Statistica Sinica **26** (2016), 1453-1478
doi:http://dx.doi.org/10.5705/ss.2013.332t

DOUBLY CONSTRAINED FACTOR MODELS WITH APPLICATIONS

Henghsiu Tsai, Ruey S. Tsay, Edward M. H. Lin and Ching-Wei Cheng

Academia Sinica, University of Chicago,
National Chiao Tung University and Purdue University

Abstract: This paper focuses on factor analysis of multivariate time series. We propose statistical methods that enable analysts to leverage their prior knowledge or substantive information to sharpen the estimation of common factors. Specifically, we consider a doubly constrained factor model that enables analysts to specify both row and column constraints of the data matrix to improve the estimation of common factors. The row constraints may represent classifications of individual subjects whereas the column constraints may show the categories of variables. We derive both the maximum likelihood and least squares estimates of the proposed doubly constrained factor model and use simulations to study the performance of the analysis in finite samples. The Akaike information criterion is used for model selection. Monthly U.S. housing start data from nine geographical divisions are used to demonstrate the application of the proposed model.

Key words and phrases: Akaike information criterion, constrained factor model, eigenvalues, factor model, housing starts, principal component analysis, seasonality.

1. Introduction

Big data have become common in statistical applications. In many situations, it is natural to entertain the data as a 2-dimensional array with row representing subjects and column denoting variables, for instance, large panel data in the econometric literature and multivariate time series data in statistics. For a specific example, consider the United States (U.S.) housing markets. The U.S. Census Bureau publishes monthly housing starts from nine geographical divisions shown in Figure 1. We employ 10 years of the data from January 1997 to December 2006. Here the data matrix $\boldsymbol{Z}$ is a 120-by-9 matrix with each column representing a division and each row denoting a particular calendar month. Figure 2 shows the time plots of the logarithms of monthly housing starts of the nine divisions. From the plots, it is clear that U.S. housing starts have strong seasonality, and that housing starts exhibit some common characteristics. It is then natural to consider seasonality (row constraints) and geographical divisions (column constraints) in searching for common factors driving the U.S. housing markets. The goal of this paper is to consider such constraints when we search for common factors in a big data set.

Factor models are widely used in econometric and statistical applications, and constrained factor models have also been studied in the literature. Bai and Ng (2002), Bai (2003), Lam, Yao, and Bathia (2011), Lam and Yao (2012), and Chang, Guo, and Yao (2013) represent multiple time series using a few common factors defined in various ways. Forni et al. (2000, 2005) generalize the static approximate factor model of Chamberlain and Rothschild (1983) to the generalized dynamic-factor model that allows for infinite dynamics and nonorthogonal idiosyncratic components. Tsai and Tsay (2010) proposed constrained and partially constrained factor models for multivariate time series analysis; they show that column constraints can be used effectively to obtain parsimonious factor models for high-dimensional series. Only column constraints are considered in that paper, however, when both row and column constraints are informative in some applications. We investigate doubly constrained factor models in this paper. The theoretical framework of the proposed model is the constrained principal component analysis of Takane and Hunter (2001), and our study focuses on estimation and applications of the proposed model. Principal component analysis was proposed originally for independent data, but it has been widely used in the time series analysis, see, for instance, Peña and Box (1987) and Tiao, Tsay, and Wang (1993).

Consider a T by N data matrix $\boldsymbol{Z}$, rows and columns of which represent subjects and variables, respectively. Let $\boldsymbol{G}$ be a T by m matrix of row constraints of rank m, and $\boldsymbol{H}$ be an N by s matrix of column constraints of rank s. Both $\boldsymbol{G}$ and $\boldsymbol{H}$ are known *a priori* based on some prior knowledge or substantive information of the problem at hand. For instance, Tsai and Tsay (2010) use $\boldsymbol{H}$ to represent the level, slope, and curvature of interest rates, and to denote the industrial classification of U.S. stocks.

Let $\boldsymbol{\omega}_1 = [\boldsymbol{\omega}_1(i,j)]$ (s by r), $\boldsymbol{\omega}_2 = [\boldsymbol{\omega}_2(i,j)]$ (N by p), and $\boldsymbol{\omega}_3 = [\boldsymbol{\omega}_3(i,j)]$ (s by q) be the loading matrices of full rank, and $\boldsymbol{E}$ (T by N) a matrix of residuals, where $p < N$, $\max\{r,q\} \le s < N$, and $q \le \min\{r,p\}$. The postulated doubly constrained factor (DCF) model for $\boldsymbol{Z} = [Z_{i,j}] = [Z_1', \cdots, Z_T']'$ is

$$\boldsymbol{Z} = \boldsymbol{F}_1\boldsymbol{\omega}_1'\boldsymbol{H}' + \boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\omega}_2' + \boldsymbol{G}\boldsymbol{F}_3\boldsymbol{\omega}_3'\boldsymbol{H}' + \boldsymbol{E}, \tag{1.1}$$

where $\boldsymbol{A}'$ denotes the transpose matrix of $\boldsymbol{A}$, $\boldsymbol{F}_1 = [F_1^{(1)'}, \ldots, F_1^{(T)'}]'$ (T by r), $\boldsymbol{F}_2 = [F_2^{(1)'}, \ldots, F_2^{(m)'}]'$ (m by p), $\boldsymbol{F}_3 = [F_3^{(1)'}, \ldots, F_3^{(m)'}]'$ (m by q), and $\boldsymbol{E} = [e_1', \ldots, e_T']'$ (T by N) with $E(e_i) = \boldsymbol{0}$, and $\mathrm{var}(e_i) = \boldsymbol{\Psi} = [\Psi(j,k)]$. We refer to the model at (1.1) as a DCF model of order (r,p,q) with r, p, and q denoting the number of common factors in $\boldsymbol{F}_1, \boldsymbol{F}_2$ and $\boldsymbol{F}_3$, respectively. For statistical factor models, one further assumes that $\boldsymbol{\Psi}$ is a diagonal matrix. In the econometric and finance literature, $\boldsymbol{\Psi}$ is not necessarily diagonal and the model becomes an approximate factor model.

For the DCF model at (1.1), the $\boldsymbol{F}_i$ are common factors. Under the model, the first term pertains to what in $\boldsymbol{Z}$ can be explained by $\boldsymbol{H}$ but not by $\boldsymbol{G}$, the second term to what can be explained by $\boldsymbol{G}$ but not by $\boldsymbol{H}$, the third term to what can be explained jointly by both $\boldsymbol{G}$ and $\boldsymbol{H}$, and the last term to what can be explained by neither $\boldsymbol{G}$ nor $\boldsymbol{H}$. Often the third term of (1.1) denotes the interaction between the constraints $\boldsymbol{G}$ and $\boldsymbol{H}$. Thus, $\boldsymbol{F}_1, \boldsymbol{F}_2$ and $\boldsymbol{F}_3$ can be interpreted as column, row, and interaction factors, respectively. Similar to the conventional factor models, the scales and orderings of the latent common factors $\boldsymbol{F}_i$ are not identifiable.

The model studied in this paper is not an approximate factor model in the sense of Chamberlain and Rothschild (1983) and Bai (2003). Our model is an extension of the traditional orthogonal factor models in the sense that the cross-section size N is fixed and finite, and $\boldsymbol{E}$, the covariance matrix of the idiosyncratic errors, is diagonal. The class of approximate factor models allows the idiosyncratic components to be 'poorly' correlated. An important property of approximate factor models is that as $N \to \infty$, if the factors are white noises and orthogonal to the idiosyncratic terms, the common components of a factor model with r factors can be recovered by the first r principal components of the covariance matrix of the observations. In this sense, the main principal components can approximate the common components when N is large. The simulations in Section 3 deal with $N = 6$ and $N = 24$, and in the application in Section 4, $N = 9$. The model studied in this paper also differs from the dynamic models in Forni et al. (2000) because it does not allow the factors to be auto-correlated.

The paper is organized as follows. In Section 2 we consider estimation of the proposed DCF model, including model selection and the common factors. We use simulations in Section 3 to investigate the efficacy of the estimation methods in finite samples. Section 4 applies the proposed analysis to the monthly U.S. housing starts, and Section 5 concludes.

2. Estimation

The proposed doubly constrained factor model at (1.1) can be estimated by least squares (LS) or maximum likelihood. In either case, we assume, for simplicity, that the row constraint $\boldsymbol{G}$ satisfies

$$\boldsymbol{G}'\boldsymbol{G} = \frac{T}{m} I_m, \tag{2.1}$$

where I_m is the $m \times m$ identity matrix. This is not a strong condition. For example, if (i) $\boldsymbol{G} = \boldsymbol{I}_m \otimes \mathbf{1}_{T/m}$, where $\mathbf{1}_{T/m}$ is the T/m-dimensional vector of 1, or if (ii) $\boldsymbol{G} = \mathbf{1}_{T/m} \otimes \boldsymbol{I}_m$, then (2.1) holds. The U.S. housing starts data follow (ii). The LS estimates are less efficient, but easier to obtain, we begin with them.

2.1. Least squares estimation

Consider the doubly constrained factor model in (1.1) subject to the following.

Assumption A.

$$\boldsymbol{F}_1'\boldsymbol{F}_1 = T\boldsymbol{I}_r, \boldsymbol{F}_2'\boldsymbol{F}_2 = m\boldsymbol{I}_p, \boldsymbol{F}_3'\boldsymbol{F}_3 = m\boldsymbol{I}_q, \boldsymbol{G}'\boldsymbol{F}_1\boldsymbol{\omega}_1' = \boldsymbol{O}, \text{ and } \boldsymbol{F}_2\boldsymbol{\omega}_2'\boldsymbol{H} = \boldsymbol{O}. \quad (2.2)$$

The least squares estimates (LSE) of $\boldsymbol{\omega}_1$, $\boldsymbol{\omega}_2$, $\boldsymbol{\omega}_3$, $\boldsymbol{F}_1$, $\boldsymbol{F}_2$, and $\boldsymbol{F}_3$ can be obtained by minimizing the objective function

$$\begin{aligned}
&l(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \boldsymbol{\omega}_3, \boldsymbol{F}_1, \boldsymbol{F}_2, \boldsymbol{F}_3) \\
&= \operatorname{tr}\{(\boldsymbol{Z} - \boldsymbol{F}_1\boldsymbol{\omega}_1'\boldsymbol{H}' - \boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\omega}_2' - \boldsymbol{G}\boldsymbol{F}_3\boldsymbol{\omega}_3'\boldsymbol{H}')(\boldsymbol{Z} - \boldsymbol{F}_1\boldsymbol{\omega}_1'\boldsymbol{H}' - \boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\omega}_2' - \boldsymbol{G}\boldsymbol{F}_3\boldsymbol{\omega}_3'\boldsymbol{H}')'\} \\
&= \operatorname{tr}\{\boldsymbol{Z}\boldsymbol{Z}' + \boldsymbol{F}_1\boldsymbol{\omega}_1'H'H\boldsymbol{\omega}_1\boldsymbol{F}_1' + \boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\omega}_2'\boldsymbol{\omega}_2\boldsymbol{F}_2'\boldsymbol{G}' + \boldsymbol{G}\boldsymbol{F}_3\boldsymbol{\omega}_3'H'H\boldsymbol{\omega}_3\boldsymbol{F}_3'\boldsymbol{G}' \\
&\quad -2\boldsymbol{Z}(H\boldsymbol{\omega}_1\boldsymbol{F}_1' + \boldsymbol{\omega}_2\boldsymbol{F}_2'\boldsymbol{G}' + H\boldsymbol{\omega}_3\boldsymbol{F}_3'\boldsymbol{G}')\}, \qquad (2.3)
\end{aligned}$$

where the second equality follows from the zero constraints of Assumption A. Taking the partial derivative of $l(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \boldsymbol{\omega}_3, \boldsymbol{F}_1, \boldsymbol{F}_2, \boldsymbol{F}_3)$ with respect to $\boldsymbol{\omega}_1$, $\boldsymbol{\omega}_2$, and $\boldsymbol{\omega}_3$, respectively, and equating the results to zero, we obtain

$$\widehat{\boldsymbol{\omega}}_1 = (\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'\boldsymbol{Z}'\boldsymbol{F}_1(\boldsymbol{F}_1'\boldsymbol{F}_1)^{-1}, \qquad (2.4)$$

$$\widehat{\boldsymbol{\omega}}_2 = \boldsymbol{Z}'\boldsymbol{G}\boldsymbol{F}_2(\boldsymbol{F}_2'\boldsymbol{G}'\boldsymbol{G}\boldsymbol{F}_2)^{-1}, \qquad (2.5)$$

$$\widehat{\boldsymbol{\omega}}_3 = (\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'\boldsymbol{Z}'\boldsymbol{G}\boldsymbol{F}_3(\boldsymbol{F}_3'\boldsymbol{G}'\boldsymbol{G}\boldsymbol{F}_3)^{-1}. \qquad (2.6)$$

Plugging $\widehat{\boldsymbol{\omega}}_1$, $\widehat{\boldsymbol{\omega}}_2$, and $\widehat{\boldsymbol{\omega}}_3$ into (2.3), and using the fact that $\boldsymbol{F}_1'\boldsymbol{F}_1 = T\boldsymbol{I}_r$, $\boldsymbol{F}_2'\boldsymbol{F}_2 = m\boldsymbol{I}_p$, $\boldsymbol{F}_3'\boldsymbol{F}_3 = m\boldsymbol{I}_q$, $\boldsymbol{G}'\boldsymbol{G} = TI_m/m$, and $tr(\mathbf{AB}) = tr(\mathbf{BA})$, we obtain the concentrated function

$$\begin{aligned}
l(\boldsymbol{F}_1, \boldsymbol{F}_2, \boldsymbol{F}_3) = \operatorname{tr}\Bigg\{&\boldsymbol{Z}\boldsymbol{Z}' - \frac{1}{T}\boldsymbol{F}_1'\boldsymbol{Z}\boldsymbol{H}(\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'\boldsymbol{Z}'\boldsymbol{F}_1 \\
&- \frac{1}{T}\boldsymbol{F}_2'\boldsymbol{G}'\boldsymbol{Z}\boldsymbol{Z}'\boldsymbol{G}\boldsymbol{F}_2 - \frac{1}{T}\boldsymbol{F}_3'\boldsymbol{G}'\boldsymbol{Z}\boldsymbol{H}(\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'\boldsymbol{Z}'G\boldsymbol{F}_3\Bigg\}. \qquad (2.7)
\end{aligned}$$

The objective function (2.7) is minimized when the second, the third, and the last term is maximized with respect to $\boldsymbol{F}_1$, $\boldsymbol{F}_2$, and $\boldsymbol{F}_3$, respectively. Applying Theorem 6 of Magnus and Neudecker (1999) or Proposition A.4 of Lütkepohl (2005), we have $\widehat{\mathbf{F}}_1 = [\mathbf{g}_1^1, \cdots, \mathbf{g}_r^1]$, where $\mathbf{g}_i^1$ is an eigenvector of the ith largest eigenvalue λ_i^1 of $\mathbf{ZH}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{Z}'$. Similarly, $\widehat{\mathbf{F}}_2 = [\mathbf{g}_1^2, \cdots, \mathbf{g}_p^2]$, where $\mathbf{g}_i^2$ is an eigenvector of the ith largest eigenvalue λ_i^2 of $\mathbf{G}'\mathbf{ZZ}'\mathbf{G}$, and $\widehat{\mathbf{F}}_3 = [\mathbf{g}_1^3, \cdots, \mathbf{g}_q^3]$, where

$\mathbf{g}_i^3$ is an eigenvector of the ith largest eigenvalue λ_i^3 of $\boldsymbol{G}'\boldsymbol{Z}\boldsymbol{H}(\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'\boldsymbol{Z}'G$. Note that the eigenvectors are standardized so that $\widehat{\mathbf{F}}_1'\widehat{\mathbf{F}}_1 = T\mathbf{I}_r$, $\widehat{\mathbf{F}}_2'\widehat{\mathbf{F}}_2 = m\mathbf{I}_p$. $\widehat{\mathbf{F}}_3'\widehat{\mathbf{F}}_3 = m\mathbf{I}_q$. The corresponding estimate of $\boldsymbol{\omega}_i$, $i = 1, 2, 3$, are computed by (2.4), (2.5), and (2.6). Specifically, by the fact that $\boldsymbol{F}_1'\boldsymbol{F}_1 = T\boldsymbol{I}_r$, $\boldsymbol{F}_2'\boldsymbol{F}_2 = m\boldsymbol{I}_p$, $\boldsymbol{F}_3'\boldsymbol{F}_3 = m\boldsymbol{I}_q$, and $\boldsymbol{G}'\boldsymbol{G} = TI_m/m$, (2.4), (2.5), and (2.6) yield

$$\widehat{\boldsymbol{\omega}}_1 = \frac{1}{T}(\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'\boldsymbol{Z}'\widehat{\mathbf{F}}_1, \tag{2.8}$$

$$\widehat{\boldsymbol{\omega}}_2 = \frac{1}{T}\boldsymbol{Z}'\boldsymbol{G}\widehat{\mathbf{F}}_2, \tag{2.9}$$

$$\widehat{\boldsymbol{\omega}}_3 = \frac{1}{T}(\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'\boldsymbol{Z}'\boldsymbol{G}\widehat{\mathbf{F}}_3. \tag{2.10}$$

The $\boldsymbol{\Psi}$ matrix is estimated by $\widehat{\boldsymbol{\Psi}} = \widehat{\boldsymbol{E}}'\widehat{\boldsymbol{E}}/T$, where $\widehat{\boldsymbol{E}} = \boldsymbol{Z} - \widehat{\boldsymbol{F}}_1\widehat{\boldsymbol{\omega}}_1'\boldsymbol{H}' - \boldsymbol{G}\widehat{\boldsymbol{F}}_2\widehat{\boldsymbol{\omega}}_2' - \boldsymbol{G}\widehat{\boldsymbol{F}}_3\widehat{\boldsymbol{\omega}}_3'\boldsymbol{H}'$. It is understood that $\widehat{\boldsymbol{\Psi}} = \mathrm{diag}(\widehat{\boldsymbol{E}}'\widehat{\boldsymbol{E}}/T)$ if $\boldsymbol{\Psi}$ is diagonal.

2.2. Maximum likelihood estimation

For maximum likelihood estimation, we assume that, for $1 \le t \le T$, $\mathrm{var}(e_t) = \boldsymbol{\Psi}$ is a diagonal N by N matrix. We further assume that $E(F_i^{(k)}) = 0$, and $\mathrm{var}(F_i^{(k)}) = \boldsymbol{I}$, the identity matrix, for $1 \le i \le 3$, and all k. We also assume $\mathrm{cov}(F_i^{(k)}, F_j^{(l)}) = 0$ for $k \ne l$ or $i \ne j$, $\mathrm{cov}(e_i, e_j) = 0$ for all $i \ne j$, $\mathrm{cov}(F_i^{(k)}, e_j) = 0$ for all i, j, and k, and e_j is an N-dimensional Gaussian random vector with mean zero and diagonal covariance matrix $\boldsymbol{\Psi}$.

For the purpose of identifiability, we adopt the approach of Anderson (2003) by imposing the restrictions that the matrices $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$, and $\boldsymbol{\Gamma}_3$ are all diagonal, where

$$\boldsymbol{\Gamma}_1 = \boldsymbol{\omega}_1'\boldsymbol{H}'\boldsymbol{\Psi}^{-1}\boldsymbol{H}\boldsymbol{\omega}_1, \quad \boldsymbol{\Gamma}_2 = \boldsymbol{\omega}_2'\boldsymbol{\Psi}^{-1}\boldsymbol{\omega}_2, \quad \boldsymbol{\Gamma}_3 = \boldsymbol{\omega}_3'\boldsymbol{H}'\boldsymbol{\Psi}^{-1}\boldsymbol{H}\boldsymbol{\omega}_3. \tag{2.11}$$

We also assume that the diagonal elements of $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$, and $\boldsymbol{\Gamma}_3$ are ordered and distinct ($\gamma_{11}^1 > \gamma_{22}^1 > \cdots > \gamma_{rr}^1$, $\gamma_{11}^2 > \gamma_{22}^2 > \cdots > \gamma_{pp}^2$, and $\gamma_{11}^3 > \gamma_{22}^3 > \cdots > \gamma_{qq}^3$), and that the first non-zero element in each column of the matrices $\boldsymbol{\omega}_i$, $i = 1, 2, 3$, is positive, so $\boldsymbol{\omega}_1$, $\boldsymbol{\omega}_2$ and $\boldsymbol{\omega}_3$ are uniquely determined; see the online supplementary material for a proof. It can readily be checked that the covariance matrix of $\mathrm{vec}(\boldsymbol{Z}')$ is $\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{I}_T \otimes \boldsymbol{A} + \boldsymbol{G}\boldsymbol{G}' \otimes \boldsymbol{B}$, where $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{\omega}_1\boldsymbol{\omega}_1'\boldsymbol{H}' + \boldsymbol{\Psi}$, and $\boldsymbol{B} = \boldsymbol{\omega}_2\boldsymbol{\omega}_2' + \boldsymbol{H}\boldsymbol{\omega}_3\boldsymbol{\omega}_3'\boldsymbol{H}'$. For the definitions of the matrix operators $\mathrm{vec}(\cdot)$ and $\otimes$, see, for example, Schott (1997).

We divide the discussion of maximum likelihood estimation into subsections to better understand the flexibility of the proposed model. Also, the existence of row constraints requires an additional condition to simplify the estimation.

2.2.1. Case 1: $\boldsymbol{\omega}_2 = \boldsymbol{\omega}_3 = \mathbf{0}$

Here the proposed model is

$$\boldsymbol{Z} = \boldsymbol{F}_1\boldsymbol{\omega}_1^{'}\boldsymbol{H}^{'} + \boldsymbol{E}, \tag{2.12}$$

which is the column constrained factor model of Tsai and Tsay (2010). An iterated procedure was proposed there to perform estimation.

2.2.2. Case 2: $\boldsymbol{\omega}_1 = \boldsymbol{\omega}_3 = \mathbf{0}$

With $\boldsymbol{\omega}_1 = \boldsymbol{\omega}_3 = 0$, the doubly constrained factor model is

$$\boldsymbol{Z} = \boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\omega}_2^{'} + \boldsymbol{E}. \tag{2.13}$$

Here the model can be estimated by an iterated procedure similar to that of Tsai and Tsay (2010). Let $\boldsymbol{Y} = (\boldsymbol{G}^{'}\boldsymbol{G})^{-1}\boldsymbol{G}^{'}\boldsymbol{Z}$, and $\boldsymbol{C}_Y = \boldsymbol{Y}'\boldsymbol{Y}/m$. The estimating procedure is as follows:

1. Compute initial estimates of the diagonal matrix $\widehat{\boldsymbol{\Psi}} = [\widehat{\Psi}(j,k)]$. Following Jöreskog (1975), set $\widehat{\Psi}(i,i) = (1 - r/(2N))/s^{ii}$, $i = 1,\ldots,N$, where s^{ii} is the ith diagonal element of $\mathbf{S}^{-1}$ and $\mathbf{S} = \mathbf{Z}'\mathbf{Z}/(T-1)$.
2. Construct the symmetric matrix $\mathbf{R}_B = \widehat{\boldsymbol{\Psi}}^{-1/2}(\mathbf{C}_Y - m\widehat{\boldsymbol{\Psi}}/T)\widehat{\boldsymbol{\Psi}}^{-1/2}$ and perform a spectral decomposition on $\mathbf{R}_B$, say $\mathbf{R}_B = \mathbf{L_B}\mathbf{W_B}\mathbf{L_B}'$, where $\mathbf{W_B} = \mathrm{diag}(\hat{\gamma}_j)$ and $\hat{\gamma}_1 > \hat{\gamma}_2 > \cdots > \hat{\gamma}_N$ are the ordered eigenvalues of $\mathbf{R}_B$.
3. Let $\widehat{\boldsymbol{\Gamma}}_B = \mathbf{W}_B$ and $\widehat{\boldsymbol{\Gamma}}_2 = \mathbf{W}_2$, where $\mathbf{W}_2$ is the left-upper $r \times r$ submatrix of $\mathbf{W}_B$. Obtain $\widehat{\boldsymbol{\omega}}_2$ from $\widehat{\boldsymbol{\Psi}}^{-1/2}\widehat{\boldsymbol{\omega}}_2 = \mathbf{L}_2$, where $\mathbf{L}_2$ consists of the first r columns of $\mathbf{L}_B$. The eigenvectors are normalized such that $\widehat{\boldsymbol{\omega}}_2'\widehat{\boldsymbol{\Psi}}^{-1}\widehat{\boldsymbol{\omega}}_2 = \widehat{\boldsymbol{\Gamma}}_2$. More precisely, $\widehat{\boldsymbol{\omega}}_2$ is a normalized version of $\widehat{\boldsymbol{\omega}}_2^* = \widehat{\boldsymbol{\Psi}}^{1/2}\mathbf{L}_2$, where the normalization is to ensure that $\widehat{\boldsymbol{\omega}}_2'\widehat{\boldsymbol{\Psi}}^{-1}\widehat{\boldsymbol{\omega}}_2 = \widehat{\boldsymbol{\Gamma}}_2$, a diagonal matrix.
4. Substitute $\widehat{\boldsymbol{\omega}}_2$ obtained in Step 3 into the objective function

$$\frac{m}{T}\ln|\widehat{Q}| + \frac{T-m}{T}\ln|\widehat{\boldsymbol{\Psi}}| + \mathrm{tr}(\boldsymbol{C}_Y\widehat{\boldsymbol{Q}}^{-1}) + \mathrm{tr}((\boldsymbol{C} - \boldsymbol{C}_Y)\widehat{\boldsymbol{\Psi}}^{-1}), \tag{2.14}$$

where $\widehat{\boldsymbol{Q}} = T\widehat{\boldsymbol{\omega}}_2\widehat{\boldsymbol{\omega}}_2'/m + \widehat{\boldsymbol{\Psi}}$, and minimize (2.14) with respect to $\widehat{\Psi}(1,1),\ldots,\widehat{\Psi}(N,N)$. A numerical search routine must be used. The resulting values $\widehat{\Psi}(1,1),\ldots,\widehat{\Psi}(N,N)$ are employed at Steps 2 and 3 to create a new $\widehat{\boldsymbol{\omega}}_2$. Steps 2, 3 and 4 are repeated until the differences between successive values of $\widehat{\boldsymbol{\omega}}_2(i,j)$ in $\widehat{\boldsymbol{\omega}}_2 = [\widehat{\boldsymbol{\omega}}_2(i,j)]$ and $\widehat{\Psi}(i,i)$ are negligible.

2.2.3. Case 3. The full model

In this case, the log-likelihood function of $\mathrm{vec}(\boldsymbol{Z}')$ is

$$\log f(\mathrm{vec}(\boldsymbol{Z}')) = -\frac{TN}{2}\log(2\pi) - \frac{1}{2}\log|\widetilde{\boldsymbol{\Sigma}}| - \frac{1}{2}\{\mathrm{vec}(\boldsymbol{Z}')\}'\widetilde{\boldsymbol{\Sigma}}^{-1}\mathrm{vec}(\boldsymbol{Z}').$$

Lemma 1. *If* $\boldsymbol{G}'\boldsymbol{G} = (T/m)I_m$, *then*
(a) $|\widetilde{\boldsymbol{\Sigma}}| = |\boldsymbol{Q}|^m|\boldsymbol{A}|^{T-m}$, *where* $\boldsymbol{Q} = \boldsymbol{A} + (T/m)\boldsymbol{B}$,
(b) $\widetilde{\boldsymbol{\Sigma}}^{-1} = \boldsymbol{I}_T \otimes \boldsymbol{A}^{-1} + \boldsymbol{G}\boldsymbol{G}' \otimes \boldsymbol{U}$, *where* $\boldsymbol{U} = (m/T)(\boldsymbol{Q}^{-1} - \boldsymbol{A}^{-1})$.
(c) $\{vec(\boldsymbol{Z}')\}'\widetilde{\boldsymbol{\Sigma}}^{-1}vec(\boldsymbol{Z}') = tr(\boldsymbol{Z}\boldsymbol{A}^{-1}\boldsymbol{Z}') + tr(\boldsymbol{Z}\boldsymbol{U}\boldsymbol{Z}'\boldsymbol{G}\boldsymbol{G}')$.

Proof. See the online supplementary material for a proof.

Recall that $\boldsymbol{Y} = (\boldsymbol{G}'\boldsymbol{G})^{-1}\boldsymbol{G}'\boldsymbol{Z}$, and $\boldsymbol{C}_Y = \boldsymbol{Y}'\boldsymbol{Y}/m$, and let $\boldsymbol{C} = \boldsymbol{Z}'\boldsymbol{Z}/T$. Using (2.1), Lemma 1 (a) and (c), the log likelihood function of $\boldsymbol{\omega}_1$, $\boldsymbol{\omega}_2$, $\boldsymbol{\omega}_3$ and $\boldsymbol{\Psi}$ given $\boldsymbol{Z}$ is

$$\begin{aligned}
&\ln L(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \boldsymbol{\omega}_3, \boldsymbol{\Psi}) \\
&= -\frac{TN}{2}\ln(2\pi) - \frac{m}{2}\ln|\boldsymbol{Q}| - \frac{T-m}{2}\ln|\boldsymbol{A}| - \frac{T^2}{2m}\mathrm{tr}(\boldsymbol{C}_Y\boldsymbol{U}) - \frac{T}{2}\mathrm{tr}(\boldsymbol{C}\boldsymbol{A}^{-1}) \\
&= -\frac{TN}{2}\ln(2\pi) - \frac{m}{2}\ln|\boldsymbol{Q}| - \frac{T-m}{2}\ln|\boldsymbol{A}| - \frac{T}{2}\mathrm{tr}(\boldsymbol{C}_Y\boldsymbol{Q}^{-1}) \\
&\quad -\frac{T}{2}\mathrm{tr}[(\boldsymbol{C} - \boldsymbol{C}_Y)\boldsymbol{A}^{-1}].
\end{aligned}$$

Thus the objective function can be written as

$$\begin{aligned}
-2\ln L(\boldsymbol{\theta}) = {}& TN\ln(2\pi) + m\ln|\boldsymbol{Q}| + (T-m)\ln|\boldsymbol{A}| + T\mathrm{tr}(\boldsymbol{C}_Y\boldsymbol{Q}^{-1}) \\
&+ T\mathrm{tr}[(\boldsymbol{C} - \boldsymbol{C}_Y)\boldsymbol{A}^{-1}],
\end{aligned} \tag{2.15}$$

where $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{\omega}_1\boldsymbol{\omega}_1'\boldsymbol{H}' + \boldsymbol{\Psi}$, $\boldsymbol{B} = \boldsymbol{\omega}_2\boldsymbol{\omega}_2' + \boldsymbol{H}\boldsymbol{\omega}_3\boldsymbol{\omega}_3'\boldsymbol{H}'$, $\boldsymbol{Q} = \boldsymbol{A} + T\boldsymbol{B}/m$, and we minimize (2.15) with respect of $\boldsymbol{\theta} = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \boldsymbol{\omega}_3, \boldsymbol{\Psi})$ to obtain the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$.

2.2.4. Case 4. $\boldsymbol{\omega}_3 = 0$

Here there is no interaction between the row and column constraints, and the model is

$$\boldsymbol{Z} = \boldsymbol{F}_1\boldsymbol{\omega}_1'\boldsymbol{H}' + \boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\omega}_2' + \boldsymbol{E}. \tag{2.16}$$

The associated objective function is

$$\begin{aligned}
-2\ln L(\boldsymbol{\theta}) = {}& TN\ln(2\pi) + m\ln|\boldsymbol{Q}| + (T-m)\ln|\boldsymbol{A}| + T\mathrm{tr}(\boldsymbol{C}_Y\boldsymbol{Q}^{-1}) \\
&+ T\mathrm{tr}((\boldsymbol{C} - \boldsymbol{C}_Y)\boldsymbol{A}^{-1}),
\end{aligned} \tag{2.17}$$

where $\boldsymbol{\theta} = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \boldsymbol{\Psi})$, $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{\omega}_1\boldsymbol{\omega}_1'\boldsymbol{H}' + \boldsymbol{\Psi}$, and $\boldsymbol{Q} = \boldsymbol{A} + T\boldsymbol{\omega}_2\boldsymbol{\omega}_2'/m$. We minimize (2.17) to obtain the estimate $\hat{\boldsymbol{\theta}}$.

2.2.5. Initial estimates for Cases 3 and 4

For Cases 1 and 2, the MLE are computed by iterated procedures. For Cases 3 and 4, no iterative procedure is available, and the MLE must be obtained by some numerical optimization method with certain initial estimates. We use the LS estimates of Subsection 2.1 as the initial estimates.

2.3. Estimation of latent factors for maximum likelihood approach

Treating the ML estimates of $\boldsymbol{\omega}_i$ as given, we can estimate the latent factors $\boldsymbol{F}_i$ by using the weighted least squares method. Specifically, given $\boldsymbol{\omega}_1$, $\boldsymbol{\omega}_2$, and $\boldsymbol{\omega}_3$, the weighted least squares estimates of $\boldsymbol{F}_1$, $\boldsymbol{F}_2$, and $\boldsymbol{F}_3$ can be obtained by minimizing $f(\boldsymbol{F}_1,\boldsymbol{F}_2,\boldsymbol{F}_3) = \mathrm{tr}(\boldsymbol{E}\boldsymbol{\Psi}^{-1}\boldsymbol{E}') = \mathrm{tr}((\boldsymbol{Z}-\boldsymbol{F}_1\boldsymbol{\omega}_1'\boldsymbol{H}'-\boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\omega}_2'-\boldsymbol{G}\boldsymbol{F}_3\boldsymbol{\omega}_3'\boldsymbol{H}')\boldsymbol{\Psi}^{-1}(\boldsymbol{Z}-\boldsymbol{F}_1\boldsymbol{\omega}_1'\boldsymbol{H}'-\boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\omega}_2'-\boldsymbol{G}\boldsymbol{F}_3\boldsymbol{\omega}_3'\boldsymbol{H}')')$. Taking the partial derivative of $f(\boldsymbol{F}_1,\boldsymbol{F}_2,\boldsymbol{F}_3)$ with respect to $\boldsymbol{F}_1$, and equating the result to zero, we obtain

$$\begin{aligned}\frac{\partial f}{\partial \boldsymbol{F}_1} &= \frac{\partial}{\partial \boldsymbol{F}_1}\mathrm{tr}(-2\boldsymbol{F}_1\boldsymbol{\omega}_1'\boldsymbol{H}'\boldsymbol{\Psi}^{-1}(\boldsymbol{Z}-\boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\omega}_2'-\boldsymbol{G}\boldsymbol{F}_3\boldsymbol{\omega}_3'\boldsymbol{H}')'+\boldsymbol{F}_1\boldsymbol{\omega}_1'\boldsymbol{H}'\boldsymbol{\Psi}^{-1}\boldsymbol{H}\boldsymbol{\omega}_1\boldsymbol{F}_1')\\ &= -2(\boldsymbol{Z}-\boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\omega}_2'-\boldsymbol{G}\boldsymbol{F}_3\boldsymbol{\omega}_3'\boldsymbol{H}')\boldsymbol{\Psi}^{-1}\boldsymbol{H}\boldsymbol{\omega}_1+2\boldsymbol{F}_1\boldsymbol{\omega}_1'\boldsymbol{H}'\boldsymbol{\Psi}^{-1}\boldsymbol{H}\boldsymbol{\omega}_1\\ &= 0. \end{aligned}\tag{2.18}$$

The second equality follows from the fact that

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{X}}\mathrm{tr}(\boldsymbol{A}\boldsymbol{X}) &= \boldsymbol{A}',\\ \frac{\partial}{\partial \boldsymbol{X}}\mathrm{tr}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}'\boldsymbol{B}) &= \boldsymbol{B}\boldsymbol{X}\boldsymbol{A}+\boldsymbol{B}'\boldsymbol{X}\boldsymbol{A}'.\end{aligned}$$

Equation (2.18) implies that

$$\boldsymbol{F}_1 = (\boldsymbol{Z}-\boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\omega}_2'-\boldsymbol{G}\boldsymbol{F}_3\boldsymbol{\omega}_3'\boldsymbol{H}')\boldsymbol{\Psi}^{-1}\boldsymbol{H}\boldsymbol{\omega}_1(\boldsymbol{\omega}_1'\boldsymbol{H}'\boldsymbol{\Psi}^{-1}\boldsymbol{H}\boldsymbol{\omega}_1)^{-1}. \tag{2.19}$$

Similarly,

$$\begin{aligned}\frac{\partial f}{\partial \boldsymbol{F}_2} &= \frac{\partial}{\partial \boldsymbol{F}_2}\mathrm{tr}(-2\boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\omega}_2'\boldsymbol{\Psi}^{-1}(\boldsymbol{Z}-\boldsymbol{F}_1\boldsymbol{\omega}_1'\boldsymbol{H}'-\boldsymbol{G}\boldsymbol{F}_3\boldsymbol{\omega}_3'\boldsymbol{H}')'+\boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\omega}_2'\boldsymbol{\Psi}^{-1}\boldsymbol{\omega}_2\boldsymbol{F}_2'\boldsymbol{G}')\\ &= -2\boldsymbol{G}'(\boldsymbol{Z}-\boldsymbol{F}_1\boldsymbol{\omega}_1'\boldsymbol{H}'-\boldsymbol{G}\boldsymbol{F}_3\boldsymbol{\omega}_3'\boldsymbol{H}')\boldsymbol{\Psi}^{-1}\boldsymbol{\omega}_2+2\boldsymbol{G}'\boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\omega}_2'\boldsymbol{\Psi}^{-1}\boldsymbol{\omega}_2\\ &= 0.\end{aligned}$$

If $\widetilde{\boldsymbol{G}} = (\boldsymbol{G}'\boldsymbol{G})^{-1}\boldsymbol{G}'$ and $\bar{\boldsymbol{G}} = G(\boldsymbol{G}'\boldsymbol{G})^{-1}\boldsymbol{G}'$, then

$$\boldsymbol{F}_2 = \widetilde{\boldsymbol{G}}(\boldsymbol{Z}-\boldsymbol{F}_1\boldsymbol{\omega}_1'\boldsymbol{H}'-\boldsymbol{G}\boldsymbol{F}_3\boldsymbol{\omega}_3'\boldsymbol{H}')\boldsymbol{\Psi}^{-1}\boldsymbol{\omega}_2(\boldsymbol{\omega}_2'\boldsymbol{\Psi}^{-1}\boldsymbol{\omega}_2)^{-1}. \tag{2.20}$$

Thirdly,

$$\begin{aligned}\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{F}_3} &= \frac{\partial}{\partial \boldsymbol{F}_3}\mathrm{tr}(-\boldsymbol{G}\boldsymbol{F}_3\boldsymbol{\omega}_3'\boldsymbol{H}'\boldsymbol{\Psi}^{-1}(\boldsymbol{Z}-\boldsymbol{F}_1\boldsymbol{\omega}_1'\boldsymbol{H}'-\boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\omega}_2')'\\ &\quad +\boldsymbol{G}\boldsymbol{F}_3\boldsymbol{\omega}_3'\boldsymbol{H}'\boldsymbol{\Psi}^{-1}\boldsymbol{H}\boldsymbol{\omega}_3\boldsymbol{F}_3'\boldsymbol{G}')\\ &= -2\boldsymbol{G}'(\boldsymbol{Z}-\boldsymbol{F}_1\boldsymbol{\omega}_1'\boldsymbol{H}'-\boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\omega}_2')\boldsymbol{\Psi}^{-1}\boldsymbol{H}\boldsymbol{\omega}_3+2\boldsymbol{G}'\boldsymbol{G}\boldsymbol{F}_3\boldsymbol{\omega}_3'\boldsymbol{H}'\boldsymbol{\Psi}^{-1}\boldsymbol{H}\boldsymbol{\omega}_3\\ &= 0.\end{aligned}$$

Therefore,

$$\boldsymbol{F}_3 = \widetilde{\boldsymbol{G}}(\boldsymbol{Z} - \boldsymbol{F}_1\boldsymbol{\omega}_1'\boldsymbol{H}' - \boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\omega}_2')\boldsymbol{\Psi}^{-1}\boldsymbol{H}\boldsymbol{\omega}_3(\boldsymbol{\omega}_3'\boldsymbol{H}'\boldsymbol{\Psi}^{-1}\boldsymbol{H}\boldsymbol{\omega}_3)^{-1}. \tag{2.21}$$

Using (2.11) and letting $\boldsymbol{\Gamma}_{12} = \boldsymbol{\Gamma}_{21}' = \boldsymbol{\omega}_1'\boldsymbol{H}'\boldsymbol{\Psi}^{-1}\boldsymbol{\omega}_2$, $\boldsymbol{\Gamma}_{13} = \boldsymbol{\Gamma}_{31}' = \boldsymbol{\omega}_1'\boldsymbol{H}'\boldsymbol{\Psi}^{-1}\boldsymbol{H}\boldsymbol{\omega}_3$, $\boldsymbol{\Gamma}_{23} = \boldsymbol{\Gamma}_{32}' = \boldsymbol{\omega}_2'\boldsymbol{\Psi}^{-1}\boldsymbol{H}\boldsymbol{\omega}_3$, $\boldsymbol{\Gamma}_{01} = \boldsymbol{Z}\boldsymbol{\Psi}^{-1}\boldsymbol{H}\boldsymbol{\omega}_1$, $\boldsymbol{\Gamma}_{02} = \boldsymbol{Z}\boldsymbol{\Psi}^{-1}\boldsymbol{\omega}_2$, and $\boldsymbol{\Gamma}_{03} = \boldsymbol{Z}\boldsymbol{\Psi}^{-1}\boldsymbol{H}\boldsymbol{\omega}_3$, (2.19), (2.20), and (2.21) become

$$\boldsymbol{F}_1 = (\boldsymbol{\Gamma}_{01} - \boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\Gamma}_{21} - \boldsymbol{G}\boldsymbol{F}_3\boldsymbol{\Gamma}_{31})\boldsymbol{\Gamma}_1^{-1}, \tag{2.22}$$

$$\boldsymbol{F}_2 = (\widetilde{\boldsymbol{G}}\boldsymbol{\Gamma}_{02} - \widetilde{\boldsymbol{G}}\boldsymbol{F}_1\boldsymbol{\Gamma}_{12} - \boldsymbol{F}_3\Gamma_{32})\boldsymbol{\Gamma}_2^{-1}, \tag{2.23}$$

$$\boldsymbol{F}_3 = (\widetilde{\boldsymbol{G}}\boldsymbol{\Gamma}_{03} - \widetilde{\boldsymbol{G}}\boldsymbol{F}_1\Gamma_{13} - \boldsymbol{F}_2\Gamma_{23})\boldsymbol{\Gamma}_3^{-1}. \tag{2.24}$$

Multiplying both sides of (2.24) by Γ_3 we obtain

$$\boldsymbol{F}_3\boldsymbol{\Gamma}_3 = \widetilde{\boldsymbol{G}}\boldsymbol{\Gamma}_{03} - \widetilde{\boldsymbol{G}}\boldsymbol{F}_1\boldsymbol{\Gamma}_{13} - \boldsymbol{F}_2\boldsymbol{\Gamma}_{23}. \tag{2.25}$$

Plugging (2.23) into (2.25) we have

$$\begin{aligned}\boldsymbol{F}_3\boldsymbol{\Gamma}_3 &= \widetilde{\boldsymbol{G}}\boldsymbol{\Gamma}_{03} - \widetilde{\boldsymbol{G}}\boldsymbol{F}_1\boldsymbol{\Gamma}_{13} - (\widetilde{\boldsymbol{G}}\boldsymbol{\Gamma}_{02} - \widetilde{\boldsymbol{G}}\boldsymbol{F}_1\boldsymbol{\Gamma}_{12} - \boldsymbol{F}_3\boldsymbol{\Gamma}_{32})\boldsymbol{\Gamma}_2^{-1}\boldsymbol{\Gamma}_{23}\\ &= \widetilde{\boldsymbol{G}}\{\boldsymbol{F}_1(\boldsymbol{\Gamma}_{12}\Gamma_2^{-1}\boldsymbol{\Gamma}_{23} - \boldsymbol{\Gamma}_{13}) + \boldsymbol{\Gamma}_{03} - \boldsymbol{\Gamma}_{02}\Gamma_2^{-1}\boldsymbol{\Gamma}_{23}\} + \boldsymbol{F}_3\boldsymbol{\Gamma}_{32}\boldsymbol{\Gamma}_2^{-1}\boldsymbol{\Gamma}_{23}.\end{aligned}$$

Subtracting both sides by $\boldsymbol{F}_3\boldsymbol{\Gamma}_{32}\boldsymbol{\Gamma}_2^{-1}\boldsymbol{\Gamma}_{23}$, and post-multiplying by $\boldsymbol{\Delta}_{32} = (\boldsymbol{\Gamma}_3 - \boldsymbol{\Gamma}_{32}\boldsymbol{\Gamma}_2^{-1}\boldsymbol{\Gamma}_{23})^{-1}$ we obtain

$$\boldsymbol{F}_3 = \widetilde{\boldsymbol{G}}\{\boldsymbol{F}_1(\boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_2^{-1}\boldsymbol{\Gamma}_{23} - \boldsymbol{\Gamma}_{13}) + \boldsymbol{\Gamma}_{03} - \boldsymbol{\Gamma}_{02}\boldsymbol{\Gamma}_2^{-1}\boldsymbol{\Gamma}_{23}\}\boldsymbol{\Delta}_{32}. \tag{2.26}$$

Similarly, multiplying both sides of (2.23) by $\boldsymbol{\Gamma}_2$ we get

$$\boldsymbol{F}_2\boldsymbol{\Gamma}_2 = \widetilde{\boldsymbol{G}}\boldsymbol{\Gamma}_{02} - \widetilde{\boldsymbol{G}}\boldsymbol{F}_1\boldsymbol{\Gamma}_{12} - \boldsymbol{F}_3\boldsymbol{\Gamma}_{32}. \tag{2.27}$$

Plugging (2.24) into (2.27) we have

$$\begin{aligned}\boldsymbol{F}_2\Gamma_2 &= \widetilde{\boldsymbol{G}}\boldsymbol{\Gamma}_{02} - \widetilde{\boldsymbol{G}}\boldsymbol{F}_1\boldsymbol{\Gamma}_{12} - (\widetilde{\boldsymbol{G}}\boldsymbol{\Gamma}_{03} - \widetilde{\boldsymbol{G}}\boldsymbol{F}_1\boldsymbol{\Gamma}_{13} - \boldsymbol{F}_2\boldsymbol{\Gamma}_{23})\boldsymbol{\Gamma}_3^{-1}\boldsymbol{\Gamma}_{32}\\ &= \widetilde{\boldsymbol{G}}\{\boldsymbol{F}_1(\boldsymbol{\Gamma}_{13}\boldsymbol{\Gamma}_3^{-1}\boldsymbol{\Gamma}_{32} - \boldsymbol{\Gamma}_{12}) + \boldsymbol{\Gamma}_{02} - \boldsymbol{\Gamma}_{03}\boldsymbol{\Gamma}_3^{-1}\boldsymbol{\Gamma}_{32}\} + \boldsymbol{F}_2\boldsymbol{\Gamma}_{23}\boldsymbol{\Gamma}_3^{-1}\boldsymbol{\Gamma}_{32}.\end{aligned}$$

Subtracting both sides by $\boldsymbol{F}_2\boldsymbol{\Gamma}_{23}\boldsymbol{\Gamma}_3^{-1}\boldsymbol{\Gamma}_{32}$, and post-multiplying by $\boldsymbol{\Delta}_{23} = (\boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_{23}\boldsymbol{\Gamma}_3^{-1}\boldsymbol{\Gamma}_{32})^{-1}$ we obtain

$$\boldsymbol{F}_2 = \widetilde{\boldsymbol{G}}\{\boldsymbol{F}_1(\boldsymbol{\Gamma}_{13}\boldsymbol{\Gamma}_3^{-1}\boldsymbol{\Gamma}_{32} - \boldsymbol{\Gamma}_{12}) + \boldsymbol{\Gamma}_{02} - \boldsymbol{\Gamma}_{03}\boldsymbol{\Gamma}_3^{-1}\boldsymbol{\Gamma}_{32}\}\boldsymbol{\Delta}_{23}. \tag{2.28}$$

Now, multiplying both sides of (2.22) by $\boldsymbol{\Gamma}_1$, we have

$$\boldsymbol{F}_1\boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_{01} - \boldsymbol{G}\boldsymbol{F}_2\boldsymbol{\Gamma}_{21} - \boldsymbol{G}\boldsymbol{F}_3\boldsymbol{\Gamma}_{31}. \tag{2.29}$$

Plugging (2.26) and (2.28) into (2.29), we obtain

$$\begin{aligned}\boldsymbol{F}_1\boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_{01} - \bar{\boldsymbol{G}}\{\boldsymbol{F}_1(\boldsymbol{\Gamma}_{13}\boldsymbol{\Gamma}_3^{-1}\boldsymbol{\Gamma}_{32} - \boldsymbol{\Gamma}_{12}) + \boldsymbol{\Gamma}_{02} - \boldsymbol{\Gamma}_{03}\boldsymbol{\Gamma}_3^{-1}\boldsymbol{\Gamma}_{32}\}\boldsymbol{\Delta}_{23}\boldsymbol{\Gamma}_{21} \\ -\bar{\boldsymbol{G}}\{\boldsymbol{F}_1(\boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_2^{-1}\boldsymbol{\Gamma}_{23} - \boldsymbol{\Gamma}_{13}) + \boldsymbol{\Gamma}_{03} - \boldsymbol{\Gamma}_{02}\boldsymbol{\Gamma}_2^{-1}\boldsymbol{\Gamma}_{23}\}\boldsymbol{\Delta}_{32}\boldsymbol{\Gamma}_{31}. \qquad (2.30)\end{aligned}$$

Pre-multiplying both sides of (2.30) by $\boldsymbol{G}'$, and noting that $\boldsymbol{G}'\bar{\boldsymbol{G}} = \boldsymbol{G}'$, we have

$$\begin{aligned}\boldsymbol{G}'\boldsymbol{F}_1\boldsymbol{\Gamma}_1 = \boldsymbol{G}'\boldsymbol{\Gamma}_{01} - \boldsymbol{G}'\{\boldsymbol{F}_1(\boldsymbol{\Gamma}_{13}\boldsymbol{\Gamma}_3^{-1}\boldsymbol{\Gamma}_{32} - \boldsymbol{\Gamma}_{12}) + \boldsymbol{\Gamma}_{02} - \boldsymbol{\Gamma}_{03}\boldsymbol{\Gamma}_3^{-1}\boldsymbol{\Gamma}_{32}\}\boldsymbol{\Delta}_{23}\boldsymbol{\Gamma}_{21} \\ -G'\{\boldsymbol{F}_1(\boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_2^{-1}\boldsymbol{\Gamma}_{23} - \boldsymbol{\Gamma}_{13}) + \boldsymbol{\Gamma}_{03} - \boldsymbol{\Gamma}_{02}\boldsymbol{\Gamma}_2^{-1}\boldsymbol{\Gamma}_{23}\}\boldsymbol{\Delta}_{32}\boldsymbol{\Gamma}_{31}. \qquad (2.31)\end{aligned}$$

One solution to (2.31) is

$$\begin{aligned}\boldsymbol{F}_1\boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}_{01} - \{\boldsymbol{F}_1(\boldsymbol{\Gamma}_{13}\boldsymbol{\Gamma}_3^{-1}\boldsymbol{\Gamma}_{32} - \boldsymbol{\Gamma}_{12}) + \boldsymbol{\Gamma}_{02} - \boldsymbol{\Gamma}_{03}\boldsymbol{\Gamma}_3^{-1}\boldsymbol{\Gamma}_{32}\}\boldsymbol{\Delta}_{23}\boldsymbol{\Gamma}_{21} \\ -\{\boldsymbol{F}_1(\boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_2^{-1}\boldsymbol{\Gamma}_{23} - \boldsymbol{\Gamma}_{13}) + \boldsymbol{\Gamma}_{03} - \boldsymbol{\Gamma}_{02}\boldsymbol{\Gamma}_2^{-1}\boldsymbol{\Gamma}_{23}\}\boldsymbol{\Delta}_{32}\boldsymbol{\Gamma}_{31}. \qquad (2.32)\end{aligned}$$

From (2.32), we obtain

$$\begin{aligned}\boldsymbol{F}_1 = \big\{\boldsymbol{\Gamma}_{01} - (\boldsymbol{\Gamma}_{02} - \boldsymbol{\Gamma}_{03}\boldsymbol{\Gamma}_3^{-1}\boldsymbol{\Gamma}_{32})\boldsymbol{\Delta}_{23}\boldsymbol{\Gamma}_{21} - (\boldsymbol{\Gamma}_{03} - \boldsymbol{\Gamma}_{02}\boldsymbol{\Gamma}_2^{-1}\boldsymbol{\Gamma}_{23})\boldsymbol{\Delta}_{32}\boldsymbol{\Gamma}_{31}\big\} \\ \big\{\boldsymbol{\Gamma}_1 + (\boldsymbol{\Gamma}_{13}\boldsymbol{\Gamma}_3^{-1}\boldsymbol{\Gamma}_{32} - \boldsymbol{\Gamma}_{12})\boldsymbol{\Delta}_{23}\boldsymbol{\Gamma}_{21} + (\boldsymbol{\Gamma}_{12}\boldsymbol{\Gamma}_2^{-1}\boldsymbol{\Gamma}_{23} - \boldsymbol{\Gamma}_{13})\boldsymbol{\Delta}_{32}\boldsymbol{\Gamma}_{31}\big\}^{-1}. \quad (2.33)\end{aligned}$$

We use (2.33) to compute $\boldsymbol{F}_1$ first, then use (2.28) to compute $\boldsymbol{F}_2$, and (2.26) to compute $\boldsymbol{F}_3$.

2.4. Model selection

In applications, the data generating process is unknown and one needs to select a proper constrained factor model based on the available data. In particular, the validity of row and/or column constraints should be verified. To this end, we consider the Akaike information criterion (AIC) (Akaike (1974)) for each of the fitted models,

$$AIC = -2\ln L(\hat{\boldsymbol{\theta}}) + 2\lambda,$$

where λ is the number of parameters of the model, and $\hat{\boldsymbol{\theta}}$ is the MLE. Our simulation study and empirical example show that AIC works well in model selection.

Tsai and Tsay (2010) used hypothesis testing to check the validity of column constraints. The testing procedure is complicated for doubly constrained factor models because it involves non-nested hypothesis testing. Thus, the model with only column constraints is not a sub-model of the one with only row constraints.

3. Simulation Study

In this section, we report some finite-sample performance of the MLE and the AIC of Subsections 3.1 and 3.3, respectively. All computations were performed using Fortran code with IMSL subroutines.

3.1. Finite sample properties of the MLE and the LSE

To evaluate the performance of the numerical optimization in finding the MLE discussed in Subsection 2.2.3 for the full model (Case 3), we considered the following data generating process.

MHG∩1: $N = 24$, $r = 2$, $p = 2$, $q = 1$, $s = 3$, $m = 12$, $\boldsymbol{G} = \mathbf{1}_{T/m} \otimes \boldsymbol{I}_m$, where $\mathbf{1}_m$ denotes the $m \times 1$ vector of ones, $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3]$, $\mathbf{h}_1 = \mathbf{1}_{24}$, $\mathbf{h}_2 = [-\mathbf{1}(6), \mathbf{0}(12), \mathbf{1}(6)]'$, $\mathbf{h}_3 = [-\mathbf{1}(6), \mathbf{0}(3), \mathbf{2}(6), \mathbf{0}(3), -\mathbf{1}(6)]'$, and $\mathbf{r}(j)$ denotes a j-dimensional row-vector of integer r, $\boldsymbol{\omega}_1 = \boldsymbol{\Psi}_0^{1/2}\boldsymbol{\Lambda}_1\text{diag}(1.2, 0.6)$, $\boldsymbol{\omega}_2 = \boldsymbol{\Psi}^{1/2}\boldsymbol{\Lambda}_2\text{diag}(0.6, 0.3)$, $\boldsymbol{\omega}_3 = 0.3\boldsymbol{\Psi}_0^{1/2}\boldsymbol{\Lambda}_3$, $\text{vec}(\boldsymbol{\Lambda}_1)$, $\text{vec}(\boldsymbol{\Lambda}_2)$, and $\text{vec}(\boldsymbol{\Lambda}_3)$ are independent random vectors from $\mathcal{N}(\mathbf{0}, \boldsymbol{I}_6)$, $\mathcal{N}(\mathbf{0}, \boldsymbol{I}_{48})$, and $\mathcal{N}(\mathbf{0}, \boldsymbol{I}_3)$, respectively, $\boldsymbol{\Psi} = \text{diag}(\Psi(j,j))$, $\Psi(j,j) = 0.1 + 0.2 \times u_i$, and u_i are i.i.d. uniform on [0,1]. Adding 0.1 to the variance avoids near-zero values (see also page 453 of Bai and Li (2012)), and $\boldsymbol{\Psi}_0 = \text{diag}(\psi_0(j,j))$, where $\{\Psi_0(1,1)\}^{-1} = \sum_{j=1}^{N}\{\Psi(j,j)\}^{-1}$, $\{\Psi_0(2,2)\}^{-1} = \sum_{j=1}^{6}\{\Psi(j,j)\}^{-1} + \sum_{j=19}^{24}\{\Psi(j,j)\}^{-1}$, and $\{\Psi_0(3,3)\}^{-1} = \{\Psi_0(2,2)\}^{-1} + 4\sum_{j=10}^{15}\{\Psi(j,j)\}^{-1}$.

We computed MLE by minimizing (2.15) using the optimizing subroutine DNCONF from FORTRAN's IMSL library. The least squares estimates of Subsection 2.1 were used as the initial values of the subroutine DNCONF. We took sample sizes $T = 24, 36, 60, 120, 240, 480$, and 960. To measure the accuracy between $\widehat{\boldsymbol{\omega}}_i$ and $\boldsymbol{\omega}_i$, for $i = 1, 2, 3$, we computed the smallest nonzero canonical correlation between them. Canonical correlation is widely used as a measure of goodness-of-fit in factor analysis; see, for example, Doz, Giannone, and Reichlin (2006), Goyal, Perignon, and Villa (2008), and Bai and Li (2012). For the estimated variances of e_i, we calculated the squared correlation between $\text{diag}(\widehat{\boldsymbol{\Psi}})$ and $\text{diag}(\boldsymbol{\Psi})$. Table 1 reports the average canonical correlations based on 1,000 repetitions for each sample size T. For comparison purpose, we also report the results for LSE in Table 1. From Table 1, both the MLEs and the LSEs show convergence to their corresponding true values as the sample size increases. In general, the MLE performs better than the LSE, except for $T = 24$.

Table 1. Finite Sample Performance of the Maximum Likelihood Estimates (MLE) and the Least Square Estimates (LSE).

		MLE				LSE			
N	T	$\boldsymbol{\omega}_1$	$\boldsymbol{\omega}_2$	$\boldsymbol{\omega}_3$	Ψ	$\boldsymbol{\omega}_1$	$\boldsymbol{\omega}_2$	$\boldsymbol{\omega}_3$	Ψ
24	24	0.6549	0.4055	0.5494	0.2293	0.7079	0.5399	0.5498	0.4636
24	36	0.8212	0.7646	0.6133	0.5362	0.7586	0.6458	0.5511	0.5640
24	60	0.8569	0.8626	0.6554	0.6661	0.8180	0.7549	0.5813	0.6745
24	120	0.8848	0.9238	0.7569	0.7996	0.8530	0.8332	0.5925	0.7837
24	240	0.9075	0.9601	0.8350	0.8925	0.8814	0.8644	0.6024	0.8508
24	480	0.9340	0.9762	0.8974	0.9440	0.9012	0.8782	0.6043	0.8866
24	960	0.9429	0.9834	0.9362	0.9706	0.9032	0.8854	0.6031	0.9069

3.2. Performance of AIC

To avoid the complications of non-nested hypothesis testing, we used AIC to check the adequacy of the column and/or row constraints. In this subsection, we consider the finite sample performance of the AIC in selecting the data generating model among Cases 1−4 below. The data generating models considered were

MH1: $\boldsymbol{\omega}_2 = \boldsymbol{\omega}_3 = 0$, and $\boldsymbol{\omega}_1$ is the same as that of model MHG∩1 (corresponding to Case 1 of Subsection 2.2.1).

MG1: $\boldsymbol{\omega}_1 = \boldsymbol{\omega}_3 = 0$, and $\boldsymbol{\omega}_2$ is the same as that of model MHG∩1 (corresponding to Case 2 of Subsection 2.2.2).

MHG1: $\boldsymbol{\omega}_3 = 0$, and $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ are the same as those of model MHG∩1 (corresponding to Case 4 of Subsection 2.2.4).

MHG∩1: $N = 6$, $r = 2$, $p = 2$, $q = 1$, $s = 2$, $m = 12$, $\boldsymbol{G} = \mathbf{1}_{T/m} \otimes \boldsymbol{I}_m$, and $\boldsymbol{H} = \boldsymbol{I}_2 \otimes \mathbf{1}_3$, $\boldsymbol{\omega}_1 = \boldsymbol{\Psi}_0^{1/2}\boldsymbol{\Lambda}_1\text{diag}(0.8, 0.6)$, $\boldsymbol{\omega}_2 = \boldsymbol{\Psi}^{1/2}\boldsymbol{\Lambda}_2\text{diag}(0.5, 0.3)$, $\boldsymbol{\omega}_3 = 0.2\boldsymbol{\Psi}_0^{1/2}\boldsymbol{\Lambda}_3$, $\boldsymbol{\Lambda}_1 = [\Lambda_a, \Lambda_b]$, $\Lambda_a = [1, 3]'$, $\Lambda_b = [3, -1]'$, $\boldsymbol{\Lambda}_2 = [\Lambda_c, \Lambda_d]$, $\Lambda_c = [2, 1, 2, 1, 2, 1]'$, $\Lambda_d = [1, 2, 1, -2, -1, -2]'$, $\boldsymbol{\Lambda}_1 = [4, 3]'$, $\boldsymbol{\Psi} = \text{diag}(0.2)$, and $\boldsymbol{\Psi}_0 = \text{diag}(0.2)$ (corresponding to Case 3 of Subsection 2.2.3).

For singly constrained factor models (Cases 1 and 2), we implemented the estimation procedures described in Subsections 2.2.1 and 2.2.2, respectively. The sample sizes employed were $T = 480$, 960, and 1,920. The experiment went as follows. First, we generated data from the above data generating process. Then, we estimated the parameters of a constrained factor model for orders (r, p, q) that satisfy the conditions of Section 1, where $0 \leq r, p, q \leq 3$. For example, $p < N$, $\max\{r, q\} \leq s < N$, and $q \leq \min\{r, p\}$. For each simulated series, we computed the AIC, and chose the order that corresponded to the smallest AIC. The percentages of the orders determined by the AIC based on 1,000 repetitions

Table 2. The frequencies of the order (r,p,q) selected by AIC The true model considered are models MH1, MG1, MHG1, and MHG∩1.

true model	MH1			MG1			MHG1			MHG∩1		
true order	(2,0,0)			(0,2,0)			(2,2,0)			(2,2,1)		
(r,p,q)\ T	480	960	1,920	480	960	1,920	480	960	1,920	480	960	1,920
(0,1,0)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(0,2,0)	0.000	0.000	0.000	0.877	0.880	0.882	0.000	0.000	0.000	0.000	0.000	0.000
(0,3,0)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(1,0,0)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(1,1,0)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(1,1,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(1,2,0)	0.000	0.000	0.000	0.116	0.108	0.108	0.000	0.000	0.000	0.000	0.000	0.000
(1,2,1)	0.000	0.000	0.000	0.004	0.008	0.005	0.000	0.000	0.000	0.000	0.000	0.000
(1,3,0)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(1,3,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(2,0,0)	0.964	0.967	0.968	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(2,1,0)	0.036	0.031	0.032	0.000	0.000	0.000	0.012	0.000	0.000	0.000	0.000	0.000
(2,1,1)	0.000	0.002	0.000	0.000	0.000	0.000	0.013	0.000	0.000	0.007	0.000	0.000
(2,2,0)	0.000	0.000	0.000	0.003	0.004	0.005	0.763	0.771	0.782	0.071	0.010	0.000
(2,2,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.180	0.229	0.218	0.810	0.878	0.891
(2,2,2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.005	0.007
(2,3,0)	0.000	0.000	0.000	0.000	0.000	0.000	0.032	0.000	0.000	0.099	0.098	0.092
(2,3,1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.009	0.010
(2,3,2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

are reported in Table 2. The results show that the AIC works well in selecting a proper doubly constrained factor model. The performance of AIC also improves with the sample size.

3.3. A comparison with unconstrained factor model

To evaluate if there is, as postulated, an advantage in using prior knowledge of the constraints in data analysis, we conducted the following experiment. The data generating process was the Model of Table 5 of Section 4. First, we generated $T+km$ data points from the true model. For $\boldsymbol{G} = \mathbf{1}_{T/m} \otimes \boldsymbol{I}_m$, let $G_{T+im+j} = G_j$, for $i = 0, \ldots, k-1$, and $j = 1, \ldots, m$, where G_j denotes the j-th row of the matrix $\boldsymbol{G}$. Second, used the first T data points to estimate the doubly constrained factor model to get $\widehat{F}_i$ and $\widehat{\boldsymbol{\omega}}_i$, $i = 1, 2, 3$. Third, for $h = 1, \ldots, km$, we computed $\hat{Z}_{T+h}$, the prediction of Z_{T+h},

$$\widehat{Z}_{T+h} = \widehat{F}_1^{(T+h)} \widehat{\boldsymbol{\omega}}_1' H' + G_{T+h} \widehat{\boldsymbol{F}}_2 \widehat{\boldsymbol{\omega}}_2' + G_{T+h} \widehat{\boldsymbol{F}}_3 \widehat{\boldsymbol{\omega}}_3' \boldsymbol{H}',$$

where $\widehat{F}_1^{(T+h)} = \sum_{j=1}^{T} \widehat{F}_1^{(j)}/T$, for $h = 1, \ldots, km$. Fourth, we computed the forecast errors $\hat{e}_{T+h} = Z_{T+h} - \widehat{Z}_{T+h}$, $h = 1, \ldots, km$. Fifth, we computed the root

Table 3. Averages (standard errors) of 1,000 repetitions of the root mean square errors of the forecasts of the DCF (doubly constrained factor) and the UCF (unconstrained factor) models.

T	480		960		1,920	
model	DCF	UCF	DCF	UCF	DCF	UCF
$k = 1$	0.9457	1.0007	0.9440	0.9936	0.9358	0.9930
	(0.1243)	(0.1087)	(0.1190)	(0.1066)	(0.1218)	(0.1065)
$k = 2$	0.9800	1.0005	0.9773	0.9960	0.9710	0.9922
	(0.1092)	(0.0874)	(0.1018)	(0.0863)	(0.1025)	(0.0882)

mean square error (RMSE) of the forecasts, namely RMSE=$[\mathrm{tr}(\widehat{E}'_{predict}\widehat{E}_{predict})/kmN]^{1/2}$, where $\widehat{E}_{predict} = [\hat{e}'_{T+1}, \cdots, \hat{e}'_{T+km}]'$. For the same data generated, we repeated the above steps by fitting an unconstrained factor (UCF) model $\boldsymbol{Z} = \boldsymbol{F}_1\boldsymbol{\omega}_1'$ to get the corresponding RMSE. We repeated the above exercise 1,000 times to get 1,000 RMSE's for each model. For the DCF model, $m = 12$, $k = 1, 2$, and $r = p = q = 2$ were used. For the UCF model, the results for $r = 3$ are reported. The sample sizes used in the simulation were $T = 480$, 960, and 1,920. The average and standard deviation of the 1,000 RMSE's for these two models are summarized in Table 3. The results show that the DCF model outperforms the UCF model if the data generating process is indeed a DCF model. Note that the forecasting results of UCF models are almost identical for $r = 1, 2$, and 3. For $r = 4$ or $r = 5$, we often encountered some numerical difficulties. Therefore, we report the results for the UCF model with $r = 3$.

4. Application

To demonstrate the application of our proposed model, we considered the total housing starts of the United States, obtained from the U.S. Census Bureau website. The data period is from January 1997 to December 2006, so that we have 120 monthly data for the nine geographical divisions of the U.S. shown in Figure 1. The LOESS regression was applied to the log transformed data before fitting the doubly constrained factor model so as to remove the trend of the series.

To specify the constraint matrix $\boldsymbol{H}$, prior experience or geographical clustering can be helpful. In this instance, we applied hierarchical clustering to the variables to specify $\boldsymbol{H}$. The result was consistent with geographical clustering. Therefore, we employed three groups for the variables (divisions), as follows.

Group 1: "New England", "Middle Atlantic", "East North Central", "West North Central".

Group 2: "South Atlantic", "East South Central", "West South Central".

Group 3: "Mountain", "Pacific".

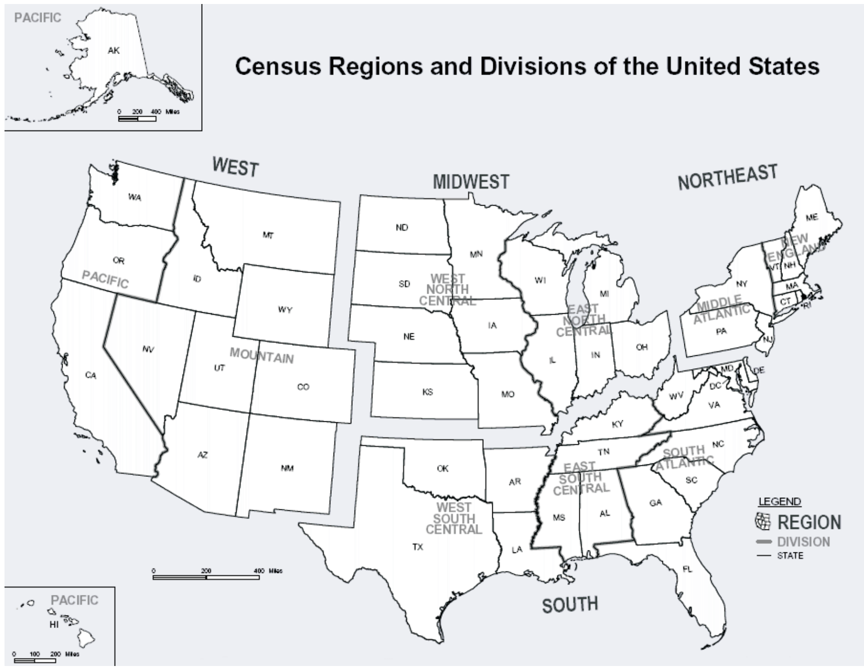


Figure 1. The census regions and divisions of the United States.

The $\boldsymbol{H}$ matrix consists of the indicator variables for the 3 groups. From Figure 1, Group 1 consists of the Northeast and Midwest of the U.S., Group 2 denotes the South, and Group 3 is the West.

The time plots in Figure 2 show that housing starts exhibit strong seasonality of period 12. Therefore, we let $\boldsymbol{G} = \mathbf{1}_{10} \otimes \boldsymbol{I}_{12}$. Consequently, we have $m = 12$, $T = 120$, $N = 9$, and $s = 3$. We considered the DCF models of order (r, p, q) with $0 \leq r, p, q \leq 3$, and $q \leq \min\{r, p\}$. Therefore, a total of 30 models were entertained. Table 4 shows the ranking of these DCF models based on the AIC criterion, where the model of order (0,0,0) is an unrestricted model. Based on the AIC criterion, the doubly constrained factor model of order (2,2,1) was selected with the model of order (2,2,2) as a close second. Model checking showed that the residuals of the fitted DCF model of order (2,2,1) had some minor serial correlations, but those of the model of order (2,2,2) were close to being white noises. Therefore, we adopted the DCF model of order (2,2,2).

Figure 3 shows the time plots of the residuals, $\widehat{\boldsymbol{E}}$, of the DCF(2,2,2) model. The left panel consists of the residuals of least square estimation, the right panel those of the maximum likelihood estimates. The two sets of residuals show similar patterns, but also certain differences. Their sample autocorrelation functions confirm that the residuals have no significant serial dependence; see Figure 4. Table 5 gives the maximum likelihood estimates and the bootstrap standard

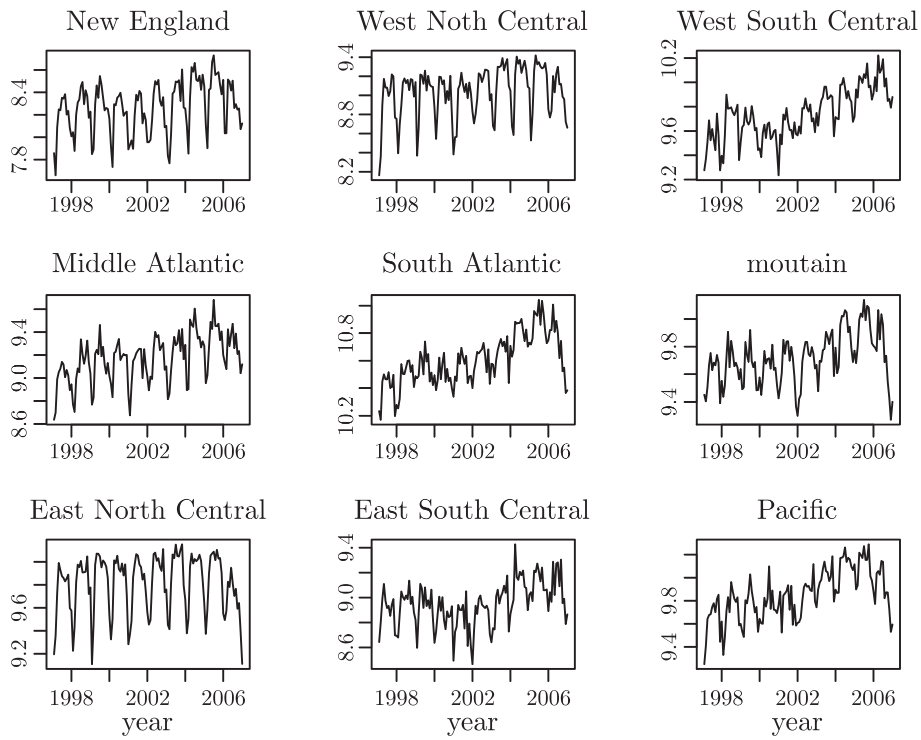


Figure 2. Time plots of monthly housing starts (in logarithms) of nine U.S. divisions: 1997-2006.

errors of the $\boldsymbol{\omega}_i$ for the selected model. The standard errors of $\boldsymbol{\omega}_2$ tend to be larger. The LSE of the $\boldsymbol{\omega}_i$ are given in Table 6. These estimates are different from those of MLE of Table 5 because different normalizations are used. Figure 5 shows the time plots of the fitted common factors. The upper three panels show the common factors obtained by the least squares method whereas the lower three panels give the corresponding results for the maximum likelihood estimation. Care must be exercised in comparing fitted common factors because their scales and orderings are not identifiable. For instance, consider the fitted common factors $\widehat{\boldsymbol{F}}_3$. The orderings seem to be interchanged between the two estimation methods. Overall, the common factors $\widehat{\boldsymbol{F}}_1$ of the maximum likelihood estimation appear to have some seasonality. We return to this point later.

4.1. Discussion

To gain insight into the decomposition of the housing starts implied by the fitted DCF model of order (2,2,2), we consider in detail the results of maximum likelihood estimation. Figures 6 to 8 show the time plots of the decompositions of the housing starts series. The plots in Figure 6 consist of $\boldsymbol{G}\widehat{\boldsymbol{F}}_2\widehat{\boldsymbol{\omega}}_2'$ of Equation(1.1).

Table 4. The rankings of AIC for the proposed constrained factor models.

Model (r,p,q)	AIC	ranks	Model (r,p,q)	AIC	ranks
(0,0,0)	-163.331	24	(3,3,0)	-366.857	10
(0,1,0)	114.146	30	(1,1,1)	-329.666	17
(0,2,0)	-18.905	28	(2,1,1)	-339.405	15
(0,3,0)	-28.427	27	(3,1,1)	-333.443	16
(1,0,0)	89.389	29	(1,2,1)	-374.615	4
(2,0,0)	-68.600	25	(2,3,1)	-373.696	6
(3,0,0)	-65.067	26	(2,3,2)	-367.696	9
(1,1,0)	-254.315	23	(1,3,1)	-363.954	11
(2,1,0)	-267.479	21	(3,2,1)	-375.021	3
(3,1,0)	-261.680	22	(3,2,2)	-374.513	5
(1,2,0)	-321.989	20	(2,2,1)	-383.749	1
(2,3,0)	-372.528	7	(2,2,2)	-380.342	2
(1,3,0)	-363.340	12	(3,3,1)	-367.881	8
(3,2,0)	-323.400	19	(3,3,2)	-361.881	13
(2,2,0)	-329.321	18	(3,3,3)	-355.881	14

Table 5. Maximum likelihood estimates of the doubly constrained factor model of order (2,2,2) for the U.S. housing starts data from 1997 to 2006.

(a) MLE of $\widehat{\boldsymbol{\omega}}_1$

$\boldsymbol{\omega}_1[,1]$	0.3051	0.4518	0.4015
(std. error)	(0.0197)	(0.0316)	(0.0423)
$\boldsymbol{\omega}_1[,2]$	0.0844	-0.0729	-0.1945
(std. error)	(0.0164)	(0.0418)	(0.0465)

(b) MLE of $\widehat{\boldsymbol{\omega}}_2$

$\boldsymbol{\omega}_2[,1]$	0.1317	0.1713	0.3943	0.3437	0.1214	0.3529	0.1641	0.1125	0.1132
(std. error)	(0.2490)	(0.2277)	(0.2685)	(0.2540)	(0.2265)	(0.2741)	(0.2568)	(0.2280)	(0.2086)
$\boldsymbol{\omega}_2[,2]$	0.2151	0.1183	0.0127	0.0123	-0.0846	-0.1398	-0.1906	-0.1661	-0.0347
(std. error)	(0.1120)	(0.0856)	(0.0969)	(0.0762)	(0.1501)	(0.1527)	(0.1517)	(0.1379)	(0.1238)

(c) MLE of $\widehat{\boldsymbol{\omega}}_3$

$\boldsymbol{\omega}_3[,1]$	0.8218	0.5770	0.7214
(std. error)	(0.1860)	(0.1891)	(0.1605)
$\boldsymbol{\omega}_3[,2]$	0.1118	-0.3868	-0.1905
(std. error)	(0.0465)	(0.1316)	(0.0906)

Since the row constraints used are monthly indicator variables, these plots signify the deterministic seasonal pattern of each housing starts series that is orthogonal to the geographical divisions. From the plots, the deterministic seasonality varies from series to series, but those of the East North Central and West North Central are similar. This seems reasonable as these two divisions are the Midwest and share close weather characteristics. New England and Middle Atlantic divisions

Table 6. Least squares estimates of the doubly constrained factor model of order (2,2,2) for the U.S. housing data from 1997 to 2006.

LSE of $\widehat{\boldsymbol{\omega}}_1$

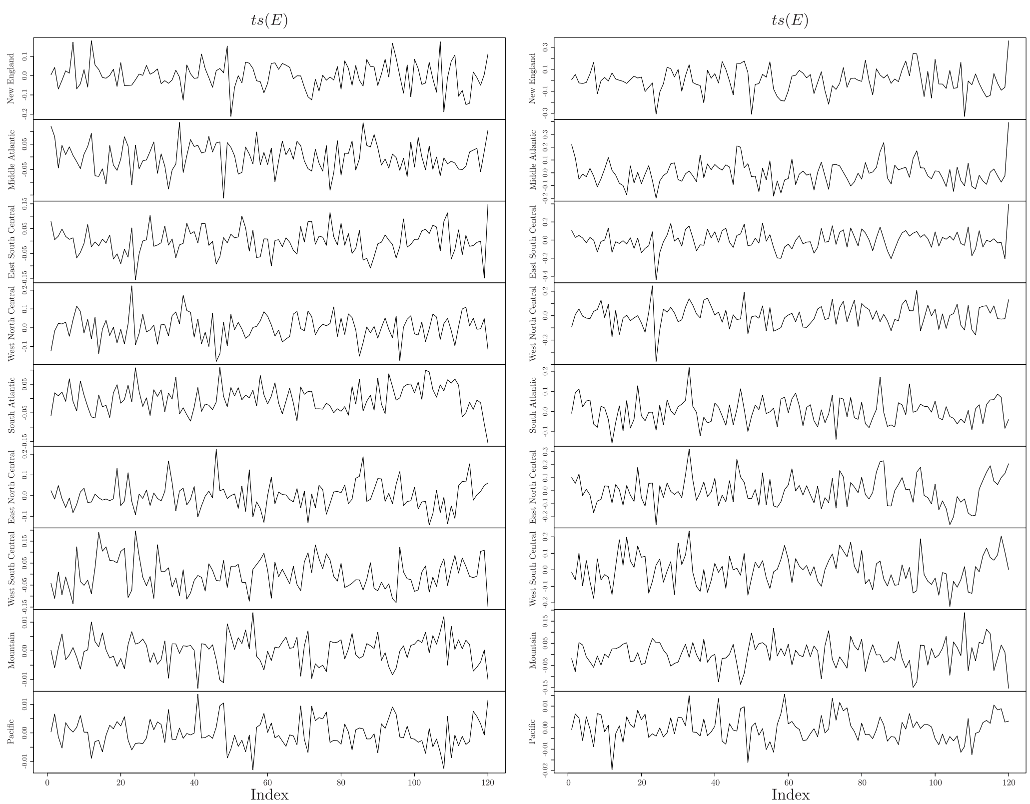
$\boldsymbol{\omega}_1[,1]$	0.0620	0.0547	0.0652
$\boldsymbol{\omega}_1[,2]$	0.0292	0.0186	-0.0434

LSE of $\widehat{\boldsymbol{\omega}}_2$

$\boldsymbol{\omega}_2[,1]$	0.0419	0.0435	-0.0424	-0.0342	0.0149	-0.0225	-0.0007	-0.0056	0.0061
$\boldsymbol{\omega}_2[,2]$	0.0236	-0.0242	0.0025	-0.0012	0.0011	-0.0018	0.0001	-0.0063	0.0063

LSE of $\widehat{\boldsymbol{\omega}}_3$

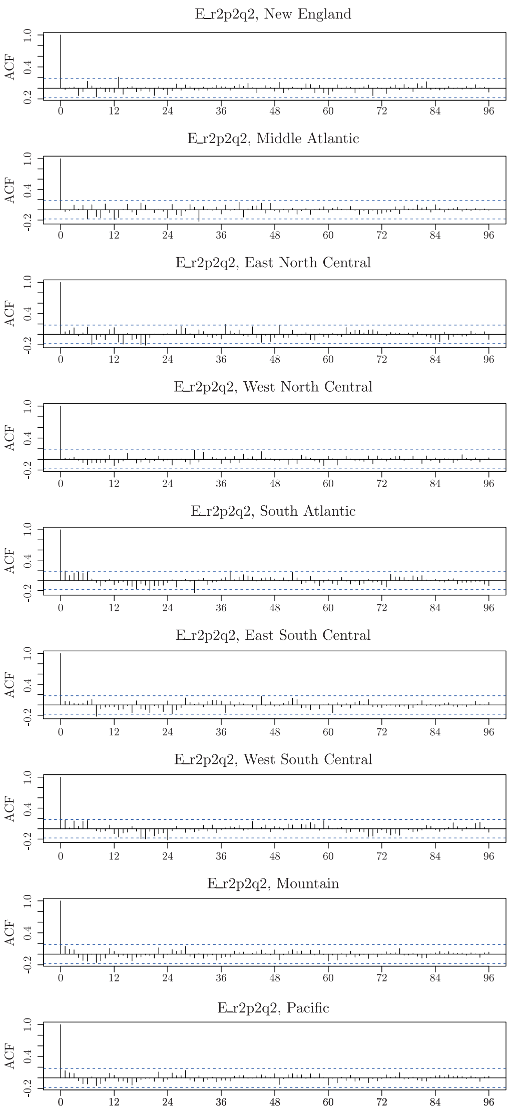
$\boldsymbol{\omega}_3[,1]$	0.1840	0.0841	0.1097
$\boldsymbol{\omega}_3[,2]$	0.0403	-0.0497	-0.0295



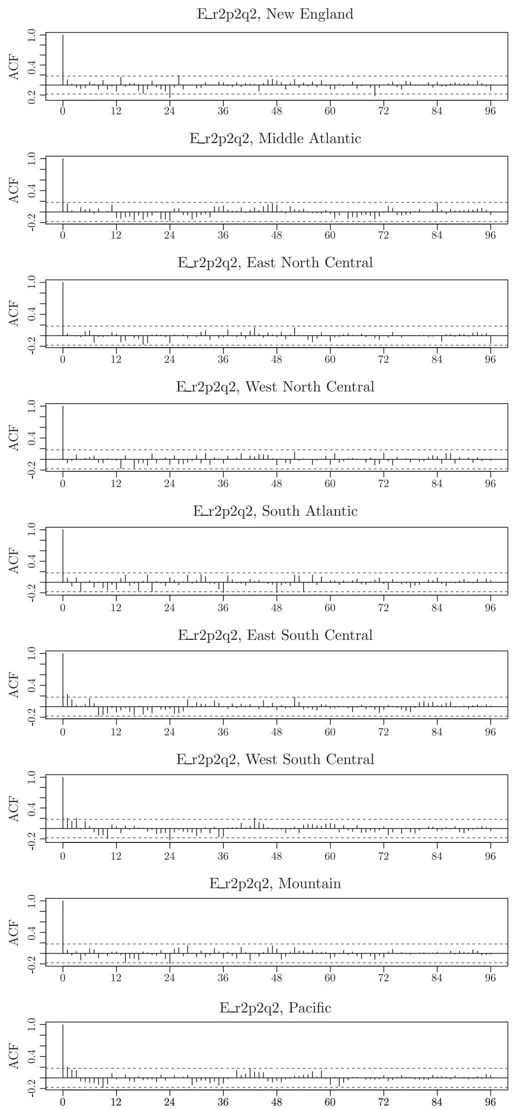
(a) residuals for LSE (b) residuals for MLE

Figure 3. Time series plots for (a) the least squares residuals and (b) the maximum likelihood residuals of the DCF model order $(r, p, q) = (2, 2, 2)$.

have their own deterministic seasonal patterns. The Mountain and West South Central also share similar deterministic seasonal pattern.



(a) ACF for LSE



(b) ACF for MLE

Figure 4. ACF for the residuals of DCF model with order $(r, p, q) = (2, 2, 2)$. Results of the least squares estimation and the maximum likelihood estimation are shown.

The plots in Figure 7 consist of $\widehat{\boldsymbol{F}}_1\widehat{\boldsymbol{\omega}}_1'\boldsymbol{H}'$ of Equation (1.1), which denotes housing variations due to the geographical locations, but orthogonal to the deterministic seasonality. The column constraints essentially pool information within each group to obtain the geographical housing variations. The series in Figure 7 also contain certain seasonality and we believe that they describe the stochastic seasonality of the three geographical groups. These stochastic seasonalities differ from group to group.

Figure 8 shows the interactions $\boldsymbol{G}\widehat{\boldsymbol{F}}_3\widehat{\boldsymbol{\omega}}_3'\boldsymbol{H}'$ between geographical grouping and deterministic seasonality in Equation (1.1). The plots show marked differences between the three interactions. For this particular example, the proposed DCF model is capable of describing the seasonal and geographical patterns of U.S. housing starts. The example demonstrates that the row and column constraints can be used to gain insight into the common structure of a multivariate time series.

5. Concluding Remarks

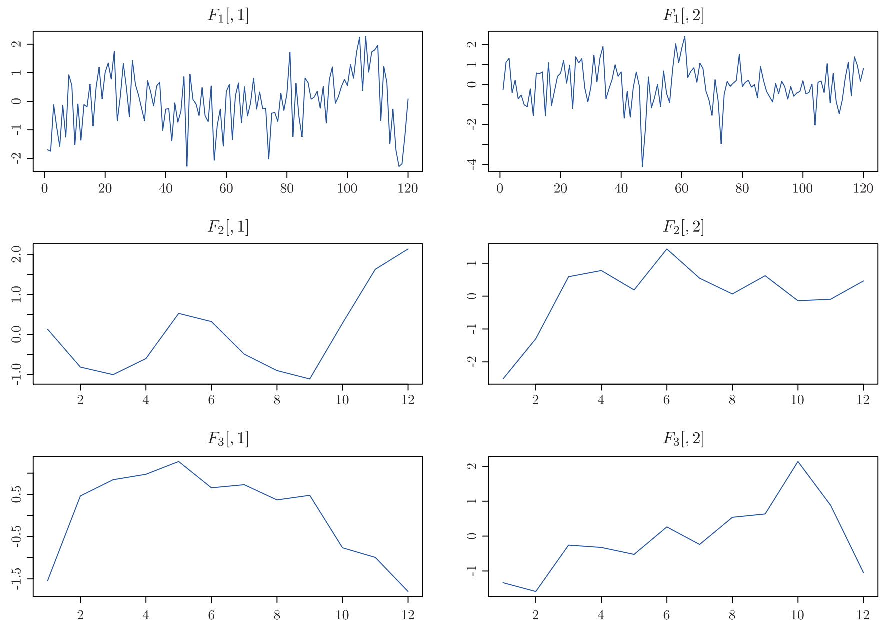
In this paper, we considered both the least squares and maximum likelihood estimations of a doubly constrained factor model, and demonstrated the proposed methods by analyzing nine U.S. monthly housing starts series. The decomposition of the housing starts series shows that the proposed model is capable of describing characteristics of the data. Much work of the constrained factor models, however, remains open. For instance, the maximum likelihood estimation is obtained under the normality assumption. In applications, such an assumption might not be valid and the innovations of Equation (1.1) may contain conditional heteroscedasticity. In addition, we only consider deterministic constraints in the paper. It is of interest to investigate the proposed analysis when the constraints are stochastic. Finally, it is also important to study the DCF models when the number of series N goes to infinity.

Supplementary Materials

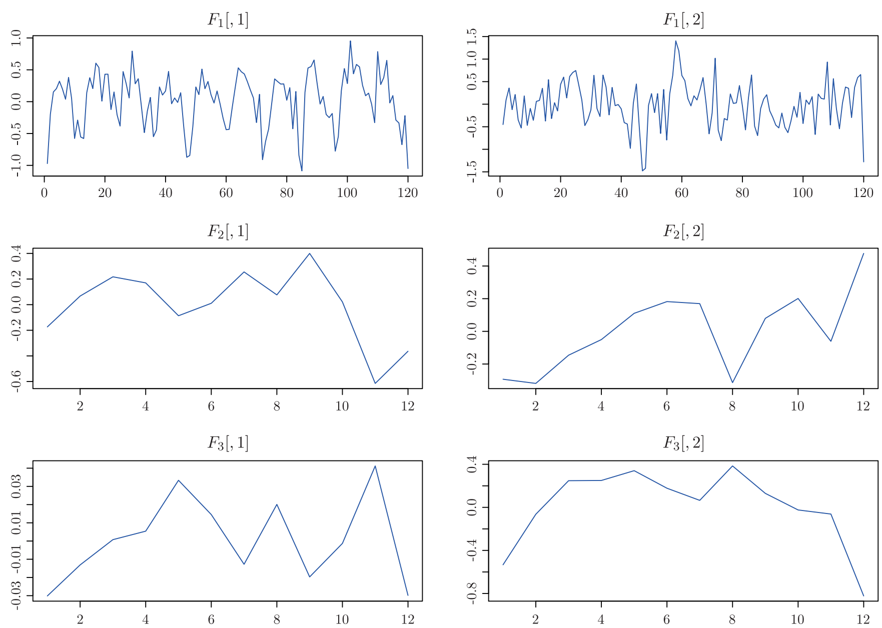
The online Supplement file contains the proofs of the identifiability of $\boldsymbol{\omega}_1$, $\boldsymbol{\omega}_2$, and $\boldsymbol{\omega}_3$ of the proposed model, and Lemma 1.

Acknowledgements

The research of H. Tsai is supported in part by Academia Sinica and the Ministry of Science and Technology (MOST 102-2118-M-001-007 -MY2) of the Republic of China and that of R. S. Tsay is supported in part by the Booth School of Business, University of Chicago. The authors would like to thank three anonymous referees, an associate editor and the Co-Editor for their helpful comments and suggestions.



(a) $\hat{F}_i$ for LSE



(b) $\hat{F}_i$ for MLE

Figure 5. Time series plots of common factors for a DCF model of order $(r, p, q) = (2, 2, 2)$ via least squares estimation and maximum likelihood estimation.

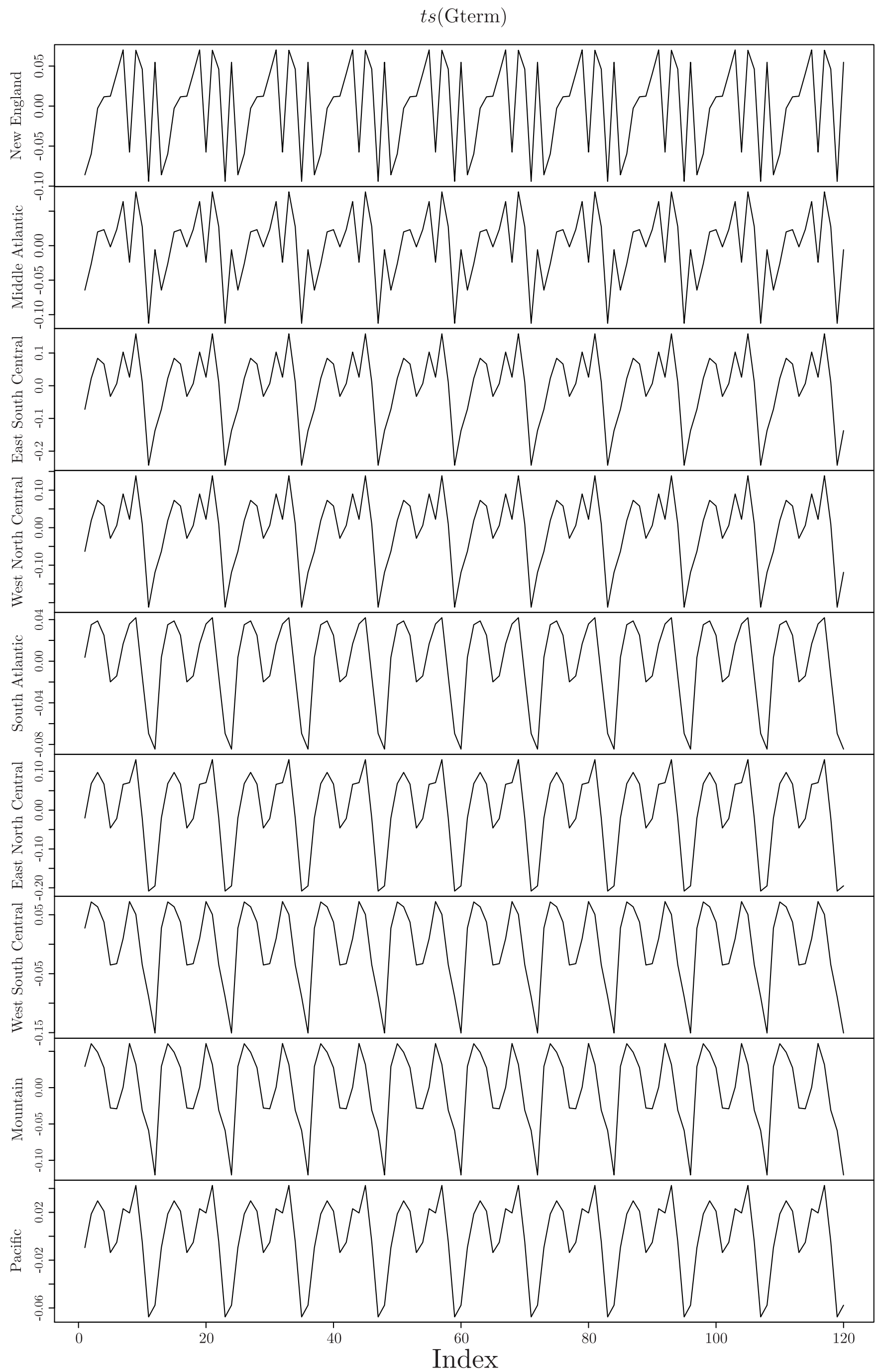


Figure 6. Time series plots for $\boldsymbol{G}\widehat{\boldsymbol{F}}_2\widehat{\boldsymbol{\omega}}_2'$ of a fitted DCF model of order (2,2,2). Maximum likelihood estimation is used.

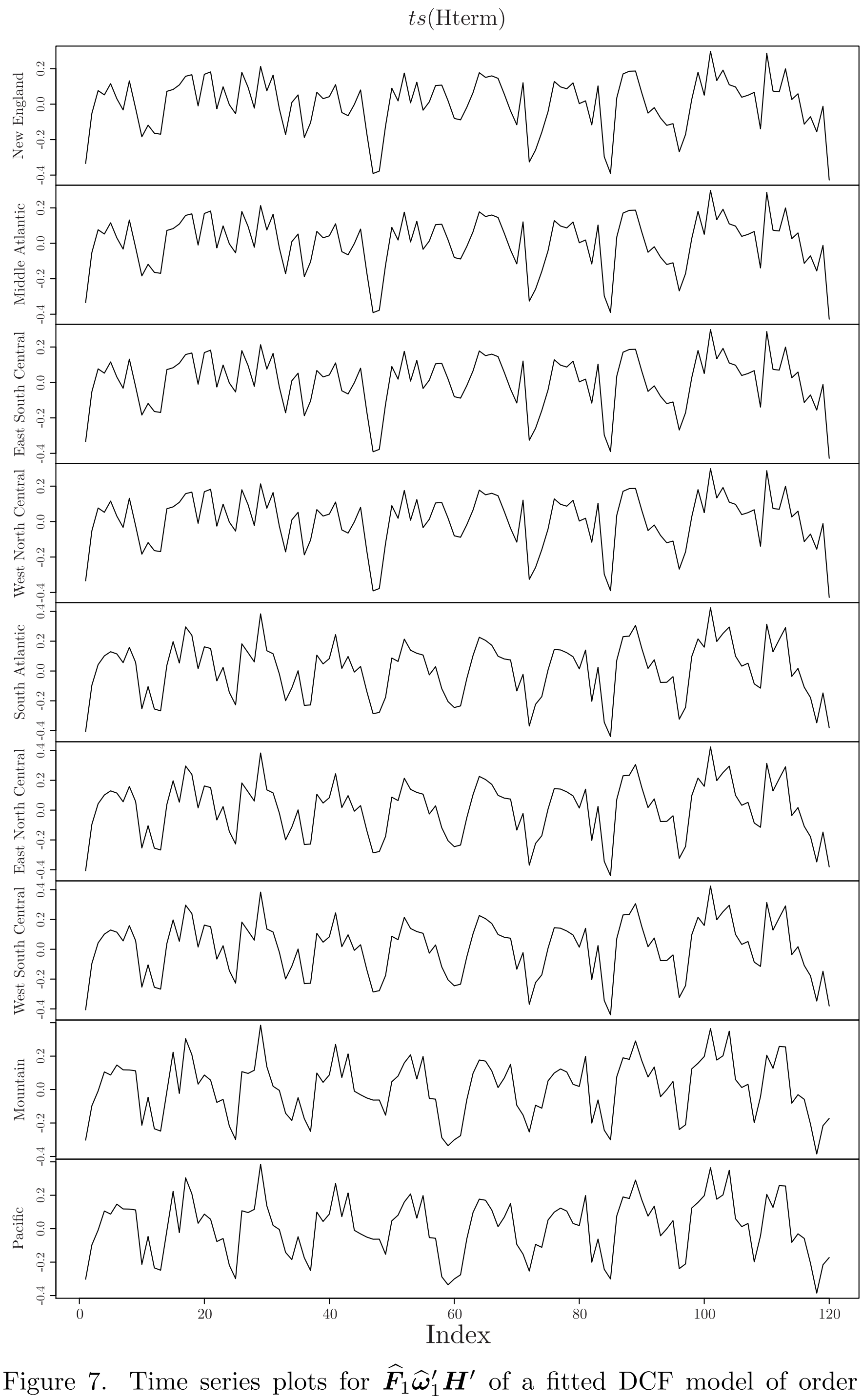


Figure 7. Time series plots for $\widehat{\boldsymbol{F}}_1\widehat{\boldsymbol{\omega}}_1'\boldsymbol{H}'$ of a fitted DCF model of order (2,2,2). Maximum likelihood estimation is used.

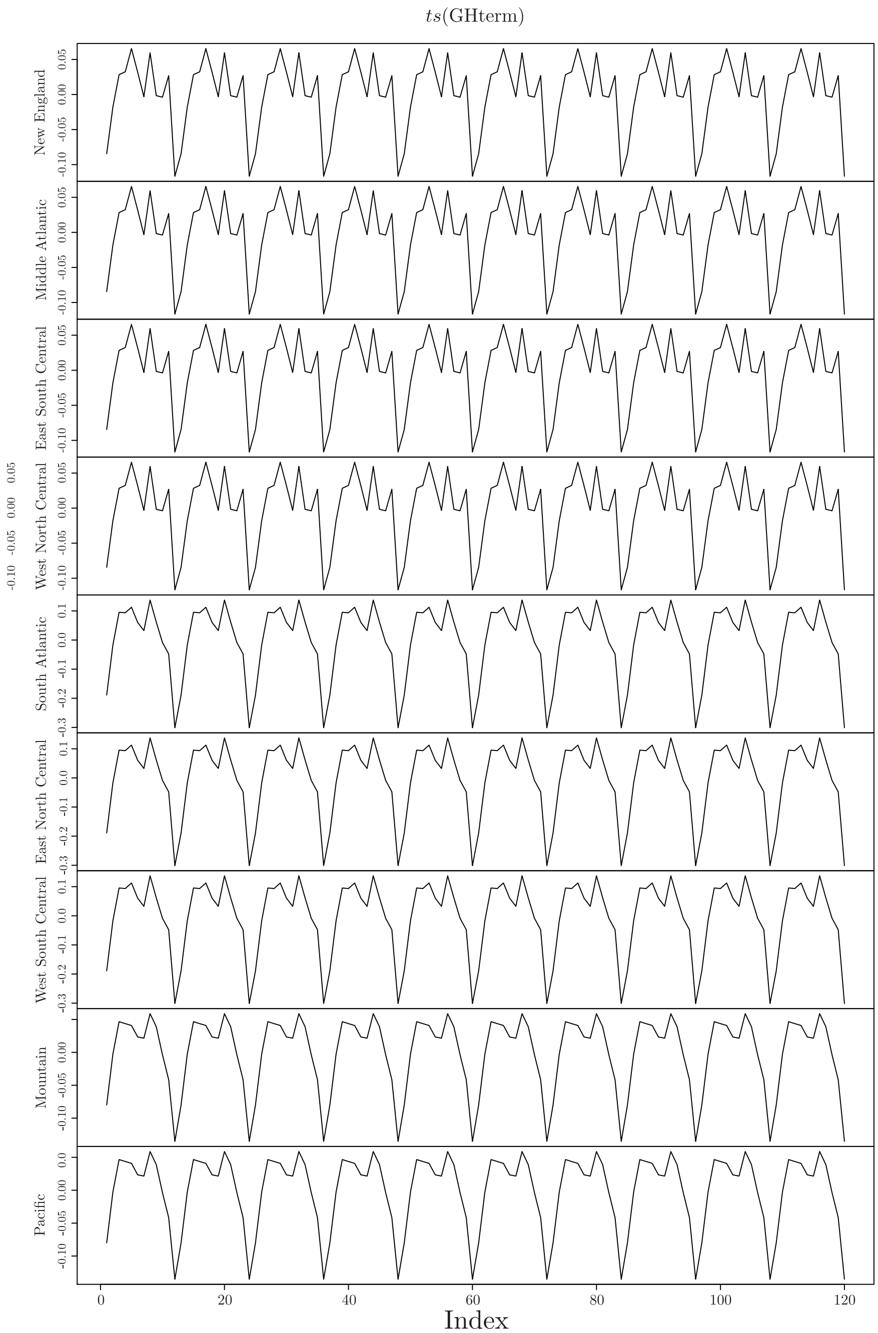


Figure 8. Time series plots for $\boldsymbol{G}\widehat{\boldsymbol{F}}_3\widehat{\boldsymbol{\omega}}_3'\boldsymbol{H}'$ of a fitted DCF model of order (2,2,2). Maximum likelihood estimation is used.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Control* **19**, 716-723.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis.* Wiley-Interscience, N.J.

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

Bai, J. and Li, K. (2012). Statistical analysis of factor models of high dimension. *Ann. Statist.* **40**, 436-465.

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

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica* **51**, 1281-1304.

Chang, J., Guo, B. and Yao, Q. (2013). High dimensional stochastic regression with latent factors, Endogeneity and Nonlinearity. arXiv:1310.1990.

Doz, C., Giannone, D. and Reichlin, L. (2006). A quasi-maximum likelihood approach for large approximate dynamic factor models. Discussion Paper 5724, CEPR.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). The generalized dynamic factor model: Identification and estimation. *Rev. Econom. Statist.* **82**, 540-554.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *J. Amer. Statist. Assoc.* **100**, 830-840.

Goyal, A., Perignon, C. and Villa, C. (2008). How common are common return factors across the NYSE and Nasdaq? *J. Finan. Econom.* **90**, 252 - 271.

Harville, D. A. (1997). *Matrix Algebra From a Statistician's Perspective.* Springer, New York.

Jöreskog, K. G. (1975). Factor analysis by least squares and maximum likelihood. In *Statistical Methods for Digital Computers.* (Edited by K. Enslein, A. Ralston and H. S. Wilf). Wiley, New York.

Lam, C. and Yao, Q. W. (2012). Factor modeling for high-dimensional time series: Inference for the number of factor. *Ann. Statist.* **40**, 694-726.

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

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* Springer-Verlag, Berlin.

Magnus, J. R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Revised Edition. Wiley, New York.

Peña, D. and Box, G. E. P. (1987). Identifying a simplifying structure in time series. *J. Amer. Statist. Assoc.* **82**, 836-843.

Schott, J. R. (1997). *Matrix Analysis for Statistics.* Wiley, New York.

Takane, Y. and Hunter, M. A. (2001). Constrained principal component analysis: A comprehensive theory. *Appl. Algebra Engineer., Commun. and Comput.* **12**, 391-419.

Tiao, G. C., Tsay, R. S. and Wang, T. C. (1993). Usefulness of linear transformations in multiple time series analysis. *Empir. Econom.* **18**, 567-593.

Tsai, H. and Tsay, R. S. (2010). Constrained factor models. *J. Amer. Statist. Assoc.* **105**, 1593-1605.

Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan.
E-mail: htsai@stat.sinica.edu.tw
Booth School of Business, University of Chicago, Chicago, IL 60637, U.S.A.
E-mail: ruey.tsay@chicagobooth.edu
Institute of Finance, National Chiao Tung University, Hsinchu 30010, Taiwan.
E-mail: m9281067@gmail.com
Department of Statistics, Purdue University, West Lafayette, IN 47907, U.S.A.
E-mail: aks43725@gmail.com

(Received November 2013; accepted October 2015)