Statistica Sinica Preprint No: SS-2025-0459	
Title	A Hybrid Framework Combining Autoregression and Common Factors for Matrix Time Series
Manuscript ID	SS-2025-0459
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0459
Complete List of Authors	Zhiyun Fan, Xiaoyu Zhang and Di Wang
Corresponding Authors	Di Wang
E-mails	di.wang@sjtu.edu.cn
Notice: Accepted author version.	

A HYBRID FRAMEWORK COMBINING AUTOREGRESSION AND COMMON FACTORS FOR MATRIX TIME SERIES

Zhiyun Fan[†], Xiaoyu Zhang[*‡], and Di Wang[*†]

† *Shanghai Jiao Tong University and* ‡ *Tongji University*

Abstract:

Matrix-valued time series are ubiquitous in modern economics and finance, yet modeling them requires navigating a trade-off between flexibility and parsimony. We propose the Matrix Autoregressive model with Common Factors (MARCF), a hybrid structured Reduced-Rank Matrix Autoregression (RRMAR) formulation that balances the flexible predictor-response subspace geometry with the parsimony of the dynamic Matrix Factor Model (MFM), in which low-dimensional latent factors follow autoregressive dynamics. While RRMAR allows arbitrary predictor and response subspaces without explicitly distinguishing their common and specific parts, MARCF explicitly characterizes the intersection of these subspaces. By decomposing the coefficient matrices into common, predictor-specific, and response-specific components, the framework accommodates distinct input and output structures while exploiting their overlap for dimension reduction. We develop a regularized gradient descent estimator that is scalable for high-dimensional data and can efficiently handle the non-convex parameter space. Theoretical analysis establishes local linear convergence of the algorithm and statistical consistency of the estimator under high-dimensional scaling. The estimation efficiency and interpretability of the proposed methods are demonstrated through simulations and an application to a global macroeconomic dataset.

Key words and phrases: Autoregression, dimension reduction, factor model, matrix-valued time series.

*Corresponding author. Email: (Zhang) xzhangck@tongji.edu.cn; (Wang) di.wang@sjtu.edu.cn

1. Introduction

Matrix-valued time series, where each observation is a matrix rather than a vector, have become ubiquitous in modern economics and finance. Prominent examples include multinational macroeconomic indicators (e.g., GDP, production, and interest rates observed across multiple countries) and financial panel data characterized by inherent two-way dependencies. Modeling such data presents a fundamental statistical challenge: one must capture complex temporal dynamics and cross-sectional correlations while respecting the intrinsic matrix structure to maintain estimation efficiency and interpretability.

Currently, the literature is dominated by two distinct modeling paradigms, each dealing with the high-dimensionality of matrix data through different structural assumptions. The first paradigm is the Matrix Autoregressive (MAR) framework, which extends vector autoregression to the matrix domain. For instance, Chen et al. (2021) introduced a bilinear MAR model, while Xiao et al. (2023) proposed the Reduced-Rank MAR (RRMAR) model. A series of extensions have been developed to capture more complex data patterns, such as the spatio-temporal MAR model (Hsu et al., 2021), envelope-based MAR model (Samadi and De Alwis, 2026), sparse MAR model (Jiang et al., 2024), and additive MAR models (Zhang, 2024), among many others. The defining feature of the RRMAR framework is its flexibility: it allows the subspace containing predictive information (the predictor subspace) to be distinct from the subspace in which the future response evolves (the response subspace) (Huang et al., 2025). While this framework allows for complex dynamics, it ignores potential structural overlaps, leading to parameter inefficiency when strong co-movements exist.

The second paradigm is the Matrix Factor Model (MFM) (Wang et al., 2019). These models assume that the high-dimensional observation is driven by a low-dimensional latent

factor process. In the literature, two primary classes of assumptions on factors are commonly adopted. The first class assumes that factors are pervasive and influence most observed series, while allowing weak serial dynamic dependence in the error processes (e.g. Stock and Watson, 2002; Bai and Ng, 2002, 2008; Chen and Fan, 2023). The second assumes that the latent factors capture all dynamic dependencies of the observed process, rendering the error process devoid of serial dependence (e.g. Lam and Yao, 2012; Wang et al., 2019; Gao and Tsay, 2022, 2023). The literature has expanded to include constrained factors (Chen et al., 2020) and tensor-based decompositions (Chang et al., 2023). To perform prediction, dynamic MFMs typically specify a low-dimensional MAR model for the latent factor process (Yu et al., 2024; Qin et al., 2025). Under this formulation, the lagged observation is projected onto the row and column loading spaces to obtain the lagged factor, and the predicted factor is mapped back to the observed matrix through the same loading spaces. Consequently, the lagged predictor and current response in the induced one-step-ahead prediction are represented in the same subspaces. While highly parsimonious, this structure can be restrictive when some directions are mainly predictive or mainly response-related.

This tension is particularly relevant in multinational macroeconomic matrix time series, where each observation can be viewed as a country-by-indicator matrix and the lagged predictor space and future response space are unlikely to be either fully aligned or unrelated. On the country side, global business-cycle movements may generate shared directions across economies, while regional or country-specific adjustment patterns may be mainly predictive or response-related. On the indicator side, broad real-activity and monetary-condition components may be shared, whereas financial-market or demand-side measures may provide additional leading information. These features suggest a natural partial overlap between

the predictor and response subspaces. In such settings, dynamic MFM may impose overly strong sharing through aligned subspaces, whereas RRMAR, although allowing general predictor and response subspaces, does not explicitly exploit their overlap, potentially reducing estimation efficiency and obscuring the interpretation.

Consequently, one would seek a hybrid approach that integrates the flexibility of MAR and the parsimony of MFM, to capture the complex dynamics of matrix-variate data. In the vector time series literature, significant efforts have been made to combine autoregressive dynamics with factor structures. A seminal example is the Factor-Augmented VAR (FAVAR) (Bernanke et al., 2005; Bai et al., 2016), which augments standard VAR systems with factors extracted from large datasets to capture broad economic movements. More recently, Miao et al. (2023) investigated high-dimensional VARs with common factors, employing low-rank constraints on the latent factor component to consistently estimate dynamics in large systems. While these approaches successfully blend factors and autoregressions, they typically treat factors as auxiliary regressors or impose low-rank structures on the vectorized process. Additionally, they do not explicitly model the relationship between the response and predictor spaces. A step in this direction was taken by Wang et al. (2023), who proposed a common basis VAR model that identifies the intersection between predictor and response subspaces. However, all these methods are fundamentally designed for vector-valued data. Direct application to matrix-valued series requires vectorization, which discards the inherent two-way structural information, such as the distinct correlations among countries (rows) and economic indicators (columns), and fails to address the specific identifiability challenges imposed by the bilinear matrix structure.

In this paper, we propose the Matrix Autoregressive model with Common Factors

(MARCF), a structured RRMAR model that provides a hybrid framework to exploit the partial overlap between the response and predictor spaces, which is not fully captured by RRMAR or MFM. Instead of treating the row and column spaces of the coefficient matrices as arbitrary low-rank subspaces, we explicitly model the intersection between the predictor and response subspaces. We decompose the relevant factor spaces into **common components** shared between the past and present, **predictor-specific components** unique to the past, and **response-specific components** unique to the present; see Section 2 for a formal introduction and interpretations of these components. Crucially, this decomposition applies simultaneously to both the row and column dimensions, allowing the model to capture not only symmetric dependencies but also complex interaction dynamics. For instance, the framework can model scenarios where dynamics are driven by a common factor in the cross-section (rows) interacting with a specific leading indicator in the variable dimension (columns). This capability allows the model to reveal asymmetric transmission channels, such as global country-level trends being predicted by specific financial variables.

This decomposition offers three key advantages. First, it refines RRMAR by explicitly parameterizing the overlap between predictor and response subspaces: $d_i = 0$ recovers the standard RRMAR structure, while $d_i = r_i$ corresponds to the aligned-subspace geometry induced by dynamic MFM. Second, exploiting this shared structure reduces the effective number of parameters by approximately $p_1 d_1 + p_2 d_2$ relative to baseline RRMAR, significantly improving estimation efficiency in high-dimensional settings. Third, it provides a transparent interpretation of the sources of variation, allowing practitioners to distinguish between persistent trends, driving forces, and reactive components in complex time series.

To estimate the model, we develop a computationally efficient regularized gradient de-

scent algorithm. Unlike existing methods that rely on alternating minimization (Chen et al., 2021; Xiao et al., 2023), our approach avoids computationally expensive matrix inversions and scales efficiently with the dimension of the data. We tackle the complex identification issues inherent in bilinear models by introducing a balanced regularization strategy, for which we establish theoretical guarantees regarding algorithmic convergence, statistical consistency, and rank selection consistency.

The remainder of the paper is organized as follows. Section 2 introduces the MARCF framework and its relationship to RRMAR and MFM. Section 3 details the estimation strategy, including the regularized loss function, gradient descent algorithm, initialization, and rank selection. Section 4 presents the theoretical properties of the estimator, including computational convergence, initialization guarantees, statistical error rates, and rank selection consistency. Section 5 reports simulation results, and Section 6 illustrates the model performance and interpretability using a multinational macroeconomic dataset. Technical proofs and supplementary details are provided in the supplementary material.

2. Model

2.1 Matrix Factor Model and Reduced-Rank MAR

Matrix-valued time series models extend classical multivariate time series methods to data with inherent two-way structure. A fundamental approach in this domain is the Matrix Factor Model (MFM). The MFM assumes that the high-dimensional observed matrix $\mathbf{Y}_t$ is driven by a low-dimensional latent process, given by

$$\mathbf{Y}_t = \mathbf{\Lambda}_1 \mathbf{F}_t \mathbf{\Lambda}_2^\top + \boldsymbol{\varepsilon}_t, \tag{2.1}$$

where $\boldsymbol{\varepsilon}_t \in \mathbb{R}^{p_1 \times p_2}$ is a matrix-valued error term. Without loss of generality, we impose the standard normalization $\boldsymbol{\Lambda}_i^\top \boldsymbol{\Lambda}_i = \mathbf{I}_{r_i}$ for $i = 1, 2$. As discussed in Section 1, the literature varies in assumptions regarding the factors and error terms. In this paper, we assume the latent factors $\mathbf{F}_t$ drive all temporal dynamics and $\boldsymbol{\varepsilon}_t$ is a white noise process.

Another fundamental approach is the Matrix Autoregressive (MAR) model, which captures bilinear dependencies in an observed time series $\mathbf{Y}_t \in \mathbb{R}^{p_1 \times p_2}$ through the recursion

$$\mathbf{Y}_t = \mathbf{A}_1 \mathbf{Y}_{t-1} \mathbf{A}_2^\top + \mathbf{E}_t, \quad t = 1, 2, \ldots, T, \tag{2.2}$$

where each $\mathbf{A}_i \in \mathbb{R}^{p_i \times p_i}$ is a coefficient matrix and $\mathbf{E}_t$ is a matrix-valued white noise process. To address the high-dimensional challenge of the parameter space, the Reduced-Rank MAR (RRMAR) model (Xiao et al., 2023) imposes low-rank constraints on the coefficient matrices, requiring $\text{rank}(\mathbf{A}_i) = r_i \ll p_i$ for $i = 1, 2$. We parameterize this constraint using the singular value decomposition (SVD) $\mathbf{A}_i = \mathbf{U}_i \mathbf{S}_i \mathbf{V}_i^\top$, where $\mathbf{U}_i \in \mathbb{R}^{p_i \times r_i}$ and $\mathbf{V}_i \in \mathbb{R}^{p_i \times r_i}$ have orthonormal columns, and $\mathbf{S}_i$ is a diagonal matrix of singular values. Although the SVD components are not unique, the column spaces of $\mathbf{U}_i$ and $\mathbf{V}_i$, as well as the corresponding subspace projection matrices $\mathbf{U}_i \mathbf{U}_i^\top$ and $\mathbf{V}_i \mathbf{V}_i^\top$, are uniquely defined.

The RRMAR and MFM frameworks are closely linked through their subspace representations. First, the low-rank structure of the RRMAR allows for a supervised factor analysis interpretation (Wang et al., 2022, 2023; Huang et al., 2025). Specifically, by pre-multiplying (2.2) by $\mathbf{U}_1^\top$ and post-multiplying by $\mathbf{U}_2$, we obtain the following low-dimensional representation of the dynamics:

$$\mathbf{U}_1^\top \mathbf{Y}_t \mathbf{U}_2 = \mathbf{S}_1 (\mathbf{V}_1^\top \mathbf{Y}_{t-1} \mathbf{V}_2) \mathbf{S}_2 + \mathbf{U}_1^\top \mathbf{E}_t \mathbf{U}_2. \tag{2.3}$$

Following Huang et al. (2025), we view $\mathbf{U}_1^\top \mathbf{Y}_t \mathbf{U}_2$ and $\mathbf{V}_1^\top \mathbf{Y}_{t-1} \mathbf{V}_2$ as the *response factor* and

predictor factor. The term "response factor" is justified because the RRMAR model implies

$$\mathbf{Y}_t = \mathbf{U}_1(\mathbf{U}_1^\top \mathbf{Y}_t \mathbf{U}_2)\mathbf{U}_2^\top + (\mathbf{E}_t - \mathbf{U}_1\mathbf{U}_1^\top \mathbf{E}_t \mathbf{U}_2\mathbf{U}_2^\top) =: \mathbf{U}_1(\mathbf{U}_1^\top \mathbf{Y}_t \mathbf{U}_2)\mathbf{U}_2^\top + \widetilde{\mathbf{E}}_t.$$

This takes the form of an MFM as in (2.1) where the low-dimensional transformation $\mathbf{U}_1^\top \mathbf{Y}_t \mathbf{U}_2$ summarizes all predictable information in $\mathbf{Y}_t$. On the other hand, the relationship in (2.3) implies that the predictor factor $\mathbf{V}_1^\top \mathbf{Y}_{t-1} \mathbf{V}_2$ contains all driving forces in the historical data necessary to predict the future response.

Consider now the case where the latent factor $\mathbf{F}_t$ in MFM follows a MAR process, as proposed by Yu et al. (2024):

$$\mathbf{F}_t = \mathbf{B}_1 \mathbf{F}_{t-1} \mathbf{B}_2^\top + \boldsymbol{\xi}_t, \tag{2.4}$$

where $\mathbf{B}_i \in \mathbb{R}^{r_i \times r_i}$ are coefficient matrices and $\boldsymbol{\xi}_t$ is low-dimensional white noise. Substituting (2.4) into (2.1), the structured dynamics become

$$\mathbf{Y}_t = (\boldsymbol{\Lambda}_1 \mathbf{B}_1 \boldsymbol{\Lambda}_1^\top)\mathbf{Y}_{t-1}(\boldsymbol{\Lambda}_2 \mathbf{B}_2^\top \boldsymbol{\Lambda}_2^\top) + \mathbf{N}_t, \tag{2.5}$$

where $\mathbf{N}_t$ is a composite noise term involving $\boldsymbol{\xi}_t$ and a moving average of $\boldsymbol{\varepsilon}_t$. Equation (2.5) reveals the implicit MAR structure of this dynamic MFM, with the effective coefficient matrices $\boldsymbol{\Lambda}_i \mathbf{B}_i \boldsymbol{\Lambda}_i^\top$. Comparing this to the RRMAR decomposition in (2.3), we observe a fundamental restriction in the MFM framework: the predictor subspace and the response subspace are identical (both spanned by $\boldsymbol{\Lambda}_i$). While this symmetry is parsimonious, it prevents the model from capturing rotational dynamics where the relevant subspaces shift between the predictor and the response. In contrast, the RRMAR model allows $\mathbf{V}_i$ and $\mathbf{U}_i$ to be distinct, offering greater flexibility but potentially ignoring structural commonalities shared by the response and predictor.

2.2 Matrix Autoregression with Common Factors

To reconcile the distinct subspace assumption of the RRMAR model with the shared subspace assumption of the MFM, we propose a hybrid framework. This framework explicitly parameterizes the intersection of the predictor and response subspaces. By controlling the degree of this overlap, we can capture shared dynamics while accommodating distinct structural features in the response and predictor spaces.

Let $\mathcal{M}(\mathbf{U}_i)$ and $\mathcal{M}(\mathbf{V}_i)$ denote the response and predictor subspaces, respectively. We assume their intersection has dimension d_i, where $0 \leq d_i \leq r_i$. The parameters d_1 and d_2 quantify the extent of structural overlap. Given these dimensions, there exist orthogonal matrices $\mathbf{O}_{i1}$ and $\mathbf{O}_{i2}$ allowing the decomposition:

$$\mathbf{U}_i\mathbf{O}_{i1} = [\mathbf{C}_i\ \mathbf{R}_i] \quad \text{and} \quad \mathbf{V}_i\mathbf{O}_{i2} = [\mathbf{C}_i\ \mathbf{P}_i].$$

Here, $\mathbf{C}_i \in \mathbb{R}^{p_i \times d_i}$ represents the *common subspace* shared by both responses and predictors. The matrix $\mathbf{R}_i \in \mathbb{R}^{p_i \times (r_i - d_i)}$ spans the *response-specific subspace*, while $\mathbf{P}_i \in \mathbb{R}^{p_i \times (r_i - d_i)}$ spans the *predictor-specific subspace*.

Substituting this decomposition into the general bilinear framework yields the *Matrix Autoregressive model with Common Factors* (MARCF):

$$\mathbf{Y}_t = [\mathbf{C}_1\ \mathbf{R}_1]\mathbf{D}_1[\mathbf{C}_1\ \mathbf{P}_1]^\top \mathbf{Y}_{t-1}[\mathbf{C}_2\ \mathbf{P}_2]\mathbf{D}_2^\top[\mathbf{C}_2\ \mathbf{R}_2]^\top + \mathbf{E}_t. \tag{2.6}$$

In this formulation, $\mathbf{D}_i = \mathbf{O}_{i1}^\top \mathbf{S}_i \mathbf{O}_{i2}$ serves as a dense mixing matrix within the low-dimensional latent space.

This specification provides a hybrid configuration between the two existing subspace paradigms. When we set $d_1 = d_2 = 0$, the common component vanishes, and the model reduces to the standard RRMAR framework. Conversely, setting $d_i = r_i$, the specific com-

ponents vanish, and the predictor and response subspaces align fully, matching the subspace geometry induced by the dynamic MFM in (2.5). By selecting d_1 and d_2 data-drivenly, the MARCF model adaptively balances the efficiency of shared factors with the flexibility of allowing the predictor and response subspaces to differ.

For interpretation, we let $\mathbf{C}_i^\top \mathbf{C}_i = \mathbf{I}_{d_i}$, $\mathbf{R}_i^\top \mathbf{R}_i = \mathbf{P}_i^\top \mathbf{P}_i = \mathbf{I}_{r_i - d_i}$, and $\mathbf{C}_i^\top \mathbf{R}_i = \mathbf{C}_i^\top \mathbf{P}_i = \mathbf{0}_{d_i \times (r_i - d_i)}$. The detailed identification conditions are formally presented in Section 2.4. By pre-multiplying (2.6) by $[\mathbf{C}_1\ \mathbf{R}_1]^\top$ and post-multiplying by $[\mathbf{C}_2\ \mathbf{R}_2]$, we have

$$\begin{bmatrix} \mathbf{C}_1^\top \mathbf{Y}_t \mathbf{C}_2 & \mathbf{C}_1^\top \mathbf{Y}_t \mathbf{R}_2 \\ \mathbf{R}_1^\top \mathbf{Y}_t \mathbf{C}_2 & \mathbf{R}_1^\top \mathbf{Y}_t \mathbf{R}_2 \end{bmatrix} = \mathbf{D}_1 \begin{bmatrix} \mathbf{C}_1^\top \mathbf{Y}_{t-1} \mathbf{C}_2 & \mathbf{C}_1^\top \mathbf{Y}_{t-1} \mathbf{P}_2 \\ \mathbf{P}_1^\top \mathbf{Y}_{t-1} \mathbf{C}_2 & \mathbf{P}_1^\top \mathbf{Y}_{t-1} \mathbf{P}_2 \end{bmatrix} \mathbf{D}_2^\top + \text{Noise}.$$

This representation extends the factor interpretation in (2.3) by explicitly characterizing the *common factors* shared by responses and predictors, alongside their specific counterparts.

Furthermore, when d_1 and d_2 are nonzero, the proposed framework achieves significant dimension reduction. The total number of free parameters in the MARCF model is given by

$$\mathrm{df}_{\mathrm{MARCF}} = p_1(2r_1 - d_1) + p_2(2r_2 - d_2) + r_1^2 + r_2^2. \tag{2.7}$$

This contrasts with $\mathrm{df}_{\mathrm{RRMAR}} = 2p_1 r_1 + 2p_2 r_2 + r_1^2 + r_2^2$ for the standard reduced-rank model. In high-dimensional settings, the MARCF model reduces the effective dimension by approximately $p_1 d_1 + p_2 d_2$, enhancing both estimation efficiency and forecasting accuracy.

2.3 Model Interpretation

To illustrate the interpretation of MARCF components, consider the multinational macroeconomic application discussed in Section 1, where rows represent countries and columns represent economic indicators.

(1) Two-way Dynamics for Rows (Country-Level) and Columns (Indicator-Level): On the country and indicator sides of the matrix autoregression, the predictor and response subspaces are decomposed into common, predictor-specific, and response-specific subspaces, summarizing distinct information of the country-indicator macroeconomic co-movement.

- **Common Subspaces ($\mathbf{C}_1$ and $\mathbf{C}_2$):** These subspaces identify systemic patterns or persistent influence shared by both predictors and responses. For the country side, the basis directions may correspond to aggregation of large advanced economies being influential on both predicting and being predicted, such as the G7 nations, or countries with distinctive but shared dynamics. For the economic indicator side, the directions can represent fundamental macroeconomic aggregates like industrial production and GDP, which are auto-correlated and central to describing the economic cycle.

- **Predictor-Specific Subspaces ($\mathbf{P}_1$ and $\mathbf{P}_2$):** These subspaces capture information summarized only for prediction. The corresponding factors contain rich information about the future direction of the economy but may be less relevant for defining the real economic state at the current moment. On the country side, the subspace may represent the additional information provided by some influential countries or small and open advanced economies, such as the United Kingdom and the Netherlands. For indicators, it may correspond to leading variables such as share prices, or informative combinations of indicators that reveal useful predictive information in addition to $\mathbf{C}_2$.

- **Response-Specific Subspaces ($\mathbf{R}_1$ and $\mathbf{R}_2$):** These capture the projection directions that are "reactive" to changes in the global economy. They may absorb global shocks and exhibit volatility in the response period, but their idiosyncratic fluctuations have little predictive power for the future global state. On the country side, the

subspace may correspond to small open economies, such as the Netherlands. For the indicator side, the directions may correspond to lagging information that reacts to past economic conditions but is slow to adjust and does not necessarily drive future cycles.

(2) Interaction Dynamics (Cross-Sectional Spillovers): Beyond the diagonal blocks, the MARCF model captures crucial cross-terms (e.g., corresponding to $\mathbf{C}_1^\top \mathbf{Y}_{t-1} \mathbf{P}_2$). These interactions allow for asymmetric dependencies where the source of predictive information differs structurally from the response it generates. For instance, the model can describe how *predictor-specific information* (e.g., forward-looking market information gathered by $\mathbf{P}_2$) in *common large economies* ($\mathbf{C}_1$) drives future *common real output* ($\mathbf{C}_2$). This capability allows the model to capture dynamic transformations where the factors driving prediction differ from those characterizing the response, contrasting with standard factor models that enforce a restrictive symmetry.

2.4 Stationarity and Identification

The MARCF model defined in (2.6) belongs to the general class of bilinear MAR processes. Following Chen et al. (2021), weak stationarity is guaranteed when the spectral radii of the coefficient matrices satisfy $\rho(\mathbf{A}_1)\rho(\mathbf{A}_2) < 1$, where $\mathbf{A}_i = [\mathbf{C}_i \ \mathbf{R}_i]\mathbf{D}_i[\mathbf{C}_i \ \mathbf{P}_i]^\top$ for $i = 1, 2$.

The model structure introduces identification challenges at two distinct levels. First, the coefficient matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ are subject to global scaling indeterminacy: for any nonzero scalar c, the pairs $(\mathbf{A}_1, \mathbf{A}_2)$ and $(c\mathbf{A}_1, c^{-1}\mathbf{A}_2)$ yield identical $\mathbf{A} = \mathbf{A}_2 \otimes \mathbf{A}_1$. In other words, the Kronecker product $\mathbf{A}$ is identifiable, but the row and column coefficient matrices are not. To resolve this, we adopt the constraint $\|\mathbf{A}_1\|_{\mathrm{F}} = \|\mathbf{A}_2\|_{\mathrm{F}}$. We define the identifiable true value of $\mathbf{A}$ as $\mathbf{A}^*$, and define the signal strength of the true model as $\phi = \|\mathbf{A}^*\|_{\mathrm{F}}^{1/2}$ such

that $\|\mathbf{A}_1^*\|_{\mathrm{F}} = \|\mathbf{A}_2^*\|_{\mathrm{F}} = \phi$.

Second, given a fixed $\mathbf{A}_i$, the decomposition into components $(\mathbf{C}_i, \mathbf{R}_i, \mathbf{P}_i, \mathbf{D}_i)$ is not unique. Specifically, for any invertible matrices $\mathbf{Q}_1 \in \mathbb{R}^{d_i \times d_i}$ and $\mathbf{Q}_2, \mathbf{Q}_3 \in \mathbb{R}^{(r_i-d_i)\times(r_i-d_i)}$, the parameter sets

$$(\mathbf{C}_i, \mathbf{R}_i, \mathbf{P}_i, \mathbf{D}_i) \quad \text{and} \quad (\mathbf{C}_i\mathbf{Q}_1, \mathbf{R}_i\mathbf{Q}_2, \mathbf{P}_i\mathbf{Q}_3, \mathrm{diag}(\mathbf{Q}_1^{-1}, \mathbf{Q}_2^{-1})\mathbf{D}_i\mathrm{diag}(\mathbf{Q}_1^{-\top}, \mathbf{Q}_3^{-\top}))$$

are equivalent. To resolve this internal indeterminacy and balance these components, we impose the following constraints:

$$\mathbf{C}_i^\top\mathbf{C}_i = b^2\mathbf{I}_{d_i}, \quad \mathbf{R}_i^\top\mathbf{R}_i = \mathbf{P}_i^\top\mathbf{P}_i = b^2\mathbf{I}_{r_i-d_i}, \quad \text{and} \quad \mathbf{C}_i^\top\mathbf{R}_i = \mathbf{C}_i^\top\mathbf{P}_i = \mathbf{0}_{d_i\times(r_i-d_i)}, \tag{2.8}$$

where $b > 0$ is a balancing scalar. The condition $\mathbf{C}_i^\top\mathbf{R}_i = \mathbf{C}_i^\top\mathbf{P}_i = \mathbf{0}_{d_i\times(r_i-d_i)}$ is imposed only on the basis representation for identifiability and an observed row or column variable may still have nonzero loadings on both common and specific basis directions. Relaxing this formulation would require a different notion of commonality, for example near-common directions characterized by small principal angles between the predictor and response subspaces. We leave this extension for future work. To ensure that $\mathbf{D}_i$ is of the same order of magnitude as $(\mathbf{C}_i, \mathbf{R}_i, \mathbf{P}_i)$, we set $b \asymp \phi^{1/3}$. This choice implies that $\|\mathbf{C}_i\|_{\mathrm{F}} \asymp \|\mathbf{R}_i\|_{\mathrm{F}} \asymp \|\mathbf{P}_i\|_{\mathrm{F}} \asymp \|\mathbf{D}_i\|_{\mathrm{F}}$ when r_i and d_i remain constant.

Under the balancing constraint in (2.8), the components are determined uniquely up to orthogonal rotations. We treat parameterizations that differ only by such rotations as equivalent. Let $\boldsymbol{\Theta}$ denote the full set of component matrices. For any two sets $\boldsymbol{\Theta}$ and $\boldsymbol{\Theta}'$ satisfying the identification conditions in (2.8), we define their distance as

$$\begin{aligned}\mathrm{dist}(\boldsymbol{\Theta}, \boldsymbol{\Theta}')^2 = \min_{\substack{\mathbf{Q}_{i,1}\in\mathbb{O}^{d_i},\\ \mathbf{Q}_{i,2},\mathbf{Q}_{i,3}\in\mathbb{O}^{r_i-d_i}}} \sum_{i=1}^{2} \Big\{ &\|\mathbf{C}_i - \mathbf{C}_i'\mathbf{Q}_{i,1}\|_{\mathrm{F}}^2 + \|\mathbf{R}_i - \mathbf{R}_i'\mathbf{Q}_{i,2}\|_{\mathrm{F}}^2 + \|\mathbf{P}_i - \mathbf{P}_i'\mathbf{Q}_{i,3}\|_{\mathrm{F}}^2 \\ &+ \|\mathbf{D}_i - \mathrm{diag}(\mathbf{Q}_{i,1}, \mathbf{Q}_{i,2})^\top\mathbf{D}_i'\mathrm{diag}(\mathbf{Q}_{i,1}, \mathbf{Q}_{i,3})\|_{\mathrm{F}}^2\Big\},\end{aligned} \tag{2.9}$$

where $\mathbb{O}^k$ denotes the set of $k \times k$ orthogonal matrices. This metric remains invariant under rotations and forms the basis of our theoretical analysis in Section 4.

Remark 1 (Interpretation vs. Estimation)**.** The parameterization with $b \asymp \phi^{1/3}$ is designed to facilitate theoretical analysis and numerical optimization. However, for model interpretation as discussed in Section 2.3, it is often more intuitive to work with orthonormal bases. One can always transform the estimated parameters to an equivalent representation where $b = 1$ (i.e., $\mathbf{C}_i^\top \mathbf{C}_i = \mathbf{I}_{d_i}$) by absorbing the scaling factors into $\mathbf{D}_i$. The interpretation of the subspaces and dynamic factors in Section 2.3 remains invariant under this rescaling.

3. Estimation and Modeling Procedure

This section presents the estimation methodology for the MARCF model (2.6). We employ a computationally efficient non-convex regularized gradient descent algorithm. We first define the optimization objective, which incorporates specific regularization terms to enforce the balanced identification constraints discussed in Section 2.4. Subsequently, we describe a theoretically guaranteed initialization procedure that exploits the subspace structures from a convex penalized estimator. Finally, we propose a data-driven procedure for sequentially selecting the model ranks (r_1, r_2) and the common subspace dimensions (d_1, d_2).

3.1 Regularized Gradient Descent Estimation

Consider the observed matrix-valued time series $\{\mathbf{Y}_t\}_{t=0}^T$. For a fixed set of structural dimensions (r_1, r_2, d_1, d_2), the model parameters are collected in $\mathbf{\Theta} = (\{\mathbf{C}_i, \mathbf{R}_i, \mathbf{P}_i, \mathbf{D}_i\}_{i=1}^2)$. Recall that the implied coefficient matrices are given by

$$\mathbf{A}_i = [\mathbf{C}_i \ \mathbf{R}_i]\mathbf{D}_i[\mathbf{C}_i \ \mathbf{P}_i]^\top, \quad \text{for } i = 1, 2. \tag{3.10}$$

Our estimation procedure minimizes the least-squares loss function:

$$\mathcal{L}(\mathbf{\Theta}) = \frac{1}{2T}\sum_{t=1}^{T}\left\|\mathbf{Y}_t - [\mathbf{C}_1\ \mathbf{R}_1]\mathbf{D}_1[\mathbf{C}_1\ \mathbf{P}_1]^\top \mathbf{Y}_{t-1}[\mathbf{C}_2\ \mathbf{P}_2]\mathbf{D}_2^\top[\mathbf{C}_2\ \mathbf{R}_2]^\top\right\|_{\mathrm{F}}^2.$$

To ensure estimation stability and satisfy the identification conditions outlined in Section 2.4, we incorporate two types of regularization. First, to resolve the global scaling indeterminacy between $\mathbf{A}_1$ and $\mathbf{A}_2$, we introduce a regularization term that encourages their Frobenius norms to be equal:

$$\mathcal{R}_1(\mathbf{\Theta}) = (\|[\mathbf{C}_1\ \mathbf{R}_1]\mathbf{D}_1[\mathbf{C}_1\ \mathbf{P}_1]^\top\|_{\mathrm{F}}^2 - \|[\mathbf{C}_2\ \mathbf{P}_2]\mathbf{D}_2^\top[\mathbf{C}_2\ \mathbf{R}_2]^\top\|_{\mathrm{F}}^2)^2 = (\|\mathbf{A}_1\|_{\mathrm{F}}^2 - \|\mathbf{A}_2\|_{\mathrm{F}}^2)^2.$$

This regularization term, consistent with the related work on low-rank matrix estimation (Tu et al., 2016; Wang et al., 2017), helps stabilize the estimation of these two coefficient matrices. Second, for a given balance scalar b, inspired by Han et al. (2022) and Wang et al. (2023), we introduce a second regularization term to enforce the identification constraints on the components in (2.8):

$$\mathcal{R}_2(\mathbf{\Theta}; b) := \sum_{i=1}^{2}\left(\|[\mathbf{C}_i\ \mathbf{R}_i]^\top[\mathbf{C}_i\ \mathbf{R}_i] - b^2\mathbf{I}_{r_i}\|_{\mathrm{F}}^2 + \|[\mathbf{C}_i\ \mathbf{P}_i]^\top[\mathbf{C}_i\ \mathbf{P}_i] - b^2\mathbf{I}_{r_i}\|_{\mathrm{F}}^2\right).$$

The full regularized objective function is

$$\overline{\mathcal{L}}(\mathbf{\Theta}; \lambda_1, \lambda_2, b) = \mathcal{L}(\mathbf{\Theta}) + \frac{\lambda_1}{4}\mathcal{R}_1(\mathbf{\Theta}) + \frac{\lambda_2}{4}\mathcal{R}_2(\mathbf{\Theta}; b), \tag{3.11}$$

where λ_1 and λ_2 are tuning parameters.

The model parameters are estimated by minimizing $\overline{\mathcal{L}}$ using a standard gradient descent algorithm, as outlined in Algorithm 1. Here, $\nabla_{\mathbf{M}}\overline{\mathcal{L}}^{(j)}$ denotes the partial gradient of $\overline{\mathcal{L}}$ with respect to any component $\mathbf{M} \in \{\mathbf{C}_i, \mathbf{R}_i, \mathbf{P}_i, \mathbf{D}_i\}_{i=1}^2$, evaluated at the parameters $\mathbf{\Theta}^{(j)}$ after j iterations. The partial gradients with respect to $\mathbf{C}_i$, $\mathbf{R}_i$, $\mathbf{P}_i$, and $\mathbf{D}_i$ are derived from the chain rule and are provided in detail in Appendix S2. Crucially, all update steps involve

only matrix multiplications and additions, avoiding expensive matrix inversions or SVDs at each iteration, making the method highly scalable to high-dimensional data.

Algorithm 1 Gradient Descent Algorithm for MARCF

input: data $\{\mathbf{Y}_t\}_{t=0}^{T}$, initial values $\boldsymbol{\Theta}^{(0)}$, r_1, r_2, d_1, d_2, hyperparameters λ_1, λ_2, b, step size η, and the maximum number of iterations J.

for $j = 0$ to $J-1$ **do**

  for $i = 1$ to 2 **do**

$$\mathbf{C}_i^{(j+1)} = \mathbf{C}_i^{(j)} - \eta \nabla_{\mathbf{C}_i} \overline{\mathcal{L}}^{(j)}, \quad \mathbf{R}_i^{(j+1)} = \mathbf{R}_i^{(j)} - \eta \nabla_{\mathbf{R}_i} \overline{\mathcal{L}}^{(j)},$$

$$\mathbf{P}_i^{(j+1)} = \mathbf{P}_i^{(j)} - \eta \nabla_{\mathbf{P}_i} \overline{\mathcal{L}}^{(j)}, \quad \mathbf{D}_i^{(j+1)} = \mathbf{D}_i^{(j)} - \eta \nabla_{\mathbf{D}_i} \overline{\mathcal{L}}^{(j)}.$$

  end for

end for

output: $\widehat{\boldsymbol{\Theta}} = (\{\mathbf{C}_i^{(J)}, \mathbf{R}_i^{(J)}, \mathbf{P}_i^{(J)}, \mathbf{D}_i^{(J)}\}_{i=1}^{2})$.

3.2 Algorithm Initialization

Since the optimization landscape of (3.11) is nonconvex, the global convergence of Algorithm 1 depends critically on reliable initialization. We propose a theoretically guaranteed initialization strategy based on a convex nuclear-norm penalized VAR estimator, the nearest Kronecker product (NKP) approximation, and spectral methods. Specifically, given the dimensions (r_1, r_2, d_1, d_2) and scalar b, the initialization proceeds in two steps:

1. Obtain a first-stage initial value from the vectorized nuclear-norm penalized VAR model:

$$\widehat{\mathbf{A}}^{\mathrm{VAR}} = \underset{\mathbf{A} \in \mathbb{R}^{p_1 p_2 \times p_1 p_2}}{\arg\min} \frac{1}{2T} \sum_{t=1}^{T} \|\mathrm{vec}(\mathbf{Y}_t) - \mathbf{A}\mathrm{vec}(\mathbf{Y}_{t-1})\|_2^2 + \lambda_0 \|\mathbf{A}\|_*,$$

where $\|\cdot\|_*$ is the matrix nuclear norm. Then, applying the NKP approximation yields

$$(\widehat{\mathbf{A}}_1^{\mathrm{KP}}, \widehat{\mathbf{A}}_2^{\mathrm{KP}}) = \underset{\|\mathbf{A}_1\|_{\mathrm{F}}=\|\mathbf{A}_2\|_{\mathrm{F}}}{\arg\min} \|\widehat{\mathbf{A}}^{\mathrm{VAR}} - \mathbf{A}_2 \otimes \mathbf{A}_1\|_{\mathrm{F}}^2. \tag{3.12}$$

As discussed in Chen et al. (2021), the optimal solution of (3.12) exists with an explicit form; see Appendix S4.4 for details. Next, for $i = 1, 2$, we compute the truncated SVD of $\widehat{\mathbf{A}}_i^{\mathrm{KP}}$, keeping only the largest r_i singular values, to obtain an initialization of the low rank transition matrices: $\widehat{\mathbf{A}}_i^{\mathrm{RR}} = \widehat{\mathbf{U}}_i \widehat{\mathbf{S}}_i \widehat{\mathbf{V}}_i^\top$. Their left and right singular subspaces estimate the total response and predictor subspaces, respectively.

2. Use spectral methods to decompose the subspaces of $\widehat{\mathbf{A}}_1^{\mathrm{RR}}$ and $\widehat{\mathbf{A}}_2^{\mathrm{RR}}$ to obtain $\boldsymbol{\Theta}^{(0)}$. We first discuss the rationale of the spectral method. Denote $\mathbf{U}_i := [\mathbf{C}_i\ \mathbf{R}_i]$ and $\mathbf{V}_i := [\mathbf{C}_i\ \mathbf{P}_i]$ with orthonormal columns. For any matrix $\mathbf{M}$ with orthonormal columns, denote $\mathcal{P}_{\mathbf{M}} := \mathbf{M}\mathbf{M}^\top$ as the subspace projection matrix and $\mathcal{P}_{\mathbf{M}}^\perp := \mathbf{I} - \mathbf{M}\mathbf{M}^\top$ as the subspace orthogonal complement projector. Straightforward algebra shows that $\mathcal{P}_{\mathbf{U}_i}\mathcal{P}_{\mathbf{V}_i}^\perp = \mathcal{P}_{\mathbf{R}_i}\mathcal{P}_{\mathbf{P}_i}^\perp$. This implies that $\mathcal{M}(\mathbf{R}_i)$ is spanned by the first $r_i - d_i$ left singular vectors of $\mathcal{P}_{\mathbf{U}_i}\mathcal{P}_{\mathbf{V}_i}^\perp$. A similar argument holds for $\mathcal{M}(\mathbf{P}_i)$. Additionally, $\mathcal{P}_{\mathbf{R}_i}^\perp \mathcal{P}_{\mathbf{P}_i}^\perp (\mathcal{P}_{\mathbf{U}_i} + \mathcal{P}_{\mathbf{V}_i}) \mathcal{P}_{\mathbf{R}_i}^\perp \mathcal{P}_{\mathbf{P}_i}^\perp = 2\,\mathcal{P}_{\mathbf{C}_i}$, implying that $\mathcal{M}(\mathbf{C}_i)$ is spanned by the leading d_i eigenvectors of the left-hand side. Based on these, we obtain the initial values as follows:

   (a) For each $i = 1, 2$, using the SVD components in the first step, calculate the top $r_i - d_i$ left singular vectors of $\mathcal{P}_{\widehat{\mathbf{U}}_i}\mathcal{P}_{\widehat{\mathbf{V}}_i}^\perp$ and $\mathcal{P}_{\widehat{\mathbf{V}}_i}\mathcal{P}_{\widehat{\mathbf{U}}_i}^\perp$, and denote them by $\widetilde{\mathbf{R}}_i$ and $\widetilde{\mathbf{P}}_i$, respectively. Calculate the top d_i eigenvectors of $\mathcal{P}_{\widetilde{\mathbf{R}}_i}^\perp \mathcal{P}_{\widetilde{\mathbf{P}}_i}^\perp \left(\mathcal{P}_{\widehat{\mathbf{U}}_i} + \mathcal{P}_{\widehat{\mathbf{V}}_i}\right) \mathcal{P}_{\widetilde{\mathbf{R}}_i}^\perp \mathcal{P}_{\widetilde{\mathbf{P}}_i}^\perp$, and denote them by $\widetilde{\mathbf{C}}_i$. Then calculate $\widetilde{\mathbf{D}}_1 = [\widetilde{\mathbf{C}}_1\ \widetilde{\mathbf{R}}_1]^\top\ \widehat{\mathbf{A}}_1^{\mathrm{RR}} [\widetilde{\mathbf{C}}_1\ \widetilde{\mathbf{P}}_1]$ and $\widetilde{\mathbf{D}}_2 = [\widetilde{\mathbf{C}}_2\ \widetilde{\mathbf{R}}_2]^\top\ \widehat{\mathbf{A}}_2^{\mathrm{RR}} [\widetilde{\mathbf{C}}_2\ \widetilde{\mathbf{P}}_2]$.

   (b) For each $i = 1, 2$, set $\mathbf{C}_i^{(0)} = b\widetilde{\mathbf{C}}_i, \mathbf{R}_i^{(0)} = b\widetilde{\mathbf{R}}_i, \mathbf{P}_i^{(0)} = b\widetilde{\mathbf{P}}_i$, and $\mathbf{D}_i^{(0)} = b^{-2}\widetilde{\mathbf{D}}_i$.

3.3 Selection of Ranks and Common Dimensions

Although our estimation methodology is applicable to any pre-specified (r_1, r_2, d_1, d_2), the appropriate selection of these parameters is critical in practice. These structural parameters define the model's complexity and determine the interpretability of the factors. We propose a two-step data-driven procedure to select (r_1, r_2) and (d_1, d_2) sequentially.

For rank selection, since r_1 and r_2 are typically much smaller than p_1 and p_2, we choose two upper bounds for the selection: $\bar{r}_1 \ll p_1$ and $\bar{r}_2 \ll p_2$. Then, we compute the initial low rank estimates as described in the first step of Section 3.2 with $\bar{r}_1$ and $\bar{r}_2$, denoted by $\widehat{\mathbf{A}}_1^{\mathrm{RR}}(\bar{r}_1)$ and $\widehat{\mathbf{A}}_2^{\mathrm{RR}}(\bar{r}_2)$, respectively. Let $\widehat{\sigma}_{i,1} \geq \widehat{\sigma}_{i,2} \geq \cdots \geq \widehat{\sigma}_{i,\bar{r}_i}$ be the singular values of $\widehat{\mathbf{A}}_i^{\mathrm{RR}}(\bar{r}_i)$. Motivated by Xia et al. (2015), we introduce a ridge-type ratio to select r_1 and r_2 separately:

$$\widehat{r}_i = \underset{1 \leqslant j < \bar{r}_i}{\arg\min} \frac{\widehat{\sigma}_{i,j+1} + s(p_1, p_2, T)}{\widehat{\sigma}_{i,j} + s(p_1, p_2, T)},$$

where we suggest $s(p_1, p_2, T) = \sqrt{p_1 p_2 \log(T)/(200T)}$. The theoretical guarantee of this procedure is given in Section 4.3, and its empirical performance is justified by the numerical experiments in Section 5.

After r_1 and r_2 are determined, the problem reduces to selecting a model in low dimensions. We use the BIC criterion to select d_1 and d_2. Let $\widehat{\mathbf{A}}(\widehat{r}_1, \widehat{r}_2, d_1, d_2)$ be the estimator of $\mathbf{A}_2 \otimes \mathbf{A}_1$ in (3.10) under $\widehat{r}_i$ and d_i. Then, we define

$$\mathrm{BIC}(d_1, d_2) = T p_1 p_2 \log\left(\sum_{t=1}^{T} \|\mathrm{vec}(\mathbf{Y}_t) - \widehat{\mathbf{A}}(\widehat{r}_1, \widehat{r}_2, d_1, d_2)\mathrm{vec}(\mathbf{Y}_{t-1})\|_2^2\right) + \mathrm{df}_{\mathrm{MARCF}} \log(T),$$

where $\mathrm{df}_{\mathrm{MARCF}}$ is defined in (2.7) with $\widehat{r}_1$ and $\widehat{r}_2$. Finally, d_1 and d_2 are chosen as:

$$(\widehat{d}_1, \widehat{d}_2) = \underset{0 \leq d_i \leq \widehat{r}_i, i=1,2}{\arg\min} \mathrm{BIC}(d_1, d_2).$$

Remark 2. In practice, BIC can also be used to select (r_1, r_2). While simultaneous selection of $(r_1,\ r_2,\ d_1,\ d_2)$ via BIC can also be considered, the search space of size $O(\bar{r}_1^2\bar{r}_2^2)$ may be computationally prohibitive. The proposed sequential approach balances statistical accuracy with computational feasibility. Rolling window cross-validation may also be used for hyperparameter tuning but incurs significantly higher computational costs.

4. Theory

In this section, we investigate the theoretical properties of the proposed estimation methodology for the MARCF model. To facilitate a clear understanding of the results, we outline a roadmap that decouples the optimization landscape from the stochastic nature of the data.

First, Section 4.1 establishes a deterministic local convergence result. Theorem 1 shows that Algorithm 1 achieves local linear convergence, conditioned on a valid initialization and the local curvature of the loss function, characterized by Restricted Strong Convexity and Smoothness (RSC and RSS, respectively). Subsequently, in Section 4.2, we provide the stochastic analysis under mild assumptions on the data-generating process. Specifically, Proposition 1 verifies the RSC/RSS condition, Proposition 2 establishes the theoretical guarantee for the initialization procedure, and Theorem 2 combines these results to give an overall statistical guarantee for the complete two-step estimation procedure. Finally, Section 4.3 proves the consistency of the rank selection procedure.

4.1 Computational Convergence Analysis

We first analyze the convergence properties of Algorithm 1. Due to the nonconvex nature of the objective function $\overline{\mathcal{L}}$, our analysis relies on the assumptions of restricted strong convexity

(RSC) and restricted strong smoothness (RSS), which characterize the curvature of the loss function along the directions of relevant low-rank matrices. These properties are defined with respect to the Kronecker product parameter matrix $\mathbf{A} = \mathbf{A}_2 \otimes \mathbf{A}_1$. Let $\widetilde{\mathcal{L}}(\mathbf{A})$ denote the least-squares loss function with respect to $\mathbf{A}$:

$$\widetilde{\mathcal{L}}(\mathbf{A}) = \frac{1}{2T} \sum_{t=1}^{T} \|\mathrm{vec}(\mathbf{Y}_t) - \mathbf{A}\mathrm{vec}(\mathbf{Y}_{t-1})\|_2^2 .$$

Definition 1 (RSC/RSS Condition)**.** The loss function $\widetilde{\mathcal{L}}(\mathbf{A})$ is said to be restricted strongly convex (RSC) with parameter α and restricted strongly smooth (RSS) with parameter β (where $0 < \alpha \leq \beta$) if, for the true value $\mathbf{A}^* = \mathbf{A}_2^* \otimes \mathbf{A}_1^*$ and any parameter matrix $\mathbf{A} = \mathbf{A}_2 \otimes \mathbf{A}_1$ admitting the MARCF structural decomposition presented in (3.10), the following inequality holds:

$$\frac{\alpha}{2} \|\mathbf{A} - \mathbf{A}^*\|_{\mathrm{F}}^2 \leqslant \widetilde{\mathcal{L}}(\mathbf{A}) - \widetilde{\mathcal{L}}(\mathbf{A}^*) - \left\langle \nabla \widetilde{\mathcal{L}}(\mathbf{A}^*), \mathbf{A} - \mathbf{A}^* \right\rangle \leqslant \frac{\beta}{2} \|\mathbf{A} - \mathbf{A}^*\|_{\mathrm{F}}^2 .$$

The statistical error is quantified by the following deviation bound.

Definition 2 (Deviation Bound)**.** For given ranks (r_1, r_2), common subspace dimensions (d_1, d_2), and the identifiable true value $\mathbf{A}^*$, the deviation bound is defined as

$$\xi(r_1, r_2, d_1, d_2) := \sup_{\substack{[\mathbf{C}_i \ \mathbf{R}_i] \in \mathbb{O}^{p_i \times r_i}, \\ [\mathbf{C}_i \ \mathbf{P}_i] \in \mathbb{O}^{p_i \times r_i}, \\ \mathbf{D}_i \in \mathbb{R}^{r_i \times r_i}, \|\mathbf{D}_i\|_{\mathrm{F}} = 1}} \left\langle \nabla \widetilde{\mathcal{L}}(\mathbf{A}^*), [\mathbf{C}_2 \ \mathbf{R}_2]\mathbf{D}_2[\mathbf{C}_2 \ \mathbf{P}_2]^\top \otimes [\mathbf{C}_1 \ \mathbf{R}_1]\mathbf{D}_1[\mathbf{C}_1 \ \mathbf{P}_1]^\top \right\rangle .$$

To quantify the estimation error, we consider $\|\mathbf{A}_2 \otimes \mathbf{A}_1 - \mathbf{A}^*\|_{\mathrm{F}}^2$, which is invariant under scaling and rotation. For the truth $\mathbf{A}_i^*$ and its decomposition components $\mathbf{C}_i^*$, $\mathbf{R}_i^*$, and $\mathbf{P}_i^*$, we follow the identification constraint in Section 2.4 to require $\|\mathbf{A}_1^*\|_{\mathrm{F}} = \|\mathbf{A}_2^*\|_{\mathrm{F}} = \phi$, $[\mathbf{C}_i^* \ \mathbf{R}_i^*]^\top [\mathbf{C}_i^* \ \mathbf{R}_i^*] = \phi^{2/3}\mathbf{I}_{r_i}$ and $[\mathbf{C}_i^* \ \mathbf{P}_i^*]^\top [\mathbf{C}_i^* \ \mathbf{P}_i^*] = \phi^{2/3}\mathbf{I}_{r_i}$, for $i = 1, 2$. As defined in (2.9), $\mathrm{dist}(\mathbf{\Theta}^{(j)}, \mathbf{\Theta}^*)^2$ measures the estimation error of the parameters after j iterations, and

naturally, $\mathrm{dist}(\mathbf{\Theta}^{(0)}, \mathbf{\Theta}^*)^2$ represents the initialization error. Let $\underline{\sigma} := \min(\sigma_{1,r_1}, \sigma_{2,r_2})$ be the smallest singular value among all the non-zero singular values of $\mathbf{A}_1^*$ and $\mathbf{A}_2^*$. Define $\kappa := \phi/\underline{\sigma}$, which quantifies the ratio between the overall signal and the minimal signal. Then, we have the following sufficient conditions for local convergence of the algorithm.

Theorem 1 (Local Linear Convergence)**.** *Suppose that the RSC/RSS condition is satisfied with α and β. If $\lambda_1 \asymp \alpha$, $\lambda_2 \asymp \kappa^{-2}\phi^{8/3}\alpha$, $b \asymp \phi^{1/3}$, and $\eta = \eta_0\phi^{-10/3}\beta^{-1}$ with η_0 being a sufficiently small positive constant, $\xi^2 \lesssim \kappa^{-6}\phi^4\alpha^3\beta^{-1}$, and the initialization error $\mathrm{dist}(\mathbf{\Theta}^{(0)}, \mathbf{\Theta}^*)^2 \le c_0\kappa^{-2}\phi^{2/3}\alpha\beta^{-1}$, where c_0 is a small constant, then for all $j \ge 1$,*

$$\mathrm{dist}(\mathbf{\Theta}^{(j)}, \mathbf{\Theta}^*)^2 \le (1 - C\eta_0\alpha\beta^{-1}\kappa^{-2})^j \mathrm{dist}(\mathbf{\Theta}^{(0)}, \mathbf{\Theta}^*)^2 + C\kappa^4\phi^{-10/3}\alpha^{-2}\xi^2(r_1, r_2, d_1, d_2),$$

$$\left\|\mathbf{A}_1^{(j)} - \mathbf{A}_1^*\right\|_{\mathrm{F}}^2 + \left\|\mathbf{A}_2^{(j)} - \mathbf{A}_2^*\right\|_{\mathrm{F}}^2 \lesssim \kappa^2(1 - C\eta_0\alpha\beta^{-1}\kappa^{-2})^j \left[\left\|\mathbf{A}_1^{(0)} - \mathbf{A}_1^*\right\|_{\mathrm{F}}^2 + \left\|\mathbf{A}_2^{(0)} - \mathbf{A}_2^*\right\|_{\mathrm{F}}^2\right] + \kappa^4\phi^{-2}\alpha^{-2}\xi^2(r_1, r_2, d_1, d_2),$$

and

$$\left\|\mathbf{A}_2^{(j)} \otimes \mathbf{A}_1^{(j)} - \mathbf{A}_2^* \otimes \mathbf{A}_1^*\right\|_{\mathrm{F}}^2 \lesssim \kappa^2(1 - C\eta_0\alpha\beta^{-1}\kappa^{-2})^j \left\|\mathbf{A}_2^{(0)} \otimes \mathbf{A}_1^{(0)} - \mathbf{A}_2^* \otimes \mathbf{A}_1^*\right\|_{\mathrm{F}}^2 + \kappa^4\alpha^{-2}\xi^2(r_1, r_2, d_1, d_2).$$

Theorem 1 confirms that, provided a good initialization, Algorithm 1 converges linearly to a neighborhood of the truth defined by the statistical noise level ξ. Importantly, the conditions on the hyperparameters do not explicitly depend on the ambient dimensions p_1, p_2 or sample size T, but rather on the signal structure (α, β, κ), highlighting the scalability of the method. The RSC/RSS and initialization conditions are justified in the next subsection.

4.2 Statistical Convergence Rates

We now establish the statistical guarantees by verifying the deterministic conditions of Theorem 1 under a probabilistic framework. We impose the following standard assumptions.

Assumption 1 (Stationarity). *The spectral radius of* $\mathbf{A}^* = \mathbf{A}_2^* \otimes \mathbf{A}_1^*$ *is strictly less than one.*

Assumption 2 (Sub-Gaussian Noise). *The white noise* $\mathbf{E}_t$ *can be represented as* $\mathrm{vec}(\mathbf{E}_t) = \boldsymbol{\Sigma}_{\mathbf{e}}^{1/2}\boldsymbol{\zeta}_t$, *where* $\boldsymbol{\Sigma}_{\mathbf{e}}$ *is a positive definite matrix,* $\{\boldsymbol{\zeta}_t\}$ *are independent and identically distributed random variables with* $\mathbb{E}[\boldsymbol{\zeta}_t] = 0$ *and* $\mathrm{Cov}(\boldsymbol{\zeta}_t) = \mathbf{I}_{p_1p_2}$. *The entries of* $\boldsymbol{\zeta}_t$, *denoted by* $\{\zeta_{it}\}_{i=1}^{p_1p_2}$, *are independent* τ^2*-sub-Gaussian, i.e.,* $\mathbb{E}\left[\exp\left(\mu\zeta_{it}\right)\right] \leq \exp\left(\tau^2\mu^2/2\right)$ *for any* $\mu \in \mathbb{R}$ *and* $1 \leq i \leq p_1p_2$.

Assumption 1 ensures the existence of a unique strictly stationary solution to the proposed model. The sub-Gaussianity condition in Assumption 2 is standard in high-dimensional time series literature, such as Zheng and Cheng (2021) and Wang et al. (2024).

To ensure the identifiability of the common versus specific subspaces, we require a minimal separation between the response-specific and predictor-specific subspaces. We quantify this separation by the $\sin\theta$ distance. For two subspaces with orthonormal bases $\mathbf{R}_i$ and $\mathbf{P}_i$, let $s_{i,1} \geq \cdots \geq s_{i,r_i-d_i} \geq 0$ be the singular values of $\mathbf{R}_i^\top\mathbf{P}_i$. Then, the canonical angles are defined as $\theta_k(\mathbf{R}_i, \mathbf{P}_i) = \arccos\left(s_{i,k}\right)$ for $1 \leq k \leq r_i - d_i$.

Assumption 3 (Minimal Gap between Specific Subspaces). *There exists a constant* $g_{\min} > 0$ *such that* $\min_{i=1,2} \sin\theta_1\left(\mathbf{R}_i^*, \mathbf{P}_i^*\right) \geq g_{\min}$.

We quantify the dependency structure of the time series following the spectral approach as in Basu and Michailidis (2015). The characteristic matrix polynomial of the MARCF model is given by $\mathcal{A}(z) = \mathbf{I} - \mathbf{A}^*z$ for $z \in \mathbb{C}$. Define $\mu_{\min}(\mathcal{A}) = \min_{|z|=1} \lambda_{\min}\left(\mathcal{A}^\dagger(z)\mathcal{A}(z)\right)$ and $\mu_{\max}(\mathcal{A}) = \max_{|z|=1} \lambda_{\max}\left(\mathcal{A}^\dagger(z)\mathcal{A}(z)\right)$, where $\mathcal{A}^\dagger(z)$ is the conjugate transpose of $\mathcal{A}(z)$.

We define the key theoretical quantities:

$$M_1 = \frac{\lambda_{\max}(\mathbf{\Sigma_e})}{\mu_{\min}^{1/2}(\mathcal{A})},\ M_2 = \frac{\lambda_{\min}(\mathbf{\Sigma_e})\,\mu_{\min}(\mathcal{A})}{\lambda_{\max}(\mathbf{\Sigma_e})\,\mu_{\max}(\mathcal{A})},\ \alpha_{\mathrm{RSC}} = \frac{\lambda_{\min}(\mathbf{\Sigma_e})}{2\mu_{\max}(\mathcal{A})},\ \text{and } \beta_{\mathrm{RSS}} = \frac{3\lambda_{\max}(\mathbf{\Sigma_e})}{2\mu_{\min}(\mathcal{A})}.$$

The following proposition verifies the RSC/RSS condition required for the local convergence theorem.

Proposition 1 (Verification of RSC/RSS). *Under Assumptions 1 and 2, if the sample size* $T \gtrsim M_2^{-2}\max(\tau^2,\tau^4)(p_1r_1+p_2r_2+4r_1^2r_2^2)$, *with probability at least* $1-C\exp(-C(p_1r_1+p_2r_2))$, *the RSC/RSS condition in Definition 1 holds with* $\alpha=\alpha_{\mathrm{RSC}}$ *and* $\beta=\beta_{\mathrm{RSS}}$.

It remains to verify the initialization condition in Theorem 1, which is proved to hold with high probability by the following proposition.

Proposition 2 (Initialization Guarantee). *Under Assumptions 1, 2, and 3, suppose that* $T \gtrsim \{M_2^{-2}\max(\tau^2,\tau^4)+\phi^{-4}\alpha_{\mathrm{RSC}}^{-2}\tau^2M_1^2r_1r_2\}p_1p_2$ *and choose* $\lambda_0 \asymp \tau M_1\sqrt{p_1p_2/T}$. *Then, the initial value* $\mathbf{\Theta}^{(0)}$ *in Section 3.2 satisfies, with probability at least* $1-C\exp(-Cp_1p_2)$,

$$\mathrm{dist}(\mathbf{\Theta}^{(0)},\mathbf{\Theta}^*)^2 \lesssim \phi^{2/3}\underline{\sigma}^{-4}g_{\min}^{-4}\alpha_{\mathrm{RSC}}^{-2}\tau^2M_1^2\frac{p_1p_2r_1r_2}{T}.$$

Consequently, if $T \gtrsim \kappa^2\underline{\sigma}^{-4}g_{\min}^{-4}\alpha_{\mathrm{RSC}}^{-3}\beta_{\mathrm{RSS}}\tau^2M_1^2p_1p_2r_1r_2$, *then the local initialization condition in Theorem 1 holds with the same probability, namely,* $\mathrm{dist}(\mathbf{\Theta}^{(0)},\mathbf{\Theta}^*)^2 \lesssim \kappa^{-2}\phi^{2/3}\alpha_{\mathrm{RSC}}\beta_{\mathrm{RSS}}^{-1}$.

Combining the preceding two propositions with the deviation bound ξ and the deterministic contraction in Theorem 1, we obtain the following guarantee for the complete two-step estimation procedure.

Theorem 2 (Statistical Guarantees of Two-Step Procedure). *Let* $\mathbf{\Theta}^{(0)}$ *be constructed as in Section 3.2 with* $\lambda_0 \asymp \tau M_1\sqrt{p_1p_2/T}$, *and let* $\mathbf{A}_1^{(J)}$ *and* $\mathbf{A}_2^{(J)}$ *be the output of Algorithm 1 after*

J iterations. Assume that Assumptions 1, 2, and 3 hold. Suppose that the hyperparameters λ_1, λ_2, b, and η are set as in Theorem 1 with $\alpha = \alpha_{\mathrm{RSC}}$ and $\beta = \beta_{\mathrm{RSS}}$. Suppose further that $T \gtrsim M_2^{-2} \max(\tau^2, \tau^4)(p_1p_2 + p_1r_1 + p_2r_2 + 4r_1^2r_2^2)$ and $T \gtrsim \kappa^2 \underline{\sigma}^{-4} g_{\min}^{-4} \alpha_{\mathrm{RSC}}^{-3} \beta_{\mathrm{RSS}} \tau^2 M_1^2 p_1 p_2 r_1 r_2$. If the number of iterations satisfies

$$J \gtrsim \frac{\log\left(g_{\min}^4 \kappa^{-2} \mathrm{df}_{\mathrm{MARCF}} / (p_1 p_2 r_1 r_2)\right)}{\log(1 - C\eta_0 \alpha_{\mathrm{RSC}} \beta_{\mathrm{RSS}}^{-1} \kappa^{-2})},$$

with probability at least $1 - C\exp\left(-C(p_1r_1 + p_2r_2)\right) - C\exp(-Cp_1p_2)$, the estimation error of the two-step estimation procedure satisfies

$$\left\|\mathbf{A}_1^{(J)} - \mathbf{A}_1^*\right\|_{\mathrm{F}}^2 + \left\|\mathbf{A}_2^{(J)} - \mathbf{A}_2^*\right\|_{\mathrm{F}}^2 \lesssim \kappa^4 \phi^{-2} \alpha_{\mathrm{RSC}}^{-2} \tau^2 M_1^2 \frac{\mathrm{df}_{\mathrm{MARCF}}}{T}$$

and

$$\left\|\mathbf{A}_2^{(J)} \otimes \mathbf{A}_1^{(J)} - \mathbf{A}_2^* \otimes \mathbf{A}_1^*\right\|_{\mathrm{F}}^2 \lesssim \kappa^4 \alpha_{\mathrm{RSC}}^{-2} \tau^2 M_1^2 \frac{\mathrm{df}_{\mathrm{MARCF}}}{T},$$

where $\mathrm{df}_{\mathrm{MARCF}}$ is the effective number of free parameters defined in (2.7).

Theorem 2 establishes that the full two-step estimator achieves the squared error rate $O_p(\mathrm{df}_{\mathrm{MARCF}}/T)$. Since $\mathrm{df}_{\mathrm{MARCF}} = \mathrm{df}_{\mathrm{RRMAR}} - (p_1d_1 + p_2d_2)$, the MARCF model offers a demonstrable variance reduction over the standard RRMAR model by exploiting the common subspace structure. In addition, the required number of iterations J grows only logarithmically with the problem dimensions p_1 and p_2, indicating the scalability of the method in high-dimensional large-scale data settings.

Remark 3 (Computational-Statistical Gap)**.** Proposition 2 requires T to scale with p_1p_2 to place the initial value in the local basin of Theorem 1, whereas the final statistical error floor in Theorem 2 is of order $p_1r_1 + p_2r_2$, revealing a gap between computational and statistical convergence. Essentially, the initialization problem amounts to initializing a VAR transition

matrix with a low-rank Kronecker product structure, and such a Kronecker product can be viewed as a fourth order low-Tucker-rank tensor (see Appendix S4.5). Existing theory on low rank tensor regression (Han et al., 2022) shows a similar gap. In view of this analogy, we suspect the gap may also be intrinsic here, and that developing a generally computationally feasible initializer with a weaker sample size requirement appears challenging.

4.3 Consistency of Rank Selection

In this subsection, we establish the consistency of rank selection under the standard high-dimensional asymptotic regime (see, e.g., Lam et al., 2011; Lam and Yao, 2012), where T, p_1, and $p_2 \to \infty$ with r_1 and r_2 remaining fixed.

Theorem 3 (Rank Selection Consistency)**.** *Under Assumptions 1 and 2, suppose that* $\bar{r}_1 > r_1, \bar{r}_2 > r_2$, $T \gtrsim \{M_2^{-2}\max(\tau^2,\tau^4) + \phi^{-4}\alpha_{\mathrm{RSC}}^{-2}\tau^2 M_1^2 r_1 r_2\}p_1p_2$, $\lambda_0 \asymp \tau M_1\sqrt{p_1p_2/T}$, $\phi^{-1}\alpha_{\mathrm{RSC}}^{-1}\tau M_1\sqrt{p_1p_2r_1r_2/T} = o(s(p_1,p_2,T))$, *and* $s(p_1,p_2,T) = o(\sigma_{i,r_i}\min_{1\le j\le r_i-1}\sigma_{i,j+1}/\sigma_{i,j})$ *for* $i = 1, 2$, *with the convention that the minimum over an empty set is one when* $r_i = 1$. *Then, we have that* $\mathbb{P}(\widehat{r}_1 = r_1, \widehat{r}_2 = r_2) \to 1$, *as* $T \to \infty$ *and* $p_1, p_2 \to \infty$.

This theorem provides sufficient conditions under which the selected ranks $\widehat{r}_1$ and $\widehat{r}_2$ converge in probability to the true ranks r_1 and r_2. In practice, if the parameters M_1, M_2, α_{RSC}, σ_{1,r_1}, and σ_{2,r_2} are bounded, and ϕ is either bounded or diverges slowly as $p_1, p_2 \to \infty$, the conditions simplify to requiring that $s(p_1,p_2,T) \to 0$ and $\sqrt{p_1p_2r_1r_2/T}/s(p_1,p_2,T) \to 0$. These simplified conditions are more tractable in applications to help ensure the reliability of the rank selection. In particular, the value of $s(p_1,p_2,T)$ suggested in Section 3.3 satisfies these requirements, ensuring consistency across a wide range of settings for p_1, p_2, and T.

For the proposed BIC procedure for selecting (d_1, d_2), the finite-sample results in Table

1 demonstrate excellent selection accuracy across the settings considered. The theoretical consistency of this procedure has not been established in this paper, which we acknowledge as a limitation of the present work. Establishing this consistency is left for future research.

5. Simulation Studies

In this section, we evaluate the finite-sample performance of the proposed MARCF model through a series of simulation experiments. The objectives are three-fold: (i) to assess the accuracy of parameter estimation and model selection, (ii) to validate the model's ability to identify structured factor-driven dynamics, and (iii) to compare its performance with existing approaches.

5.1 Experiment I: Estimation Accuracy and Model Selection

In the first experiment, we generate data from the MARCF model defined in (2.6). We first assess the reliability of the proposed procedures for selecting the model ranks and common dimensions. Then, we compare the estimation errors of the MARCF model with two competing approaches: the full-rank MAR model (Chen et al., 2021) and the RRMAR model (Xiao et al., 2023).

We fix the model ranks at $r_1 = 3$ and $r_2 = 2$, and consider all combinations where $d_1 \in \{0, 1, 2, 3\}$ and $d_2 \in \{0, 1, 2\}$. Two regimes are examined: $(p_1, p_2, T) = (20, 10, 500)$ and $(p_1, p_2, T) = (30, 20, 1000)$. For each configuration, the coefficient matrices are generated according to the MARCF decomposition. Specifically, for $i = 1, 2$, $\mathbf{C}_i$ is drawn as a random matrix with orthonormal columns, while $\mathbf{R}_i$ and $\mathbf{P}_i$ are obtained by QR orthonormalization after projecting independent Gaussian random matrices onto the orthogonal complement of

$\mathbf{C}_i$. Candidate draws with $\sin\theta_1 < 0.8$ are discarded. The transition matrix is generated as $\mathbf{D}_i = \mathbf{O}_{i,1}^\top \mathbf{S}_i \mathbf{O}_{i,2}$, with random orthogonal $\mathbf{O}_{i,1}$ and $\mathbf{O}_{i,2}$ and diagonal $\mathbf{S}_i$ whose entries are sampled from $U(0.8, 1)$ when $d_i = r_i$ and from $U(0.9, 1.1)$ otherwise. We resample until $\rho(\mathbf{A}_1) \cdot \rho(\mathbf{A}_2) < 1$, generate $\mathrm{vec}(\mathbf{E}_t) \overset{i.i.d.}{\sim} N(\mathbf{0}, \mathbf{I})$, and simulate $\{\mathbf{Y}_t\}_{t=0}^T$ from (2.6).

For each pair of (d_1, d_2), the MARCF model is estimated using the full procedure proposed in Section 3, including rank selection, common dimension selection, initialization, and final gradient descent estimation (see Appendix S1 of the supplementary material for details). In the initialization step, λ_0 is set by a fixed scaling rule throughout the simulations. For rank selection, we set $\bar{r}_1 = \bar{r}_2 = 8$ with $s(p_1, p_2, T) = \sqrt{p_1 p_2 \log(T)/(200T)}$. The RRMAR model is estimated by both RR.LS and RR.CC methods in Xiao et al. (2023) with selected ranks. Based on preliminary experiments, we observed that the model performance is insensitive to variations in λ_1, λ_2, and b. Accordingly, we set these parameters to 1. We set $\eta = 0.001$ and run Algorithm 1 until convergence or for a maximum of 1000 iterations. No divergence was observed under these hyperparameter settings.

The left panel of Table 1 reports the selection accuracy for (r_1, d_1, r_2, d_2) across the 24 simulation configurations. A replication is counted as successful only when all four parameters are correctly selected. The proposed procedure achieves near-perfect selection accuracy: the accuracy is 100% in 22 cases and 99.8% in the remaining two, corresponding to only one incorrect selection out of 500 replications in each case. These results indicate that the proposed selection procedure is reliable across different common dimension structures.

Figure 1 reports the relative estimation errors of the competing methods, defined as $\|\widehat{\mathbf{A}}_2 \otimes \widehat{\mathbf{A}}_1 - \mathbf{A}_2^* \otimes \mathbf{A}_1^*\|_{\mathrm{F}} / \|\mathbf{A}_2^* \otimes \mathbf{A}_1^*\|_{\mathrm{F}}$. For each method, the error bar represents the interquartile range of the 500 replications, and the cross marker denotes the median. Across

all configurations, the full MAR estimator has substantially larger errors, reflecting the cost of ignoring the low-rank structure. The initializer also has larger errors than the refined low-rank estimators, which is expected since it is not optimized under the final MARCF objective. The two RRMAR estimators, RR.LS and RR.CC, exhibit nearly indistinguishable performance throughout the experiment. When $d_1 = d_2 = 0$, the MARCF model reduces to the standard RRMAR model. Accordingly, MARCF, RR.LS, and RR.CC have comparable estimation errors. However, as the common dimensions increase, the error distribution of MARCF shifts downward and separates from those of RR.LS and RR.CC. The improvement is most pronounced when $d_1 = r_1$ and $d_2 = r_2$, where MARCF attains the smallest estimation error. This trend empirically validates the efficiency gains obtained by explicitly modeling the overlap between the predictor and response subspaces.

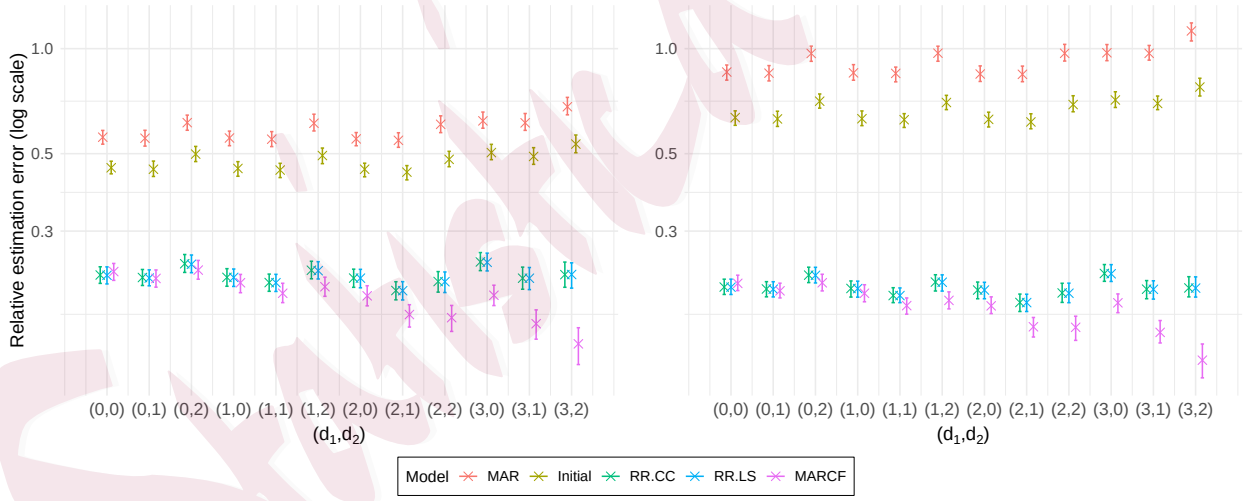

Figure 1: Relative estimation error based on 500 replications. Left panel: $(p_1, p_2, T) = (20, 10, 500)$; right panel: $(p_1, p_2, T) = (30, 20, 1000)$.

5.2 Experiment II: Identification of Factor-Driven Dynamics

In the second experiment, we examine whether the MARCF model can correctly identify a pure factor-driven structure when the data are generated from the dynamic MFM defined in (2.1) and (2.4). The goal is to evaluate the proportion of trials in which the proposed modeling procedure correctly identifies the dynamic MFM structure, and to compare the estimation errors of the factor loading projection matrices under the MARCF model and the MFM model of Wang et al. (2019).

We set $(p_1, p_2, T) = (16, 16, 800)$ and consider cases where $r_1 = d_1 \in \{1, 2, 3, 4\}$ and $r_2 = d_2 \in \{1, 2, 3, 4\}$. The data are generated as follows:

1. For $i = 1, 2$, generate the loading matrix $\mathbf{\Lambda}_i \in \mathbb{R}^{p_i \times r_i}$ with random orthonormal columns. Generate the factor transition matrix $\mathbf{B}_i \in \mathbb{R}^{r_i \times r_i}$ in the same way as $\mathbf{D}_i$ in Experiment I, and resample until $\rho(\mathbf{B}_1) \cdot \rho(\mathbf{B}_2) < 1$.

2. Generate two independent white noise processes $\mathbf{E}_t \in \mathbb{R}^{p_1 \times p_2}$ and $\boldsymbol{\xi}_t \in \mathbb{R}^{r_1 \times r_2}$ with $\mathrm{vec}(\mathbf{E}_t) \overset{\text{i.i.d.}}{\sim} N_{p_1 p_2}(\mathbf{0}, \mathbf{I})$ and $\mathrm{vec}(\boldsymbol{\xi}_t) \overset{\text{i.i.d.}}{\sim} N_{r_1 r_2}(\mathbf{0}, \mathbf{I})$. Construct $\mathbf{W}_t = \mathbf{E}_t - \mathbf{\Lambda}_1 \mathbf{\Lambda}_1^\top \mathbf{E}_t \mathbf{\Lambda}_2 \mathbf{\Lambda}_2^\top$ to eliminate the moving average term in the composite noise of (2.5).

3. Generate the factors as $\mathbf{F}_t = \mathbf{B}_1 \mathbf{F}_{t-1} \mathbf{B}_2^\top + \boldsymbol{\xi}_t$ and the observation as $\mathbf{Y}_t = \mathbf{\Lambda}_1 \mathbf{F}_t \mathbf{\Lambda}_2^\top + \mathbf{W}_t$.

The training procedure and tuning parameters are the same as in Experiment I. For the MFM benchmark, we use the TIPUP method from Chen et al. (2022) with the true ranks.

The right panel of Table 1 reports the percentage of replications in which the MFM pattern is correctly identified, i.e., $\widehat{r}_1 = \widehat{d}_1 = r_1$ and $\widehat{r}_2 = \widehat{d}_2 = r_2$. The proposed procedure also achieves near-perfect selection accuracy on this MFM-generated data. It correctly identifies the full structure in all 500 replications for 15 of the 16 cases, and attains a success

Table 1: Selection accuracy (%) over 500 replications. The left and right column blocks of MARCF data correspond to $(p_1, p_2, T) = (20, 10, 500)$ and $(30, 20, 1000)$, respectively.

(a) MARCF-generated data

	$d_2 = 0$	$d_2 = 1$	$d_2 = 2$	$d_2 = 0$	$d_2 = 1$	$d_2 = 2$
$d_1 = 0$	100.0	100.0	100.0	100.0	100.0	100.0
$d_1 = 1$	100.0	100.0	100.0	100.0	100.0	100.0
$d_1 = 2$	100.0	99.8	100.0	100.0	100.0	100.0
$d_1 = 3$	100.0	99.8	100.0	100.0	100.0	100.0

(b) MFM-generated data

	$r_2 = 1$	$r_2 = 2$	$r_2 = 3$	$r_2 = 4$
$r_1 = 1$	97.6	100.0	100.0	100.0
$r_1 = 2$	100.0	100.0	100.0	100.0
$r_1 = 3$	100.0	100.0	100.0	100.0
$r_1 = 4$	100.0	100.0	100.0	100.0

rate of 97.6% in the remaining case $(r_1, r_2) = (1, 1)$. These results show that the MARCF selection procedure can reliably recognize the fully factor-driven structure when the data are generated from an MFM.

Figure 2 presents the log estimation errors of the MARCF model and MFM, where each pair of (r_1, r_2) is evaluated over 500 replications. Only trials with correctly identified parameters are included. The error metric is defined as $\log(\|\widehat{\mathbf{\Lambda}}_i(\widehat{\mathbf{\Lambda}}_i^\top\widehat{\mathbf{\Lambda}}_i)^{-1}\widehat{\mathbf{\Lambda}}_i^\top - \mathbf{\Lambda}_i^*\mathbf{\Lambda}_i^{*\top}\|_{\mathrm{F}})$ for $i = 1, 2$, which corresponds to the estimation errors for the column spaces $\mathcal{M}(\mathbf{\Lambda}_1^*)$ and $\mathcal{M}(\mathbf{\Lambda}_2^*)$. For $\mathcal{M}(\mathbf{\Lambda}_1^*)$ (the left panel), the MARCF model shows slightly smaller errors when $r_2 = 1$, and significantly smaller errors when $r_2 \geq 2$, especially for $r_2 \geq 3$. Similar trends hold for $\mathcal{M}(\mathbf{\Lambda}_2^*)$. Overall, MARCF performs comparably to or significantly better than MFM, demonstrating its effectiveness and flexibility.

6. Real Example: Quarterly Macroeconomic Data

To illustrate the empirical utility of the MARCF framework, we apply it to a multinational macroeconomic dataset obtained from the Organisation for Economic Co-operation and De-

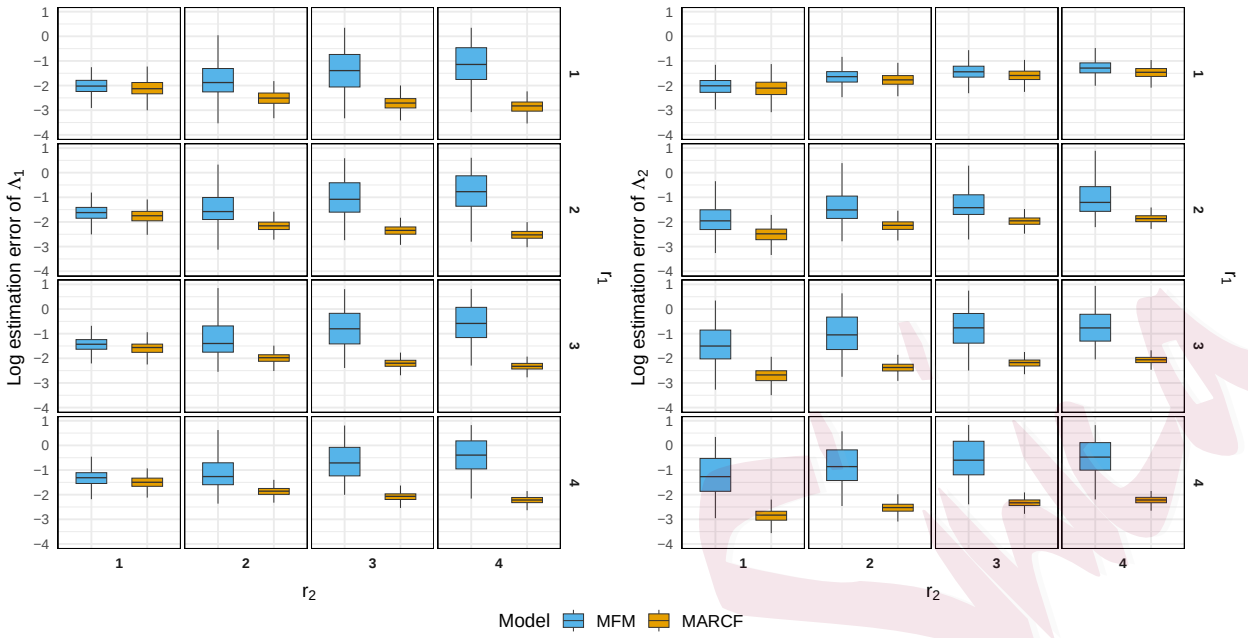

Figure 2: Log estimation error of $\mathcal{M}(\mathbf{\Lambda}_1^*)$ (left panel) and $\mathcal{M}(\mathbf{\Lambda}_2^*)$ (right panel).

velopment (OECD; https://www.oecd.org/). The dataset contains 9 key economic indicators for 12 countries observed over 92 quarters, from 1997-Q1 to 2019-Q4. The processed data and code are available at https://github.com/zhiyun-fan/MARCF for reproducibility.

The economic indicators are Gross Domestic Product (GDP), final consumption expenditure of households and non-profit institutions serving households (HCONS), exports of goods and services (EXP), imports of goods and services (IMP), total industrial production excluding construction (PTEC), total manufacturing production (PTM), three-month interbank interest rates (IR3), unemployment (UNEMP), and share prices (SHARE). The countries are the United States (USA), Canada (CAN), the United Kingdom (GBR), Australia (AUS), Germany (DEU), France (FRA), the Netherlands (NLD), New Zealand (NZL), Austria (AUT), Sweden (SWE), Finland (FIN), and Poland (POL). Prior to analysis, the series were seasonally adjusted, transformed to promote stationarity (using logarithmic differences or first differences where appropriate), and standardized to zero mean and unit

variance.

6.1 Model Selection and Estimation

The MARCF model structure is selected by BIC, as discussed in Section 3.3. We first fit RRMAR models using RR.LS, for which BIC selected total ranks $r_1 = r_2 = 3$. Conditional on these ranks, the selected common subspace dimensions are $d_1 = d_2 = 2$. The presence of both common components ($d_i > 0$) and specific components ($r_i > d_i$) suggests a hybrid subspace structure in the data. For interpretation, the estimated basis matrices $[\widehat{\mathbf{C}}_i\ \widehat{\mathbf{R}}_i]$ and $[\widehat{\mathbf{C}}_i\ \widehat{\mathbf{P}}_i]$ are first orthonormalized and then rotated using varimax.

6.2 Interpretation of Global Economic Dynamics

We dissect the estimated dynamics by analyzing the basis matrices $\widehat{\mathbf{C}}_i$, $\widehat{\mathbf{R}}_i$, and $\widehat{\mathbf{P}}_i$ associated with the row (country) and column (indicator) subspaces.

Country-Level Dynamics (Row Space). Figure 3 reports the entries of the varimax-rotated basis matrices on the country side.

- Common Directions: The first common direction, C_1–1, is broadly and positively supported by a group of major and open advanced economies, including USA, CAN, GBR, DEU, FRA, SWE, and FIN. The corresponding common factor can be interpreted as a persistent global macroeconomic state, which plays an important role both in predicting the global trends and being predicted. The second common direction, C_1–2, is dominated by POL, suggesting a distinct country-side pattern alongside C_1–1. The strong loading of POL (0.92) suggests that the POL-related information is mostly aligned between the predictor and response spaces, rather than being side-specific.

- Specific Directions: The predictor-specific direction, P_1–1, captures predictive information beyond the common directions. Its loading pattern differs from the first common direction: unlike C_1–1, which loads the major economies in the same direction, P_1–1 contrasts FIN, NLD, and DEU with GBR, AUS, USA, and AUT. The response-specific direction, R_1–1, is dominated by NZL, whose large loading (0.71) suggests a more reactive role. In addition, the angle between R_1–1 and P_1–1 is 59.63°, indicating clear separation between the response- and predictor-specific country-side subspaces and supporting the need to model them separately.

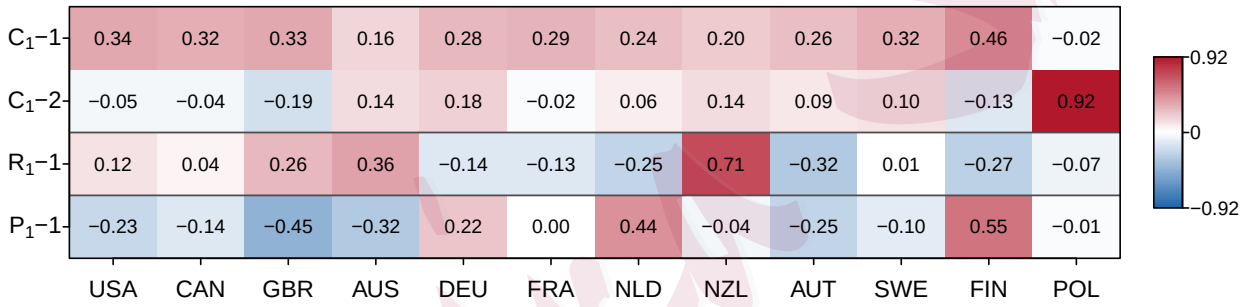

Figure 3: Estimates of common and specific basis matrices over countries.

Indicator-Level Dynamics (Column Space). Figure 4 reports the entries of the varimax-rotated basis matrices on the indicator side.

- Common Directions: C_2–1 summarizes a broad real-activity block covering GDP, consumption, external trade, and production. UNEMP loads on the opposite side, consistent with its countercyclical relation to real economic activity. We therefore interpret this direction as an aggregate business-cycle state shared across indicators. The second direction, C_2–2, is dominated by IR3 and is interpreted as a monetary-condition

component. Together, these two common directions capture the shared real-activity and monetary-condition dynamics on both sides of the autoregressive relation.

- Specific Directions: P_2–1 loads mainly on SHARE and GDP, moderately on HCONS and UNEMP, and oppositely on IMP and PTM. We interpret this direction as a forward-looking signal, reflecting financial-market and aggregate-demand information beyond the common real-activity component. R_2–1 is dominated by SHARE and UNEMP, suggesting an additional financial- and labor-market adjustment to global economic changes. Note that SHARE has large entries in both $\widehat{\mathbf{R}}_2$ and $\widehat{\mathbf{P}}_2$, but small entries in $\widehat{\mathbf{C}}_2$. This suggests that stock-market information is not simply summarized by a single shared indicator direction, but enters the predictor and response spaces in different side-specific ways. This is reflected by its co-loading patterns: SHARE is aligned with IMP and PTM in $\widehat{\mathbf{R}}_2$, but opposed to them in $\widehat{\mathbf{P}}_2$. Moreover, R_2–1 and P_2–1 form an angle of 50.37°, showing clear separation on the indicator side as well.

Figure 4: Estimates of basis matrices over economic indicators.

Interaction Dynamics. To illustrate the cross terms, we focus on the estimated C–P factor block $\widehat{\mathbf{F}}_{t-1}^{CP} := \widehat{\mathbf{C}}_1^\top \mathbf{Y}_{t-1} \widehat{\mathbf{P}}_2 = (f_1, f_2)^\top \in \mathbb{R}^{2\times 1}$, which represents the interaction between

common-country signals and predictor-specific indicator signals. Its fitted transmission to the future **C**–**C** response block is the upper-left 2×2 block of

$$\widehat{\mathbf{Z}}_t := \widehat{\mathbf{D}}_1 \begin{pmatrix} \mathbf{0} & \widehat{\mathbf{F}}^{CP}_{t-1} \\ \mathbf{0} & 0 \end{pmatrix} \widehat{\mathbf{D}}_2^{\top}.$$

We report the fitted transmission coefficients below.

$$\widehat{\mathbf{Z}}_t^{CP \to CC} = \begin{pmatrix} 0.78 f_1 + 0.24 f_2 & 0.40 f_1 + 0.12 f_2 \\ 0.17 f_1 + 0.38 f_2 & 0.09 f_1 + 0.19 f_2 \end{pmatrix}.$$

The largest coefficient appears in the upper-left entry, indicating that predictor-specific financial and aggregate-demand signals observed along the first common-country direction are most strongly mapped to the main common real-activity response coordinate. The two scalar coordinates play different roles: f_1 mainly contributes to the first common-country response row, while f_2 contributes relatively more to the second row. Most importantly, the predictor and response differ on the indicator side: the predictive signal comes from the predictor-specific indicator subspace, but is transmitted to the common response subspace. This provides a concrete example of the asymmetric transformation.

6.3 Forecasting Performance

Finally, we evaluate the out-of-sample forecasting performance. We compute one-step-ahead forecast errors using 16 rolling windows covering the period from 2016-Q1 to 2019-Q4. We compare MARCF against various baselines: (1) entrywise univariate AR(1) and ARMA(1,1) models (iAR and iARMA, respectively); (2) penalized vectorized VAR models, including ℓ_1 penalty (SparseVAR) and ℓ_2 penalty (RidgeVAR); (3) RRMAR model and dynamic MFM with the same ranks as MARCF. The RRMAR model is estimated by both RR.LS and

RR.CC, and the dynamic MFM follows Yu et al. (2024), which fits a MAR(1) model for the latent factor process.

Table 2 reports the mean and median sums of squared one-step-ahead out-of-sample prediction errors. MARCF delivers the best forecasting performance among all competing methods. Its improvement over the entrywise univariate models and the vectorized VAR benchmarks highlights the benefit of preserving the matrix structure. Moreover, MARCF improves upon RRMAR and dynamic MFM, suggesting that the predictive dynamics in this macroeconomic series are not fully captured by either a low-rank matrix autoregression without explicit common subspace structure or a pure factor-driven representation. Instead, the data appear to exhibit partially shared predictor and response subspaces, together with distinct subspace-specific dynamics. This evidence supports the empirical relevance of the MARCF framework, which is designed to capture such intermediate structures.

Table 2: Means and medians of the out-of-sample forecast SSE across 16 rolling windows.

	MARCF	iAR	iARMA	SparseVAR	RidgeVAR	RR.LS	RR.CC	MFM
Mean	**36.06**	37.17	37.71	38.52	38.11	36.50	37.44	37.79
Median	**33.79**	37.80	38.39	34.47	38.40	35.16	34.60	34.23

7. Conclusion

In this paper, we have introduced the Matrix Autoregressive model with Common Factors (MARCF), a hybrid structured RRMAR formulation for high-dimensional matrix time series that exploits the partial overlap between the response and predictor spaces. By explicitly parameterizing the intersection, MARCF addresses the tension between RRMAR, which

does not explicitly exploit subspace overlap, and dynamic MFM, which imposes fully aligned predictor-response subspaces. Our contributions can be summarized as follows:

- **Methodological Innovation:** We proposed a flexible decomposition that separates dynamics into common (shared), predictor-specific, and response-specific components. This allows the model to adaptively capture the "subspace alignment" of the data, providing a balance between the flexible subspace geometry of RRMAR and parsimony inspired by MFMs, while enabling the modeling of complex interaction dynamics.

- **Scalable Estimation:** We developed a regularized gradient descent algorithm together with a feasible initialization procedure. By incorporating balanced regularization terms ($b \asymp \phi^{1/3}$), our approach resolves the scaling indeterminacies inherent in bilinear systems and avoids computationally expensive matrix operations associated with alternating least squares, ensuring scalability to large-scale datasets.

- **Theoretical Rigor:** We established theoretical foundations for the proposed framework, including rank selection consistency, initialization, and convergence analysis for the full two-step estimation procedure. Our analysis decouples the geometric landscape of the loss function from the stochastic properties of the data, proving the estimator achieves the optimal convergence rate governed by the effective degrees of freedom.

The empirical application to global macroeconomic indicators highlights the practical value of this framework. The MARCF model not only improves forecasting accuracy but also provides interpretable insights into the global economy.

Several promising directions remain for future research. First, while we focused on dense subspace structures, incorporating sparsity constraints on the basis matrices $(\mathbf{C}, \mathbf{R}, \mathbf{P})$ could

further enhance interpretability and efficiency, particularly when p_1 and p_2 are very large. Second, extending this framework to tensor-valued time series of order $K \geq 3$ would allow for the modeling of more complex structures, though this introduces additional algebraic challenges regarding tensor rank selection. Finally, developing inferential tools, such as confidence intervals for the estimated factors or impulse response functions, remains an important challenge due to the non-standard asymptotics of singular subspace identification.

Supplementary Material

The online supplementary material provides additional details on the estimation procedure and proofs of the theoretical results.

Acknowledgements

We thank the editor, the associate editor, and the referees for their constructive comments and helpful suggestions. Xiaoyu Zhang's research is supported by National Natural Science Foundation of China (Grant No. 12601528), National Key R&D Program of China (Grant No. 2024YFA1015700), and Fundamental Research Funds for the Central Universities (Grant No. 22120260347). Di Wang's research is supported by National Natural Science Foundation of China (Grant Nos. 12301352, 72495121, 72671185).

References

Bai, J., Li, K., and Lu, L. (2016). Estimation and inference of FAVAR models. *Journal of Business & Economic Statistics*, 34(4):620–641.

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

Bai, J. and Ng, S. (2008). Large dimensional factor analysis. *Foundations and Trends® in Econometrics*, 3(2):89–163.

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, 43(4):1535 – 1567.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly journal of economics*, 120(1):387–422.

Chang, J., He, J., Yang, L., and Yao, Q. (2023). Modelling matrix time series via a tensor CP-decomposition. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(1):127–148.

Chen, E. Y. and Fan, J. (2023). Statistical inference for high-dimensional matrix-variate factor models. *Journal of the American Statistical Association*, 118(542):1038–1055.

Chen, E. Y., Tsay, R. S., and Chen, R. (2020). Constrained factor models for high-dimensional matrix-variate time series. *Journal of the American Statistical Association*, 115(530):775–793.

Chen, R., Xiao, H., and Yang, D. (2021). Autoregressive models for matrix-valued time series. *Journal of Econometrics*, 222(1, Part B):539–560.

Chen, R., Yang, D., and Zhang, C.-H. (2022). Factor models for high-dimensional tensor time series. *Journal of the American Statistical Association*, 117(537):94–116.

Gao, Z. and Tsay, R. S. (2022). Modeling high-dimensional time series: A factor model with dynamically dependent factors and diverging eigenvalues. *Journal of the American Statistical Association*, 117(539):1398–1414.

Gao, Z. and Tsay, R. S. (2023). A two-way transformed factor model for matrix-variate time series. *Econometrics and Statistics*, 27:83–101.

Han, R., Willett, R., and Zhang, A. R. (2022). An optimal statistical and computational framework for generalized tensor estimation. *The Annals of Statistics*, 50(1):1 – 29.

Hsu, N.-J., Huang, H.-C., and Tsay, R. S. (2021). Matrix autoregressive spatio-temporal models. *Journal of Computational and Graphical Statistics*, 30(4):1143–1155.

Huang, F., Lu, K., Zheng, Y., and Li, G. (2025). Supervised factor modeling for high-dimensional linear time series. *Journal of Econometrics*, 249:105995.

Jiang, H., Shen, B., Li, Y., and Gao, Z. (2024). Regularized estimation of high-dimensional matrix-variate autoregressive models. *arXiv preprint arXiv:2410.11320*.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *The Annals of Statistics*, 40(2):694 – 726.

Lam, C., Yao, Q., and Bathia, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika*, 98(4):901–918.

Miao, K., Phillips, P. C. B., and Su, L. (2023). High-dimensional VARs with common factors. *Journal of Econometrics*, 233(1):155–183.

Qin, L., Wang, Y., Zhu, Y., and Shia, B.-C. (2025). Bayesian dynamic matrix factor models. *Journal of Business & Economic Statistics*, 43(4):1170–1182.

Samadi, S. Y. and De Alwis, T. P. (2026). Envelope matrix autoregressive models. *Journal of business & economic statistics*, 44(2):397–412.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Tu, S., Boczar, R., Simchowitz, M., Soltanolkotabi, M., and Recht, B. (2016). Low-rank solutions of linear matrix equations via Procrustes Flow. In *Proceedings of The 33rd International Conference on Machine Learning*, volume 48, pages 964–973.

Wang, D., Liu, X., and Chen, R. (2019). Factor models for matrix-valued high-dimensional time series. *Journal of Econometrics*, 208(1):231–248.

Wang, D., Zhang, X., Li, G., and Tsay, R. (2023). High-dimensional vector autoregression with common response and predictor factors. *arXiv preprint arXiv:2203.15170*.

Wang, D., Zheng, Y., and Li, G. (2024). High-dimensional low-rank tensor autoregressive time series modeling. *Journal of Econometrics*, 238(1):105544.

Wang, D., Zheng, Y., Lian, H., and Li, G. (2022). High-dimensional vector autoregressive time series modeling via tensor decomposition. *Journal of the American Statistical Association*, 117(539):1338–1356.

Wang, L., Zhang, X., and Gu, Q. (2017). A unified computational and statistical framework for nonconvex low-rank matrix estimation. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54, pages 981–990.

Xia, Q., Xu, W., and Zhu, L. (2015). Consistently determining the number of factors in multivariate volatility modelling. *Statistica Sinica*, 25(3):1025–1044.

Xiao, H., Han, Y., Chen, R., and Liu, C. (2023). Reduced-rank autoregressive models for matrix time series. *Journal of Business & Economic Statistics*. To appear.

Yu, R., Chen, R., Xiao, H., and Han, Y. (2024). Dynamic matrix factor models for high dimensional time series. *arXiv preprint arXiv:2407.05624*.

Zhang, H.-F. (2024). Additive autoregressive models for matrix valued time series. *Journal of Time Series Analysis*, 45(3):398–420.

Zheng, Y. and Cheng, G. (2021). Finite-time analysis of vector autoregressive models under linear restrictions. *Biometrika*, 108(2):469–489.

Zhiyun Fan

School of Mathematical Sciences, Shanghai Jiao Tong University, China

E-mail: fzy2023@sjtu.edu.cn

Xiaoyu Zhang

School of Mathematical Sciences and MOE Key Laboratory of Intelligent Computing and Applications, Tongji University, China

E-mail: xzhangck@tongji.edu.cn

Di Wang

School of Mathematical Sciences, Shanghai Jiao Tong University, China

E-mail: di.wang@sjtu.edu.cn