Journal of Econometrics 252 (2025) 106105

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Matrix-valued factor model with time-varying main effects

Clifford Lam [a], *, Zetai Cen [b]

[a] Department of Statistics, London School of Economics and Political Science, United Kingdom
[b] School of Mathematics, University of Bristol, United Kingdom

ARTICLE INFO

Keywords:
Large-scale dependent data
Time-varying row and column effects
MEFM and FM interchange
Sufficiency of FM over MEFM
Tucker decomposition

ABSTRACT

We introduce the matrix-valued time-varying Main Effects Factor Model (MEFM). MEFM is a generalization to the traditional matrix-valued factor model (FM). We give rigorous definitions of MEFM and its identifications, and propose estimators for the time-varying grand mean, row and column main effects, and the row and column factor loading matrices for the common component. Rates of convergence for different estimators are spelt out, with asymptotic normality shown. The core rank estimator for the common component is also proposed, with consistency of the estimators presented. As time series, the row and column main effects $\{\boldsymbol{\alpha}_t\}$ and $\{\boldsymbol{\beta}_t\}$ can be non-stationary without affecting the estimation accuracy of our estimators. The number of main effects factors contributing to row or column main effects is also consistently estimated by our proposed estimators. We propose a test for testing if FM is sufficient against the alternative that MEFM is necessary, and demonstrate the power of such a test in various simulation settings. We also demonstrate numerically the accuracy of our estimators in extended simulation experiments. A set of NYC Taxi traffic data is analyzed and our test suggests that MEFM is indeed necessary for analyzing the data against a traditional FM.

1. Introduction

Matrix-valued time series factor models, a generalization of vector time series factor models (Bai, 2003; Stock and Watson, 2002), have been utilized a lot for dimension reduction and prediction in recent years in fields such as finance, economics, medical science and meteorology, to name but a few. This is a subject still in its infancy, but important earlier theoretical and methodological developments include (Wang et al., 2019), Chen et al. (2020), Chen and Fan (2023) and He et al. (2024), which are all on matrix-valued factor models using the Tucker decomposition for the common component. Chang et al. (2023) uses the CP decomposition for the common component, while Guan (2023) also considers CP decomposition of the common component but taking in covariates in the loadings. Beyond factor modeling, Chen et al. (2021), Wu and Bi (2023) and H.-F. (2024) propose autoregressive and moving average models for matrix-valued time series data. See Tsay (2023) for a comprehensive review of matrix-valued time series analysis.

With Tucker decomposition, a matrix-valued time series factor model (FM) can be written as

$$\mathbf{Y}_t = \boldsymbol{\mu} + \mathbf{R}\mathbf{F}_t\mathbf{C}' + \mathbf{E}_t, \tag{1.1}$$

where $\mathbf{Y}_t \in \mathbb{R}^{p\times q}$ is the observed matrix at time t, $\mu \in \mathbb{R}^{p\times q}$ is the mean matrix, $\mathbf{R} \in \mathbb{R}^{p\times k_r}$ and $\mathbf{C} \in \mathbb{R}^{q\times k_c}$ are the row and column factor loading matrices respectively, $\mathbf{F}_t \in \mathbb{R}^{k_r\times k_c}$ is the core factor matrix at time t, and finally $\mathbf{E}_t \in \mathbb{R}^{p\times q}$ is the noise matrix at time t. If we set $\mathbf{R} := (\boldsymbol{\alpha}_{p\times r}, \widetilde{\mathbf{R}}_{p\times(k_r-r-\ell)}, \mathbf{1}_{p\times\ell})$, $\mathbf{C} := (\mathbf{1}_{q\times r}, \widetilde{\mathbf{C}}_{q\times(k_c-r-\ell)}, \boldsymbol{\beta}_{q\times\ell})$ and $\mathbf{F}_t := \mathrm{diag}((\mathbf{g}_t)_{r\times r}, (\widetilde{\mathbf{F}}_t)_{(k_r-r-\ell)\times(k_c-r-\ell)}, (\mathbf{h}_t)_{\ell\times\ell})$ (He et al.,

---
* Corresponding author.
E-mail addresses: C.Lam2@lse.ac.uk (C. Lam), Z.Cen@lse.ac.uk (Z. Cen).

https://doi.org/10.1016/j.jeconom.2025.106105
Received 7 June 2024; Received in revised form 16 June 2025; Accepted 20 September 2025

Available online 26 September 2025
0304-4076/

2023), where $\mathbf{1}_{m\times n}$ is a matrix of ones of size $m \times n$, then (1.1) becomes

$$\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\alpha}\mathbf{g}_t\mathbf{1}_{r\times q} + \mathbf{1}_{p\times \ell}\mathbf{h}_t\boldsymbol{\beta}' + \widetilde{\mathbf{R}}\widetilde{\mathbf{F}}_t\widetilde{\mathbf{C}}' + \mathbf{E}_t. \quad (1.2)$$

If the rows of $\mathbf{Y}_t$ represent different countries and the columns represent different economic indicators, then since the jth row of $\boldsymbol{\alpha}\mathbf{g}_t\mathbf{1}_{r\times q}$ is $\boldsymbol{\alpha}_{j\cdot}\mathbf{g}_t\mathbf{1}_{r\times q}$, where $\boldsymbol{\alpha}_{j\cdot}$ is the jth row of $\boldsymbol{\alpha}$, it means that each element in the jth row is the same, with value $\boldsymbol{\alpha}_{j\cdot}\mathbf{g}_t\mathbf{1}_r$. Hence we can argue that $\mathbf{g}_t$ represents common global factors affecting all countries, although each country loads differently on $\mathbf{g}_t$. Similarly, $\mathbf{h}_t$ represents latent economic states across different economic indicators. The term $\widetilde{\mathbf{R}}\widetilde{\mathbf{F}}_t\widetilde{\mathbf{C}}'$ can be viewed as an interaction term, while $\boldsymbol{\alpha}\mathbf{g}_t\mathbf{1}_{r\times q}$ the country main effects, and $\mathbf{1}_{p\times\ell}\mathbf{h}_t\boldsymbol{\beta}'$ the economic states' main effects.

Three problems arise upon inspecting (1.1) and (1.2) however. Firstly, for (1.1) to transform to (1.2), $\mathbf{R}$ and $\mathbf{C}$ are both of reduced rank. In the literature for model (1.1), we always need $\mathbf{R}$ and $\mathbf{C}$ to be of full rank at least asymptotically (see for example, Assumption (B2) in He et al. (2023) or Equation (8) in Chen and Fan (2023)) for estimation purpose.

Secondly, model (1.2) is not general enough, unless r and ℓ can be large. For example, if r is small, each country is driven only by few global common factors affecting all countries, on top of the factors in $\widetilde{\mathbf{F}}_t$. This will not be a problem, if not for the fact that there can be latent common factors that only drive a small group of countries/economic indicators. For instance, there can be a few small European countries which do not share global common factors with the majority of European countries, but with other middle-Eastern countries. Such "grouping" of countries usually comes with their corresponding groups of unique factors. These unique factors become "weak" country effects, shared only among "small" number of countries, essentially inflating the value of r while inducing a sparse $\boldsymbol{\alpha}$. Constraint factor modeling in Chen et al. (2020) can certainly help, but we do not always know the exact group of countries which share latent common factors.

The final problem is related to the second one. The inability of (1.2) to accommodate "weak" country/economic states effects originates from the fact that the common component in (1.1), $\widetilde{\mathbf{R}}\widetilde{\mathbf{F}}_t\widetilde{\mathbf{C}}'$, contains only pervasive factors, which is essentially assumed across all past works in factor models for matrix-valued time series. In a tensor setting however, Cen and Lam (2025) and Chen and Lam (2024) have both allowed weak factors in the common component in the factor model.

One way to generalize (1.2) to address all aforementioned problems is to note that

$$\boldsymbol{\alpha}\mathbf{g}_t\mathbf{1}_{r\times q} = (\boldsymbol{\alpha}\mathbf{g}_t\mathbf{1}_r)\mathbf{1}_q' =: \boldsymbol{\alpha}_t\mathbf{1}_q', \quad \mathbf{1}_{p\times\ell}\mathbf{h}_t\boldsymbol{\beta}' = \mathbf{1}_p(\boldsymbol{\beta}\mathbf{h}_t'\mathbf{1}_\ell)' =: \mathbf{1}_p\boldsymbol{\beta}_t',$$

where $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ are the time-varying row and column main effects respectively. If we are able to estimate the two vectors $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ without any low-rank constraints as in the equation above, then the second problem is naturally solved. As an independent interest however, we develop consistent estimators for the number of factors affecting the row/column main effects in Section 4.7, allowing the number of such factors to be diverging as quick as the number of rows/columns of the data matrix $\mathbf{Y}_t$. Formally allowing for weak factors in the row and column loading matrices, like those in Lam and Yao (2012) for a vector factor model, solves the third problem. Finally, with these problems solved, we can go back to assuming full rank row and column factor loading matrices (see Assumption (L1) in Section 4.1 in this paper) to solve the first problem.

In this paper, we contribute to the literature in several important ways. Firstly, we generalize model (1.2) to (3.1) which is the time-varying main effects factor model (MEFM), incorporating all relaxations described in the previous paragraph. Such an MEFM model is more general than FM in (1.1) since we "derived" model (1.2) from (1.1) allowing $\mathbf{R}$ and $\mathbf{C}$ to be of reduced rank, when we are in fact restricted to only full rank $\mathbf{R}$ and $\mathbf{C}$ in (1.1) in real usage in order for both loading matrices to be identifiable and estimable. MEFM further enhances the generality over (1.2) by allowing r and ℓ to be as large as p and q respectively, essentially allowing weak row/column factors incorporated in $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ respectively. See also the explanations in Section 3.1 where MEFM can be written as FM but has to incorporate diverging number of factors when r and ℓ are indeed diverging as p and q respectively. In addition, $\{\boldsymbol{\alpha}_t\}$ and $\{\boldsymbol{\beta}_t\}$ are allowed to be non-stationary (except when we need to consistently estimate r and ℓ in Theorem 10). These show that MEFM in (3.1) is more general than FM in (1.1).

In doing so, we also consider the following model:

$$\mathbf{Y}_t = \mu_t\mathbf{1}_p\mathbf{1}_q' + \boldsymbol{\alpha}\mathbf{g}_t\mathbf{1}_q' + (\boldsymbol{\beta}\mathbf{h}_t\mathbf{1}_p')' + \mathbf{A}_r\mathbf{F}_t\mathbf{A}_c' + \mathbf{E}_t, \quad (1.3)$$

which is (3.1) but with $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ there restricted to $\boldsymbol{\alpha}\mathbf{g}_t$ and $\boldsymbol{\beta}\mathbf{h}_t$ respectively, so that the row and column main effects are generated by $(\mathbf{g}_t)_{r\times 1}$ and $(\mathbf{h}_t)_{\ell\times 1}$ respectively. We provide consistent estimators for r and ℓ for users to see if the row/column main effects are generated by r/ℓ global common factors affecting all rows/columns. See more explanations and details in Section 4.7. See also the end of Remark 1 in Section 3.2 as well.

Secondly, we provide estimation and inference methods and the corresponding theoretical guarantees, on top of a separate ratio-based method for identifying the core rank of $\mathbf{F}_t$, with consistency proved . Third and perhaps the most important of all, we provide a statistical test to test if FM in (1.1), with $\mathbf{R}$ and $\mathbf{C}$ both of full rank, is sufficient against the more general MEFM in (3.1). A rejected null hypothesis of FM being sufficient for the data then means that there are row and/or column main effects that is not of a low rank structure like those in (1.2), essentially pointing to the existence of "weak" main effects.

The rest of the paper is organized as follows. Section 2 introduces the notations used in this paper. Section 3 introduces MEFM formally, laying down important identification conditions and estimation methodologies for all the components in the model. Section 4 presents the assumptions for MEFM and the consistency and asymptotic normality results for its estimators. The test for FM versus MEFM is detailed in Section 4.5, while the core rank estimator for $\mathbf{F}_t$ is presented in Section 4.6. Finally, Section 5 presents our extensive simulation results and details the NYC Taxi traffic data analysis, pinpointing the presence of weak hourly main effects in the data. Our method is available in the R package `MEFM`, with instruction in its reference manual on CRAN. All proofs of the theorems are relegated to the supplementary materials of this paper.

2. Notations

Throughout this paper, we use the lower-case letter, bold lower-case letter and bold capital letter, i.e., $a, \mathbf{a}, \mathbf{A}$, to denote a scalar, a vector and a matrix respectively. We also use $a_i, A_{ij}, \mathbf{A}_{i\cdot}, \mathbf{A}_{\cdot i}$ to denote, respectively, the ith element of $\mathbf{a}$, the (i,j)-th element of $\mathbf{A}$, the ith row vector (as a column vector) of $\mathbf{A}$, and the ith column vector of $\mathbf{A}$. We use $\circ$ to denote the Hadamard product. We use $a \asymp b$ to denote $a = O(b)$ and $b = O(a)$, while $a \asymp_P b$ to denote $a = O_P(b)$ and $b = O_P(a)$. A random variable X is sub-Gaussian with variance proxy σ^2, denoted as $X \sim \text{subG}(\sigma^2)$, if $\mathbb{E}[\exp(s(X - \mathbb{E}[X]))] \le \exp(s^2\lambda^2/2)$ for all $s \in \mathbb{R}$. A random variable X is sub-exponential with parameter λ, denoted as $X \sim \text{subE}(\lambda)$, if $\mathbb{E}[\exp(s(X - \mathbb{E}[X]))] \le \exp(s^2\lambda^2/2)$ for all $|s| \le 1/\lambda$.

Given a positive integer m, define $[m] := \{1, 2, \dots, m\}$. The vector $\mathbf{1}_m$ denotes a vector of ones of length m. The ith largest eigenvalue of a matrix $\mathbf{A}$ is denoted by $\lambda_i(\mathbf{A})$. The notation $\mathbf{A} \succcurlyeq 0$ (resp. $\mathbf{A} \succ 0$) means that $\mathbf{A}$ is positive semi-definite (resp. positive definite). We use $\mathbf{A}'$ to denote the transpose of $\mathbf{A}$, and $\text{diag}(\mathbf{A})$ to denote a diagonal matrix with the diagonal elements of $\mathbf{A}$, while $\text{diag}(\{a_1, \dots, a_n\})$ or $\text{diag}(\mathbf{a})$ represents the diagonal matrix with $\{a_1, \dots, a_n\}$ or the elements in the vector $\mathbf{a}$ on the diagonal, respectively.

Some norm notations. For a given set, we denote by $|\cdot|$ its cardinality. We use $\|\cdot\|$ to denote the spectral norm of a matrix or the L_2 norm of a vector, and $\|\cdot\|_F$ to denote the Frobenius norm of a matrix. We use $\|\cdot\|_{\max}$ to denote the maximum absolute value of the elements in a vector or a matrix. The notations $\|\cdot\|_1$ and $\|\cdot\|_\infty$ denote the L_1 and L_∞-norm of a matrix respectively, defined by $\|\mathbf{A}\|_1 := \max_j \sum_i |(\mathbf{A})_{ij}|$ and $\|\mathbf{A}\|_\infty := \max_i \sum_j |(\mathbf{A})_{ij}|$. WLOG, we always assume the eigenvalues of a matrix are arranged by descending orders, and so are their corresponding eigenvectors.

3. Model and estimation

3.1. Main effect matrix factor model

We propose the time-varying **M**ain **E**ffect matrix **F**actor **M**odel (MEFM) such that for $t \in [T]$,

$$\mathbf{Y}_t = \mu_t \mathbf{1}_p \mathbf{1}_q' + \boldsymbol{\alpha}_t \mathbf{1}_q' + \mathbf{1}_p \boldsymbol{\beta}_t' + \mathbf{C}_t + \mathbf{E}_t, \tag{3.1}$$

where $\mathbf{Y}_t$ is a $p \times q$ observed matrix at time t, μ_t is a scalar representing the grand mean of $\mathbf{Y}_t$, $\boldsymbol{\alpha}_t \in \mathbb{R}^p$ and $\boldsymbol{\beta}_t \in \mathbb{R}^q$ are the row and column main effects at time t, respectively. The common component $\mathbf{C}_t := \mathbf{A}_r \mathbf{F}_t \mathbf{A}_c'$ is latent, where $\mathbf{F}_t \in \mathbb{R}^{k_r \times k_c}$ is the core factor series with unknown number of factors k_r and k_c, and $\mathbf{A}_r$ and $\mathbf{A}_c$ are the row and column factor loading matrices, with size $p \times k_r$ and $q \times k_c$, respectively. Lastly, $\mathbf{E}_t$ is the idiosyncratic noise series with the same dimension as $\mathbf{Y}_t$.

Unlike FM in (1.2), the main effects $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ in MEFM are not restricted to be of low rank, which significantly improves the flexibility of FM, and allows for a test of FM in (1.1) in the end. In fact, setting concatenated matrices $\ddot{\mathbf{A}}_r = (\mathbf{I}_p, \mathbf{A}_r, \mathbf{1}_p)$ and $\ddot{\mathbf{A}}_c = (\mathbf{1}_q, \mathbf{A}_c, \mathbf{I}_q)$, block matrix

$$\ddot{\mathbf{F}}_t = \begin{pmatrix} \boldsymbol{\alpha}_t & 0 & 0 \\ 0 & \mathbf{F}_t & 0 \\ \mu_t & 0 & \boldsymbol{\beta}_t' \end{pmatrix},$$

then we can read (3.1) as

$$\mathbf{Y}_t = \ddot{\mathbf{A}}_r \ddot{\mathbf{F}}_t \ddot{\mathbf{A}}_c' + \mathbf{E}_t.$$

However, we observe that the dimension of the factor series is now $(1 + p + k_r) \times (1 + q + k_c)$, and hence there is not much dimension reduction for $\mathbf{Y}_t$, and both $\ddot{\mathbf{A}}_r$ and $\ddot{\mathbf{A}}_c$ have no full column ranks. This observation suggests again that MEFM is more general than FM, and numerical results in Section 5 actually show that even an approximate estimation by FM in general comes at a cost of using very large number of factors.

Given the above motivation of MEFM, we point out that the form of MEFM can be obtained by FM in general, see Remark 1 for details. For generality purpose, $\mathbf{Y}_t$ can have non-zero mean but we can always demean the data as the sample mean is not our main parameter of interest. The right hand side of (3.1) is entirely latent and hence we propose Assumption (IC1) below to identify the grand mean and the row and column effects.

(IC1) (Identification) *For any $t \in [T]$, we assume $\mathbf{1}_p' \boldsymbol{\alpha}_t = \mathbf{1}_q' \boldsymbol{\beta}_t = 0$, $\mathbf{1}_p' \mathbf{A}_r = 0$ and $\mathbf{1}_q' \mathbf{A}_c = \mathbf{0}$.*

Condition (IC1) clarifies both the meaning and the differences between main effects and core factor effects from the common component $\mathbf{C}_t$. Here we focus our explanations on the row main effects since the column main effects have similar interpretations. The term $\boldsymbol{\alpha}_t \mathbf{1}_p'$ in model (3.1) means that for the ith row, the main effect for each column stays the same at $\alpha_{i,t}$, which is the definition of row main effects in a contingency table. The condition $\mathbf{1}_p' \boldsymbol{\alpha}_t = 0$ means that the level of the main effects $\alpha_{i,t}$ are relative, since all non-balanced main effects at time t are absorbed into μ_t, the grand mean at time t. The core factor effects from $\mathbf{C}_t = \mathbf{A}_r \mathbf{F}_t \mathbf{A}_c'$, however, cannot represent any main effects. For, if the ith row of $\mathbf{C}_t$ represents row main effects, then this effect is the average sum of the elements of the ith row, which is $q^{-1} \mathbf{A}_{r,i\cdot} \mathbf{F}_t \mathbf{A}_c' \mathbf{1}_q = 0$ by the identification condition $\mathbf{A}_c' \mathbf{1}_q = 0$ in (IC1). Hence in a sense, (IC1) helps identify the main effects as purely relative, while the common components is identified as the residual interaction effects between row and column variables after the grand mean and the row and column main effects are identified.

As a simple example, consider time series of GDP (in percentage) for a group of countries (row) versus categorization by different service sectors (column), including financial, retail, hospitality etc. Condition (IC1) helps visualize if there are group of countries having positive/negative effects at a certain time irrespective of which service sector we are looking at, indicating a "country" effect of a relatively higher/lower level of service sector GDP compared to other countries irrespective of sectors. A similar "sector" effect can also be observed to see if there is a particular service sector that is higher/lower than other GDP percentages irrespective of countries. The residual interaction effects between countries and sectors are then summarized in the common component $\mathbf{C}_t$, which is main-effect free with the help of $\mathbf{A}_r'\mathbf{1}_p = 0$ and $\mathbf{A}_c'\mathbf{1}_q = 0$ in (IC1). If there are any factors in $\mathbf{C}_t$ that are "localized" to certain countries or sectors, their mean levels would have been absorbed into the main effects of those countries or sectors.

An obvious advantage of (IC1) where $\mathbf{1}_p'\boldsymbol{\alpha}_t = \mathbf{1}_q'\boldsymbol{\beta}_t = 0$ is easy comparisons of the levels of row/column effects, since high levels will usually end up with positive main effects, and vice versa. Yet these conditions are not unique for identifications. In fact, if we want to visualize a time series of a certain country's main effect over time, and determine if it has extended periods of insignificant main effect levels compared to other countries, we would want to sparsify the time series and perhaps use penalized estimation on $\boldsymbol{\alpha}_t$ over time. However, this does not make sense since low levels mean negative or largely negative main effects under (IC1), which is not suitable for penalization (towards 0). Another identification, where we set the minimum value in $\boldsymbol{\alpha}_t$ to be 0, is a more effective identification for such a purpose. Such a pursuit is left in a future project.

We require further identification between the factors and the factor loading matrices. To do this, we normalize the loading matrices to $\mathbf{Q}_r = \mathbf{A}_r\mathbf{Z}_r^{-1/2}$ and $\mathbf{Q}_c = \mathbf{A}_c\mathbf{Z}_c^{-1/2}$, where $\mathbf{Z}_r = \mathrm{diag}(\mathbf{A}_r'\mathbf{A}_r)$ and $\mathbf{Z}_c = \mathrm{diag}(\mathbf{A}_c'\mathbf{A}_c)$, measuring the sparsity of each column of loading matrices and hence the factor strength. For example, $\mathbf{F}_t$ pervasive in the jth row will have the jth column of $\mathbf{A}_r$ dense and hence the jth diagonal entry of $\mathbf{Z}_r$ will be of order p. For technical details, see Assumption (L1). We leave the identification to Section 4.1. Assumption (IC1) also facilitates the estimation of μ_t, $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$, and we discuss in the next section how to estimate the grand mean, the row and column effects, and the row and column factor loading matrices in (3.1).

3.2. Estimation of the main effects and factor components

The factor structure is hidden in $\mathbf{Y}_t$ and we need to estimate the time-varying grand mean and main effects first. For the grand mean, right-multiplying by $\mathbf{1}_q$ and left-multiplying by $\mathbf{1}_p'$ on both sides of (3.1) results in $\mathbf{1}_p'\mathbf{Y}_t\mathbf{1}_q = pq\mu_t + \mathbf{1}_p'\mathbf{E}_t\mathbf{1}_q$ by Assumption (IC1). Hence for each $t \in [T]$, we obtain the moment estimator for the time-varying grand mean as

$$\hat{\mu}_t := \mathbf{1}_p'\mathbf{Y}_t\mathbf{1}_q/pq. \tag{3.2}$$

Also, right-multiplying by $\mathbf{1}_q$ and left-multiplying by $\mathbf{1}_p'$ lead respectively to $\mathbf{Y}_t\mathbf{1}_q = q\mu_t\mathbf{1}_p + q\boldsymbol{\alpha}_t + \mathbf{E}_t\mathbf{1}_q$ and $\mathbf{1}_p'\mathbf{Y}_t = p\mu_t\mathbf{1}_q' + p\boldsymbol{\beta}_t' + \mathbf{1}_p'\mathbf{E}_t$. Therefore, we obtain the time-varying row and column effect estimators as

$$\hat{\boldsymbol{\alpha}}_t := q^{-1}\mathbf{Y}_t\mathbf{1}_q - \hat{\mu}_t\mathbf{1}_p, \quad \hat{\boldsymbol{\beta}}_t' := p^{-1}\mathbf{1}_p'\mathbf{Y}_t - \hat{\mu}_t\mathbf{1}_q'. \tag{3.3}$$

Finally, we introduce the following to estimate the factor structure,

$$\begin{aligned}\hat{\mathbf{L}}_t &:= \mathbf{Y}_t - \hat{\mu}_t\mathbf{1}_p\mathbf{1}_q' - \hat{\boldsymbol{\alpha}}_t\mathbf{1}_q' - \mathbf{1}_p\hat{\boldsymbol{\beta}}_t' = \mathbf{Y}_t + (pq)^{-1}\mathbf{1}_p'\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_p\mathbf{1}_q' - q^{-1}\mathbf{Y}_t\mathbf{1}_q\mathbf{1}_q' - p^{-1}\mathbf{1}_p\mathbf{1}_p'\mathbf{Y}_t \\ &= \mathbf{M}_p\mathbf{Y}_t\mathbf{M}_q,\end{aligned} \tag{3.4}$$

where $\mathbf{M}_m := \mathbf{I}_m - m^{-1}\mathbf{1}_m\mathbf{1}_m'$ for any positive integer m. From the above, $\hat{\mathbf{L}}_t\hat{\mathbf{L}}_t'$ admits $\mathbf{1}_p$ in its null space, and $\hat{\mathbf{L}}_t'\hat{\mathbf{L}}_t$ admits $\mathbf{1}_q$ instead. The factor structure can hence be estimated, with $\hat{\mathbf{Q}}_r$ constructed as the eigenvectors of $T^{-1}\sum_{t=1}^T \hat{\mathbf{L}}_t\hat{\mathbf{L}}_t'$ corresponding to the first k_r largest eigenvalues, and $\hat{\mathbf{Q}}_c$ the eigenvectors of $T^{-1}\sum_{t=1}^T \hat{\mathbf{L}}_t'\hat{\mathbf{L}}_t$ corresponding to the first k_c largest eigenvalues.

We can then estimate the factor time series $\mathbf{F}_{Z,t} = \mathbf{Z}_r^{1/2}\mathbf{F}_t\mathbf{Z}_c^{1/2}$, and the common component $\mathbf{C}_t$, respectively as

$$\hat{\mathbf{F}}_{Z,t} := \hat{\mathbf{Q}}_r'\hat{\mathbf{L}}_t\hat{\mathbf{Q}}_c = \hat{\mathbf{Q}}_r'\mathbf{Y}_t\hat{\mathbf{Q}}_c, \quad \hat{\mathbf{C}}_t := \hat{\mathbf{Q}}_r\hat{\mathbf{F}}_{Z,t}\hat{\mathbf{Q}}_c' = \hat{\mathbf{Q}}_r\hat{\mathbf{Q}}_r'\mathbf{Y}_t\hat{\mathbf{Q}}_c\hat{\mathbf{Q}}_c'. \tag{3.5}$$

Finally, the residuals $\mathbf{E}_t$ is estimated by

$$\hat{\mathbf{E}}_t := \hat{\mathbf{L}}_t - \hat{\mathbf{C}}_t. \tag{3.6}$$

Remark 1. *Suppose we have a traditional matrix-valued factor model such that $\acute{\mathbf{Y}}_t = \acute{\mathbf{C}}_t + \acute{\mathbf{E}}_t$ where $\acute{\mathbf{Y}}_t$, $\acute{\mathbf{C}}_t$, and $\acute{\mathbf{E}}_t$ are $p \times q$ matrices representing the observation, common component, and noise, respectively. Suppose also $\acute{\mathbf{C}}_t = \mathbf{A}_r\mathbf{F}_t\mathbf{A}_c'$. Then we can construct*

$$\acute{\mu}_t := (pq)^{-1}\mathbf{1}_p'\acute{\mathbf{C}}_t\mathbf{1}_q, \quad \acute{\boldsymbol{\alpha}}_t := q^{-1}\acute{\mathbf{C}}_t\mathbf{1}_q - \acute{\mu}_t\mathbf{1}_p = q^{-1}\mathbf{M}_p\acute{\mathbf{C}}_t\mathbf{1}_q, \quad \acute{\boldsymbol{\beta}}_t := p^{-1}\acute{\mathbf{C}}_t'\mathbf{1}_p - \acute{\mu}_t\mathbf{1}_q = p^{-1}\mathbf{M}_q\acute{\mathbf{C}}_t'\mathbf{1}_p.$$

Hence we can express FM in the following MEFM form satisfying (IC1):

$$\acute{\mathbf{Y}}_t = \acute{\mu}_t\mathbf{1}_p\mathbf{1}_q' + \acute{\boldsymbol{\alpha}}_t\mathbf{1}_q' + \mathbf{1}_p\acute{\boldsymbol{\beta}}_t' + (\acute{\mathbf{C}}_t - \acute{\mu}_t\mathbf{1}_p\mathbf{1}_q' - \acute{\boldsymbol{\alpha}}_t\mathbf{1}_q' - \mathbf{1}_p\acute{\boldsymbol{\beta}}_t') + \acute{\mathbf{E}}_t,$$

where

$$\acute{\mathbf{C}}_t - \acute{\mu}_t\mathbf{1}_p\mathbf{1}_q' - \acute{\boldsymbol{\alpha}}_t\mathbf{1}_q' - \mathbf{1}_p\acute{\boldsymbol{\beta}}_t' = (\mathbf{M}_p\mathbf{A}_r)\mathbf{F}_t(\mathbf{M}_q\mathbf{A}_c)',$$

is the common component. Since $\mathbf{M}_m\mathbf{1}_m = 0$, it is easy to see that

$$\mathbf{1}_p'(\mathbf{M}_p\mathbf{A}_r) = 0, \quad \mathbf{1}_q'(\mathbf{M}_q\mathbf{A}_c) = 0.$$

It is also easy to verify that $\mathbf{1}_p'\acute{\alpha}_t = \mathbf{1}_q'\acute{\beta}_t = 0$. *Hence a traditional matrix-valued factor model can be expressed as MEFM in (3.1) to satisfy (IC1). Such MEFM is also consistent with model (1.3), since we can rewrite*

$$\acute{\alpha}_t = q^{-1}\mathbf{M}_p\mathbf{A}_r\mathbf{F}_t\mathbf{A}_c'\mathbf{1}_q = \boldsymbol{\alpha}\mathbf{g}_t, \quad \text{where}$$
$$\boldsymbol{\alpha} := \mathbf{M}_p\mathbf{A}_r diag^{1/2}(\mathrm{Var}(q^{-1}\mathbf{F}_t\mathbf{A}_c'\mathbf{1}_q)), \quad \mathbf{g}_t := diag^{-1/2}(\mathrm{Var}(q^{-1}\mathbf{F}_t\mathbf{A}_c'\mathbf{1}_q))(q^{-1}\mathbf{F}_t\mathbf{A}_c'\mathbf{1}_q),$$

so that $\mathbf{g}_t$ *has independent elements with mean 0 and variance 1, if* $\{\mathbf{F}_t\}$ *satisfies Assumption (F1) in Section 4.1 below. Similar treatments for* $\acute{\beta}_t$.

Remark 2 (*Computational Complexity*). *Although the proposed MEFM has more parameters than FM (for the same* k_r *and* k_c *for both MEFM and FM), the estimation procedure remains efficient. First of all, estimating each grand mean by (3.2) requires* $O(pq+q)$ *operations if we compute* $\mathbf{1}_p'\mathbf{Y}_t$ *first or* $O(pq+p)$ *if* $\mathbf{Y}_t\mathbf{1}_q$ *is computed first. Hence the computation could be empirically optimized, although both imply computational complexity of the order of* pq. *For both row and column main effect estimators by (3.3),* $O(pq)$ *operations are required. Next, computing* $\widehat{\mathbf{L}}_t$ *from (3.4) requires* $O(pq(p+q))$ *operations, and hence the computational complexity for obtaining* $T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'$ *and* $T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t'\widehat{\mathbf{L}}_t$ *are of order* Tp^2q *and* Tpq^2, *respectively.*

Note that once the grand mean and main effects are computed, the remaining steps are the same as estimating the factor structure in FM, and hence for simplicity, the loading estimators $\widehat{\mathbf{Q}}_r$ *and* $\widehat{\mathbf{Q}}_c$ *are then obtained in* $O(p^3)$ *and* $O(q^3)$ *operations using the conventional computational complexity of SVD for Hermitian matrices, albeit more advanced algorithms could be employed (Banks et al., 2023). To compute each* $\widehat{\mathbf{F}}_{Z,t}$ *in (3.5), it takes* $O(pq)$ *operations theoretically, but we may again optimize the empirical runtime by first computing* $\widehat{\mathbf{Q}}_r'\mathbf{Y}_t$ *if* $q<p$ *or* $\mathbf{Y}_t\widehat{\mathbf{Q}}_c$ *otherwise. Finally,* $\widehat{\mathbf{C}}_t$ *and* $\widehat{\mathbf{E}}_t$ *from (3.5) and (3.6) can be obtained in* $O(pq)$ *operations. Overall, estimating all parameters in MEFM has the rate of* $O(Tp^2q+Tpq^2+p^3+q^3)$, *which is the same as that in estimating the parameters in FM. Hence, besides being more general than FM, MEFM enjoys the same computational complexity in parameter estimation.*

4. Assumptions and theoretical results

4.1. Assumptions

A set of assumptions on the factor structure is imposed below, and in particular, we allow factors to have different strengths, as in Lam and Yao (2012) and Chen and Lam (2024).

(M1) (Alpha mixing) *The elements in* $\mathbf{vec}(\mathbf{F}_t)$ *and* $\mathbf{vec}(\mathbf{E}_t)$ *are* α*-mixing. A vector process* $\{\mathbf{x}_t : t = 0, \pm1, \pm2, \dots\}$ *is* α*-mixing if, for some* $\gamma > 2$, *the mixing coefficients satisfy the condition that*

$$\sum_{h=1}^{\infty} \alpha(h)^{1-2/\gamma} < \infty,$$

where $\alpha(h) = \sup_\tau \sup_{A\in\mathcal{H}_{-\infty}^{\tau}, B\in\mathcal{H}_{\tau+h}^{\infty}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|$ *and* $\mathcal{H}_\tau^s$ *is the* σ*-field generated by* $\{\mathbf{x}_t : \tau \le t \le s\}$.

(F1) (Time Series in $\mathbf{F}_t$) *There is* $\mathbf{X}_{f,t}$ *the same dimension as* $\mathbf{F}_t$, *such that* $\mathbf{F}_t = \sum_{w\ge0} a_{f,w}\mathbf{X}_{f,t-w}$. *The time series* $\{\mathbf{X}_{f,t}\}$ *has i.i.d. elements with mean* 0 *and variance* 1, *with uniformly bounded fourth order moments. The coefficients* $a_{f,w}$ *are such that* $\sum_{w\ge0} a_{f,w}^2 = 1$ *and* $\sum_{w\ge0}|a_{f,w}| \le c$ *for some constant* c.

(L1) (Factor strength) *We assume that* $\mathbf{A}_r$ *and* $\mathbf{A}_c$ *are of full rank and independent of factors and errors series. Furthermore, as* $p, q \to \infty$,

$$\mathbf{Z}_r^{-1/2}\mathbf{A}_r'\mathbf{A}_r\mathbf{Z}_r^{-1/2} \to \boldsymbol{\Sigma}_{A,r}, \quad \mathbf{Z}_c^{-1/2}\mathbf{A}_c'\mathbf{A}_c\mathbf{Z}_c^{-1/2} \to \boldsymbol{\Sigma}_{A,c}, \tag{4.1}$$

where $\mathbf{Z}_r = \mathrm{diag}(\mathbf{A}_r'\mathbf{A}_r)$, $\mathbf{Z}_c = \mathrm{diag}(\mathbf{A}_c'\mathbf{A}_c)$, *and both* $\boldsymbol{\Sigma}_{A,r}$ *and* $\boldsymbol{\Sigma}_{A,c}$ *are positive definite with all eigenvalues bounded away from 0 and infinity. We assume* $(\mathbf{Z}_r)_{jj} \asymp p^{\delta_{r,j}}$ *for* $j\in[k_r]$ *and* $1/2 < \delta_{r,k_r} \le \cdots \le \delta_{r,2} \le \delta_{r,1} \le 1$. *Similarly, we assume* $(\mathbf{Z}_c)_{jj} \asymp p^{\delta_{c,j}}$ *for* $j\in[k_c]$, *with* $1/2 < \delta_{c,k_c} \le \cdots \le \delta_{c,2} \le \delta_{c,1} \le 1$.

With Assumption (L1), we can denote $\mathbf{Q}_r := \mathbf{A}_r\mathbf{Z}_r^{-1/2}$ and $\mathbf{Q}_c := \mathbf{A}_c\mathbf{Z}_c^{-1/2}$. Hence $\mathbf{Q}_r'\mathbf{Q}_r \to \boldsymbol{\Sigma}_{A,r}$ and $\mathbf{Q}_c'\mathbf{Q}_c \to \boldsymbol{\Sigma}_{A,c}$.

(E1) (Decomposition of $\mathbf{E}_t$) *We assume that*

$$\mathbf{E}_t = \mathbf{A}_{e,r}\mathbf{F}_{e,t}\mathbf{A}_{e,c}' + \boldsymbol{\Sigma}_\epsilon \circ \boldsymbol{\epsilon}_t, \tag{4.2}$$

where $\mathbf{F}_{e,t}$ *is a matrix of size* $k_{e,r}\times k_{e,c}$, *containing independent elements with mean* 0 *and variance* 1. *The matrix* $\boldsymbol{\epsilon}_t \in \mathbb{R}^{p\times q}$ *contains independent and identically distributed elements with mean 0 and variance 1, with* $\{\boldsymbol{\epsilon}_t\}$ *independent of* $\{\mathbf{F}_{e,t}\}$. *The matrix* $\boldsymbol{\Sigma}_\epsilon$ *contains the standard deviations of the corresponding elements in* $\boldsymbol{\epsilon}_t$, *and has elements uniformly bounded away from 0 and infinity.*
Moreover, $\mathbf{A}_{e,r}$ *and* $\mathbf{A}_{e,c}$ *are (approximately) sparse matrices with sizes* $p\times k_{e,r}$ *and* $q\times k_{e,c}$ *respectively, such that* $\|\mathbf{A}_{e,r}\|_1, \|\mathbf{A}_{e,c}\|_1 = O(1)$, *with* $k_{e,r}, k_{e,c} = O(1)$.

(E2) (Time Series in $\mathbf{E}_t$) *There is $\mathbf{X}_{e,t}$ the same dimension as $\mathbf{F}_{e,t}$, and $\mathbf{X}_{\epsilon,t}$ the same dimension as ϵ_t, such that $\mathbf{F}_{e,t} = \sum_{w\geq 0} a_{e,w}\mathbf{X}_{e,t-w}$ and $\epsilon_t = \sum_{w\geq 0} a_{\epsilon,w}\mathbf{X}_{\epsilon,t-w}$, with $\{\mathbf{X}_{e,t}\}$ and $\{\mathbf{X}_{\epsilon,t}\}$ independent of each other. $\{\mathbf{X}_{e,t}\}$ has independent elements while $\{\mathbf{X}_{\epsilon,t}\}$ has i.i.d. elements, and all elements have mean zero with unit variance and uniformly bounded fourth order moments. Both $\{\mathbf{X}_{e,t}\}$ and $\{\mathbf{X}_{\epsilon,t}\}$ are independent of $\{\mathbf{X}_{f,t}\}$ from (F1).*
The coefficients $a_{e,w}$ and $a_{\epsilon,w}$ are such that $\sum_{w\geq 0} a_{e,w}^2 = \sum_{w\geq 0} a_{\epsilon,w}^2 = 1$ and $\sum_{w\geq 0}|a_{e,w}|, \sum_{w\geq 0}|a_{\epsilon,w}| \leq c$ for some constant c.

(R1) (Rate assumptions) *We assume that,*

$$T^{-1}p^{2(1-\delta_{r,k_r})}q^{1-2\delta_{c,1}} = o(1), \quad p^{1-2\delta_{r,k_r}}q^{2(1-\delta_{c,1})} = o(1),$$
$$T^{-1}q^{2(1-\delta_{c,k_c})}p^{1-2\delta_{r,1}} = o(1), \quad q^{1-2\delta_{c,k_c}}p^{2(1-\delta_{r,1})} = o(1).$$

Assumption (F1) introduces serial dependence into the factors, and (E1) and (E2) introduce both cross-sectional and temporal dependence in the noise. The factor structure depicted by (F1), (E1) and (E2) is the same as the one in Cen and Lam (2025). Note that although Assumption (M1) also features in serial dependence, it is mainly used to construct asymptotic normality of estimators.

By Assumption (L1), we have $\mathbf{A}_r\mathbf{F}_t\mathbf{A}_c' = \mathbf{Q}_r\mathbf{Z}_r^{1/2}\mathbf{F}_t\mathbf{Z}_c^{1/2}\mathbf{Q}_c'$, so we aim to estimate $(\mathbf{Q}_r, \mathbf{Q}_c, \mathbf{F}_{Z,t})$ where $\mathbf{F}_{Z,t} := \mathbf{Z}_r^{1/2}\mathbf{F}_t\mathbf{Z}_c^{1/2}$. Unlike the traditional approximate factor models which assumes all factors are pervasive, we allow factors to have different strength similar to Lam and Yao (2012) and Chen and Lam (2024). To be precise, a column of $\mathbf{A}_r$ (resp. $\mathbf{A}_c$) is dense (i.e., a pervasive factor) if the corresponding $\delta_{r,j} = 1$ (resp. $\delta_{c,j} = 1$), otherwise the column represents a weak factor as it is sparse to certain extent.

Due to the presence of potentially weak factors, we require rate conditions in Assumption (R1) for consistency to hold. If all factors are pervasive (R1) holds trivially. We point out that the first and second (or the third and fourth) conditions in (R1) are exactly the same as the first and third conditions of Assumption (R1) in Cen and Lam (2025) for matrix time series.

Remark 3. *The lack of assumptions on time series dynamics for the main effects $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ is not by accident. In fact, we allow $\{\boldsymbol{\alpha}_t\}$ and $\{\boldsymbol{\beta}_t\}$ to be non-stationary (apart from Theorem 10 which uses (ME1) in Section 4.7, an assumption essentially restricting $\{\boldsymbol{\alpha}_t\}$ and $\{\boldsymbol{\beta}_t\}$ to be stationary). This also highlights that our proposed MEFM is more general than FM, since the core factors in FM are assumed stationary. When MEFM is rewritten as FM, if the main effects are non-stationary, then the core factors in the rewritten FM are inevitably non-stationary, violating the essential assumption of stationary core factors for estimation purpose. See also Fig. 12 for our NYC taxi data analysis, showing clearly that $\{\boldsymbol{\beta}_t\}$ estimated is not stationary.*

4.2. Identification of the model

With Assumptions (IC1) and (L1), the model (3.1) is identified according to Theorem 1 below.

Theorem 1 (*Identification*). *With Assumption (IC1), each μ_t, $\boldsymbol{\alpha}_t$, and $\boldsymbol{\beta}_t$ can be identified. The common component is hence identified, and if (L1) is also satisfied, the factor structure is identified up to some rotation such that $(\mathbf{Q}_r, \mathbf{Q}_c, \mathbf{F}_{Z,t}) = (\mathbf{Q}_r\mathbf{M}_r, \mathbf{Q}_c\mathbf{M}_c, \mathbf{M}_r^{-1}\mathbf{F}_{Z,t}\mathbf{M}_c^{-1})$ for some invertible matrices $\mathbf{M}_r \in \mathbb{R}^{k_r\times k_r}$ and $\mathbf{M}_c \in \mathbb{R}^{k_c\times k_c}$.*

4.3. Rate of convergence for various estimators

To present the consistency of the loading estimators, define

$$\mathbf{H}_r := T^{-1}\widehat{\mathbf{D}}_r^{-1}\widehat{\mathbf{Q}}_r'\mathbf{Q}_r\sum_{t=1}^T(\mathbf{F}_{Z,t}\mathbf{Q}_c'\mathbf{Q}_c\mathbf{F}_{Z,t}'), \tag{4.3}$$
$$\mathbf{H}_c := T^{-1}\widehat{\mathbf{D}}_c^{-1}\widehat{\mathbf{Q}}_c'\mathbf{Q}_c\sum_{t=1}^T(\mathbf{F}_{Z,t}'\mathbf{Q}_r'\mathbf{Q}_r\mathbf{F}_{Z,t}), \tag{4.4}$$

where $\widehat{\mathbf{D}}_r := \widehat{\mathbf{Q}}_r'(T^{-1}\sum_{t=1}^T\widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t')\widehat{\mathbf{Q}}_r$ is the $k_r\times k_r$ diagonal matrix of eigenvalues of $T^{-1}\sum_{t=1}^T\widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'$, and similarly $\widehat{\mathbf{D}}_c := \widehat{\mathbf{Q}}_c'(T^{-1}\sum_{t=1}^T\widehat{\mathbf{L}}_t'\widehat{\mathbf{L}}_t)\widehat{\mathbf{Q}}_c$ is the $k_c\times k_c$ diagonal matrix of eigenvalues of $T^{-1}\sum_{t=1}^T\widehat{\mathbf{L}}_t'\widehat{\mathbf{L}}_t$.

Theorem 2. *Under Assumptions (IC1), (M1), (F1), (L1), (E1), (E2) and (R1), we have*

$$(\widehat{\mu}_t - \mu_t)^2 = O_P(p^{-1}q^{-1}), \tag{4.5}$$
$$p^{-1}\|\widehat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t\|^2 = O_P(q^{-1}), \tag{4.6}$$
$$q^{-1}\|\widehat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t\|^2 = O_P(p^{-1}), \tag{4.7}$$
$$p^{-1}\|\widehat{\mathbf{Q}}_r - \mathbf{Q}_r\mathbf{H}_r'\|_F^2 = O_P\Big(T^{-1}p^{1-2\delta_{r,k_r}}q^{1-2\delta_{c,1}} + p^{-2\delta_{r,k_r}}q^{2(1-\delta_{c,1})}\Big), \tag{4.8}$$
$$q^{-1}\|\widehat{\mathbf{Q}}_c - \mathbf{Q}_c\mathbf{H}_c'\|_F^2 = O_P\Big(T^{-1}q^{1-2\delta_{c,k_c}}p^{1-2\delta_{r,1}} + q^{-2\delta_{c,k_c}}p^{2(1-\delta_{r,1})}\Big). \tag{4.9}$$

From Theorem 2, the consistency for the loading matrix estimators requires Assumption (R1). If all factors are pervasive, the (squared) convergence rates for the row (resp. column) loading matrix will be $\max(1/(Tpq), 1/p^2)$ (resp. $\max(1/(Tpq), 1/q^2)$), which are consistent with those in Chen and Fan (2023) after the same normalization of the loading matrices.

Theorem 3. *Under the assumptions in Theorem 2, we have the following:*

1. *The error of the estimated factor series has rate*

$$\|\widehat{\mathbf{F}}_{Z,t} - (\mathbf{H}_r^{-1})'\mathbf{F}_{Z,t}\mathbf{H}_c^{-1}\|_F^2 = O_P\big(p^{1-\delta_{r,k_r}} q^{1-\delta_{c,k_c}} + T^{-1} p^{1+2\delta_{r,1}-2\delta_{r,k_r}} q^{1-\delta_{c,1}} + p^{1+\delta_{r,1}-3\delta_{r,k_r}} q^{2-\delta_{c,1}} + T^{-1} q^{1+2\delta_{c,1}-2\delta_{c,k_c}} p^{1-\delta_{r,1}} + q^{1+\delta_{c,1}-3\delta_{c,k_c}} p^{2-\delta_{r,1}}\big).$$

2. *For any $t \in [T], i \in [p], j \in [q]$, the squared error of the estimated individual common component is*

$$(\widehat{C}_{t,ij} - C_{t,ij})^2 = O_P\big(p^{1-2\delta_{r,k_r}} q^{1-2\delta_{c,k_c}} + T^{-1} p^{1+2\delta_{r,1}-3\delta_{r,k_r}} q^{1-\delta_{c,1}-\delta_{c,k_c}} + p^{1+\delta_{r,1}-4\delta_{r,k_r}} q^{2-\delta_{c,1}-\delta_{c,k_c}} + T^{-1} q^{1+2\delta_{c,1}-3\delta_{c,k_c}} p^{1-\delta_{r,1}-\delta_{r,k_r}} + q^{1+\delta_{c,1}-4\delta_{c,k_c}} p^{2-\delta_{r,1}-\delta_{r,k_r}}\big).$$

We state the above results separating from Theorem 2 since they have used some arguments from the proof of Theorem 5. If all factors are pervasive, it is clear that individual common components are consistent with rate $(pq)^{-1/2}+T^{-1/2}(q^{-1/2}+p^{-1/2})+p^{-1}+q^{-1} = \max(1/(Tq)^{1/2}, 1/(Tp)^{1/2}, 1/p, 1/q)$. This rate coincides with Theorem 4 of Chen and Fan (2023) for instance.

4.4. Asymptotic normality of estimators

We present the asymptotic normality of various estimators in this section, together with the estimation of the corresponding covariance matrices for practical inferences. But before that, we need three more assumptions.

(L2) (Eigenvalues) *The eigenvalues of the $k_r \times k_r$ matrix $\boldsymbol{\Sigma}_{A,r}\mathbf{Z}_r$ from Assumption (L1) are distinct, and so are those of the $k_c \times k_c$ matrix $\boldsymbol{\Sigma}_{A,c}\mathbf{Z}_c$.*

(AD1) *There is a constant C such that for any $k \in [K], j \in [d_k]$, as $p, q, T \to \infty$,*

$$\sqrt{\frac{1}{Tqp^{\delta_{r,1}}}} \cdot \mathbb{E}\Big\{\Big\|\mathbf{H}_r^* \sum_{i=1}^{p} \mathbf{Q}_{r,i\cdot} \sum_{t=1}^{T} (\mathbf{C}_t\mathbf{E}_t')_{ij}\Big\|\Big\} \geq C > 0,$$
$$\sqrt{\frac{1}{Tpq^{\delta_{c,1}}}} \cdot \mathbb{E}\Big\{\Big\|\mathbf{H}_c^* \sum_{i=1}^{q} \mathbf{Q}_{c,i\cdot} \sum_{t=1}^{T} (\mathbf{C}_t'\mathbf{E}_t)_{ij}\Big\|\Big\} \geq C > 0,$$

where $\mathbf{H}_r^* := \mathrm{tr}(\mathbf{A}_c'\mathbf{A}_c)^{1/2} \cdot \mathbf{D}_r^{-1/2}(\boldsymbol{\Gamma}_r^*)'\mathbf{Z}_r^{1/2}$ with $\mathbf{D}_r := \mathrm{tr}(\mathbf{A}_c'\mathbf{A}_c)\mathrm{diag}\{\lambda_1(\mathbf{A}_r'\mathbf{A}_r), \dots, \lambda_{k_r}(\mathbf{A}_r'\mathbf{A}_r)\}$, and $\boldsymbol{\Gamma}_r^*$ is the eigenvector matrix of $\mathrm{tr}(\mathbf{A}_c'\mathbf{A}_c) \cdot p^{-\delta_{r,k_r}} q^{-\delta_{c,1}} \mathbf{Z}_r^{1/2}\boldsymbol{\Sigma}_{A,r}\mathbf{Z}_r^{1/2}$. Similarly, we have $\mathbf{H}_c^* := \mathrm{tr}(\mathbf{A}_r'\mathbf{A}_r)^{1/2} \cdot \mathbf{D}_c^{-1/2}(\boldsymbol{\Gamma}_c^*)'\mathbf{Z}_c^{1/2}$, with $\mathbf{D}_c := \mathrm{tr}(\mathbf{A}_r'\mathbf{A}_r)\mathrm{diag}\{\lambda_1(\mathbf{A}_c'\mathbf{A}_c), \dots, \lambda_{k_c}(\mathbf{A}_c'\mathbf{A}_c)\}$, and $\boldsymbol{\Gamma}_c^*$ is the eigenvector matrix of $\mathrm{tr}(\mathbf{A}_r'\mathbf{A}_r) \cdot q^{-\delta_{c,k_c}} p^{-\delta_{r,1}} \mathbf{Z}_c^{1/2}\boldsymbol{\Sigma}_{A,c}\mathbf{Z}_c^{1/2}$.

(R2) We have

$$T^{-1}p^{1+2\delta_{r,1}-3\delta_{r,k_r}} q^{1-\delta_{c,1}-\delta_{c,k_c}}, \quad p^{1+\delta_{r,1}-4\delta_{r,k_r}} q^{2-\delta_{c,1}-\delta_{c,k_c}},$$
$$T^{-1}q^{1+2\delta_{c,1}-3\delta_{c,k_c}} p^{1-\delta_{r,1}-\delta_{r,k_r}}, \quad q^{1+\delta_{c,1}-4\delta_{c,k_c}} p^{2-\delta_{r,1}-\delta_{r,k_r}} = o(1).$$

Assumption (AD1) appears in Cen and Lam (2025) as well. This assumption facilitates the proof of the asymptotic normality of each row of $\widehat{\mathbf{Q}}_r$ and $\widehat{\mathbf{Q}}_c$, by asserting that in the decomposition of $\widehat{\mathbf{Q}}_r - \mathbf{Q}_r\mathbf{H}_r$ (resp. $\widehat{\mathbf{Q}}_c - \mathbf{Q}_c\mathbf{H}_c$), certain terms are dominating others even in the lower bound, and hence is truly dominating rather than just having the upper bounds dominating other upper bounds as in the proofs of similar theorems in the broader literature of factor models. Assumption (R2) is needed to make sure that the estimated common component $\widehat{\mathbf{C}}_t$ is consistent element-wise (see Theorem 3). This is satisfied automatically when all factors are pervasive, for instance.

Let $\Sigma_{\epsilon,ij}$ be the (i,j) entry of $\boldsymbol{\Sigma}_\epsilon$ in Assumption (E1).

Theorem 4. *Let all assumptions in Theorem 2 hold. Assume also for $i \in [p]$ and $j \in [q]$,*

$$\gamma_{\alpha,i}^2 := \lim_{q\to\infty} \frac{1}{q}\sum_{j=1}^{q} \Sigma_{\epsilon,ij}^2, \quad \gamma_{\beta,j}^2 := \lim_{p\to\infty} \frac{1}{p}\sum_{i=1}^{p} \Sigma_{\epsilon,ij}^2, \quad \gamma_\mu^2 := \lim_{p,q\to\infty} \frac{1}{pq} \sum_{i\in[p],j\in[q]} \Sigma_{\epsilon,ij}^2.$$

Then for each $t \in [T]$,

$$\sqrt{pq}(\hat{\mu}_t - \mu_t) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \gamma_\mu^2).$$

Take a finite integer m, and integers $i_1 < i_2 < \cdots < i_m$ with $i_\ell \in [p]$. Define $\theta_{\alpha,t} := (\alpha_{t,i_1}, \dots, \alpha_{t,i_m})'$ and similarly for $\hat{\theta}_{\alpha,t}$, where $\alpha_{t,i}$ is the ith element of $\boldsymbol{\alpha}_t$. Then for a fixed $t \in [T]$,

$$\sqrt{q}(\hat{\theta}_{\alpha,t} - \theta_{\alpha,t}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, diag(\gamma_{\alpha,i_1}^2, \dots, \gamma_{\alpha,i_m}^2)).$$

Similarly, take integers $j_1 < \cdots < j_m$ where $j_\ell \in [q]$. Define $\theta_{\beta,t} := (\beta_{t,j_1}, \dots, \beta_{t,j_m})'$ and similarly for $\widehat{\theta}_{\beta,t}$, where $\beta_{t,j}$ is the jth element of $\boldsymbol{\beta}_t$. Then for a fixed $t \in [T]$,

$$\sqrt{p}(\widehat{\theta}_{\beta,t} - \theta_{\beta,t}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, diag(\gamma_{\beta,j_1}^2, \dots, \gamma_{\beta,j_m}^2)).$$

Moreover, for $i \in [p]$ and $j \in [q]$, if the rate for $\hat{C}_{t,ij} - C_{t,ij}$ in Theorem 3 is $o(1)$, then

$$\hat{\gamma}^2_{\alpha,i} := q^{-1}(\hat{\mathbf{E}}_t\hat{\mathbf{E}}_t')_{ii}, \quad \hat{\gamma}^2_{\beta,j} := p^{-1}(\hat{\mathbf{E}}_t'\hat{\mathbf{E}}_t)_{jj}, \quad \hat{\gamma}^2_{\mu} := p^{-1}\sum_{i=1}^p \hat{\gamma}^2_{\alpha,i} = q^{-1}\sum_{j=1}^q \hat{\gamma}^2_{\beta,j}$$

are consistent estimators for $\gamma^2_{\alpha,i}$, $\gamma^2_{\beta,j}$ and γ^2_μ respectively under Assumption (R2), so that

$$\sqrt{pq}\,\hat{\gamma}_\mu^{-1}(\hat{\mu}_t - \mu_t) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$
$$\sqrt{q}\,\mathrm{diag}(\hat{\gamma}^{-1}_{\alpha,i_1},\ldots,\hat{\gamma}^{-1}_{\alpha,i_m})(\hat{\boldsymbol{\theta}}_{\alpha,t} - \boldsymbol{\theta}_{\alpha,t}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\mathbf{I}_m),$$
$$\sqrt{p}\,\mathrm{diag}(\hat{\gamma}^{-1}_{\beta,j_1},\ldots,\hat{\gamma}^{-1}_{\beta,j_m})(\hat{\boldsymbol{\theta}}_{\beta,t} - \boldsymbol{\theta}_{\beta,t}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\mathbf{I}_m).$$

Recall from Remark 1 that FM can be expressed in MEFM, and hence the ability to make inferences on the elements of $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ does not facilitate a test for the necessity of MEFM over FM. For such a test, please see Section 4.5. Theorem 4 gives us the ability to infer on the level of row and column main effects at each time point, which is important if we have target comparisons we want to make for these effects. For instance, if each row represents a country, we can easily compare the main effects at time t for the first country against the average of the second and third simply by considering $\mathbf{g} := (1,-1/2,-1/2)'$, $\boldsymbol{\theta}_{\alpha,t} := (\alpha_{t,1},\alpha_{t,2},\alpha_{t,3})'$ and using Theorem 4 to arrive at

$$\sqrt{q}\,(\mathbf{g}'\mathrm{diag}(\hat{\gamma}^2_{\alpha,1},\hat{\gamma}^2_{\alpha,2},\hat{\gamma}^2_{\alpha,3})\mathbf{g})^{-1/2}\mathbf{g}'(\hat{\boldsymbol{\theta}}_{\alpha,t} - \boldsymbol{\theta}_{\alpha,t}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Theorem 5. *Let all the assumptions under Theorem 2 hold, in addition to (AD1) and (L2). Suppose k_r and k_c are fixed and $p,q,T\to\infty$. If $Tq = o(p^{\delta_{r,1}+\delta_{r,k_r}})$, we have*

$$(Tp^{2\delta_{r,k_r}-\delta_{r,1}}q^{2\delta_{c,1}-1})^{1/2}\cdot(\hat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r\mathbf{Q}_{r,j\cdot}) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, T^{-1}p^{2\delta_{r,k_r}-\delta_{r,1}}q^{2\delta_{c,1}-1}\cdot \mathbf{D}_r^{-1}\mathbf{H}_r^*\boldsymbol{\Xi}_{r,j}(\mathbf{H}_r^*)'\mathbf{D}_r^{-1}\big),$$
$$\text{where}\quad \boldsymbol{\Xi}_{r,j} := \operatorname*{plim}_{p,q,T\to\infty}\mathrm{Var}\Big(\sum_{i=1}^p \mathbf{Q}_{r,i\cdot}\sum_{t=1}^T(\mathbf{C}_t\mathbf{E}_t')_{ij}\Big).$$

On the other hand, if $Tp = o(q^{\delta_{c,1}+\delta_{c,k_c}})$, we have

$$(Tq^{2\delta_{c,k_c}-\delta_{c,1}}p^{2\delta_{r,1}-1})^{1/2}\cdot(\hat{\mathbf{Q}}_{c,j\cdot} - \mathbf{H}_c\mathbf{Q}_{c,j\cdot}) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, T^{-1}q^{2\delta_{c,k_c}-\delta_{c,1}}p^{2\delta_{r,1}-1}\cdot \mathbf{D}_c^{-1}\mathbf{H}_c^*\boldsymbol{\Xi}_{c,j}(\mathbf{H}_c^*)'\mathbf{D}_c^{-1}\big),$$
$$\text{where}\quad \boldsymbol{\Xi}_{c,j} := \operatorname*{plim}_{p,q,T\to\infty}\mathrm{Var}\Big(\sum_{i=1}^q \mathbf{Q}_{c,i\cdot}\sum_{t=1}^T(\mathbf{C}_t'\mathbf{E}_t)_{ij}\Big).$$

Theorem 5 is essentially Theorem 3 of Cen and Lam (2025) when $K = 2$ and $\eta = 0$ (no missing values), having the same rate of convergence under potentially weak factors. Hence our MEFM estimation procedure has successfully estimated and removed all time-varying main effects and grand mean, leaving the estimation of the common component exactly the same as in FM. The proof (in the supplementary materials) revolves around the decompositions of $(\hat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r\mathbf{Q}_{r,j\cdot})$ and $(\hat{\mathbf{Q}}_{c,j\cdot} - \mathbf{H}_c\mathbf{Q}_{c,j\cdot})$ into sums so that we can use a version of central limit theorem for α-mixing sequences, which is Theorem 2.21 in Fan and Yao (2003).

4.4.1. Estimation of the asymptotic covariance matrix for factor loading estimators

To practically use Theorem 5 for inference, we need to estimate the covariance matrices for $\hat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r\mathbf{Q}_{r,j\cdot}$ and $\hat{\mathbf{Q}}_{c,j\cdot} - \mathbf{H}_c\mathbf{Q}_{c,j\cdot}$. We use the heteroscedasticity and autocorrelation consistent (HAC) estimators (Newey and West, 1987) based on $\{\hat{\mathbf{D}}_r,\hat{\mathbf{Q}}_r,\hat{\mathbf{C}}_t,\hat{\mathbf{E}}_t\}_{t\in[T]}$ and $\{\hat{\mathbf{D}}_c,\hat{\mathbf{Q}}_c,\hat{\mathbf{C}}_t,\hat{\mathbf{E}}_t\}_{t\in[T]}$, respectively.

For $\hat{\mathbf{Q}}_{r,j\cdot} - \mathbf{H}_r\mathbf{Q}_{r,j\cdot}$, with η_r such that $\eta_r\to\infty$ and $\eta_r/(Tp^{2\delta_{r,k_r}-\delta_{r,1}}q^{2\delta_{c,1}-1})^{1/4}\to 0$, define an HAC estimator

$$\hat{\boldsymbol{\Sigma}}^{HAC}_{r,j} := \mathbf{D}_{r,0,j} + \sum_{\nu=1}^{\eta_r}\Big(1 - \frac{\nu}{1+\eta_r}\Big)\Big(\mathbf{D}_{r,\nu,j} + \mathbf{D}_{r,\nu,j}'\Big), \quad \text{where}$$
$$\mathbf{D}_{r,\nu,j} := \sum_{t=1+\nu}^T\Big\{\sum_{i=1}^p\Big(T^{-1}\hat{\mathbf{D}}_r^{-1}\hat{\mathbf{Q}}_r'\sum_{s=1}^T\hat{\mathbf{C}}_s\hat{\mathbf{C}}_{s,i\cdot}\Big)(\hat{\mathbf{C}}_t\hat{\mathbf{E}}_t')_{ij}\Big\}\Big\{\sum_{i=1}^p\Big(T^{-1}\hat{\mathbf{D}}_r^{-1}\hat{\mathbf{Q}}_r'\sum_{s=1}^T\hat{\mathbf{C}}_s\hat{\mathbf{C}}_{s,i\cdot}\Big)(\hat{\mathbf{C}}_{t-\nu}\hat{\mathbf{E}}_{t-\nu}')_{ij}\Big\}'.$$

Similarly for $\hat{\mathbf{Q}}_{c,j\cdot} - \mathbf{H}_c\mathbf{Q}_{c,j\cdot}$, with η_c such that $\eta_c\to\infty$ and $\eta_c/(Tq^{2\delta_{c,k_c}-\delta_{c,1}}p^{2\delta_{r,1}-1})^{1/4}\to 0$, define

$$\hat{\boldsymbol{\Sigma}}^{HAC}_{c,j} := \mathbf{D}_{c,0,j} + \sum_{\nu=1}^{\eta_c}\Big(1 - \frac{\nu}{1+\eta_c}\Big)\Big(\mathbf{D}_{c,\nu,j} + \mathbf{D}_{c,\nu,j}'\Big), \quad \text{where}$$
$$\mathbf{D}_{c,\nu,j} := \sum_{t=1+\nu}^T\Big\{\sum_{i=1}^q\Big(T^{-1}\hat{\mathbf{D}}_c^{-1}\hat{\mathbf{Q}}_c'\sum_{s=1}^T\hat{\mathbf{C}}_s'\hat{\mathbf{C}}_{s,\cdot i}\Big)(\hat{\mathbf{C}}_t'\hat{\mathbf{E}}_t)_{ij}\Big\}\Big\{\sum_{i=1}^q\Big(T^{-1}\hat{\mathbf{D}}_c^{-1}\hat{\mathbf{Q}}_c'\sum_{s=1}^T\hat{\mathbf{C}}_s'\hat{\mathbf{C}}_{s,\cdot i}\Big)(\hat{\mathbf{C}}_{t-\nu}'\hat{\mathbf{E}}_{t-\nu})_{ij}\Big\}'.$$

Theorem 6. *Let all the assumptions under Theorem 2 hold, in addition to (L2), (AD1) and (R2). Suppose k_r and k_c are fixed and $p,q,T\to\infty$. If $Tq = o(p^{\delta_{r,1}+\delta_{r,k_r}})$, then*

1. $\hat{\mathbf{D}}_r^{-1}\hat{\boldsymbol{\Sigma}}_{r,j}^{HAC}\hat{\mathbf{D}}_r^{-1}$ *is consistent for* $\mathbf{D}_r^{-1}\mathbf{H}_r^*\Xi_{r,j}(\mathbf{H}_r^*)'\mathbf{D}_r^{-1}$;
2. $T\cdot\big(\hat{\boldsymbol{\Sigma}}_{r,j}^{HAC}\big)^{-1/2}\hat{\mathbf{D}}_r(\hat{\mathbf{Q}}_{r,j\cdot}-\mathbf{H}_r\mathbf{Q}_{r,j\cdot})\xrightarrow{\mathcal{D}}\mathcal{N}(0,\mathbf{I}_{k_r})$.

On the other hand, if $Tp=o(q^{\delta_{c,1}+\delta_{c,k_c}})$, *then*

3. $\hat{\mathbf{D}}_c^{-1}\hat{\boldsymbol{\Sigma}}_{c,j}^{HAC}\hat{\mathbf{D}}_c^{-1}$ *is consistent for* $\mathbf{D}_c^{-1}\mathbf{H}_c^*\Xi_{c,j}(\mathbf{H}_c^*)'\mathbf{D}_c^{-1}$;
4. $T\cdot\big(\hat{\boldsymbol{\Sigma}}_{c,j}^{HAC}\big)^{-1/2}\hat{\mathbf{D}}_c(\hat{\mathbf{Q}}_{c,j\cdot}-\mathbf{H}_c\mathbf{Q}_{c,j\cdot})\xrightarrow{\mathcal{D}}\mathcal{N}(0,\mathbf{I}_{k_c})$.

Remark 4. *In this remark, we discuss the computational details of the asymptotic normality depicted in Theorem 4 and Theorem 5, respectively. In Theorem 4, it is straightforward in computing each* $\hat{\gamma}_{\alpha,i}^2$, $\hat{\gamma}_{\beta,j}^2$ *and* $\hat{\gamma}_{\mu}^2$ *from their definitions, and hence to construct the covariance matrices for* $\hat{\theta}_{\alpha,t}$, $\hat{\theta}_{\alpha,t}$ *and* $\hat{\mu}_t$. *Given* $\hat{\mathbf{E}}_t$, *the computational complexity is* $O(p^2q)$ *for* $\hat{\gamma}_{\alpha,i}^2$ *and* $O(q^2p)$ *for* $\hat{\gamma}_{\beta,i}^2$; *for* $\hat{\gamma}_{\mu}^2$, *we should optimize and compute* $\hat{\gamma}_{\mu}^2=p^{-1}\sum_{i=1}^p\hat{\gamma}_{\alpha,i}^2$ *if* $p<q$ *so that the computational complexity is* $O(p^3q)$, *and otherwise compute* $\hat{\gamma}_{\mu}^2=q^{-1}\sum_{j=1}^q\hat{\gamma}_{\beta,j}^2$ *with run time of order* $O(q^3p)$.

For the covariance matrix in Theorem 5, we may apply Theorem 6 and it remains to compute $\hat{\boldsymbol{\Sigma}}_{r,j}^{HAC}$ *and* $\hat{\boldsymbol{\Sigma}}_{c,j}^{HAC}$. *To specify* η_r *and* η_c *in practice, by the rate requirement, we use* $\eta_c,\eta_r=\lfloor c(Tpq)^{1/4}\rfloor$ *for some constant* $c\in(0,1)$ *and we suggest* $c=1/5$ *which is used in our numerical results later, and works well in various settings. Lastly, given* η_r, $\hat{\mathbf{D}}_r$, $\hat{\mathbf{Q}}_r$, $\hat{\mathbf{C}}_t$ *and* $\hat{\mathbf{E}}_t$, *the computational complexity to obtain* $\hat{\boldsymbol{\Sigma}}_{r,j}^{HAC}$ *is* $O(\eta_rTp^2q)$ *by computing and storing* $\sum_{s=1}^T\hat{\mathbf{C}}_s\hat{\mathbf{C}}_s'$; *similarly, it takes* $O(\eta_cTq^2p)$ *operations to compute* $\hat{\boldsymbol{\Sigma}}_{c,j}^{HAC}$.

4.5. *Testing the sufficiency of FM versus MEFM*

In the last section, we introduce how to make inferences on various parameters of MEFM. However, to test if FM is sufficient against our proposed MEFM, simple inferences on the model parameters are not enough in the face of Remark 1. Formally, we want to test, for the time horizon $t\in[T]$,

$$H_0:\text{FM is sufficient over } t\in[T]\quad\longleftrightarrow\quad H_1:\text{MEFM is needed over } t\in[T].$$

The above problem is complicated by the fact that, in Section 3.1, we have seen that MEFM can always be expressed as FM if we are willing to potentially consider a large number of factors. So, how "large" an increase in the number of factors do we consider unacceptable?

Remark 1 tells us that a special form of MEFM can be expressed back in FM:

$$\mathbf{Y}_t=\mu_t\mathbf{1}_p\mathbf{1}_q'+\boldsymbol{\alpha}_t\mathbf{1}_q'+\mathbf{1}_p\boldsymbol{\beta}_t'+\mathbf{M}_p\acute{\mathbf{C}}_t\mathbf{M}_q+\mathbf{E}_t=\mathbf{A}_r\mathbf{F}_t\mathbf{A}_c'+\mathbf{E}_t,\quad t\in[T],$$

where $\acute{\mathbf{C}}_t:=\mathbf{A}_r\mathbf{F}_t\mathbf{A}_c'$ and

$$\mu_t:=(pq)^{-1}\mathbf{1}_p'\acute{\mathbf{C}}_t\mathbf{1}_q,\quad \boldsymbol{\alpha}_t:=q^{-1}\mathbf{M}_p\acute{\mathbf{C}}_t\mathbf{1}_q,\quad \boldsymbol{\beta}_t:=p^{-1}\mathbf{M}_q\acute{\mathbf{C}}_t'\mathbf{1}_p.$$

If $\mathbf{A}_r$ has rank k_r satisfying Assumption (L1) and $\mathbf{A}_c$ has rank k_c, the potential rank of $\mathbf{M}_p\mathbf{A}_r$ is k_r-1 (when a column in $\mathbf{A}_r$ is parallel to $\mathbf{1}_p$), and that of $\mathbf{M}_q\mathbf{A}_c$ is k_c-1 (when a column in $\mathbf{A}_c$ is parallel to $\mathbf{1}_q$), demonstrating that FM can have an increase in the number of factors, albeit still finite.

Another special example is when both $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ are zero, but $\mu_t\neq 0$. Then we can write MEFM as

$$\mathbf{Y}_t=\mu_t\mathbf{1}_p\mathbf{1}_q'+\mathbf{A}_r\mathbf{F}_t\mathbf{A}_c'+\mathbf{E}_t=(\mathbf{A}_r,\mathbf{1}_p)\begin{pmatrix}\mathbf{F}_t & 0\\ \mathbf{0}' & \mu_t/(pq)\end{pmatrix}\begin{pmatrix}\mathbf{A}_c'\\ \mathbf{1}_q'\end{pmatrix}+\mathbf{E}_t,$$

which is FM with loading matrices $(\mathbf{A}_r,\mathbf{1}_p)$ and $(\mathbf{A}_c,\mathbf{1}_q)$ respectively, and an increase by 1 for both the number of row and column factors.

In light of the above examples, we deem FM sufficient if and only if the number of factors in the FM is still finite and any model variables satisfy the Assumptions in Section 4.1.

To be able to test H_0 against H_1, define $\check{\mathbf{E}}_t$ to be the residual matrix after a fitting of FM (a similar procedure to fitting MEFM but treating μ_t, $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ as zero), with

$$\check{\mathbf{E}}_t:=\mathbf{Y}_t-\check{\mathbf{C}}_t,\quad\text{where}\quad \check{\mathbf{C}}_t:=\check{\mathbf{A}}_r\check{\mathbf{A}}_r'\mathbf{Y}_t\check{\mathbf{A}}_c\check{\mathbf{A}}_c',$$

with $\check{\mathbf{A}}_r$ and $\check{\mathbf{A}}_c$ the $p\times\ell_r$ and $q\times\ell_c$ eigenmatrices of $\sum_{t=1}^T\mathbf{Y}_t\mathbf{Y}_t'$ and $\sum_{t=1}^T\mathbf{Y}_t'\mathbf{Y}_t$ respectively.

Theorem 7. *Let all assumptions in Theorem 2 hold, on top of (R2). Also assume that* $\hat{C}_{t,ij}-C_{t,ij}=o_P(\min(p^{-1/2},q^{-1/2}))$ *in Theorem 3. Suppose* k_r,k_c,ℓ_r *and* ℓ_c *are all fixed and known. Then under* H_0, *for each* $t\in[T]$, *we have*

$$\frac{(\hat{\mathbf{E}}_t\hat{\mathbf{E}}_t')_{ii}-\sum_{j=1}^q\Sigma_{\epsilon,ij}^2}{\sqrt{\sum_{j=1}^q\mathrm{Var}(\epsilon_{t,ij}^2)\Sigma_{\epsilon,ij}^4}},\ \frac{(\check{\mathbf{E}}_t\check{\mathbf{E}}_t')_{ii}-\sum_{j=1}^q\Sigma_{\epsilon,ij}^2}{\sqrt{\sum_{j=1}^q\mathrm{Var}(\epsilon_{t,ij}^2)\Sigma_{\epsilon,ij}^4}}\xrightarrow{\mathcal{D}}Z_{i,t}\xrightarrow{\mathcal{D}}\mathcal{N}(0,1)\ \ \textit{for each}\ \ i\in[p];$$

$$\frac{(\hat{\mathbf{E}}_t'\hat{\mathbf{E}}_t)_{jj}-\sum_{i=1}^p\Sigma_{\epsilon,ij}^2}{\sqrt{\sum_{i=1}^p\mathrm{Var}(\epsilon_{t,ij}^2)\Sigma_{\epsilon,ij}^4}},\ \frac{(\check{\mathbf{E}}_t'\check{\mathbf{E}}_t)_{jj}-\sum_{i=1}^p\Sigma_{\epsilon,ij}^2}{\sqrt{\sum_{i=1}^p\mathrm{Var}(\epsilon_{t,ij}^2)\Sigma_{\epsilon,ij}^4}}\xrightarrow{\mathcal{D}}W_{j,t}\xrightarrow{\mathcal{D}}\mathcal{N}(0,1)\ \ \textit{for each}\ \ j\in[q],$$

where $Z_{h,t}$ is independent of $Z_{\ell,t}$ and $W_{h,t}$ is independent of $W_{\ell,t}$ for $h \neq \ell$. The same asymptotic results hold true under H_1 for $(\widehat{\mathbf{E}}_t\widehat{\mathbf{E}}_t')_{ii}$ and $(\widehat{\mathbf{E}}_t'\widehat{\mathbf{E}}_t)_{jj}$ respectively for $i \in [p], j \in [q]$.

The assumption $\widehat{C}_{t,ij} - C_{t,ij} = o_P(\min(p^{-1/2}, q^{-1/2}))$ is satisfied, for instance, when all factors are pervasive and T, p, q are of the same order. Theorem 7 tells us that for each $t \in [T]$, both

$$x_{\alpha,t} := \max_{i\in[p]} \widehat{\gamma}^2_{\alpha,i} = \max_{i\in[p]}\{q^{-1}(\widehat{\mathbf{E}}_t\widehat{\mathbf{E}}_t')_{ii}\}, \quad y_{\alpha,t} := \max_{i\in[p]} \check{\gamma}^2_{\alpha,i} := \max_{i\in[p]}\{q^{-1}(\check{\mathbf{E}}_t\check{\mathbf{E}}_t')_{ii}\} \tag{4.10}$$

are distributed approximately the same for large q under H_0, and $x_{\alpha,t}$ in particular is distributed the same no matter under H_0 or H_1. Similarly, define

$$x_{\beta,t} := \max_{j\in[q]} \widehat{\gamma}^2_{\beta,j} = \max_{j\in[q]}\{p^{-1}(\widehat{\mathbf{E}}_t'\widehat{\mathbf{E}}_t)_{jj}\}, \quad y_{\beta,t} := \max_{j\in[q]} \check{\gamma}^2_{\beta,j} := \max_{j\in[q]}\{p^{-1}(\check{\mathbf{E}}_t'\check{\mathbf{E}}_t)_{jj}\}, \tag{4.11}$$

which are distributed approximately the same for large p under H_0 from Theorem 7, and $x_{\beta,t}$ in particular is distributed the same no matter under H_0 or H_1. To utilize Theorem 7 in testing H_0, we impose an additional assumption on the core factor and idiosyncratic noise as follows.

(E3) (Tail condition in $\mathbf{F}_t$ and $\mathbf{E}_t$) *Each element in the time series $\{\mathbf{X}_{f,t}\}$, $\{\mathbf{X}_{e,t}\}$ and $\{\mathbf{X}_{\epsilon,t}\}$ has sub-Gaussian tail.*

This assumption allows us to make convergence statements in quantiles to be defined in Theorem 8 below.

Define $\mathbb{F}_{x,\alpha}$, $\mathbb{F}_{y,\alpha}$, $\mathbb{F}_{x,\beta}$ and $\mathbb{F}_{y,\beta}$ the empirical cumulative distribution functions for $\{x_{\alpha,t}\}_{t\in[T]}$, $\{y_{\alpha,t}\}_{t\in[T]}$, $\{x_{\beta,t}\}_{t\in[T]}$ and $\{y_{\beta,t}\}_{t\in[T]}$ respectively:

$$\begin{aligned}
\mathbb{F}_{x,\alpha}(c) &:= \frac{1}{T}\sum_{t=1}^T \mathbb{1}\{x_{\alpha,t} \le c\}, \quad \mathbb{F}_{y,\alpha}(c) := \frac{1}{T}\sum_{t=1}^T \mathbb{1}\{y_{\alpha,t} \le c\},\\
\mathbb{F}_{x,\beta}(c) &:= \frac{1}{T}\sum_{t=1}^T \mathbb{1}\{x_{\beta,t} \le c\}, \quad \mathbb{F}_{y,\beta}(c) := \frac{1}{T}\sum_{t=1}^T \mathbb{1}\{y_{\beta,t} \le c\}.
\end{aligned} \tag{4.12}$$

Theorem 8. *Let Assumption (E3) and all the assumptions in Theorem 7 hold. Moreover, we assume for simplicity of presentation that all factors are pervasive. Define for $0 < \theta < 1$,*

$$\widehat{q}_{x,\alpha}(\theta) := \inf\{c \mid \mathbb{F}_{x,\alpha}(c) \ge \theta\}, \quad \widehat{q}_{x,\beta}(\theta) := \inf\{c \mid \mathbb{F}_{x,\beta}(c) \ge \theta\},$$

Then under H_0, as $T, p, q \to \infty$, we have for each $t \in [T]$,

$$\mathbb{P}_{y,\alpha}[y_{\alpha,t} > \widehat{q}_{x,\alpha}(\theta)] \le 1 - \theta + O_P\Big\{\Big(\frac{1}{\sqrt{p}} + \frac{1}{\sqrt{q}} + \frac{1}{\sqrt{T}} + \frac{\sqrt{q}}{p} + \sqrt{\frac{q}{Tp}}\Big)\log^2(T)\log(p)\log^2(q)\Big\},$$

$$\mathbb{P}_{y,\beta}[y_{\beta,t} > \widehat{q}_{x,\beta}(\theta)] \le 1 - \theta + O_P\Big\{\Big(\frac{1}{\sqrt{p}} + \frac{1}{\sqrt{q}} + \frac{1}{\sqrt{T}} + \frac{\sqrt{p}}{q} + \sqrt{\frac{p}{Tq}}\Big)\log^2(T)\log^2(p)\log(q)\Big\},$$

where $\mathbb{P}_{y,\alpha}$ and $\mathbb{P}_{y,\beta}$ are empirical probability measures induced by $\mathbb{F}_{y,\alpha}$ and $\mathbb{F}_{y,\beta}$ respectively.

The assumption of pervasive factors is for the ease of presentation of the rate added to the two probability statements above. But if some factors are weaker, then the convergence rate of the common components will be adversely affected, and the rate in the probability statements above will be inflated.

With Theorem 8, we can test H_0 at significance level $1-\theta$ asymptotically using the test statistics $y_{\alpha,t}$ and $y_{\beta,t}$, and rejection rules $y_{\alpha,t} \ge \widehat{q}_{x,\alpha}(\theta)$ and $y_{\beta,t} \ge \widehat{q}_{x,\beta}(\theta)$ respectively. Since we have $y_{\alpha,t}$ and $y_{\beta,t}$ for $t \in [T]$, we can assess the significance level under H_0 by calculating

$$\text{Significance levels} = T^{-1}\sum_{t=1}^T \mathbb{1}\{y_{\alpha,t} \ge \widehat{q}_{x,\alpha}(\theta)\}, \quad T^{-1}\sum_{t=1}^T \mathbb{1}\{y_{\beta,t} \ge \widehat{q}_{x,\beta}(\theta)\},$$

and see if they are close to $1-\theta$. If H_0 is not true, then if any $\alpha_{t,i}$ is large, we expect $y_{\alpha,t}$ to be large. Or, if any $\beta_{t,j}$ is large, we expect $y_{\beta,t}$ to be large.

In practice for testing H_0 against H_1, we estimate k_r and k_c, and set $\ell_r = k_r + 1$ and $\ell_c = k_c + 1$ in light of the previous argument on how a special form of MEFM can be expressed back in FM. For estimation of k_r and k_c, see Section 4.6.

4.6. Estimation of the number of factors

From (3.4), we have $T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'$ essentially the row sample covariance matrix and $T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t'\widehat{\mathbf{L}}_t$ the column sample covariance matrix. We then propose the eigenvalue-ratio estimators for the number of factors as

$$\widehat{k}_r := \arg\min_j \Big\{ \frac{\lambda_{j+1}\big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\big) + \xi_r}{\lambda_j\big(T^{-1}\sum_{t=1}^T \widehat{\mathbf{L}}_t\widehat{\mathbf{L}}_t'\big) + \xi_r}, \ j \in [\lfloor p/2 \rfloor] \Big\}, \quad \xi_r \asymp pq\big[(Tq)^{-1/2} + p^{-1/2}\big], \tag{4.13}$$

$$\hat{k}_c := \arg\min_j \left\{ \frac{\lambda_{j+1}\big(T^{-1}\sum_{t=1}^T \hat{\mathbf{L}}_t'\hat{\mathbf{L}}_t\big)+\xi_c}{\lambda_j\big(T^{-1}\sum_{t=1}^T \hat{\mathbf{L}}_t'\hat{\mathbf{L}}_t\big)+\xi_c},\ j\in[\lfloor q/2\rfloor] \right\},\quad \xi_c \asymp pq\big[(Tp)^{-1/2}+q^{-1/2}\big]. \tag{4.14}$$

Ratio-based estimators are widely studied by researchers. For example, an eigenvalue-ratio estimator is considered in Lam and Yao (2012) and Ahn and Horenstein (2013), while a cumulative eigenvalue ratio estimator is proposed by Zhang et al. (2024). Our proposed estimator is similar to the perturbed eigenvalue-ratio estimators as in Pelger (2019). Technically, we can minimize (4.13) (resp. (4.14)) over any $j\in[p]$ (resp. $j\in[q]$), but it is very reasonable to assume $k_r\le p/2$ and $k_c\le q/2$ in all applications of factor models. The correction terms ξ_r and ξ_c are added to stabilize the ratio so that consistency follows from the theorem below.

Theorem 9. *Under Assumptions (IC1), (M1), (F1), (L1), (E1), (E2) and (R1), we have the following.*

1. $\hat{k}_r$ *is a consistent estimator of* k_r *if*

$$\begin{cases} p^{1-\delta_{r,k_r}}q^{1-\delta_{c,1}}[(Tq)^{-1/2}+p^{-1/2}] = o(p^{\delta_{r,j+1}-\delta_{r,j}}),\ j\in[k_r-1] \text{ with } k_r\ge 2;\\ p^{1-\delta_{r,1}}q^{1-\delta_{c,1}}[(Tq)^{-1/2}+p^{-1/2}] = o(1),\ k_r=1.\end{cases}$$

2. $\hat{k}_c$ *is a consistent estimator of* k_c *if*

$$\begin{cases} q^{1-\delta_{c,k_c}}p^{1-\delta_{r,1}}[(Tp)^{-1/2}+q^{-1/2}] = o(q^{\delta_{c,j+1}-\delta_{c,j}}),\ j\in[k_c-1] \text{ with } k_c\ge 2;\\ q^{1-\delta_{c,1}}p^{1-\delta_{r,1}}[(Tp)^{-1/2}+q^{-1/2}] = o(1),\ k_c=1.\end{cases}$$

The extra rate conditions in the theorem are due to existence of potential weak factors and are trivially satisfied for pervasive factors. The theorem is similar to the consistency result in Cen and Lam (2025) for matrix-valued factor models, and this implies that the number of factors in MEFM can be well estimated just as in the case of FM.

4.7. Estimating the number of factors for the row/column main effects

In order to see if model (3.1) can be written as model (1.3), essentially exploring how many global row/column common factors is contributing to $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ respectively, we develop in this section two eigenvalue-ratio estimators for estimating r and ℓ in model (1.3):

$$\hat{r} := \arg\min_j \left\{ \frac{\lambda_{j+1}(T^{-1}\sum_{t=1}^T \hat{\boldsymbol{\alpha}}_t\hat{\boldsymbol{\alpha}}_t')+q^{-1/2}}{\lambda_j(T^{-1}\sum_{t=1}^T \hat{\boldsymbol{\alpha}}_t\hat{\boldsymbol{\alpha}}_t')+q^{-1/2}},\ j\in[p-1] \right\}, \tag{4.15}$$

$$\hat{\ell} := \arg\min_j \left\{ \frac{\lambda_{i+1}(T^{-1}\sum_{t=1}^T \hat{\boldsymbol{\beta}}_t\hat{\boldsymbol{\beta}}_t')+p^{-1/2}}{\lambda_i(T^{-1}\sum_{t=1}^T \hat{\boldsymbol{\beta}}_t\hat{\boldsymbol{\beta}}_t')+p^{-1/2}},\ j\in[q-1] \right\}. \tag{4.16}$$

We first present some extra assumptions regarding $\{\mathbf{g}_t\}$ and $\{\mathbf{h}_t\}$ in model (1.3).

(ME1) (Time Series in $\mathbf{g}_t$ and $\mathbf{h}_t$) *There are* $\mathbf{x}_{g,t}\in\mathbb{R}^r$ *and* $\mathbf{x}_{h,t}\in\mathbb{R}^\ell$ *each with independent entries having mean 0 and variance 1 such that* $\mathbf{g}_t=\sum_{w\ge 0}a_{g,w}\mathbf{x}_{g,t-w}$ *and* $\mathbf{h}_t=\sum_{w\ge 0}a_{h,w}\mathbf{x}_{h,t-w}$, *with both* $\{\mathbf{g}_t\}$ *and* $\{\mathbf{h}_t\}$ *invertible. The time series* $\{\mathbf{x}_{g,t}\}$ *contains independent random vectors while being independent of* $\{\mathbf{E}_t\}$. *The same goes for* $\{\mathbf{x}_{h,t}\}$. The coefficients $\{a_{g,w}\}$ and $\{a_{h,w}\}$ satisfy $\sum_{w\ge 0}a_{g,w}^2=\sum_{w\ge 0}a_{h,w}^2=1$ and $\sum_{w\ge n}|a_{g,w}|,\ \sum_{w\ge n}|a_{h,w}|=O(n^{-\alpha})$ for some constant $\alpha>1$.

(ME2) In Assumption (E2), the sequence $\{a_{\epsilon,w}\}_{w\in\mathbb{N}\cup\{0\}}$ is such that $\sum_{w\ge 0}w|a_{\epsilon,w}|<\infty$. Also, if z_t is an element in $\mathbf{x}_{g,t}$ or $\mathbf{x}_{h,t}$,

One important note is that $\{\mathbf{g}_t\}$ and $\{\mathbf{h}_t\}$ are possibly correlated because $\{\mathbf{x}_{g,t}\}$ and $\{\mathbf{x}_{h,t}\}$ may not be independent of each other, although they are both independent of the noise series $\{\mathbf{E}_t\}$. This is important in light of FM being able to be expressed as MEFM in Remark 1. Assumption (ME2) makes sure that a random matrix theory can be used when p/T or q/T approach finite constants so as to bound the largest eigenvalues of certain noise related sample covariance matrices.

Theorem 10. *Let Assumptions (IC1), (M1), (F1), (E1), (E2), (ME1) and (ME2) hold. Suppose also* $r/T\to c_r\in(0,1)$ *(or* $r<\infty$*),* $\ell/T\to c_\ell\in[0,1)$ *(or* $\ell<\infty$*), and for* $i\in[r]$, $j\in[\ell]$,

$$\lambda_i(\boldsymbol{\alpha}'\boldsymbol{\alpha})\asymp p^{2\gamma_\alpha},\quad \lambda_j(\boldsymbol{\beta}'\boldsymbol{\beta})\asymp q^{2\gamma_\beta},$$

where $\gamma_\alpha,\gamma_\beta\in[0,1/2]$. *Then as* $\min(p,q,T)\to\infty$ *with* $p/T\to c_p\in(0,\infty)$ *and* $q/T\to c_q\in(0,\infty)$, $\hat{r}$ *and* $\hat{\ell}$ *are consistent estimators for* r *and* ℓ *respectively.*

The assumption for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are parallel to Assumption (L1) for the factor loading matrices $\mathbf{A}_r$ and $\mathbf{A}_c$, with the row main effects factor loading matrix $\boldsymbol{\alpha}$ assumed to have the same factor strength γ_α. The same holds true for $\boldsymbol{\beta}$ with the same factor strength γ_β.

The assumption r/T and ℓ/T approach constants less than 1 is such that, together with Assumption (ME1), we can use a random matrix theory for general linear processes to bound the largest and smallest eigenvalues of the sample covariance matrices of $\{\mathbf{g}_t\}_{t\in[T]}$ or $\{\mathbf{h}_t\}_{t\in[T]}$. Note that if $T\asymp p\asymp q$, then we are essentially allowing r/p and ℓ/q to approach constants less than or equal to 1. This is important since if the main effects $\boldsymbol{\alpha}_t$ or $\boldsymbol{\beta}_t$ in model (3.1) are all "weak", e.g. most countries' main effects are of different dynamics from others, then the number of global common factors contributing to $\boldsymbol{\alpha}_t$ or $\boldsymbol{\beta}_t$ is of the same order as p or q respectively. At the same time, if most of the elements in $\boldsymbol{\alpha}_t$ or $\boldsymbol{\beta}_t$ are zero, meaning that $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$ are sparse, then it corresponds to γ_α or γ_β being closer to 0. These practical situations are all allowed in Theorem 10.

5. Numerical results

5.1. Simulation

We demonstrate the performance of our estimators in this section. We will experiment different settings to assess consistency results as described in Theorems 2 and 3, followed by the asymptotic normality of our estimators in Theorems 4 and 5, where the covariance matrices can be constructed by their consistent estimators by Theorem 4 and Theorem 6, respectively. We then showcase the results for the rank estimators described in Theorem 9. We also show the performance of the estimators $\hat{r}$ and $\hat{\ell}$ in Theorem 10 for the number of factors contributing to the main effects in Section 5.1.5. As it is a first to consider matrix factor model with time-varying grand mean and main effects, we unveil the differences between MEFM and FM using numerical results that will illustrate Theorem 7.

For the data generating process, we use Assumptions (E1), (E2) and (F1) to generate general linear processes for the noise and factor series in model (3.1). To be precise, the elements in $\mathbf{F}_t$ are independent standardized AR(5) with AR coefficients 0.7, 0.3, −0.4, 0.2, and −0.1. The elements in $\mathbf{F}_{e,t}$ and ϵ_t are generated similarly, but their AR coefficients are (−0.7, −0.3, −0.4, 0.2, 0.1) and (0.8, 0.4, −0.4, 0.2, −0.1) respectively. The standard deviation of each element in ϵ_t is generated by i.i.d. $|\mathcal{N}(0,1)|$. To test how robust our method is under heavy-tailed distribution, we consider two distributions for the innovation process in generating $\mathbf{F}_t$, $\mathbf{F}_{e,t}$ and ϵ_t: (1) i.i.d. $\mathcal{N}(0,1)$; (2) i.i.d. t_3.

The row factor loading matrix $\mathbf{A}_r$ is generated with $\mathbf{A}_r = \mathbf{M}_p\mathbf{U}_r\mathbf{B}_r$, where each entry of $\mathbf{U}_r \in \mathbb{R}^{p\times k_r}$ is i.i.d. $\mathcal{N}(0,1)$, and $\mathbf{B}_r \in \mathbb{R}^{k_r\times k_r}$ is diagonal with the jth diagonal entry being $p^{-\zeta_{r,j}}$, $0 \le \zeta_{r,j} \le 0.5$. Pervasive (strong) factors have $\zeta_{r,j} = 0$, while weak factors have $0 < \zeta_{r,j} \le 0.5$. Note that $\mathbf{M}_p$ is defined in (3.4) so that (IC1) is satisfied. In a similar way, the column factor loading matrix $\mathbf{A}_c$ is generated independently. Each entry of $\mathbf{A}_{e,r} \in \mathbb{R}^{p\times k_{e,r}}$ is i.i.d. $\mathcal{N}(0,1)$ and has independent probability of 0.95 being set exactly to 0, and $\mathbf{A}_{e,c}$ is generated similarly. We fix $k_{e,r} = k_{e,c} = 2$ throughout the section.

For any $t \in [T]$, we generate $\mu_t = \upsilon_{\mu,t}$, $\boldsymbol{\alpha}_t = \mathbf{M}_p\mathbf{v}_{\alpha,t}$ and $\boldsymbol{\beta}_t = \mathbf{M}_q\mathbf{v}_{\beta,t}$, where $\upsilon_{\mu,t}$ is $\mathcal{N}(m_\mu,\sigma_\mu^2)$, each element of $\mathbf{v}_{\alpha,t}$ is i.i.d. $\mathcal{N}(m_\alpha,\sigma_\alpha^2)$ and that of $\mathbf{v}_{\beta,t}$ is i.i.d. $\mathcal{N}(m_\beta,\sigma_\beta^2)$. We set $m_\mu = m_\alpha = m_\beta = 0$ and $\sigma_\mu = \sigma_\alpha = \sigma_\beta = 1$, and every experiment in this section is repeated 1000 times unless specified otherwise.

5.1.1. Accuracy of various estimators

To assess the accuracy of our estimators, we define the relative mean squared errors (MSE) for μ_t, $\boldsymbol{\alpha}_t$, $\boldsymbol{\beta}_t$ and $\mathbf{C}_t$ as the following, respectively,

$$\text{relative MSE}_\mu = \frac{\sum_{t=1}^T(\mu_t-\hat{\mu}_t)^2}{\sum_{t=1}^T \mu_t^2},\quad \text{relative MSE}_\alpha = \frac{\sum_{t=1}^T\|\boldsymbol{\alpha}_t-\hat{\boldsymbol{\alpha}}_t\|^2}{\sum_{t=1}^T\|\boldsymbol{\alpha}_t\|^2},$$
$$\text{relative MSE}_\beta = \frac{\sum_{t=1}^T\|\boldsymbol{\beta}_t-\hat{\boldsymbol{\beta}}_t\|^2}{\sum_{t=1}^T\|\boldsymbol{\beta}_t\|^2},\quad \text{relative MSE}_{\mathbf{C}} = \frac{\sum_{t=1}^T\|\mathbf{C}_t-\hat{\mathbf{C}}_t\|_F^2}{\sum_{t=1}^T\|\mathbf{C}_t\|_F^2}.$$

For measuring the accuracy of our factor loading matrix estimators, we use the column space distance,

$$\mathcal{D}(\mathbf{Q},\hat{\mathbf{Q}}) = \left\|\mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}' - \hat{\mathbf{Q}}(\hat{\mathbf{Q}}'\hat{\mathbf{Q}})^{-1}\hat{\mathbf{Q}}'\right\|,$$

for any given $\mathbf{Q}$ and $\hat{\mathbf{Q}}$, which is a common measure in the literature such as (Chen et al., 2022) and Chen and Fan (2023).

We consider the following settings:

(Ia) $T = 100$, $p = q = 40$, $k_r = 1$, $k_c = 2$. All factors are pervasive with $\zeta_{r,j} = \zeta_{c,j} = 0$. All innovation processes in constructing $\mathbf{F}_t$, $\mathbf{F}_{e,t}$ and ϵ_t are i.i.d. standard normal.
(Ib) Same as (Ia), but one factor is weak with $\zeta_{r,1} = 0.2$ and $\zeta_{c,1} = 0.2$. Set also $m_\alpha = -2$.
(Ic) Same as (Ia), but all innovation processes are i.i.d. t_3.
(Id) Same as (Ib), but $T = 100$, $p = q = 80$ and $\sigma_\alpha = 2$.
(Ie) Same as (Id), but $T = 200$.
(IIa–e) Same as (Ia) to (Ie) respectively, except that we generate $\mathbf{F}_t$, $\mathbf{F}_{e,t}$ and ϵ_t using white noise rather than AR(5).

Setting (IIa) to (IIe) are to investigate how temporal dependence in the noise affects our results. .

We report the boxplots of accuracy measures for our estimators from Fig. 1 to Fig. 6. Note first that stronger temporal dependence leads to larger variance of our estimators in general. The serial dependence mainly undermines the performance of our loading matrix estimators as shown in Figs. 5 and 6, which in turn affects our common component estimator (see Fig. 2).

Considering the comparisons among (Ia) to (Ie), we see that relative MSE$_\mu$ can be improved by increasing the spatial dimensions, but is not affected by weak factors. Similar results can be seen from Fig. 3 and Fig. 4 for relative MSE$_\alpha$ and relative MSE$_\beta$. The detrimental effects of heavy-tailed innovation processes in Setting (Ic) are most reflected in the corresponding boxplots in Fig. 4.

Weak factors can be detrimental to the accuracy of the factor loading matrix estimators, as can be seen by the significant rise in the factor loading space errors from Setting (Ia) to (Ib) in Figs. 5 and 6. In fact, $\hat{k}_c$ barely captures the second factor under Setting (Ib) and (IIb). See Section 5.1.2 for details. Comparing Setting (Ib) with (Id), Fig. 5 and 6 show that increase in data dimensions slightly improves our factor loading matrix estimators, which is consistent to the simulation results in Wang et al. (2019) for instance.

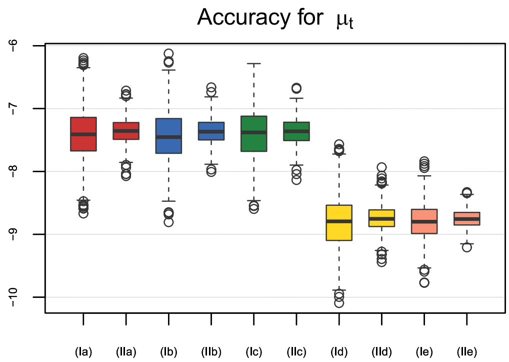

Fig. 1. Plot of the relative MSE for μ_t (in log-scale) from Settings (Ia) to (Ie), comparing with (IIa) to (IIe).

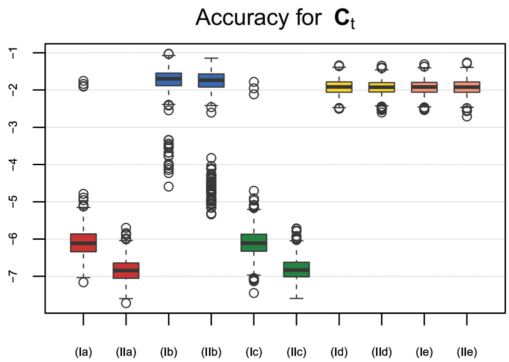

Fig. 2. Plot of the relative MSE for $\mathbf{C}_t$ (in log-scale) from Settings (Ia) to (Ie), comparing with (IIa) to (IIe).

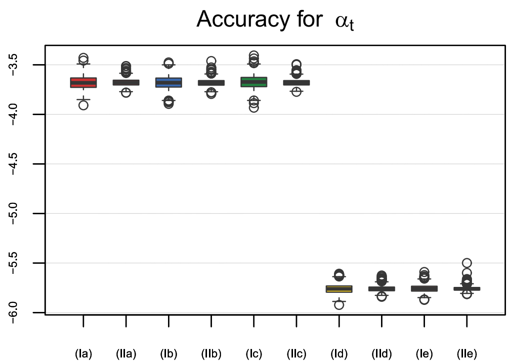

Fig. 3. Plot of the relative MSE for $\boldsymbol{\alpha}_t$ (in log-scale) from Settings (Ia) to (Ie), comparing with (IIa) to (IIe).

5.1.2. Performance for the estimation of the number of factors k_r and k_c

In this section, we demonstrate the performance of our estimators for the number of factors, as described in Theorem 9. First, we set $\xi_r = pq[(Tq)^{-1/2} + p^{-1/2}]/5$ and $\xi_c = pq[(Tp)^{-1/2} + q^{-1/2}]/5$, so that the conditions for ξ_r and ξ_c in (4.13) and (4.14) are respectively satisfied. A wide range of values other than $1/5$ for ξ_r and ξ_c are experimented, but $1/5$ is working the best in vast majority of settings, and hence we do not recommend treating this as a tuning parameter.

We present the results for each of the following settings:

- (IIIa) $k_r = k_c = 3$. All factors are pervasive with $\zeta_{r,j} = \zeta_{c,j} = 0$ for all $j \in [3]$. All innovation processes involved are i.i.d. standard normal.
- (IIIb) Same as (IIIa), but some factors are weak with $\zeta_{r,1} = \zeta_{c,1} = \zeta_{c,2} = 0.2$.
- (IIIc) Same as (IIIa), but all factors are weak with $\zeta_{r,j} = \zeta_{c,j} = 0.2$ for all $j \in [3]$.

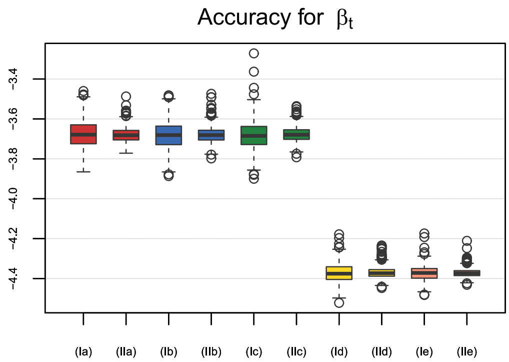

Fig. 4. Plot of the relative MSE for β_t (in log-scale) from Settings (Ia) to (Ie), comparing with (IIa) to (IIe).

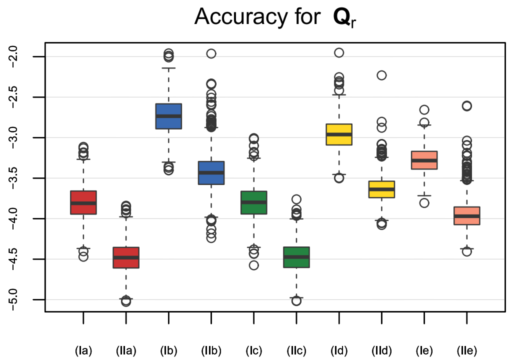

Fig. 5. Plot of the row space distance $\mathcal{D}(\mathbf{Q}_r, \widehat{\mathbf{Q}}_r)$ (in log-scale) from Settings (Ia) to (Ie), comparing with (IIa) to (IIe).

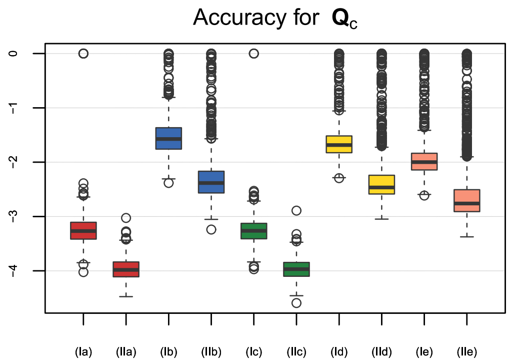

Fig. 6. Plot of the column space distance $\mathcal{D}(\mathbf{Q}_c, \widehat{\mathbf{Q}}_c)$ (in log-scale) from Settings (Ia) to (Ie), comparing with (IIa) to (IIe).

We experiment the above settings with various (p, q) pairs among $(10, 10)$, $(10, 20)$ and $(20, 20)$, with the choice $T = 0.5 \cdot pq$ or $T = pq$. The setup is similar to Wang et al. (2019) and Chen and Fan (2023), but we use smaller sets of dimensions since the accuracy of our estimators are approaching 1 with larger dimensions, which reveal little intricacies among different settings.

From the results in Table 1, our eigenvalue-ratio estimators is working well with MEFM. The accuracy of $\widehat{k}_r$ and $\widehat{k}_c$ suffers from the existence of weak factors, which is also seen in traditional FM (see for instance (Chen and Lam, 2024) and Cen and Lam (2025)). In particular, the accuracy of our estimators drops significantly as we move from Setting (IIIa) to (IIIc), and in general large dimensions are beneficial to our estimation. Lastly, note that although we have two weak factors in the column loading matrix while there is only one weak factor in the row loading matrix, the correct proportion of $\widehat{k}_c$ is much larger than that of $\widehat{k}_r$ for $(p, q) = (10, 20)$. This hints at the importance of data dimensions over factor strength, which can also be seen from the fact that the results for $(p, q) = (20, 20)$ under Setting (IIIc) are comparable with those for $(p, q) = (10, 10)$ under Setting (IIIa).

Table 1
Results for Settings (IIIa)–(IIIc). Each cell reports the frequency of $(\hat{k}_r, \hat{k}_c)$ under the setting in the corresponding column. The true number of factors is $(k_r, k_c) = (3, 3)$, and the cells corresponding to correct estimations are bolded.

$(\hat{k}_r, \hat{k}_c)$	$p, q = 10, 10$		$p, q = 10, 20$		$p, q = 20, 20$	
	$T = .5pq$	$T = pq$	$T = .5pq$	$T = pq$	$T = .5pq$	$T = pq$
	Setting (IIIa)					
(2, 3)	0.121	0.112	0.128	0.11	0	0.004
(3, 2)	0.124	0.111	0.004	0.003	0.001	0.001
(3, 3)	**0.583**	**0.659**	**0.833**	**0.855**	**0.999**	**0.995**
other	0.172	0.118	0.035	0.032	0	0
	Setting (IIIb)					
(2, 3)	0.135	0.13	0.23	0.257	0.228	0.149
(3, 2)	0.079	0.096	0.024	0.017	0.022	0.02
(3, 3)	**0.136**	**0.17**	**0.289**	**0.347**	**0.556**	**0.637**
other	0.65	0.604	0.457	0.379	0.194	0.194
	Setting (IIIc)					
(2, 3)	0.082	0.085	0.218	0.254	0.089	0.096
(3, 2)	0.075	0.124	0.04	0.035	0.088	0.089
(3, 3)	**0.073**	**0.096**	**0.209**	**0.257**	**0.614**	**0.646**
other	0.77	0.695	0.533	0.454	0.209	0.169

5.1.3. Asymptotic normality

We numerically demonstrate the asymptotic normality results in Theorems 4 and 5 in this section. For the ease of demonstration, we consider $t = 10$ only for the asymptotic distribution of $\hat{\mu}_t$, $\hat{\boldsymbol{\theta}}_{\alpha,t} = (\hat{\alpha}_{t,1}, \hat{\alpha}_{t,2}, \hat{\alpha}_{t,3})'$ and $\hat{\boldsymbol{\theta}}_{\beta,t} = (\hat{\beta}_{t,1}, \hat{\beta}_{t,2}, \hat{\beta}_{t,3})'$, and for $\hat{\boldsymbol{\theta}}_{\alpha,t}$ and $\hat{\boldsymbol{\theta}}_{\beta,t}$ we will only report results for the third component. We will also demonstrate the asymptotic normality for $(\hat{\mathbf{Q}}_c)_{1\cdot}$ and present the results for $(\hat{\mathbf{Q}}_c)_{11}$, i.e., the first entry of the first row in the column loading matrix estimator. To consistently estimate its covariance matrix, we use Theorem 6 with $\eta_c = \lfloor (Tpq)^{1/4}/5 \rfloor$.

We use heavy-tailed innovations to investigate the robustness of our results, hence Setting (Ic) is adapted except that we generate $\mathbf{F}_t$, $\mathbf{F}_{e,t}$ and ϵ_t using AR(1) with coefficient -0.2. Due to the different rates of convergence in Theorems 4 and 5, we specify different dimensions (T, p, q) in the following settings:

$$\hat{\mu}_t : (80, 100, 100), \quad \hat{\boldsymbol{\theta}}_{\alpha,t} : (60, 60, 300), \quad \hat{\boldsymbol{\theta}}_{\beta,t} : (60, 300, 60), \quad (\hat{\mathbf{Q}}_c)_{1\cdot} : (60, 60, 300),$$

where the dimension setting for $(\hat{\mathbf{Q}}_c)_{1\cdot}$ is to align with the rate conditions in Theorem 5 that $Tp/q^2 \to 0$ under pervasive factors. Each setting is repeated 400 times, and we present the histograms of our four estimators in Fig. 7.

Our plots stand as empirical evidence of Theorem 4, 5 and 6. It might worth noting that the spread of the normalized empirical density for $\hat{\beta}_{10,3}$ is slightly larger than expected by comparing with the superimposed standard normal. The same problem is not seen in the histogram for $\hat{\alpha}_{10,3}$. With true $(k_r, k_c) = (1, 2)$, the common component estimation using $(p, q) = (300, 60)$ is worse than that using $(p, q) = (60, 300)$ due to insufficient column dimension relative to k_c. Hence it leads to worse estimators for errors and $(\hat{\gamma}_{\beta,1}^{-1}, \hat{\gamma}_{\beta,2}^{-1}, \hat{\gamma}_{\beta,3}^{-1})$ under $(p, q) = (300, 60)$. Hence inference performances on the time-varying row and column effect estimators are affected by the latent number of factors.

5.1.4. Testing MEFM versus FM

We demonstrate numerical results for Theorem 8 in this section. We consider the two scenarios below:

1. *(Global effect.)* The entries of at least one of $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ are in general non-zero for each t.
2. *(Local effect.)* The entries of at least one of $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ are sparse for each t, i.e., given any t, there are some non-zero entries in at least one of $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ with all other entries zero.

Throughout this section, we generate the time-varying grand mean and main effects using Rademacher random variables such that $v_{\mu,t}$ is i.i.d. Rademacher multiplied by some u_μ and each entry of $\mathbf{v}_{\alpha,t}$, $\mathbf{v}_{\beta,t}$ is i.i.d. Rademacher multiplied by some u_α, u_β respectively, recalling that $\mu_t = v_{\mu,t}$, $\boldsymbol{\alpha}_t = \mathbf{M}_p \mathbf{v}_{\alpha,t}$ and $\boldsymbol{\beta}_t = \mathbf{M}_q \mathbf{v}_{\beta,t}$. Hence, setting $u_\mu = u_\alpha = u_\beta = 0$ corresponds to generating a traditional FM. We set $k_r = k_c = 2$, and consider the following settings:

(IVa) $T = p = q = 40$. All factors are pervasive with $\zeta_{r,j} = \zeta_{c,j} = 0$. All innovation processes in constructing $\mathbf{F}_t$, $\mathbf{F}_{e,t}$ and ϵ_t are i.i.d. standard normal. Set $u_\mu = u_\beta = 0$, and we select u_α from 0.1, 0.5, 1.

(IVb) Same as (IVa), but fix $u_\alpha = 0.1$ and select u_β from 0.1, 0.5, 1.

(IVc) Same as (IVa), except that $u_\alpha = 1$, and when generating $\boldsymbol{\alpha}_t = \mathbf{M}_p \mathbf{v}_{\alpha,t}$ as specified previously, we only keep the first u_{local} entries of $\mathbf{v}_{\alpha,t}$ as non-zero where u_{local} is selected from 2, 5, 10.

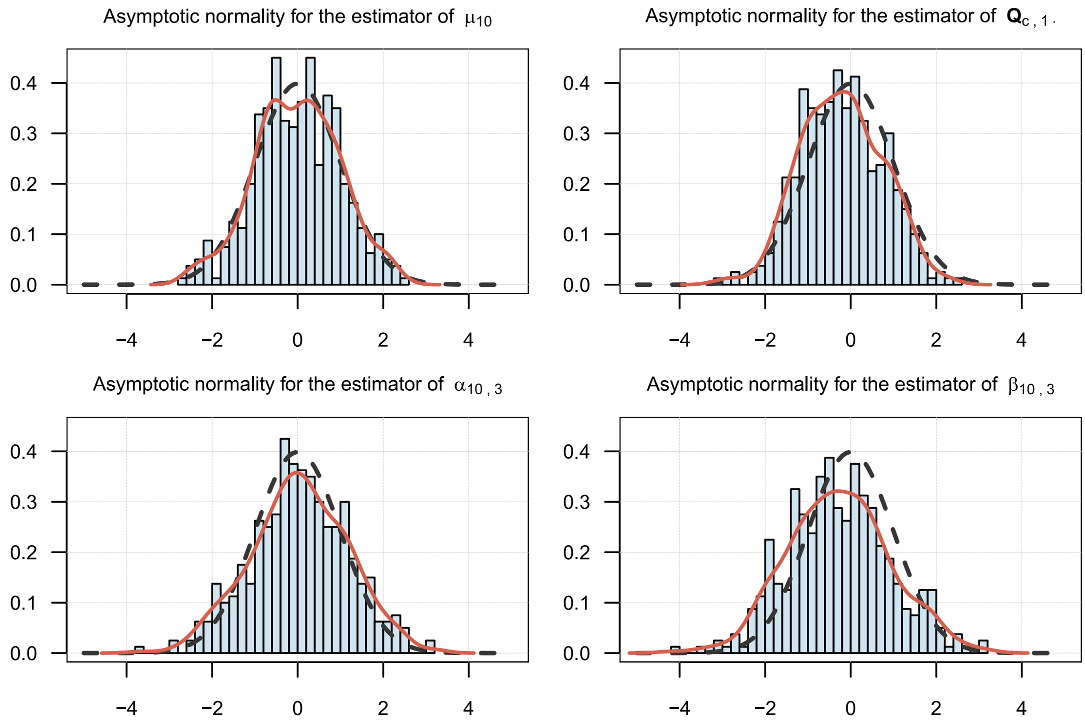

Fig. 7. Histograms of $\sqrt{pq}\,\hat{\gamma}_{\mu}^{-1}(\hat{\mu}_{10}-\mu_{10})$ (top-left), $[T\,(\hat{\boldsymbol{\Sigma}}_{c,1}^{HAC})^{-1/2}\hat{\mathbf{D}}_c(\hat{\mathbf{Q}}_{c,1\cdot}-\mathbf{H}_1^a\mathbf{Q}_{c,1\cdot})]_1$ (top-right), $\sqrt{q}\,[\mathrm{diag}(\hat{\gamma}_{\alpha,1}^{-1},\hat{\gamma}_{\alpha,2}^{-1},\hat{\gamma}_{\alpha,3}^{-1})(\hat{\boldsymbol{\theta}}_{\alpha,10}-\boldsymbol{\theta}_{\alpha,10})]_3$ (bottom-left), and $\sqrt{p}\,[\mathrm{diag}(\hat{\gamma}_{\beta,1}^{-1},\hat{\gamma}_{\beta,2}^{-1},\hat{\gamma}_{\beta,3}^{-1})(\hat{\boldsymbol{\theta}}_{\beta,10}-\boldsymbol{\theta}_{\beta,10})]_3$ (bottom-right). In each panel, the curve (in red) is the empirical density, and the density curve for $\mathcal{N}(0,1)$ (in black, dotted) is also superimposed.

Table 2
Results for Settings (IVa)–(IVc). Each cell reports the mean and standard deviation (subscripted), both multiplied by 100. The parameters for Settings (IVa), (IVb) and (IVc) are u_α, u_β and u_{local}, respectively. Setting (IVa) with $u_\alpha = 0$ is reported in the first column, representing the size of the test.

Parameter	Size	Setting (IVa)			Setting (IVb)			Setting (IVc)		
	0	0.1	0.5	1	0.1	0.5	1	2	5	10
$\mathbf{reject}_\alpha$	$5_{(4)}$	$11_{(7)}$	$63_{(31)}$	$96_{(15)}$	$13_{(8)}$	$53_{(30)}$	$86_{(23)}$	$37_{(17)}$	$77_{(24)}$	$85_{(27)}$
$\mathbf{reject}_\beta$	$5_{(4)}$	$11_{(7)}$	$52_{(28)}$	$87_{(22)}$	$13_{(8)}$	$62_{(32)}$	$96_{(16)}$	$14_{(8)}$	$28_{(16)}$	$48_{(26)}$

Setting (IVa) and (IVb) are designed for testing global effects, and Setting (IVc) for local effects. For each setting, we construct $y_{\alpha,t}$, $y_{\beta,t}$ and use $\theta = 0.95$ in Theorem 8. Each experiment is repeated 400 times and we report both $\mathbf{reject}_\alpha := T^{-1}\sum_{t=1}^{T}\mathbb{1}\{y_{\alpha,t} \geq \hat{q}_{x,\alpha}(0.95)\}$ and $\mathbf{reject}_\beta := T^{-1}\sum_{t=1}^{T}\mathbb{1}\{y_{\beta,t} \geq \hat{q}_{x,\beta}(0.95)\}$.

As explained under Theorem 8, we expect $\mathbf{reject}_\alpha$ and $\mathbf{reject}_\beta$ to be close to $1-\theta = 0.05$ if FM is sufficient. From Table 2, our proposed test works well since it suggests FM is insufficient as we strengthen $\boldsymbol{\alpha}_t$ or $\boldsymbol{\beta}_t$. In particular, even if the signal of $\boldsymbol{\alpha}_t$ is not strong enough such as $u_\alpha = 0.1$, Setting (IVb) shows that additional signals from $\boldsymbol{\beta}_t$ allows us to reject the use of FM. The comparison between $\mathbf{reject}_\alpha$ and $\mathbf{reject}_\beta$ is indicative of which effect is stronger. According to the results for (IVc) in the table, our test is capable of detecting local effect such that $\mathbf{reject}_\alpha$ is far from 0.05 even when only two entries in $\boldsymbol{\alpha}_t$ are non-zero.

Extensive experiments on different dimensions, factor strengths or grand mean magnitudes are performed. All indicate similar interpretations as the above settings and hence the results are not shown here. The power curves for Setting (IVa) are also presented in Fig. 8 to support the use of our test, with $(T, p, q) = (60, 80, 80)$ and u_α ranging from 0.02 to 1. The $\mathbf{reject}_\alpha$ statistic is more powerful in general than the $\mathbf{reject}_\beta$ one since the $\mathbf{reject}_\beta$ statistic is more sensitive to signals from the column main effects $\boldsymbol{\beta}_t$, which is 0 in Setting (IVa). Besides, we also show the power curve for local effect in Fig. 9, for Setting (IVc) except that $(T, p, q) = (60, 80, 80)$ and we generate $\boldsymbol{\alpha}_t$ as described in the caption. Both power curves show that the test is able to reject the use of FM if signals from the time-varying main effects are large, either globally or locally. Similar to Fig. 8, the $\mathbf{reject}_\beta$ statistic gains power only slowly since $\boldsymbol{\beta}_t$ is zero, and the low power is exacerbated from the fact that the signals from $\boldsymbol{\alpha}_t$ are only local. This prompts us to look at both $\mathbf{reject}_\alpha$ and $\mathbf{reject}_\beta$ in practice. Finally, in both figures, when u_α is close to 0.02 or $\tilde{u}_{local}$ close to 0, the value of the power curves are all very close to 0.05, which is exactly what we want for the size of the tests.

To investigate the robustness of our proposed testing procedure, we consider different variants of Settings (IVa)–(IVc), described as follows:

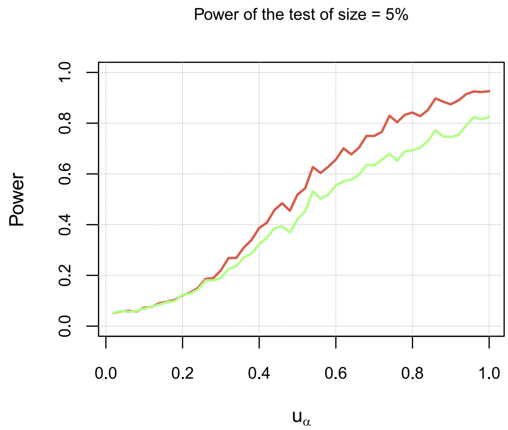

Fig. 8. Statistical power curve of testing the null hypothesis that FM is sufficient for the given series, against the alternative that MEFM is necessary. Each power value is computed as the average over 400 runs of $\mathbf{reject}_\alpha$ (in red) and $\mathbf{reject}_\beta$ (in green) under Setting (IVa) except that $(T, p, q) = (60, 80, 80)$.

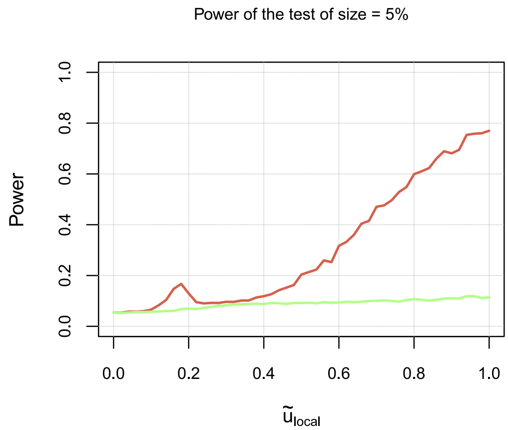

Fig. 9. Statistical power curve of testing the null hypothesis that FM is sufficient for the given series, against the alternative that MEFM is necessary. Refer to Fig. 8 for how the power is computed. The data is generated under Setting (IVc) except that $(T, p, q) = (60, 80, 80)$ and $\boldsymbol{\alpha}_t$ is generated as $\boldsymbol{\alpha}_1 = \widetilde{u}_{local}\,(1, 1, -2, 0, \ldots, 0)'$, $\boldsymbol{\alpha}_2 = \widetilde{u}_{local}\,(1, 2, -3, 0, \ldots, 0)'$, $\boldsymbol{\alpha}_3 = \widetilde{u}_{local}\,(2, -5, 3, 0, \ldots, 0)'$ and $\boldsymbol{\alpha}_{3\ell+i} = \boldsymbol{\alpha}_i$ for ℓ a positive integer and $i = 1, 2, 3$, so that each $\boldsymbol{\alpha}_t$ has non-zero entries only in the first three indices.

(Va–c) Same as (IVa)–(IVc), but the innovations in constructing $\mathbf{F}_t$, $\mathbf{F}_{e,t}$ and ϵ_t are i.i.d. t_3.
(VIa–c) Same as (IVa)–(IVc), but all factors are weak with $\zeta_{r,j} = \zeta_{c,j} = 0.3$.
(VIIa–c) Same as (IVa)–(IVc), except that $T = p = q = 30$.
(VIIIa–c) Same as (IVa)–(IVc), except that the number of core factors are over-estimated by 2.

Essentially, Settings (Va)–(Vc) violate the tail condition in Assumption (E3), Settings (VIa)–(VIc) imply that all factor strengths are below $1/2$ and hence breaks Assumption (R1), Settings (VIIa)–(VIIc) showcase the finite-sample behavior of the test, and Settings (VIIIa)–(VIIIc) represent misspecification of the factor structure. Table 3 reports the test size and power for each setting.

Each setting in Table 3 represents a scenario with certain assumptions violated, so it is unsurprising that the power drops when the varying parameter is small, implying weak signals. Nevertheless, it is interesting that under Settings (VIa)–(VIc), i.e., when all factors are too weak for the factor structure to be consistently estimated, or under Settings (VIIIa)–(VIIIc) where the number of factors is misspecified, the test are more powerful for a majority of parallel settings in Table 2, while the size remains satisfying. This might benefit from the fact that the test is based on comparing estimated residuals and stays robust even if the factor structure is estimated poorly both under MEFM and FM. Overall, although all the test sizes are slightly inflated, the test power show similar patterns as in Table 2.

Table 3
Results for Settings (Va)–(Vc), (VIa)–(VIc), (VIIa)–(VIIc) and (VIIIa)–(VIIIc). Refer to Table 2 for the details of each entry.

	Size	Setting (Va)			Setting (Vb)			Setting (Vc)		
Parameter	0	0.1	0.5	1	0.1	0.5	1	2	5	10
reject$_\alpha$	$9_{(5)}$	$10_{(6)}$	$60_{(28)}$	$94_{(18)}$	$11_{(6)}$	$48_{(28)}$	$86_{(23)}$	$34_{(17)}$	$73_{(24)}$	$84_{(26)}$
reject$_\beta$	$9_{(5)}$	$10_{(6)}$	$47_{(25)}$	$85_{(24)}$	$11_{(7)}$	$58_{(30)}$	$96_{(16)}$	$13_{(6)}$	$23_{(14)}$	$41_{(24)}$
	Size	Setting (VIa)			Setting (VIb)			Setting (VIc)		
Parameter	0	0.1	0.5	1	0.1	0.5	1	2	5	10
reject$_\alpha$	$9_{(5)}$	$10_{(6)}$	$66_{(26)}$	$99_{(2)}$	$11_{(6)}$	$52_{(26)}$	$91_{(12)}$	$17_{(11)}$	$58_{(17)}$	$87_{(14)}$
reject$_\beta$	$9_{(5)}$	$10_{(6)}$	$52_{(25)}$	$91_{(13)}$	$11_{(7)}$	$65_{(27)}$	$99_{(2)}$	$13_{(7)}$	$23_{(13)}$	$44_{(22)}$
	Size	Setting (VIIa)			Setting (VIIb)			Setting (VIIc)		
Parameter	0	0.1	0.5	1	0.1	0.5	1	2	5	10
reject$_\alpha$	$9_{(6)}$	$9_{(6)}$	$60_{(26)}$	$97_{(13)}$	$11_{(7)}$	$49_{(25)}$	$87_{(21)}$	$34_{(16)}$	$71_{(24)}$	$86_{(24)}$
reject$_\beta$	$9_{(6)}$	$9_{(9)}$	$49_{(24)}$	$86_{(21)}$	$11_{(8)}$	$59_{(26)}$	$97_{(15)}$	$14_{(9)}$	$29_{(16)}$	$53_{(24)}$
	Size	Setting (VIIIa)			Setting (VIIIb)			Setting (VIIIc)		
Parameter	0	0.1	0.5	1	0.1	0.5	1	2	5	10
reject$_\alpha$	$9_{(5)}$	$10_{(6)}$	$73_{(22)}$	$99_{(2)}$	$12_{(6)}$	$56_{(23)}$	$92_{(9)}$	$13_{(6)}$	$45_{(13)}$	$87_{(10)}$
reject$_\beta$	$9_{(5)}$	$10_{(6)}$	$55_{(22)}$	$92_{(10)}$	$12_{(7)}$	$73_{(22)}$	$99_{(1)}$	$12_{(6)}$	$21_{(11)}$	$43_{(19)}$

Table 4
Proportion of estimated number of factors for the main effects, under various settings. Each cell reports the frequency of $(\hat{r}, \hat{\ell})$ under the corresponding setting over 1000 runs. The cells corresponding to correct estimations are bolded.

$(\hat{r}, \hat{\ell})$	$(p,q)=(20,20)$		$(p,q)=(20,40)$		$(p,q)=(40,40)$	
	$T=p+q$	$T=2(p+q)$	$T=p+q$	$T=2(p+q)$	$T=p+q$	$T=2(p+q)$
	$(r,\ell)=(2,2)$					
$(1,2)$	0	0	0	0	0	0
$(2,1)$	.002	0	0	0	0	0
$(2,2)$	**.998**	**1**	**1**	**1**	**1**	**1**
others	0	0	0	0	0	0
	$(r,\ell)=(6,6)$					
$(5,6)$	.011	.005	0	0	0	0
$(6,5)$	.004	.008	0	0	0	0
$(6,6)$	**.982**	**.987**	**1**	**1**	**1**	**1**
others	.003	0	0	0	0	0
	$(r,\ell)=(18,18)$					
$(17,18)$	.002	.012	.301	.342	0	0
$(18,17)$	.003	.013	0	0	0	0
$(18,18)$	**.001**	**.001**	**.067**	**.098**	**1**	**1**
others	.994	.974	.632	.560	0	0

5.1.5. Performance on estimating the number of factors for the main effects

We demonstrate the numerical performance of $\hat{r}$ and $\hat{\ell}$ in (4.15) and (4.16). For simplicity, we fix $k_r = k_c = 2$ and all core factors to be pervasive. The data is generated as described at the beginning of Section 5.1 with Gaussian innovations, except that the main effects are obtained by $\boldsymbol{\alpha}_t = \mathbf{M}_p \boldsymbol{\alpha} \mathbf{g}_t$ and $\boldsymbol{\beta}_t = \mathbf{M}_q \boldsymbol{\beta} \mathbf{h}_t$ with $\boldsymbol{\alpha} \in \mathbb{R}^{p\times r}$, $\boldsymbol{\beta} \in \mathbb{R}^{q\times \ell}$, and each element in $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\mathbf{g}_t$ and $\mathbf{h}_t$ being i.i.d. $\mathcal{N}(0,1)$.

The number of factors for the row and column main effects (r,ℓ) is set as $(2,2)$, $(6,6)$ and $(18,18)$ respectively. For each setting, we experiment various dimensions (p,q) from $(20,20)$, $(20,40)$ and $(40,40)$, each with $T=p+q$ and $T=2(p+q)$ respectively. The results are presented in Table 4, which corroborates the consistency of $\hat{r}$ and $\hat{\ell}$ in Theorem 10. We also see that if r or ℓ are too close to p and q, accuracy of $\hat{r}$ or $\hat{\ell}$ can suffer.

5.1.6. Comparison of estimation accuracy between MEFM and FM

To demonstrate the advantages of MEFM over FM, we also compare the estimation accuracy of MEFM and various FM. In particular, we construct $\{\mathbf{Y}_t\}$ based on the same data generating process in Section 5.1.5, which represents MEFM with low-rank main effects. All settings in Table 4 are experimented, except that settings with $(r,\ell)=(18,18)$ are replaced by models without main effects (denoted as $(r,\ell)=(0,0)$). Note that as (r,ℓ) increases, the underlying model drifts away from FM to MEFM. Furthermore, we also consider data with general stationary main effects as described at the beginning of Section 5.1, where each main effect entry is essentially independent standard normal. Finally to examine the effects of non-stationary main effects on estimation accuracy, we added two more cases in Table 5. One is for $\{\boldsymbol{\alpha}_t\}$ and $\{\boldsymbol{\beta}_t\}$ to have a structural change point at time $T/2$, where the loadings in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are regenerated. Another is a random walk for each entry of the main effects with independent standard normal as innovations.

Table 5
Relative MSE of residuals for various settings. Among the methods, the rank of core factor used for MEFM is $(2,2)$, while ranks for FM1, FM2, FM3 are $(3,3)$, $(5,5)$, $(7,7)$, respectively. Each cell reports the mean and standard deviation (subscripted) of the measure over 1000 runs.

Relative	$(p,q)=(20,20)$		$(p,q)=(20,40)$		$(p,q)=(40,40)$	
MSE	$T=p+q$	$T=2(p+q)$	$T=p+q$	$T=2(p+q)$	$T=p+q$	$T=2(p+q)$
	$(r,\ell)=(0,0)$					
MEFM	$.878_{(.010)}$	$.885_{(.007)}$	$.913_{(.005)}$	$.917_{(.003)}$	$.944_{(.002)}$	$.946_{(.002)}$
FM1	$.961_{(.006)}$	$.969_{(.004)}$	$.980_{(.003)}$	$.984_{(.002)}$	$.990_{(.001)}$	$.992_{(.001)}$
FM2	$.867_{(.015)}$	$.893_{(.014)}$	$.929_{(.013)}$	$.943_{(.014)}$	$.965_{(.009)}$	$.971_{(.010)}$
FM3	$.753_{(.015)}$	$.794_{(.014)}$	$.863_{(.013)}$	$.889_{(.014)}$	$.931_{(.009)}$	$.944_{(.010)}$
	$(r,\ell)=(2,2)$					
MEFM	$.878_{(.010)}$	$.885_{(.007)}$	$.913_{(.005)}$	$.917_{(.003)}$	$.944_{(.002)}$	$.946_{(.002)}$
FM1	$3.58_{(.555)}$	$3.67_{(.540)}$	$3.94_{(.507)}$	$3.99_{(.510)}$	$4.23_{(.449)}$	$4.31_{(.436)}$
FM2	$.920_{(.008)}$	$.928_{(.005)}$	$.959_{(.003)}$	$.964_{(.002)}$	$.979_{(.002)}$	$.982_{(.001)}$
FM3	$.803_{(.014)}$	$.828_{(.014)}$	$.895_{(.012)}$	$.910_{(.014)}$	$.947_{(.009)}$	$.954_{(.010)}$
	$(r,\ell)=(6,6)$					
MEFM	$.878_{(.010)}$	$.885_{(.007)}$	$.913_{(.005)}$	$.917_{(.003)}$	$.944_{(.002)}$	$.946_{(.002)}$
FM1	$8.92_{(1.12)}$	$9.19_{(1.06)}$	$9.82_{(1.01)}$	$10.0_{(1.00)}$	$10.7_{(.881)}$	$10.9_{(.835)}$
FM2	$4.22_{(.495)}$	$4.44_{(.495)}$	$5.01_{(.483)}$	$5.22_{(.500)}$	$5.75_{(.446)}$	$5.98_{(.447)}$
FM3	$1.79_{(.187)}$	$1.90_{(.201)}$	$2.25_{(.199)}$	$2.36_{(.213)}$	$2.65_{(.203)}$	$2.78_{(.216)}$
	General Stationary Main Effects					
MEFM	$.878_{(.010)}$	$.885_{(.007)}$	$.913_{(.005)}$	$.917_{(.003)}$	$.944_{(.002)}$	$.946_{(.002)}$
FM1	$2.61_{(.073)}$	$2.63_{(.057)}$	$2.71_{(.056)}$	$2.73_{(.046)}$	$2.80_{(.039)}$	$2.82_{(.033)}$
FM2	$2.13_{(.057)}$	$2.23_{(.049)}$	$2.36_{(.048)}$	$2.44_{(.043)}$	$2.55_{(.034)}$	$2.61_{(.030)}$
FM3	$1.73_{(.046)}$	$1.87_{(.042)}$	$2.06_{(.042)}$	$2.17_{(.089)}$	$2.32_{(.031)}$	$2.42_{(.028)}$
	$(r,\ell)=(2,2)$ with change at $T/2$					
MEFM	$.878_{(.010)}$	$.885_{(.007)}$	$.913_{(.005)}$	$.917_{(.003)}$	$.944_{(.002)}$	$.946_{(.002)}$
FM1	$4.02_{(.516)}$	$4.08_{(.457)}$	$4.31_{(.454)}$	$4.33_{(.423)}$	$4.52_{(.408)}$	$4.57_{(.364)}$
FM2	$1.72_{(.170)}$	$1.79_{(.159)}$	$1.96_{(.172)}$	$2.00_{(.165)}$	$2.14_{(.164)}$	$2.20_{(.154)}$
FM3	$.861_{(.010)}$	$.869_{(.008)}$	$.928_{(.005)}$	$.934_{(.004)}$	$.964_{(.002)}$	$.967_{(.001)}$
	Random Walk Main Effects					
MEFM	$.878_{(.010)}$	$.885_{(.007)}$	$.913_{(.005)}$	$.917_{(.003)}$	$.944_{(.002)}$	$.946_{(.002)}$
FM1	$8.08_{(1.19)}$	$11.5_{(1.76)}$	$10.46_{(1.54)}$	$15.1_{(1.49)}$	$12.4_{(1.41)}$	$19.2_{(1.32)}$
FM2	$3.53_{(.269)}$	$5.95_{(.502)}$	$5.29_{(.365)}$	$8.87_{(.673)}$	$7.26_{(.420)}$	$12.0_{(.795)}$
FM3	$1.96_{(.095)}$	$3.18_{(.177)}$	$2.93_{(.126)}$	$5.02_{(.243)}$	$4.07_{(.159)}$	$7.22_{(.307)}$

For comparison, we estimate each model using MEFM and three FM's such that the number of factors in FM is (k_r+k_0, k_c+k_0) with $k_0=1$, 3 and 5. The FM's are denoted by FM1, FM2 and FM3, respectively. We measure the estimation accuracy by relative MSE of residuals defined as $\sum_{t=1}^{T}\|\widehat{\mathbf{E}}_t\|_F^2/\sum_{t=1}^{T}\|\mathbf{E}_t\|_F^2$, where $\widehat{\mathbf{E}}_t$ represents the estimated residual at time t of each corresponding model. Table 5 shows that, MEFM performs similarly as FM when the main effects have low-rank structures, and largely outperforms FM when (r,ℓ) increases.

In particular, as (r,ℓ) increases, it is inevitable that the estimation of FM requires larger number of factors to account for all genuine factors, while MEFM performs exactly the same due to its model structure and identification. When the number of main effect common factors is large, it becomes impractical for FM to appropriately model the data. This also strengthen our reasoning that MEFM is more general. Moreover, note that while using more factors in FM can alleviate the relative MSE, MEFM remains superior by its more stable estimation.

Lastly, for the general stationary main effects where the main effects are not driven by any global common factors (i.e., essentially $r=p$ and $\ell=q$), the performance of MEFM dominates all FM experimented. For the two scenarios with non-stationary main effects, FM using more factors cannot effectively improve its performance, in particular for the random walk setting, while MEFM remains stable with similarly good performance as other settings.

5.2. Real data analysis

5.2.1. NYC taxi traffic

We analyze a set of taxi traffic data in New York city in this example. The data includes all individual taxi rides operated by Yellow Taxi in New York City, published at

https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page.

For simplicity, we only consider the rides within Manhattan Island, which comprises most of the data. The dataset contains 842 million trip records within the period of January 1, 2013 to December 31, 2022. Each trip record includes features such as pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts. Our example here focuses on the drop-off dates/times and locations.

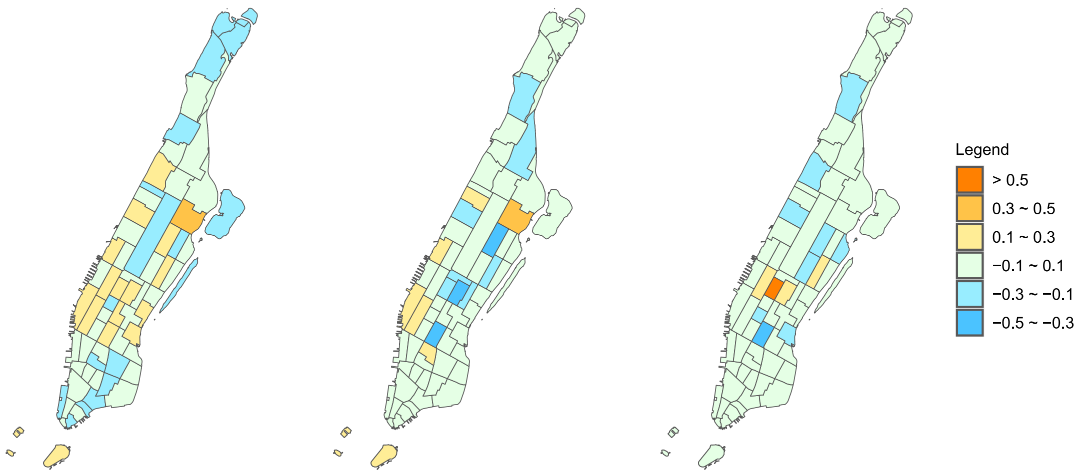

Fig. 10. Estimated loading on three dropoff factors using MEFM, i.e. $\hat{\mathbf{Q}}_{1,\cdot 1}$ (left), $\hat{\mathbf{Q}}_{1,\cdot 2}$ (middle) and $\hat{\mathbf{Q}}_{1,\cdot 3}$ (right).

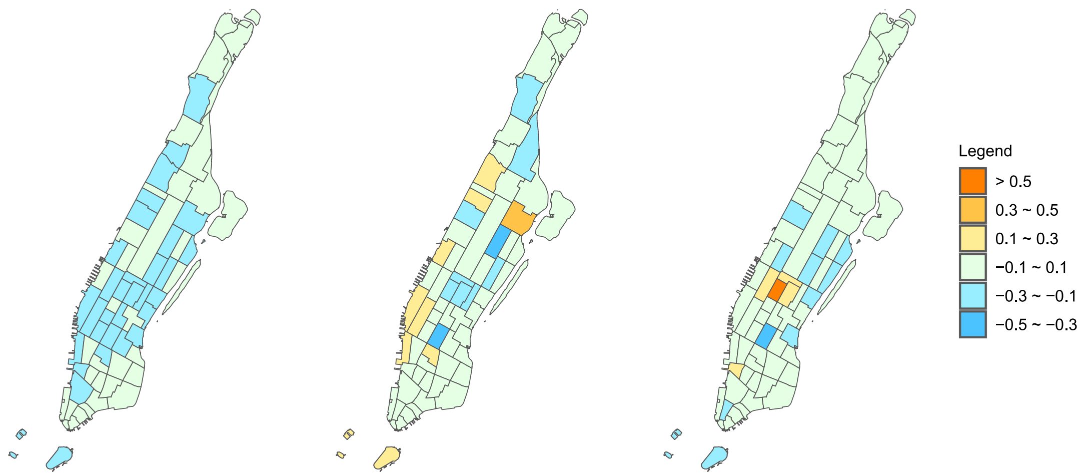

Fig. 11. Estimated loading on three dropoff factors using FM, similar to Fig. 10.

To classify the drop-off locations in Manhattan, they are coded according to 69 predefined zones in the dataset. Moreover, each day is divided into 24 hourly periods to represent the drop-off times each day, with the first hourly period from 0 a.m. to 1 a.m. The total number of rides moving among the zones within each hour are recorded, yielding data $\mathbf{Y}_t \in \mathbb{R}^{69\times 24}$ each day, where $y_{i_1,i_2,t}$ is the number of trips to zone i_1 and the pick-up time is within the i_2-th hourly period on day t.

We consider the non-business-day series which is 1,133 days long, within the period of January 1, 2013 to December 31, 2022. Using MEFM, the estimated rank of the core factors is $(2, 2)$ according to our proposed eigenvalue ratio estimator. As mentioned in Section 4.5, we therefore use $(3, 3)$ as the number of factors to estimate FM and test if FM is sufficient. We compute $\textbf{reject}_\alpha = 0.064$ and $\textbf{reject}_\beta = 0.133$ which are defined in Section 5.1.4. They should be close to $1 - \theta = 0.05$ according to Theorem 8 if FM is sufficient. Hence we reject the use of traditional FM due to the signals in $\hat{\boldsymbol{\beta}}_t$.

To compare MEFM with FM, we use core rank $(3, 3)$ to estimate MEFM for the rest of this section. Figs. 10 and 11 illustrate the heatmaps of the estimated loading columns on the three dropoff factors using MEFM and FM, respectively. From both heatmaps, we can identify the first factor as active areas, the second as dining and sports areas and the third as downtown areas. The three factors are similar to their corresponding counterparts, except that the first factor estimated using MEFM is more indicative on the active areas to taxi traffic in Manhattan by its emphasized orange zone which corresponds to East Harlem.

To gain further understanding on the taxi traffic, we show the scaled $\hat{\mathbf{Q}}_2$ by MEFM and FM in Tables 6 and 7, respectively. We can see that for the rush hours from 6 p.m. to 11 p.m., the estimated loadings almost vanish for MEFM, which is consistent with the fact that $\hat{\boldsymbol{\beta}}_t$ captures the common hour effect on Manhattan life style. This also provides an intuition why the time-varying column/hour effect is strong, since in non-business days, the way that daily hours affecting the taxi traffic can change drastically over time as compared to the same when Manhattan zones are considered. For demonstration purpose, we plot both $\hat{\beta}_{t,2}$ and $\hat{\beta}_{t,18}$ in Fig. 12, where the former series features the mid-night effects and the latter features the night-life effects. Both series demonstrate

Table 6
Estimated loading matrix $\widehat{\mathbf{Q}}_2$ using MEFM, after scaling. Magnitudes larger than 6 are highlighted in red.

0_{am}		2		4		6		8		10		12_{pm}		2		4		6		8		10		12_{am}
1	−2	−5	−6	-7	-7	-7	−6	−5	−3	0	3	5	6	6	5	5	4	4	5	5	2	0	0	−1
2	6	5	3	1	−1	−4	−5	−6	-7	-7	−6	−5	−3	−2	−1	−2	−1	−1	2	5	8	6	6	7
3	−1	-13	-9	−6	−2	2	4	5	6	4	2	−2	−4	−5	−4	−3	−2	−1	0	2	5	4	7	9

Table 7
Estimated loading matrix $\widehat{\mathbf{Q}}_2$ using FM, after scaling. Magnitudes larger than 5 are highlighted in red.

0_{am}		2		4		6		8		10		12_{pm}		2		4		6		8		10		12_{am}
1	−5	−5	−4	−3	−2	−1	−1	−1	−2	−3	−4	−5	−5	−6	−5	−5	−5	−5	-6	-6	-6	−5	−5	−5
2	5	7	7	5	4	2	0	−2	−4	-6	-6	-6	-6	−5	−4	−4	−3	−3	−2	1	4	4	5	6
3	1	-13	-10	-9	-6	−3	−1	0	1	0	−1	−3	−3	−3	−3	−2	−1	0	3	6	6	6	8	11

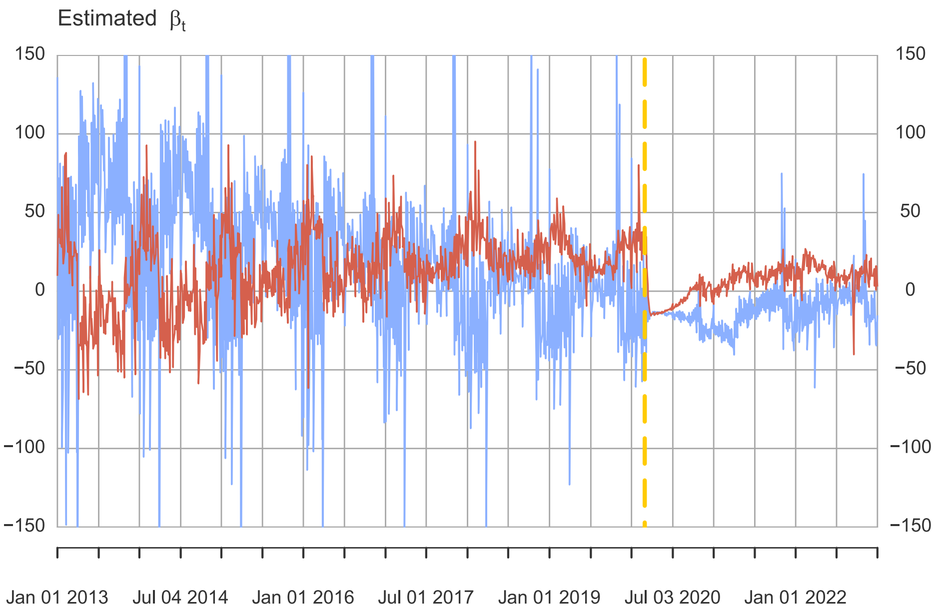

Fig. 12. Plot of the estimated hour effects for periods from 1 a.m. to 2 a.m. (in blue) and from 5 p.m. to 6.p.m. (in red). The date for the first confirmed case of COVID-19 in New York is also shown (dotted yellow vertical line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

obvious seasonality before COVID-19 as indicated on the plot. It is clear that the onset of COVID-19 serves as a change point for both series, and hence both are non-stationary. We also present Fig. 14 to provide a complete view of the estimated hour effects, and also Fig. 13 to provide the two hourly effects in Fig. 12 but relative to the estimated grand mean. They clearly deliver four pieces of information as follows:

(1) The usual rush hours, manifested by the high ridges, are 11 a.m.–1 p.m. and 0 a.m.–1 a.m.
(2) The traffic spikes in the midnight are seasonal in each October (except in 2020), corroborated by the blue time series in Figs. 12 and 13. This is potentially due to Manhattan's vibrant Halloween activities, e.g., the iconic Village Halloween Parade which actually was only canceled in October 2020.
(3) The pattern of hour effects almost vanished after COVID-19 measures in March 2020, except that the October spike revived after 2020 but with smaller magnitudes. This reflects the relative number of taxi travels in different hours follows a much different dynamics compared to before March 2020, when NYC started to shutdown, irrespective of zones in Manhattan. This can be seen in Fig. 12 also. In particular from Fig. 12 after March 2020, relative number of taxi travels between 1 to 2 a.m. is almost always lower than between 5 to 6 p.m. (apart from Octobers from 2021 onwards), which is not the case before March 2020.
(4) From Fig. 13, both hourly effects are gradually increasing relative to the grand mean until March 2020, showing that taxi traffic is becoming relatively more concentrated in these two hours irrespective of zones. They both start to drop and vary significantly less after March 2020, and gradually drop to close to 0 after 2022. These align with COVID-19 measures taking place when many activities start to cease, irrespective of the hour of the day.

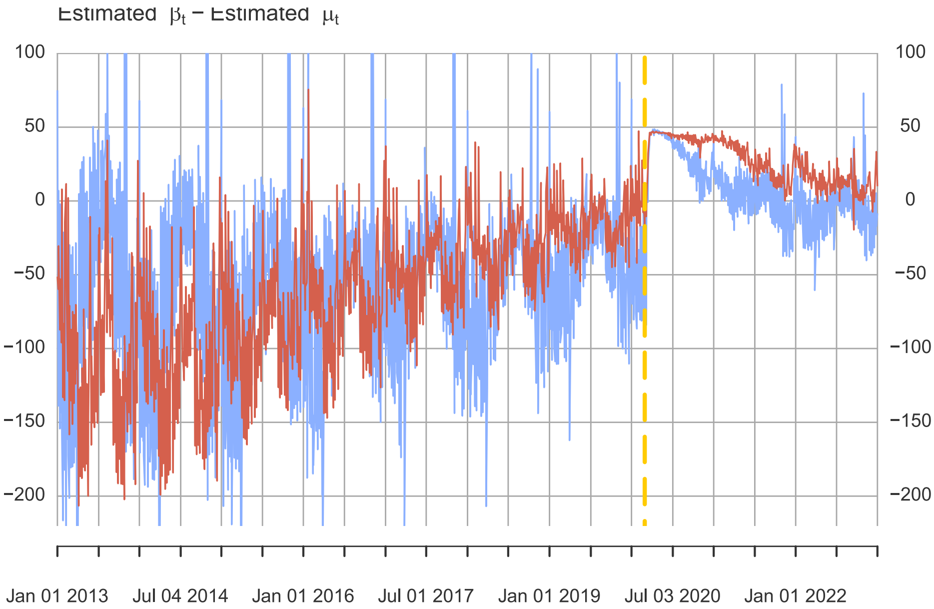

Fig. 13. Plot of the estimated hour effects minus the estimated grand mean for periods from 1 a.m. to 2 a.m. (in blue) and from 5 p.m. to 6.p.m. (in red). The date for the first confirmed case of COVID-19 in New York is also shown (dotted yellow vertical line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

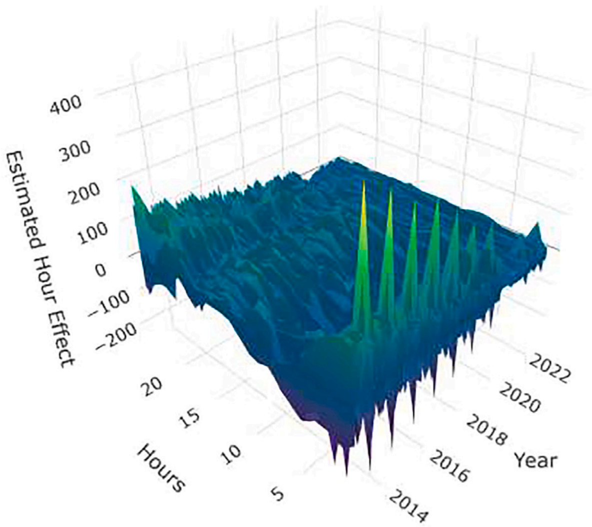

Fig. 14. A 3D-plot of the estimated hour effects.

To further illustrate the strength of MEFM, we compare both estimation and prediction accuracy of MEFM with core factor rank $(2, 2)$ to different FM's with ranks $(3, 3)$, $(3, 4)$, $(4, 3)$, $(4, 4)$, $(4, 5)$ and $(5, 4)$. We measure the estimation performance by the root mean squared error (RMSE) of the fitted data computed as

$$\sqrt{\frac{\sum_{t=1}^{1133} \|\hat{\mathbf{Y}}_t - \mathbf{Y}_t\|_F^2}{24 \cdot 69 \cdot 1133}},$$

where $\hat{\mathbf{Y}}_t$ is the fitted matrix at time t of the corresponding model. For prediction, we employ a rolling window of 500 timestamps. The core factors are modeled by VAR, whereas the grand mean and each entry of the main effects is modeled by AR. For both VAR and AR, the order is selected by minimizing AIC over all possible fitted models with order from 1–10. The prediction accuracy is

Table 8
Performance comparison between MEFM and FM on the taxi data.

Factor rank	MEFM	FM						
	(2, 2)	(3, 3)	(3, 4)	(4, 3)	(4, 4)	(4, 5)	(5, 4)	(5, 5)
RMSE	36.5	41.0	38.2	40.0	36.7	34.7	35.3	33.1
RMSPE	36.3	37.2	36.3	36.4	36.3	36.8	36.1	39.7

measured by the root mean squared prediction error (RMSPE), computed as

$$\sqrt{\frac{\sum_{t=501}^{1133} \|\widehat{\mathbf{Y}}_t - \mathbf{Y}_t\|_F^2}{24 \cdot 69 \cdot 633}},$$

where $\widehat{\mathbf{Y}}_t$ is the prediction based on parameters estimated from $\{\mathbf{Y}_{t-500}, \dots, \mathbf{Y}_{t-1}\}$. The results reported in Table 8 show that MEFM has comparable performance, both in estimation and prediction, to FM with more number of factors used. In particular, MEFM with rank $(2, 2)$ dominates FM with number of factors $(4, 4)$. On the other hand, although FM's with ranks $(4, 5)$, $(5, 4)$ and $(5, 5)$ estimate the data more accurately, their soaring out-of-sample prediction errors imply a potential overfitting and hence MEFM appears to be a more appropriate model.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

Proofs of all the theorems in this paper can be found in the supplement of this paper at http://stats.lse.ac.uk/lam/Supp-MEFM.pdf. Instruction in using our R package MEFM can be found http://stats.lse.ac.uk/lam/A-short-introduction-to-MEFM.html here.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2025.106105.

References

Ahn, S.C., Horenstein, A.R., 2013. Eigenvalue ratio test for the number of factors. Econometrica 81 (3), 1203–1227.
Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71 (1), 135–171.
Banks, J., Garza-Vargas, J., Kulkarni, A., Srivastava, N., 2023. Pseudospectral Shattering, the Sign Function, and Diagonalization in Nearly Matrix Multiplication Time. vol. 23, (6), Foundations of Computational Mathematics, pp. 1959–2047.
Cen, Z., Lam, C., 2025. Tensor time series imputation through tensor factor modelling. J. Econometrics 249, 105974.
Chang, J., He, J., Yang, L., Yao, Q., 2023. Modelling matrix time series via a tensor CP-decomposition. J. R. Stat. Soc. Ser. B Stat. Methodol. 85 (1), 127–148.
Chen, E.Y., Fan, J., 2023. Statistical inference for high-dimensional matrix-variate factor models. J. Amer. Statist. Assoc. 118 (542), 1038–1055.
Chen, W., Lam, C., 2024. Rank and factor loadings estimation in time series tensor factor model by pre-averaging. Ann. Stat. 52 (1), 364–391.
Chen, E.Y., Tsay, R.S., Chen, R., 2020. Constrained factor models for high-dimensional matrix-variate time series. J. Amer. Statist. Assoc. 115 (530), 775–793.
Chen, R., Xiao, H., Yang, D., 2021. Autoregressive models for matrix-valued time series. J. Econometrics 222 (1, Part B), 539–560, Annals Issue: Financial Econometrics in the Age of the Digital Economy.
Chen, R., Yang, D., Zhang, C.-H., 2022. Factor models for high-dimensional tensor time series. J. Amer. Statist. Assoc. 117 (537), 94–116.
Fan, J., Yao, Q., 2003. Nonlinear Time Series: Nonparametric and Parametric Methods. Springer-Verlag, New York.
Guan, L., 2023. Smooth and probabilistic parafac model with auxiliary covariates. J. Comput. Graph. Statist. 1–13.
H.-F., Zhang, 2024. Additive autoregressive models for matrix valued time series. J. Time Series Anal. 45 (3), 398–420.
He, Y., Kong, X., Trapani, L., Yu, L., 2023. One-way or two-way factor model for matrix sequences? J. Econometrics 235 (2), 1981–2004.
He, Y., Kong, X., Yu, L., Zhang, X., Zhao, C., 2024. Matrix factor analysis: From least squares to iterative projection. J. Bus. Econom. Statist. 42 (1), 322–334.
Lam, C., Yao, Q., 2012. Factor modeling for high-dimensional time series: Inference for the number of factors. Ann. Statist. 40 (2), 694–726.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix.. Econ. 55 (3), 703–708.
Pelger, M., 2019. Large-dimensional factor modeling based on high-frequency observations. J. Econometrics 208 (1), 23–42, Special Issue on Financial Engineering and Risk Management.
Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. J. Amer. Statist. Assoc. 97 (460), 1167–1179.
Tsay, R.S., 2023. Matrix-variate time series analysis: A brief review and some new developments. Int. Stat. Rev. http://dx.doi.org/10.1111/insr.12558.
Wang, D., Liu, X., Chen, R., 2019. Factor models for matrix-valued high-dimensional time series. J. Econometrics 208 (1), 231–248, Special Issue on Financial Engineering and Risk Management.
Wu, S., Bi, P., 2023. Autoregressive moving average model for matrix time series. Stat. Theory Relat. Fields 7 (4), 318–335.
Zhang, B., Pan, G., Yao, Q., Zhou, W., 2024. Factor modeling for clustering high-dimensional time series. J. Amer. Statist. Assoc. 119 (546), 1252–1263.