalent if \( c \neq 0 \).

In Model (1), we assume that \( vec(U_t) \sim WN(0, \Sigma_e) \) and is independent of the factor process \( vec(F_t) \). That is, \( \{U_t\}_{t=1}^T \) is a white noise matrix-variate time series and the common fundamental factors \( F_t \) drive all dynamics and co-movement of \( Y_t \). \( \Lambda \) and \( \Gamma \) reflect the importance of common factors and their interactions. Wang et al. (2018) provide several interpretations of the loading matrices \( \Lambda \) and \( \Gamma \). Essentially, \( \Lambda \) (\( \Gamma \)) can be viewed as the row (column) loading matrix that reflects how each row (column) in \( Y_t \) depends on the factor matrix \( F_t \). The interaction between the row and column is introduced through the multiplication of these terms.

The definition of common factors in Model (1) is similar to that of Lam et al. (2011). This decomposition facilitates model identification in finite samples and simplifies the procedure of model identification and statistical inference. However, under the definition, both the “common factors” defined in the traditional factor models and the serially correlated idiosyncratic components will be identified as factors. The method in Wang et al. (2018) 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 \( \hat{F}_t \) becomes difficult because of the uncertainty and dependence involved in the estimates of the loading matrices \( \Lambda \) and \( \Gamma \).
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

\[ Y_t = H_R R C' H_C + U_t, \tag{2} \]

where \( H_R \) and \( H_C \) are pre-specified full column-rank \( p_1 \times m_1 \) and \( p_2 \times m_2 \) constraint matrices, respectively, and \( R \) and \( C \) are \( m_1 \times k_1 \) row loading matrix and \( m_2 \times k_2 \) column loading matrix, respectively. For meaningful constraints, we assume \( k_1 \leq m_1 << p_1 \) and \( k_2 \leq m_2 << p_2 \). Compared with the matrix factor model in (1), we set \( \Lambda = H_R R \) and \( \Gamma = H_C C \) with \( H_R \) and \( H_C \) given. The number of parameters in the left loading matrix \( R \) is \( m_1 k_1 \), smaller than \( p_1 k_1 \) of the unconstrained model. The number of parameters in the column loading matrix \( C \) also decreases from \( p_2 k_2 \) to \( m_2 k_2 \). The constraint matrices \( H_R \) and \( H_C \) 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 8 companies, which can be grouped according to their industrial classification (Tech and Retail) and also their market capitalization (Large and Medium). The two groups form \( 2 \times 2 \) combinations as shown in Table 1.

| Industry | Market Cap |
|----------|------------|
| I1. Tech | C1. Large  |
|          | Apple, Microsoft |
| I2. Retail | C2. Medium |
|          | Broade, FireEye |
|          | Walmart, Target |
|          | JC Penny, Kohl’s |

Table 1: Groups of companies by industry and market capitalization

Constraint matrix \( H_R^{(1)} \) in Table 2 utilizes only industrial classification. To combine both industrial classification and market cap information, we first consider an additive model constraint on the \( 8 \times k_1 \) \( (k_1 \leq 3) \) loading matrix \( \Lambda \) in Model (1). The additive model constraint means that the \( i \)-th row of \( \Lambda \), that is, the loadings of \( k_1 \) row factors on the \( i \)-th variable, must assume the form

\[ \lambda_{i} = u_{j} + v_{l}, \]

where the \( i \)-th variable falls in group \( (Industry_j, MarketCap_l) \), \( k_1 \)-dimensional
vectors $\mathbf{u}_j$ and $\mathbf{v}_l$ are the loadings of $k_1$ row factors on the $j$-th market cap group and $l$-th industrial group, respectively. The most obvious way to express the additive model constraint is to use row constraints $\mathbf{H}_R^{(2)}$ in Table 2. Then, in the constrained matrix factor model (2), $\mathbf{H}_R = \mathbf{H}_R^{(2)}$ and $\mathbf{R} = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_1, \mathbf{v}_2)'$.

$$
\begin{array}{c}
\begin{array}{c}
\begin{array}{ccc}
1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0
\end{array}
\end{array}
\end{array}
$$

$$
\begin{array}{c}
\begin{array}{c}
\begin{array}{ccc}
1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0
\end{array}
\end{array}
\end{array}
$$

$$
\begin{array}{c}
\begin{array}{c}
\begin{array}{ccc}
1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0
\end{array}
\end{array}
\end{array}
$$

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

Further, we consider the constraint incorporating an interaction term between industry and market cap grouping information. Now the $i$-th row of $\mathbf{A}$ has the form $\lambda_i = \mathbf{u}_j + \mathbf{v}_l + \alpha_{ij} \mathbf{w}$, where $\mathbf{w}$ is the $k_1$-dimensional interaction vector containing loadings of $k_1$ row factors and $\alpha_{ij}$ is the interaction term determined by $\mathbf{u}_j$ and $\mathbf{v}_l$ jointly. For example,

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

In this case, for the constrained matrix factor model (2), $\mathbf{H}_R = \mathbf{H}_R^{(3)}$ and $\mathbf{R} = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_1, \mathbf{v}_2, \mathbf{w})'$. Note that $\mathbf{H}_R^{(2)}$ and $\mathbf{H}_R^{(3)}$ 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 $\mathbf{H}_R^{(2)}$ and $\mathbf{H}_R^{(3)}$ 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 & 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 $\tau$ denotes time to maturity. [Diebold & Li (2006)] and [Diebold et al. (2006)] interpret the Nelson-Siegel representation as a dynamic latent factor model where $\beta_1$, $\beta_2$, and $\beta_3$ 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 \( \lambda \) 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 \( \{Y_t\}_{t=1}^T \), where rows of \( Y_t \) represent time to maturity and columns of \( Y_t \) 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, \( H_R = [h_1, h_2, h_3] \), where \( h_1 = (1, 1, 1, 1)' \), \( h_2 = (1, 1, 0, -1, -1)' \) and \( h_3 = (-1, 0, 2, 0, -1) \). 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
\[ Y_t = H_{R_1} R_1 F_{1t} C_{1}' H_{C_1} + H_{R_2} R_2 F_{2t} C_{2}' H_{C_2} + U_t. \] (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 as natural resource based and manufacture based economies, and the corresponding factors may not interact with each other.

Note that (3) can be rewritten as (2), with \( H_R = \begin{bmatrix} H_{R_1} & H_{R_2} \end{bmatrix}, \ H_C = \begin{bmatrix} H_{C_1} & H_{C_2} \end{bmatrix}, \ R = \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix}, \ C = \begin{bmatrix} C_1 & 0 \\ 0 & C_2 \end{bmatrix} \) and \( F_t = \begin{bmatrix} F_{1t} & 0 \\ 0 & F_{2t} \end{bmatrix} \).

Hence (3) is a special case of (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 \( H_{R_1} \) and \( H_{R_2} \) do not have to be orthogonal. Neither does the pair \( H_{C_1} \) and \( H_{C_2} \). 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 \( H_{R_1} \) and \( H_{R_2} \). Since we can always transform
non-orthogonal constraint matrices to some orthogonal constraint matrices, we shall focus on the case when \( H_{R_1} \) and \( H_{R_2} \) (or \( H_{C_1} \) and \( H_{C_2} \)) 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

\[
Y_t = \left[ H_{R_1} R_1 \ A_2 \right] \left[ \begin{array}{c} F_{11,t} \\ F_{21,t} \end{array} \right] \left[ \begin{array}{c} C_{1} H'_{C_1} \\ C_{2} H'_{C_2} \end{array} \right] + U_t, \tag{4}
\]

where \( H_{R_1}, R_1, H_{C_1} \) and \( C_1 \) are defined similarly as those in (3). \( F_{ij,t} \)'s are common matrix factors corresponding to the interactions of the row and column loading space spanned by the columns of \( H_R \) and \( H_C \) and their complements, \( A_2 \) is \( p_1 \times q_1 \) row loading matrix and \( \Gamma_2 \) is a \( p_2 \times q_2 \) column loading matrix. Again, we have \( q_1 < p_1, q_2 < p_2 \) and \( \text{vec}(F_{ij,t}) \)'s are independent with \( \text{vec}(U_t) \). We assume that \( H'_{R_1} A_2 = 0 \) and \( H'_{C_1} \Gamma_2 = 0 \), because all the row loadings that are in the space of \( H_{R_1} \) and all the column loadings that are in the space of \( H_{C_1} \) could be absorbed into the first parts of loading matrices. Thus, we could explicitly rewrite the model as

\[
Y_t = \left[ H_{R_1} R_1 \ H_{R_2} R_2 \right] \left[ \begin{array}{c} F_{11,t} \\ F_{21,t} \end{array} \right] \left[ \begin{array}{c} C_{1} H'_{C_1} \\ C_{2} H'_{C_2} \end{array} \right] + U_t, \tag{5}
\]

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

In the special case when \( F_{21,t} = 0 \) and \( F_{12,t} = 0 \), Model (4) can be further simplified as

\[
Y_t = H_{R_1} R_1 F_{11,t} C_{1} H'_{C_1} + H_{R_2} R_2 F_{22,t} C_{2} H'_{C_2} + U_t. \tag{6}
\]

Model (6) is different from the multi-term model of (3) in that the matrix \( H_{R_2} \) in (5) is induced from \( H_{R_1} \) while the \( H_{R_2} \) in (3) is an informative constraint, with a lower dimension.

In the special case when \( H_{C_1} = I_{p_1} \) (there is no column constraint), Model (5) becomes

\[
Y_t = \left[ H_{R_1} R_1 \ H_{R_2} R_2 \right] \left[ \begin{array}{c} F_{11,t} \\ F_{21,t} \end{array} \right] C' + U_t,
\]

where \( F_{1,t} = [F_{11,t}, F_{12,t}] \) and \( F_{2,t} = [F_{21,t}, F_{22,t}] \). The left loading matrix still spans the entire \( p_1 \) dimensional space, but the first part of loading matrix \( R_1 \) has a clearer interpretation.

The partially constrained matrix factor model (5) incorporates partial information \( H_{R_1} \) and \( H_{C_1} \) 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 \( H_{R_1} \) and \( H_{C_1} \) and their
complements, the number of parameters in (5) is the same as that in the unconstrained model (1). However, as shown by 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) and Hallin & Liška (2007), Hallin & Liška (2011) considered $n$-dimensional random variable $x = [y' z']'$ with sub-panel vectors $y \in \mathbb{R}^{n_y}$ and $z \in \mathbb{R}^{n_z}$ 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 paper. Both approaches produce certain block structures in the loading matrix. Consider the vector factor model case. With two subpanels, the model becomes

$$
\begin{bmatrix}
  y \\
  z
\end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & 0 \\
  A_{21} & 0 & A_{23}
\end{bmatrix} \begin{bmatrix}
  f_1 \\
  f_2 \\
  f_3
\end{bmatrix} + \begin{bmatrix}
  \epsilon_y \\
  \epsilon_z
\end{bmatrix}.
$$

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

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

where $I$'s and $0$'s are identity and zero matrices of proper dimensions. However, our current estimation procedure is not able to force certain submatrices in $R$ 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 two-folds. Firstly, the model is capable of identifying, from the complement spaces of $H_R$ and $H_C$, the factors that are unknown to researchers. In this case, the dimensions of $F_{22,t}$ are typically much smaller than those of $F_{11,t}$ even though the loading matrices $R_2$ and $C_2$ 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 & 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, $F_{11,t}$ represents the factor
matrix with row and column factors affecting the observed matrix-variate time series in the way as specified by the constraints \( H_R \) and \( H_C \) completely. Consider the multinational macroeconomic index example. If \( H_R \) is built from the country classification information, how the rows in \( F_{11,t} \) affect the observations can be completely explained by the country groups instead of individual countries and the row factors in \( F_{11,t} \) 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, \( H_R = [h_1, h_2, h_3] \), where \( h_1 = (1, 1, 1, 1)' \), \( h_2 = (1, 1, 0, -1, -1)' \), and \( h_3 = (-1, 0, 2, 0, -1) \). Then the estimation method in Section 3 is capable of isolating \( H_{R1}R_{11}F_{11,t}C_{11}'H_{C1} \), 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 \( O_1 \) and \( O_2 \) be two invertible matrices of size \( k_1 \times k_1 \) and \( k_2 \times k_2 \). Then the triples \((R, F_t, C)\) and \((RO_1, O^{-1}_1F_t, O^{-1}_2, O_2C)\) are equivalent under Model (2). Here, we may assume that the columns of \( R \) and \( C \) are orthonormal, that is, \( R'R = I_{k_1} \) and \( C'C = I_{k_2} \), where \( I_d \) denotes the \( d \times d \) identity matrix. Even with these constraints, \( R, F_t \) and \( C \) are not uniquely determined in (2), as aforementioned replacement is still valid for any orthonormal \( O \). However, the column spaces of the loading matrices \( R \) and \( C \) are uniquely determined. Hence, in the following sections, we focus on the estimation of the column spaces of \( R \) and \( C \). We denote the row and column factor loading spaces by \( M(R) \) and \( M(C) \), respectively. For simplicity, we suppress the matrix column space notation and use the matrix notation directly.

3.1 Orthogonal Constraints

We start with the estimation of the constrained matrix-variate factor model (2). The approach follows the ideas of Tsai & Tsay (2010) and Wang et al. (2018). In what follows, we illustrate the estimation procedure for the column space of \( R \). The column space of \( C \) can be obtained similarly from the transpose of \( Y_t \)’s. For ease in representation, we assume that the process \( F_t \) has mean \( 0 \), and the observation \( Y_t \)’s are centered and standardized throughout the paper.

Suppose we have orthogonal constraints \( H_{R}'H_{R} = I_{m_1} \) and \( H_{C}'H_{C} = I_{m_2} \). Define the trans-
formation $X_t = H^t_R Y_t H_C$. It follows from (2) that

$$X_t = RF_t C' + E_t, \quad t = 1, 2, \ldots, T,$$

(7)

where $E_t = H^t_R U_t H_C$.

This transformation projects the observed matrix time series into the constrained space. For example, if $H_R$ is the orthonormal matrix corresponding to the group constraint $H^{(1)}_R$ of Table 2, then $H^t_R Y_t$ is a $2 \times p_2$ matrix, with the first row being the normalized average of the rows of $Y_t$ 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 $E_t$ remains to be a white noise process, the estimation method in Wang et al. (2018) directly applies to the transformed $m_1 \times m_2$ matrix time series $X_t$ in Model (7). For completeness, we outline briefly the procedure. See Wang et al. (2018) for details.

To facilitate the estimation, we use the QR decomposition $R = Q_1 W_1$ and $C = Q_2 W_2$. The estimation of column spaces of $R$ and $C$ is equivalent to the estimation of column spaces of $Q_1$ and $Q_2$. Thus, Model (7) can be re-expressed as

$$X_t = RF_t C' + E_t = Q_1 Z_t Q_2' + E_t, \quad t = 1, 2, \ldots, T,$$

(8)

where $Z_t = W_1 F_t W'_2$, $Q'_1 Q_1 = I_{m_1}$, and $Q'_2 Q_2 = I_{m_2}$.

Let $h$ be a positive integer. For $i, j = 1, 2, \ldots, m_2$, define

$$\Omega_{qijk}(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} \text{Cov}(Z_t Q_2_{2,i}, Z_t+h, Q_2_{2,j}),$$

(9)

and

$$\Omega_{xij}(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} \text{Cov}(X_t, X_{t+h, j}),$$

(10)

which can be interpreted as the auto-cross-covariance matrices at lag $h$ between column $i$ and column $j$ of $\{Z_t Q_2'\}^T_{t=1}$ and $\{X_t\}^T_{t=1}$, respectively. For $h > 0$, $\Omega_{xij}(h)$ defined in (10) does not involve the covariance terms incurred by $E_t$ because of the whiteness condition.

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

$$M = \sum_{h=1}^{h_0} \sum_{i=1}^{m_2} \sum_{j=1}^{m_2} \Omega_{xij}(h) \Omega_{xij}(h)' = Q_1 \left\{ \sum_{h=1}^{h_0} \sum_{i=1}^{m_2} \sum_{j=1}^{m_2} \Omega_{qijk}(h) \Omega_{qijk}(h)' \right\} Q_1'.$$

(11)

Since $M$ and the matrix sandwiched by $Q_1$ and $Q_1'$ are positive definite matrices, Equation (11) implies that the eigen-space of $M$ is the same as the column space of $Q_1$ if the middle term is full rank (Condition 2 in Appendix A. Hence, $M(Q_1)$ can be estimated by the space spanned by the eigenvectors of the sample version of $M$. The normalized eigenvectors $q_1, \ldots, q_{k_1}$ corresponding to the $k_1$ nonzero eigenvalues of $M$ are uniquely defined up to a sign change. Thus
\( Q_1 \) is uniquely defined by \( \hat{Q}_1 = (q_1, \ldots, q_{k_1}) \) up to a sign change. We estimate \( \hat{Q}_1 = (\hat{q}_1, \ldots, \hat{q}_{k_1}) \) as a representative of \( M(Q_1) \) or \( M(R) \).

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 \( M \) in (11) as the following

\[
\hat{M} = \sum_{h=1}^{h_0} \sum_{m_2=1}^{h_0} \sum_{i=1}^{m_2} \sum_{j=1}^{m_2} \hat{\Omega}_{x,ij}(h)\hat{\Omega}_{x,ij}(h)', \quad \hat{\Omega}_{x,ij}(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} X_{t,i}X_{t+h,j}.' (12)
\]

Then, \( M(Q_1) \) can be estimated by \( M(\hat{Q}_1) \), where \( \hat{Q}_1 = (\hat{q}_1, \ldots, \hat{q}_{k_1}) \) and \( \hat{q}_i \) is an eigenvector of \( \hat{M} \), corresponding to its \( i \)-th largest eigenvalue. The \( Q_2 \) is defined similarly for the column loading matrix \( C \) and \( M(\hat{Q}_2) \) and \( \hat{Q}_2 \) can be estimated with the same procedure to the transpose of \( X_t \). Consequently, we estimate the normalized factors and residuals, respectively, by \( \hat{Z}_t = \hat{Q}_1'X_t\hat{Q}_2 \) and \( \hat{U}_t = Y_t - H_R\hat{Q}_1\hat{Z}_t\hat{Q}_2' H_C' \).

The above estimation procedure assumes that the number of row factors \( k_1 \) is known. To determine \( k_1 \), Wang et al. (2018) used the eigenvalue ratio-based estimator of Lam & Yao (2012). Let \( \hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_{m_1} \geq 0 \) be the ordered eigenvalues of \( \hat{M} \). The ratio-based estimator for \( k_1 \) is defined as

\[
\hat{k}_1 = \arg \min_{1 \leq j \leq K} \left( \frac{\hat{\lambda}_{j+1}}{\hat{\lambda}_j} \right),
\]

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 \( \hat{M} \)-matrix corresponding to the transpose of \( X_t \).

Although the estimation procedure on the transformed series \( X_t \) is exactly the same as that of Wang et al. (2018), the asymptotic properties of the estimator are different due to the transformation, as shown in Section 3.1 and \( X_t \) is of lower dimension.

### 3.2 Nonorthogonal Constraints

If the constraint matrix \( H_R \) (or \( H_C \)) is not orthogonal, we can perform column orthogonalization and standardization, similar to that in Tsai & Tsay (2010). Specifically, we obtain

\[
H_R = \Theta_R K_R,
\]

where \( \Theta_R \) is an orthonormal matrix and \( K_R \) is a \( m_1 \times m_1 \) upper triangular matrix with nonzero diagonal elements. \( H_C = \Theta C K_C \) can be obtained in the same way.

Letting \( X_t = \Theta_R' Y_t \Theta_C, \quad R^* = K_R R, \) and \( C^* = K_C C \), we have

\[
X_t = R^* F_t C^* + E_t, \quad t = 1, 2, \ldots, T,
\]

where \( E_t = \Theta_R' U_t \Theta_C \). Since \( E_t \) remains a white noise process, we apply the same estimation method as that in Section 3.1 to obtain \( \hat{Q}_1^* \) and \( \hat{Q}_2^* \) as the representatives of \( M(\hat{R}^*) \) and \( M(\hat{C}^*) \). Then the estimators of \( R \) and \( C \) are \( \hat{R} = K_R^{-1} \hat{Q}_1^* \) and \( \hat{C} = K_C^{-1} \hat{Q}_2^* \). Note that \( K_R \) and \( K_C \) 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 $H_{R_1}$ and $H_{R_2}$ (or $H_{C_1}$ and $H_{C_2}$) are orthogonal, we obtain, for $t = 1, 2, \ldots, T,$

$$H'_{R_1} Y_t H_{C_1} = R_1 F_{1,t} C_1' + H'_{R_1} U_t H_{C_1},$$

$$H'_{R_2} Y_t H_{C_2} = R_2 F_{2,t} C_2' + H'_{R_2} U_t H_{C_2},$$

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

**Remark 3.** For multi-term constrained model (3), $H_{R_1}$ and $H_{R_2}$ (or $H_{C_1}$ and $H_{C_2}$) may not necessarily be orthogonal. In this case, we illustrate the estimation procedure for the column loadings. Define projection matrices $P_{H_{R_1}^\perp} = I - H_{R_1} H_{R_1}'$ and $P_{H_{R_2}^\perp} = I - H_{R_2} H_{R_2}'$, which represent the projections onto the spaces perpendicular to the column spaces of $H_{R_1}$ and $H_{R_2}$, respectively. Left multiplying Equation (3) by $P_{H_{R_2}^\perp}$, respectively, and taking transpose of the resulting matrices, we have

$$Y_t P_{H_{R_2}^\perp} = H_{C_1} C_1 F_{1,t} R_1' H_{R_1}' P_{H_{R_2}^\perp} + U_t P_{H_{R_2}^\perp},$$

$$Y_t P_{H_{R_1}^\perp} = H_{C_2} C_2 F_{2,t} R_2' H_{R_2}' P_{H_{R_1}^\perp} + U_t P_{H_{R_1}^\perp},$$

where $P_{H_{R_2}^\perp} U_t$ and $P_{H_{R_1}^\perp} U_t$ are white noises. The column loading estimators $\hat{C}_1$ and $\hat{C}_2$ can be obtained by applying the procedure described in Section 3.1 to $H'_{C_1} Y_t P_{H_{R_2}^\perp}$ and $H'_{C_2} Y_t P_{H_{R_1}^\perp}$, respectively. Note that the $p_1 \times m_1$ matrix $P_{H_{R_2}^\perp} H_{R_1}$ 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 $\hat{C}_1$ will depend on $\|P_{H_{R_2}^\perp} H_{R_1} R_1\|_2^2$ instead of $\|H_{R_1} R_1\|_2^2$.

3.4 Partially Constrained Matrix Factor Model

For the partially constrained matrix factor model (5), we assume that $H'_{R_1} H_{R_2} = 0$ and $H'_{C_1} H_{C_2} = 0$. Define the transformation $X_{t}^{(l)} = H'_{R_1} Y_t H_{C_k}$ for $l, k = 1, 2$. Then the transformed data follow the structure,

$$X_{t}^{(l)} = R_l F_{l k,t} C_k' + E_{t}^{(l)}, \quad l, k = 1, 2,$$

where $E_{t}^{(l)} = H'_{R_1} U_t H_{C_k}$ remains a white noise process.
Let $M^{(lk)}$ represent the $M$ matrix defined in (11) for each $X^{(lk)}_t$, $l, k = 1, 2$. Define $M^{(l)} = \sum_{k=1}^{2} M^{(lk)}$ for $l = 1, 2$, then
\[
M^{(l)} = Q_1^{(l)} \left\{ \sum_{k=1}^{2} \sum_{h=1}^{m_2} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \Omega^{(lk)}_{zi,j}(h) \Omega^{(lk)}_{zi,j}(h)' \right\} Q_1^{(l)}, \quad l = 1, 2,
\]
has the same column space as that of $R_l$, for $l = 1, 2$, respectively.

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

This procedure effectively projects the observed matrix time series $Y_t$ into four orthogonal subspaces, based on the constraints obtained from the domain knowledge or some empirical procedure. Because $X^{(lk)}_t$, $l, k = 1, 2$ 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 $X^{(11)}_t$ and $X^{(12)}_t$ 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 $F_{lk,t}, l, k = 1, 2$ 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 $F_{11,t}$ is no weaker than that of the remainder factor $F_{22,t}$. Improved estimators of $\hat{R}_l$, $l = 1, 2$, can be obtained by applying eigen-decomposition on the sample version of $M^{(l1)}$ defined similarly to (12). Improved estimators of $\hat{C}_k$, $k = 1, 2$, can be obtained by using the same procedure on the transposes of $X^{(lk)}_t$ 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 $\|A\|_2, \|A\|_F$ and $\|A\|_{\min}$ denote the spectral norm, Frobenius norm, and the smallest nonzero singular value of $A$, respectively. When $A$ is a square matrix, we denote by $tr(A)$, $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ the trace, maximum and minimum eigenvalues of the matrix $A$, 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 et al. (2018) 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 \( \mathbb{H} \). Asymptotic properties of nonorthogonal, multi-term, and partially constrained matrix factor model are trivial extensions.

Several regularity conditions (Conditions 1 to 5) are listed in the Appendix. They are similar to those in Wang et al. (2018) and are used to derive the limiting behavior of \( \mathbb{H} \) towards its population version. The following condition requires some discussion.

**Condition 6. Factor Strength.** There exist constants \( \delta_1 \) and \( \delta_2 \) in \([0, 1]\) such that \( \| \mathbb{H} \mathbb{R} \|^2_2 \asymp p_1^{1-\delta_1} \times \| \mathbb{H} \mathbb{C} \|^2_2 \asymp p_2^{1-\delta_2} \asymp \| \mathbb{H} \mathbb{C} \|^2_{\min} \).

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

The rates \( \delta_1 \) and \( \delta_2 \) 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 \( Y_t \) 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 \( Y_t \) about the factors grows more slowly than the noises introduced as \( p_1 \) increases. The smaller the \( \delta_1 \)’s, the stronger the factors. In the strong factor case, the loading matrix is dense. See Lam et al. (2011) for further discussions.

If we restrict \( \mathbb{H} \mathbb{R} \) to be orthonormal, \( \| \mathbb{H} \mathbb{R} \|^2_2 = \| \mathbb{R} \|^2_2 \asymp p_1^{1-\delta_1} \) and there is an interplay between \( \mathbb{H} \mathbb{R} \) and \( \mathbb{R} \) as \( p_1 \) increases. In order for \( \mathbb{H} \mathbb{R} \) to remain orthonormal, when \( p_1 \) increases, each element of \( \mathbb{H} \mathbb{R} \) decreases at the rate of \( p_1^{-1/2} \). At the same time, each element of \( \mathbb{R} \) on average increases at the rate of \( \sqrt{p_1^{1-\delta_1}/m_1} \). The column factor loading \( \| \mathbb{H} \mathbb{C} \|^2_2 \) behaves in the same way.

As \( p_1 \) and \( p_2 \) increase, each element of the transformed error \( \mathbb{E}_t \) remains a growth rate of 1 under Condition 3 (see Lemma 1 in Appendix A, but the dimension of \( \mathbb{E}_t \) 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 \( \mathbb{H} \mathbb{R} \) and \( \mathbb{H} \mathbb{C} \), but clearly how \( m_1 \) and \( m_2 \) increase with \( p_1, p_2 \) is also important because it controls the signal-noise ratio in the constrained model.

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 \( m_1 p_1^{1+\delta_1} m_2 p_2^{1+\delta_2} T^{-1/2} = o(1) \), as \( m_1, p_1, m_2, p_2, \) and
$T$ go to $\infty$, it holds that

$$
||\hat{Q}_1 - Q_1||_2 = O_p \left( \max \left( T^{-1/2}, \frac{m_1}{p_1^{1-\delta_1}}, \frac{m_2}{p_2^{1-\delta_2}} T^{-1/2} \right) \right),
$$
$$
||\hat{Q}_2 - Q_2||_2 = O_p \left( \max \left( T^{-1/2}, \frac{m_1}{p_1^{1-\delta_1}}, \frac{m_2}{p_2^{1-\delta_2}} T^{-1/2} \right) \right).
$$

**Remark 5.** The convergence rate for the unconstrained model is $p_1^{1-\delta_1} p_2^{1-\delta_2} T^{-1/2}$ in [Wang et al. (2018)](#). The rates for the constrained model under different relations between $m_1 m_2$ and $p_1 p_2$ are shown in Table 3.

| Rate | $m_1 m_2 \asymp p_1 p_2$ | $p_1^{1-\delta_1} p_2^{1-\delta_2} = o(m_1 m_2)$ | $m_1 m_2 = o(1)$ |
|------|-----------------|---------------------------------|-----------------|
| Rate | $p_1^{\delta_1} p_2^{\delta_2} T^{-1/2}$ | $m_1 m_2 p_1^{1-\delta_1} p_2^{1-\delta_2} T^{-1/2}$ | $T^{-1/2}$ |

Table 3: Convergence rate of the loading space estimator.

For strong factors with $\delta_1 = \delta_2 = 0$, the convergence rates are the same for the constrained and unconstrained models. However, $m_1 p_1^{1+\delta_1} m_2 p_2^{1+\delta_2} T^{-1/2} = o(1)$ 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 $m_1 m_2$ 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 $\delta_1$ and $\delta_2$ to $\delta_1 - (1 - \alpha_1)$ and $\delta_2 - (1 - \alpha_2)$, respectively. Hence, the constraints are particularly useful for weak strength cases.

When $m_1 m_2 = O(p_1^{1-\delta_1} p_2^{1-\delta_2})$, 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 (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 $m_1 m_2 \asymp p_1 p_2$, the convergence rates of the constrained and unconstrained models are the same. A special case is when $m_1 = c_1 p_1$ and $m_2 = c_2 p_2$, that is, the dimensions of the constrained loading spaces increase with $p$’s linearly.
Remark 6. Under some conditions the convergence rates in Theorem 1 may improve significantly. For example, if $\Sigma_u \equiv \text{Var}(\text{vec}(U_t))$ 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. $H^{(1)}_R$ in Section 2.1), then each element in $E_t$ is a group average. $\text{Var}(E_{t,ij})$ is smaller by a factor of $m_1 m_2 p_1 p_2$ and goes to zero when $m_1 m_2 p_1 p_2 = o(1)$.

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 et al. (2008) in (11). This will lead to faster convergence rates. See Section 3.2 of Chang et al. (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 $M$ matrix in (11) has $k_1$ distinct positive eigenvalues, then the eigenvalues $\{\hat{\lambda}_1, \ldots, \hat{\lambda}_{m_1}\}$ of $\hat{M}$, sorted in the descending order, satisfy

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

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

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

Theorem 2 shows that the estimators of the nonzero eigenvalues of $M$ 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 $M$ 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 $\Delta_{\mu T}^{\lambda} \equiv p_1^{2-\delta_1} p_2^{2-\delta_2} T^{-1/2}$ for the non-zero eigenvalues and $p_1^{\delta_1} p_2^{\delta_2} T^{-1/2} \cdot \Delta_{\mu T}^{\lambda}$ for the zero eigenvalues, respectively. See Wang et al. (2018). The rates for the constrained model under different relations between $m_1 m_2$ and $p_1 p_2$ are shown in Table 4.
\begin{table}
\centering
\begin{tabular}{|c|c|c|c|}
\hline
$\lambda_j$ & $m_1m_2 \asymp p_1p_2$ & $p_1^{1-\delta_1}p_2^{1-\delta_2} = o(m_1m_2)$ & $m_1m_2 = o(p_1^{1-\delta_1}p_2^{1-\delta_2})$ \\
\hline
Zero & $p_1^{\delta_1}p_2^{\delta_2}T^{-1/2} \cdot \Delta^\lambda_{pT}$ & $(m_1m_2)^2p_1^{\delta_1}p_2^{\delta_2}T^{-1/2} \cdot \Delta^\lambda_{pT}$ & $p_1^{\delta_1}p_2^{\delta_2}T^{-1/2} \cdot \Delta^\lambda_{pT}$ \\
\hline
Non-zero & $\Delta^\lambda_{pT}$ & $m_1m_2p_1^{-1}p_2^{-1} \cdot \Delta^\lambda_{pT}$ & $p_1^{-\delta_1}p_2^{-\delta_2} \cdot \Delta^\lambda_{pT}$ \\
\hline
Ratio & $p_1^{\delta_1}p_2^{\delta_2}T^{-1/2}$ & $m_1m_2p_1^{-1+\delta_1}p_2^{-1+\delta_2}T^{-1/2}$ & $T^{-1/2}$ \\
\hline
\end{tabular}
\caption{Convergence rate of estimators for non-zero and zero eigenvalues of $M$.}
\end{table}

In the cases of strong factors or weak factors with $m_1m_2 \asymp p_1p_2$, our result is the same as that of Wang et al. (2018). In all other cases, the gap between the convergence rates of nonzero and zero eigenvalues of $M$ is larger in the constrained case.

Let $S_t$ be the dynamic signal part of $Y_t$, i.e. $S_t = H_RRF_tC'H_C = H_RQ_1Z_tQ_2'H_C$. From the discussion in Section 3.1, $S_t$ can be estimated by

$$\hat{S}_t = H_R\hat{Q}_1\hat{Z}_t\hat{Q}_2'H_C.$$ 

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

**Theorem 3.** Under Conditions 1-6 and $m_1p_1^{-1+\delta_1}m_2p_2^{-1+\delta_2}T^{-1/2} = o(1)$, we have

$$\frac{1}{\sqrt{p_1p_2}}\|\hat{S}_t - S_t\|_2 = O_p\left(\max\left(p_1^{-\delta_1/2}p_2^{-\delta_2/2}, m_1p_1^{-1+\delta_1/2}m_2p_2^{-1+\delta_2/2}\right) \cdot \frac{1}{\sqrt{T}} + \frac{1}{\sqrt{p_1p_2}}\right),$$

$$= \begin{cases} 
O_p\left(p_1^{-\delta_1/2}p_2^{-\delta_2/2}T^{-1/2} + p_1^{-1/2}p_2^{-1/2}\right), & \text{if } m_1m_2 = O_p(p_1^{1-\delta_1}p_2^{1-\delta_2}), \\
O_p\left(m_1p_1^{-1+\delta_1/2}m_2p_2^{-1+\delta_2/2}T^{-1/2} + p_1^{-1/2}p_2^{-1/2}\right), & \text{otherwise}.
\end{cases}$$

Theorem 3 shows that as long as $m_1m_2 = o(p_1p_2)$ 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 et al. (2018). When $m_1m_2 = O_p(p_1^{1-\delta_1}p_2^{1-\delta_2})$, 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 $U_t$. Using Gaussian noises shows similar results.
As noted in Section 3, the row and column factor loading matrices $\Lambda = H_R R$ and $\Gamma = H_C C$ are only identifiable up to a linear space spanned by its columns. Following Lam et al. (2011) and Wang et al. (2018), we adopt the discrepancy measure used by Chang et al. (2015): for two orthogonal matrices $O_1$ and $O_2$ of size $p \times q_1$ and $p \times q_2$, then the difference between the two linear spaces $M(O_1)$ and $M(O_2)$ is measured by

$$
D(M(O_1), M(O_2)) = \left(1 - \frac{1}{\max(q_1, q_2)} tr(O_1 O_1^T O_2 O_2^T)\right)^{1/2}.
$$

(15)

Clearly, $D(M(O_1), M(O_2))$ assumes values in $[0,1]$. It equals to 0 if and only if $M(O_1) = M(O_2)$ and equals to 1 if and only if $M(O_1) \perp M(O_2)$. If $O_1$ and $O_2$ are vectors, (15) is the cosine similarity measure. We report this space distance $D(\cdot, \cdot)$ as a measurement of the discrepancy between estimated and true loading spaces.

5.1 Case 1. Orthogonal Constraints

In this case, the observed data $Y_t$’s are generated according to Model (2),

$$
Y_t = H_R R F_t C' H_C' + U_t, \quad t = 1, \ldots, T,
$$

under the following design.

The latent factor process $F_t$ is of dimension $k_1 \times k_2 = 3 \times 2$. The entries of $F_t$ follow $k_1 k_2$ independent AR(1) processes with Gaussian white noise $N(0,1)$ innovations. Specifically, $vec(F_t) = \Phi_F vec(F_{t-1}) + \epsilon_t$ with $\Phi_F = diag(-0.5, 0.6, 0.8, -0.4, 0.7, 0.3)$. The dimensions of the constrained row and column loading spaces are $m_1 = 12$ and $m_2 = 3$, respectively. Hence, $R$ is $12 \times 3$ and $C$ is $3 \times 2$. The entries of $R$ and $C$ are independently sampled from the uniform distribution $U(-p_{i-1}^{-\beta_i/2} \sqrt{m_i/p_i}, p_i^{-\beta_i/2} \sqrt{m_i/p_i})$ for $i = 1, 2$, respectively, so that the condition on the factor strength is satisfied. The disturbance $U_t = \Psi^{1/2} \Xi_t$ is a white noise process, where the elements of $\Xi_t$ are independent random variables of Student-$t$ distribution with five degrees of freedom and the matrix $\Psi^{1/2}$ is chosen so that $U_t$ has a Kronecker product covariance structure $cov(vec(U_t)) = \Gamma_2 \otimes \Gamma_1$, where $\Gamma_1$ and $\Gamma_2$ are of size $p_1 \times p_1$ and $p_2 \times p_2$ respectively. For $\Gamma_1$ and $\Gamma_2$, 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 $(\delta_1, \delta_2)$ among $(0,0), (0.5,0), (0.5,0.5)$. For each pair of $\delta_i$’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 0.5$p_1 p_2$, $p_1 p_2$, $1.5p_1 p_2$ and $2p_1 p_2$. For each combination of the parameters, we use 500 realizations. And we use $h_0 = 1$ for all simulations. The estimation error of $M(\hat{Q}_i)$ is defined as $D(\hat{Q}_i, Q_i)$, where the distance $D$ is defined in (15).

The row constraint matrix $H_R$ is a $p_1 \times 12$ orthogonal matrix. For $p_1 = 20$, $H_R$ is assumed to be a block diagonal matrix $I_4 \otimes D$, where $I_k$ is the identify matrix of dimension $k$ and $D = [d_1, d_2, d_3]$
is a $5 \times 3$ matrix with $d'_1 = (1, 1, 1, 1, 1)/\sqrt{5}$, $d'_2 = (-1, -1, 0, 1, 1)/2$, $d'_3 = (-1, 0, 2, 0, -1)/\sqrt{6}$. These three $d'_j$ 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 $D$, using $d'_1 = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1)/\sqrt{10}$, $d'_2 = (-1, -1, -1, -1, 0, 0, 1, 1, 1, 1)/\sqrt{8}$ and $d'_3 = (-1, -1, 0, 2, 2, 0, 0, -1, -1)/\sqrt{12}$.

The column constraint matrix $H_C$ is a $p_2 \times 3$ orthogonal matrix. For $p_2 = 20$, the three columns of $H_C$ are generated as $h_{c,1} = [1/\sqrt{7}, 0_7, 0_6]'$, $h_{c,2} = [0_7, 1/\sqrt{7}, 0_6]'$, $h_{c,3} = [0_7, 0_7, 1_6/\sqrt{6}]'$, where $0_k$ 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, i.e. $(\delta_1, \delta_2) = (0, 0)$, 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, i.e. $(\delta_1, \delta_2) = (0.5, 0)$, 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, i.e. $(\delta_1, \delta_2) = (0.5, 0.5)$, 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 box-plots of the estimation errors in estimating the loading spaces of $Q = Q_2 \otimes Q_1$ using the correct number of factors. The gray boxes are for the constrained models. From the plots, it is seen that when both row and column factors are strong, i.e. $(\delta_1, \delta_2) = (0, 0)$, and the number of factors is properly estimated, the mean and standard deviation of the estimation errors $D(\hat{Q}, Q)$ are small for both models, but the constrained model has a smaller mean estimation error. When row factors are weak, i.e. $(\delta_1, \delta_2) = (0.5, 0)$, and the 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 $D(\hat{Q}_i, Q_i)$ 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
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. Table on top is for $k_1$, while table on the bottom is for $k_2$. 

| $k_1$ | $T = 0.5 \ p_1 \ p_2$ | $T = p_1 \ p_2$ | $T = 1.5 \ p_1 \ p_2$ | $T = 2 \ p_1 \ p_2$ |
|-------|-------------------|----------------|----------------|----------------|
| $\delta_1$ | $\delta_2$ | $p_1$ | $p_2$ | $f_u$ | $f_c$ | $f_u$ | $f_c$ | $f_u$ | $f_c$ | $f_u$ | $f_c$ |
| 0 | 0 | 20 | 20 | 0.264 | 0.942 | 0.728 | 0.996 | 0.952 | 1 | 0.996 | 1 |
| 0.5 | 0 | 20 | 40 | 0.734 | 1 | 0.998 | 1 | 1 | 1 | 1 | 1 |
| 0 | 0 | 40 | 20 | 0.786 | 0.994 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0.5 | 0.5 | 40 | 40 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| $k_2$ | $T = 0.5 \ p_1 \ p_2$ | $T = p_1 \ p_2$ | $T = 1.5 \ p_1 \ p_2$ | $T = 2 \ p_1 \ p_2$ |
|-------|-------------------|----------------|----------------|----------------|
| $\delta_1$ | $\delta_2$ | $p_1$ | $p_2$ | $f_u$ | $f_c$ | $f_u$ | $f_c$ | $f_u$ | $f_c$ | $f_u$ | $f_c$ |
| 0 | 0 | 20 | 20 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0.5 | 0 | 20 | 40 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0 | 0 | 40 | 20 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0.5 | 0.5 | 40 | 40 | 0.988 | 0.996 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0 | 0 | 20 | 20 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0.5 | 0 | 20 | 40 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0 | 0 | 40 | 20 | 0.02 | 0.762 | 0.104 | 0.962 | 0.272 | 0.992 | 0.58 | 1 |
| 0.5 | 0 | 20 | 40 | 0 | 0.956 | 0.09 | 0.998 | 0.472 | 1 | 0.85 | 1 |
| 0 | 0 | 40 | 20 | 0 | 0.952 | 0.026 | 1 | 0.196 | 1 | 0.572 | 1 |
| 0.5 | 0.5 | 40 | 40 | 0 | 1 | 0.03 | 1 | 0.438 | 1 | 0.906 | 1 |
(m_2) is smaller than that of the constrained row space (m_1). Intuitively, after transformation (7), the ratio of the effective column factor strength ∥C∥_2^2 and noise level ∥E_t∥_2^2 is larger than the ratio of the effective row factor strength ∥R∥_2^2 and noise level ∥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.

![Figure 1: Box-plots of the estimation accuracy measured by D(\hat{Q}, 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 for plotted values.](image)

To investigate the performance of estimation under different choices of \( h_0 \), which is the number of lags used in (11), we change the underlying generating model of \( \text{vec}(F_t) \) to a VAR(2) process without the lag-1 term, \( \text{vec}(F_t) = \Phi_F \text{vec}(F_{t-2}) + \epsilon_t \). 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 \( \text{vec}(F_t) \), and hence \( \text{vec}(Y_t) \), has zero auto-covariance matrix at lag 1, \( \hat{\mathbf{M}} \) under \( h_0 = 1 \) contains 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 auto-covariance matrices of the data. See Condition 2 in Appendix A. In practice, one can select \( h_0 \) by examining the sample cross-correlation matrices of \( Y_t \).
In this form, the true number of factors is $k = \delta_H$. The case of orthogonal constraints is used. The results are based on 500 simulations.

### Table 6: Means and standard deviations (in parentheses) of the estimation accuracy measured by $D(\hat{Q}, Q)$ of equation (15) for constrained factor models.

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.

#### 5.2 Case 2. Partial Orthogonal Constraints

In this case, the observed data $Y_t$’s are generated using Model (5):

$$Y_t = H_R R_1 F_t C_1’ H_C’ + L_R R_2 G_t C_2’ L_C’ + U_t, \quad t = 1, \ldots, T.$$  

Parameter settings of the first part $H_R R_1 F_t C_1’ H_C’$ are the same as those in Case 1. The latent factor process $G_t$ is of dimension $q_1 \times q_2 = 5 \times 4$. The entries of $G_t$ follow $q_1 q_2$ independent $AR(1)$ processes with Gaussian white noise $\mathcal{N}(0,1)$ innovations, $vec(G_t) = \Phi_G vec(G_{t-1}) + \epsilon_t$ with $\Phi_G$ being a diagonal matrix with entries $(-0.7, 0.5, -0.2, 0.9, 0.1, 0.4, 0.6, -0.5, 0.7, 0.7, -0.4, 0.4, 0.4, -0.6, -0.6, -0.6, -0.3, 0.2, -0.4)$. The row loading matrix $L_R R_2$ is a $20 \times 5$ orthogonal matrix, satisfying $H_R’ L_R = 0$. The column loading matrix $L_C C_2$ is a $20 \times 4$ orthogonal matrix, satisfying $H_C’ L_C = 0$. The entries of $R_2$ and $C_2$ are random draws from the uniform distribution between $-p_i^{-1/2} / \sqrt{p_i / (p_i - m_i)}$ and $p_i^{-1/2} / \sqrt{p_i / (p_i - m_i)}$ for $i = 1, 2$, respectively, so that the conditions on factor strength are satisfied. Factor strength is controlled by the $\delta_i$’s.

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

$$Y_t = (H_R R_1 L_R R_2) \begin{pmatrix} F_t & 0 \\ 0 & G_t \end{pmatrix} \begin{pmatrix} C_1’ H_C’ \\ C_2’ L_C’ \end{pmatrix} + U_t, \quad t = 1, \ldots, T.$$  

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 $(H_C’ C_1 L_C C_2) \otimes (H_R R_1 L_R R_2)$. 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...
|       | $p_1$ | $p_2$ | $h_0 = 1$ | $h_0 = 2$ | $h_0 = 3$ | $h_0 = 4$ |
|-------|-------|-------|-----------|-----------|-----------|-----------|
| $f_c$ | 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      |
| $D_c$ | 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)|
| $D_u$ | 20    | 20    | 4.37(1.29)| 0.51(0.07)| 0.53(0.07)| 0.53(0.08)|
|       | 20    | 40    | 4.30(1.30)| 0.34(0.04)| 0.35(0.04)| 0.35(0.04)|
|       | 40    | 20    | 4.36(1.31)| 0.36(0.04)| 0.37(0.04)| 0.37(0.05)|
|       | 40    | 40    | 4.34(1.34)| 0.24(0.02)| 0.24(0.03)| 0.25(0.03)|

Table 7: Performance of estimation under different choices of $h_0$ when $\text{vec}(F_t) = \Phi_F \text{vec}(F_{t-2}) + \epsilon_t$. Metrics reported are relative frequencies of correctly estimating $k$, means and standard deviations (in parentheses) of the estimation accuracy measured by $D(\hat{Q}, Q)$. Means and standard deviations are multiplied by 10 for ease in presentation. $f_u$ and $f_c$ denote unconstrained and constrained models.

$\text{vec}(F_t)$ and $\text{vec}(G_t)$ denote the same frequency metric for the first matrix factor $F_t$ and second matrix factor $G_t$ of the constrained model. The number of factors in $F_t$ is estimated with a higher accuracy because the dimension of constrained loading space for $F_t$ is $m_1m_2 = 36$, which is smaller than that for $G_t$, $(p_1 - m_1)(p_2 - m_2) = 136$. The result again confirms the theoretical results in Section 4. Note that Table 8 only contains selected combinations of factor strength parameters $\delta_i$’s ($i = 1, \ldots, 4$). The results of all combinations of factor strength are given in Table 18 in Appendix C.

Figure 2 and Figure 3 present box-plots 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.
Table 8: Relative frequencies of correctly estimating the number of factors for partially constrained factor models. The full table including all combinations is presented in Table 18 in Appendix C.

Figure 2: The strong factors case. Box-plots of the estimation accuracy measured by $D(\hat{Q}, Q)$ 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 for the plotted values.

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 in order to determine an appropriate constrained factor model. The numbers of factors ($k_1$, $k_2$) can be determined by...
any existing methods, such as those in Lam & Yao (2012) and Wang et al. (2018). 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 in ??.

Detailed descriptions of the dataset and the transformation used are given in Table 15 and 16 of Appendix B. 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
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 Table 15 and 16 of Appendix C.

index by its factor loadings. A hierarchical clustering algorithm (Xu & Wunsch [2005] Murtagh & 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 \textit{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 6 groups: (i) United States and Canada; (ii) New Zealand and Australia; (iii) Norway; (iv) Ireland, Denmark, and United Kingdom; (v) Finland and Sweden; (vi) France, Netherlands, Austria, and Germany. The grouping more or less follows geographical 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 5 categories: (i) GDP, production of total industry excluding construction, and production of total manufacturing; (ii) long-term government bond yields and 3-month interbank rates and yields; (iii) total CPI and CPI of Food; (iv) CPI of Energy; (v) total exports value and total imports value in goods. Again, the grouping agrees with common economic knowledge.

![Country Clustering](image)

![Macro-Index Clustering](image)

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

Figure 5: Macroeconomic series: Clustering loading matrices

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

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 time spans. We calculate two matrices \( \hat{\mathbf{M}}_1 \) and \( \hat{\mathbf{M}}_2 \) according to (12) for the two disconnected time spans separately. The matrix \( \hat{\mathbf{M}} \) is re-defined as the sum of \( \hat{\mathbf{M}}_1 \) and \( \hat{\mathbf{M}}_2 \). Loading matrices and latent dimensions are estimated from this newly defined \( \hat{\mathbf{M}} \) 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
| Model | Loading | Row | USA | CAN | NZL | AUS | NOR | IRL | DNK | GBR | FIN | SWE | FRA | NLD | AUT | DEU |
|-------|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| $R_{\text{unc}}$ | $\hat{R}$ | 1 | 3 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| | | 2 | 4 | 2 | 5 | 5 | 1 | 0 | 1 | 0 | -3 | -1 | -2 | -2 | -2 | -3 |
| | | 3 | 3 | 6 | -2 | -2 | 4 | -5 | -3 | -1 | 1 | 0 | -1 | 1 | 0 | 0 |
| | | 4 | -4 | -3 | 0 | 2 | 8 | -1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| $R_{\text{con}}$ | $\tilde{R}^*H_R^*$ | 1 | 1 | 1 | 2 | 2 | 3 | 3 | 3 | 4 | 4 | 3 | 3 | 3 | 3 | 3 |
| | | 2 | 5 | 5 | 3 | 3 | 4 | 0 | 0 | 0 | -2 | -2 | -2 | -2 | -2 | -2 |
| | | 3 | -1 | -1 | 5 | 5 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -1 | -1 |
| | | 4 | -4 | -4 | 3 | 3 | 6 | -2 | -2 | -2 | 1 | 1 | -1 | -1 | -1 | -1 |

| Model | Loading | Row | CPI:Food | CPI:Ener | CPI:Tot | IR:Long | IR:3-Mon | P:TIEC | P:TM | GDP | IT:Ex | IT:Im |
|-------|---------|-----|----------|----------|---------|----------|----------|---------|-------|------|-------|-------|
| $C_{\text{unc}}$ | $\tilde{C}$ | 1 | 1 | 4 | 2 | 4 | 3 | 3 | 3 | 4 | 4 | 4 |
| | | 2 | 5 | 3 | 6 | -1 | 1 | -3 | -4 | -4 | 0 | 0 |
| | | 3 | 5 | -1 | 2 | -1 | 1 | 4 | 4 | 3 | -4 | -4 |
| | | 4 | 0 | -1 | -2 | 7 | 5 | -2 | -2 | 0 | -3 | -3 |
| $C_{\text{con}}$ | $\tilde{C}^*H_C^*$ | 1 | 6 | -2 | 6 | 4 | 4 | 0 | 0 | 0 | -2 | -2 |
| | | 2 | 0 | 0 | 0 | 3 | 3 | 5 | 5 | 5 | 3 | 3 |
| | | 3 | -3 | 3 | -3 | 5 | 5 | -3 | -3 | -3 | 1 | 1 |
| | | 4 | 3 | 5 | 3 | -1 | -1 | -2 | -2 | -2 | 5 | 5 |

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

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 contains the descriptions of variables used and their definitions, the 200 companies and their corresponding industry group and sector information. Data are arranged in matrix-variate time series format. At each $t$, we observe a $16 \times 200$ matrix, whose rows represent financial variables and columns represent companies. Thus we have $T = 40$, $p_1 = 16$ and $p_2 = 200$. The total number of time series is 3,200. Following the convention in eigenanalysis, we standardize the
| Model     | # Factor 1 | # Factor 2 | RSS    | RSS/TSS | # Parameters |
|-----------|------------|------------|--------|---------|--------------|
| Full      | (6,5)      |            | 570.50 | 0.449   | 134          |
| Constrained | (6,5)     |            | 560.31 | 0.442   | 61           |
| Partial   | (6,5)      | (6,5)      | 454.41 | 0.358   | 134          |
| Full      | (5,5)      |            | 613.26 | 0.482   | 120          |
| Constrained | (5,5)     |            | 604.63 | 0.477   | 55           |
| Partial   | (5,5)      | (5,5)      | 516.27 | 0.407   | 120          |
| Full      | (4,5)      |            | 658.15 | 0.517   | 106          |
| Constrained | (4,5)     |            | 649.85 | 0.512   | 49           |
| Partial   | (4,5)      | (4,5)      | 576.94 | 0.454   | 106          |
| Full      | (4,4)      |            | 729.46 | 0.573   | 96           |
| Constrained | (4,4)     |            | 721.96 | 0.568   | 44           |
| Partial   | (4,4)      | (4,4)      | 657.13 | 0.517   | 96           |
| Full      | (3,4)      |            | 787.80 | 0.620   | 82           |
| Constrained | (3,4)     |            | 768.64 | 0.605   | 38           |
| Partial   | (3,4)      | (3,4)      | 719.46 | 0.567   | 82           |
| Full      | (3,3)      |            | 868.43 | 0.684   | 72           |
| Constrained | (3,3)     |            | 852.76 | 0.671   | 33           |
| Partial   | (3,3)      | (3,3)      | 813.16 | 0.640   | 72           |

Table 11: Results of 10-fold CV of out-of-sample performance for the multinational macroeconomic indexes. The numbers shown are average over the cross validation, where RSS and TSS stand for residual and total sum of squares, respectively.

individual series before applying factor analysis. This data set was used in Wang et al. (2018) for an unconstrained matrix factor model.

The column constraint matrix $H_C$ 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 $H_C$ is $200 \times 51$. 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 $H_R = I_{16}$.

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 3,200 time series is typical 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.

| Model     | # Factor 1 | # Factor 2 | RSS    | RSS/SST | # parameters |
|-----------|------------|------------|--------|---------|--------------|
| Full      | (4,10)     | (4,12)     | 8140.32| 0.869   | 2064         |
|           | (4,19)     |            | 7990.04| 0.853   | 2464         |
| Constrained| (4,10)   | (4,10)     | 8062.63| 0.861   | 574          |
| Partial   | (4,10)     | (4,10)     | 7969.83| 0.851   | 936          |
|           | (4,20)     |            | 7539.68| 0.805   | 4064         |
| Full      | (4,27)     | (4,27)     | 7261.49| 0.775   | 5464         |
|           | (4,39)     |            | 6872.18| 0.734   | 7864         |
| Constrained| (4,20)  | (4,20)     | 7646.70| 0.816   | 1084         |
| Partial   | (4,20)     | (4,20)     | 7292.06| 0.779   | 2191         |
|           | (4,20)     | (4,19)     | 6815.96| 0.728   | 3979         |
| Full      | (5,10)     | (5,12)     | 7849.34| 0.838   | 2480         |
|           | (5,19)     |            | 7420.04| 0.792   | 3880         |
| Constrained| (5,10) | (5,10)     | 7942.95| 0.848   | 590          |
| Partial   | (5,10)     | (5,10)     | 7849.40| 0.838   | 968          |
|           | (5,20)     | (5,20)     | 7357.80| 0.786   | 1627         |
|           | (5,20)     | (5,19)     | 6595.03| 0.704   | 4011         |
| Full      | (5,30)     | (5,30)     | 6960.70| 0.743   | 6080         |
|           | (5,34)     |            | 6813.93| 0.727   | 6880         |
| Constrained| (5,30) | (5,30)     | 7184.53| 0.767   | 1610         |
| Partial   | (5,30)     | (5,30)     | 6997.21| 0.747   | 2286         |
|           | (5,30)     | (5,29)     | 5936.64| 0.634   | 6011         |

Table 12: Summary of 10-fold CV of out-of-sample analysis for the 16 corporate financial measurements for each of 200 companies. The numbers shown are average over the cross validation and RSS and TSS denote, respectively, the residual and total sum of squares.
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 et al. (2018) carried out a clustering analysis on the ME and BE/ME loading matrices after rotation. Their results suggest $H_R = [h_{R_1}, h_{R_2}, h_{R_3}]$, where $h_{R_1} = [1(5)/\sqrt{5}, 0(5)]$, $h_{R_2} = [0(5), 1(4)/2, 0]$, and $h_{R_3} = [0(9), 1]$. Therefore, ME factors are classified into three groups of smallest 5 ME’s, middle 4 ME’s, and the largest ME, respectively. For cases when we need 4 row constraints, we redefine $h_{R_2} = [0(5), 1(3)/\sqrt{3}, 0(2)]$ and add a fourth column $h_{R_4} = [0(8), 1, 0]$. For column constraints, $H_C = [h_{C_1}, h_{C_2}, h_{C_3}]$, where $h_{C_1} = [1, 0(9)]$, $h_{C_2} = [0, 1(3)/\sqrt{3}, 0(6)]$, $h_{C_3} = [0(4), 1(6)]$. Therefore, BE/ME factors are divided into three groups of the smallest BE/ME’s, middle 3 BE/ME’s, and the 6 largest BE/ME, respectively. For cases when we need 4 column constraints, we redefine $h_{C_3} = [0(4), 1(4)/2, 0(2)]$ and add a fourth column $h_{C_4} = [0(8), 1(2)]$.

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
Table 13: Estimates of the loading matrices of constrained and unconstrained matrix factor modes for Fama-French 10×10 portfolio returns. The loading matrices are varimax rotated and normalized for ease in comparison.

| Model | Loading | Column | Rotated Estimated Loadings |
|-------|---------|--------|---------------------------|
| $\tilde{R}'$ | 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 |
| $\tilde{R}'H'_R$ | 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 |
| $\tilde{C}'$ | 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 |
| $\tilde{C}'H'_C$ | 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 |

model uses a much smaller number of parameters than the unconstrained model.

7 Summary and Discussion

This paper established a general framework for incorporating domain or prior knowledge induced linear constraints in the matrix factor model. We developed efficient estimation procedures for 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 data sets, where the constrained factor models outperform their unconstrained counterparts in explaining the variabilities of the data using out-of-sample 10-fold cross validation and in factor interpretation.

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. Firstly, a natural question is how we know the existence of weak factors in real data. Lam & Yao (2012) utilized a two-step approach to facilitate the discovery of weak factors that may be masked by strong factors. They run 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
| Model      | # Factor 1 | # Factor 2 | RSS      | RSS/SST | # Parameters |
|------------|------------|------------|----------|---------|--------------|
| (3,3)      |            |            | 3064.40  | 0.500   | 60           |
| Full       | (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           |
| (2,2)      |            |            | 3473.32  | 0.567   | 40           |
| Full       | (2,3)      |            | 3269.50  | 0.533   | 50           |
|            | (2,4)      |            | 3152.63  | 0.514   | 60           |
| Constrained| (2,2)      |            | 3535.56  | 0.577   | 16           |
| Partial    | (2,2)      | (2,2)      | 3415.25  | 0.557   | 40           |
|            | (2,2)      | (2,1)      | 3486.15  | 0.569   | 33           |

Table 14: Performance of out-of-sample 10-fold CV of constrained and unconstrained factor models using Fama-French $10 \times 10$ portfolio return series, where $RSS$ and $RSS/TSS$ denote, respectively, the residual and total sum of squares.

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 sub-panels or block structure is a common situation where weak factors arise. Hallin & 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 & Tsay (2010). Data containing sub-panels can also be cast into matrix observations by putting sub-panels 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.
References

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

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

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

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

Bickel, P. J., Levina, E. et al. (2008), ‘Covariance regularization by thresholding’, *The Annals of Statistics* **36**(6), 2577–2604.

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

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

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

Chang, J., Guo, B. & Yao, Q. (2018), ‘Principal component analysis for second-order stationary vector time series’, *Annals of Statistics* **46**, 2094–2124.

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

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

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

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

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2000), ‘The generalized dynamic-factor model: Identification and estimation’, *Review of Economics and Statistics* **82**(4), 540–554.
Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2004), ‘The generalized dynamic factor model consistency and rates’, *Journal of Econometrics* **119**(2), 231–255.

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

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

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

Hallin, M. & Liška, R. (2007), ‘Determining the number of factors in the general dynamic factor model’, *Journal of the American Statistical Association* **102**(478), 603–617.

Hallin, M. & Liška, R. (2011), ‘Dynamic factors in the presence of blocks’, *Journal of Econometrics* **163**(1), 29–41.

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

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

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

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

Murtagh, F. & Legendre, P. (2014), ‘Ward’s hierarchical agglomerative clustering method: which algorithms implement ward’s criterion?’, *Journal of Classification* **31**(3), 274–295.

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

Pan, J. & Yao, Q. (2008), ‘Modelling multiple time series via common factors’, *Biometrika* **95**(2), 365–379.

Rudebusch, G. D. & Wu, T. (2008), ‘A macro-finance model of the term structure, monetary policy and the economy’, *The Economic Journal* **118**(530), 906–926.
Tibshirani, R. (1996), ‘Regression shrinkage and selection via the lasso’, *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 267–288.

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

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

Wang, D., Liu, X. & Chen, R. (2018), ‘Factor models for matrix-valued high-dimensional time series’, *Journal of Econometrics*.

Xu, R. & Wunsch, D. (2005), ‘Survey of clustering algorithms’, *IEEE Transactions on Neural Networks* **16**(3), 645–678.

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

Zou, H. (2006), ‘The adaptive lasso and its oracle properties’, *Journal of the American Statistical Association* **101**(476), 1418–1429.