Rank determination in tensor factor model†∗

Yuefeng Han, Rong Chen and Cun-Hui Zhang

Department of Statistics, Rutgers University, Piscataway, NJ 08854, USA
e-mail: yuefeng.han@rutgers.edu; rongchen@stat.rutgers.edu; czhang@stat.rutgers.edu

Abstract: Factor model is an appealing and effective analytic tool for high-dimensional time series, with a wide range of applications in economics, finance and statistics. This paper develops two criteria for the determination of the number of factors for tensor factor models where the signal part of an observed tensor time series assumes a Tucker decomposition with the core tensor as the factor tensor. The task is to determine the dimensions of the core tensor. One of the proposed criteria is similar to information based criteria of model selection, and the other is an extension of the approaches based on the ratios of consecutive eigenvalues often used in factor analysis for panel time series. Theoretically results, including sufficient conditions and convergence rates, are established. The results include the vector factor models as special cases, with an additional convergence rates. Simulation studies provide promising finite sample performance for the two criteria.

MSC2020 subject classifications: Primary 62H25, 62H12; secondary 62F07.

Keywords and phrases: High-dimensional tensor data, factor model, rank determination, eigenvalues, Tucker decomposition.

Received March 2021.

1. Introduction


† Han’s research is supported in part by National Science Foundation grant IIS-1741390.
Zhang’s research is supported in part by NSF grants DMS-1721495, IIS-1741390, CCF-1934924 and DMS-2052949 Chen’s research is supported in part by National Science Foundation grants DMS-1503409, DMS-1737857, IIS-1741390 and DMS-2052949

Cun-Hui Zhang is the corresponding author.
are also used to evaluate the impacts of various policies; see, e.g., Bai, Li and Ouyang (2014), Ouyang and Peng (2015) and Li and Bell (2017). Recently, large matrix or tensor (multi-dimensional array) data has become ubiquitous. Wang, Liu and Chen (2019) proposed a matrix factor model and applied it to matrix-valued financial data. Chen, Yang and Zhang (2021) analyzed the multi-category import-export network data via tensor factor model.

A critical step in building a factor model is to correctly specify the number of factors used in the model. Estimation and forecasting procedures are all dependent on the number of factors. Moreover, in some cases the number of factors may have some crucial economic interpretations and important theoretical consequences. For example, in finance and macroeconomics, it provides the number of sources of nondiversifiable risk or the fundamental shocks driving the macroeconomic dynamics. See, Forni and Reichlin (1998), Stock and Watson (2016), Giannone and Reichlin (2006), among others.

Over the past decades, many methods have been developed to determine the number of common factors needed for modelling high dimensional vector time series. The most widely studied approach is to utilize the behavior of the eigenvalues of the covariance matrix (see, e.g., Bai and Ng (2002)), or the singular values of the autocovariance matrix (see, e.g., Lam and Yao (2012)). By the definition of factor models, the eigenvalues or the singular values corresponding to the systematic components must increase with the number of cross-sectional units. The rest of the eigenvalues, which represents idiosyncratic components, stay bounded or remain to be zero. In the static factor model, Bai and Ng (2002) proposed to estimate the number of factors by separating diverging eigenvalues from the rest using threshold functions, in the form of an information criterion. Alternative criteria based on random matrix theory have been studied in Kapetanios (2010) and Onatski (2010) for the static factor model. Specifically Kapetanios (2010) developed sequential tests and employed a subsampling method to obtain an approximation of the asymptotic distribution of the estimated eigenvalues. Onatski (2010) constructed tests based on the empirical distribution of the eigenvalues. In addition, Onatski (2012) proposed an alternative estimator using the difference of consecutive eigenvalues. Bai and Ng (2007) and Amengual and Watson (2007) extended the work of Bai and Ng (2002) to the restricted dynamic factor model. Hallin and Liška (2007) further extended the framework to the generalized dynamic factor model through thresholding eigenvalues of the spectral density matrix. They also proposed a data-dependent method to adjust the multiplicative constant of the penalty function. Alessi, Barigozzi and Capasso (2010) introduced a tuning multiplicative constant in the penalty for dealing with approximate factor models. Kong (2017) employed similar ideas to study continuous time factor model with high frequency data. Li, Li and Shi (2017) modified Bai and Ng (2002)’s procedure to the case that the number of factors is allowed to increase with the sample size. Trapani (2018) proposed a randomized sequential test procedure to determine the number of factors.

An alternative approach is to study the ratio of each pair of adjacent eigenvalues, with the insight that ratio of the smallest eigenvalue among these cor-
responding to the system component and the largest eigenvalue among these corresponding to the idiosyncratic component goes to infinity. Under stationary conditions, Ahn and Horenstein (2013) developed such an estimator based on the sample covariance matrix. Lam and Yao (2012) used such a ratio based estimator based on singular values of the autocovariance matrix, under an alternative definition of factor models proposed in Pan and Yao (2008) and Lam, Yao and Bathia (2011).

Other than the eigenvalue-based methods, Ye and Weiss (2003) developed an eigenvector based order determination procedure. Luo and Li (2016) proposed a new estimator that combines both the eigenvalues and the bootstrap eigenvector variability. Jung, Lee and Ahn (2018) suggested to sequentially test skewness of the squared lengths of residual scores that are obtained by removing leading principal components. However, these works assumed that the data are temporally independent, which are unlikely to hold for economic data.

These studies all focus on panel (vector) time series. Recently there is a growing interest in analyzing matrix- or tensor-valued time series, as such time series is encountered more and more frequently in applications, including Fama-French 10 by 10 series (Wang, Liu and Chen, 2019), a set of economic indicator series among a set of countries (Chen, Xiao and Yang, 2021), multi-category international trading volume series (Chen and Chen, 2019, Hoff, 2011), multi-type international action counts among a group of countries (Hoff, 2015), sequence of realized covariance matrices (Kim and Fan, 2019, Lunde, Shephard and Sheppard, 2016), sequence of gray-scale face recognition images (Chen and Fan, 2021), dynamic networks (Barabási and Albert, 1999, Jiang, Li and Yao, 2020), dynamic human brain transcriptome data (Liu, Yuan and Zhao, 2017), multivariate spatial-temporal climate series (Chen et al., 2020), neuroimaging data (Zhang, 2019, Zhou, Li and Zhu, 2013). Factor model is again developed as an effective dimension reduction tool (Chen, Yang and Zhang, 2021, Han et al., 2020, Wang, Liu and Chen, 2019). Same as for the vector factor models, it is important to determine the number of factors in these models.

In this paper, we consider the determination of the dimension of the core tensor factor in the tensor factor model in Chen, Yang and Zhang (2021) and Han et al. (2020), which assumes the form

\[ X_t = M_t + \mathcal{E}_t = \mathcal{F}_t \times_1 A_1 \times_2 \cdots \times_K A_K + \mathcal{E}_t. \]

Similar to Lam and Yao (2012), the noise tensor \( \mathcal{E}_t \) is assumed to be a white tensor process with potentially strong contemporary correlations among the elements of the noise tensor, and all common dynamics is absorbed in the signal process \( M_t \). This model setting is different from the approximate factor model in Bai and Ng (2002) and the dynamic factor model in Hallin and Liška (2007), in which the noise process is allowed to have weak auto-correlations, but with strong restriction on the contemporary correlation.

Chen, Yang and Zhang (2021) and Han et al. (2020) studied the estimation procedures of the tensor factor model, assuming the ranks of the core tensor \( \mathcal{F}_t \) is given, with some ad hoc rank determination suggestions. In this paper we
formally propose two criteria for specifying the ranks of the core factor process, which we name “the information criterion” (IC) and “the eigenvalue ratio” (ER). They are all based on examining the eigenvalues of the sample cross-autocorrelation of the observed tensor time series, utilizing the whiteness property of the noise process. The IC estimators aim at truncating eigenvalues, which is similar to the information criteria in vector factor models (e.g., Bai and Ng (2002) and Hallin and Liška (2007)). The ER estimators are obtained by minimizing the ratio of two adjacent eigenvalues arranged in ascending order, extending the standard ER estimator in Lam and Yao (2012) and Wang, Liu and Chen (2019) with an added small penalty term in both the numerator and denominator of the ratio. The penalty term behaves like a lower bound correction to the true zero eigenvalues. We adopt similar ideas of the TOPUP and TIPUP procedures of Chen, Yang and Zhang (2021), and their corresponding iterative versions, iTOPUP and iTIPUP, of Han et al. (2020), to construct sample auto-cross-moments. Our theoretical and empirical investigations show that estimators based on the iterative algorithms are much better than that based on the non-iterative ones, as the iterative algorithms significantly improve the estimation accuracy of the eigenvalues. The finite sample properties of the IC and ER criteria are also good. The empirical evidences show that the best estimators in tensor factor model are the IC and ER estimators based on iTIPUP, under some mild conditions on the level of signal cancellation typically associated with the TIPUP based procedures.

This paper is organized as follows. Section 3.1 briefly describes the tensor factor model and the corresponding estimation procedures proposed in Chen, Yang and Zhang (2021) and Han et al. (2020). Section 3.2 introduces the criteria for determining the ranks of the core tensor factor process, and their iterative versions. Section 4 investigates theoretical properties of the proposed estimators. Section 5 presents simulation studies of the finite sample properties of the proposed methods. Real data analysis is given in Section 6. Discussions are provided in Section 7. All technical details are relegated to the Supplementary Material.

2. General order determination criteria of semipositive definite matrices

In this section, we first propose two general order determination criteria based on the properties of the estimated eigenvalues of a semipositive matrix.

Let $\hat{W}$ be a $p \times p$ symmetric and non-negative definite matrix, which is a sample version of a true $p \times p$ symmetric and non-negative definite matrix. We assume $W = \mathbb{E}\hat{W}$. Also let $\hat{\lambda}_j$ be the eigenvalues of $\hat{W}$ such that $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_p$. Let $\lambda_1 \geq \ldots \geq \lambda_r > \lambda_{r+1} = \ldots = \lambda_p = 0$ be the eigenvalues of $W$. Note that the rank of $W$ is $r$.

Let $m^* < p$ be a predefined upper bound and functions $G(\hat{W})$ and $H(\hat{W})$ be some appropriate positive penalty functions. We propose the following two
quantities

\[
\text{IC}(\hat{W}) = \arg\min_{0 \leq m \leq m^*} \left\{ \sum_{j=m+1}^{p} \hat{\lambda}_j + mG(\hat{W}) \right\}
\]

(1)

\[
\text{ER}(\hat{W}) = \arg\min_{1 \leq m \leq m^*} \frac{\hat{\lambda}_{m+1} + H(\hat{W})}{\hat{\lambda}_m + H(\hat{W})}.
\]

(2)

The first criterion in (1) is similar to an information criterion as its first term mimics the residual sum of squares of using a rank \( m \) matrix to approximate the matrix \( \hat{W} \) while the second term \( mG(\hat{W}) \) penalizes the model complexity \( m \). We will call it \text{the information criterion (IC)}. The second criterion in (2) uses the ratio of two adjacent eigenvalues of \( \hat{W} \), with a small penalty term \( H(\hat{W}) \) added to both the numerator and denominator. We will call it \text{the eigen-ratio criterion (ER)}.

**Remark 1** (\text{The information criterion}). Note that, for a given \( m \), the principle components can be viewed as solutions of an optimization problem in which the “sum of squared residuals” is minimized,

\[
\hat{U}_m = \arg\min_{\hat{U}_m} \text{tr} \left\{ \left( I - U_m U_m^\top \right) \hat{W} \right\}.
\]

(3)

Note that \( \text{tr} \left\{ \left( I - \hat{U}_m \hat{U}_m^\top \right) \hat{W} \right\} = \sum_{j=m+1}^{p} \hat{\lambda}_k \). It plays the role of residual sum of squares classically appearing in information criterion methods. Criterion (1) has a structure comparable to that of Bai and Ng (2002) and Hallin and Liška (2007). For vector factor models, the method proposed in Bai and Ng (2002) is the same of the IC criterion with \( \hat{W} \) being the sample covariance matrix, while that in Hallin and Liška (2007) used spectral density matrix estimation. The penalty \( mG(\hat{W}) \) is intimately related to the rate of convergence of the non-divergent eigenvalues, when \( \hat{W} \) is estimated from a set of data with diverging dimensions, and balances between overestimation and underestimation.

**Remark 2** (\text{Eigen-ratio criterion}). Different from the standard ER estimator in Lam and Yao (2012), we add a penalty term \( H(\hat{W}) \) to both the numerator and denominator. The intuition behind \( H(\hat{W}) \) is as follows. Since \( \hat{W} \) is a noisy version (an estimator) of \( W \) of rank \( r \), all estimated eigenvalues \( \hat{\lambda}_j \) \( (r+1 \leq j \leq p) \) correspond to the zero eigenvalues of \( W \). Hence the ratio \( \hat{\lambda}_{j+1}/\hat{\lambda}_j \) \( (j > r) \) theoretically can be arbitrary small. The penalty \( H(\hat{W}) \) provides a lower bound correction to \( \hat{\lambda}_j \) \( (r+1 \leq j \leq p) \). When it is of a proper order, we can ensure that the ratio \( (\hat{\lambda}_{m+1} + H(\hat{W}))/(\hat{\lambda}_m + H(\hat{W})) \) goes to zero when \( m = r \) (the true rank), while all other such ratios are asymptotically bounded from below. In vector factor models, Ahn and Horenstein (2013) exploited the ratio of eigenvalues of sample covariance matrix to determine the number of factors. Non-divergent eigenvalues therein are bounded below by a positive number asymptotically, as
long as the eigenvalues of covariance matrix of idiosyncratic noises are bounded away from zero. Our criterion (2) has a similar flavor.

Here we show a consistency result for the general estimator. To be more precise, let $\hat{W}^{(n)}$ and $W^{(n)}$ be two sequences of semi-positive symmetric matrices, with $W^{(n)} = EW^{(n)}$ and $n$ be an index associated with the sample size and dimension. Also let $\hat{\lambda}_j^{(n)}$ be the eigenvalues of $\hat{W}^{(n)}$ such that $\hat{\lambda}_1^{(n)} \geq \hat{\lambda}_2^{(n)} \geq \ldots \geq \hat{\lambda}_p^{(n)}$. Let $\lambda_1^{(n)} \geq \ldots \geq \lambda_r^{(n)} > \lambda_{r+1}^{(n)} = \ldots = \lambda_p^{(n)} = 0$ be the eigenvalues of $W$. We assume $\lambda_r^{(n)} \to \infty$ as $n \to \infty$. The following proposition provides the sufficient conditions for the consistency of the IC and ER estimators in (1) and (2), respectively. It provides a guideline of choosing proper penalty functions $G(\cdot)$ and $H(\cdot)$ in order determination for any generic $\hat{W}^{(n)}$.

**Proposition 1.** Assume $|\hat{\lambda}_1^{(n)} - \lambda_1^{(n)}| = o_p(\lambda_1^{(n)}) + O_p(\gamma_n)$ and $|\hat{\lambda}_r^{(n)} - \lambda_r^{(n)}| = o_p(\lambda_r^{(n)}) + O_p(\gamma_n)$, and $|\hat{\lambda}_j^{(n)} - \lambda_j^{(n)}| = O_p(\beta_n)$ for all $j > r$. Then,

(i) $\mathbb{P}(\text{IC}(\hat{W}^{(n)}) = r) \to 1$, provided that $(G(\hat{W}^{(n)}) + \gamma_n)/\hat{\lambda}_r^{(n)} \to 0$ and $G(\hat{W}^{(n)})/\beta_n \to \infty$;

(ii) $\mathbb{P}(\text{ER}(\hat{W}^{(n)}) = r) \to 1$, provided that $(H(\hat{W}^{(n)}) + \beta_n)/((\lambda_1^{(n)})^2/\lambda_1^{(n)}) \to 0$, $\gamma_n/\lambda_1^{(n)} \to 0$ and $H(\hat{W}^{(n)})/(\beta_n^2/\lambda_1^{(n)}) \to \infty$.

In the conditions of Proposition 1, $\gamma_n$ represents the convergence rate of the sample eigenvalues corresponding to the non-zero eigenvalues of $W^{(n)}$, and $\beta_n$ represents the rate of the sample eigenvalues corresponding to the zero eigenvalues of $W^{(n)}$. For example, under the strong factor model of Lam and Yao (2012)’s setting, $\gamma_n = p^2T^{-1/2}$ and $\beta_n = p^2T^{-1}$, where $T$ is the sample size.

**Remark 3.** In our model setting, $\lambda_1^{(n)} = \ldots = \lambda_p^{(n)} = 0$ and our objective is to separate the zero and non-zero eigenvalues. The proposition holds for general spiked eigenvalue detection as well. Specifically, let $\lambda_1^{(n)} \geq \ldots \geq \lambda_r^{(n)} > \gamma_n > \lambda_{r+1}^{(n)} \geq \ldots \geq \lambda_p^{(n)} \geq 0$ be the eigenvalues of $W^{(n)}$, where $\lambda_{r+1}^{(n)}, \ldots, \lambda_p^{(n)}$ are called non-spiked eigenvalues (Cai, Han and Pan, 2020). Again it is assumed that $\lambda_r^{(n)} \to \infty$ as $n \to \infty$. Then Proposition 1 holds when $\gamma_n$ and $\beta_n$ are the convergence rate of the sample eigenvalues corresponding to the spiked and non-spiked eigenvalues of $W^{(n)}$, respectively. Note that the approaches of Bai and Ng (2002), Amengual and Watson (2007), Hallin and Liška (2007), Lam and Yao (2012) and Ahn and Horenstein (2013) all fit in this generic setting or its variants, with various forms of the penalty functions $G(\hat{W})$ and $H(\hat{W})$ to distinguish $\hat{\lambda}_r^{(n)}$ from $\hat{\lambda}_{r+1}^{(n)}$. For example, Bai and Ng (2002) suggest to use $G_1 = p^{-1}T^{-1}(p + T) \log(pT(p + T)^{-1})$, $G_2 = p^{-1}T^{-1}(p + T) \log(\min\{p, T\})$, or $G_3 = \max\{p^{-1}, T^{-1}\} \log(\min\{p, T\})$, where $T$ is the sample size and $p$ is the number of variables.

**Remark 4.** When the dimensions $d_k$ are large, estimating eigenvalues of a matrix using its sample version is in general very difficult and potentially inaccurate. However, to determine the number of factors, only the leading eigenvalues
need to be estimated relatively accurately to achieve the purpose, which requires relatively mild conditions on the sample version of the matrix.

For the specific problems such as the tensor factor model problem we focus there, a detailed analysis of the rates $\gamma_n$ and $\beta_n$ is needed to construct the penalty functions $G(\cdot)$ and $H(\cdot)$ and to establish the consistency of the rank estimators. In fact one can establish the convergence rate with a more detailed analysis beyond the simple consistency results in Proposition 1, as we will do for the tensor factor model.

3. Order determination criteria for tensor factor models

3.1. The model

Here we briefly introduce the tensor factor model setup in Chen, Yang and Zhang (2021) and Han et al. (2020). A tensor factor model can be written as

$$X_t = M_t + \mathcal{E}_t = \mathcal{F}_t \times_1 A_1 \times_2 \cdots \times_K A_K + \mathcal{E}_t, \quad (4)$$

where $X_t \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is the observed tensor at time $t$, the core tensor $\mathcal{F}_t$ is the unobserved latent tensor factor process of dimension $r_1 \times \cdots \times r_K$, $A_k$ are the deterministic loading matrix of size $d_k \times r_k$ and $r_k \ll d_k$, and $\mathcal{E}_t$ is the idiosyncratic noise components of $X_t$, which is assumed to be a white process. Here the $k$-mode product of $X \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ with a matrix $U \in \mathbb{R}^{d_k' \times d_k}$, denoted as $X \times_k U$, is an order $K$-tensor of size $d_1 \times \cdots \times d_k - 1 \times d_k' \times d_k + 1 \times \cdots \times d_K$ such that

$$(X \times_k U)_{i_1,\ldots,i_{k-1},j,i_{k+1},\ldots,i_K} = \sum_{i_k=1}^{d_k} X_{i_1,i_2,\ldots,i_{k-1},j,i_{k+1},\ldots,i_K} U_{j,i_k}.$$  

The core tensor $\mathcal{F}_t$ is usually much smaller than $X_t$ in dimension. We also assume that the rank of $A_k$ is $r_k$. Otherwise $X_t$ in (4) may be expressed equivalently with a lower-dimensional factor process. The parameters $r_1,\ldots,r_K$ are assumed to be fixed but unknown. For more details of the tensor factor model (4), see Chen, Yang and Zhang (2021) and Han et al. (2020).

It is obvious that the loading matrices $A_k$ are not identifiable in Model (4). Model (4) is unchanged if we replace $(A_1,\ldots,A_K, \mathcal{F}_t)$ by $(A_1 H_1,\ldots,A_K H_K, \mathcal{F}_t \times_{k=1}^{K} H_k^{-1})$ for any invertible $r_k \times r_k$ matrix $H_k$. However, the linear space spanned by the columns of $A_k$, called the factor loading space, is uniquely defined. Assume $A_k$ has a SVD representation $A_k = U_k A_k V_k^T$. Then, the factor loading space of $A_k$ can be represented by the orthogonal projection $P_k$,

$$P_k = P_{A_k} := A_k (A_k^T A_k)^{-1} A_k^T = U_k U_k^T. \quad (5)$$

3.2. Rank selection criteria for tensor factor models

The two criteria introduced in Section 2 can be used to estimate the number of factors in the tensor factor model (4), using properly constructed matrices $W$ and $\hat{W}$. Particularly, we will study the following four constructions.
Since they have been proposed and used for loading space estimation in Chen, Yang and Zhang (2021) and Han et al. (2020), we adopt the same names to represent them.

**I (TOPUP):** Let

\[
\text{TOPUP}_k(X_{1:T}) = \text{mat}_1 \left( \sum_{t=h+1}^{T} \frac{\text{mat}_k(X_{t-h}) \otimes \text{mat}_k(X_t)}{T-h}, \ h = 1, \ldots, h_0 \right),
\]

and

\[
\hat{W}_k = \hat{W}_k(X_{1:T}) := \text{TOPUP}_k(X_{1:T})(\text{TOPUP}_k(X_{1:T}))^T,
\]

where \(\otimes\) is the tensor product such that, for any \(A \in \mathbb{R}^{m_1 \times m_2 \times \cdots \times m_K}\) and \(B \in \mathbb{R}^{r_1 \times r_2 \times \cdots \times r_N}\),

\[
(A \otimes B)_{i_1 \ldots i_K j_1 \ldots j_N} = (A)_{i_1 \ldots i_K} (B)_{j_1 \ldots j_N},
\]

and \(\text{mat}_k\) is the tensor unfolding (into a matrix) operation along mode-\(k\) of a tensor. Here we emphasize that \(\hat{W}_k\) is constructed using \(X_{1:T} = (X_1, \ldots, X_T)\). The constant \(h_0\) is a (small) predetermined integer and the sum over \(h\) in \(\hat{W}_k\) is to accumulate the information from different time lags \(h\). The rank of its population version can be shown to be \(r_k\) under certain conditions, hence we can use the IC and ER estimators presented in Section 2 to determine \(r_k\).

**II (TIPUP):** Define a \(d_k \times (d_k h_0)\) matrix as

\[
\text{TIPUP}_k(X_{1:T}) = \text{mat}_1 \left( \sum_{t=h+1}^{T} \frac{\text{mat}_k(X_{t-h}) \text{mat}_k^T(X_t)}{T-h}, \ h = 1, \ldots, h_0 \right),
\]

which replaces the tensor product in \(\text{mat}_1(\text{TOPUP}_k(X_{1:T}))\) by the inner product. Let

\[
\hat{W}'_k = \hat{W}'_k(X_{1:T}) := (\text{TIPUP}_k(X_{1:T}))(\text{TIPUP}_k(X_{1:T})^T).
\]

**III and IV (iTOPUP and iTIPUP):** Han et al. (2020) proposed an iterative procedure to estimate \(\hat{U}_k, k = 1, \ldots, K\) in (5), based on either TOPUP or TIPUP procedure. Briefly, at \(i\)-th iteration, suppose we have obtained an estimate of the ranks \(r_k^{(i-1)}\) \((k = 1, \ldots, K)\) and their corresponding \(\hat{U}_k^{(i-1)}\) at \((i-1)\)-th iteration, we calculate the orthogonal projections of \(X_t, 1 \leq t \leq T\), to obtain

\[
Z_{k,t}^{(i)} = X_t \times_1 (\hat{U}_1^{(i)})^T \times_2 \cdots \times_{k-1} (\hat{U}_{k-1}^{(i)})^T \times_{k+1} (\hat{U}_{k+1}^{(i)})^T \cdots \times_K (\hat{U}_K^{(i)})^T.
\]

Note that \(Z_{k,t}^{(i)}\) uses projection of \(X_t\) on all modes, except mode-\(k\). The initial ranks \(r_k^{(0)}\) \((k = 1, \ldots, K)\) and their corresponding \(\hat{U}_k^{(0)}\) can be obtained
through the non-iterative TOPUP and TIPUP procedure. Let $Z_{k,1:T}^{(i)} = (Z_{k,1}^{(i)}, \ldots, Z_{k,T}^{(i)})$, and define

$$
\begin{align*}
\hat{W}_k^{(i)} &= \hat{W}_k(Z_{k,1:T}^{(i)}) := \text{TOPUP}_k(Z_{k,1:T}^{(i)})(\text{TOPUP}_k(Z_{k,1:T}^{(i)}))^\top, \\
\hat{W}_k^{(∗)} &= \hat{W}_k(Z_{k,1:T}^{(i)}) := \text{TOPUP}_k(Z_{k,1:T}^{(i)})(\text{TOPUP}_k(Z_{k,1:T}^{(i)}))^\top.
\end{align*}
$$

The iterative procedure is motivated by the observation that $Z^{(j)}$ is a $r_1 \ldots r_{k-1}d_k r_{k+1} \ldots r_K$ tensor, much smaller than $X_t$, which is $d_1 \ldots d_K$ tensor. Hence $A_k$ can be estimated more accurately if all $A_j$ ($j \neq k$) are given in advance or can be estimated accurately, since the convergence rate now depends on $(d_k/r_k) \prod_{k=1}^K r_i$ rather than $\prod_{k=1}^K d_i$.

The IC and ER estimators are constructed by replacing $\hat{W}$ in (1) and (2) with $\hat{W}_k, \hat{W}_k^{(∗)}, \hat{W}_k^{(i)}$ and $\hat{W}_k^{(∗)}$. This yields eight different criteria, summarized in Table 1. Again, we use the same names of the procedures as that in Chen, Yang and Han et al. (2021) to represent the various constructions of $\hat{W}$. For the iterative procedures, we start with an initial rank estimate $r_k^{(0)}$, $k = 1, \ldots, K$ and estimate the ranks through iteration until convergence. See remark below for setting the initial ranks and the stopping criteria.

**Remark 5.** In our theories, we fix $m^*$ in (1) and (2) as a finite constant. However, in practice, we may use, for example, $m^* = p/2$. We do not recommend to extend the search up to $p$, as the minimum eigenvalue is likely to be practically 0, especially when $T$ is small and $d_k$ is large.

**The choice of the penalty function** $G(\cdot)$ and $H(\cdot)$: Both criteria essentially try to distinguish the smallest (true) non-zero eigenvalue from the true zero eigenvalue using noisy estimators of the eigenvalues. Hence the penalty function is closely related to the amount of error in the eigenvalue estimation and the strength of the smallest (true) non-zero eigenvalue. We consider the following penalty functions $G(\cdot) = g_k(d,T)$:

$$
\begin{align*}
g_{k,1}(d,T) &= \frac{h_0 d^{2-2\nu}}{T} \log \left( \frac{dT}{d+T} \right), \\
g_{k,2}(d,T) &= \frac{h_0 d^{2-2\nu}}{T} \left( \frac{1}{T} + \frac{1}{d} \right) \log \left( \frac{dT}{d+T} \right),
\end{align*}
$$

<table>
<thead>
<tr>
<th>Estimation method used</th>
<th>IC</th>
<th>ER</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-iterative TOPUP</td>
<td>$\hat{r}_k^{(\text{IC})} = \hat{r}_k^{(0)}(\text{IC}) = \text{IC}($</td>
<td>$\hat{r}_k^{(\text{ER})} = \hat{r}_k^{(0)}(\text{ER}) = \text{ER}($</td>
</tr>
<tr>
<td>non-iterative TIPUP</td>
<td>$\hat{r}_k^{(\text{IC})} = \hat{r}_k^{(0)}(\text{IC}) = \text{IC}($</td>
<td>$\hat{r}_k^{(\text{ER})} = \hat{r}_k^{(0)}(\text{ER}) = \text{ER}($</td>
</tr>
<tr>
<td>i-th iteration of iTOPUP</td>
<td>$\hat{r}_k^{(i)}(\text{IC}) = \text{IC}($</td>
<td>$\hat{r}_k^{(i)}(\text{ER}) = \text{ER}($</td>
</tr>
<tr>
<td>i-th iteration of iTIPUP</td>
<td>$\hat{r}_k^{(i)}(\text{IC}) = \text{IC}($</td>
<td>$\hat{r}_k^{(i)}(\text{ER}) = \text{ER}($</td>
</tr>
</tbody>
</table>
Determining the number of factors

\[ g_{k,3}(d,T) = \frac{h_0 d^{2-2\nu}}{T} \log \{ \min\{d,T\} \}, \]
\[ g_{k,4}(d,T) = h_0 d^{2-2\nu} \left( \frac{1}{T} + \frac{1}{d} \right) \log \{ \min\{d,T\} \}, \]
\[ g_{k,5}(d,T) = h_0 d^{2-2\nu} \left( \frac{1}{T} + \frac{1}{d} \right) \log \{ \min\{d_k,T\} \}, \]  

(7)

where \( d = \prod_{k=1}^{K} d_k \) and \( \nu \) is a tuning parameter. Ideally \( \nu \) should be chosen to be the strength of the weakest factor (see Assumption IV in Section 4.1), though in practice we usually do not know its precise value. A more thorough discussion on this issue will be given later in Remark 10. Note that only \( g_{k,5} \) involves \( k \).

For the eigen-ratio criterion, we consider the following penalty function \( H(\cdot) = h_k(d,T) \):

\[ h_{k,1}(d,T) = c_0 h_0, \quad h_{k,2}(d,T) = h_0 d^2 \frac{d}{T^2}, \quad h_{k,3}(d,T) = h_0 d^2 \frac{T^2}{d_k^2}, \]
\[ h_{k,4}(d,T) = \frac{h_0 d^2}{T^2 d_k^2} + \frac{h_0 d^2}{T^2}, \quad h_{k,5}(d,T) = \frac{h_0 d^2}{T^2 d_k^2} + \frac{h_0 d d_k}{T^2}. \]  

(8)

where \( c_0 \) is a small constant, e.g. \( c_0 = 0.1 \). Note that the penalty functions scale with \( h_0 \), because the strength of divergent eigenvalues increases with \( h_0 \). Our theoretical analysis indicates that a better penalty function \( H(\cdot) \) should also involve the strengths of the factors, similar to \( G(\cdot) \). However, the function \( G(\cdot) \) has a much wider allowable range, and in most of the situations a simple constant function \( h_{k,1} \) is sufficient. A more detailed discussion will be given later in Remark 10.

More considerations of the iterative procedure: The non-iterative procedures estimate \( r_k (k = 1, \ldots, K) \) individually. The accuracy of \( \hat{r}_k \) does not depend on the accuracy of the estimation of the ranks in other directions. On the other hand, the iterative procedures estimate all the ranks simultaneously, hence the accuracy of estimated rank in one direction depends on that in all other directions. The iterative algorithm improves the estimation accuracy of the eigenvalues and the principal subspace because the projected tensor \( Z_{k,t}^{(i)} \) in (6) is of lower dimensional than \( X_t \). Figure 1 numerically shows that the iterative algorithms improve the accuracy of estimated true zero eigenvalues over the non-iterative algorithms. In iterative algorithms, one would need to specify all the ranks \( \hat{r}_k^{(i)} (k = 1, \ldots, K) \) in each iteration. Intuitively, an overestimated \( \hat{r}_k^{(i)} > r_k \) would still produce consistent estimators, since the non-iterative procedure, using \( \hat{r}_k^{(i)} = d_k \) for all other directions, is consistent. Our theoretical results shown later confirm that, if \( \hat{r}_k^{(i-1)} \) used is larger than the true \( r_k \), the iterative algorithm warrant the consistency of the IC and ER estimators at \( i \)-th iteration. However, an underestimated \( \hat{r}_k^{(i)} < r_k \) would potentially result in loss of signal strength hence negatively impacting the estimation in other dimensions. A precise quantification of the impact requires a more detailed investigation. But the
numerical studies show that for iteration \( i > 1 \), the performance is relatively robust by using the order obtained by the IC or ER criteria. This is partially due to the theoretical justification that, for fixed ranks \( r_k \), iterative algorithm only needs one iteration to achieve the ideal convergence rate for the estimation of the eigenvalues (see Theorem 2 later). In fact, if one has a priori information about a possible maximum (fixed) ranks of the core factor process, one could use such ranks in the iterative algorithms accordingly.

For iTOPUP procedure, we use \( \hat{r}_k^{(i)} = \hat{r}_k^{(i)}(\text{IC}) \) or \( \hat{r}_k^{(i)} = \hat{r}_k^{(i)}(\text{ER}) \) after the initial iteration \( i \geq 1 \). However, one needs to use a more conservative estimator of the rank for the initial step since \( \hat{r}_k^{(0)}(\text{IC}) \) and \( \hat{r}_k^{(0)}(\text{ER}) \) tend to be inaccurate. We suggest to use \( \hat{r}_k^{(0)} = \min\{2\hat{r}_k^{(0)}(\text{IC}), \hat{r}_k^{(0)}(\text{IC}) + 3\} \) or \( \hat{r}_k^{(0)} = \min\{2\hat{r}_k^{(0)}(\text{ER}), \hat{r}_k^{(0)}(\text{ER}) + 3\} \) by default, unless one has prior knowledge of the number of the factors. iTIPUP procedure is similar. Although it is safer to use larger initial ranks \( r_k^{(0)} \), it is often not necessary to be extremely conservative, as the initial loss of signal strength of using a rank too small can be corrected later through iterations.

In the iterative algorithms, iteration is not stopped until the convergence of both the rank estimators and the loading space estimators. Theoretical prop-
properties in Section 4 only state consistency results in each iteration step. This stopping rule is mainly suggested by simulation study. In practice, we may stop the algorithm when the estimated number of factors in current iteration is the same as that in previous iteration.

3.3. Some discussions

Remark 6. When the means of the factor processes deviate from zeros by a large margin, there is often one or several dominating factors corresponding to these non-zero means, as observed by Brown (1989), while the factors associated with the covariances become weak factors and are more difficult to identify. Assuming that the factor tensor process does not change its dimension after the deterministic means are removed, i.e. the factor tensor process is not a constant in any of its dimensions, then one should always demean the data in practice for the determination of the dimension of the factors, though not necessary for the estimation of the loading spaces, as shown by Chen and Fan (2021) that aggregating the first and second moments of the data may improve the estimation accuracy of the factor loading matrices.

Remark 7. When some $r_k = 1$, although the factor process has a reduced number of tensor modes, the proposed IC and ER methods should work well in identifying $r_k = 1$ cases. If there is no factor structure ($r_1 = \ldots = r_K = 0$), the proposed IC methods still can select the zero rank. But the ER methods need a slight modification, by constructing a new mock eigenvalue $\hat{\lambda}_{k,0}$ using the rate of the spiked eigenvalues. In the current paper, we focus on the case that $r_k$ is fixed and $d_k$ diverges. If $r_k = d_k$ along one (or even all) dimension(s), the theoretical results in Section 4 can be extended. In this case we will need to modify the IC and ER methods using the developed convergence rates of zero eigenvalues.

Remark 8 (Improved penalization for IC approaches). The information criterion (1) has the property, exploited by Hallin and Liška (2007) in the context of dynamic factor models, that a penalty function $G(\cdot) = g_k(d,T)$ leads to a consistent estimate of $r_k$ if and only if $cg_k(d,T)$ does, where $c$ is an arbitrary positive real number. Thus, multiplying the penalty by $c$ has no influence on the asymptotic performance of the identification method. However, for given finite $d$ and $T$, the value of a penalty function $g_k(d,T)$ satisfying (1) can be arbitrarily small or arbitrarily large, and this indeterminacy can affect the actual result quite dramatically. The procedures in Hallin and Liška (2007) and Alessi, Barigozzi and Capasso (2010) can also be used in tensor factor model to robustify the IC approach with an empirically optimal choice of $c$.

Specifically, following Hallin and Liška (2007) and Alessi, Barigozzi and Capasso (2010), we generate a sequence of subsamples of sizes $(d_{1,j},\ldots,d_{K,j},T_j)$ with $j = 0,\ldots,J$ such that $d_{k,0} = 0 < d_{k,1} < d_{k,2} < \cdots < d_{k,J} = d_k$ and $T_0 = 0 < T_1 \leq T_2 \leq \cdots \leq T_J = T$, where $d_k$ is the original data dimension of tensor mode $k$ and $T$ is the original sample size, $1 \leq k \leq K$. For any $j$, we
obtain an estimated rank $\hat{r}_{k,c,j}$ of $r_k$, which is a non-increasing function of $c$. Assume $r_k > 0$. The behavior of $\hat{r}_{k,c,j}$, as a function of $j$, is different for different values of $c$. If $c > 0$ and small, in practice, as $j$ increases, $\hat{r}_{k,c,j}$ would increase to the maximum rank $m^*$ considered in (1), and we tends to overestimate $r_k$. On the other hand, when $c$ is very large, $\hat{r}_{k,c,j}$ tends to zero for any $j$, and $r_k$ is underestimated. Due to the monotonicity of $\hat{r}_{k,c,j}$ as a function of $c$, there must exist a range of “moderate” values of $c$ such that $\hat{r}_{k,c,j}$ is a stable function of the subsample size $(d_{1,j}, ..., d_{K,j}, T_j)$. The stability can be measured by the empirical variance of $\hat{r}_{k,c,j}$ as a function of $j$,

$$S_{k,c} = \frac{1}{J} \sum_{j=1}^{J} \left( \hat{r}_{k,c,j} - \frac{1}{J} \sum_{j=1}^{J} \hat{r}_{k,c,j} \right)^2.$$  

The optimal $c$ is then chosen to minimizes $S_{k,c}$.

4. Assumptions and asymptotic properties

4.1. Assumptions and notation

We introduce some notations first. Let $d = \prod_{k=1}^{K} d_k$ and $d_{-k} = d/d_k$. For a matrix $A = (a_{ij}) \in \mathbb{R}^{m \times n}$, write the SVD as $A = U \Sigma V^\top$, where $\Sigma = \text{diag}(\sigma_1(A), \sigma_2(A), ..., \sigma_{\min(m,n)}(A))$, with the singular values $\sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_{\min(m,n)}(A) \geq 0$ in descending order. The matrix Frobenius norm can be denoted as $\|A\|_F = (\sum_{ij} a_{ij}^2)^{1/2} = (\sum_{i=1}^{\min(m,n)} \sigma_i^2(A))^{1/2}$. Define the spectral norm

$$\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2 = \sigma_1(A).$$

The tensor Hilbert Schmidt norm for a tensor $A \in \mathbb{R}^{m_1 \times m_2 \times \cdots \times m_K}$ is defined as

$$\|A\|_{\text{HS}} = \sqrt{\sum_{i_1=1}^{m_1} \cdots \sum_{i_K=1}^{m_K} (A)_{i_1,\ldots,i_K}^2}.$$  

Define the tensor operator norm for an order-$4$ tensor $A \in \mathbb{R}^{m_1 \times m_2 \times m_3 \times m_4}$,

$$\|A\|_{\text{op}} = \max \left\{ \sum_{i_1,i_2,i_3,i_4} u_{i_1,i_2} \cdot u_{i_3,i_4} \cdot (A)_{i_1,i_2,i_3,i_4} : \|U_1\|_F = \|U_2\|_F = 1 \right\},$$

where $U_1 = (u_{i_1,i_2}) \in \mathbb{R}^{m_1 \times m_2}$ and $U_2 = (u_{i_3,i_4}) \in \mathbb{R}^{m_3 \times m_4}$. Define order-$4$ tensors

$$\Theta_{k,h} = \frac{\sum_{t=h+1}^{T} \text{mat}_k(M_{t-h}) \otimes \text{mat}_k(M_t)}{T - h} \in \mathbb{R}^{d_k \times d_{-k} \times d_k \times d_{-h}},$$
\[\Phi_{k,h} = \sum_{t=h+1}^{T} \frac{\text{mat}_k(F_{t-h}) \otimes \text{mat}_k(F_t)}{T-h} \in \mathbb{R}^{r_k \times r_k \times r_k \times r_k},\]

\[\Phi_{k,h}^{(\text{cano})} = \sum_{t=h+1}^{T} \frac{\text{mat}_k(M_{t-h} \times K_{k=1} U_k^\top) \otimes \text{mat}_k(M_{t-h} \times K_{k=1} U_k^\top)}{T-h} \in \mathbb{R}^{r_k \times r_k \times r_k \times r_k},\]

where \(\otimes\) is the tensor product and \(U_k\) is from the SVD form of \(A_k = U_k A_k V_k^\top\).

We view \(\Phi_{k,h}^{(\text{cano})}\) as the canonical version of the auto-covariance of the factor process. Similarly define

\[\Theta_{k,h}^* = \sum_{t=h+1}^{T} \frac{\text{mat}_k(M_{t-h}) \text{mat}_k^\top(M_t)}{T-h} \in \mathbb{R}^{d_k \times d_k},\]

\[\Phi_{k,h}^* = \sum_{t=h+1}^{T} \frac{\text{mat}_k(F_{t-h}) \text{mat}_k^\top(F_t)}{T-h} \in \mathbb{R}^{r_k \times r_k},\]

\[\Phi_{k,h}^{*(\text{cano})} = U_k^\top \Theta_{k,h}^* U_k = \sum_{t=h+1}^{T} \frac{\text{mat}_k(M_{t-h} \times K_{k=1} U_k^\top) \text{mat}_k^\top(M_{t-h} \times K_{k=1} U_k^\top)}{T-h} \in \mathbb{R}^{r_k \times r_k}.\]

Write \(\Phi_{k,1:h_0} = (\Phi_{k,h}, h = 1, \ldots, h_0)\) and \(\Phi_{k,1:h_0}^* = (\Phi_{k,h}^*, h = 1, \ldots, h_0)\). Denote \(\mathbb{E}(\cdot) = \mathbb{E}(\cdot | \{F_1, \ldots, F_T\})\). Let \(\tau_{k,m}\) be the \(m\)-th largest singular value of \(\mathbb{E}(\text{TIPUP}_k(X_{1:T}))\),

\[\tau_{k,m} = \sigma_m(\mathbb{E}(\text{TIPUP}_k(X_{1:T}))) = \sigma_m(\text{mat}_1(\Theta_{k,1:h_0}^*)).\]

Similarly, let \(\tau_{k,m}^* = \sigma_m(\mathbb{E}(\text{TOPUP}_k(X_{1:T}))) = \sigma_m(\text{mat}_1(\Theta_{k,1:h_0}^*)).\)

For simplicity, we write \(U_k = U_{k,r_k}\) and \(\hat{U}_k = \hat{U}_{k,r_k}\).

To facilitate consistency properties of the proposed procedures, we impose the following assumptions.

**Assumption I.** The error process \(\mathcal{E}_t\) are independent Gaussian tensors, condition on the factor process \(\{F_t, t \in \mathbb{Z}\}\). In addition, there exists some constant \(\sigma > 0\), such that

\[\mathbb{E}(u^\top \text{vec}(\mathcal{E}_t))^2 \leq \sigma^2 \|u\|_2^2, \quad u \in \mathbb{R}^d.\]

**Assumption II.** Assume the factor process \(F_t\) satisfies the strong \(\alpha\)-mixing condition such that

\[\alpha(h) \leq \exp(-c_0 h^{\theta_1})\]

for some constant \(c_0 > 0\) and \(0 < \theta_1 \leq 1\), where

\[\alpha(h) = \sup_t \left\{ \left| \mathbb{P}(A \cap B) - \mathbb{P}(A) \mathbb{P}(B) \right| : A \in \sigma(F_s, s \leq t), B \in \sigma(F_s, s \geq t + h) \right\} ,\]
Assumption III. For any $u_k \in \mathbb{R}^r$ with $\|u_k\|_2 = 1$ and $1 \leq k \leq K$,
\[
\max_i \mathbb{P} \left( \|F_i \times u_1 \times u_2 \times \cdots \times u_K\|_2 \geq x \right) \leq c_1 \exp \{ -c_2x^{\theta_2} \}, \tag{10}
\]
where $c_1, c_2$ are some positive constants and $0 < \theta_2 \leq 2$.

Assumption IV. Assume $r_1, ..., r_K$ are fixed. There exist some constants $\delta_0, \delta_1$ with $0 \leq \delta_0 \leq \delta_1 \leq 1$, such that $\|A_k\|_2 \approx d_k^{(1-\delta_0)/2}$ and $\sigma_{\text{tr}}(A_k) \approx d_k^{(1-\delta_1)/2}$ for all $1 \leq k \leq K$.

Assumption V. Assume that $h_0$ is fixed, and
(a) (TOPUP related): $\mathbb{E}[\text{mat}_1(\Phi_{k,1,h_0})]$ is of rank $r_k$ for $1 \leq k \leq K$.
(b) (TIPUP related): $\mathbb{E}[\text{mat}_1(\Phi_{k,1,h_0}^{(\text{cano})})]$ is of rank $r_k$ for $1 \leq k \leq K$.

Assumption I is the same assumption used in Chen, Yang and Zhang (2021) and Han et al. (2020). This assumption corresponds to the white noise assumption of Lam, Yao and Bathia (2011), Lam and Yao (2012). It allows substantial contemporaneous correlation among the entries of $\mathcal{E}_t$. Note that the normality assumption, which ensures fast convergence rates in our analysis, is imposed for technical convenience. In fact we only need to impose the sub-Gaussian condition. Assumption II allows a very general class of time series models, including causal ARMA processes with continuously distributed innovations; see also Bradley (2005), Fan and Yao (2003), Rosenblatt (2012), Tong (1990), Tsay (2005), Tsay and Chen (2018), among others. The restriction $\theta_1 \leq 1$ is introduced only for presentation convenience. Assumption III requires that the tail probability of any orthonormal projection of $F_i$ decay exponentially fast. In particular, when $\theta_2 = 2$, $F_i$ is sub-Gaussian.

Assumption IV is similar to the signal strength condition of Lam and Yao (2012), and the pervasive condition on the factor loadings (e.g., Stock and Watson (2002) and Bai (2003)). It plays a key role in identifying the common factors and idiosyncratic noises in (4). Indices $\delta_0, \delta_1$ are measures of the strength of factors, or the rate of signal strength growth as the dimension $d_k$ grows. When $\delta_0 = \delta_1 = 0$, the factors are called strong factors; otherwise, the factors are called weak factors. In particular, $\delta_0$ represents the strength of the strongest factors and $\delta_1$ the strength of the weakest factors.

Remark 9 (Signal cancellation). Assumption V guarantees that there is no redundant tensor direction in $F_i$ when combined with $A_k$’s. It is related to certain signal cancellation phenomenon which is rare for TOPUP procedures but may occur among TIPUP related procedures. Consider the case of $k = 1$ and $K = 2$. We write the factor process in the canonical form as $F_i^{(\text{cano})} = U_1^T M_i U_2 = (f_{i,j,t})_{d_1 \times d_2}$, and $\phi_{i,j_1,j_2,h} = \sum_{t=h+1}^{T-h} f_{i,j_1,j_2,t} f_{i,j_1,j_2,t}/(T-h)$ as the time average cross product between fibers $f_{i,j_1,j_2,t}$ and $f_{i,j_1,j_2,t}$ of the factor process (in canonical form). Then $\|\Theta_{1,h}\|^2_F = \sum_{i_1,j_1,j_2,j_3} \phi_{i_1,j_1,j_2,h} \phi_{i_2,j_2,j_3,h}$ and $\|\Theta_{1,h}^{\text{cano}}\|^2_F = \|\phi_k^{(\text{cano})}\|^2_F = \sum_{i_1,j_1} \sum_{j_2,j_3} \phi_{i_1,j_1,j_2,j_3,h}$. Note that the summation $\sum_{j=1}^{r_2} \phi_{i_1,j_1,j_2,j_3,h}$ is subject to potential cancellation among its terms for
In the extreme cases, \( \mathbb{E}[\text{mat}_1(\Theta_{k,1:h_0}^{(\text{cano})})] \) may not have full rank \( r_k \) and thus the signal strength \( \tau_{k,r_k}^* \) can be much smaller than the order \( d^{1-\delta} \). In Assumption V(b), we rule out the possibility of such severe signal cancellation. In practice, Han et al. (2020) suggest to examine the patterns of the estimated singular values under different lag \( h \) values. If there is no severe signal cancellation, we would expect that the pattern of \( h_0^{-1/2} \tau_{k,r_k,h_0} \) would be similar to that of \( h_0^{-1/2} \tau_{k,r_k,h_0}^* \) under different \( h_0 \). Here we emphasize that \( \tau_{k,r_k,h_0} \) and \( \tau_{k,r_k,h_0}^* \) depend on \( h_0 \), though in other places when \( h_0 \) is fixed we will omit \( h_0 \) in the notation. Severe signal cancellation would make the patterns different, since \( h_0^{-1/2} \tau_{k,r_k,h_0}^* \) suffers signal cancellation but \( h_0^{-1/2} \tau_{k,r_k,h_0} \) does not. See the discussion in Han et al. (2020). On the other hand, Assumption V(a) is sufficient to guarantee that \( \mathbb{E}[\Theta_{k,1:h_0}] \) and \( A_k \) have the same rank \( r_k \).

Instead of Assumptions II to V, Chen, Yang and Zhang (2021) and Han et al. (2020) imposed conditions on \( \|\Theta_{k,0}\|_{\text{op}}, \|\Theta_{k,0}^*\|_2^2, \tau_{k,r_k} \) and \( \tau_{k,r_k}^* \) in order to allow \( r_k \) to increase with \( d_k \). The following proposition establishes a connection between these two types of assumptions when the rank \( r_k \) is fixed.

**Proposition 2.** Suppose that Assumptions I to V hold. Let \( 1/\theta_1 = 1/\theta_1 + 2/\theta_2 \). Then, in an event \( \Omega_0 \) with probability at least \( 1 - T \exp(-C_1T^\delta) - \exp(-C_2T) \),

\[
\|\Theta_{k,0}\|_{\text{op}} \asymp \|\Theta_{k,0}^*\|_2 \asymp d^{1-\delta_0},
\]

where \( C_1, C_2 > 0 \).

### 4.2. Theoretical properties for IC and ER estimators

In this section, we shall present theoretical properties of the IC and ER estimators using non-iterative TIPUP, non-iterative TIPUP, iTOPUP and iTIPUP. We first introduce some quantities related to the estimation errors of the estimated eigenvalues of the four different methods. For non-iterative TIPUP, define

\[
\beta_k = (d^2d_k^{-1} + d^{3/2-3\delta_0/2+\delta_1} + d_k^{1/2}d^{3/2-2\delta_0+\delta_1} + d_k d^{1-\delta_0})T^{-1},
\]

\[
\gamma_k = d^{3/2-3\delta_0/2}T^{-1/2} + (d^2-\delta_0 + d^2d_k^{-1} + d^{3/2-3\delta_0/2+\delta_1} + d_k^{1/2}d^{3/2-2\delta_0+\delta_1})T^{-1},
\]

For non-iterative TIPUP, define

\[
\beta_k^* = (d^{1-\delta_0+\delta_1} + d_k d^{1-2\delta_0+\delta_1})T^{-1},
\]

\[
\gamma_k^* = (d^{3/2-3\delta_0/2} + d^{3/2-\delta_0}d_k^{-1/2})T^{-1/2} + (d^{1-\delta_0+\delta_1} + d_k d^{1-2\delta_0+\delta_1})T^{-1},
\]

Similarly, we use \( \tilde{\beta}_k, \tilde{\gamma}_k \) and \( \tilde{\beta}_k^*, \tilde{\gamma}_k^* \) for iTOPUP and iTIPUP, respectively, where

\[
\tilde{\beta}_k = (d_k d^{1-\delta_0/2} + d_k d^{1-2\delta_0+\delta_1})T^{-1}.
\]
\[
\begin{align*}
\tilde{\gamma}_k &= d^{3/2-3\delta_0/2}T^{-1/2} + d_k d^{1-2\delta_0+\delta_1} T^{-1}, \\
\tilde{\beta}^*_k &= d_k d^{1-2\delta_0+\delta_1} T^{-1}, \\
\gamma^*_k &= d^{3/2-3\delta_0/2}T^{-1/2} + d_k d^{1-2\delta_0+\delta_1} T^{-1}.
\end{align*}
\]

It is clear that \(\beta_k, \gamma_k, \beta^*_k\) and \(\gamma^*_k\) dominate \(\tilde{\beta}_k, \tilde{\gamma}_k, \tilde{\beta}^*_k\) and \(\tilde{\gamma}^*_k\), respectively. Actually, \(\beta_k\) and \(\gamma_k\) (resp. \(\beta^*_k, \gamma^*_k\), or \(\tilde{\beta}_k, \tilde{\gamma}_k\), or \(\tilde{\beta}^*_k, \tilde{\gamma}^*_k\)) correspond to the \(\beta_n\) and \(\gamma_n\) sequence in Proposition 1.

We impose the following set of conditions to ensure that the estimation error of the divergent eigenvalue is much smaller than the true smallest non-zero eigenvalue, and the estimation error of the zero eigenvalues is relatively small. The \(\gamma\)’s above are part of the estimation errors of the divergent eigenvalues, and the \(\beta\)’s are the estimation errors of the (true) zero eigenvalues, for the four different estimation method. Note that \(d^{2-2\delta_1}\) is the growth rate of the smallest non-zero eigenvalues corresponding to the weakest factors. See also Theorem 2 below.

**Assumption VI** (Rate condition).

(a) \(\max_{1 \leq k \leq K} \{a_k\} = o(d^{2-2\delta_1})\)

(b) \(\max_{1 \leq k \leq K} \{b_k\} = o(d^{2-2\delta_0-4\delta_1})\).

The sequences \(a_k\) and \(b_k\) will be one of the \(\gamma_k\) and \(\beta_k\) sequences defined above, respectively, based on the estimators.

We will impose the following sufficient conditions on the penalty function \(g_k(\cdot)\) and \(h_k(\cdot)\).

**Assumption VII** (Sufficient condition on the penalty functions).

(a) \(b_k \prec \min_k \{g_k(d,T)\} \leq \max_k \{g_k(d,T)\} \prec d^{2-2\delta_1}\),

(b) \(d^{2\delta_1-2\delta_0} b^2_k \ll \min_k \{h_k(d,T)\} \leq \max_k \{h_k(d,T)\} \ll d^{2+2\delta_0-4\delta_1}\),

where \(\ll\) indicates that there is a constant \(C\) such that \(\ll\) uniformly, and \(\ll\) indicates \(\ll = o(\ll)\). The sequence \(b_k\) will be specified for different estimators.

**Remark 10** (Penalty functions). The penalty functions \(g_k\) and \(h_k\) enter the consistency theorem below through Assumption VII. They do not have direct impact on the convergence rate of the rank estimators, as long as the condition is satisfied. Indirectly their choice interacts with the required sample size \(T\) and dimension \(d\). Roughly speaking, Assumption VII(a) dictates that the penalty \(g_k(d,T\) should be less than the smallest diverging eigenvalue, but large enough to correctly truncate the estimated true zero eigenvalues. The \(b_k\) sequence in the assumption is taken to be one of the \(\beta_k, \beta^*_k, \tilde{\beta}_k\) and \(\tilde{\beta}^*_k\) defined above, according to the procedure used. In practice we generally do not know the the factor strengths \(\delta_0\) and \(\delta_1\), hence may not always be able to specify a \(g_k(d,T)\) that satisfies the condition. However, the range between the upper and lower bounds is quite wide in most of the cases, especially with the additional \(T^{-1}\) term in the lower bound. All of the suggested \(g_k(d,T)\) listed in (7) satisfy the condition, if
\( \nu = \delta_1 \). Our experiments shows that setting \( \nu = 0 \) in (7) is sufficient in most of the cases. Only when the true \( \delta_1 \) is very large (extreme weak factors), the results become sensitive to the selection of \( \nu \). In such cases, a data driven procedure similar to that in Hallin and Liška (2007) for vector factor models may be used to estimate \( \delta_1 \). Its property for tensor factor model may need further investigation.

One can also study the pattern of the rank estimates under different \( \nu \).

The condition imposed on the penalty function \( h_k(d, T) \) in Assumption VII(b) is even weaker. The upper bound goes to infinity but the lower bound goes to zero, except when both \( \delta_0 \) and \( \delta_1 \) are large, and \( T \) is of smaller order than \( d \). Hence in most of the cases a (small) constant function is sufficient. If \( r_k \) is the true rank, the function \( h_k(d, T) \) is designed to adjust the ratio of eigenvalues \( \hat{\lambda}_{k,j+1}/\hat{\lambda}_{k,j} \), \( j > r_k \) to be bounded below and to be around 1 (as we add \( h_k(d, T) \) on both the numerator and denominator) so that they do not accidentally be smaller than \( \hat{\lambda}_{k,r_k+1}/\hat{\lambda}_{k,r_k} \) (a number that goes to 0). Hence intuitively we do not expect the impact of \( h_k(d, T) \) to be large, which is confirmed by our empirical study. The suggested functions in (8) all satisfy the Assumption VII(b), except the extreme weak factor cases, for which the ER estimators do not perform well under any penalty function.

Assumption VII only provides broad guidance asymptotically. There is no general unique optimal penalty function. Note that if \( g_k(d, T) \) is an appropriate penalty function, then \( cg_k(d, T) \) is appropriate as well asymptotically. The same property holds for the approaches of Bai and Ng (2002, 2007), Amengual and Watson (2007), Hallin and Liška (2007) and Li, Li and Shi (2017). This creates potential problems in practice with given \( d \) and \( T \). Under certain circumstances, the empirical performance of IC estimators may heavily depend on the threshold function chosen among many alternatives; see the discussion in Hallin and Liška (2007).

The following is a set of different sample size conditions for different settings. They ensure sufficiently large sample size \( T \) so that the non-iterative (true rank) factor loading space estimator based on TOPUP or TIPUP is consistent (for (a) and (b)), or has a relatively small error (for (c) and (d)).

**Assumption VIII** (Condition on the sample size).

(a) \( (d^{1/2} - \delta_0/2 + d^{1/2} d_k^{-1/2})T^{-1/2} = o(1), \quad 1 \leq k \leq K. \)
(b) \( (d_k^{1/2} d^{1/2} - \delta_0/2 - 1/2 + d^{1/2} - 1/2)T^{-1/2} = o(1), \quad 1 \leq k \leq K. \)
(c) \( (d^{1/2} - \delta_0/2 + d^{1/2} d_k^{-1/2})T^{-1/2} + d_k^{1/2} d^{1/2} - 1/2 T^{-1} \leq C, \quad 1 \leq k \leq K. \)
(d) \( (d_k^{1/2} d^{1/2} - 5\delta_0/2 - 1/2 + d^{1/2} - 2\delta_0 - 1/2)T^{-1/2} \leq C, \quad 1 \leq k \leq K. \)

Theorem 1 presents the asymptotic properties of the IC and ER estimators in (1) and (2) based on non-iterative TOPUP, non-iterative TIPUP, iTOPUP and iTIPUP.

**Theorem 1.** Let \( 1/\vartheta = 1/\theta_1 + 2/\theta_2 \) as in Proposition 2. For the various rank determination estimators, if their corresponding conditions listed in Table 2 hold,
Table 2
Summary of conditions needed for Theorem 1

<table>
<thead>
<tr>
<th>IC Estimator</th>
<th>Model (I)-(IV)</th>
<th>Rank (V)</th>
<th>Rate 1 (VI)(a)</th>
<th>Rate 2 (VI)(b)</th>
<th>Penalty γk(·) (VII)(a)</th>
<th>Size (VIII)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOPUP r(I)C</td>
<td>yes (a)</td>
<td>a_k = γk</td>
<td>b_k = β_k</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIPUP r*(IC)</td>
<td>yes (b)</td>
<td>γ_k</td>
<td>β_k</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>iTOPUP r(1)(IC)</td>
<td>yes (a)</td>
<td>γ_k</td>
<td>β_k</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>iTIPUP r*(1)(IC)</td>
<td>yes (b)</td>
<td>γ_k</td>
<td>β_k</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 2
Assumptions

\[
P(\hat{r}_k = r_k, 1 \leq k \leq K) \geq 1 - \sum_{k=1}^{K} e^{-d_k} - T \exp(-C_1 T^0) - \exp(-C_2 T),
\]

with \( C_1, C_2 > 0 \).

In addition to the consistency of the rank estimators, we also have the following more detailed properties of the estimated eigenvalues. Let \( \lambda_{k,j} \) be the eigenvalues of \( \hat{W}_k \) (defined in Section 3.1) such that \( \lambda_{k,1} \geq \lambda_{k,2} \geq \ldots \geq \lambda_{k,d_k}, 1 \leq k \leq K \). Also let \( \lambda_{k,j} \) be the eigenvalues of population version \( E\hat{W}_k \) such that \( \lambda_{k,1} \geq \ldots \geq \lambda_{k,r_k} > \lambda_{k,r_k+1} = \ldots = \lambda_{k,d_k} = 0 \). Similarly, define \( \hat{\lambda}_{k,j}^*, \lambda_{k,j}^*, \hat{\lambda}_{k,j}^{(i)}, \lambda_{k,j}^{(i)}, \hat{\lambda}_{k,j}^{(i)} \) and \( \lambda_{k,j}^{(i)} \) as the eigenvalues of \( \hat{W}_k^*, E\hat{W}_k^*, \hat{W}_k^{(i)}, E\hat{W}_k^{(i)}, \hat{W}_k^{(i)}, E\hat{W}_k^{(i)} \), \( i \geq 1 \), respectively.

**Theorem 2.** Suppose the same conditions (I-V, VIII) in Theorem 1 hold. In an event with probability approaching 1 (as \( T \to \infty \) and \( d \to \infty \)), the following holds.

(i). For estimating the true zero eigenvalues, we have, for \( j \geq r_k \) and \( i \geq 1 \),

\[
\hat{\lambda}_{k,j} = O(\beta_k), \quad \hat{\lambda}_{k,j}^* = O(\beta_k^*), \quad \hat{\lambda}_{k,j}^{(i)} = O(\beta_k), \quad \text{and} \quad \hat{\lambda}_{k,j}^{(i)} = O(\beta_k^*).
\]

(ii). For estimating the non-zero eigenvalues, we have, for all \( 1 \leq j \leq r_k \) and \( i \geq 1 \),

\[
|\hat{\lambda}_{k,j} - \lambda_{k,j}| = O(T^{-1/2} d^{2+\delta_1-5\delta_0/2} + T^{-1/2} d^{1/2-\delta_1} - 2\delta_0 + \gamma_k),
\]
\[
|\hat{\lambda}_{k,j}^* - \lambda_{k,j}^*| = O(T^{-1/2} d^{1/2} d^{1/2} - 5\delta_0/2 + T^{-1/2} d^{1/2+\delta_1} - 2\delta_0 + \gamma_k^*),
\]
\[
|\hat{\lambda}_{k,j}^{(i)} - \lambda_{k,j}^{(i)}| = O(T^{-1/2} d^{1/2} d^{1/2} - 5\delta_0/2 + \gamma_k),
\]
\[
|\hat{\lambda}_{k,j}^{(i)} - \lambda_{k,j}^{(i)}| = O(T^{-1/2} d^{1/2} d^{1/2} - 5\delta_0/2 + \gamma_k^*).
\]
Remark 11 (Strong factor cases). To illustrate the theorem, we consider the strong factor case $\delta_0 = \delta_1 = 0$. Here the quantities (11)-(18) can be simplified to
\[
\beta_k \asymp T^{-1}d_k^{-1}d^2 + T^{-1}d_k^{1/2}d^{3/2}, \quad \gamma_k \asymp T^{-1}d^2 + T^{-1/2}d^{3/2}, \\
\beta_k^* \asymp \tilde{\beta}_k \asymp \tilde{\beta}_k^* \asymp T^{-1}d_k d, \\
\gamma_k^* \asymp \tilde{\gamma}_k \asymp \tilde{\gamma}_k^* \asymp T^{-1}d_k d + T^{-1/2}d^{3/2}.
\]
The sample size conditions in Assumption VIII all reduces to $T \to \infty$. And the penalty function conditions (Assumption VII) is equivalent to (a) $b_k \prec g_k(d,T) \prec d^2$, (b) $d^{-2}b_k^2 \ll h_k(d,T) \ll d^2$, $1 \leq k \leq K$. Thus, we shall expect similar performance for non-iterative TIPUP, iTOPUP and iTIPUP, but the non-iterative TOPUP may be worse.

Remark 12 (The vector factor models). Theorem 1 and 2 hold for vector factor models by setting $K = 1$ and $d = d_1$. In such a case, TOPUP is the same as TIPUP. More specifically, assuming all factors have the same strength ($\delta_0 = \delta_1$), a common assumption used in the literature, Theorem 2 reduces to $|\hat{\lambda}_{k,j} - \lambda_{k,j}| = O_p(T^{-1/2}d^{-3\delta_1/2})$ for $1 \leq j \leq r_1$ and $\hat{\lambda}_{k,j} = O_p(T^{-1}d^{-2\delta_0})$ for $j > r_1$ for the vector factor model case. This is the same as the convergence rate of the estimated eigenvalues derived in Lam and Yao (2012), though our improved technical proof removed the restrictive conditions that $T = O(d)$ and all the non-zero eigenvalues are distinct. In addition, Theorem 1 provides the rate of convergence of the rank estimators.

Remark 13. Note that our model setting is different from that used in Bai and Ng (2002) and Hallin and Liška (2007) where covariance matrix or spectral density matrix are used, instead of the auto-co-moment we use here. It is possible to extend our approach to identify the number of factors in these models, by setting $h_0 = 0$ in the construction of $\tilde{W}$ and using an extension of Proposition 1 discussed in Remark 3. The main difference is that, in our model and with auto-co-moments, we are trying to separate non-zero and zero eigenvalues in the underlying $W$, while in approximate factor model and $h_0 = 0$, one would be trying to separate spiked and non-spiked eigenvalues. Hence a detailed analysis of the corresponding $\gamma_n$ and $\beta_n$ in Proposition 1 will be needed.

Remark 14 (Sample size requirement comparison). The sample size required for the non-iterative estimators as shown in Assumptions VIII(a,b) is of higher order than that for the iterative estimators as in Assumptions VIII(c,d). This is because, when the true ranks are used, TOPUP and TIPUP require a larger sample size to consistently estimate the true loading spaces $A_k$ than the iTOPUP and iTIPUP procedures which require only a sufficiently “good” initial estimator of the loading space, but not necessarily a consistent one (Han et al., 2020). Similarly, the required sample size condition for TIPUP in Assumption VIII(b) is much weaker than that for TOPUP in Assumption VIII(a). In this regard, iterative procedures are better than the non-iterative ones, and TIPUP based procedures are better than TOPUP based ones.

Remark 15 (Convergence rate comparison). The convergence rates of the estimated eigenvalues in the iterative methods are faster than that in the non-
Comparison of the \( \delta \)’s in the convergence rate for estimating the true zero eigenvalues, under Assumption VIII(a)-(d).

<table>
<thead>
<tr>
<th>( \delta_0 ) condition</th>
<th>( \delta_1 ) and ( T ) condition</th>
<th>comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta_0 &gt; 1/K )</td>
<td>( d_1 &gt; 3d_0/2 )</td>
<td>( \beta^*_k \gg \bar{\beta}_k \gg \beta_k )</td>
</tr>
<tr>
<td></td>
<td>( d_0/2 + 1/K &lt; \delta_1 &lt; 3d_0/2 )</td>
<td>( \beta^*_k \gg \bar{\beta}_k \gg \beta_k )</td>
</tr>
<tr>
<td></td>
<td>( \delta_1 = d_0/2 + 1/K \geq d_0 )</td>
<td>( \beta^*_k \gg \bar{\beta}_k \gg \beta_k )</td>
</tr>
<tr>
<td></td>
<td>( d_0 / 2 + 1/K &lt; d_0 )</td>
<td>( \beta^*_k \gg \bar{\beta}_k \gg \beta_k )</td>
</tr>
<tr>
<td>( \delta_0 \leq 1/K )</td>
<td>( \delta_1 &gt; 3d_0/2 )</td>
<td>( \beta^*_k \gg \bar{\beta}_k \gg \beta_k )</td>
</tr>
<tr>
<td></td>
<td>( \delta_1 &lt; 3d_0/2 )</td>
<td>( \beta^*_k \gg \bar{\beta}_k \gg \beta_k )</td>
</tr>
</tbody>
</table>

Comparison of convergence rate for estimating the true zero eigenvalues

<table>
<thead>
<tr>
<th>( \delta_0 ) condition</th>
<th>( \delta_1 ) condition</th>
<th>comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta_0 &gt; 1/K )</td>
<td>( d_1 )</td>
<td>( \beta^*_k \gg \bar{\beta}_k \gg \beta_k )</td>
</tr>
<tr>
<td></td>
<td>( d_0/2 + 1/K &lt; \delta_1 &lt; 3d_0/2 )</td>
<td>( \beta^*_k \gg \bar{\beta}_k \gg \beta_k )</td>
</tr>
<tr>
<td></td>
<td>( \delta_1 = d_0/2 + 1/K \geq d_0 )</td>
<td>( \beta^*_k \gg \bar{\beta}_k \gg \beta_k )</td>
</tr>
<tr>
<td></td>
<td>( d_0 / 2 + 1/K &lt; d_0 )</td>
<td>( \beta^*_k \gg \bar{\beta}_k \gg \beta_k )</td>
</tr>
<tr>
<td>( \delta_0 \leq 1/K )</td>
<td>( \delta_1 &gt; 3d_0/2 )</td>
<td>( \beta^*_k \gg \bar{\beta}_k \gg \beta_k )</td>
</tr>
<tr>
<td></td>
<td>( \delta_1 &lt; 3d_0/2 )</td>
<td>( \beta^*_k \gg \bar{\beta}_k \gg \beta_k )</td>
</tr>
</tbody>
</table>

iterative methods, especially when there are weak factors in the model. Moreover, the rate of TIPUP related procedures is also faster than that of TOPUP related procedures. This can be seen by comparing \( \beta_k, \gamma_k, \bar{\gamma}_k, \bar{\beta}_k \) with \( \tilde{\beta}_k, \tilde{\gamma}_k, \tilde{\beta}_k, \tilde{\gamma}_k \).

For example, consider the case that all \( d_k \) are of the same order and \( K > 1 \). The following table shows the comparison of the convergence rate (\( \beta \)’s) of the estimated true zero eigenvalues, where \( \prec \) and \( \prec \prec \) are defined in Assumption VII.

From Tables 3, it is clear that \( \beta^*_k \) is always the smallest, and \( \bar{\beta}_k \) is the largest. Similarly, Table 4 shows that \( \tilde{\gamma}_k \) and \( \tilde{\gamma}_k \) are always the smallest among these four \( \gamma \)’s.

For the IC estimators with a fixed penalty functions \( g_k(d, T) \), a faster convergence rate of the eigenvalue estimators make the sufficient condition Assumption VII(a) easier to satisfy with smaller sample size \( T \) and/or dimension \( d_k \), \( 1 \leq k \leq K \). Similarly, for the ER estimators, faster rates for estimating the true zero eigenvalue increase the gap between the estimated divergent eigenvalues and the estimated true zero eigenvalue, leading to better performance of the ER estimators.

Combining the discussion in Remarks 14 and 15, we can conclude that in general the iterative procedures are better than the non-iterative ones and the TIPUP based procedures are better than the TOPUP ones, assuming no signal cancellation when TIPUP is used (see Remark 9).
5. Simulation Study

In this section, we compare the empirical performance of the proposed methods and their variants under various simulation setups. We consider the identification of the number of factors based on the non-iterative TIPUP and TOPUP methods (denoted as initial estimators), the one step iterative methods (denoted as one-step estimators) and the iterative procedures after convergence (denoted as final estimator). In the iterative algorithm, at i-th iteration \(i > 1\), we use the order obtained by the IC or ER criteria. Iteration is stopped when both the rank estimators and the loading space estimators converge. We also check the performance of different choices of penalty function \(g_k(d,T)\) and \(h_k(d,T)\). Specifically, we consider penalty functions (7) and (8), and denote them as IC1-IC5, ER1-ER5, respectively. The empirical performance of IC1-IC5 (resp. ER1-ER5) are very similar, thus we only present IC2 and ER1 in this section. The detailed comparison are shown in Appendix B.

The simulation study consists of three parts. The first part is designed to investigate the overall performance of our methods and their comparisons under models with different factor strength. As the strength of the weakest factors is unknown, we by default set \(\nu = 0\) for all the penalty function \(g_k(d,T)\) in (7). In the second part, we investigate the case in which some factors have a dominantly strong explanatory power. The third part is the case in which we use \(\nu = \delta_1\) in (7) for all IC estimators, when some factors are weak. For each case, we compute the proportion of correct identification of the rank of the factor processes or the root mean squared errors (RMSEs) of the rank estimates from 1000 simulated data sets. In Section 5.4, we study the selection of the optimal constant \(c\) in IC criteria, using the method proposed in Remark 8.

The simulation uses the following matrix factor model:

\[
X_t = A_1 F_t A_2^\top + E_t.
\]

Here, \(E_t\) is white and is generated according to \(E_t = \Psi_1^{1/2} Z_t \Psi_2^{1/2}\), where \(\Psi_1, \Psi_2\) are the column and row covariance matrices with the diagonal elements being 1 and all off diagonal elements being 0.2. All of the elements in the \(d_1 \times d_2\) matrix \(Z_t\) are i.i.d \(N(0,1)\). This type of model of \(E_t\) has been proposed and studied in the literature, see, for example Hafner, Linton and Tang (2020), Hoff (2011), Linton and Tang (2020). The entries \(f_{ijt}\) in the factor matrix \(F_t\) were drawn from independent univariate AR(1) model \(f_{ijt} = \phi_{ij} f_{ij(t-1)} + \epsilon_{ijt}\) with standard \(N(0,1)\) innovation.

5.1. Part I: Determining strong and weak factors, using \(\nu = 0\) in \(g_k(\cdot)\)

In the first part, the following three models are studied:

(M1). Set \(r_1 = r_2 = 5\). The univariate \(f_{ijt}\) follows AR(1) with AR coefficient...
Table 5
Proportion of correct identification of rank \( r \) using IC2 and ER1 estimators based on both TOPUP and TIPUP procedures for Model M1, over 1000 replications

<table>
<thead>
<tr>
<th>( T )</th>
<th>IC</th>
<th>ER</th>
<th>IC</th>
<th>ER</th>
<th>IC</th>
<th>ER</th>
<th>IC</th>
<th>ER</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_1 = d_2 = 20 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.10</td>
<td>1 1 1</td>
<td>0.75</td>
<td>1 1 1</td>
<td>0.85</td>
<td>1 1 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>0.99</td>
<td>1 1 1</td>
<td>1 1 1</td>
<td>1 1 1</td>
<td>1 1 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>1</td>
<td>1 1 1</td>
<td>1 1 1</td>
<td>1 1 1</td>
<td>1 1 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>1</td>
<td>1 1 1</td>
<td>1 1 1</td>
<td>1 1 1</td>
<td>1 1 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( d_1 = d_2 = 40 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>1 1 1</td>
<td>0.47</td>
<td>1 1 1</td>
<td>0.48</td>
<td>1 1 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>0.03</td>
<td>1 1 1</td>
<td>0.98</td>
<td>1 1 1</td>
<td>1 1 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.46</td>
<td>1 1 1</td>
<td>1 1 1</td>
<td>1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>1</td>
<td>1 1 1</td>
<td>1 1 1</td>
<td>1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( d_1 = d_2 = 80 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>1 1 1</td>
<td>0.66</td>
<td>1 1 1</td>
<td>0.68</td>
<td>1 1 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>0.03</td>
<td>1 1 1</td>
<td>0.98</td>
<td>1 1 1</td>
<td>1 1 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.46</td>
<td>1 1 1</td>
<td>1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>0.70</td>
<td>1 1 1</td>
<td>1 1 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( \phi_{ij} \), where

\[
\begin{pmatrix}
(\phi_{ij}) = \\
0.8 & 0.5 & 0.3 & 0.3 \\
0.5 & 0.8 & 0.3 & 0.3 \\
0.3 & 0.5 & 0.8 & 0.3 \\
0.3 & 0.3 & 0.5 & 0.8 \\
\end{pmatrix}
\] ; \hspace{1cm} (19)

All elements of \( A_1 \) and \( A_2 \) are i.i.d N(0,1).

(M2). Set \( r_1 = r_2 = 5 \). The univariate \( f_{ijt} \) follows AR(1) with AR coefficient \( \phi_{ij} \), where \( \phi_{ij} \) is defined in (19). The elements of the first two columns of \( A_1 \) and \( A_2 \) are i.i.d N(0,1) and the elements of the last three columns of \( A_1 \) and \( A_2 \) are i.i.d \( N(0,1)/d_0^{0.2} \) and \( N(0,1)/d_0^{0.2} \), respectively.

(M3). Same setting as in Model M2, except the elements of \( A_1 \) and \( A_2 \) are i.i.d \( N(0,1)/d_0^{0.3} \) and \( N(0,1)/d_0^{0.3} \).

All the factors in Models M1 are strong factors. Model M2 is the case in which four \((2 \times 2)\) factors are strong \((\delta_0 = 0)\), twelve factors are weak factor with strength 0.2 and the rest nine factors are weak factor with strength \( \delta_1 = 0.4 \). Model M3 is the case in which all the factors are very weak factors with strength \( \delta_0 = \delta_1 = 0.6 \). Models M2 and M3 are designed to examine the effects of weak factors on the estimators. We choose a set of data dimensions to be \((d_1, d_2) = (20, 20), (40, 40), (80, 80)\) and the sample size to be \( T = 100, 300, 500, 1000 \). Again, in this first part of simulation, we fix \( h_0 = 1 \) and set \( \nu = 0 \) in the penalty function in (7), under the assumption that all factors are strong, even though some of factors simulated are weak (e.g., true \( \delta_1 = 0.4 \) in Model M2).

For Model M1, the results in Table 5 show clearly that, using TOPUP and IC, the initial estimator behaves very poorly even for large sample sizes. On the other hand, the one step estimator uniformly and significantly outperforms the non-iterative initial estimator. In addition, the final estimator performs the
Determining the number of factors

Table 6
Proportion of correct identification of rank $r$ using IC2 and ER1 estimators based on both TOPUP and TIPUP procedures for Model M2, over 1000 replications

<table>
<thead>
<tr>
<th>$T$</th>
<th>$d_1 = d_2 = 20$</th>
<th>$d_1 = d_2 = 40$</th>
<th>$d_1 = d_2 = 80$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial estimator</td>
<td>one step estimator</td>
<td>final estimator</td>
</tr>
<tr>
<td></td>
<td>IC</td>
<td>ER</td>
<td>IC</td>
</tr>
<tr>
<td>100</td>
<td>0.50</td>
<td>0.01</td>
<td>0.29</td>
</tr>
<tr>
<td>300</td>
<td>0.73</td>
<td>0.35</td>
<td>0.01</td>
</tr>
<tr>
<td>500</td>
<td>0.58</td>
<td>0.79</td>
<td>0</td>
</tr>
<tr>
<td>1000</td>
<td>0.46</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

We also observe that the performance improves as the dimension increases, except that $d_1 = 40$ is not as good as $d_1 = 20$ when $T = 100$. This improvement is due to the fact that when with strong factors ($\delta_0 = \delta_1 = 0$), larger dimension ($d_1, d_2$) provides more data points and information on the rank $r_k$. With the same settings, IC criterion based on TIPUP determines the ranks perfectly, indicating that it is uniformly better than IC criterion based on TOPUP. This is partially due to the fact that non-iterative and iterative TIPUP procedures estimate the loading matrices and the eigenvalues more accurately than the corresponding TOPUP procedures. More interestingly, with the same setting, the ER estimators based on both TOPUP and TIPUP procedures perform perfectly.

For Model M2, Tables 6 reports the proportion of correct rank identification using IC and ER estimators based on TOPUP and TIPUP procedures. It is seen that, for small sample sizes ($T = 100$ and 300), the performance of IC estimators deteriorate when we use iterative procedures, which may indicate that the sample size $T$ and dimension $d$ do not meet the required Assumption VII(a) in Theorem 1. In addition, for $T = 100$, the IC estimators using TIPUP procedures do not work at all, though they are better when $T = 300$. This is due to the existence of weak factors and the fact that we use the default $\nu = 0$ in (7). When the sample size is small, the estimators tend to identify the strong factors while miss the weak factors as their corresponding eigenvalues are relatively small and comparable to the penalty function $g_k(\cdot)$. The IC estimators using TOPUP procedures performed much better for small sample sizes. The performance also becomes worse as $d_1$ and $d_2$ increases, also due to the existence of weak factors. The ER estimators based on TIPUP procedures show almost perfect accuracy, even there are weak factors in the model. We note that, even with weak factors, the performance remains almost the same with larger ($d_1, d_2$). Again, accuracy improves by using the iterative procedure. The ER estimator based on TIPUP procedures are much better than all the other estimators for
Table 7 shows the more detailed identification results using IC2 and ER1 estimators for Model M2 over 1000 replications, based on TOPUP and TIPUP procedures. The sample size is $T = 300$ and $(d_1, d_2) = (80, 80)$. The true rank pair is $(5,5)$.

<table>
<thead>
<tr>
<th>$(\hat{r}_1, \hat{r}_2)$</th>
<th>initial estimator</th>
<th>one step estimator</th>
<th>final estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IC2-TOP</td>
<td>IC2-TIP</td>
<td>IC2-TOP</td>
</tr>
<tr>
<td>(4,4)</td>
<td>0</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>(4,5)</td>
<td>0.01</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>(5,4)</td>
<td>0.24</td>
<td>0.29</td>
<td>0.35</td>
</tr>
<tr>
<td>(5,5)</td>
<td>0.75</td>
<td>0.71</td>
<td>0.61</td>
</tr>
</tbody>
</table>

| $(\hat{r}_1, \hat{r}_2)$ | ER1-TOP           | ER1-TIP            | ER1-TOP        | ER1-TIP        |
|-------------------------|-------------------|--------------------|----------------|
| (2,2)                   | 1                 | 0                  | 0.15           | 0.19           |
| (2,5)                   | 0                 | 0                  | 0.20           | 0.13           |
| (5,2)                   | 0                 | 0                  | 0.14           | 0.07           |
| (5,5)                   | 0                 | 1                  | 0.51           | 0.61           |

Model M2.

For Model M3 with all very weak factors ($\delta_0 = \delta_1 = 0.6$), Table 8 reports RMSEs of the ER1 estimators. The results of using IC estimators (not shown here) are significantly worse than that of the ER estimators due to the difficulty of IC estimators in dealing with weak factors when $\nu = 0$ is used. It is seen from the table that ER estimators based on TIPUP procedure outperform that based on TOPUP procedure. We also see that the iterative algorithms improves the performance very significantly under this very weak factor case. The performance varies with the change of $(d_1, d_2)$ in a non-standard way, as the performance with $(40, 40)$ seems to be better than that with $(20, 20)$ and $(80, 80)$. We note that in weak factor cases, a large $d_k$ will potentially reduce the accuracy of the estimator of $r_k$, since the signal level on the $k$-th dimension becomes weaker. On the other hand, a larger $d_i$ ($i \neq k$) in the other dimension potentially improves the estimation of $r_k$, since we have more “repeated” observations to be used for...
Determining the number of factors

Table 8

Root mean squared errors (RMSEs) of the ER1 estimators based on TOPUP and TIPUP procedures for Model M3, averaging over 1000 replications

<table>
<thead>
<tr>
<th>$T$</th>
<th>initial estimator</th>
<th>one step estimator</th>
<th>final estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ER1-TOP</td>
<td>ER1-TIP</td>
<td>ER1-TOP</td>
</tr>
<tr>
<td>100</td>
<td>2.15</td>
<td>1.50</td>
<td>1.84</td>
</tr>
<tr>
<td>300</td>
<td>1.17</td>
<td>0.60</td>
<td>1.07</td>
</tr>
<tr>
<td>500</td>
<td>1.02</td>
<td>0.30</td>
<td>1.04</td>
</tr>
<tr>
<td>1000</td>
<td>1.00</td>
<td>0.07</td>
<td>0.95</td>
</tr>
</tbody>
</table>

$d_1 = d_2 = 40$

<table>
<thead>
<tr>
<th>$T$</th>
<th>initial estimator</th>
<th>one step estimator</th>
<th>final estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>1.01</td>
<td>1.19</td>
<td>1.03</td>
</tr>
<tr>
<td>300</td>
<td>1.00</td>
<td>0.27</td>
<td>1.00</td>
</tr>
<tr>
<td>500</td>
<td>1.00</td>
<td>0.17</td>
<td>1.00</td>
</tr>
<tr>
<td>1000</td>
<td>1.00</td>
<td>0.04</td>
<td>0.99</td>
</tr>
</tbody>
</table>

$d_1 = d_2 = 80$

<table>
<thead>
<tr>
<th>$T$</th>
<th>initial estimator</th>
<th>one step estimator</th>
<th>final estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>2.71</td>
<td>2.78</td>
<td>2.06</td>
</tr>
<tr>
<td>300</td>
<td>1.00</td>
<td>0.99</td>
<td>1.01</td>
</tr>
<tr>
<td>500</td>
<td>1.00</td>
<td>0.32</td>
<td>1.00</td>
</tr>
<tr>
<td>1000</td>
<td>1.00</td>
<td>0.15</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 9

Proportion of the estimated rank pair ($r_1, r_2$) of the ER1 estimators for Model M3 over 1000 replications, when $T = 300$ and $(d_1, d_2) = (80, 80)$. The true rank pair is $(5, 5)$.

<table>
<thead>
<tr>
<th>$(r_1, r_2)$</th>
<th>initial estimator</th>
<th>one step estimator</th>
<th>final estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>(5,5)</td>
<td>0</td>
<td>0.51</td>
<td>0</td>
</tr>
<tr>
<td>(6,5)</td>
<td>0</td>
<td>0.07</td>
<td>0</td>
</tr>
<tr>
<td>(6,6)</td>
<td>1</td>
<td>0.09</td>
<td>0.98</td>
</tr>
<tr>
<td>(7,5)</td>
<td>0</td>
<td>0.15</td>
<td>0</td>
</tr>
<tr>
<td>Others</td>
<td>0</td>
<td>0.18</td>
<td>0.02</td>
</tr>
</tbody>
</table>

estimating $r_k$.

Table 9 shows the relative frequency of different estimated ranks of the ER1 estimator based on both TIPUP and TOPUP procedures, for the case of $T = 300$ and $(d_1, d_2) = (80, 80)$ under Model M3 with the true rank $(5, 5)$. It is seen that the ER estimators tend to overestimate the number of factors, when all the factors are weak. All ER1-TOPUP estimators essentially identify $(6, 6)$ as the rank. The non-iterative ER1 estimator based on TIPUP procedure can overestimate the ranks by a large margin. However, the iterations can gradually correct the over-estimation.

In summary, the first part of simulation shows that the ER estimators and IC estimator based on TIPUP procedures perform very well when all the factors are strong. The ER estimators significantly outperform the IC estimators when some or all factors are weak. Different from the results shown in Lam and Yao (2012) for vector factor models with both strong and weak factors, in tensor factor models, the ER estimators (based on TIPUP procedure) are
able to determine the correct number of factors in many cases. The results also show that the iterative procedure significantly improves the performance except the IC estimators based on TOPUP in Model M2, which may due to that the sample size $T$ and dimension $d$ do not meet the required Assumption VII(a) in Theorem 1. It also shows that the estimators based on TIPUP perform better than that based on TOPUP in general. Hence, when some factors are weak, the iterative ER estimators based on TIPUP are the choice.

5.2. Part II: The case of dominating strong factors

The second part of our simulation examines the effects of dominate strong factors on the IC and ER estimators. The data are generated from the following model, (M4). Set $r_1 = r_2 = 2$. The univariate $f_{jt}$ follows AR(1) with AR coefficient $\phi_{11} = 0.98$ and $\phi_{12} = \phi_{21} = \phi_{22} = 0.15$; The elements of the loading matrices $A_1$ and $A_2$ are i.i.d $N(0, 1)$.

We fix $d_1 = d_2 = 40$ and $T = 200$. Again, we use IC2 and ER1 for demonstration, and assume $\nu = 0$ in the penalty function in (7). Although all of the four factors are strong factors, the strongly imbalanced signal strength in $F_t$ makes one of the factors dominating the others in explanatory power. Table 10 reports the relative frequencies of estimated rank pairs over 1000 replications. It is seen that the ER estimators are very likely to pick up only the dominate factor, although iterations significantly improves the accuracy. The IC estimators perform much better in this case. And over all, estimators based on TIPUP perform better than the corresponding estimators using TOPUP. Overall, IC-TIPUP performs the best in this case.

<table>
<thead>
<tr>
<th>$(r_1, r_2)$</th>
<th>initial estimator</th>
<th>one step estimator</th>
<th>final estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IC2-TOP</td>
<td>IC2-TIP</td>
<td>IC2-TOP</td>
</tr>
<tr>
<td>$(1,1)$</td>
<td>0</td>
<td>0.02</td>
<td>0</td>
</tr>
<tr>
<td>$(1,2)$</td>
<td>0.05</td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td>$(2,1)$</td>
<td>0.01</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>$(2,2)$</td>
<td>0.22</td>
<td>0.85</td>
<td>0.31</td>
</tr>
<tr>
<td>$(3,3)$</td>
<td>0.54</td>
<td>0</td>
<td>0.54</td>
</tr>
<tr>
<td>Others</td>
<td>0.18</td>
<td>0</td>
<td>0.12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$(r_1, r_2)$</th>
<th>initial estimator</th>
<th>one step estimator</th>
<th>final estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ER1-TOP</td>
<td>ER1-TIP</td>
<td>ER1-TOP</td>
</tr>
<tr>
<td>$(1,1)$</td>
<td>1</td>
<td>0.878</td>
<td>0.88</td>
</tr>
<tr>
<td>$(1,2)$</td>
<td>0</td>
<td>0.065</td>
<td>0</td>
</tr>
<tr>
<td>$(2,1)$</td>
<td>0</td>
<td>0.017</td>
<td>0.04</td>
</tr>
<tr>
<td>$(2,2)$</td>
<td>0</td>
<td>0.040</td>
<td>0.08</td>
</tr>
</tbody>
</table>
5.3. Part III: Using the correct penalty function in the IC estimators

The third part of the simulation considers the impact of penalty function selection. We consider Model M2 and M3 again, with \( d_1 = d_2 = 40 \), but use the true weakest factor strength \( \delta_1 \) as \( \nu \) in the penalty function \( g_k(T,d) \) in (7) instead assuming strong factor and use \( \nu = 0 \) as in Section 5.1. To compare with Tables 6 for M2 and 8 for M3, we report the proportion of correct rank identification using IC2 estimators in Table 11 for M2 and RMSEs of the IC2 estimators in Table 12 for M3.

Comparing Tables 6 and 11, it is seen that the performance of IC2 estimators using TIPUP improve by using the correct \( \nu \) in the penalty function. We notice that the initial and one step IC estimators using TOPUP with the correct \( \nu \) in the penalty term actually under-perform the ones with the incorrect \( \nu \). The reason for this unusual behavior of TOPUP is unclear. It might be that the magnitude of the estimation of true zero eigenvalue is much larger in this mixed weak and strong factor case, as shown in Figure 1. Hence, reducing the penalty term by using \( \nu = \delta_1 = 0.4 \) resulting in severe over-estimation. However, the TIPUP estimators estimate the true zero eigenvalues more accurately, hence are able to take advantage of the more accurate penalty term.

For Model M3 with all weak factors \( (\delta_0 = \delta_1 = 0.6) \), all IC estimators using the wrong penalty function \( g_k(T,d) \) with \( \nu = 0 \) in (7) identified \((1,1)\) rank pair in all 1000 simulations for all sample sizes. Comparing it with the result shown in Table 12, the importance of using the right \( \nu \) in the penalty function is obvious in this all-weak factor case. Table 12 also shows that IC estimator

<table>
<thead>
<tr>
<th>( T )</th>
<th>IC2-TOP</th>
<th>IC2-TIP</th>
<th>IC2-TOP</th>
<th>IC2-TIP</th>
<th>IC2-TOP</th>
<th>IC2-TIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0</td>
<td>1</td>
<td>0.01</td>
<td>1</td>
<td>0.65</td>
<td>1</td>
</tr>
<tr>
<td>300</td>
<td>0</td>
<td>1</td>
<td>0.29</td>
<td>1</td>
<td>0.90</td>
<td>1</td>
</tr>
<tr>
<td>500</td>
<td>0</td>
<td>1</td>
<td>0.62</td>
<td>1</td>
<td>0.96</td>
<td>1</td>
</tr>
<tr>
<td>1000</td>
<td>0.43</td>
<td>1</td>
<td>0.95</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( T )</th>
<th>IC2-TOP</th>
<th>IC2-TIP</th>
<th>IC2-TOP</th>
<th>IC2-TIP</th>
<th>IC2-TOP</th>
<th>IC2-TIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>1.01</td>
<td>0.91</td>
<td>0.91</td>
<td>0.92</td>
<td>0.98</td>
<td>0.92</td>
</tr>
<tr>
<td>300</td>
<td>1.00</td>
<td>0.03</td>
<td>0.87</td>
<td>0.03</td>
<td>0.23</td>
<td>0.03</td>
</tr>
<tr>
<td>500</td>
<td>1.00</td>
<td>0</td>
<td>0.43</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1000</td>
<td>1.00</td>
<td>0</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
using TIPUP outperforms that using TOPUP, when the right penalty function is used. Comparing Tables 8 and 12, the iterative IC estimators out-perform the ER estimators when \( T = 300, 500, 1000 \). It shows the great potential of IC estimators using a proper \( \nu \) in the penalty function under the weak factor cases. As mentioned earlier, more investigation is needed to determine the proper value for \( \nu \).

5.4. Selection of the optimal \( c \) in IC criteria

To study the empirical property of the optimal constant \( c \) discussed in Remark 8, we simulate data from Model M1 with \( r_1 = r_2 = 5 \), and set \( d_1 = d_2 = 80 \), \( T = 300 \) and set the upper bound of the rank as \( m^* = 10 \). In Figure 2, we show

![Graphs showing the robustified IC criterion using iTOPUP related methods.](image)

(a) Tensor mode \( k = 1 \)

(b) Tensor mode \( k = 2 \)

*Fig 2. Robustified IC criterion using iTOPUP related methods*
respectively the behavior of \( \hat{r}_{k,c,j} \) as a function of \((d_{1,j}, d_{2,j})\), and of \( \hat{r}_{k,c,j} \) and of \( S_{k,c} \) as functions of \( c \), when setting \( T_1 = T_2 = \cdots = T_J = T \) and \( d_{1,j} = d_{2,j} \). The rank \( r_k \) of the factor process can be determined by considering the mapping \( c \rightarrow S_{k,c} \) and choosing \( \hat{r}_{k,c,J} = \hat{r}_{k,c,J} \), where \( \hat{c} \) belongs to an interval of \( c \) implying \( S_{k,c} \approx 0 \) and therefore the value of \( \hat{r}_{k,c,J} \) is a constant function of \( c \). Similar to Hallin and Liška (2007) and Alessi, Barigozzi and Capasso (2010), we see that the second stability interval always delivers an estimated number \( \hat{r}_{k,c,J} \) which is closer to the true \( r_k = 5 \) than the number suggested by the other intervals. That is, the smallest values of \( c \) for which \( \hat{r}_{k,c,j} \) is also close to a constant function of \( j, j \leq J \). Note that the first stability interval always corresponds to the predefined upper bound \( m^* \) and it is thus a non-admissible solution.

6. Real data analysis

In this section, we illustrate the proposed procedures using the Fama–French 10 by 10 monthly return series as an example. According to ten levels of market capital (size) and ten levels of book to equity ratio (BE), stocks are grouped into 100 portfolios. The sampling period used in this excises is from January 1964 to December 2015 for a total of 624 months. There are overall 62,400 observed monthly returns used in this analysis. The data is from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Similar to Wang, Liu and Chen (2019), we subtract from each of the series their corresponding monthly excess market return.

The data was used by Wang, Liu and Chen (2019) to demonstrate the estimation of a matrix factor model using a non-iterative TOPUP procedure. They used an estimator similar to our non-iterative TOPUP based ER estimator without a penalty term to estimate the number of factors and found the estimator suggested \((1,1)\) as the rank – a single factor in the model. In the end they demonstrate the model using \((2,2)\) as the ranks for demonstration of the estimation procedures.

Figure 3 shows the 1st to 3rd largest eigenvalues of using the non- iterative TIPUP and TOPUP procedures, \( h_{0}^{-1}\tau_{k,m_k} \) and \( h_{0}^{-1}\hat{r}_{k,m_k} \) (\( m_k \leq 3 \)), under different lag values \( h_0 \). The pattern of eigenvalues using the iterative TIPUP and TOPUP are similar, thus is omitted. It can be seen from panel (a) of Figure 3 that, using TIPUP procedure, the 1st and 2nd largest eigenvalues reach their maximum value at \( h_0 = 1 \), and tends to decrease as \( h_0 \) increases. However, the 3rd largest eigenvalue reaches its maximum at \( h_0 = 2 \). In contrast, from panel (b) of Figure 3, using TOPUP procedure, the 1st to 3rd largest eigenvalues reach the maximum at \( h_0 = 1 \). The difference of the patterns of estimated singular values indicates possible severe signal cancellation when using \( h_0 = 1 \), according to the suggestions in Han et al. (2020). Hence, we choose \( h_0 = 2 \), and consider IC and ER estimators based on the iterative TIPUP procedure.

Figure 4 shows the estimated rank \( \hat{r}_k \) of the core factor process with different number of iterations, using IC2(TIPUP) with \( \nu = 0 \) (left figure) and ER1(TIPUP) (right figure). It is seen that the iterative algorithms converge very quickly.
Fig 3. The 1st to 3rd largest estimated eigenvalues using the non-iterative TIPUP and TOPUP procedures, $h_0^{-1} \tau_{k,m_k}^2$ and $h_0^{-1} \hat{\tau}_{k,m_k}^2$ ($m_k \leq 3, k \leq 2$), under different maximum lag values $h_0$. The line marked as size is for the row factors, corresponding to the levels of market capital (size), and the line marked as BE is for the column factors, corresponding to the levels of book to equity ratio (BE).

Fig 4. Estimated rank $\hat{r}_k$ of the core factor process against the number of iterations for the Fama-French 10 by 10 series, using IC2 (left) and ER1 (right) estimator based on TIPUP procedure.
Table 13
Estimated ranks using different IC criteria and \( \nu \) for the Fama-French return series

<table>
<thead>
<tr>
<th>( \nu )</th>
<th>0</th>
<th>0.05</th>
<th>0.1</th>
<th>0.15</th>
<th>0.2</th>
<th>0.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC1, IC3</td>
<td>(3,3)</td>
<td>(4,3)</td>
<td>(4,3)</td>
<td>(4,4)</td>
<td>(4,4)</td>
<td>(4,4)</td>
</tr>
<tr>
<td>IC2, IC4, IC5</td>
<td>(2,2)</td>
<td>(2,2)</td>
<td>(2,2)</td>
<td>(2,2)</td>
<td>(3,3)</td>
<td>(3,3)</td>
</tr>
</tbody>
</table>

Fig 5. The eigenvalues \( \tau_{k,m}^2 \), \( 1 \leq m_k \leq d_k \), for each \( 1 \leq k \leq 2 \), using initial TIPUP procedure

Table 13 shows the estimated rank pairs using different IC(TIPUP) estimators and different \( \nu \) parameter for the penalty function. It is seen that IC1 and IC3 tend to select larger models. These rank estimates do not change when we use \( h_0 = 3 \) and 4.

On the other hand, ER1-ER5 in (8) produce exactly the same rank estimate \((1, 3)\) using \( h_0 = 2 \). But these rank estimates change to \((1, 1)\) when we use \( h_0 = 3 \) and 4. Figure 5 shows the estimated eigenvalues \( \tau_{k,m}^2 \), \( 1 \leq m_k \leq d_k \), using the non-iterative initial TIPUP procedure, for \( k = 1 \) (size factor) and \( k = 2 \) (BE factor). It is seen that for the size factor, the largest eigenvalue is more than 20 folds larger than the second largest eigenvalue. As simulation results in Section 5.2 show, in such an unbalanced case, the ER estimator may find it difficult to find the gap between the true non-zero eigenvalues and the true zero eigenvalues, based on the ratio of the eigenvalues. On the other hand, the IC estimator may fare better in such cases since it is based on the level of estimation error of the true zero eigenvalue.

Overall, it seems that \((2, 2)\) or \((3, 3)\) are possibly good choices. More detailed analysis, include goodness-of-fit measures, prediction performance and result interpretation, is needed.

7. Discussions

In this paper, we develop two rank identification estimators, in an attempt to fill a gap on modelling tensor factor model in the literature. Non-iterative and iterative IC and ER estimators, based on similar ideas of the TOPUP and TIPUP
procedures of Chen, Yang and Zhang (2021) and Han et al. (2020) are considered. Theoretical analysis shows that the iterative estimators are much better than the non-iterative estimators. We show that in general the estimators based on TIPUP procedures are better than that based on the TOPUP procedure, due to its fast convergence rate of the estimated eigenvalues under proper conditions. However, in situations when TIPUP procedures also lead to significant signal cancellation, extra care needs to be taken, including increasing the maximum lag $h_0$ in the procedure.

Simulation studies are conducted to compare the finite sample performance of the estimators using the non-iterative and iterative estimation procedures. The results show that the ER estimators based on both TIPUP and TOPUP procedures, and the IC estimators based on TIPUP procedures generally perform very well when all the factors are strong. The ER estimators are better than the IC estimators when some factors are weak, unless one chooses the precise tuning parameter $\nu$ in the IC penalty function, which is a difficult task. When some dominant factors have unrealistically high explanatory power, the ER estimators may not perform well. But the IC estimators still work very well, since the factors are strong. In summary, IC estimator based on iTIPUP shall be used to estimate the number of strong factors, while ER estimators based on iTIPUP are likely to capture weak factors.

Appendix A: Proofs

It suffices to consider $K = 2$ as the TOPUP and TIPUP begin with mode-$k$ matrix unfolding. We observe a matrix time series with $X_t = A_1 F_t A_2^\top + E_t = G_t + E_t \in \mathbb{R}^{d_1 \times d_2}$. Let $U_1, U_2$ be the left $r_1, r_2$ singular vectors of $A_1$ and $A_2$, respectively. Recall $\odot$ is kronecker product and $\otimes$ is tensor product. Without loss of generality, we only consider the case $k = 1$. Denote $E(\cdot) = \mathbb{E}(\cdot | \{F_1, \ldots, F_T\})$. For simplicity, write

\begin{align*}
\mathcal{M}(m, \hat{U}_{1,m}) &= \frac{1}{d^2} \text{tr} \left\{ \left( I - \hat{U}_{1,m} \hat{U}_{1,m}^\top \right) \hat{W}_k \right\}, \\
\mathcal{M}^*(m, \hat{U}_{1,m}) &= \frac{1}{d^2} \text{tr} \left\{ \left( I - \hat{U}_{1,m} \hat{U}_{1,m}^\top \right) \hat{W}_k^* \right\}, \\
\mathcal{M}^{(i)}(m, \hat{U}_{1,m}) &= \frac{1}{d^2} \text{tr} \left\{ \left( I - \hat{U}_{1,m} \hat{U}_{1,m}^\top \right) \hat{W}_k^{(i)} \right\}, \\
\mathcal{M}^{*,(i)}(m, \hat{U}_{1,m}) &= \frac{1}{d^2} \text{tr} \left\{ \left( I - \hat{U}_{1,m} \hat{U}_{1,m}^\top \right) \hat{W}_k^{*,(i)} \right\}.
\end{align*}

Although we present Theorem 1 for the IC and ER estimators together, the proofs of the IC and ER estimators using the same estimation procedure are very similar. Thus, we first prove Theorem 1 based on TOPUP, and then move to iTOPUP, TIPUP and iTIPUP in sequence.
A.1. Proof of Theorem 1 for non-iterative TOPUP

Lemma 1.

Proof. Note that $U_1 U_1^T A_1 = A_1$, $U_2 U_2^T A_2 = A_2$ and $\|U_1\|_2 = \|U_2\|_2 = 1$. It follows that

$$
\left\| \frac{1}{T} \sum_{t=h+1}^{T} \text{mat}_1 \left( A_1 F_{t-h} A_2^T \otimes A_1 F_t A_2^T \right) \right\|_2^2 \leq \|\Theta_{1,0}\|_2 \cdot \|\Theta_{1,0}\|_{op}
$$

(24)

Then (24) follows.
Lemma 2. Let $U_{1\perp} \in \mathbb{R}^{d_1 \times (d_1 - r_1)}$ be the orthonormal complement of $U_1$. Denote $U_{1\perp} = (U_{1j})_{r_1 + 1 \leq j \leq d_1}$ and $U_1 = (U_{1j})_{1 \leq j \leq r_1}$, where $U_{1j} \in \mathbb{R}^{d_1}$. Denote $\eta_d = d^{\delta_1 - \delta_0 / 2} + d^{\delta_1 - 1 / 2}$. Define $\hat{U}_1$ be the estimated top $r_1$ left singular space of TOPUP$_1(x_{1:T})$. Suppose Assumptions I, II, III, IV and V(a) hold. Then, in an event $\Omega_{11} \cap \Omega_0$ with $\mathbb{P}(\Omega_{11}) \geq 1 - e^{-d^2 / 2}$ and $\mathbb{P}(\Omega_0) \geq 1 - T \exp(-C_1 T^0) - \exp(-C_2 T)$ with $C_1, C_2 > 0, 1/\theta_1 = 1/\theta_1 + 2/\theta_2$, there exist a matrix $\hat{U}_1 \in \mathbb{R}^{d_1 \times r_1}$ with $\hat{U}_1 \hat{U}_1^\top = \hat{U}_1 \hat{U}_1^\top$, such that

\[
\|\hat{U}_1 - U_1 \hat{U}_1^\top \|_2 \leq C\left(\frac{\eta_d}{\sqrt{T}}\right), \tag{25}
\]

\[
\|U_1 U_1^\top \hat{U}_1 \|_2 \leq C\left(\frac{\eta_d}{\sqrt{T}}\right), \tag{26}
\]

for all $r_1 + 1 \leq j \leq d_1$ and some $C > 0$.

Proof. Under Assumptions I, II, III, IV, V(a) and Proposition 2, as the derivation of Theorem 1 in Han et al. (2020), in an event $\Omega_{11} \cap \Omega_0$ with $\mathbb{P}(\Omega_{11}) \geq 1 - e^{-d^2 / 2}$ and $\mathbb{P}(\Omega_0) \geq 1 - T \exp(-C_1 T^0) - \exp(-C_2 T)$,

\[
\|\hat{U}_1 \hat{U}_1^\top - U_1 U_1^\top \|_2 \leq C\left(\frac{\eta_d}{\sqrt{T}}\right). \tag{27}
\]

Applying Lemma 17 and Theorem 1 in Chen, Yang and Zhang (2021), for sufficient large $T$, there exists a matrix $Q \in \mathbb{R}^{(d_1 - r_1) \times r_1}$ such that

\[
\|Q\|_2 \leq \|\hat{U}_1 \hat{U}_1^\top - U_1 U_1^\top\|_2
\]

and $\hat{U}_1 = (U_1 + U_{1\perp}Q)(I + Q^\top Q)^{-1/2} \in \mathbb{R}^{d_1 \times r_1}$ is an estimator for $U_1$ with $\hat{U}_1 \hat{U}_1^\top = \hat{U}_1 \hat{U}_1^\top$. Elementary calculation shows that

\[
\|\hat{U}_1 - U_1\|_2 = \|(U_1(I - (I + Q^\top Q)^{1/2}) + U_{1\perp}Q)(I + Q^\top Q)^{-1/2}\|_2 \\
\leq \|I - (I + Q^\top Q)^{1/2}\|_2 + \|Q\|_2 \leq 2\|Q\|_2.
\]

It follows that

\[
\|\hat{U}_1 - U_1 \hat{U}_1^\top \hat{U}_1\|_2 = \|\hat{U}_1 \hat{U}_1^\top \hat{U}_1 - U_1 \hat{U}_1^\top \hat{U}_1\|_2 \\
\leq \|\hat{U}_1 - U_1\|_2 \|\hat{U}_1^\top \hat{U}_1\|_2 \leq 2\|\hat{U}_1 \hat{U}_1^\top - U_1 U_1^\top\|_2.
\]

In view of (27), we have (25).

For (26), notice that in the event $\Omega_{11} \cap \Omega_0$, for all $r_1 + 1 \leq j \leq d_1$,

\[
\|U_1 U_1^\top \hat{U}_1\|_2 = \|\hat{U}_1 \hat{U}_1^\top \hat{U}_1 - U_1 \hat{U}_1^\top \hat{U}_1\|_2 \\
\leq \|\hat{U}_1 \hat{U}_1^\top - U_1 U_1^\top\|_2 \leq C\left(\frac{\eta_d}{\sqrt{T}}\right).
\]

\[\square\]
Lemma 3. Suppose Assumptions I, II, III, IV and V(a) hold. Then, in an event 
$\Omega_{12} \cap \Omega_0$ with $\mathbb{P}(\Omega_{12}) \geq 1 - e^{-d_2/2}$ and $\mathbb{P}(\Omega_0) \geq 1 - T \exp(-C_1T^{\phi}) - \exp(-C_2T)$ 
with $C_1, C_2 > 0, 1/\phi = 1/\theta_1 + 2/\theta_2$, we have

\[
\frac{1}{T-h} \sum_{t=h+1}^{T} \text{mat}_1(A_1F_{t-h}A_2^T \otimes E_t) \leq C \left( \frac{d^{1/2}d_1^{1/2-d_0/2}}{\sqrt{T}} \right),
\]

(28)

\[
\frac{1}{T-h} \sum_{t=h+1}^{T} \text{mat}_1(A_1F_{t-h}A_2^T \otimes U_1^T E_t U_2) \leq C \left( \frac{d^{1/2}d_1^{1/2-d_0/2}}{\sqrt{T}} \right),
\]

(29)

\[
\frac{1}{T-h} \sum_{t=h+1}^{T} \text{mat}_1(E_{t-h} \otimes A_1F_tA_2^T) \leq C \left( \frac{(\sqrt{d_1} + \sqrt{d/d_1})d^{1/2-d_0/2}}{\sqrt{T}} \right),
\]

(30)

\[
\frac{1}{T-h} \sum_{t=h+1}^{T} \text{mat}_1(E_{t-h}U_2 \otimes A_1F_tA_2^T) \leq C \left( \frac{d^{1/2}d_1^{1/2-d_0/2}}{\sqrt{T}} \right),
\]

(31)

\[
\frac{1}{T-h} \sum_{t=h+1}^{T} U_1^T \text{mat}_1(E_{t-h} \otimes A_1F_tA_2^T) \leq C \left( \frac{d^{1/2}d_1^{1/2-d_0/2}}{\sqrt{T}} \right),
\]

(32)

\[
\frac{1}{T-h} \sum_{t=h+1}^{T} \text{mat}_1(E_{t-h} \otimes E_t) \leq C \left( \frac{d}{\sqrt{d_1T}} \right),
\]

(33)

\[
\frac{1}{T-h} \sum_{t=h+1}^{T} \text{mat}_1(E_{t-h}U_2 \otimes A_t) \leq C \left( \frac{d^{1/2}}{\sqrt{T}} \right),
\]

(34)

\[
\frac{1}{T-h} \sum_{t=h+1}^{T} \text{mat}_1(E_{t-h} \otimes U_1^T E_t U_2) \leq C \left( \frac{d^{1/2}}{\sqrt{T}} \right),
\]

(35)

\[
\frac{1}{T-h} \sum_{t=h+1}^{T} \text{mat}_1(E_{t-h}U_2 \otimes U_1^T E_t U_2) \leq C \left( \frac{d^{1/2}}{\sqrt{T}} \right),
\]

(36)

\[
\frac{1}{T-h} \sum_{t=h+1}^{T} U_1^T \text{mat}_1(E_{t-h} \otimes U_1^T E_t U_2) \leq C \left( \frac{d^{1/2}}{\sqrt{d_1T}} \right),
\]

(37)

\[
\frac{1}{T-h} \sum_{t=h+1}^{T} U_1^T \text{mat}_1(E_{t-h}U_2 \otimes U_1^T E_t U_2) \leq C \left( \frac{1}{\sqrt{T}} \right),
\]

(38)

for some constant positive $C$ depending on $K$ only.

Proof. Note that $U_1U_1^T A_1 = A_1$ and $U_2U_2^T A_2 = A_2$. Let

\[
\|\Delta_1\|_2 := \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} \text{mat}_1(A_1F_{t-h}A_2^T \otimes E_t) \right\|_2
\]
\[ Y. \text{Han et al.} \]

\[ = \left\| \frac{1}{T - h} \sum_{t=h+1}^{T} U_1^\top \mat_1(A_1 F_{t-h} A_2^\top U_2 \otimes E_t) \right\|_2. \]

By Theorem 1 in Chen, Yang and Zhang (2021),

\[ \mathbb{E} \| \Delta_1 \|_2 \leq \frac{\sigma \sqrt{2T(\sqrt{r_1} + \sqrt{r_2 d_1 d_2})}}{T - h} \| \Theta_{1,0}^* \|_2^{1/2}. \]

Elementary calculation shows that

\[ \left\| \sum_{t=h+1}^{T} \mat_1(A_1 F_{t-h} A_2^\top \otimes E_t) \right\|_2 - \left\| \sum_{t=h+1}^{T} \mat_1(A_1 F_{t-h} A_2^\top \otimes E_t^*) \right\|_2 \]

\[ \leq \left\| \mat_1(A_1 A_2^\top \otimes I_{d_1}), \ldots, \mat_1(A_1 A_2^\top \otimes I_{d_1}) \right\|_2 \]

\[ \leq \sqrt{T} \| \Theta_{1,0}^* \|_2^{1/2} \left\| \begin{pmatrix} E_{h+1} - E_{h+1}^* \\ \vdots \\ E_T - E_T^* \end{pmatrix} \right\|_F. \]

That is, \( \left\| \sum_{t=h+1}^{T} \mat_1(A_1 F_{t-h} A_2^\top \otimes E_t) \right\|_2 \) is a \( \sigma \sqrt{T} \| \Theta_{1,0}^* \|_2^{1/2} \) Lipschitz function in \((E_1, \ldots, E_T)\). Then, by Gaussian concentration inequalities for Lipschitz functions,

\[ \mathbb{P} \left( \left\| \sum_{t=h+1}^{T} \mat_1(A_1 F_{t-h} A_2^\top \otimes E_t) \right\|_2 - \frac{\sigma (2T)^{1/2}(\sqrt{r_1} + \sqrt{d_1 d_2 r_2})}{T - h} \| \Theta_{1,0}^* \|_2^{1/2} \right) \]

\[ \geq \frac{\sigma \sqrt{T}}{T - h} \| \Theta_{1,0}^* \|_2^{1/2} x \leq 2e^{-\frac{x^2}{2}}. \]

This implies that with \( x \approx \sqrt{d} \), in an event \( \Omega_\alpha \) with at least probability \( 1 - e^{-d_2/6} \),

\[ \| \Delta_1 \|_2 \leq \frac{C_0 \sqrt{T d}}{T - h} \| \Theta_{1,0}^* \|_2^{1/2}, \quad (39) \]

where \( C_0 > 0 \) depends on \( K \) only. Then, using Proposition 2, in the event \( \Omega_\alpha \cap \Omega_\beta \), (28) follows. Similar arguments yield (29) in the event \( \Omega_\alpha \cap \Omega_\beta \).
Let
\[ \|\Delta_2\|_2 := \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} \text{mat}_1(E_{t-h} \otimes A_1F_tA_2^\top) \right\|_2 \]
\[ = \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} \text{mat}_1(E_{t-h} \otimes U_1^\top A_1F_tA_2^\top U_2) \right\|_2. \]

By Theorem 1 in Chen, Yang and Zhang (2021),
\[ \mathbb{E}\|\Delta_2\|_2 \leq \frac{\sigma \sqrt{2T}((\sqrt{d_1}+\sqrt{d_2})n_0)}{T-h}\|\Theta_{1,0}\|_{op}^{1/2}. \]

Then, using Proposition 2, in an event \( \Omega_b \cap \Omega_0 \) with \( \mathbb{P}(\Omega_b) \geq 1 - e^{-d_2}/6, (30) \) follows from the same argument as the above step. Similarly, in the event \( \Omega_b \cap \Omega_0 \), we can obtain (31) and (32).

We split the sum into two terms over the index sets, \( S_1 = \{(h, 2h] \cup (3h, 4h] \cup \cdots \} \cap (h, T] \) and its complement \( S_2 \) in \( (h, T] \), so that \( \{E_{t-h}, t \in S_a\} \) is independent of \( \{E_t, t \in S_a\} \) for each \( a = 1, 2 \). Let \( n_a = |S_a| \). By Lemma 19(ii), for any \( x > 0 \),
\[ \mathbb{P}\left( \left\| \sum_{t \in S_a} \text{mat}_1(E_{t-h} \otimes E_t) \right\|_S \geq d_1 \sqrt{d_2} + 2d_2 \sqrt{d_1} n_a + x^2 + \sqrt{n_a}x + 3\sqrt{d_1}d_2x \right) \leq 2e^{-x^2/2}. \]

With \( x \asymp \sqrt{d_2} \) and some constant \( C_1 \) depending on \( K \) only, we have
\[ \mathbb{P}\left( \|\Delta_3\|_2 \geq \frac{C_1d_1 \sqrt{d_2} + C_2d_2 \sqrt{d_1}T}{T} \right) \leq e^{-d_2}/6. \]

Then, as in the derivation of \( \|\Delta_4\|_2 \) in the proof of Theorem 1 in Chen, Yang and Zhang (2021), in an event \( \Omega_c \) with \( \mathbb{P}(\Omega_c) \geq 1 - e^{-d_2}/6, (33) \) follows. Similar arguments yield (34), (35), (36), (37) and (38) in the event \( \Omega_c \). Set \( \Omega_{12} = \Omega_a \cap \Omega_b \cap \Omega_c \), then \( \mathbb{P}(\Omega_{12}) \geq 1 - e^{-d_2}/2. \)

**Lemma 4.** Suppose Assumptions I, II, III, IV and V(a) hold. In an event \( \Omega_1 \) with \( \mathbb{P}(\Omega_1) \geq 1 - e^{-d_2} - T \exp(-C_1T^{\vartheta}) - \exp(-C_2T) \) with \( C_1, C_2 > 0, 1/\vartheta = 1/\vartheta_1 + 2/\vartheta_2, \) for any fixed \( m \) with \( m > r_1, \mathcal{M}(r_1, \hat{U}_1) - \mathcal{M}(m, \hat{U}_{1,m}) \leq C\beta_{d,T}, \) where \( C > 0, \)

\[ \beta_{d,T} = \frac{1}{Td_1^{1/2}} + \frac{d_1}{Td_1^{1/2} + \delta_0} + \frac{d_1^{1/2} \eta_d}{Td_1^{1/2} + 3\delta_0} + \frac{\eta_d}{T^{3/2}d_1^{1/2} + 4\delta_0}, \]

and \( \eta_d = \delta^{1/2}d_1^{1/2} + d_1^{1/2}d_1^{-1/2}. \)
Proof. Let \( \Omega_1 = \Omega_{11} \cap \Omega_{12} \cap \Omega_0 \), where \( \Omega_{11}, \Omega_{12} \) are defined in Lemma 2 and 3, respectively, and \( \Omega_0 \) is the event in Proposition 2. Let \( U_{1,r} = U_1 \) and \( U_{1,m} = (U_{11}, U_{12}, \ldots, U_{1m}) \).

\[
\begin{align*}
|M(m, \hat{U}_{1,m}) - M(r_1, \hat{U}_1)| &\leq |M(m, \hat{U}_{1,m}) - M(r_1, U_1)| \\
&+ |M(r_1, \hat{U}_1) - M(r_1, U_1)| \\
&\leq 2 \max_{r_1 < m \leq m_1} |M(m, \hat{U}_{1,m}) - M(r_1, U_1)|.
\end{align*}
\]

As \( m_1 \) is fixed, it is sufficient to prove for each \( m \) with \( m > r_1 \), in the event \( \Omega_1 \),

\[
|\mathcal{M}(m, \hat{U}_{1,m}) - \mathcal{M}(r_1, U_1)| \leq C \beta_{d,T}.
\]

(41)

In the following, we shall only work on \( \Omega_1 \).

Elementary calculation shows that

\[
\begin{align*}
d^2 \cdot M(m, \hat{U}_{1,m}) &- d^2 \cdot M(r_1, U_1) \\
&= \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^T \text{tr}\left\{ \left(U_1 U_1^\top - \hat{U}_{1,m} \hat{U}_{1,m}^\top \right) \text{mat}_1(X_{t-h} \otimes X_t) \right. \\
&\quad \left. \cdot \text{mat}_1^\top(X_{s-h} \otimes X_s) \right\} \\
&+ \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^T \text{tr}\left\{ \text{mat}_1^\top(A_1 F_{s-h} A_2^\top \otimes X_s) \left(U_1 U_1^\top - \hat{U}_{1,m} \hat{U}_{1,m}^\top \right) \right. \\
&\quad \left. \cdot \text{mat}_1(A_1 F_{t-h} A_2^\top \otimes X_t) \right\} \\
&+ \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^T \text{tr}\left\{ \text{mat}_1^\top(E_{s-h} \otimes X_s) \left(U_1 U_1^\top - \hat{U}_{1,m} \hat{U}_{1,m}^\top \right) \right. \\
&\quad \left. \cdot \text{mat}_1(E_{t-h} \otimes X_t) \right\} \\
&+ \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^T \text{tr}\left\{ \text{mat}_1^\top(A_1 F_{s-h} A_2^\top \otimes X_s) \left(U_1 U_1^\top - \hat{U}_{1,m} \hat{U}_{1,m}^\top \right) \right. \\
&\quad \left. \cdot \text{mat}_1(E_{t-h} \otimes X_t) \right\} \\
&+ \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^T \text{tr}\left\{ \text{mat}_1^\top(E_{s-h} \otimes X_s) \left(U_1 U_1^\top - \hat{U}_{1,m} \hat{U}_{1,m}^\top \right) \right. \\
&\quad \left. \cdot \text{mat}_1(A_1 F_{t-h} A_2^\top \otimes X_t) \right\} \\
&:= \text{I} + \text{II} + \text{III} + \text{IV}.
\end{align*}
\]

Note that \( \text{III} = \text{IV} \).
We consider each term in turn. Note that

\[
\tilde{\mathbb{I}} = \mathbb{I} := \mathbb{I} = \mathbb{I} = \mathbb{I} = \mathbb{I} = \mathbb{I} = \mathbb{I} = \mathbb{I} = \mathbb{I} = \mathbb{I} = \mathbb{I} = \mathbb{I} = \mathbb{I} = \mathbb{I}
\]

We first consider II.

\[
\mathbb{I} = \sum_{h=1}^{\delta_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^T(E_{s-h} \otimes X_s) \left( U_1 U_1^T - \hat{U}_1 \hat{U}_1^T \right) \right\} 
\]

\[
\cdot \text{mat}_1(E_{t-h} \otimes X_t) \right\}
\]

\[
\mathbb{I} = \sum_{h=1}^{\delta_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^T(E_{s-h} \otimes E_s) \left( U_1 U_1^T - \hat{U}_1 \hat{U}_1^T \right) \right\} 
\]

\[
\cdot \text{mat}_1(E_{t-h} \otimes E_t) \right\}
\]

\[
\mathbb{I} + \mathbb{I}_2 + 2 \mathbb{I}_3.
\]

We consider each term in turn. Note that

\[
\hat{U}_1 \hat{U}_1^T - U_1 U_1^T = \left( \hat{U}_1 - U_1 \hat{U}_1^T \hat{U}_1 \right) \left( \hat{U}_1 - U_1 \hat{U}_1^T \hat{U}_1 \right)^T 
\]

\[
+ U_1 \hat{U}_1^T \hat{U}_1 \left( \hat{U}_1 - U_1 \hat{U}_1^T \hat{U}_1 \right)^T 
\]

\[
+ \left( \hat{U}_1 - U_1 \hat{U}_1^T \hat{U}_1 \right) \left( U_1 \hat{U}_1^T \hat{U}_1 \right)^T,
\]

where \( \hat{U}_1 \) is defined in Lemma 2. Then, by Lemma 2 and 3,

\[
\mathbb{I}_1 = \sum_{h=1}^{\delta_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^T(E_{s-h} \otimes E_s) \left( U_1 U_1^T - \hat{U}_1 \hat{U}_1^T \right) \right\} 
\]

\[
\cdot \text{mat}_1(E_{t-h} \otimes E_t) \right\}
\]

\[
+ \sum_{h=1}^{\delta_0} \sum_{j=r_1+1}^{m} \frac{1}{(T-h)^2} \text{tr}\left\{ \text{mat}_1^T(E_{s-h} \otimes E_s) \hat{U}_j \hat{U}_j^T \right\} 
\]

\[
\cdot \text{mat}_1(E_{t-h} \otimes E_t) \right\}
\]

\[
\leq r_1 \sum_{h=1}^{\delta_0} \left\| \frac{1}{(T-h)} \sum_{t=h+1}^{T} \text{mat}_1(E_{t-h} \otimes E_t) \right\|_2
\]
by the fact $\eta_d = o(\sqrt{T})$.

$$\Pi_2 = \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^\top (E_{s-h} \otimes A_1 F_s A_2^\top) \left( U_{1} U_{1}^\top - \hat{U}_1 \hat{U}_1^\top \right) \right\} \cdot \text{mat}_1(E_{t-h} \otimes A_1 F_t A_2^\top) \}$$
$$\sum_{h=1}^{h_0} \sum_{j=r_1+1}^{m} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^\top (E_{s-h} \otimes A_1 F_s A_2^\top) \hat{U}_{1j} \hat{U}_{1j}^\top \right\} \cdot \text{mat}_1(E_{t-h} \otimes A_1 F_t A_2^\top) \}$$
$$= O \left( \frac{(d_1 + d/d_1)d^{1-\delta_0}}{T} \right).$$

$$\Pi_3 = \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^\top (E_{s-h} \otimes U_{1}^\top A_1 F_s A_2^\top U_2) \left( U_{1} U_{1}^\top - \hat{U}_1 \hat{U}_1^\top \right) \right\} \cdot \text{mat}_1(E_{t-h} \otimes U_{1}^\top E_t U_2) \}$$
$$\sum_{h=1}^{h_0} \sum_{j=r_1+1}^{m} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^\top (E_{s-h} \otimes U_{1}^\top A_1 F_s A_2^\top U_2) \hat{U}_{1j} \hat{U}_{1j}^\top \right\} \cdot \text{mat}_1(E_{t-h} \otimes U_{1}^\top E_t U_2) \}$$
$$\leq (m + O_T(1)) \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,h=1}^{T} \text{tr}\left\{ \text{mat}_1^\top (E_{s-h} \otimes U_{1}^\top A_1 F_s A_2^\top U_2) \right\} \cdot \text{mat}_1(E_{t-h} \otimes U_{1}^\top E_t U_2) \right\}$$
$$\sum_{s,t=h+1}^{T} \text{mat}_1(E_{s-h} \otimes U_{1}^\top E_t U_2) \right\}$$
$$= O \left( \frac{d_1^2}{d_1 T} + \frac{d_1 d^{1-\delta_0}}{T} \right).$$

by Lemma 2 and 3. Combining the bounds of $\Pi_1, \Pi_2$ and $\Pi_3$, we have

$$\Pi = O \left( \frac{d^2}{d_1 T} + \frac{d_1 d^{1-\delta_0}}{T} \right). \quad (43)$$
Next, we consider $\text{III.}$

$$ \text{III} = \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^\top (A_1 F_{s-h} A_2^\top \otimes X_s) \left( U_1 U_1^\top - \hat{U}_{1,m} \hat{U}_{1,m}^\top \right) \right\}$$

$$\cdot \text{mat}_1 (E_{t-h} \otimes X_t) \left\{ \right\}$$

$$= \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^\top (A_1 F_{s-h} A_2^\top \otimes A_1 F_s A_2^\top) \right\}$$

$$\cdot \left( U_1 U_1^\top - \hat{U}_{1,m} \hat{U}_{1,m}^\top \right) \text{mat}_1 (E_{t-h} \otimes A_1 F_s A_2^\top) \left\{ \right\}$$

$$+ \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^\top (A_1 F_{s-h} A_2^\top \otimes E_s) \left( U_1 U_1^\top - \hat{U}_{1,m} \hat{U}_{1,m}^\top \right) \right\}$$

$$\cdot \text{mat}_1 (E_{t-h} \otimes E_t) \left\{ \right\}$$

$$+ \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^\top (A_1 F_{s-h} A_2^\top \otimes A_1 F_s A_2^\top) \right\}$$

$$\cdot \left( U_1 U_1^\top - \hat{U}_{1,m} \hat{U}_{1,m}^\top \right) \text{mat}_1 (E_{t-h} \otimes A_1 F_s A_2^\top) \left\{ \right\}$$

$$+ \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^\top (A_1 F_{s-h} A_2^\top \otimes A_1 F_s A_2^\top) \right\}$$

$$\cdot \left( U_1 U_1^\top - \hat{U}_{1,m} \hat{U}_{1,m}^\top \right) \text{mat}_1 (E_{t-h} \otimes E_t) \left\{ \right\}$$

$$:= \text{III}_1 + \text{III}_2 + \text{III}_3 + \text{III}_4.$$

Again, we bound each term in turn. By Lemma 1, 2, 3,
\begin{align*}
+ & \sum_{h=1}^{h_0} \frac{m-r_1}{(T-h)^2} \sum_{s,t=h+1}^{T} \left\| \text{mat}_1(A_1F_{s-h}A_2^T U_2 \otimes A_1F_sA_2^T) \right\|_2 \\
\cdot & \left\| \text{mat}_1(E_{t-h}U_2 \otimes A_1F_sA_2^T) \right\|_2 \left\| U_1U_1^T \hat{U}_1 \right\|_2 \\
= & O \left( \frac{d_1^{1/2}d_2^{1/2-3\delta_0/2}k_2}{T} \right),
\end{align*}

where \( \eta_d = d_1^{3-\delta_0/2} + d_1^{1-1/2}. \)

\[ \text{III}_2 = \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \left\| \text{mat}_1(A_1F_{s-h}A_2^T U_2 \otimes E_s) \right\|_2 \\
\cdot \left\| \text{mat}_1(E_{t-h}U_2 \otimes E_t) \right\|_2 \left\| U_1^T\hat{U}_1^T - \hat{U}_{1,m}^T \hat{U}_{1,m}^T \right\|_2 \\
\leq & O \left( \frac{d_1^{1/2}d_2^{1/2}k_2}{\sqrt{T}} \right) = O \left( \frac{d_1^{1/2}d_2^{1/2}k_2}{T^{3/2}} \right). \]

Similarly,

\[ \text{III}_3 = \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \left\| \text{mat}_1(A_1F_{s-h}A_2^T U_2 \otimes U_1^T E_s U_2) \right\|_2 \\
\cdot \left\| U_1^T - \hat{U}_{1,m}^T \hat{U}_{1,m}^T \right\|_2 \\
\cdot \left\| \text{mat}_1(E_{t-h}U_2 \otimes U_1^T A_1F_sA_2^T U_2) \right\|_2 \\
= O \left( \frac{d_1^{1/2}d_2^{1/2}k_2}{T^{3/2}} \right), \]

and

\[ \text{III}_4 = \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \left\| \text{mat}_1(A_1F_{s-h}A_2^T U_2 \otimes U_1^T A_1F_sA_2^T U_2) \right\|_2 \\
\cdot \left\| U_1^T - \hat{U}_{1,m}^T \hat{U}_{1,m}^T \right\|_2 \\
\cdot \left\| \text{mat}_1(E_{t-h}U_2 \otimes U_1^T E_t U_2) \right\|_2 \\
= O \left( \frac{d_1^{1/2}d_2^{1/2}k_2}{T} \right). \]
Combing the bounds of \(\text{III}_1, \text{III}_2, \text{III}_3\) and \(\text{III}_4\), we have

\[
\text{III} = O \left( \frac{d_1^{1/2} d_5^{3/2 - 3\delta_0/2} \eta_d}{T} + \frac{d_3^{1/2} d_4^{3/2 - 3\delta_0/2} \eta_d}{T^3/2} \right),
\]

(44)

In view of (43) and (44),

\[
|\text{II} + \text{III} + \text{IV}| \leq C_1 \left( \frac{d^2}{d_1 T} + \frac{d_1 d_3^{1/2} d_4^{3/2 - 3\delta_0/2} \eta_d}{T} + \frac{d_3^{1/2} d_4^{3/2 - 3\delta_0/2} \eta_d}{T^3/2} \right)
\]

(45)

\[
\leq C_2 \left( d^2 \beta_{d,T} \right).
\]

Next, we consider \(I\).

\[
I = \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^T \text{tr} \left\{ \text{mat}_1^T \left( A_1 F_{s-h} A_2^T \otimes X_s \right) \left( U_1 U_1^T - \hat{U}_1, m \hat{U}_1^T \right) \right\}
\]

\[
\cdot \text{mat}_1 \left( A_1 F_{t-h} A_2^T \otimes X_t \right) \}
\]

\[
= \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^T \text{tr} \left\{ \text{mat}_1^T \left( A_1 F_{s-h} A_2^T \otimes X_s \right) \left( I - \hat{U}_1, m \hat{U}_1^T \right) \right\}
\]

\[
\cdot \text{mat}_1 \left( A_1 F_{t-h} A_2^T \otimes X_t \right) \}
\]

\[
= \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \text{tr} \left\{ \sum_{s=h+1}^T \text{mat}_1^T \left( A_1 F_{s-h} A_2^T \otimes X_s \right) \left( I - \hat{U}_1, m \hat{U}_1^T \right) \right\}
\]

\[
\cdot \sum_{t=h+1}^T \text{mat}_1 \left( A_1 F_{t-h} A_2^T \otimes X_t \right) \}
\]

\[
\geq 0,
\]

using the fact that \(I - \hat{U}_1, m \hat{U}_1^T\) is positive semi-definite, \(x^T \hat{U}_1, m \hat{U}_1^T x \leq x^T x\).

Using the definition of our optimization target, we have \(\mathcal{M}(m, \hat{U}_1, m) \leq \mathcal{M}(r_1, U_1) \leq \mathcal{M}(r_1, U_1)\). It follows that \(I + II + III + IV \leq 0\). As \(I \geq 0\), by (45), \(I = O(d^2 \beta_{d,T})\). In summary,

\[
\left| \mathcal{M}(m, \hat{U}_1, m) - \mathcal{M}(r_1, U_1) \right| \leq C \left( \beta_{d,T} \right).
\]

Lemma 5. Suppose Assumptions I, II, III, IV and V(a) hold. In an event \(\Omega_1\) with \(\mathbb{P}(\Omega_1) \geq 1 - e^{-d_2} - T \exp(-C_1 T^{3/2}) - \exp(-C_2 T)\) with \(C_1, C_2 \geq 0, 1/\theta = 1/\theta_1 + 1/\theta_2\), there exist constant \(c_m > 0\) and \(C > 0\), such that \(\mathcal{M}(m, \hat{U}_1, m) - \mathcal{M}(r_1, \hat{U}_1) \geq (c_m + o(1))d^{-2\delta_1} + C \gamma_{d,T}\), where

\[
\gamma_{d,T} = \frac{1}{T d_{80}^2} + \frac{1}{T d_1} + \frac{1}{T^{1/2} d_{41}^{1/2} / 2 + 3\delta_0 / 2} + \frac{d_1^{1/2} \eta_d}{T d_{41}^{1/2} + 3\delta_0 / 2},
\]

and \(\eta_d = d^{\delta_1 - \delta_0 / 2} + d^{\delta_1} d_1^{-1/2}\).
Proof. Let $\Omega_1 = \Omega_{11} \cap \Omega_{12} \cap \Omega_0$, where $\Omega_{11}, \Omega_{12}$ are defined in Lemma 2 and 3, respectively, and $\Omega_0$ is the event in Proposition 2. In the following, we shall only work on $\Omega_1$.

Let $U_{1,r_1} = U_1$ and $U_{1,m} = (U_{11}, U_{12}, ..., U_{1m})$.

$$d^2 \cdot \mathcal{M}(m, \hat{U}_{1,m}) - d^2 \cdot \mathcal{M}(r_1, \hat{U}_1)$$
$$= \sum_{h=1}^{\theta_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^T \text{tr} \left\{ \left( \hat{U}_1 \hat{U}_1^\top - \hat{U}_{1,m} \hat{U}_{1,m}^\top \right) \cdot \text{mat}_1(X_{t-h} \otimes X_t) \right\}$$
$$= \sum_{h=1}^{\theta_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^T \text{tr} \left\{ \text{mat}_1^\top (A_1 F_{s-h} A_2^\top \otimes X_s) \hat{U}_{1j} \hat{U}_{1j}^\top \right\}$$
$$\quad \cdot \text{mat}_1(A_1 F_{t-h} A_2^\top \otimes X_t) \}$$
$$+ \sum_{h=1}^{\theta_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^T \text{tr} \left\{ \text{mat}_1^\top (E_{s-h} \otimes X_s) \hat{U}_{1j} \hat{U}_{1j}^\top \right\}$$
$$\quad \cdot \text{mat}_1(E_{t-h} \otimes X_t) \}$$
$$+ \sum_{h=1}^{\theta_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^T \text{tr} \left\{ \text{mat}_1^\top (A_1 F_{s-h} A_2^\top \otimes X_s) \hat{U}_{1j} \hat{U}_{1j}^\top \right\}$$
$$\quad \cdot \text{mat}_1(A_1 F_{t-h} A_2^\top \otimes X_t) \}$$
$$+ \sum_{h=1}^{\theta_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^T \text{tr} \left\{ \text{mat}_1^\top (E_{s-h} \otimes X_s) \hat{U}_{1j} \hat{U}_{1j}^\top \right\}$$
$$\quad \cdot \text{mat}_1(A_1 F_{t-h} A_2^\top \otimes X_t) \}$$
$$:= I + II + III + IV.$$

Note that $III = IV$.

We first consider $I$.

$$I = \sum_{h=1}^{\theta_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^T \text{tr} \left\{ \text{mat}_1^\top (A_1 F_{s-h} A_2^\top \otimes A_1 F_s A_2^\top) \hat{U}_{1j} \hat{U}_{1j}^\top \right\}$$
$$\quad \cdot \text{mat}_1(A_1 F_{t-h} A_2^\top \otimes A_1 F_s A_2^\top) \}$$
$$+ \sum_{h=1}^{\theta_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^T \text{tr} \left\{ \text{mat}_1^\top (A_1 F_{s-h} A_2^\top \otimes E_s) \hat{U}_{1j} \hat{U}_{1j}^\top \right\}$$
$$\quad \cdot \text{mat}_1(A_1 F_{t-h} A_2^\top \otimes E_t) \}$$
\[ + 2 \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \mathbf{mat}_1^T (A_1 F_{s-h} A_2^T \otimes E_s) \hat{U}_{1j} \hat{U}_{1j}^T \right. \\
\left. \cdot \mathbf{mat}_1 (A_1 F_{t-h} A_2^T \otimes A_1 F_t A_2^T) \right\} \]
\[ := I_1 + I_2 + 2I_3. \]

For \( I_1 \),
\[ I_1 = \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \hat{U}_{1j}^T \mathbf{mat}_1 (A_1 F_{t-h} A_2^T \otimes A_1 F_t A_2^T) \right. \\
\left. \cdot \mathbf{mat}_1^T (A_1 F_{s-h} A_2^T \otimes A_1 F_s A_2^T) \hat{U}_{1j} \right\} \]
\[ = \sum_{j=m+1}^{r_1} \text{tr}\left\{ \hat{U}_{1j}^T \mathbf{mat}_1 (\Theta_{1,h}) \mathbf{mat}_1^T (\Theta_{1,h}) U_1 U_1^T \hat{U}_{1j} \right\} \]
\[ \geq \sum_{j=m+1}^{r_1} \lambda_{\min} \left( \sum_{h=1}^{h_0} \mathbf{U}_{1j}^T \mathbf{mat}_1 (\Theta_{1,h}) \mathbf{mat}_1^T (\Theta_{1,h}) U_1 \right) \text{tr}(\hat{U}_{1j}^T U_1 U_1^T \hat{U}_{1j}) \]
\[ = \sigma_{r_1}^2 \left( U_1^T \mathbf{E} \mathbf{mat}_1 (\text{TOPUP}_1) \right) \sum_{j=m+1}^{r_1} \hat{U}_{1j}^T U_1 U_1^T \hat{U}_{1j}. \]

Let \( W_1 A_0 W_2^T \) be singular value decomposition of \( U_1^T \hat{U}_1 \), where
\[ \Lambda_0 = \begin{pmatrix} \phi_1 & \cdots & \phi_{r_1} \end{pmatrix}, \]
with \( \phi_1 \geq \phi_2 \geq \cdots \geq \phi_{r_1} \). Note that \( W_1 W_1^T = W_1^T W_1 = W_2 W_2^T = W_2^T W_2 = I_{r_1} \). By Lemma 16,
\[ \phi_{r_1}^2 = \sigma_{\min}(U_1^T \hat{U}_1) = 1 - \| \sin(\Theta(\hat{U}_1, U_1)) \|_2^2 \geq 1 - \| \hat{U}_1 U_1^T - U_1 U_1^T \|_2^2. \]

It follows that
\[ \hat{U}_{1j}^T U_1 U_1^T \hat{U}_{1j} = e_j^T \hat{U}_{1j}^T U_1 U_1^T \hat{U}_{1j} e_j = e_j^T W_2 A_0 W_1^T W_1 A_0 W_2^T e_j = e_j^T W_2 A_0^2 W_2^T e_j \]
\[ \geq \lambda_{\min}(A_0^2) e_j^T W_2 W_2^T e_j = \phi_{r_1}^2 \]
\[ \geq 1 - \| \hat{U}_1 U_1^T - U_1 U_1^T \|_2^2, \]
where \( e_j \) is a \( r_1 \times 1 \) vector with 1 at \( j \)-th element and 0 otherwise. Thus, we can obtain
\[ I_1 \geq \sigma_{r_1}^2 \left( U_1^T \mathbf{E} \mathbf{mat}_1 (\text{TOPUP}_1) \right) (r_1 - m) \left( 1 - \| \hat{U}_1 U_1^T - U_1 U_1^T \|_2 \right). \]

By Proposition 2, in the event \( \Omega_0 \), \( \sigma_{r_1} \left( U_1^T \mathbf{E} \mathbf{mat}_1 (\text{TOPUP}_1) \right) = \tau_{r_1} \), with \( 0 \leq \delta_0 \leq \delta_1 < 1 \). Hence, there exists a constant \( c_m > 0 \) such that in the
event $\Omega_1$, 

$$I_1 \geq (c_m + o(1))d^{2-2\delta_1}. \quad (48)$$

By Lemma 3, 

$$I_2 = \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}^\top_1 (A_1 F_{s-h} A_2^\top \otimes E_s) \hat{U}_{1j} \hat{U}_{1j}^\top \right\}$$

$$= O \left( \frac{d^{2-\delta_1}}{T} \right),$$

and 

$$I_3 = \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}^\top_1 (A_1 F_{s-h} A_2^\top \otimes U_1^\top E_s U_2) \hat{U}_{1j} \hat{U}_{1j}^\top \right\}$$

$$= O \left( \frac{d^{3/2-3\delta_1/2}}{T^{1/2}} \right).$$

This implies that 

$$I_2 + I_3 = O \left( \frac{d^{2-\delta_1}}{T} + \frac{d^{3/2-3\delta_1/2}}{T^{1/2}} \right). \quad (49)$$

Next, we consider $I_2$.

$$I_2 = \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}^\top_1 (E_{s-h} \otimes E_s) \hat{U}_{1j} \hat{U}_{1j}^\top \right\}$$

$$+ \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}^\top_1 (E_{s-h} \otimes A_1 F_s A_2^\top) \hat{U}_{1j} \hat{U}_{1j}^\top \right\}$$

$$+ 2 \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}^\top_1 (E_{s-h} \otimes A_1 F_s A_2^\top) \hat{U}_{1j} \hat{U}_{1j}^\top \right\}$$

$$:= I_{11} + I_{12} + 2I_{13}. $$
By Lemma 3,

$$\Pi_1 = O\left( \frac{d^2}{Td_1} \right)$$

Using the same decomposition (42),

$$\Pi_2 = \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T - h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^\top (E_{s-h} \otimes A_1 F_s A_2^\top) \right. \right.$$  

$$\cdot \left( \tilde{U}_{1j} - U_1 \tilde{U}_1^\top \tilde{U}_{1j} \right) \left( \tilde{U}_{1j} - U_1 \tilde{U}_1^\top \tilde{U}_{1j} \right)^\top$$  

$$\cdot \left. \text{mat}_1 (E_{t-h} \otimes A_1 F_t A_2^\top) \right\}$$

$$+ 2 \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T - h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^\top (E_{s-h} \otimes A_1 F_s A_2^\top) \right. \right.$$  

$$\cdot \left( \tilde{U}_{1j} - U_1 \tilde{U}_1^\top \tilde{U}_{1j} \right) \left( U_1 \tilde{U}_1^\top \tilde{U}_{1j} \right)^\top$$  

$$\cdot \left. \text{mat}_1 (E_{t-h} \otimes A_1 F_t A_2^\top) \right\}$$

$$+ \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_2} \frac{1}{(T - h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^\top (E_{s-h} \otimes A_1 F_s A_2^\top) \right. \right.$$  

$$\cdot \left( U_1 \tilde{U}_1^\top \tilde{U}_{1j} \right) \left( U_1 \tilde{U}_1^\top \tilde{U}_{1j} \right)^\top$$  

$$\cdot \left. \text{mat}_1 (E_{t-h} \otimes A_1 F_t A_2^\top) \right\}$$

$$= O\left( \frac{d_1 d_1 - h_0}{T^2} + \frac{d_2 - h_0}{Td_1} \right)$$

Similarly,

$$\Pi_3 = \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T - h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^\top \left( E_{s-h} \otimes U_1^\top \right) A_1 F_s A_2^\top U_2 \right. \right.$$  

$$\cdot \left( \tilde{U}_{1j} - U_1 \tilde{U}_1^\top \tilde{U}_{1j} \right) \left( \tilde{U}_{1j} - U_1 \tilde{U}_1^\top \tilde{U}_{1j} \right)^\top$$  

$$\cdot \left. \text{mat}_1 (E_{t-h} \otimes U_1^\top \hat{E} t U_2) \right\}$$

$$+ \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T - h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^\top \left( E_{s-h} \otimes U_1^\top \right) A_1 F_s A_2^\top U_2 \right. \right.$$  

$$\cdot \left( U_1 \tilde{U}_1^\top \tilde{U}_{1j} \right) \left( U_1 \tilde{U}_1^\top \tilde{U}_{1j} \right)^\top$$  

$$\cdot \left. \text{mat}_1 (E_{t-h} \otimes U_1^\top \hat{E} t U_2) \right\}$$

$$+ \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_2} \frac{1}{(T - h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^\top \left( E_{s-h} \otimes U_1^\top \right) A_1 F_s A_2^\top U_2 \right. \right.$$  

$$\cdot \left. \text{mat}_1 (E_{t-h} \otimes U_1^\top \hat{E} t U_2) \right\}$$
Again, we bound each term in turn. Using the decomposition (42), adopting same arguments in the proof of \( \Pi \), we have,

\[
\Pi = O\left(\frac{d^2}{T d_1^3} + \frac{d_1^{1/2}d_1^{-\delta_0/2}n_2^2}{T^2}\right).
\]

Next, we consider \( \Pi \).

\[
\Pi = \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{\text{mat}_1^{\top}(A_1 F_{s-h} A_2^\top \otimes E_s) \hat{U}_{1j} \hat{U}_{1j}^\top \right\} \cdot \text{mat}_1(E_{t-h} \otimes E_t) \right\}
\]

\[
+ \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{\text{mat}_1^{\top}(A_1 F_{s-h} A_2^\top \otimes A_1 F_{t} A_2^\top) \hat{U}_{1j} \hat{U}_{1j}^\top \right\} \cdot \text{mat}_1(E_{t-h} \otimes A_1 F_{t} A_2^\top) \right\}
\]

\[
+ \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{\text{mat}_1^{\top}(A_1 F_{s-h} A_2^\top \otimes A_1 F_{t} A_2^\top) \hat{U}_{1j} \hat{U}_{1j}^\top \right\} \cdot \text{mat}_1(E_{t-h} \otimes A_1 F_{t} A_2^\top) \right\}
\]

\[
+ \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{\text{mat}_1^{\top}(A_1 F_{s-h} A_2^\top \otimes A_1 F_{t} A_2^\top) \hat{U}_{1j} \hat{U}_{1j}^\top \right\} \cdot \text{mat}_1(E_{t-h} \otimes E_t) \right\}
\]

\[
:= \Pi_1 + \Pi_2 + \Pi_3 + \Pi_4.
\]

Again, we bound each term in turn. Using the decomposition (42), adopting same arguments in the proof of \( \Pi \), we have,

\[
\Pi_1 = O\left(\frac{d^{3/2-\delta_0/2}}{T}\right),
\]

\( \left(\begin{array}{c}
\Pi_1 \\
\Pi_2 \\
\Pi_3 \\
\Pi_4
\end{array}\right)
\]


\[ III_2 = O\left(\frac{d^{3/2-3\delta_0/2}}{T^{1/2}} + \frac{d^{1/2}d^{3/2-3\delta_0/2}T}{T}\right), \]

\[ III_3 = O\left(\frac{d^{1-\delta_0}}{T} + \frac{d^{1/2}d^{1-\delta_0}\eta_d}{T^{3/2}}\right), \]

\[ III_4 = O\left(\frac{d^{1-\delta_0}}{T^{1/2}} + \frac{d^{1/2}d^{1-\delta_0}\eta_d}{T}\right). \]

This implies that

\[ III = O\left(\frac{d^{3/2-\delta_0/2}}{T^{1/2}} + \frac{d^{1/2}d^{3/2-\delta_0/2}T}{T}\right). \] (51)

Employing (48), (49), (50) and (51), in the event \( \Omega_1 \), there exists a constant \( c_m > 0 \),

\[ M(m, \hat{U}_{1,m}) - M(r_1, \hat{U}_1) \]

\[ \geq (c_m + o(1))d^{-2\delta} + C\left(\frac{1}{Td^{\delta_0}} + \frac{1}{T^{1/2}d^{1/2+3\delta_0/2}} + \frac{1}{T^{1/2+d^{1/2+3\delta_0/2}}}\right), \] (52)

\[ \square \]

**Proof of Theorem 1 for non-iterative TOPUP.** Let \( \Omega_1 = \Omega_{11} \cap \Omega_{12} \cap \Omega_0 \), where \( \Omega_{11}, \Omega_{12} \) are defined in Lemma 2 and 3, respectively, and \( \Omega_0 \) is the event in Proposition 2. Then \( \Omega_1 \) is the same event in Lemma 4 and 5, with \( P(\Omega_1) \geq 1 - e^{-d_2} - T \exp(-C_1T^\delta) - \exp(-C_2T) \) and \( C_1, C_2 > 0, 1/\theta = 1/\theta_1 + 2/\theta_2 \). In the following, we shall only work on \( \Omega_1 \).

(i) (IC estimator using non-iterative TOPUP) We shall prove in the event \( \Omega_1, M(m, \hat{U}_{1,m}) + m \cdot d^{-2}g_1(d, T) \geq M(r_1, \hat{U}_1) + r_1 \cdot d^{-2}g_1(d, T) \) for all \( m \neq r_1 \) and \( m \leq m_1 \). For \( m < r_1 \), it is sufficient to show \( M(m, \hat{U}_{1,m}) - M(r_1, \hat{U}_1) \geq (r_1 - m) d^{-2}g_1(d, T) > 0 \). This is followed by Lemma 5 and the second part of Assumption VII(a). For \( m > r_1 \), as \( M(m, \hat{U}_{1,m}) \leq M(r_1, \hat{U}_1) \), it is sufficient to prove \( M(r_1, \hat{U}_1) - M(m, \hat{U}_{1,m}) \leq (m-r_1)d^{-2}g_1(d, T) \). By Lemma 4, \( M(r_1, \hat{U}_1) - M(m, \hat{U}_{1,m}) \leq C(\beta_{d,T}) \). Then the first case is followed by the second part of Assumption VII(a).

(ii) (ER estimator using non-iterative TOPUP) By Lemma 4, for all \( j > r_1, \) \( \hat{\lambda}_j \leq C_{11}(d^2\beta_{d,T}) \). Lemma 5 implies that \( \hat{\lambda}_j \geq (c_m + o(1))d^{2-2\delta} + C_{12}(d^{2}\gamma_{d,T}) \) for all \( 1 \leq j \leq r_1 \). Employing similar arguments in the proof of Lemma 5, we can obtain \( \hat{\lambda}_1 \leq (c_m^* + o(1))d^{2-2\delta} + C_{13}(d^{2}\gamma_{d,T}) \), for some constant \( c_m^* > 0 \). Provided that \( h_1(d, T) = o(d^{2+2\delta_0-4\delta}) = o(d^{2-2\delta_0}) \) and \( \gamma_{d,T} = o(d^{-2\delta_0}) \),

\[ \frac{\hat{\lambda}_{j+1} + h_1(d, T)}{\hat{\lambda}_j + h_1(d, T)} \geq \frac{h_1(d, T)}{h_1(d, T) + C_{14}(d^2\beta_{d,T})}, \] for \( j > r_1 \),
Theorem 1 for iTOPUP

Lemma 6. Suppose Assumptions I, II, III, IV and V(a) hold. Assume that \( \|U_2 U_2^\top - U_2 U_2^\top_o \|_2 \leq c_0 \sqrt{d_1/d_2} \) for some positive constant \( c_0 \). Then, in an event \( \Omega_{22} \cap \Omega_0 \) with \( \mathbb{P}(\Omega_{22}) \geq 1 - e^{-d_2/2} \) and \( \mathbb{P}(\Omega_0) \geq 1 - T \exp(-C_1 T^4) - \exp(-C_2 T) \) with \( C_1, C_2 > 0, 1/\vartheta = 1/\theta_1 + 2/\theta_2 \), for all fixed \( m \) such that \( m \geq r_2 \), we have

\[
\begin{align*}
\| \frac{1}{T-h} \sum_{t=h+1}^T \text{mat}_3(A_{1}F_{t-h}A_2^\top \hat{U}_{2,m} \otimes E_t \hat{U}_{2,m}) \|_2 & \leq C \left( d_1^{1/2}d_1^{1/2-\delta_0/2} \right), \\
\| \frac{1}{T-h} \sum_{t=h+1}^T \text{mat}_3(A_{1}F_{t-h}A_2^\top \hat{U}_{2,m} \otimes U_1^\top E_t \hat{U}_{2,m}) \|_2 & \leq C \left( d_1^{1/2}d_1^{1/2-\delta_0/2} \right), \\
\| \frac{1}{T-h} \sum_{t=h+1}^T \text{mat}_3(E_{t-h} \hat{U}_{2,m} \otimes A_1 F_{t} A_2^\top \hat{U}_{2,m}) \|_2 & \leq C \left( d_1^{1/2}d_1^{1/2-\delta_0/2} \right), \\
\| \frac{1}{T-h} \sum_{t=h+1}^T U_1^\top \text{mat}_3(E_{t-h} \hat{U}_{2,m} \otimes A_1 F_{t} A_2^\top \hat{U}_{2,m}) \|_2 & \leq C \left( d_1^{1/2}d_1^{1/2-\delta_0/2} \right), \\
\| \frac{1}{T-h} \sum_{t=h+1}^T \text{mat}_3(E_{t-h} \hat{U}_{2,m} \otimes E_t \hat{U}_{2,m}) \|_2 & \leq C \left( d_1^{1/2} \right), \\
\| \frac{1}{T-h} \sum_{t=h+1}^T \text{mat}_3(\hat{U}_{t-1} \hat{U}_{2,m} \otimes U_1^\top E_t \hat{U}_{2,m}) \|_2 & \leq C \left( d_1^{1/2} \right), \\
\| \frac{1}{T-h} \sum_{t=h+1}^T U_1^\top \text{mat}_3(E_{t-h} \hat{U}_{2,m} \otimes U_1^\top E_t \hat{U}_{2,m}) \|_2 & \leq C \left( \frac{1}{\sqrt{T}} \right),
\end{align*}
\]

for some constant positive \( C \) depending on \( K \) only.

Proof. Recall \( \hat{U}_2 = \hat{U}_{2,r_2} = [\hat{U}_{21}, \ldots, \hat{U}_{2r_2}] \). Write \( U_{2,m} = [U_2, U_2^m] \) for \( m > r_2 \). Let \( U_{21} \) be complement part of \( U_2 \), namely \( [U_2, U_{21}] \in O_{d_2} \). If \( m \leq r_2 \), then
\[ \| \hat{U}_{2,m}^T U_{2,1}^2 \|_2 \leq \| \hat{U}_2^T U_{2,1} \|_2 = \| \hat{U}_2^T - U_2^T \|_2. \]

If \( m > r_2 \), then \( \hat{U}_{2,m}^T U_{2,m} = \hat{U}_2^T \hat{U}_2 \) \( \pm \sum_{j=r+1}^m \hat{U}_2 \hat{U}_2^T = \hat{U}_2^T + \hat{U}_2^T \). Furthermore, for any matrix \( B \in \mathbb{R}^{a \times d} \),
\[ \| B \hat{U}_{2,m} \|_2 = \| B \hat{U}_2, B \hat{U}_2^2 \|_2 \leq \sqrt{2} \left( \| B \hat{U}_2 \|_2 + \| B \hat{U}_2^2 \|_2 \right). \]

Note that
\[ \| U_2^T \hat{U}_2^T \hat{U}_2^2 \|_2 = \| U_2^T \hat{U}_2^2 \|_2 \leq \| U_2^T \hat{U}_2^T - U_2 U_2^T \|_2. \]

Adopting similar procedures in the proof of Theorem 1 in Han et al. (2020), we can show Lemma 6.

**Lemma 7.** Suppose Assumptions I, II, III, IV and V(a) hold. Let \( r_1 \leq r_{i1} \leq m_1 < d_i \) for all \( 1 \leq j \leq i-1 \), and \( m_1 = O(r_1) \). There exists an event \( \Omega_2 \) such that \( \mathbb{P}(\Omega_2) \geq 1 - e^{-d_2} - T \exp(-C_1 T^\delta) - \exp(-C_2 T) \), \( C_1, C_2 > 0, 1/\delta = 1/\theta_1 + 2/\theta_2 \) and \( \Omega_2 \) is independent of iteration number \( i \). Then, at \( i \)-th iteration, in the event \( \Omega_2 \), for any fixed \( m > r_1 \), \( M^{(i)}(r_1, \hat{U}_1) - M^{(i)}(m, \hat{U}_{1,m}) \leq C \beta_{d,T}, \)
where \( C > 0 \),
\[ \beta_{d,T} = \frac{d_1}{T d^{1+\delta_0/2}} + \frac{d_1^{1/2} \eta_{i}^{(i)}}{T d^{1/2+3\delta_0/2}}, \]
and \( \eta_{i}^{(i)} = d_1^{1/2} d^{\delta_1-\delta_0/2-1/2} + d_k^{1/2} d^{\delta_1-1}. \)

**Proof.** Note that by Han et al. (2020), in the event \( \Omega_2 \), at \( i \)-th iteration \( (i \geq 1) \), for the iTOPUP estimator, assume that \( \| \hat{U}_k^T \hat{U}_k^T - U_k U_k^T \|_2 \leq C_0 \| \hat{U}_k^T \|^{1/2} \), \( 1 \leq k \leq K \), for some \( C_0 > 0 \). The proof is omitted. Thus, it is similar to Lemma 4. \( \square \)

**Lemma 8.** Suppose Assumptions I, II, III, IV and V(a) hold. Let \( r_1 \leq r_{i1} \leq m_1 < d_i \) for all \( 1 \leq j \leq i-1 \), and \( m_1 = O(r_1) \). There exists an event \( \Omega_2 \) such that \( \mathbb{P}(\Omega_2) \geq 1 - e^{-d_2} - T \exp(-C_1 T^\delta) - \exp(-C_2 T) \), \( C_1, C_2 > 0, 1/\delta = 1/\theta_1 + 2/\theta_2 \) and \( \Omega_2 \) is independent of iteration number \( i \). Then, at \( i \)-th iteration, in the event \( \Omega_2 \), for any \( m > r_1 \), there exist constants \( c_m > 0 \) and \( C > 0 \), such that \( M^{(i)}(m, \hat{U}_{1,m}) - M^{(i)}(r_1, \hat{U}_1) \geq (c_m + o(1))d^{-2\delta_1} + C \gamma_{d,T}, \)
where
\[ \gamma_{d,T} = \frac{d_1}{T d^{1+\delta_0}} + \frac{1}{T^{1/2} d^{1+3\delta_0/2}} + \frac{d_1^{1/2} \eta_{i}^{(i)}}{T d^{1/2+3\delta_0/2}}, \]
and \( \eta_{i}^{(i)} = d_k^{1/2} d^{\delta_1-\delta_0/2-1/2} + d_k^{1/2} d^{\delta_1-1}. \)

**Proof.** Note that by Han et al. (2020), in the event \( \Omega_2 \), at \( i \)-th iteration \( (i \geq 1) \), for the iTOPUP estimator, assume that \( \| \hat{U}_k^T \hat{U}_k^T - U_k U_k^T \|_2 \leq C_0 \| \hat{U}_k^T \|^{1/2} \), \( 1 \leq k \leq K \), for some \( C_0 > 0 \). The proof is omitted, as it is similar to Lemma 5. \( \square \)

**Proof of Theorem 1 for iTOPUP.** Employing similar arguments in the proofs of Theorem 1 for non-iterative TOPUP we can show Theorem 1 for iTOPUP. \( \square \)
A.3. Proof of Theorem 1 for non-iterative TIPUP

Lemma 9. Let $U_{1\perp} \in \mathbb{R}^{d_1 \times (d_1 - r_1)}$ be the orthonormal complement of $U_1$. Denote $U_{1\perp} = (U_{1j})_{1 \leq j \leq d_1}$ and $U_1 = (U_{1j})_{1 \leq j \leq r_1}$, where $U_{1j} \in \mathbb{R}^{d_1}$. Denote $\eta^*_d = d^{1/2}d_1^{-1/2} - \delta_0/2 - 1/2 + d_1^{-1/2}$. Define $\hat{U}_1$ be the estimated top $r_1$ left singular space of TIPUP$_1(X_1; T)$. Suppose Assumptions I, II, III, IV and $V(b)$ hold. Then, in an event $\Omega_{31} \cap \Omega_0$ with $\mathbb{P}(\Omega_{31}) \geq 1 - e^{-d_2/2}$ and $\mathbb{P}(\Omega_0) \geq 1 - T \exp(-C_1T^d) - \exp(-C_2T)$ with $C_1, C_2 > 0$, $1/\vartheta = 1/\theta_1 + 2/\theta_2$, there exist a matrix $\hat{U}_1 \in \mathbb{R}^{d_1 \times r_1}$ with $\hat{U}_1 \hat{U}_1^T = U_1U_1^T$, such that

$$\|\hat{U}_1 - U_1\hat{U}_1^T\|_2 \leq C \left( \frac{\eta^*_d}{\sqrt{T}} \right),$$

(60)

$$\|U_1U_1^T\hat{U}_1\|_2 \leq C \left( \frac{\eta^*_d}{\sqrt{T}} \right),$$

(61)

for all $r_1 + 1 \leq j \leq d_1$ and $C > 0$.

The proof is similar to Lemma 2, thus is omitted.

Lemma 10. Suppose Assumptions I, II, III, IV and $V(b)$ hold. Then, in an event $\Omega_{32} \cap \Omega_0$ with $\mathbb{P}(\Omega_{32}) \geq 1 - e^{-d_1/2}$ and $\mathbb{P}(\Omega_0) \geq 1 - T \exp(-C_1T^d) - \exp(-C_2T)$ with $C_1, C_2 > 0$, $1/\vartheta = 1/\theta_1 + 2/\theta_2$, we have

$$\left\| \frac{1}{T - h} \sum_{t = h + 1}^{T} A_1 F_{t-h} A_2^T E_t^T \right\|_2 \leq C \left( \frac{d^{1/2}d_1^{1/2} - \delta_0/2}{\sqrt{T}} \right),$$

(62)

$$\left\| \frac{1}{T - h} \sum_{t = h + 1}^{T} E_{t-h} A_2^T A_1^T \right\|_2 \leq C \left( \frac{d^{1/2}d_1^{1/2} - \delta_0/2}{\sqrt{T}} \right),$$

(63)

$$\left\| \frac{1}{T - h} \sum_{t = h + 1}^{T} A_1 F_{t-h} A_2^T E_t^T U_1 \right\|_2 \leq C \left( \frac{d^{1/2} - \delta_0/2}{\sqrt{T}} \right),$$

(64)

$$\left\| \frac{1}{T - h} \sum_{t = h + 1}^{T} U_1^T E_{t-h} A_2^T F_i^T A_1^T \right\|_2 \leq C \left( \frac{d^{1/2} - \delta_0/2}{\sqrt{T}} \right),$$

(65)

$$\left\| \frac{1}{T - h} \sum_{t = h + 1}^{T} E_{t-h} E_t^T \right\|_2 \leq C \left( \frac{d^{1/2}}{\sqrt{T}} \right),$$

(66)

$$\left\| \frac{1}{T - h} \sum_{t = h + 1}^{T} U_1^T E_{t-h} E_t^T U_1 \right\|_2 \leq C \left( \frac{d^{1/2}}{\sqrt{T}} \right),$$

(67)

for some constant positive $C$ depending on $K$ only.

Proof. By Theorem 2 in Chen, Yang and Zhang (2021), under Assumption I,

$$\mathbb{E} \left\| \frac{1}{T - h} \sum_{t = h + 1}^{T} A_1 F_{t-h} A_2^T E_t^T \right\|_2 \leq \frac{\sigma \sqrt{8T d_1}}{T - h} \|\Theta^*_1, 0\|_2^{1/2}. $$
Elementary calculation shows that
\[
\left\| \sum_{t=h+1}^{T} A_1 F_{t-h}^T A_2^T E_t^T \right\|_2 - \left\| \sum_{t=h+1}^{T} A_1 F_{t-h}^T A_2^T E_t^T \right\|_2 \\
\leq \left\| \left( A_1 F_1^T, \ldots, A_1 F_{T-h}^T \right) \left( E_{h+1}^T - E_{h+1}^T \right) \ldots \left( E_T^T - E_T^T \right) \right\|_2 \\
\leq \left\| \left( A_1 F_1^T, \ldots, A_1 F_{T-h}^T \right) \right\|_{1/2} \left\| \left( E_{h+1}^T - E_{h+1}^T \right) \ldots \left( E_T^T - E_T^T \right) \right\|_F \\
\leq \sqrt{T} \left\| \Theta_{1,0}^* \right\|_{1/2} \left\| \left( E_{h+1}^T - E_{h+1}^T \right) \ldots \left( E_T^T - E_T^T \right) \right\|_F.
\]

That is, \( \left\| \sum_{t=h+1}^{T} A_1 F_{t-h}^T A_2^T E_t^T \right\|_2 \) is a \( \sigma \sqrt{T} \left\| \Theta_{1,0}^* \right\|_{1/2} \) Lipschitz function in \( (E_1, \ldots, E_T) \). Then, by Gaussian concentration inequalities for Lipschitz functions,
\[
P \left( \left\| \sum_{t=h+1}^{T} \frac{A_1 F_{t-h}^T A_2^T E_t^T}{T-h} \right\|_2 - \frac{\sigma (8T d_1)^{1/2}}{T-h} \left\| \Theta_{1,0}^* \right\|_{1/2} \geq \frac{\sigma \sqrt{T}}{T-h} \left\| \Theta_{1,0}^* \right\|_{1/2} x \right) \\
\leq 2 e^{-\frac{x^2}{2}}.
\]

As \( T \geq 4h_0 \) and \( K = 2 \), this implies with \( x \propto \sqrt{d_1} \) that in an event \( \Omega_a \) with at least probability 1 - \( e^{-d_1/6} \),
\[
\left\| \frac{1}{T-h} \sum_{t=h+1}^{T} A_1 F_{t-h}^T A_2^T E_t^T \right\|_2 \leq \frac{C_0 \sigma \sqrt{T d_1}}{T-h} \left\| \Theta_{1,0}^* \right\|_{1/2}, \quad (68)
\]

with a constant \( C_0 \) depending on \( K \) only. Then, by Proposition 2, in the event \( \Omega_a \cap \Omega_0, \) (62) follows. Similar arguments yield (63), (64) and (65) in the event \( \Omega_a \cap \Omega_0 \).

We split the sum into two terms over the index sets, \( S_1 = \{(h,2h] \cup (3h,4h] \cup \cdots \} \cap (h,T] \) and its complement \( S_2 \) in \( (h,T] \), so that \( \{E_{t-h}, t \in S_a\} \) is independent of \( \{E_t, t \in S_a\} \) for each \( a = 1, 2 \). Let \( n_a = |S_a| d_2 \). Define \( G_a = (E_{t-h}, t \in S_a) \in \mathbb{R}^{d_1 \times n_a} \) and \( H_a = (E_t, t \in S_a) \in \mathbb{R}^{d_1 \times n_a} \). Then, \( G_a, H_a \) are two independent Gaussian matrices. Note that
\[
\left\| \sum_{t \in S_a} E_{t-h} E_t^T \right\|_2 \left\| \sum_{t \in S_a} H_{t} E_t^T \right\|_2 = \left\| G_a H_a^T \right\|_2.
\]
Moreover, by Assumption I, \( \text{Var}(u^T \text{vec}(G_a)) \leq \sigma^2 \) and \( \text{Var}(u^T \text{vec}(H_a)) \leq \sigma^2 \) for all unit vectors \( u \in \mathbb{R}^d \), so that by Lemme 19(i)\\n\\n\[
\mathbb{P} \left\{ \|G_a H_a^T\|_2^2 / \sigma^2 \geq d_1 + 2\sqrt{d_1} n_a + x(x + 2\sqrt{m_a} + 2\sqrt{d_1}) \right\} \leq 2e^{-x^2/2}, \quad x > 0.
\]

Thus, as \( \sum_{a=1}^2 n_a = d_2(T - h) \), it follows from the above inequality that with \( h_0 \leq T/4 \), \( x \geq \sqrt{d_1} \) and some constant \( C_0 \) depending on \( K \) only,

\[
\left\| \sum_{t=h+1}^T E_{t-h} E_{t}^T \right\|_2^2 \leq C_0^2 \left( d_1 + \sqrt{d_1 d_2 T} \right) / T - h,
\]

in an event \( \Omega_c \) with at least probability \( 1 - e^{-d_1}/6 \). Then, as in the derivation of \( \|\Delta_1\|_2 \) in the proof of Theorem 2 in Chen, Yang and Zhang (2021), in the event \( \Omega_c, (66) \) follows. Similar arguments yield (67) in the event \( \Omega_c \). Set \( \Omega_{32} = \Omega_a \cap \Omega_c \), then \( \mathbb{P}(\Omega_{32}) \geq 1 - e^{-d_1}/2 \). \( \square \)

**Lemma 11.** Suppose Assumptions I, II, III, IV and \( V(b) \) hold. In an event \( \Omega_3 \) with \( \mathbb{P}(\Omega_3) \geq 1 - e^{-d_1} - e^{-d_2} - T \exp(-C_1 T^0) - \exp(-C_2 T) \) with \( C_1, C_2 > 0, 0, 1/\theta = 1/\theta_1 + 2/\theta_2, \) for any fixed \( m \) with \( m > r_1 \), \( \mathcal{M}^*(r_1, U_1) - \mathcal{M}^*(m, U_{1,m}) \leq C \beta^*_d, \) where \( C > 0, \)

\[
\beta^*_d = \frac{1}{T d} + \frac{d_1}{T d^{1+\delta_0}} + \frac{d_1^{1/2} \eta^*_d}{T d^{1/2+\delta_0/2}} + \frac{\eta^*_d}{T d^{1/2+\delta_0}},
\]

and \( \eta^*_d = d^{1/2} \delta_1^{-1/2} - 2/1/2 + d^{1/2} \delta_1^{-1/2}. \)

**Proof.** Let \( \Omega_3 = \Omega_{31} \cap \Omega_{32} \cap \Omega_0 \), where \( \Omega_{31}, \Omega_{32} \) are defined in Lemma 9 and 10, respectively, and \( \Omega_0 \) is the event in Proposition 2. Let \( U_{1,r_1} = U_1 \) and \( U_{1,m} = (U_{11}, U_{12}, ..., U_{1m}) \).

\[
\left| \mathcal{M}^*(m, \hat{U}_{1,m}) - \mathcal{M}^*(r_1, \hat{U}_1) \right| \leq \left| \mathcal{M}^*(m, \hat{U}_{1,m}) - \mathcal{M}^*(r_1, U_1) \right| + \left| \mathcal{M}^*(r_1, \hat{U}_1) - \mathcal{M}^*(r_1, U_1) \right| \leq 2 \max_{r_1 < m \leq m_1} \left| \mathcal{M}^*(m, \hat{U}_{1,m}) - \mathcal{M}^*(r_1, U_1) \right|.
\]

As \( m_1 \) is fixed, it is sufficient to prove for each \( m \) with \( m > r_1 \),

\[
\left| \mathcal{M}^*(m, \hat{U}_{1,m}) - \mathcal{M}^*(r_1, U_1) \right| \leq C \beta^*_d.
\]

(70)

In the following, we shall only work on \( \Omega_3 \).

Elementary calculation shows that

\[
\begin{align*}
& d^2 \cdot \mathcal{M}^*(m, \hat{U}_{1,m}) - d^2 \cdot \mathcal{M}^*(r_1, U_1) \\
= & \sum_{h=1}^{h_0} \frac{1}{(T - h)^2} \sum_{s,t=h+1}^T \text{tr} \left\{ \left( U_{11} U_{11}^T - \hat{U}_{1,m} \hat{U}_{1,m}^T \right) X_{t-h} X_t^T X_s X_{s-h} \right\}
\end{align*}
\]
Determining the number of factors

\[ = \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr} \left\{ X_s A_2 F_{s-h}^T A_1^T \left( U_1 U_1^T - \hat{U}_{1,m} \hat{U}_{1,m}^T \right) A_1 F_{t-h} A_2^T X_t^T \right\} \]

\[ + \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr} \left\{ X_s E_{s-h}^T \left( U_1 U_1^T - \hat{U}_{1,m} \hat{U}_{1,m}^T \right) E_{t-h} X_t^T \right\} \]

\[ + \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr} \left\{ X_s A_2 F_{s-h}^T A_1^T \left( U_1 U_1^T - \hat{U}_{1,m} \hat{U}_{1,m}^T \right) E_{t-h} X_t^T \right\} \]

\[ + \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr} \left\{ X_s E_{s-h}^T \left( U_1 U_1^T - \hat{U}_{1,m} \hat{U}_{1,m}^T \right) A_1 F_{t-h} A_2^T X_t^T \right\} \]

\[ := I + II + III + IV. \]

Note that III = IV.

We first consider II.

\[ II = \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr} \left\{ X_s E_{s-h}^T \left( U_1 U_1^T - \hat{U}_{1,m} \hat{U}_{1,m}^T \right) E_{t-h} X_t^T \right\} \]

\[ = \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr} \left\{ E_s E_{s-h}^T \left( U_1 U_1^T - \hat{U}_{1,m} \hat{U}_{1,m}^T \right) E_{t-h} X_t^T \right\} \]

\[ + \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr} \left\{ A_1 F_s A_2^T E_{s-h} \left( U_1 U_1^T - \hat{U}_{1,m} \hat{U}_{1,m}^T \right) E_{t-h} X_t^T \right\} \]

\[ + 2 \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr} \left\{ A_1 F_s A_2^T E_{s-h} \left( U_1 U_1^T - \hat{U}_{1,m} \hat{U}_{1,m}^T \right) E_{t-h} X_t^T \right\} \]

\[ := II_1 + II_2 + 2II_3. \]

By (42), Lemma 9 and 10,

\[ II_1 \leq r_1 \sum_{h=1}^{h_0} \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} E_{t-h} E_t^T \right\|_2^2 \]

\[ \cdot \left( \left\| \hat{U}_1 - U_1 U_1^T \hat{U}_1 \right\|_2^2 + \left\| \hat{U}_1 - U_1 U_1^T \hat{U}_1 \right\|_2^2 \left\| U_1 U_1^T \hat{U}_1 \right\|_2^2 \right) \]

\[ + (m - r_1) \sum_{h=1}^{h_0} \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} E_{t-h} E_t^T \right\|_2^2 = O \left( \frac{d}{T} \right), \]
using the fact \( \eta_1^* = o(\sqrt{T}) \).

\[
\Pi_2 \leq r_1 \sum_{h=1}^{h_0} \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} E_{t-h} A_2 F_t^T A_1^T \right\| \left\| \hat{U}_1 - U_1 \hat{U}_1^T \hat{U}_1 \right\|_2^2 + 2r_1 \sum_{h=1}^{h_0} \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} E_{t-h} A_2 F_t^T A_1^T \right\| \left\| \frac{1}{T-h} \sum_{s=h+1}^{T} U_1^T E_{s-h} A_2 F_s^T A_1^T \right\|_2^2 \\
\cdot \left\| \hat{U}_1 - U_1 \hat{U}_1^T \hat{U}_1 \right\|_2^2 + (m - r_1) \sum_{h=1}^{h_0} \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} E_{t-h} A_2 F_t^T A_1^T \right\| \left\| \frac{1}{T-h} \sum_{s=h+1}^{T} U_1^T E_{s-h} A_2 F_s^T A_1^T \right\|_2^2 \\
= O \left( \frac{d_1 d_1^{1 - \delta_0}}{T} \right).
\]

\[
\Pi_3 \leq r_1 \sum_{h=1}^{h_0} \left\| \frac{1}{T-h} \sum_{s=h+1}^{T} E_{s-h} A_2 F_s^T A_1^T \right\| \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} E_{t-h} E_t^T \right\|_2^2 \\
\cdot \left\| \hat{U}_1 - U_1 \hat{U}_1^T \hat{U}_1 \right\|_2^2 + r_1 \sum_{h=1}^{h_0} \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} E_{t-h} E_t^T \right\| \left\| \frac{1}{T-h} \sum_{s=h+1}^{T} U_1^T E_{s-h} A_2 F_s^T A_1^T \right\|_2^2 \\
\cdot \left\| \hat{U}_1 - U_1 \hat{U}_1^T \hat{U}_1 \right\|_2^2 + r_1 \sum_{h=1}^{h_0} \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} E_{t-h} E_t^T \right\| \left\| \frac{1}{T-h} \sum_{s=h+1}^{T} U_1^T E_{s-h} A_2 F_s^T A_1^T \right\|_2^2 \\
\cdot \left\| \hat{U}_1 - U_1 \hat{U}_1^T \hat{U}_1 \right\|_2^2 + (m - r_1) \sum_{h=1}^{h_0} \left\| \frac{1}{T-h} \sum_{s=h+1}^{T} E_{s-h} A_2 F_s^T A_1^T \right\| \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} E_{t-h} E_t^T \right\|_2^2 \\
= O \left( \frac{d_1^{1/2} d_1^{1 - \delta_0/2}}{T} \right).
\]

Combining the bounds of \( \Pi_1, \Pi_2 \) and \( \Pi_3 \), we have

\[
\Pi = O \left( \frac{d_1 + d_1^{d_1^{1 - \delta_0}}}{T} \right). \tag{71}
\]

Next, we consider \( \Pi_2 \).

\[
\Pi_2 = \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,i=h+1}^{T} \text{tr} \left\{ X_s A_2 F_s^T A_1^T \left( U_1 U_1^T - \hat{U}_1 \hat{U}_1^T \right) E_{i-h} X_i^T \right\}
\]
Moreover, \[ III \leq 1 \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr} \left\{ A_1 F_s A_2^T A_2 F_{s-h} A_1^T \left( U_1 U_1^T - \hat{U}_{1,m} \hat{U}_{1,m}^T \right) \right\} \cdot E_{t-h} A_2 F_t^T A_1^T \]

\[ + \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr} \left\{ E_s A_2 F_{s-h} A_1^T \left( U_1 U_1^T - \hat{U}_{1,m} \hat{U}_{1,m}^T \right) E_{t-h} E_t^T \right\} \]

\[ + \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr} \left\{ E_s A_2 F_{s-h} A_1^T \left( U_1 U_1^T - \hat{U}_{1,m} \hat{U}_{1,m}^T \right) \cdot E_{t-h} A_2 F_t^T A_1^T \right\} \]

\[ + \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr} \left\{ A_1 F_s A_2^T A_2 F_{s-h} A_1^T \left( U_1 U_1^T - \hat{U}_{1,m} \hat{U}_{1,m}^T \right) \cdot E_{t-h} E_t^T \right\} \]

\[ := III_1 + III_2 + III_3 + III_4. \]

Note that \( A_1^T U_1 U_1^T = A_1^T \). Similarly, by (42), Lemma 9 and 10,

\[ III_1 \leq 3r_1 \sum_{h=1}^{h_0} \frac{1}{T-h} \sum_{s=t=h+1}^{T} A_1 F_s A_2^T A_2 F_{s-h} A_1^T \] \[ \cdot \left\| \hat{U}_1 - U_1 \hat{U}_1^T \hat{U}_1 \right\|_2 \right\] \[ + (m-r_1) \sum_{h=1}^{h_0} \frac{1}{T-h} \sum_{s=t=h+1}^{T} A_1 F_s A_2^T A_2 F_{s-h} A_1^T \] \[ \cdot \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} E_{t-h} A_2 F_t^T A_1^T \right\|_2 \left\| \hat{U}_1 U_1^T \hat{U}_{1j} \right\|_2 \]

\[ = O \left( \frac{d_1^{3/2} q^{3/2} - 3 \beta \delta^{2} n_4}{T} \right). \]

Moreover,

\[ III_2 \leq 3r_1 \sum_{h=1}^{h_0} \frac{1}{T-h} \sum_{s=t=h+1}^{T} E_s A_2 F_{s-h} A_1^T \] \[ \cdot \left\| \hat{U}_1 - U_1 \hat{U}_1^T \hat{U}_1 \right\|_2 \right\] \[ + r_1 \sum_{h=1}^{h_0} \left\| \frac{1}{T-h} \sum_{s=t=h+1}^{T} E_s A_2 F_{s-h} A_1^T \right\|_2 \]
\begin{align*}
\| \frac{1}{T-h} \sum_{t=h+1}^{T} E_{t-h} E_t^T \|_2 \| U_1 U_1^T \|_2 & = O \left( \frac{d_{1/2} d_{1-\delta_0/2} \eta_d^*}{T^{3/2}} \right). \\
\| \frac{1}{T-h} \sum_{t=h+1}^{T} E_{t-h} A_2 F_t^T A_1^T U_1 \|_2 \| \dot{U}_1 - U_1 \dot{U}_1^T \|_2 & = O \left( \frac{d_{1/2} d_{1-\delta_0} \eta_d^*}{T^{3/2}} \right). \\
\| \frac{1}{T-h} \sum_{t=h+1}^{T} E_{t-h} A_2 F_t^T A_1^T U_1 \|_2 \| U_1 U_1^T \dot{U}_1 \|_2 & = O \left( \frac{d_{1/2} d_{1-\delta_0} \eta_d^*}{T^{3/2}} \right). \\
\| \frac{1}{T-h} \sum_{t=h+1}^{T} E_{t-h} A_2 F_t^T A_1^T U_1 \|_2 \| U_1 U_1^T \dot{U}_1 \|_2 & = O \left( \frac{d_{1/2} d_{1-\delta_0} \eta_d^*}{T^{3/2}} \right). \\
\| \frac{1}{T-h} \sum_{t=h+1}^{T} E_{t-h} A_2 F_t^T A_1^T U_1 \|_2 \| U_1 U_1^T \dot{U}_1 \|_2 & = O \left( \frac{d_{1/2} d_{1-\delta_0} \eta_d^*}{T^{3/2}} \right). \\
\| \frac{1}{T-h} \sum_{t=h+1}^{T} E_{t-h} A_2 F_t^T A_1^T U_1 \|_2 \| U_1 U_1^T \dot{U}_1 \|_2 & = O \left( \frac{d_{1/2} d_{1-\delta_0} \eta_d^*}{T^{3/2}} \right).
\end{align*}

Combining III, III, III and IV, we have

\begin{equation}
\text{III} = O \left( \frac{d_{1/2} d_{1-\delta_0} \eta_d^*}{T} + \frac{d_{1/2} d_{1-\delta_0} \eta_d^*}{T^{3/2}} + \frac{d_{1/2} d_{1-\delta_0} \eta_d^*}{T} + \frac{d_{1/2} d_{1-\delta_0} \eta_d^*}{T} \right). \tag{72}
\end{equation}

In view of (71) and (72),

\begin{equation}
\text{II + III + IV} = O \left( \frac{d}{T} + \frac{d_{1/2} d_{1-\delta_0} \eta_d^*}{T} + \frac{d_{1/2} d_{1-\delta_0} \eta_d^*}{T} + \frac{d_{1/2} d_{1-\delta_0} \eta_d^*}{T} \right). \tag{73}
\end{equation}
Next, we consider $I$.

$$I = \sum_{h=1}^{b_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr} \left\{ X_s A_2 F_{s-h}^T A_1^T \left( U_{1} U_{1}^T - \hat{U}_{1,m} \hat{U}_{1,m}^T \right) A_1 F_{t-h} A_2^T X_t^T \right\}$$

$$= \sum_{h=1}^{b_0} \frac{1}{(T-h)^2} \text{tr} \left\{ X_s A_2 F_{s-h}^T A_1^T \left( I - \hat{U}_{1,m} \hat{U}_{1,m}^T \right) \sum_{t=h+1}^{T} A_1 F_{t-h} A_2^T X_t^T \right\}$$

$$\geq 0,$$

using the fact that $I - \hat{U}_{1,m} \hat{U}_{1,m}^T$ is positive semi-definite, $x^T \hat{U}_{1,m} \hat{U}_{1,m}^T x \leq x^T x$.

Using the definition of our optimization target, we have $\mathcal{M}^*(m, \hat{U}_{1,m}) < \mathcal{M}^*(r_1, \hat{U}_1) \leq \mathcal{M}^*(r_1, U_1)$. It follows that $I + II + III + IV \leq 0$. As $I \geq 0$, by (73), $I = O(d^2 \beta^2_{d,T})$. In summary,

$$\left| \mathcal{M}^*(m, \hat{U}_{1,m}) - \mathcal{M}^*(r_1, U_1) \right| \leq C \beta^2_{d,T}. \quad \square$$

**Lemma 12.** Suppose Assumptions I, II, III, IV and $V(h)$ hold. In an event $\Omega_3$ with $\mathbb{P}(\Omega_3) \geq 1 - e^{-d_1} - e^{-d_2} - T \exp(-C_1 T^{\vartheta}) - \exp(-C_2 T)$ with $C_1, C_2 > 0, 1/\vartheta = 1/\theta_1 + 2/\theta_2$, for any $m$ with $m < r_1$, there exist constants $c_m > 0$ and $C > 0$, such that $\mathcal{M}^*(m, \hat{U}_{1,m}) - \mathcal{M}^*(r_1, \hat{U}_1) \geq (c_m + o(1))d^{-2k_1} + C \gamma^*_{d,T}$, where

$$\gamma_{d,T} = \frac{d_1}{T d_1^{1+\delta_0}} + \frac{1}{T d_1^{1/2}} + \frac{1}{T^{1/2} d_1^{1/2+3\delta_0}/2} + \frac{1}{T^{1/2} d_1^{1/2+\delta_0} d_1^{1/2}}$$

$$+ \frac{d_1^{1/2} \eta^*_d}{T d_1^{1/2+3\delta_0}/2} + \frac{\eta^*_d}{T d_1^{1/2+\delta_0}},$$

and $\eta^*_d = d_1^{1/2} d_1^{1-\delta_0/2-1/2} + d_1^{1-1/2}$.

**Proof.** Let $\Omega_3 = \Omega_3 \cap \Omega_3 \cap \Omega_0$, where $\Omega_3 \cap \Omega_3$ are defined in Lemma 9 and 10, respectively, and $\Omega_0$ is the event in Proposition 2. In the following, we shall only work on $\Omega_3$.

Let $U_{1,r_1} = U_1$ and $U_{1,m} = (U_{11}, U_{12}, ..., U_{1m})$.

$$d^2 \cdot \mathcal{M}^*(m, \hat{U}_{1,m}) - d^2 \cdot \mathcal{M}^*(r_1, \hat{U}_1)$$

$$= \sum_{h=1}^{b_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr} \left\{ \left( \hat{U}_{1} \hat{U}_{1}^T - \hat{U}_{1,m} \hat{U}_{1,m}^T \right) X_{t-h} X_t^T X_{s-h} X_s^T \right\}$$

$$= \sum_{h=1}^{b_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr} \left\{ X_s A_2 F_{s-h}^T A_1^T \hat{U}_{1j} \hat{U}_{1j}^T A_1 F_{t-h} A_2^T X_t^T \right\}$$

$$+ \sum_{h=1}^{b_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr} \left\{ X_s E_{s-h} \hat{U}_{1j} \hat{U}_{1j}^T E_{t-h} X_t^T \right\}$$

$$+ \sum_{h=1}^{b_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr} \left\{ X_s A_2 F_{s-h}^T A_1^T \hat{U}_{1j} \hat{U}_{1j}^T E_{t-h} X_t^T \right\}$$

$$+ \sum_{h=1}^{b_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr} \left\{ X_s A_2 F_{s-h}^T A_1^T \hat{U}_{1j} \hat{U}_{1j}^T E_{t-h} X_t^T \right\}$$
\begin{align*}
& + \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \operatorname{tr}\left\{ X_s E_{s-h}^T \hat{U}_{1j} \hat{U}_{1j}^T A_1 F_{t-h} A_2^T X_t^T \right\} \\
& := I + II + III + IV.
\end{align*}

Note that III = IV.

We first consider I.

\begin{align*}
I &= \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \operatorname{tr}\left\{ A_1 F_s A_2 F_{t-h} A_1^T \hat{U}_{1j} \hat{U}_{1j}^T \cdot A_1 F_{t-h} A_2 F_{t} A_1^T \right\} \\
& + \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \operatorname{tr}\left\{ E_s A_2 F_{t-h} A_1^T \hat{U}_{1j} \hat{U}_{1j}^T A_1 F_{t-h} A_2^T E_t \right\} \\
& + 2 \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \operatorname{tr}\left\{ E_s A_2 F_{t-h} A_1^T \hat{U}_{1j} \hat{U}_{1j}^T A_1 F_{t-h} A_2^T A_2 F_{t} A_1^T \right\} \\
& := I_1 + I_2 + I_3.
\end{align*}

For I_1,

\begin{align*}
I_1 &= \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \operatorname{tr}\left\{ \hat{U}_{1j}^T A_1 F_{t-h} A_2^T A_2 F_{t} A_1^T A_1 F_{s} A_2^T \cdot A_2 F_{s-h} A_1^T \hat{U}_{1j} \right\} \\
& = \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \operatorname{tr}\left\{ \hat{U}_{1j}^T U_{1j} U_{1j}^T \Theta_{s,h}^* \Theta_{t,h}^* U_{1j}^T \hat{U}_{1j} \right\} \\
& \geq \lambda_{\min} \left( \sum_{h=1}^{h_0} U_{1j}^T \Theta_{s,h}^* \Theta_{t,h}^* U_{1j} \right) \operatorname{tr}\left\{ \sum_{j=m+1}^{r_1} \hat{U}_{1j}^T U_{1j} U_{1j}^T \hat{U}_{1j} \right\} \\
& = \sigma_{r_1}^2 (U_1^T \mathbb{E}(\text{TIPUP}_1)) \sum_{j=m+1}^{r_1} \hat{U}_{1j}^T U_{1j} U_{1j}^T \hat{U}_{1j} \\
& \geq \sigma_{r_1}^2 (U_1^T \mathbb{E}(\text{TIPUP}_1)) (r_1 - m) (1 - \| \hat{U}_{1j}^T - U_{1j}^T \|_2^2),
\end{align*}

where the last step follows from (46). By Proposition 2, in the event \Omega_0, \sigma_{r_1} (U_1^T \mathbb{E}(\text{TIPUP}_1)) = \tau \sim d_1 - \delta_1, with 0 \leq \delta_0 \leq \delta_1 \leq 1. Hence, there exists a constant \( c_m > 0 \) such that in the event \Omega_3,

\begin{equation}
I_1 \geq (c_m + o(1)) d_1^{2-2\delta_1}.
\end{equation}

By Lemma 10,

\begin{align*}
I_2 &= \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{T-h} \sum_{s=h+1}^{T} E_s A_2 F_{t-h} A_1^T \left\{ \frac{1}{T-h} \sum_{s=h+1}^{T} E_s A_2 F_{t-h} A_1^T \right\}_2^2 \\
& = O \left( \frac{d_1 d_1^{1-\delta_0}}{T} \right).
\end{align*}
By Lemma 10,

Applying (42),

This implies that

\[ I_2 + I_3 = O\left(\frac{d^{3/2 - 3\delta_0/2}}{T^{1/2}}\right). \] (75)

Next, we consider \( I_2 \).

\[
I_2 = \sum_{h=1}^{b_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ E_s E_{s-h} \hat{U}_1 \hat{U}_1^\top E_{t-h} E_t^\top \right\} \\
+ \frac{b_0}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ A_1 F_s A_2^\top E_s E_{s-h} \hat{U}_1 \hat{U}_1^\top E_{t-h} A_2 F_t^\top A_1^\top \right\} \\
+ 2 \sum_{h=1}^{b_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ A_1 F_s A_2^\top E_s E_{s-h} \hat{U}_1 \hat{U}_1^\top E_{t-h} E_t^\top \right\} \\
:= I_{11} + I_{12} + 2I_3.
\]

By Lemma 10,

\[
I_1 = \sum_{h=1}^{b_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ E_s E_{s-h} \hat{U}_1 \hat{U}_1^\top E_{t-h} E_t^\top \right\} = O\left(\frac{d}{T}\right).
\]

Applying (42),

\[
I_2 = \sum_{h=1}^{b_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ A_1 F_s A_2^\top E_s E_{s-h}(\hat{U}_{1j} - U_1 \hat{U}_1^\top \hat{U}_{1j}) \right\} \cdot (\hat{U}_{1j} - U_1 \hat{U}_1^\top \hat{U}_{1j})^\top E_{t-h} A_2 F_t^\top A_1^\top \\
+ 2 \sum_{h=1}^{b_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ A_1 F_s A_2^\top E_s E_{s-h}(\hat{U}_{1j} - U_1 \hat{U}_1^\top \hat{U}_{1j}) \right\} \cdot (U_1 \hat{U}_1^\top \hat{U}_{1j})^\top E_{t-h} A_2 F_t^\top A_1^\top \\
+ \sum_{h=1}^{b_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ A_1 F_s A_2^\top E_s E_{s-h}(U_1 \hat{U}_1^\top \hat{U}_{1j}) \right\} \cdot (U_1 \hat{U}_1^\top \hat{U}_{1j})^\top E_{t-h} A_2 F_t^\top A_1^\top \\
+ \sum_{h=1}^{b_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ A_1 F_s A_2^\top E_s E_{s-h}(U_1 \hat{U}_1^\top \hat{U}_{1j}) \right\} \cdot (U_1 \hat{U}_1^\top \hat{U}_{1j})^\top E_{t-h} A_2 F_t^\top A_1^\top
\]
\[ \Pi_3 = \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ A_1 F_s A_2^T E_{s-h}^T (\hat{U}_{1j} - U_1 \hat{U}_{1j}^T \hat{U}_{1j}) \right\} \cdot (\hat{U}_{1j} - U_1 \hat{U}_{1j}^T \hat{U}_{1j})^T E_{t-h} E_t^T \right\} \]

\[ + \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ A_1 F_s A_2^T E_{s-h}^T (U_1 \hat{U}_{1j}^T \hat{U}_{1j}) \right\} \cdot (U_1 \hat{U}_{1j}^T \hat{U}_{1j})^T E_{t-h} E_t^T \right\} \]

\[ + \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ U_1^T A_1 F_s A_2 A_2^T E_{s-h}^T (U_1 \hat{U}_{1j}^T \hat{U}_{1j}) \right\} \cdot (U_1 \hat{U}_{1j}^T \hat{U}_{1j})^T E_{t-h} E_t^T \right\} \]

\[ = O \left( \frac{d_1^2 \delta_0^2 \eta_d^2}{T^2} + \frac{d_1^2 \delta_0^2 \eta_d^2}{T^2} + \frac{d_1^1 \delta_0^2 \eta_d^2}{T^2} \right). \]

Combining \( \Pi_1, \Pi_2 \) and \( \Pi_3 \),

\[ \Pi = O \left( \frac{d}{T} + \frac{d_1^1 \delta_0^2 \eta_d^2}{T^2} \right). \] (76)

Next, we consider \( \Pi_3 \).

\[ \Pi_3 = \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ E_s A_2 A_2^T A_1^T \hat{U}_{1j} \hat{U}_{1j}^T E_{t-h} E_t^T \right\} \]

\[ + \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ A_1 F_s A_2^T A_2^T A_1^T \hat{U}_{1j} \cdot \hat{U}_{1j} E_{t-h} A_2 F_t^T A_1^T \right\} \]

\[ + \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ E_s A_2 F_s A_2 A_2^T A_1^T \hat{U}_{1j} \hat{U}_{1j}^T E_{t-h} A_2 F_t^T A_1^T \right\} \]

\[ + \sum_{h=1}^{h_0} \sum_{j=m+1}^{r_1} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ A_1 F_s A_2^T A_2 F_s A_2^T A_1^T \hat{U}_{1j} \hat{U}_{1j}^T E_{t-h} E_t^T \right\} \]
Determining the number of factors

\( \text{III} := \text{III}_1 + \text{III}_2 + \text{III}_3 + \text{III}_4. \)

We bound each term in turn. Applying (42), using same arguments in the proof of II, we have,

\[
\text{III}_1 = O \left( \frac{d_1^{1/2} d_1^{1-\delta_0/2}}{T} \right),
\]

\[
\text{III}_2 = O \left( \frac{d_1^{1/2} d_3^{1/2-3\delta_0/2} \eta_d^*}{T^{1/2}} + \frac{d_3^{3/2-3\delta_0/2}}{T^{1/2}} \right),
\]

\[
\text{III}_3 = O \left( \frac{d_1^{1/2} d_1^{1-\delta_0} \eta_d^*}{T^{3/2}} + \frac{d_1^{1-\delta_0}}{T} \right),
\]

\[
\text{III}_4 = O \left( \frac{d_3^{3/2-\delta_0} \eta_d^*}{T} + \frac{d_3^{3/2-\delta_0}}{T^{1/2} d_1^{1/2}} \right).
\]

This implies that

\[
\text{III} = O \left( \frac{d_1^{1/2} d_1^{1-\delta_0/2}}{T} + \frac{d_1^{1/2} d_3^{1/2-3\delta_0/2} \eta_d^*}{T^{1/2}} + \frac{d_3^{3/2-3\delta_0/2}}{T^{1/2}} + \frac{d_1^{1-\delta_0}}{T} 
+ \frac{d_3^{3/2-\delta_0}}{T^{1/2} d_1^{1/2}} \right). \tag{77}
\]

Employing (74), (75), (76) and (77), in the event \( \Omega_3 \), there exist constants \( c_m > 0 \) and \( C > 0 \), such that

\[
\mathcal{M}^*(m, \hat{U}_1, m) - \mathcal{M}^*(r_1, \hat{U}_1) 
\geq (c_m + o(1)) d^{-2\delta_1} + C \left( \frac{d_1}{T d^{1/2+\delta_0}} + \frac{1}{T d^{1/2+3\delta_0/2}} + \frac{1}{T^{1/2} d^{1/2+\delta_0} d_1^{1/2}} 
+ \frac{d_1^{1/2} \eta_d^*}{T d^{1/2+3\delta_0/2}} + \frac{\eta_d^*}{T d^{1/2+\delta_0}} \right). \tag{78}
\]

Proof of Theorem 1 for non-iterative TIPUP. The proof is similar to Theorem 1 for non-iterative TOPUP. For the IC estimators base on non-iterative TIPUP, it is followed by Lemma 11, 12 and Assumption VII(a).

Applying similar arguments in the proof of Theorem 1 for non-iterative TOPUP, the IC estimators base on non-iterative TIPUP in Theorem 1 follow from Lemma 11, 12 and Assumption VII(b).

A.4. Proof of Theorem 1 for iTIPUP

Lemma 13. Suppose Assumptions I, II, III, IV and V(b) hold. Assume that \( \| \hat{U}_2 \hat{U}_2^\top - U_2 U_2^\top \|_2 \leq c_0 \sqrt{d_1/d_2} \) for some positive constant \( c_0 \). Then, in an event
Lemma 14. Suppose Assumptions I, II, III, IV and V(b) hold. Let \( r_1 \leq r_1^{(j)} \leq m_1 < d_1 \) for all \( 1 \leq j \leq i - 1 \), and \( m_1 = O(r_1) \). There exists an event \( \Omega_4 \) such that \( \mathbb{P}(\Omega_4) \geq 1 - e^{-d_1} - e^{-d_2} - T \exp(-C_1T^0) - \exp(-C_2T) \), \( C_1, C_2 > 0, 1/\vartheta = 1/\theta_1 + 2/\theta_2 \) and \( \Omega_4 \) is independent of iteration number \( i \). Then, in the event \( \Omega_4 \), at \( i \)-th iteration, for any fixed \( m \) with \( m > r_1 \), \( \mathcal{M}^{s(i)}(r_1, \hat{U}_1) - \mathcal{M}^{s(i)}(m, \hat{U}_{1,m}) \leq C_{\beta_d,T}^{\ast} \), where \( C > 0 \),

\[
\beta_{d,T}^{\ast} = \frac{d_1}{Td^{1+\delta_0}} + \frac{d_1^{1/2} \eta_{k}^{\ast(i)}}{Td^{1/2+3\delta_0/2}},
\]

and \( \eta_{k}^{\ast(i)} = d_1^{1/2} d_{1}^{\delta_1-\delta_0/2-1/2} + d_1^{1/2} d_{1}^{\delta_1-1} \).

Lemma 15. Suppose Assumptions I, II, III, IV and V(b) hold. Let \( r_1 \leq r_1^{(j)} \leq m_1 < d_1 \) for all \( 1 \leq j \leq i - 1 \), and \( m_1 = O(r_1) \). There exists an event \( \Omega_4 \) such that \( \mathbb{P}(\Omega_4) \geq 1 - e^{-d_1} - e^{-d_2} - T \exp(-C_1T^0) - \exp(-C_2T) \), \( C_1, C_2 > 0, 1/\vartheta = 1/\theta_1 + 2/\theta_2 \) and \( \Omega_4 \) is independent of iteration number \( i \). Then, in the event \( \Omega_4 \), at \( i \)-th iteration, for any \( m \) with \( m < r_1 \), there exist constants \( c_m > 0 \) and \( C > 0 \), such that \( \mathcal{M}^{s(i)}(m, \hat{U}_{1,m}) - \mathcal{M}^{s(i)}(r_1, \hat{U}_1) \geq (c_m + o(1))d^{-2\delta_1} + C_{\gamma_d,T}^{\ast}, \)

where

\[
\gamma_{d,T}^{\ast} = \frac{d_1}{Td^{1+\delta_0}} + \frac{1}{Td^{1/2+3\delta_0/2}} + \frac{d_1^{1/2} \eta_{k}^{\ast(i)}}{Td^{1/2+3\delta_0/2}},
\]

and \( \eta_{k}^{\ast(i)} = d_1^{1/2} d_{1}^{\delta_1-\delta_0/2-1/2} + d_1^{1/2} d_{1}^{\delta_1-1} \).
A.5. Proof of propositions and Theorem 2

Proof of Proposition 2. Let $A_k$ has SVD $A_k = U_k A_k V_k^\top$. As $|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \leq 1/4 \leq e^{-1}$ for all event pairs $\{A, B\}$,

$$
\sup_t \left\{ \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) : A \in \sigma(F_{s+t} \otimes F_s, s \leq t, \tau \leq \tau_0), B \in \sigma(F_s, s \geq t + h) \right\}
$$

$$
\leq \sup_t \left\{ \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) : A \in \sigma(F_s, s \leq t + \tau_0), B \in \sigma(F_s, s \geq t + h) \right\}
$$

$$
\leq \exp \left\{ - \max\{1, c_0(h - \tau_0)^\theta_1\} \right\}
$$

where $c_0 = c_0(1 + c_0^{1/\theta_1} \tau_0)^{-\theta_1}$.

Since $r_k$ is fixed, $\mathbb{E}(F_{i,k,t-h} F_{j,l,t}) = O(1)$ for any $0 \leq h \leq T/4$, $1 \leq i, j \leq r_1$, $1 \leq k, l \leq r_2$. Pick $a > 0$, $x > \mathbb{E}(F_{i,k,t-h} F_{j,l,t})$, then

$$
\mathbb{P}(\left| F_{i,k,t-h} F_{j,l,t} - \mathbb{E}F_{i,k,t-h} F_{j,l,t} \right| > x)
$$

$$
= \mathbb{P}(\left| F_{i,k,t-h} F_{j,l,t} - \mathbb{E}F_{i,k,t-h} F_{j,l,t} \right| > x, |F_{i,k,t-h}| > x^a)
$$

$$
+ \mathbb{P}(\left| F_{i,k,t-h} F_{j,l,t} - \mathbb{E}F_{i,k,t-h} F_{j,l,t} \right| > x, |F_{i,k,t-h}| \leq x^a)
$$

$$
\leq \mathbb{P}(\left| F_{i,k,t-h} \right| > x^a) + \mathbb{P}(\left| F_{j,l,t} \right| > x^{1-a})
$$

$$
\leq c_1 \exp \left( -c_2 x^{a \theta_2} \right) + c_1 \exp \left( -c_2 x^{(1-a) \theta_2} \right)
$$

$$
\leq 2c_1 \exp \left( -c_2 x^{\theta_2/2} \right).
$$

Let $1/\theta = 1/\theta_1 + 2/\theta_2$. Hence, by Theorem 1 in Merlevède, Peligrad and Rio (2011), for any $\|u\|_2 = \|v\|_2 = 1$,

$$
\mathbb{P}\left( \frac{1}{T-h} \sum_{t=h+1}^{T} u^\top (F_{i,t-h} F_{i,t}^\top - \mathbb{E}F_{i,t-h} F_{i,t}^\top) v \bigg| > x \right)
$$

$$
\leq (T-h) \exp \left( -c_3(T-h)^\theta x^\theta \right) + \exp \left( -c_4(T-h)x^2 \right).
$$

By Lemma 18(i),

$$
\mathbb{P}\left( \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} (F_{i,t-h} F_{i,t}^\top - \mathbb{E}F_{i,t-h} F_{i,t}^\top) \right\|_2 > x \right)
$$
\[
\leq 5^r \cdot \mathbb{P} \left( \left| \frac{1}{T-h} \sum_{t=h+1}^{T} u^\top (F_{t-h}F_{t}^\top - \mathbb{E}F_{t-h}F_{t}^\top) v \right| > x/2 \right) \\
\leq 5^r (T-h) \exp \left( -c_3(T-h)^{\vartheta}(x/2)^{\vartheta} \right) + 5^r \exp \left( -c_4(T-h)x^2/4 \right).
\]

As \( h \leq T/4 \), choosing \( x = 2^{-1} \sigma_{r_1} \left( (T-h)^{-1} \sum_{t=h+1}^{T} \mathbb{E}F_{t-h}F_{t}^\top \right) \vee 1 \), in an event \( \Omega_0 \) with probability at least \( 1 - T \exp(-C_1T^\vartheta) - \exp(-C_2T) \),

\[
\left\| \frac{1}{T-h} \sum_{t=h+1}^{T} (F_{t-h}F_{t}^\top - \mathbb{E}F_{t-h}F_{t}^\top) \right\|_2 \leq \frac{1}{2} \sigma_{r_1} \left( \frac{1}{T-h} \sum_{t=h+1}^{T} \mathbb{E}F_{t-h}F_{t}^\top \right) \vee 1.
\]

(85)

By triangle inequality,

\[
\left\| \frac{1}{T-h} \sum_{t=h+1}^{T} \mathbb{E}M_{t-h}M_t^\top \right\|_2 - \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} \left( M_{t-h}M_t^\top - \mathbb{E}M_{t-h}M_t^\top \right) \right\|_2 \\
\leq \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} M_{t-h}M_t^\top \right\|_2 \\
\leq \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} \mathbb{E}M_{t-h}M_t^\top \right\|_2 + \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} \left( M_{t-h}M_t^\top - \mathbb{E}M_{t-h}M_t^\top \right) \right\|_2.
\]

In the event \( \Omega_0 \), by (85) and Assumption V(b),

\[
\left\| \frac{1}{T-h} \sum_{t=h+1}^{T} \left( M_{t-h}M_t^\top - \mathbb{E}M_{t-h}M_t^\top \right) \right\|_2 \\
\leq d^{1-\delta_0} \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} (F_{t-h}F_{t}^\top - \mathbb{E}F_{t-h}F_{t}^\top) \right\|_2 \\
\leq \frac{1}{2} \sigma_{r_1} \left( \frac{1}{T-h} \sum_{t=h+1}^{T} \mathbb{E}M_{t-h}M_t^\top \right).
\]

It follows that

\[
\| \Theta_{1,h}^\top \|_2 = \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} M_{t-h}M_t^\top \right\|_2 \asymp \left\| \frac{1}{T-h} \sum_{t=h+1}^{T} \mathbb{E}M_{t-h}M_t^\top \right\|_2 \asymp d^{1-\delta_0},
\]

for \( h = 0 \) and some \( 1 \leq h \leq h_0 \leq T/4 \). Similarly, by (85) and Assumption V(b), for some \( 1 \leq h \leq h_0 \leq T/4 \),

\[
\sigma_{r_1} \left( \frac{1}{T-h} \sum_{t=h+1}^{T} M_{t-h}M_t^\top \right) \asymp d^{1-\delta_1},
\]
Then,
\[ \tau_{1,r_1}^* \approx d^{1-\delta_1}. \]
Hence, in the event \( \Omega_0 \) with probability at least \( 1 - C_1 \exp(-C_2T^\delta) - C_3 \exp(-C_2T) \),
\[ \| \Theta_{i,0}^* \|_2 \approx d^{1-\delta_0}, \quad \tau_{1,r_1}^* \approx d^{1-\delta_1}. \quad (86) \]
Similarly, applying arguments of proving Lemma 5 in Wang, Liu and Chen (2019), we can show, in an event with probability at least \( 1 - T \exp(-C_1T^\delta) - \exp(-C_2T) \),
\[ \| \Theta_{1,0} \|_{\text{op}} \approx d^{1-\delta_0}, \quad \tau_{1,r_1} \approx d^{1-\delta_1}. \quad (87) \]

**Proof of Theorem 2.** We only consider the case of non-iterative TOPUP, as the other cases will be similar. Let \( \eta_d = d^{3-\delta_0/2} + d\delta_1 d_1^{-1/2} \). For any \( 0 \leq m < n < d_1 \),
\[ \sum_{j=m+1}^{n} \hat{\lambda}_{1,j} = d^2 \mathcal{M}(m, \hat{U}_{1,m}) - d^2 \mathcal{M}(n, \hat{U}_{n}). \quad (88) \]
By Lemma 4, in the same event \( \Omega_1 \), \( \hat{\lambda}_{r_1+1} = O(d^2 \beta_{d,T}) \). As \( \hat{\lambda}_{r_1+1} \geq \hat{\lambda}_{r_1+2} \geq \ldots \geq \hat{\lambda}_{d_1} \), part (i) follows.
Set \( n = m + 1 \leq r_1 \) in (88). Employing the same arguments in the proof of Lemma 5, we can show \( I_2 + I_3 + II + III = O(d^2 \beta_{d,T}) \) in the same event \( \Omega_1 \). For \( I_1 \),
\[ I_1 = \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^\top (A_1 F_s A_2^\top \otimes A_1 F_s A_2^\top) \hat{U}_{1m} \hat{U}_{1m}^\top \right. \]
\[ \cdot \left. \text{mat}_1 (A_1 F_{t-h} A_2^\top \otimes A_1 F_t A_2^\top) \right\} \]
\[ = \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^\top (A_1 F_s A_2^\top \otimes A_1 F_s A_2^\top) (U_{1m} U_{1m}^\top - \hat{U}_{1m} \hat{U}_{1m}^\top) \right. \]
\[ \cdot \left. \text{mat}_1 (A_1 F_{t-h} A_2^\top \otimes A_1 F_t A_2^\top) \right\} \]
\[ + \sum_{h=1}^{h_0} \frac{1}{(T-h)^2} \sum_{s,t=h+1}^{T} \text{tr}\left\{ \text{mat}_1^\top (A_1 F_s A_2^\top \otimes A_1 F_s A_2^\top) U_{1m} U_{1m}^\top \right. \]
\[ \cdot \left. \text{mat}_1 (A_1 F_{t-h} A_2^\top \otimes A_1 F_t A_2^\top) \right\} \]
\[ := I_{11} + I_{12}. \]
By (42),
\[ I_{11} \leq \| U_{1m} U_{1m}^\top - \hat{U}_{1m} \hat{U}_{1m}^\top \|_2 \left\| \sum_{h=1}^{h_0} U_{1m}^\top \text{mat}_1 (\Theta_{1,h}) \text{mat}_1^\top (\Theta_{1,h}) U_{1m} \right\|_2.
Suppose Lemma 17. Let Lemma 16. Appendix B: Technical lemmas

1794

Y. Han et al.

is an orthogonal matrix such that span$(Q_1)$ is an invariant subspace of $A$, where $Q_1 \in \mathbb{R}^{n \times r}$ and $Q_2 \in \mathbb{R}^{n \times (n-r)}$. Partition the matrices $Q^\top AQ$ and $Q^\top EQ$ as follows:

$$Q^\top AQ = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}, \quad Q^\top EQ = \begin{pmatrix} E_{11} & E_{12}^\top \\ E_{21} & E_{22} \end{pmatrix}.$$  

If $\text{sep}(D_1, D_2) = \min_{\lambda \in \lambda(D_1), \mu \in \lambda(D_2)} |\lambda - \mu| > 0$, where $\lambda(M)$ denotes the set of eigenvalues of the matrix $M$, and $\|E\|_2 \leq \text{sep}(D_1, D_2)/\delta$, then there exists a matrix $P \in \mathbb{R}^{(n-r) \times r}$ with

$$\|P\|_2 \leq \frac{4}{\text{sep}(D_1, D_2)} \|E_{21}\|_2$$
such that the columns of $\tilde{Q}_1 = (Q_1 + Q_2 P)(I + P^T P)^{-1/2}$ define an orthonormal basis for a subspace that is invariant for $A + E$.

**Proof.** See Theorem 8.1.10 in Golub and Van Loan (2012).

**Lemma 18.** Let $d, d_j, d_*, r \leq d \land d_j$ be positive integers, $\epsilon > 0$ and $N_{d, \epsilon} = \lfloor (1 + 2/\epsilon)^d \rfloor$.

(i) For any norm $\| \cdot \|$ in $\mathbb{R}^d$, there exist $M_j \in \mathbb{R}^d$ with $\|M_j\| \leq 1$, $j = 1, \ldots, N_{d, \epsilon}$, such that $\max_{\|M\| \leq 1} \min_{1 \leq j \leq N_{d, \epsilon}} \|M - M_j\| \leq \epsilon$. Consequently, for any linear mapping $f$ and norm $\| \cdot \|_*$,

$$\sup_{M \in \mathbb{R}^d, \|M\| \leq 1} \|f(M)\|_* \leq 2 \max_{1 \leq j \leq N_{d, \epsilon/2}} \|f(M_j)\|_*.$$  

(ii) Given $\epsilon > 0$, there exist $U_j \in \mathbb{R}^{d \times r}$ and $V_j \in \mathbb{R}^{d \times r}$ with $\|U_j\|_2 \vee \|V_j\|_2 \leq 1$ such that

$$\max_{M \in \mathbb{R}^{d \times d'}, \|M\|_2 \leq 1, \text{rank}(M) \leq r} \min_{1 \leq j \leq N_{d, \epsilon/2}^r} \|M - U_j V_j^T\|_2 \leq \epsilon.$$  

Consequently, for any linear mapping $f$ and norm $\| \cdot \|_*$ in the range of $f$,

$$\sup_{M, M' \in \mathbb{R}^{d \times d'}, \|M - M'\|_2 \leq \epsilon/2^{1+\delta}} \|f(M) - f(M')\|_* \leq 2 \max_{1 \leq j \leq N_{d, \epsilon/2}^r} \|f(U_j V_j^T)\|_*.$$  

(iii) Given $\epsilon > 0$, there exist $U_{j,k} \in \mathbb{R}^{d_k \times r_k}$ and $V_{j',k} \in \mathbb{R}^{d_k' \times r_k}$ with $\|U_{j,k}\|_2 \vee \|V_{j',k}\|_2 \leq 1$ such that

$$\min_{M_k \in \mathbb{R}^{d_k \times d_k'}, \|M_k\|_2 \leq 1 \land \text{rank}(M_k) \leq r_k} \max_{j_k \leq N_{d_k, r_k}^r, r_k} \min_{j_k' \leq N_{d_k', r_k}^r, r_k} \|\sum_{k=2}^{K} M_k - \sum_{k=2}^{K} (U_{j_k,k} V_{j_k,k}^T)\|_2 \leq \epsilon(K - 1).$$

For any linear mapping $f$ and norm $\| \cdot \|_*$ in the range of $f$,

$$\sup_{M_k, M_k' \in \mathbb{R}^{d_k \times d_k'}, \|M_k - M_k'\|_2 \leq \epsilon/(2K - 2)} \|f(\sum_{k=2}^{K} M_k) - f(\sum_{k=2}^{K} M_k')\|_* \leq 2 \max_{M_k \in \mathbb{R}^{d_k \times d_k'}, \|M_k\|_2 \leq 1} \|f(\sum_{k=2}^{K} M_k)\|_*.$$  

and

$$\sup_{M_k \in \mathbb{R}^{d_k \times d_k'}, \|M_k\|_2 \leq 1} \|f(\sum_{k=2}^{K} M_k)\|_* \leq 2 \max_{1 \leq j_k \leq N_{d_k, r_k}^r, r_k} \min_{1 \leq j_k' \leq N_{d_k', r_k}^r, r_k} \|f(\sum_{k=2}^{K} U_{j_k,k} V_{j_k,k}^T)\|_*.$$  

**Proof.** See Lemma 1 in Han et al. (2020).

**Lemma 19.** (i) Let $G \in \mathbb{R}^{d_1 \times n}$ and $H \in \mathbb{R}^{d_2 \times n}$ be two centered independent Gaussian matrices such that $\mathbb{E}(u^\top \text{vec}(G))^2 \leq \sigma^2$ and $u \in \mathbb{R}^{d_1 n}$ and $\mathbb{E}(v^\top \text{vec}(H))^2 \leq \sigma^2$. Then,
\[ \|GH^T\|_2 \leq \sigma^2(\sqrt{d_1d_2} + \sqrt{d_3n} + \sqrt{d_2n}) + \sigma^2(x + 2\sqrt{n} + \sqrt{d_1} + \sqrt{d_2}) \quad (92) \]

with at least probability \(1 - 2e^{-x^2/2}\) for all \(x \geq 0\).

(ii) Let \(G_i \in \mathbb{R}^{d_1 \times d_2}, H_i \in \mathbb{R}^{d_3 \times d_4}, i = 1, \ldots, n\), be independent centered Gaussian matrices such that \(E(u^T \text{vec}(G_i))^2 \leq \sigma^2 \forall u \in \mathbb{R}^{d_1d_2}\) and \(E(v^T \text{vec}(H_i))^2 \leq \sigma^2 \forall v \in \mathbb{R}^{d_3d_4}\). Then,

\[
\left\| \text{mat}_1 \left( \sum_{i=1}^n G_i \otimes H_i \right) \right\|_2 \leq \sigma^2(\sqrt{d_1n} + \sqrt{d_1d_3d_4} + \sqrt{nd_2d_3d_4}) + \sigma^2(x + \sqrt{n} + \sqrt{d_1} + \sqrt{d_2} + \sqrt{d_3d_4}) \quad (93)\]

with at least probability \(1 - 2e^{-x^2/2}\) for all \(x \geq 0\).

Proof. See Lemma 2 in Han et al. (2020).

\]

### Appendix C: Additional simulation results

In this section, we show detailed comparison among IC1-IC5, and ER1-ER5 for the first part and third part simulation in Section 5. We also study a strong factor model with \(r_1 = r_2 = 2\).

(M0). Set \(r_1 = r_2 = 2\). The univariate \(f_{ijt}\) follows AR(1) with AR coefficient \(\phi_{11} = \phi_{22} = 0.8\) and \(\phi_{12} = \phi_{21} = 0.3\); All elements of \(A_1\) and \(A_2\) are i.i.d \(N(0,1)\).

For Model M0 with small \(r_1 = r_2 = 2\) and all strong factors, all versions of our methods show perfect accuracy for all sample sizes except \(T = 100\) when some versions show error rates less than 1%. For Model M1, the results in Table 14 show clearly that, among all five penalty functions, IC2 and IC4 seem to be slightly better than the others. With the same settings, the determination of

### Table 14

Proportion of correct identification of rank \(r\) using IC estimators based on TOPUP procedures for Model M1, over 1000 replications

<table>
<thead>
<tr>
<th>(d_1 = d_2 = 20)</th>
<th>(d_1 = d_2 = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(T)</td>
<td>IC1</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>200</td>
<td>0</td>
</tr>
<tr>
<td>300</td>
<td>0</td>
</tr>
<tr>
<td>500</td>
<td>0</td>
</tr>
<tr>
<td>1000</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(d_1 = d_2 = 40)</th>
<th>(d_1 = d_2 = 80)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(T)</td>
<td>IC1</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>200</td>
<td>0</td>
</tr>
<tr>
<td>300</td>
<td>0</td>
</tr>
<tr>
<td>500</td>
<td>0</td>
</tr>
<tr>
<td>1000</td>
<td>0</td>
</tr>
</tbody>
</table>
the ranks are perfect using IC and TIPUP, or using ER estimators. Thus, we omitted the tables.

For Model M2, Tables 15 and 16 report the proportion of correct rank identification using IC estimators based on TOPUP and TIPUP procedures, respectively. Tables 17 and 18 report the proportion of correct rank identification using ER estimators based on TOPUP and TIPUP procedures, respectively.

### Table 15
Proportion of correct identification of rank \( r \) using IC estimators based on TOPUP procedures for Model M2, over 1000 replications

<table>
<thead>
<tr>
<th>( r )</th>
<th>IC1</th>
<th>IC2</th>
<th>IC3</th>
<th>IC4</th>
<th>IC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_1 = d_2 = 2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.050</td>
<td>0.050</td>
<td>0.050</td>
<td>0.050</td>
<td>0.050</td>
</tr>
<tr>
<td>200</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
</tr>
<tr>
<td>300</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
</tr>
<tr>
<td>500</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
</tr>
<tr>
<td>1000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>( d_1 = d_2 = 4 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
</tr>
<tr>
<td>200</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
</tr>
<tr>
<td>300</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
</tr>
<tr>
<td>500</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
</tr>
<tr>
<td>1000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>( d_1 = d_2 = 8 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
</tr>
<tr>
<td>200</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
</tr>
<tr>
<td>300</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
</tr>
<tr>
<td>500</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
</tr>
<tr>
<td>1000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

### Table 16
Proportion of correct identification of rank \( r \) using IC estimators based on TIPUP procedures for Model M2, over 1000 replications

<table>
<thead>
<tr>
<th>( r )</th>
<th>IC1</th>
<th>IC2</th>
<th>IC3</th>
<th>IC4</th>
<th>IC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_1 = d_2 = 2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.050</td>
<td>0.050</td>
<td>0.050</td>
<td>0.050</td>
<td>0.050</td>
</tr>
<tr>
<td>200</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
</tr>
<tr>
<td>300</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
</tr>
<tr>
<td>500</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
</tr>
<tr>
<td>1000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>( d_1 = d_2 = 4 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
</tr>
<tr>
<td>200</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
</tr>
<tr>
<td>300</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
</tr>
<tr>
<td>500</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
</tr>
<tr>
<td>1000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>( d_1 = d_2 = 8 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
</tr>
<tr>
<td>200</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
</tr>
<tr>
<td>300</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
</tr>
<tr>
<td>500</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
<td>0.990</td>
</tr>
<tr>
<td>1000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

For Model M3 with all very weak factors (\( \delta_0 = \delta_1 = 0.6 \)), Tables 19 and 20 report RMSEs of the ER estimators. The results of using IC estimators are not shown, as it is impossible to detect weak factors if we set \( \delta_1 = 0 \) in \( g_k \).

In summary, the first part of simulation shows that in IC estimators, IC2 and IC4 seem to perform slightly better than IC1, IC3 and IC5 in Table 14, while it is reversed in Tables 15 and 16. The difference is not significant though. In ER estimators, the choice of the penalty function also seems to have a limited impact on the results. In most cases, ER1 and ER2 are slightly better.
### Table 17
Proportion of correct identification of rank \( r \) using ER estimators based on TOPUP procedures for Model M2, over 1000 replications

<table>
<thead>
<tr>
<th>( T )</th>
<th>( \hat{d}_1 )</th>
<th>( \hat{d}_2 )</th>
<th>( \hat{d}_3 )</th>
<th>( \hat{d}_4 )</th>
<th>( \hat{d}_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.029</td>
<td>0.029</td>
<td>0.029</td>
<td>0.029</td>
<td>0.029</td>
</tr>
<tr>
<td>200</td>
<td>0.008</td>
<td>0.008</td>
<td>0.008</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>300</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>500</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>1000</td>
<td>0.004</td>
<td>0.004</td>
<td>0.004</td>
<td>0.004</td>
<td>0.004</td>
</tr>
</tbody>
</table>

### Table 18
Proportion of correct identification of rank \( r \) using ER estimators based on TIPUP procedures for Model M2, over 1000 replications

<table>
<thead>
<tr>
<th>( T )</th>
<th>( \hat{d}_1 )</th>
<th>( \hat{d}_2 )</th>
<th>( \hat{d}_3 )</th>
<th>( \hat{d}_4 )</th>
<th>( \hat{d}_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.029</td>
<td>0.029</td>
<td>0.029</td>
<td>0.029</td>
<td>0.029</td>
</tr>
<tr>
<td>200</td>
<td>0.008</td>
<td>0.008</td>
<td>0.008</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>300</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>500</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>1000</td>
<td>0.004</td>
<td>0.004</td>
<td>0.004</td>
<td>0.004</td>
<td>0.004</td>
</tr>
</tbody>
</table>

### Table 19
Root mean squared errors (RMSEs) of the ER estimators based on TOPUP procedures for Model M3, averaging over 1000 replications

<table>
<thead>
<tr>
<th>( T )</th>
<th>( \hat{d}_1 )</th>
<th>( \hat{d}_2 )</th>
<th>( \hat{d}_3 )</th>
<th>( \hat{d}_4 )</th>
<th>( \hat{d}_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>2.135</td>
<td>2.29</td>
<td>2.15</td>
<td>2.15</td>
<td>2.17</td>
</tr>
<tr>
<td>200</td>
<td>1.68</td>
<td>1.76</td>
<td>1.68</td>
<td>1.68</td>
<td>1.73</td>
</tr>
<tr>
<td>300</td>
<td>1.17</td>
<td>1.17</td>
<td>1.17</td>
<td>1.19</td>
<td>1.07</td>
</tr>
<tr>
<td>500</td>
<td>1.02</td>
<td>1.02</td>
<td>1.02</td>
<td>1.02</td>
<td>1.02</td>
</tr>
<tr>
<td>1000</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

\( d_1 = d_2 = 40 \)

\( d_3 = d_4 = 40 \)

\( d_5 = d_6 = 50 \)
### Table 20

| \( t \) | \( \text{ER1} \) | \( \text{ER2} \) | \( \text{ER3} \) | \( \text{ER4} \) | \( \text{ER5} \) | \( \text{ER1} \) | \( \text{ER2} \) | \( \text{ER3} \) | \( \text{ER4} \) | \( \text{ER5} \) | \( \text{ER1} \) | \( \text{ER2} \) | \( \text{ER3} \) | \( \text{ER4} \) | \( \text{ER5} \) | \( \text{ER1} \) | \( \text{ER2} \) | \( \text{ER3} \) | \( \text{ER4} \) | \( \text{ER5} \) |
| 100 | 1.50 | 1.76 | 1.50 | 1.50 | 1.51 | 1.14 | 1.63 | 1.15 | 1.15 | 1.20 | 0.45 | 1.19 | 0.45 | 0.45 | 0.52 | 0.20 | 0.35 | 0.20 | 0.20 | 0.20 |
| 200 | 0.84 | 1.00 | 0.84 | 0.84 | 0.86 | 0.61 | 0.84 | 0.61 | 0.61 | 0.65 | 0.20 | 0.35 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 |
| 300 | 0.60 | 0.66 | 0.60 | 0.60 | 0.61 | 0.43 | 0.50 | 0.43 | 0.43 | 0.44 | 0.12 | 0.14 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 |
| 500 | 0.30 | 0.33 | 0.30 | 0.30 | 0.30 | 0.17 | 0.18 | 0.17 | 0.18 | 0.17 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 |
| 5000 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |

### Table 21

| \( t \) | \( \text{IC1} \) | \( \text{IC2} \) | \( \text{IC3} \) | \( \text{IC4} \) | \( \text{IC5} \) | \( \text{IC1} \) | \( \text{IC2} \) | \( \text{IC3} \) | \( \text{IC4} \) | \( \text{IC5} \) | \( \text{IC1} \) | \( \text{IC2} \) | \( \text{IC3} \) | \( \text{IC4} \) | \( \text{IC5} \) | \( \text{IC1} \) | \( \text{IC2} \) | \( \text{IC3} \) | \( \text{IC4} \) | \( \text{IC5} \) |
| 100 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.002 | 0.001 | 0.004 | 0.012 | 0.001 | 0.056 | 0.056 | 0.056 | 0.056 | 0.056 | 0.056 | 0.056 |
| 300 | 0.04 | 0.01 | 0.04 | 0.01 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.994 | 0.994 | 0.994 | 0.994 | 0.994 | 0.994 | 0.994 | 0.994 |
| 500 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.106 | 0.619 | 0.168 | 0.689 | 0.006 | 0.960 | 0.962 | 0.962 | 0.962 | 0.962 | 0.962 | 0.962 |
| 1000 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.152 | 0.947 | 0.494 | 0.974 | 0.226 | 0.990 | 0.997 | 0.993 | 0.998 | 0.998 | 0.998 | 0.998 |

### Table 22

| \( t \) | \( \text{IC1} \) | \( \text{IC2} \) | \( \text{IC3} \) | \( \text{IC4} \) | \( \text{IC5} \) | \( \text{IC1} \) | \( \text{IC2} \) | \( \text{IC3} \) | \( \text{IC4} \) | \( \text{IC5} \) | \( \text{IC1} \) | \( \text{IC2} \) | \( \text{IC3} \) | \( \text{IC4} \) | \( \text{IC5} \) | \( \text{IC1} \) | \( \text{IC2} \) | \( \text{IC3} \) | \( \text{IC4} \) | \( \text{IC5} \) |
| 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 300 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 500 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

In the third part of the simulation, Table 21 shows the proportion of correct rank identification of the IC estimators in (7) with given \( \delta_1 \) for Model M2. Moreover, Table 22 reports the RMSEs of the IC estimators in (7) with known \( \delta_1 \) for Model M3. Again, given \( \delta_1 \), the performance of all IC1-IC5 is also very similar. In some cases, IC2 and IC4 are slightly better.
References


LI, H., LI, Q. and SHI, Y. (2017). Determining the number of factors when the number of factors can increase with sample size. *Journal of Econometrics* **197** 76–86. MR3598647


Determining the number of factors


