

## Statistica Sinica Preprint No: SS-2024-0181

<b>Title</b>	Mean Independent Component Analysis for Multivariate Time Series
<b>Manuscript ID</b>	SS-2024-0181
<b>URL</b>	<a href="http://www.stat.sinica.edu.tw/statistica/">http://www.stat.sinica.edu.tw/statistica/</a>
<b>DOI</b>	10.5705/ss.202024.0181
<b>Complete List of Authors</b>	Chung Eun Lee and Zeda Li
<b>Corresponding Authors</b>	Chung Eun Lee
<b>E-mails</b>	chungeun.lee@baruch.cuny.edu
Notice: Accepted author version.	

# MEAN INDEPENDENT COMPONENT ANALYSIS FOR MULTIVARIATE TIME SERIES

Chung Eun Lee and Zeda Li

*Paul H. Chook Department of Information Systems and Statistics, Baruch College*

*Abstract:* In this article, we introduce the mean independent component analysis for multivariate time series to reduce the parameter space. In particular, we seek for a contemporaneous linear transformation that detects univariate mean independent components so that each component can be modeled separately. The mean independent component analysis is flexible in the sense that no parametric model or distributional assumptions are made. We propose a unified framework to estimate the mean independent components from a data with a fixed dimension or a diverging dimension. We estimate the mean independent components by the martingale difference divergence so that the mean dependence across components and across time is minimized. The approach is extended to the group mean independent component analysis by imposing a group structure on the mean independent components. We further introduce a method to identify the group structure when it is unknown. The consistency of both proposed methods is established. Extensive simulations and a real data illustration for community mobility is provided to demonstrate the efficacy of our method.

*Key words and phrases:* Conditional mean, Dimension reduction, High dimensional time series, Nonlinear dependence.

---

# 1 Introduction

Multivariate time series data is prevalent in a variety of fields, such as economics, business, engineering, climate science, and biomedicine. Often, high-dimensional time series data are encountered, where the modeling task becomes more challenging since the modeling complexity increases dramatically with the dimension of the series. Due to this fact, the dimension reduction is essential and becomes the core of multivariate time series analysis.

A primary concern in modeling a multivariate time series is searching for an elaborative but simplified underlying structure that substantially reduces the dimension of the parameter space. Many statistical methods have been proposed to reduce the dimension of the parameter space. On one hand, a factor model structure is assumed to achieve the dimension reduction on the data; see, for example, Peña & Box (1987), Bai & Ng (2002), Forni et al. (2005), Pan & Yao (2008), Lam & Yao (2012) among others. On the other hands, there is a line of research imposing a simplified structure after applying a linear transformation to the data which includes the canonical correlation analysis (Box & Tiao 1977), the principal components analysis (Stock & Watson 2002), the scalar component analysis (Tiao & Tsay 1989), the dynamic orthogonal component analysis (Matteson & Tsay 2011). Furthermore, there is an extensive literature on the blind source separation in the signal processing. Such methods include the independent component analysis (ICA) (Back & Weigend 1997, Cardoso 1998, Hyvarinen et al. 2001, among others), the second-order blind source separation (Tong et al. 1992, Belouchrani et al. 1997, Miettinen et al. 2016, among others). More recently, Chang

---

et al. (2018) have extended the principal components analysis (PCA) and sought for several lower-dimensional series that have no cross-correlations across different time lags. However, all the aforementioned approaches assume that the latent components are uncorrelated or independent, which is either highly simplified or extremely restrictive for multivariate time series. For a simple example, let the underlying component  $\mathbf{X}_t = (x_{1,t}, \dots, x_{p,t})^\top \in \mathbb{R}^p$  be a stationary process, where  $x_{i,t} = \eta_{t-i+1}$  for  $i = 1, \dots, p$ . On one hand, the latent independent components do not exist unless the sequence  $\{\eta_t, \eta_{t-1}, \dots, \eta_{t-p+1}\}$  is independent. This independence condition seems to be a strong constraint for a time series to be satisfied. On the other hand, let  $\eta_t$  be uncorrelated but have temporal dependence beyond linear; see Lobato et al. (2002) and Shao & Wu (2007) for examples. Then imposing uncorrelatedness on latent components may overlook this important dynamics and highly simplify the dependence of the components, which may not be optimal. To this end, it would be ideal to search for the latent components that satisfy more constrained assumption than uncorrelatedness but less restrictive than independence.

In this article, we propose an alternative approach that searches for novel latent components which better align with the modeling and forecasting in practice, since the conditional mean is considered. This differentiates our method from most of the existing methods. We focus on seeking the components that are mutually mean independent across time so that each component can be modeled separately by univariate time series models, and reduce the number of parameters in modeling. More specifically, our goal is to estimate the orthogo-

nal matrix  $\mathbf{A}$  that decomposes the data  $\mathbf{Y}_t \in \mathbb{R}^p$  into the mean independent components,

$\mathbf{X}_t = (x_{1,t}, \dots, x_{p,t})^\top$  such that

$$\mathbf{A}^\top \mathbf{Y}_t = \mathbf{X}_t, \quad \mathbb{E} \left( x_{i,t} \mid \mathcal{F}_t^{(-i)} \right) = \mathbb{E}(x_{i,t}) \text{ almost surely for } i = 1, \dots, p,$$

where  $\mathcal{F}_t^{(-i)} = \sigma \left\{ \mathbf{X}_t^{(-i)}, \mathbf{X}_{t-1}^{(-i)}, \dots \right\}$  is the current and past information of  $\mathbf{X}_t^{(-i)} \in \mathbb{R}^{p-1}$

which contains all the elements of  $\mathbf{X}_t$  except  $x_{i,t}$ , and the components of  $\mathbf{X}_t$  are mutually mean independent across time. Thus  $\{x_{i,t-1}, x_{i,t-2}, \dots\}$  are sufficient to model the behavior of  $x_{i,t}$  for each  $i$  and supports to build a separate univariate time series model with the least amount of loss on the conditional mean of  $x_{i,t}$ . This is especially important because the conditional mean is the optimal predictor in the mean squared error sense. Hence, compared to the existing approaches, our mean independent component analysis can produce more accurate forecasting. Inspired by the method in Chang et al. (2018), we further extend our approach by imposing a group structure that segments the time series into several mean independent lower-dimensional subseries and propose a method and an algorithm to estimate the unknown group structure. Since the mean independence is the relationship that lies between the uncorrelatedness and the independence, our approaches naturally bridge the gap between the PCA approaches and the ICA approaches. Furthermore, no parametric model or distributional assumptions are imposed on the conditional mean  $\mathbb{E}(x_{i,t} \mid x_{i,t-1}, x_{i,t-2}, \dots)$ , thus our approaches are model-free in this aspect. Moreover, our approaches can detect both linear and nonlinear dependence. Consequently they can be widely applicable to real-world data exhibiting nonlinear dependence including the economical data such as interest rates, exchange rates, economic price data (Tsay 1998, Teräsvirta et al. 2010, Wang et al. 2022),

---

the environmental data such as temperature, precipitation, and sea ice dynamics, (Munch et al. 2023, Marwan et al. 2021), and the physiological signals such as gait variability and heart rate variability (Calderón-Juárez et al. 2023, Li 2023). Our approaches are built upon the metric called martingale difference divergence (MDD) of Shao & Zhang (2014) that perfectly captures the mean dependence between two vectors. To the best of our knowledge, the theoretical investigation of MDD for the large dimension has been mostly developed under the i.i.d. data. However, there are relatively fewer studies of MDD for the large dimension under the dependent data. In this article, we establish the asymptotic results of our methods under both fixed dimension and diverging dimension, which seems to be new to the literature and requires totally different techniques.

The rest of the article is organized as follows. We provide a brief review of the martingale difference divergence and the martingale difference correlation in Section 2. We introduce our mean independent component analysis in Section 3 and introduce the group mean independent component analysis in Section 4 with the theoretical investigations. All the numerical results are summarized in Section 5. In Section 6, we present a real data application and Section 7 concludes. All the proofs and additional theoretical results, simulation results, real data application results, and figures are gathered in the supplementary material.

## 2 Review of Martingale Difference Divergence

To introduce our new approaches, we briefly review the martingale difference divergence (MDD) and its scale invariant metric, the martingale difference correlation (MDC), in Shao & Zhang (2014) and Park et al. (2015). For  $\mathbf{U} \in \mathbb{R}^r$  and  $\mathbf{V} \in \mathbb{R}^q$ , where  $r$  and  $q$  are positive integers, the metrics MDD and MDC measure the mean dependence between two vectors  $\mathbf{V}$  and  $\mathbf{U}$ , i.e.,  $\mathbb{E}(\mathbf{V} | \mathbf{U}) = \mathbb{E}(\mathbf{V})$  almost surely. More specifically, if  $\mathbb{E}(\|\mathbf{V}\|^2 + \|\mathbf{U}\|^2) < \infty$  where  $\|\cdot\|$  is the Euclidean norm, then  $\text{MDD}^2(\mathbf{V} | \mathbf{U}) = -\mathbb{E} \left[ \{\mathbf{V} - \mathbb{E}(\mathbf{V})\}^\top \{\mathbf{V}' - \mathbb{E}(\mathbf{V}')\} \|\mathbf{U} - \mathbf{U}'\| \right] \in \mathbb{R}$ ,  $\text{MDC}^2(\mathbf{V} | \mathbf{U}) = \frac{\text{MDD}^2(\mathbf{V} | \mathbf{U})}{\{\text{tr}(\widehat{\text{var}}(\mathbf{V}))\}^{1/2} \{\widehat{\text{dvar}}(\mathbf{U})\}^{1/2}} \in [0, 1]$ , where  $(\mathbf{V}', \mathbf{U}')$  is an independent copy of  $(\mathbf{V}, \mathbf{U})$  and  $\widehat{\text{dvar}}(\mathbf{U})$  is the distance variance of  $\mathbf{U}$  in Székely et al. (2007). The key property of MDD and MDC is that

$$\text{MDD}^2(\mathbf{V} | \mathbf{U}) = 0, \text{MDC}^2(\mathbf{V} | \mathbf{U}) = 0 \text{ if and only if } \mathbb{E}(\mathbf{V} | \mathbf{U}) = \mathbb{E}(\mathbf{V}) \text{ almost surely} \quad (2.1)$$

thus they fully characterize the mean independence of  $\mathbf{V}$  on  $\mathbf{U}$ . For given random vectors  $(\mathbf{U}_i, \mathbf{V}_i)_{i=1}^n$  from the joint distribution of  $(\mathbf{U}, \mathbf{V})$ , the sample  $\text{MDD}^2$  and the sample  $\text{MDC}^2$

$$\begin{aligned} \text{are defined as } \widehat{\text{MDD}}^2(\mathbf{V} | \mathbf{U}) &= \frac{-1}{n^2} \sum_{i,j=1}^n (\mathbf{V}_i - \bar{\mathbf{V}})^\top (\mathbf{V}_j - \bar{\mathbf{V}}) \|\mathbf{U}_i - \mathbf{U}_j\|, \widehat{\text{MDC}}^2(\mathbf{V} | \mathbf{U}) = \\ &= \frac{\widehat{\text{MDD}}^2(\mathbf{V} | \mathbf{U})}{\{\text{tr}(\widehat{\text{var}}(\mathbf{V}))\}^{1/2} \{\widehat{\text{dvar}}(\mathbf{U})\}^{1/2}}, \text{ where } \bar{\mathbf{V}} = \frac{1}{n} \sum_{k=1}^n \mathbf{V}_k, \text{tr}(\widehat{\text{var}}(\mathbf{V})) = \text{tr} \left( \frac{1}{n} \sum_{k=1}^n (\mathbf{V}_k - \bar{\mathbf{V}}) (\mathbf{V}_k - \bar{\mathbf{V}})^\top \right)^2 \\ &= \text{tr} \left( \frac{1}{n^2} \sum_{i,j=1}^n A_{i,j}^2 \right), A_{i,j} = a_{i,j} - a_{i,\cdot} - a_{\cdot,j} + a_{\cdot,\cdot}, a_{i,j} = \frac{1}{2} \|\mathbf{V}_i - \mathbf{V}_j\|^2, a_{i,\cdot} = \frac{1}{n} \sum_{j=1}^n a_{i,j}, \\ &a_{\cdot,j} = \frac{1}{n} \sum_{i=1}^n a_{i,j}, a_{\cdot,\cdot} = \frac{1}{n^2} \sum_{i,j=1}^n a_{i,j}, \widehat{\text{dvar}}(\mathbf{U}) = \frac{1}{n^2} \sum_{i,j=1}^n B_{i,j}^2, B_{i,j} = b_{i,j} - b_{i,\cdot} - b_{\cdot,j} + b_{\cdot,\cdot}, \\ &b_{i,j} = \|\mathbf{U}_i - \mathbf{U}_j\|, \text{ and } b_{i,\cdot}, b_{\cdot,j}, b_{\cdot,\cdot} \text{ are defined similarly to } a_{i,\cdot}, a_{\cdot,j} \text{ and } a_{\cdot,\cdot}. \end{aligned}$$

## 3 Mean Independent Component Analysis

### 3.1 Mean Independent Component Model

We shall first briefly review the existing methods, the dynamic orthogonal components in Matteson & Tsay (2011) and the second-order blind source separation in Belouchrani et al. (1997), which are closely related to our approach. The existing approaches assume that the data  $\mathbf{Y}_t = \mathbf{B}\mathbf{Z}_t$ , where  $\mathbb{E}(\mathbf{Y}_t) = \mathbb{E}(\mathbf{Z}_t) = \mathbf{0}$ ,  $\text{var}(\mathbf{Y}_t) = \text{var}(\mathbf{Z}_t) = \mathbf{I}$ ,  $\mathbf{B} = (\mathbf{B}_{\cdot 1}, \dots, \mathbf{B}_{\cdot p})$  is an orthogonal matrix,  $\mathbf{Z}_t = (z_{1,t}, \dots, z_{p,t})^\top \in \mathbb{R}^p$  are the underlying components that are uncorrelated to each other, i.e.,  $\text{cov}(z_{i,t}, z_{j,t-h}) = 0$  for  $i \neq j$ , all lags  $h = 0, 1, 2, \dots$ .

Next, we consider distinct components and introduce an alternative approach. In particular,

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t, \quad (3.2)$$

where  $\mathbb{E}(\mathbf{Y}_t) = \mathbb{E}(\mathbf{X}_t) = \mathbf{0}$ ,  $\text{var}(\mathbf{Y}_t) = \text{var}(\mathbf{X}_t) = \mathbf{I}$ ,  $\mathbf{A} = (\mathbf{A}_{\cdot 1}, \dots, \mathbf{A}_{\cdot p})$  is an orthogonal matrix,  $\mathbf{X}_t = (x_{1,t}, \dots, x_{p,t})^\top$  are the mean independent components, i.e.,  $\mathbb{E}\{x_{i,t} \mid \mathcal{F}_t^{(-i)}\} = \mathbb{E}(x_{i,t})$  almost surely,  $\mathcal{F}_t^{(-i)} = \sigma\{\mathbf{X}_t^{(-i)}, \mathbf{X}_{t-1}^{(-i)}, \dots\}$ ,  $\mathbf{X}_t^{(-i)} = (x_{1,t}, \dots, x_{i-1,t}, x_{i+1,t}, \dots, x_{p,t})^\top$ .

Similar to the other existing methods, the orthogonal matrix  $\mathbf{A}$  and the mean independent components  $\mathbf{X}_t$  are not unique in terms of signed permutation. In other words,  $\mathbf{Y}_t = \mathbf{A}\mathbf{P}_\pm^\top (\mathbf{P}_\pm \mathbf{X}_t)$  produces the same data  $\mathbf{Y}_t$ , where  $\mathbf{P}_\pm$  is a signed permutation matrix. However, in modeling or forecasting perspectives, discovering the mean independent components up to a signed permutation is sufficient. Therefore, we seek to identify the mean independent components up to a signed permutation; see Section 2.4 in Matteson & Tsay (2011).

---

**Remark 1** We shall make several remarks on the comparison between our method and the existing methods. It is worth mentioning that our approach is essentially different from the dynamic orthogonal components in Matteson & Tsay (2011) and the second-order blind source separation in Belouchrani et al. (1997). All approaches aim to reduce the number of parameters involved in the modeling. Therefore, all methods seek for the orthogonal matrix  $\mathbf{A}$  or  $\mathbf{B}$  so that each component can be equipped with a separate univariate model. However, Matteson & Tsay (2011) and Belouchrani et al. (1997) search for serially uncorrelated components by adopting covariances. Unless the data is Gaussian or solely generated from linear time series models, separating components through the linear metric is not sufficient. If the forecasting or modeling is the primary concern, we can alternatively separate the components by examining the contributions toward the conditional mean of each components, and seek for serially mean independent components  $\mathbf{X}_t$ . Since the conditional mean is the best predictor in terms of the mean squared error, we shall lose less prediction accuracy by focusing on the mean independent components. Also, it is worth pointing out that our approach summarizes both the linear and nonlinear dependence of components based on their contribution toward the conditional mean. Hence, our approach can be more robust and flexible to the dependence and the distribution of the data. Due to these facts, our approach can be viewed as a nonlinear analog of the dynamic orthogonal components in Matteson & Tsay (2011) or the second-order blind source separation in Belouchrani et al. (1997). Following the existing approaches, we assume  $\text{var}(\mathbf{Y}_t) = \text{var}(\mathbf{X}_t) = \mathbf{I}$ . Our approach can still remain valid when  $\text{var}(\mathbf{Y}_t) \neq \mathbf{I}$ . Un-

der the model assumption in (3.2), we have  $\mathbf{Y}_t = \text{var}(\mathbf{Y}_t)^{1/2} \tilde{\mathbf{Y}}_t = \text{var}(\mathbf{Y}_t)^{1/2} \mathbf{A} \mathbf{X}_t = \mathbf{C} \mathbf{X}_t$ , where  $\tilde{\mathbf{Y}}_t = \text{var}(\mathbf{Y}_t)^{-1/2} \mathbf{Y}_t$  and  $\mathbf{C} = \text{var}(\mathbf{Y}_t)^{1/2} \mathbf{A}$ . With a consistent estimator of  $\text{var}(\mathbf{Y}_t)$  denoted as  $\widehat{\text{var}}(\mathbf{Y}_t)$  and the assumption that the smallest eigenvalue of  $\text{var}(\mathbf{Y}_t)$  is bounded below by a positive constant, the estimated orthogonal matrix  $\widehat{\mathbf{A}}$  obtained after applying our approach to  $\widehat{\text{var}}(\mathbf{Y}_t)^{-1/2} \mathbf{Y}_t$  is consistent, especially under the fixed  $p$ . The estimator of  $\mathbf{C}$  is then given by  $\widehat{\mathbf{C}} = \widehat{\text{var}}(\mathbf{Y}_t)^{1/2} \widehat{\mathbf{A}}$  and provides the estimated mean independent components. In terms of the implementation, our approach fully measures the mean dependence through MDD to estimate the hidden components which is a moment-based method. Thus, comparing to the existing methods, our approach is also simple to implement; see Section 3.2 for the objective function.

### 3.2 Estimation of Mean Independent Components

Our specific goal is to seek  $\mathbf{A}$  that decomposes our data  $\mathbf{Y}_t$  into mean independent components  $\mathbf{X}_t$ . Notice that the mean independence between  $x_{i,t}$  and  $x_{j,t-h}$ ,  $h \geq 0$  can be fully captured by MDD due to its property in (2.1). Based on our model in (3.2), it implies that  $\text{MDD}^2(\mathbf{A}_{\cdot i}^\top \mathbf{Y}_t \mid \mathbf{A}_{\cdot j}^\top \mathbf{Y}_{t-h}) = 0$ ,  $i \neq j$ ,  $h = 0, 1, \dots$ . Motivated by this fact, we propose to search  $\mathbf{A}$  by minimizing the objective function,

$$\mathcal{S}_{h_0}(\tilde{\mathbf{A}}) = \left( \frac{1}{\text{tr}\{\tilde{\mathbf{A}}\tilde{\mathbf{A}}^\top\}} \right) \sum_{h=0}^{h_0} \sum_{i=1}^p \sum_{j \neq i} \text{MDD}^2(\tilde{\mathbf{A}}_{\cdot i}^\top \mathbf{Y}_t \mid \tilde{\mathbf{A}}_{\cdot j}^\top \mathbf{Y}_{t-h}), \quad (3.3)$$

where  $\mathcal{S}_{h_0}(\tilde{\mathbf{A}}) \geq 0$  and it becomes zero at the true  $\mathbf{A}$ , i.e.,  $\mathcal{S}_{h_0}(\mathbf{A}) = 0$ . Hence, our estimator

$$\widehat{\mathbf{A}} \text{ is } \widehat{\mathbf{A}} = \underset{\tilde{\mathbf{A}}^\top \tilde{\mathbf{A}} = \tilde{\mathbf{A}} \tilde{\mathbf{A}}^\top = \mathbf{I}}{\text{argmin}} \widehat{\mathcal{S}}_{h_0}(\tilde{\mathbf{A}}), \quad (3.4)$$

---

where  $\widehat{S}_{h_0}(\widetilde{\mathbf{A}}) = \left( \frac{1}{\text{tr}\{\widetilde{\mathbf{A}}\widetilde{\mathbf{A}}^\top\}} \right) \sum_{h=0}^{h_0} \sum_{i=1}^p \sum_{j \neq i} \widehat{\text{MDD}}^2(\widetilde{\mathbf{A}}_{\cdot i}^\top \mathbf{Y}_t \mid \widetilde{\mathbf{A}}_{\cdot j}^\top \mathbf{Y}_{t-h})$ .

**Remark 2** We shall make several remarks. We note that the denominator of the objective function in (3.3) does not affect the optimization procedure. Thus, in practice, the denominator has less impact on the estimation. However, the denominator is required to show the theoretical properties when  $p$  diverges. More specifically, the denominator is necessary in order to obtain Lipschitz continuity of  $S_{h_0}$  with respect to the distance in (3.5); see Lemma 1 in the supplementary material. Furthermore, selecting  $h_0$  has been a common question in the literature and proposing a method to select  $h_0$  is beyond the scope of this paper. As mentioned in the literature, generally, relatively small  $h_0$  is used since major dependence is often at the short time lag for stationary time series and more noises can be added if large  $h_0$  is selected; see Lam & Yao (2012) and Wang et al. (2019). Lastly, it is possible to provide weights for difference lags and construct the objective function similar to Matteson & Tsay (2011). However, we use our objective function in (3.4) considering the close connection with the method in Chang et al. (2018) which is described in Section 4. Also, we observe that our approach with the equal weights in (3.4) provides encouraging performance under our simulation study; see Section 5.

To optimize the objective function in (3.4), we adopt the parameterization estimation approach for  $\widetilde{\mathbf{A}}$  in Matteson & Tsay (2011) and Belouchrani et al. (1997). Specifically, we set  $\widetilde{\mathbf{A}} = \prod_{1 \leq i < j \leq p} \Gamma_{ij}(\theta_{ij})$ , where  $\Gamma_{ij}(\theta_{ij})$  is an identity matrix with  $(i, i)$ th and  $(j, j)$ th elements replaced by  $\cos(\theta_{ij})$ , and  $(i, j)$ th and  $(j, i)$ th elements replaced by  $\sin(\theta_{ij})$  and  $-\sin(\theta_{ij})$ , respectively. Note that our objective function is locally convex. By following the common

practice, we try multiple starting points to avoid local minima. Specifically, we employ Latin hypercube sampling to uniformly sample the space of  $\theta_{ij}$  to generate 200 multiple initial values. We select the best one in terms of our objective function to begin the process. When  $p$  is large, we suggest estimating the mean independent components sequentially so that the optimization is performed with reasonable computational time. However, it is important to note that estimation uncertainty accumulates at each stage of the sequential process compared to the joint estimation. Here, we have a tradeoff between computational complexity and statistical efficiency. For more details regarding estimating the mean independent components sequentially, we refer readers to Matteson & Tsay (2017) in section 2.4.

Next, we shall introduce the asymptotic results for our estimation procedure for both fixed and diverging  $p$ . Since  $\mathbf{A}$  is not identifiable in terms of signed permutation, we adopt the following distance from Chang et al. (2018) which can measure the closeness of two orthogonal matrices and establish the consistency of our estimator. After arranging the orders of the columns based on the similarity for  $\mathbf{B} = (\mathbf{B}_{\cdot 1}, \dots, \mathbf{B}_{\cdot p}) \in \mathbb{R}^{p \times p}$  and  $\mathbf{C} = (\mathbf{C}_{\cdot 1}, \dots, \mathbf{C}_{\cdot p}) \in \mathbb{R}^{p \times p}$ , the  $\mathcal{D}$ -distance is defined as

$$\mathcal{D}(\mathbf{B}, \mathbf{C}) = \left\{ p - \sum_{i=1}^p \text{tr}(\mathbf{B}_{\cdot i} \mathbf{B}_{\cdot i}^\top \mathbf{C}_{\cdot i} \mathbf{C}_{\cdot i}^\top) \right\}^{1/2}. \quad (3.5)$$

Notice that for any orthogonal matrices  $\mathbf{B}$  and  $\mathbf{C}$ , the distance is nonnegative,  $\mathcal{D}(\mathbf{B}, \mathbf{C}) \geq 0$ . Moreover, if  $\mathbf{B}$  is a column permutation of  $\mathbf{C}$ , then  $\mathcal{D}(\mathbf{B}, \mathbf{C}) = 0$ . Thus, smaller values of  $\mathcal{D}$  indicate greater similarity between the orthogonal matrices  $\mathbf{B}$  and  $\mathbf{C}$ . We make the following assumptions, under which we derive the consistency of our estimator.

---

**Assumption 1** Let  $\mathcal{H}_{\mathcal{D}}$  be the space that consists of orthogonal matrices.

1. There exists an orthogonal matrix  $\mathbf{A} \in \mathcal{H}_{\mathcal{D}}$  that minimizes the objective function  $\mathcal{S}_{h_0}(\cdot)$ . Furthermore, the minimum value of  $\mathcal{S}_{h_0}(\cdot)$  is obtained at an orthogonal matrix  $\tilde{\mathbf{A}}$  if and only if  $\mathcal{D}(\mathbf{A}, \tilde{\mathbf{A}}) = 0$ .
2. When  $p$  is fixed, the data  $(\mathbf{Y}_t)_{t=1}^n$  is strictly stationary and  $\beta$ -mixing process. There exists  $\delta > 0$  such that  $\mathbb{E}\|\mathbf{Y}_t\|^{6+3\delta} < \infty$  and the  $\beta$ -mixing coefficients  $\beta(k) = O(k^{-(2+\delta')/\delta'})$  for  $\delta' \in (0, \delta)$ .
3. When  $p$  diverges with  $n$ , the set  $\mathcal{H}_{\mathcal{D}}$  is the union of  $\mathcal{H}_p$ , where  $\mathcal{H}_p$  is the set containing all  $p \times p$  orthogonal matrices.
4. When  $p$  diverges, the data  $(\mathbf{Y}_t)_{t=1}^n$  is strictly stationary and  $\beta$ -mixing process. There exists  $\delta > 0$  such that  $\mathbb{E}\left\|\frac{1}{\sqrt{p}}\mathbf{Y}_t\right\|^{6+6\delta} < C$  for all  $p$ , where  $C$  is a finite constant. The  $\beta$ -mixing coefficients satisfy  $\sum_{k=1}^{\infty} k\beta(k)^{\delta/(1+\delta)} < \infty$ .

The first condition in Assumption 1 ensures that  $\mathbf{A}$  is the unique minimizer of  $\mathcal{S}_{h_0}$  in the sense of  $\mathcal{D}$ . This condition is necessary to show the theoretical properties, and it is also assumed in Matteson & Tsay (2011). The second and fourth conditions are used to bound certain moments when  $p$  is fixed or  $p$  diverges, where similar  $\beta$ -mixing condition in the fourth condition is often used in time series literature; see Hjellvik et al. (1996) and Dette & Spreckelsen (2004). The third condition is assumed so that  $\hat{\mathbf{A}}$  is well defined and measurable when  $p$  diverges and guarantees the consistency of  $\hat{\mathbf{A}}$  in terms of  $\mathcal{D}$  if  $\hat{\mathcal{S}}_{h_0}(\tilde{\mathbf{A}})$  converges

uniformly to  $\mathcal{S}_{h_0}(\tilde{\mathbf{A}})$ ; see the proof of Theorem 1 in the supplementary material for details.

**Theorem 1** 1. Let  $h_0$  and  $p$  be fixed integers. Under the first and second conditions in Assumption 1, we have  $\mathcal{D}(\mathbf{A}, \hat{\mathbf{A}}) \rightarrow^p 0$  as  $n \rightarrow \infty$ .

2. Let  $h_0$  be a fixed integer and  $p$  grow with  $n$  such that  $p\sqrt{p}/\sqrt{n} \rightarrow 0$  as  $n \rightarrow \infty$ . Under the first, third, and fourth conditions in Assumption 1, we have  $\mathcal{D}(\mathbf{A}, \hat{\mathbf{A}}) \rightarrow^p 0$  as  $n \rightarrow \infty$ .

Theorem 1 indicates that when the dimension  $p$  is fixed or  $p$  grows with  $n$  such that  $p\sqrt{p}/\sqrt{n} \rightarrow 0$ , our estimator gets closer to the true  $\mathbf{A}$  in terms of the  $\mathcal{D}$ -distance as sample size increases.

It is worth mentioning that the assumptions are made towards the data  $\mathbf{Y}_t$  but not on the mean independent components  $\mathbf{X}_t$  to study the asymptotic properties. It seems hard to obtain the convergence rate for  $\mathcal{D}(\mathbf{A}, \hat{\mathbf{A}})$  under the current assumptions since the main theorems we used only guarantee  $\mathcal{D}(\mathbf{A}, \hat{\mathbf{A}}) \rightarrow 0$ . However, it is possible to extend the theorem so that our approach is theoretically justified under a weaker assumption for  $p$  by following the proof in Chang et al. (2018). In particular,  $p$  may grow as long as  $p(\log(p))^{4/\gamma'}/n \rightarrow 0$ , where  $\gamma'^{-1} = 1 + r_2^{-1}$  for  $r_2 \in (0, 1]$  comes from the mixing condition in Assumption S1. We summarize this result in Theorem S1, available in the supplementary material. Also, in the literature, often a parametric model or distributional assumptions are made to derive the theoretical results. However, our approach does not assume a parametric model or a distribution of the data. In terms of the theoretical investigations, the existing methods studied asymptotic properties for fixed  $p$  while our approach examines both fixed  $p$  and growing  $p$ . Moreover,

the theoretical properties of MDD is mostly investigated for i.i.d. data. However, we derive the asymptotic results of MDD with dependent data, which is relatively less studied in the literature. Notice that our objective function depends on  $n$  when  $p$  diverges. Thus, showing the consistency is far from trivial and requires new techniques.

## 4 Group Mean Independent Component Analysis

### 4.1 Group Mean Independent Component Model

In many applications, the observed data might be produced from components that have a certain group structure, resulting in lower-dimensional subseries. Inspired by the work of the PCA for time series (TS-PCA) in Chang et al. (2018), we extend the mean independent component analysis with a group structure the so-called group mean independent component analysis. More precisely, we consider

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t, \quad (4.6)$$

where  $\mathbb{E}(\mathbf{Y}_t) = \mathbb{E}(\mathbf{X}_t) = \mathbf{0}$ ,  $\text{var}(\mathbf{Y}_t) = \text{var}(\mathbf{X}_t) = \mathbf{I}$ ,  $\mathbf{A} = (\mathbf{A}_1, \dots, \mathbf{A}_m)$  is an orthogonal matrix,  $\mathbf{A}_i \in \mathbb{R}^{p \times p_i}$ ,  $\sum_{i=1}^m p_i = p$ ,  $\mathbf{X}_t = \left\{ (\mathbf{X}_t^{(1)})^\top, \dots, (\mathbf{X}_t^{(m)})^\top \right\}^\top \in \mathbb{R}^p$  is the group mean independent components,  $\mathbf{X}_t^{(i)} = \left( x_{1,t}^{(i)}, x_{2,t}^{(i)}, \dots, x_{p_i,t}^{(i)} \right)^\top \in \mathbb{R}^{p_i}$ , i.e.,

$$\mathbb{E} \left\{ \mathbf{X}_t^{(i)} \mid \mathcal{F}_t^{(-i)} \right\} = \mathbb{E}(\mathbf{X}_t^{(i)}) \text{ almost surely,} \quad (4.7)$$

here,  $\mathcal{F}_t^{(-i)} = \sigma \left\{ \mathbf{X}_t^{(-i)}, \mathbf{X}_{t-1}^{(-i)}, \dots \right\}$ ,  $\mathbf{X}_t^{(-i)} = \left\{ (\mathbf{X}_t^{(1)})^\top, \dots, (\mathbf{X}_t^{(i-1)})^\top, (\mathbf{X}_t^{(i+1)})^\top, \dots, (\mathbf{X}_t^{(m)})^\top \right\}^\top \in \mathbb{R}^{p-p_i}$ . As mentioned in Chang et al. (2018), the orthogonal matrix  $\mathbf{A}$  is not identifiable

---

but  $\{\text{span}(\mathbf{A}_i)\}_{i=1}^m$  are unique; see Section 2.1 in Chang et al. (2018). Thus, our goal is to estimate the identifiable spaces  $\{\text{span}(\mathbf{A}_i)\}_{i=1}^m$ .

**Remark 3** We shall make a few remarks. While TS-PCA in Chang et al. (2018) segments elements of  $\mathbf{X}_t$  by the uncorrelatedness, our approach divides the groups of  $\mathbf{X}_t$  by the mean independence in (4.7). Because of this fact, our method has several distinct features. In particular, since the TS-PCA separates the groups through the autocovariance matrices, it suffers from the same limitations mentioned in Remark 1. Thus, our approach can be more robust and flexible to different forms of dependence and different distributions of the data. Furthermore, as mentioned in Chang et al. (2018), TS-PCA cannot separate uncorrelated components if different blocks share the same eigenvalues of their linear matrix. However, our approach bypasses the limitation of the existing method since our approach directly estimates the group mean independent components by optimizing an objective function which considers the mean dependence between two groups. Also, if there is prior knowledge of the group, our approach can easily incorporate the known group structure. Lastly, the correlatedness implies the mean dependence, which suggests a nested structure for a particular  $i$ th group, i.e.,  $\text{span}(\bar{\mathbf{A}}_i) \subseteq \text{span}(\mathbf{A}_i), i \in \{1, 2, \dots, m\}$ , after the same alignments for  $\mathbf{A}$  and  $\bar{\mathbf{A}}$ , where  $\bar{\mathbf{A}} = (\bar{\mathbf{A}}_1, \dots, \bar{\mathbf{A}}_l)$  is the orthogonal matrix that generates the block components that are uncorrelated in TS-PCA. Due to this fact, for a given  $p = \dim(\mathbf{Y}_t)$ , it is possible for our approach to have a different number of groups compared to TS-PCA. More specifically, it is possible that the TS-PCA selects more groups than our approach, i.e.,  $m \leq l$ . To better

understand the differences, we provide one example.

**Example 1** Suppose the data  $\mathbf{Y}_t$  is generated by two groups of mean independent components, i.e.,  $\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t = \mathbf{A} \begin{pmatrix} \mathbf{X}_t^{(1)} \\ \mathbf{X}_t^{(2)} \end{pmatrix}$ , where one group  $\mathbf{X}_t^{(1)} = (x_{1,t}^{(1)}, \dots, x_{p_1,t}^{(1)})^\top$  has components which are mutually uncorrelated but mean dependent across time and the other group  $\mathbf{X}_t^{(2)} = (x_{1,t}^{(2)}, \dots, x_{p-p_1,t}^{(2)})^\top$  has components that are correlated and mean dependent.

Under the model assumption in TS-PCA, the true group structure becomes  $(1, 1, \dots, 1, p - p_1)$  which indicates that  $\{x_{i,t}^{(1)}\}$ ,  $i = 1, \dots, p_1$  construct  $p_1$  different groups while  $\mathbf{X}_t^{(2)}$  form a single group. Thus, there are  $l = p_1 + 1$  different groups for TS-PCA. On the other hand, under our framework, there are  $m = 2$  different groups  $(\mathbf{X}_t^{(1)}, \mathbf{X}_t^{(2)})$  and the true group structure is  $(p_1, p - p_1)$ . Hence, it is possible to have different number of groups and different group structures for the existing method and our approach.

## 4.2 Estimation of Group Mean Independent Components

Suppose the group structure  $(p_1, \dots, p_m)$  are known. Similar to the mean independent components in Section 3, we have  $\text{MDD}^2(\mathbf{X}_t^{(i)} | \mathbf{X}_{t-h}^{(j)}) = \sum_{k=1}^{p_i} \text{MDD}^2(x_{k,t}^{(i)} | \mathbf{X}_{t-h}^{(j)}) = 0$ ,  $i \neq j$ ,  $h = 0, 1, \dots$ , due to (4.7) and this is equivalent to  $\text{MDD}^2(\mathbf{A}_i^\top \mathbf{Y}_t | \mathbf{A}_j^\top \mathbf{Y}_{t-h}) = \sum_{k=1}^{p_i} \text{MDD}^2((\mathbf{A}_i)_{\cdot k}^\top \mathbf{Y}_t | \mathbf{A}_j^\top \mathbf{Y}_{t-h}) = 0$ ,  $i \neq j$ ,  $h = 0, 1, \dots$ , where  $(\mathbf{A}_i)_{\cdot k}$  is the  $k$ th column in  $\mathbf{A}_i$ . By this fact, we define the objective function as

$$\mathcal{G}_{h_0}(\tilde{\mathbf{A}}) = \left( \frac{1}{\text{tr}(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^\top) m} \right)^{1/2} \sum_{h=0}^{h_0} \sum_{i=1}^m \sum_{k=1}^{p_i} \sum_{j \neq i} \text{MDD}^2((\tilde{\mathbf{A}}_i)_{\cdot k}^\top \mathbf{Y}_t | \tilde{\mathbf{A}}_j^\top \mathbf{Y}_{t-h}), \quad (4.8)$$

where  $\mathcal{G}_{h_0}(\tilde{\mathbf{A}}) \geq 0$  and it becomes zero at the true orthogonal matrix  $\mathbf{A}$  with the true group structure, i.e.,  $\mathcal{G}_{h_0}(\mathbf{A}) = 0$ . Notice that the denominator is different from the one in  $\mathcal{S}_{h_0}$ . This is mainly due to the fact that there are different number of elements that construct  $\mathcal{G}_{h_0}$ . In other words, while there are  $h_0 p(p-1)$  different MDD<sup>2</sup> added in  $\mathcal{S}_{h_0}$ , there are  $h_0 p(m-1)$  different MDD<sup>2</sup> that construct  $\mathcal{G}_{h_0}$ . We observe that the case  $m = p$  corresponds to the mean independent components, under which the objective function  $\mathcal{G}_{h_0}(\tilde{\mathbf{A}})$  reduces to  $\mathcal{S}_{h_0}(\tilde{\mathbf{A}})$ . Consequently, mean independent components can be viewed as a special case of group mean independent components.

For the estimation of  $\mathbf{A}$ , we shall utilize

$$\hat{\mathcal{G}}_{h_0}(\tilde{\mathbf{A}}) = \left( \frac{1}{\text{tr}(\tilde{\mathbf{A}}\tilde{\mathbf{A}}^\top)^m} \right)^{1/2} \sum_{h=0}^{h_0} \sum_{i=1}^m \sum_{j \neq i} \widehat{\text{MDD}}^2 \left( \tilde{\mathbf{A}}_i^\top \mathbf{Y}_t \mid \tilde{\mathbf{A}}_j^\top \mathbf{Y}_{t-h} \right), \text{ and search for } \mathbf{A}$$

$$\hat{\mathbf{A}} = \underset{\tilde{\mathbf{A}}^\top \tilde{\mathbf{A}} = \tilde{\mathbf{A}} \tilde{\mathbf{A}}^\top = \mathbf{I}}{\text{argmin}} \hat{\mathcal{G}}_{h_0}(\tilde{\mathbf{A}}). \quad (4.9)$$

Since  $\{\text{span}(\mathbf{A}_i)\}_{i=1}^m$  are identifiable, we shall establish the consistency for our estimator  $\mathbf{A}$  with  $\mathcal{D}$ -distance in Section 3.2, i.e.,  $\mathcal{D}(\mathbf{B}, \mathbf{C}) = \left\{ p - \sum_{i=1}^m \text{tr}(\mathbf{B}_i \mathbf{B}_i^\top \mathbf{C}_i \mathbf{C}_i^\top) \right\}^{1/2}$ , where  $\mathbf{B} = (\mathbf{B}_1, \dots, \mathbf{B}_m)$ ,  $\mathbf{C} = (\mathbf{C}_1, \dots, \mathbf{C}_m)$ , and  $\mathbf{B}_i, \mathbf{C}_i \in \mathbb{R}^{p \times p_i}$  are semi-orthogonal matrices.

**Theorem 2** 1. Let  $h_0, m$ , and  $p$  be fixed integers. Assume that  $(p_1, \dots, p_m)$  are known.

Under the first and second conditions in Assumption 1 with respect to  $\mathcal{G}_{h_0}$ , we have

$$\mathcal{D}(\mathbf{A}, \hat{\mathbf{A}}) \xrightarrow{p} 0 \text{ as } n \rightarrow \infty.$$

2. Let  $h_0$  be fixed integers. Assume that  $(p_1, \dots, p_m)$  are known. Let  $p$  grow with  $n$

such that  $p\sqrt{m}/\sqrt{n} \rightarrow 0$  as  $n \rightarrow \infty$ . Under the first, third, and fourth conditions in

Assumption 1 with respect to  $\mathcal{G}_{h_0}$ , we have  $\mathcal{D}(\mathbf{A}, \hat{\mathbf{A}}) \xrightarrow{p} 0$  as  $n \rightarrow \infty$ .

It is worth noting that Theorem 2 remains valid for  $m \rightarrow \infty$ . Furthermore, the growth rate of  $p$  in Theorem 2 coincides with the one in Theorem 1. This is expected since the mean independent components emerge as a special case of the group mean independent components when  $m = p$ .

In practice, the true group structure  $(p_1, \dots, p_m)$  may not be given and needs to be estimated. We shall follow the maximum cross correlation approach in Chang et al. (2018), which is stemmed from the ratio-based estimator in the factor model literature. We first select  $r$  which is the number of two pairs of components that have at least one time lag showing a significant mean dependence. By searching for the pairs that have significant serial mean dependence, we can estimate  $(p_1, \dots, p_m)$ . More specifically, we estimate  $r$  by

$$\hat{r} = \operatorname{argmax}_{1 \leq j \leq c_0 p_0} \frac{\hat{M}_j}{\hat{M}_{j+1}}, \quad (4.10)$$

where  $\hat{M}_1 \geq \hat{M}_2 \geq \dots \geq \hat{M}_{p_0}$  are every pairs of  $\hat{M}(i, j)$  in the descending order,

$$\hat{M}(i, j) = \max_{0 \leq h \leq h_0} \left\{ \widehat{\text{MDC}}^2(\hat{x}_{i,t} | \hat{x}_{j,t-h}), \widehat{\text{MDC}}^2(\hat{x}_{j,t} | \hat{x}_{i,t-h}) \right\}, \quad (4.11)$$

and  $\hat{\mathbf{X}}_t = \hat{\mathbf{A}}^\top \mathbf{Y}_t = (\hat{x}_{1,t}, \dots, \hat{x}_{p,t})^\top$ ,  $c_0 = 0.75$  which is recommended in Chang et al. (2018),  $p_0 = p(p-1)/2$  is the total number of two pairs of  $p$  components, Once  $r$  is estimated, we define an undirected graph  $G = (V, E)$ , where  $V = \{1, 2, \dots, p\}$  is the vertex set and  $E$  is the set of edges such that  $e_{i,j} = e_{j,i} = 1$  if  $\hat{M}(i, j) \in (\hat{M}_1, \dots, \hat{M}_{\hat{r}})$  or  $e_{i,j} = e_{j,i} = 0$  if  $\hat{M}(i, j) \notin (\hat{M}_1, \dots, \hat{M}_{\hat{r}})$ . Two components  $\hat{x}_{i,t}$  and  $\hat{x}_{j,t}$  belong to the same group if the vertices  $i$  and  $j$  are directly or indirectly connected, i.e.,  $e_{i,j} = 1$  or there exists  $\{v_1, v_2, \dots, v_w\} \subset V$  such that  $e_{i,v_1} = e_{v_1,v_2} = \dots = e_{v_w,j} = 1$ . Algorithm 1 summarizes the entire process for estimating our group mean independent components without prior

---

knowledge of the group structure.

**Remark 4** We shall make several remarks on Algorithm 1 and our group mean independent component analysis. If the group information is known, the group structure is directly incorporated into the objective function (4.9) by pre-specifying the number of groups  $m$  and  $p_1, \dots, p_m$ . We suggest estimating semi-orthogonal matrix  $\mathbf{A}_i, i = 1, \dots, m$ , sequentially using the optimization solver with orthogonality constraint proposed by Wen & Yin (2013), which has an efficient computational time; see Wen & Yin (2013) for more details. If the group structure is unknown, we suggest using Algorithm 1. Step 2 - Step 5 estimates the group structure and Step 6 estimates  $\mathbf{A}$ . We use our mean independent components  $\widehat{\mathbf{X}}_t^{(0)}$  in Section 3 as an initial estimate for the group mean independent components. The intention is to group different components into one group that still show the mean dependence in  $\widehat{\mathbf{X}}_t^{(0)}$ , where components are as mutually mean independent as possible. This strategy is often used in the literature; see Stögbauer et al. (2004), Cardoso (1998), Gruber et al. (2009) among others. Furthermore, it is desirable to conduct the false discovery rate-based multiple testing with MDD to assess the mean independence for every two pairs of components. However, constructing a test is very nontrivial since we only have  $\widehat{\mathbf{X}}_t$  and the true  $\mathbf{X}_t$  is unknown, thus there will be an estimation effect. Also, each test shall involve a bootstrap procedure which requires intensive computations. In this sense, the ratio-based approach in Algorithm 1 is more computationally efficient to estimate the group structure. We further note that there is a new method for the estimation of  $r$ ; see the variant of ratio-based estimator discussed in Chang et al. (2015) and Han et al. (2024). We use the ratio-based estimator in (4.10) by

---

following the practice of the implementation in Chang et al. (2018). Algorithm 1 generally requests several iterations to converge and shows an improvement after each iteration. For instance, the algorithm stops after 2 or 3 iterations under our simulation examples. Besides, the estimated  $\widehat{\mathbf{A}}^{(1)}$  and  $(\widehat{p}_1, \dots, \widehat{p}_{\widehat{m}})$  obtained through Step 1 - Step 6 generate reasonable results in our simulation study and they are not greatly different from the final estimates after a few iterations until the algorithm converges. Lastly, the theory to show the consistency of our estimator of the group structure is very challenging since it involves showing the consistency of  $\widehat{r}$  and  $(\widehat{p}_1, \dots, \widehat{p}_{\widehat{m}})$ . As mentioned in the existing literature, showing the consistency of the estimator  $\widehat{r}$  is not straightforward because it is hard to show that the maximum of the ratio between  $\widehat{M}_j$  and  $\widehat{M}_{j+1}$  is occurred when  $j = r$ . Even if there is a new approach that estimates  $r$  with the consistency, the challenge still remains for showing the consistency of  $(\widehat{p}_1, \dots, \widehat{p}_{\widehat{m}})$ . To establish this consistency, it is necessary to show  $P(\widehat{e}_{i,j} \neq e_{i,j} | \widehat{r} = r) \rightarrow 0$  for any  $i, j$ . However, this is nontrivial, and we cannot adopt the arguments from Chang et al. (2018) and Han et al. (2024) since our estimation procedure and metric differ from theirs. Notice that the estimation procedure of our approach and the one in Chang et al. (2018) are fundamentally different. While our approach optimizes an objective function to search  $\mathbf{A}$ , the method in Chang et al. (2018) applies eigen decomposition to their linear matrix. Thus,  $\widehat{\mathbf{M}}(i, j)$  in (4.11) has more complicated form which cannot be expressed by the multiplication of a particular matrix and vectors. Thus,  $|\widehat{\mathbf{M}}(i, j) - \mathbf{M}(i, j)|$  cannot be controlled by the difference between a matrix and its sample counterpart. Also, unlike the methods in Chang

et al. (2018) and Han et al. (2024), it is hard to obtain the convergence rate of  $\|\widehat{\mathbf{A}} - \mathbf{A}\|_2$ ,  $\|\cdot\|_2$  is a spectral norm, for our approach which makes the theoretical analysis even more challenging. We shall leave this for future work.

## 5 Numerical Simulations

### 5.1 Mean Independent Component Analysis

We examine the finite sample performance of our approach for mean independent component analysis (MICA) in Section 3 and consider two examples with various distributions. In particular, the stationary data is generated by  $\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t$ , where the orthogonal matrix  $\mathbf{A}$  is generated through `randortho` function in `pracma` R package by following Stewart (1980), and  $\mathbf{X}_t = (x_{1,t}, \dots, x_{p,t})^\top$  is generated by

$$\text{Example I: } x_{i,t} = \phi_i x_{i,t-1} + \epsilon_t, \text{ for } i = 1, \dots, p,$$

$$\text{Example II: } x_{i,t} = \phi_i x_{i,t-1} - \phi_i^{-1} \epsilon_{t-1} + \epsilon_t, \text{ for } i = 1, \dots, p,$$

here  $\{\phi_i\}_{i=1}^p$  are randomly generated from `Uniform(0.5, 0.9)`. For Example I,  $\epsilon_t$  is generated from i.i.d. standard normal, student's t distribution with 3 degrees of freedom, or exponential distribution with  $\lambda = 1$ . For Example II,  $\epsilon_t$  is generated from all distributions considered in Example I except for the standard normal so that each component  $x_{i,t}$  is not independent across time. Similar to the simulation study in Matteson & Tsay (2011), we set  $p = 5, 10$ ,  $n = 50, 100, 200$ , and  $h_0 = 1, 2$ , where the results for  $h_0 = 2$  is reported in the supplementary material which shows similar results as the ones in Table 1. Also, the finite sample

performance for larger dimension is summarized in the supplementary material. We note that  $\mathbf{X}_t$  are mutually mean independent and fall into the framework (3.2). Example I with standard normal  $\epsilon_t$  is adopted from Matteson & Tsay (2011) and the data has linear dependence since all components are generated from linear time series models. For Example II, the data has no linear dependence but has a dependence beyond linear since  $\mathbf{X}_t$  is generated from the all-pass ARMA(1,1) model which is a well known example for a uncorrelated process yet possesses temporal dependence; see Lobato et al. (2002). As a result, the existing autocovariance-based approaches may be insufficient to detect the important dynamics since linear metric is limited to measuring linear dependence and cannot capture other forms of dependence.

We apply our approach and compare with the existing approaches, the dynamic orthogonal components (DOC) in Matteson & Tsay (2011) and the second-order blind source separation (SOBI) in Belouchrani et al. (1997), which search for the uncorrelated components. In order to compare the accuracy of the estimation of  $\mathbf{A}$ , we compute the scaled version of the distance between two orthogonal matrices so that  $\mathcal{D}^2(\mathbf{A}, \hat{\mathbf{A}}) \in [0, 1]$ , i.e.,  $\mathcal{D}^2(\mathbf{A}, \hat{\mathbf{A}}) = \frac{1}{p} \sum_{i=1}^p \left\{ 1 - \text{tr} \left( \mathbf{A}_{\cdot i} \mathbf{A}_{\cdot i}^\top \hat{\mathbf{A}}_{\cdot i} \hat{\mathbf{A}}_{\cdot i}^\top \right) \right\}$  where smaller  $\mathcal{D}^2$  indicates more accurate result.

Table 1 summarizes the results based on 500 replications. Overall, both existing methods and our approach produce more accurate dimension reduction results when  $n$  increases. For Example I with a standard normal  $\epsilon_t$ , we observe that three approaches produce comparable results but the DOC and the SOBI methods perform slightly better than our method. Better

performances from the DOC and the SOBI are expected since the data is generated from Gaussian linear time series and the entire dependence can be captured by the covariances. However, our approach outperforms the existing methods when the data is generated by a student's t distribution or an exponential distribution. For Example II, our approach is superior to the DOC and the SOBI methods for all sample sizes and different distributions. This demonstrates the robustness of our approach for the dependence and the distribution.

## 5.2 Group Mean Independent Component Analysis

In this section, we study the finite sample performance for the group mean independent component analysis (GMICA) in Section 4. We consider two examples, where Example I is adopted from Chang et al. (2018). More specifically, the data is generated by  $\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t$ ,

where  $\mathbf{A}$  is defined in Section 5.1, and  $\mathbf{X}_t = \left\{ (\mathbf{X}_t^{(1)})^\top, (\mathbf{X}_t^{(2)})^\top, (\mathbf{X}_t^{(3)})^\top \right\}^\top = (x_{1,t}, \dots, x_{p,t})^\top$ ,

$m = 3$ , is generated by  $x_{i,t} = z_{t+i-1}^{(1)}$ ,  $i \leq p/2$ ,  $x_{i,t} = z_{t+i-p/2-1}^{(2)}$ ,  $p/2 + 1 \leq i \leq p \cdot 5/6$ ,

$x_{i,t} = z_{t+i-p \cdot 5/6-1}^{(3)}$ ,  $p \cdot 5/6 + 1 \leq i \leq p$ ,

Example I:  $z_t^{(1)} = 0.5z_{t-1}^{(1)} + 0.3z_{t-2}^{(1)} + \epsilon_t^{(1)} - 0.9\epsilon_{t-1}^{(1)} + 0.3\epsilon_{t-2}^{(1)} + 1.2\epsilon_{t-3}^{(1)} + 1.3\epsilon_{t-4}^{(1)}$ ,

$z_t^{(2)} = 0.8z_{t-1}^{(2)} - 0.5z_{t-2}^{(2)} + \epsilon_t^{(2)} + \epsilon_{t-1}^{(2)} + 0.8\epsilon_{t-2}^{(2)} + 1.8\epsilon_{t-3}^{(2)}$ ,

and  $z_t^{(3)} = -0.7z_{t-1}^{(3)} - 0.5z_{t-2}^{(3)} + \epsilon_t^{(3)} - \epsilon_{t-1}^{(3)} - 0.8\epsilon_{t-2}^{(3)}$ .

Example II:  $z_t^{(1)} = 0.9z_{t-1}^{(1)} - 0.9^{-1}\epsilon_{t-1}^{(1)} + \epsilon_t^{(1)}$ ,  $z_t^{(2)} = 0.5z_{t-1}^{(2)} - 0.5^{-1}\epsilon_{t-1}^{(2)} + \epsilon_t^{(2)}$ ,

and  $z_t^{(3)} = 0.1z_{t-1}^{(3)} - 0.1^{-1}\epsilon_{t-1}^{(3)} + \epsilon_t^{(3)}$ .

Similar to examples in Section 5.1,  $\epsilon_t^{(1)}$ ,  $\epsilon_t^{(2)}$ ,  $\epsilon_t^{(3)}$  are generated from standard normal, student's t distribution with 3 degrees of freedom, or exponential distribution with  $\lambda = 1$ . We further note that  $(\mathbf{X}_t^{(1)}, \mathbf{X}_t^{(2)}, \mathbf{X}_t^{(3)})$  are mutually uncorrelated and mean independent in Ex-

ample I and each series  $z_t^{(i)}$ ,  $i = 1, 2, 3$  has linear dependence. To check the performance for the dependence beyond linear dependence, we replace  $(z_t^{(1)}, z_t^{(2)}, z_t^{(3)})$  in Example I with the series generated from all-pass ARMA(1,1) for Example II, where each series is uncorrelated; see Lobato et al. (2002). By following the simulation study in Chang et al. (2018), we set  $p = 6, 12, n = 200, 500, 1000$  and  $h_0 = 1, 5$ , where the results for  $h_0 = 1$  is reported in the supplementary material and shows similar results for our approach. Also, the finite sample performance for larger dimension is summarized in the supplementary material.

We compare the performance of our approach in comparison with the existing method, TS-PCA in Chang et al. (2018), which searches for the uncorrelated low-dimensional components. For the fair comparison, we treat the group information is unknown and estimate both the group structure and the separation matrix  $\mathbf{A}$  with our objective function using Algorithm 1 with  $c_0 = 0.75$  and no iterations since the method of Chang et al. (2018) does not iteratively search for  $\mathbf{A}$ . In other words, we shall estimate the group structure and  $\mathbf{A}$  through Step 1 - Step 6 in Algorithm 1. By following Chang et al. (2018), we compute  $\pi \in [0, 1]$  the proportions of correct segmentations, and compute the scaled version of the  $\mathcal{D}^2$ -distance when the method correctly group the components; see the supplemental material of Chang et al. (2018), i.e.,  $\pi = \frac{\sum_{l=1}^R \mathbf{1}\{\widehat{p}_{1,l}, \dots, \widehat{p}_{\widehat{m},l} = (p_1, \dots, p_m)\}}{R}$ ,  $\mathcal{D}^2(\mathbf{A}, \widehat{\mathbf{A}}) = \frac{1}{m} \sum_{i=1, \dots, m} \left( 1 - \frac{1}{p_i} \text{tr} \left( \mathbf{A}_i \mathbf{A}_i^\top \widehat{\mathbf{A}}_i \widehat{\mathbf{A}}_i^\top \right) \right)$ , where  $R$  is the number of the replications,  $\{\widehat{p}_{i,l}\}_{i=1}^{\widehat{m}}$  is the estimate of the number of components for each group for  $l$ th replicate. Also, notice that a higher  $\pi$  and a smaller  $\mathcal{D}^2$  indicate more accurate result.

---

Table 2 presents the results based on 500 replications. From Table 2, we observe that our approach outperforms the TS-PCA for both Example I and Example II in terms of higher  $\pi$  and smaller  $\mathcal{D}^2$ . The difference between our approach and the TS-PCA becomes more prominent for Example II. In Example II, we consider latent components that are white noise but have nonlinear dependence. Therefore, the existing method with a covariance matrix cannot detect the components. This could be a part of the reason for the larger  $\mathcal{D}^2$ -distances produced by the existing approach.

## 6 Real Data Illustration

In this section, we illustrate our approach by forecasting Google’s community mobility reports during COVID-19. The data contains daily changes in mobility trends compared to a baseline established on February 15, 2020. It captures movements across 6 categories of locations. It is essential to produce accurate predictions on the community mobility as it is related to a virus spread and support to combat the virus.

In our analysis, we consider community mobility reports of eight states in Northeastern of the U.S., including Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont, New York, and New Jersey. By following Zhen & Wang (2023), we also consider the average of the mobility of all 6 categories in the particular day and state from January 1, 2021 to February 28, 2022, thus we have  $\mathbf{Y}_t \in \mathbb{R}^8$  of length  $n = 424$ . In order to make the data stationary, we first detrend each time series by using polynomial smoothing to re-

move the trend and standardize the time series with  $\text{var}(\mathbf{Y}_t)^{-1/2}$  so that the data falls into the framework (3.2) or (4.6). The auto and cross correlations of  $\mathbf{Y}_t$  which is reported in the supplementary material exhibits weekly periodicity. Hence we set  $h_0 = 7, 14$  and apply both MICA and GMICA. We first divide the data into a training set and a testing set, where the data from January 1, 2021 to December 31, 2021 (365 data points) is assigned as a training set and the remaining two months of observations starting from January 1, 2022 to February 28, 2022 (59 data points) are assigned as a testing set. To compare the prediction accuracy, we compute the mean squared prediction error (MSPE) where smaller MSPE indicates more precise prediction, i.e.,  $\text{MSPE} = \frac{\|\hat{\mathbf{Y}}_{n_0+j+q}^{(q)} - \mathbf{Y}_{n_0+j+q}\|^2}{p}$ , where  $\hat{\mathbf{Y}}_{n_0+j+q}^{(q)}$  is the  $q$ -step ahead forecasting of  $\mathbf{Y}_{n_0+j+q}$ , where  $n_0 = 365$ ,  $j = 1, \dots, 58$ , and  $q = 1, 2$ . To produce  $\hat{\mathbf{Y}}_{n_0+j+q}^{(q)}$ , we follow the approach in Chang et al. (2018) which is described by the following steps. (1) We first apply the proposed MICA and GMICA to the training set as described in our simulations, and obtain  $\hat{\mathbf{A}}$  and  $(\hat{p}_1, \dots, \hat{p}_{\hat{m}})$  which are fixed afterward. (2) We estimate  $\{\hat{\mathbf{X}}_t\}_{t=1}^{n_0+j}$  by  $\{\hat{\mathbf{A}}^\top \mathbf{Y}_t\}_{t=1}^{n_0+j}$ . (3) Due to the existence of the seasonal correlations, we fit appropriate univariate or multivariate seasonal ARMA  $(d_1, 0) \times (d_2, 0)_7$  models to each component or group of  $\hat{\mathbf{X}}_t$ , and obtain  $q$ -step ahead predictions of the components,  $\hat{\mathbf{X}}_{n_0+j+q}^{(q)}$ , where the orders  $d_1$  and  $d_2$  are determined by AIC. (4) We finally generate the prediction for  $\mathbf{Y}_{n_0+j+q}$  by  $\hat{\mathbf{Y}}_{n_0+j+q}^{(q)} = \hat{\mathbf{A}} \hat{\mathbf{X}}_{n_0+j+q}^{(q)}$ . In addition to our MICA and GMICA, we apply 6 different approaches to compare the prediction accuracy. In particular, we further apply the existing methods TS-PCA of Chang et al. (2018), DOC of Matteson & Tsay (2011), SOBI

---

of Belouchrani et al. (1997) to this data, where the predictions are generated by following the same procedure above. Also, we consider the parametric models by fitting univariate seasonal ARMA model (sUARMA) for each components in  $\mathbf{Y}_t$  and fitting seasonal vector ARMA model (sVARMA) to  $\mathbf{Y}_t$  by following step (3). Lastly, we consider fitting a refined VAR model (refVAR) to the data  $\mathbf{Y}_t$  with a thresholding value equals to 1, and the order selected by AIC.

Table 3 summarizes the prediction accuracy for all methods. From Table 3, our GMICA estimates 6 different groups, where the largest group has 3 components. On the other hand, TS-PCA generates 7 different groups. Thus, the number of groups are different for TS-PCA and our method, where the number of groups for TS-PCA is larger. This aligns with our argument in Remark 3. In terms of the predictions, our GMICA generates notably smaller MSPE than the ones of the existing methods including TS-PCA, DOC, SOBI, and parametric models sUARMA, sVARMA, and refVAR. In particular, MSPE of GMICA is roughly 10% lower than MSPE of TS-PCA, DOC, SOBI on average for  $q = 1$ . Also, the difference of MSPE becomes more prominent by comparing GMICA and parametric models sUARMA, sVARMA, and refVAR. We note that sUARMA, sVARMA, and refVAR produce the same prediction performance under  $h_0 = 7$  and  $h_0 = 14$  since those approaches do not depend on  $h_0$ . Moreover, GMICA performs better than MICA. We believe this is due to the extra grouping step in the GMICA and forming groups for the components seems to be helpful for this data. Notice that GMICA forms groups if the past or current values of another compo-

---

nents affect the conditional mean of the other components. Thus, if the main interest is the prediction, it would be more beneficial to seek the group mean independent components and build separate time series model for each group since GMICA incurs the minimum loss on the conditional mean which is the optimal predictor in terms of the mean squared error sense. We conjecture that this could be one reason for better prediction performance of GIMCA. It is interesting to observe that our MICA also produces more accurate predictions than the other existing methods such as TS-PCA, DOC, SOBI, sUARMA, sVARMA, and refVAR in terms of smaller MSPE. It is worth mentioning that the improvement in forecasting for our approaches GMICA, MICA compared to the existing methods TS-PCA, DOC, SOBI is essentially due to the different estimator of  $\mathbf{A}$ , since we apply the same procedure to generate the predictions. For this data, our approach effectively separates the components that lead to more accurate forecastings. We additionally compare the prediction performances focusing on each category which are reported in our supplementary material. We observe that GMICA or MICA consistently produce smaller MSPE across all categories.

We further report the auto and cross correlations of  $\hat{\mathbf{X}}_t$  for our GMICA and the existing approach TS-PCA in our supplementary material in order to check the uncorrelatedness across different groups. By comparing Figures S2 and Figures S3, it appears that our GMICA shows weaker linear dependence between two components that belong to different groups. For instance, the third component of TS-PCA seems to be correlated with the first and second components which are in a separate group. This could be another part of the reason that our

---

GMICA can produce more accurate forecastings for this data. We observe a similar phenomenon in Figure S2 from the auto and cross correlations of  $\widehat{\mathbf{X}}_t$  for our MICA and the DOC and the SOBI which are reported in the supplementary material.

## 7 Summary and Conclusions

We propose new approaches to model the multivariate time series data by seeking the latent mean independent components. Once the mean independent components are identified, a separate analysis or model can be built for each component. This reduces huge number of parameters involved in the modeling. Compared to the existing methods, our approach can be more robust and flexible to different forms of dependence and different distributions of the data, which is demonstrated in our simulation study. We further introduce a natural extension to the group mean independent component analysis by following the idea of Chang et al. (2018). To justify the validity of our method, we provide the asymptotic properties under two scenarios: fixed  $p$  and growing  $p$ . In addition to encouraging finite sample performance, we further illustrate the advantage of our approaches with a real data application that can generate more accurate predictions.

We conclude this article by discussing several future research topics. We introduce a procedure to estimate the group structure. Our simulation results indicate that the method identifying the group structure produces reasonable results, however there is no theoretical justification. It would be helpful to provide the theoretical property for the estimation of the

---

group structure. Furthermore, imposing the sparsity assumption on  $\mathbf{A}$  for the group mean independent component analysis is an interesting direction and we conjecture that the sparse method may allow  $p$  to grow faster than  $p\sqrt{m}/\sqrt{n} \rightarrow 0$  as  $n \rightarrow \infty$ . It requires an updated objective function, along with a different estimation procedure. Lastly, we often observe matrix or tensor time series data, and the dimension reduction for matrix or tensor time series has received considerable attention; see Wang et al. (2019), Chen et al. (2020), Chen et al. (2021) for the recent development for the dimension reduction on matrix or tensor time series. It would be useful to extend our approach for matrix or tensor time series to reduce the number of the parameters. The research along these directions are well underway.

## Supplementary Material

Supplementary material available online includes technical proofs of theoretical results and state additional theorem with its proof, and reports additional simulations, real data application results, and figures.

## Acknowledgements

The authors thank the Editor, the Associate Editor, and two referees for their constructive comments and suggestions that led to substantial improvements. Dr. Lee's research is supported by NSF grant DMS-2532852. Dr. Li's research is supported by NSF grant DMS-

---

2418850.

## References

- Back, A. D. & Weigend, A. S. (1997), 'A first application of independent component analysis to extracting structure from stock returns', *International journal of neural systems* **8**(04), 473–484.
- Bai, J. & Ng, S. (2002), 'Determining the number of factors in approximate factor models', *Econometrica* **70**(1), 191–221.
- Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F. & Moulines, E. (1997), 'A blind source separation technique using second-order statistics', *IEEE Transactions on signal processing* **45**(2), 434–444.
- Box, G. E. & Tiao, G. C. (1977), 'A canonical analysis of multiple time series', *Biometrika* **64**(2), 355–365.
- Calderón-Juárez, M., González-Gómez, G. H., Echeverría, J. C. & Lerma, C. (2023), 'Revisiting nonlinearity of heart rate variability in healthy aging', *Scientific Reports* **13**(1), 13185.
- Cardoso, J.-F. (1998), Multidimensional independent component analysis, in 'Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)', Vol. 4, IEEE, pp. 1941–1944.

- 
- 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’, *The Annals of Statistics* **46**(5), 2094–2124.
- Chen, E. Y., Tsay, R. S. & Chen, R. (2020), ‘Constrained factor models for high-dimensional matrix-variate time series’, *Journal of the American Statistical Association* **115**, 775–793.
- Chen, R., Yang, D. & Zhang, C.-H. (2021), ‘Factor models for high-dimensional tensor time series’, *Journal of the American Statistical Association* pp. 1–23.
- Dette, H. & Spreckelsen, I. (2004), ‘Some comments on specification tests in nonparametric absolutely regular processes’, *Journal of Time Series Analysis* **25**(2), 159–172.
- Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2005), ‘The generalized dynamic factor model: one-sided estimation and forecasting’, *Journal of the American statistical association* **100**(471), 830–840.
- Gruber, P., Gutch, H. W. & Theis, F. J. (2009), Hierarchical extraction of independent subspaces of unknown dimensions, in ‘International Conference on Independent Component Analysis and Signal Separation’, Springer, pp. 259–266.
- Han, Y., Chen, R., Zhang, C.-H. & Yao, Q. (2024), ‘Simultaneous decorrelation of matrix time series’, *Journal of the American Statistical Association* **119**(546), 957–969.

- 
- Hjellvik, V., Yao, Q. & Tjøstheim, D. (1996), Linearity testing using local polynomial approximation. discussion paper 60, Technical report, Sonderforschungsbereich 373, Humboldt-Universität zu Berlin, Spandauerstr. 1, 10178, Berlin.
- Hyvarinen, A., Karhunen, J. & Oja, E. (2001), ‘Independent component analysis [m]’, *New York: A Wiley Interscience Publication* pp. 175–182.
- 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.
- Li, Z. (2023), ‘Robust conditional spectral analysis of replicated time series’, *Statistics and Its Interface* **16**(1), 81–96.
- Lobato, I. N., Nankervis, J. C. & Savin, N. (2002), ‘Testing for zero autocorrelation in the presence of statistical dependence’, *Econometric Theory* **18**(3), 730–743.
- Marwan, N., Donges, J. F., Donner, R. V. & Eroglu, D. (2021), ‘Nonlinear time series analysis of palaeoclimate proxy records’, *Quaternary Science Reviews* **274**, 107245.
- Matteson, D. S. & Tsay, R. S. (2011), ‘Dynamic orthogonal components for multivariate time series’, *Journal of the American Statistical Association* **106**(496), 1450–1463.
- Matteson, D. S. & Tsay, R. S. (2017), ‘Independent component analysis via distance covariance’, *Journal of the American Statistical Association* **112**(518), 623–637.

- 
- Miettinen, J., Illner, K., Nordhausen, K., Oja, H., Taskinen, S. & Theis, F. J. (2016), ‘Separation of uncorrelated stationary time series using autocovariance matrices’, *Journal of Time Series Analysis* **37**(3), 337–354.
- Munch, S. B., Rogers, T. L., Symons, C. C., Anderson, D. & Pennekamp, F. (2023), ‘Constraining nonlinear time series modeling with the metabolic theory of ecology’, *Proceedings of the National Academy of Sciences* **120**(12), e2211758120.
- Pan, J. & Yao, Q. (2008), ‘Modelling multiple time series via common factors’, *Biometrika* **95**(2), 365–379.
- Park, T., Shao, X. & Yao, S. (2015), ‘Partial martingale difference correlation’, *Electronic Journal of Statistics* **109**, 1492–1517.
- Peña, D. & Box, G. E. (1987), ‘Identifying a simplifying structure in time series’, *Journal of the American statistical Association* **82**(399), 836–843.
- Shao, X. & Wu, W. B. (2007), ‘Asymptotic spectral theory for nonlinear time series’.
- Shao, X. & Zhang, J. (2014), ‘Martingale difference correlation and its use in high-dimensional variable screening’, *Journal of the American Statistical Association* **109**, 1302–1318.
- Stewart, G. W. (1980), ‘The efficient generation of random orthogonal matrices with an application to condition estimators’, *SIAM Journal on Numerical Analysis* **17**(3), 403–409.

- 
- Stock, J. H. & Watson, M. W. (2002), 'Forecasting using principal components from a large number of predictors', *Journal of the American statistical association* **97**(460), 1167–1179.
- Stögbauer, H., Kraskov, A., Astakhov, S. A. & Grassberger, P. (2004), 'Least-dependent-component analysis based on mutual information', *Physical Review E* **70**(6), 066123.
- Székely, G. J., Rizzo, M. L. & Bakirov, N. K. (2007), 'Measuring and testing dependence by correlation of distances'.
- Teräsvirta, T., Tjøstheim, D. & Granger, C. W. (2010), *Modelling nonlinear economic time series*, Oxford University Press.
- Tiao, G. C. & Tsay, R. S. (1989), 'Model specification in multivariate time series', *Journal of the Royal Statistical Society: Series B (Methodological)* **51**(2), 157–195.
- Tong, L., Inouye, Y. & Liu, R.-W. (1992), 'A finite-step global convergence algorithm for the parameter estimation of multichannel ma processes', *IEEE Transactions on signal processing* **40**(10), 2547–2558.
- Tsay, R. S. (1998), 'Testing and modeling multivariate threshold models', *journal of the american statistical association* **93**(443), 1188–1202.
- Wang, D., Liu, X. & Chen, R. (2019), 'Factor models for matrix-valued high-dimensional time series', *Journal of econometrics* **208**(1), 231–248.

---

Wang, G., Zhu, K. & Shao, X. (2022), ‘Testing for the martingale difference hypothesis in multivariate time series models’, *Journal of Business & Economic Statistics* **40**(3), 980–994.

Wen, Z. & Yin, W. (2013), ‘A feasible method for optimization with orthogonality constraints’, *Mathematical Programming* **142**(1-2), 397–434.

Zhen, Y. & Wang, J. (2023), ‘Non-negative tensor completion for dynamic counterfactual prediction on covid-19 pandemic’, *Annals of Applied Statistics* .

---

**Algorithm 1:** Algorithm to estimate the group mean independent components and the

group structure

---

**Data:**  $\mathbf{Y}_t = (y_{1,t}, \dots, y_{p,t})^\top$

**Result:**  $\hat{\mathbf{A}} = (\hat{\mathbf{A}}_1, \dots, \hat{\mathbf{A}}_m)$ ,  $\hat{\mathbf{X}}_t = (\hat{x}_{1,t}, \dots, \hat{x}_{p,t})^\top$ , and  $(\hat{p}_1, \dots, \hat{p}_{\hat{m}})$ .

**Step 1:** Begin the algorithm by setting  $m = p$  and  $p_1 = \dots = p_m = 1$ . Obtain an initial estimate  $\hat{\mathbf{A}}^{(0)}$  by minimizing  $\hat{\mathcal{S}}_{h_0}(\cdot)$  in (3.4) and estimate the components by

$$\hat{\mathbf{X}}_t^{(0)} = (\hat{\mathbf{A}}^{(0)})^\top \mathbf{Y}_t.$$

**Step 2:** Compute  $\hat{\mathbf{M}}(i, j)$  defined in (4.11) for every two pairs of components and arrange  $\hat{\mathbf{M}}(i, j)$  in the descending order.

**Step 3:** Select  $r$  through the ratio-based estimator in (4.10) with prepecified  $c_0$  and collect  $(\hat{\mathbf{M}}_1, \dots, \hat{\mathbf{M}}_{\hat{r}})$ , where  $\hat{\mathbf{M}}_k$  is the  $k$ th largest  $\hat{\mathbf{M}}(i, j)$ .

**Step 4:** Based on the collected  $(\hat{\mathbf{M}}_1, \dots, \hat{\mathbf{M}}_{\hat{r}})$ , create an undirected graph  $G = (V, E)$ , where  $V = \{1, 2, \dots, p\}$  is the vertex set and  $E$  is the set of edges such that  $e_{i,j} = e_{j,i} = 1$  if  $\hat{\mathbf{M}}(i, j) \in (\hat{\mathbf{M}}_1, \dots, \hat{\mathbf{M}}_{\hat{r}})$  or  $e_{i,j} = e_{j,i} = 0$  if  $\hat{\mathbf{M}}(i, j) \notin (\hat{\mathbf{M}}_1, \dots, \hat{\mathbf{M}}_{\hat{r}})$ .

**Step 5:** Based on the graph in Step 4, estimate the group structure,  $(\hat{p}_1, \dots, \hat{p}_{\hat{m}})$ . For instance, two components  $i$  and  $j$  belong to the same group if the vertices  $i$  and  $j$  are directly connected or indirectly connected, i.e.,  $e_{i,j} = 1$  or there exists

$$\{v_1, v_2, \dots, v_w\} \subset V \text{ such that } e_{i,v_1} = e_{v_1,v_2} = \dots = e_{v_w,j} = 1.$$

**Step 6:** Estimate  $\hat{\mathbf{A}}^{(1)}$  by minimizing  $\hat{\mathcal{G}}_{h_0}(\cdot)$  in (4.9) with  $(\hat{p}_1, \dots, \hat{p}_{\hat{m}})$  obtained from Step 5. Permute  $\hat{\mathbf{A}}^{(1)}$  based on the estimated group structure and estimate the components by

$$\hat{\mathbf{X}}_t^{(1)} = (\hat{\mathbf{A}}^{(1)})^\top \mathbf{Y}_t, \text{ where } \hat{\mathbf{A}}^{(1)} \text{ is the permuted matrix.}$$

**Step 7:** Repeat Step 2 - Step 6 until the estimated group structure,  $(\hat{p}_1, \dots, \hat{p}_{\hat{m}})$ , does not change and  $\|\hat{\mathbf{A}}^{(i+1)} - \hat{\mathbf{A}}^{(i)}\|_F < \epsilon$ , where  $\hat{\mathbf{A}}^{(i+1)}$  and  $\hat{\mathbf{A}}^{(i)}$  are the estimates of  $\mathbf{A}$  after  $i$ th and  $(i - 1)$ th iterations, and  $\epsilon$  is the prespecified tolerance.

---

Table 1: Reported are the average and the standard deviation of  $\mathcal{D}^2$ -distance with  $h_0 = 1$  based on 500 replications. Three methods are compared: our method (MICA), the method in Matteson & Tsay (2011, DOC), and the method in Belouchrani et al. (1997, SOBI).

Dist.	$p$	Methods	$n = 50$	$n = 100$	$n = 200$	
Normal	5	MICA	0.193 (0.069)	0.172 (0.071)	0.147 (0.074)	
		DOC	0.181 (0.071)	0.161 (0.068)	0.127 (0.066)	
		SOBI	0.169 (0.064)	0.151 (0.068)	0.124 (0.068)	
	10	MICA	0.374 (0.050)	0.349 (0.054)	0.323 (0.055)	
		DOC	0.362 (0.051)	0.324 (0.056)	0.289 (0.062)	
		SOBI	0.368 (0.050)	0.321 (0.060)	0.291 (0.059)	
	Example I	$t$	MICA	0.140 (0.068)	0.102 (0.061)	0.075 (0.056)
			DOC	0.201 (0.084)	0.166 (0.078)	0.119 (0.075)
			SOBI	0.216 (0.087)	0.164 (0.082)	0.129 (0.079)
10		MICA	0.317 (0.062)	0.269 (0.059)	0.231 (0.055)	
		DOC	0.394 (0.063)	0.350 (0.071)	0.286 (0.074)	
		SOBI	0.405 (0.065)	0.362 (0.072)	0.301 (0.074)	
exp		5	MICA	0.153 (0.071)	0.109 (0.065)	0.067 (0.050)
			DOC	0.230 (0.068)	0.214 (0.070)	0.176 (0.072)
			SOBI	0.187 (0.075)	0.147 (0.072)	0.107 (0.065)
	10	MICA	0.350 (0.058)	0.314 (0.058)	0.268 (0.056)	
		DOC	0.385 (0.052)	0.358 (0.054)	0.319 (0.058)	
		SOBI	0.371 (0.050)	0.330 (0.060)	0.278 (0.060)	
	Example II	$t$	MICA	0.137 (0.076)	0.075 (0.058)	0.037 (0.037)
			DOC	0.281 (0.093)	0.270 (0.088)	0.271 (0.080)
			SOBI	0.268 (0.086)	0.270 (0.081)	0.261 (0.084)
10		MICA	0.292 (0.072)	0.188 (0.075)	0.099 (0.057)	
		DOC	0.452 (0.066)	0.444 (0.068)	0.439 (0.072)	
		SOBI	0.448 (0.053)	0.442 (0.054)	0.440 (0.053)	
exp		5	MICA	0.161 (0.072)	0.074 (0.052)	0.021 (0.024)
			DOC	0.258 (0.076)	0.263 (0.085)	0.253 (0.080)
			SOBI	0.268 (0.086)	0.270 (0.081)	0.261 (0.084)
	10	MICA	0.349 (0.060)	0.261 (0.064)	0.132 (0.052)	
		DOC	0.437 (0.057)	0.435 (0.057)	0.423 (0.055)	
		SOBI	0.448 (0.053)	0.442 (0.054)	0.440 (0.053)	

Table 2: Reported are the proportions of correct segmentations,  $\pi$ , the average and the standard deviation of  $\mathcal{D}^2$ -distance when estimated groups are correct with  $h_0 = 5$  based on 500 replications. Two methods are compared: our method (GMICA), the method in Chang et al. (2018, TS-PCA).

Dist.	$p$	Method	$n = 200$		$n = 500$		$n = 1000$		
			$\pi$	$\mathcal{D}^2$	$\pi$	$\mathcal{D}^2$	$\pi$	$\mathcal{D}^2$	
Normal	6	GMICA	0.684	0.028 (0.020)	0.884	0.011 (0.008)	0.976	0.005 (0.004)	
		TS-PCA	0.682	0.098 (0.056)	0.886	0.040 (0.036)	0.970	0.021 (0.026)	
	12	GMICA	0.364	0.060 (0.070)	0.756	0.018 (0.017)	0.882	0.004 (0.009)	
		TS-PCA	0.152	0.206 (0.072)	0.288	0.119 (0.043)	0.460	0.092 (0.038)	
	Example I	6	GMICA	0.638	0.021 (0.029)	0.858	0.007 (0.006)	0.974	0.004 (0.003)
			TS-PCA	0.612	0.118 (0.065)	0.842	0.055 (0.049)	0.938	0.028 (0.027)
12		GMICA	0.414	0.050 (0.022)	0.714	0.030 (0.057)	0.816	0.005 (0.010)	
		TS-PCA	0.164	0.206 (0.066)	0.372	0.146 (0.065)	0.462	0.102 (0.067)	
exp		6	GMICA	0.640	0.023 (0.028)	0.880	0.008 (0.005)	0.956	0.004 (0.003)
			TS-PCA	0.630	0.104 (0.055)	0.842	0.042 (0.027)	0.938	0.022 (0.016)
	12	GMICA	0.454	0.064 (0.071)	0.752	0.030 (0.031)	0.790	0.001 (0.005)	
		TS-PCA	0.138	0.217 (0.063)	0.338	0.144 (0.051)	0.426	0.102 (0.045)	
Example II	6	GMICA	0.608	0.067 (0.085)	0.702	0.037 (0.057)	0.728	0.002 (0.011)	
		TS-PCA	0.158	0.229 (0.061)	0.160	0.198 (0.071)	0.132	0.171 (0.066)	
	12	GMICA	0.306	0.073 (0.043)	0.468	0.009 (0.023)	0.660	0.005 (0.014)	
		TS-PCA	0.098	0.247 (0.059)	0.088	0.215 (0.054)	0.096	0.194 (0.059)	
	exp	6	GMICA	0.768	0.067 (0.062)	0.920	0.027 (0.029)	1.000	0.001 (0.006)
			TS-PCA	0.136	0.234 (0.064)	0.088	0.214 (0.056)	0.064	0.200 (0.056)
12		GMICA	0.354	0.096 (0.047)	0.556	0.017 (0.031)	0.808	0.007 (0.018)	
		TS-PCA	0.086	0.267 (0.046)	0.044	0.233 (0.053)	0.050	0.221 (0.057)	

Table 3: Average MSPE for community mobility data.

Method		$h_0 = 7$	$h_0 = 14$
	Groups	6 Groups (3,1,1,1,1,1)	6 Groups (3,1,1,1,1,1)
GMICA	MSPE ( $q = 1$ )	1.003	1.002
	MSPE ( $q = 2$ )	0.960	0.952
	Groups	8 Groups	8 Groups
MICA	MSPE ( $q = 1$ )	1.011	1.023
	MSPE ( $q = 2$ )	0.993	0.989
	Groups	7 Groups (2,1,1,1,1,1,1)	7 Groups (2,1,1,1,1,1,1)
TS-PCA	MSPE ( $q = 1$ )	1.112	1.139
	MSPE ( $q = 2$ )	1.164	1.041
	Groups	8 Groups	8 Groups
DOC	MSPE ( $q = 1$ )	1.127	1.123
	MSPE ( $q = 2$ )	1.042	1.064
	Groups	8 Groups	8 Groups
SOBI	MSPE ( $q = 1$ )	1.122	1.155
	MSPE ( $q = 2$ )	1.036	1.045
	Groups	8 Groups	8 Groups
sUARMA	MSPE ( $q = 1$ )	1.235	1.235
	MSPE ( $q = 2$ )	1.281	1.281
	Groups	1 Group	1 Group
sVARMA	MSPE ( $q = 1$ )	1.159	1.159
	MSPE ( $q = 2$ )	1.231	1.231
	Groups	1 Group	1 Group
refVAR	MSPE ( $q = 1$ )	1.583	1.583
	MSPE ( $q = 2$ )	2.353	2.353