Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Rejoinder

Rong Chen, Dan Yang & Cun-Hui Zhang

To cite this article: Rong Chen, Dan Yang & Cun-Hui Zhang (2022) Rejoinder, Journal of the American Statistical Association, 117:537, 128-132, DOI: 10.1080/01621459.2022.2035099

To link to this article: https://doi.org/10.1080/01621459.2022.2035099

Published online: 24 Mar 2022.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2022, VOL. 117, NO. 537, 128–132: Theory and Methods Discussion
https://doi.org/10.1080/01621459.2022.2035099

Rejoinder

Rong Chen[a], Dan Yang[b], and Cun-Hui Zhang[a]

[a]Department of Statistics, Rutgers University, Piscataway, NJ; [b]Faculty of Business and Economics, The University of Hong Kong, Hong Kong

We are honored to have this lively discussion of our work, and we thank the distinguished discussants for their encouragements, insightful comments and new perspectives. In the following we make an attempt to respond to what we perceive as being the major points raised by the discussants.

1. Structured Tensor Factor Model with Main and Interaction Factors

Professor Peña proposed an interesting and important extension of the tensor factor model and touched on two important issues: structured factor process and estimation with strong and weak factors.

Separating the main effects and interactions has been shown to be an effective and important tool in the analysis of multi-classification data in many fields. It is important in tensor data analysis as well. Often, the main effects are “stronger” than the interactions, more important to be discovered, and easier to interpret. The concept of “main factors” and “interacting factors” are interesting and important.

Following the perspective offered by Peña, here we discuss a more comprehensive and, in a way, more unified modeling strategy, and discuss its various implications.

We note that the main effect model in Mode-1 (dimension d_1) proposed by Peña can be written in our notation as a special case of

$$\mathcal{X}_t = \mathcal{F}_t \times_1 \boldsymbol{A}_1 \times_2 \cdots \times_k \boldsymbol{A}_k + \mathcal{E}_t,$$

where $\mathcal{F}_t$ is a dimension r_1 vector (or a $r_1 \times 1 \times \cdots \times 1$ tensor), $\boldsymbol{A}_1$ is a $d_1 \times r_1$ matrix, and for $k = 2, \ldots, K$, $\boldsymbol{A}_k$ is a nonzero d_k dimensional vector. In this case, all mode-1 fibers share the same factors, with loading matrices proportional to $\boldsymbol{A}_1$. Peña's discussion is focused on the case where $\boldsymbol{A}_k = (1, \ldots, 1)^\top$, $k = 2, \ldots, K$.

Adopting this presentation, model (5) in Peña's discussion can be extended to include main effect, two-way interactions, and up to K-way interaction, in a unified fashion.

$$\mathcal{X}_t = \sum_{i=1}^{K} \mathcal{F}^{(1)}_{t,i} \times_{k=1}^{K} \boldsymbol{A}^{(1)}_{k,i} + \sum_{i_1<i_2}^{K} \mathcal{F}^{(2)}_{t,i_1,i_2} \times_{k=1}^{K} \boldsymbol{A}^{(2)}_{k,i_1,i_2} + \cdots + \mathcal{F}^{(K)}_t \times_{k=1}^{K} \boldsymbol{A}^{(K)}_k. \tag{1}$$

Here $\mathcal{F}^{(1)}_{t,i}$ is a $r^{(1)}_i$ dimensional vector representing the main effect on the i-th mode, with $\boldsymbol{A}^{(1)}_{i,i}$ a $d_i \times r^{(1)}_i$ loading matrix, and the other $\boldsymbol{A}^{(1)}_{k,i}$ being d_k dimensional vector for $k \neq i$. Similarly, $\mathcal{F}^{(2)}_{t,i_1,i_2}$ is a $r^{(2)}_{i_1} \times r^{(2)}_{i_2}$ matrix representing the 2-way interaction factors between the i_1-th and i_2-th modes, $\boldsymbol{A}^{(2)}_{k,i_1,i_2}$ is a $d_k \times r^{(2)}_k$ matrix when $k = i_1$ or i_2, and $\boldsymbol{A}^{(2)}_{k,i_1,i_2}$ is a vector of dimension d_k for $k \notin \{i_1, i_2\}$. The last term is the interaction of all K modes, where $\mathcal{F}^{(K)}_t$ is a $r^{(K)}_1 \times \cdots \times r^{(K)}_K$ tensor, and $\boldsymbol{A}^{(K)}_k$ is a $d_k \times r^{(K)}_k$ loading matrix.

Note that Model (1) can be written as

$$\mathcal{X}_t = \mathcal{F}^*_t \times_1 \boldsymbol{A}^*_1 \times_2 \cdots \times_k \boldsymbol{A}^*_k + \boldsymbol{E}_t$$

where $\mathcal{F}^*_t$ is a block diagonal tensor consisting of $(\mathcal{F}^{(1)}_{t,1}, \ldots, \mathcal{F}^{(1)}_{t,K}, \mathcal{F}^{(2)}_{t,1,1}, \ldots, \mathcal{F}^{(K)}_t)$, and $\boldsymbol{A}^*_i$ is a block diagonal matrix consisting of all of corresponding loading matrices on the i-mode. Although the above model is a special case of the tensor factor model in our paper, the block diagonal structures in the factor process and the loading matrices have important implications.

First, the estimation of such a model structure requires special treatment. The general TIPUP and TOPUP procedures do not work here, as only the loading spaces can be estimated. Although it may be possible to develop a procedure to identify the block diagonal structure based on the estimated loading space, it may not be as efficient as directly utilizing the special structure in estimation. Peña's proposed sequential estimation procedure is an interesting approach to this difficult problem. Each term in Model (1) can be sequentially estimated using TIPUP or TOPUP with the given factor tensor dimension, based on the residual obtained in the previous step. Theoretical investigation on the properties of such a sequential estimation procedure is needed.

Second, the proposed sequential estimation procedure may also provide a valuable solution to the problem of having mixed strong and weak factors in the model. Joint estimation and the determination of the number of factors often encounter difficulties in such cases as strong factors often dominate the signal, resulting in inaccurate selection of the number of factors or inaccurate estimation of the weak factors. These issues may become more severe when the main effects are much stronger than the interactions. Hopefully, the sequential approach would alleviate some of the difficulties in separating strong and weak factors, and provide a way to determine the numbers of factors of different strengths. Again, theoretical investigation is needed.

CONTACT Rong Chen rongchen@stat.rutgers.edu Department of Statistics, Rutgers University, Piscataway, NJ

Third, as Model (1) encompasses many sub-models, a model selection procedure becomes more important. Sequential rank selection (with the possibility of zero rank) is one possibility, as Peña suggested. Certain information criteria based procedure can also be useful and demands further investigation.

Peña also pointed out the potential main effects in the import–export data. His results, which are very convincing, demonstrate the great potential of the structured tensor factor model in broader applications. The clustering idea Peña mentioned is interesting, as it relaxes the rigid row/column classifications in tensor time series. Even within the tensor factor model, cluster structure within a mode (e.g., columns) can be built in Model (1) above, by adding additional terms with specific constraints in the loading matrices. It adds a new dimension to the tensor factor model.

2. Relationship between TIPUP and the Partial Trace Operator

Professors Tang and Linton (TL) pointed out a very interesting perspective in tensor estimation procedure. It is based on the Krnonecker product form of the covariance matrix of the vectorized tensor under the tensor factor model. The keen observation is that, if $\boldsymbol{C} = \boldsymbol{A} \odot \boldsymbol{B}$, then $\boldsymbol{C}$ has a block structure, with its (i,j)-th block in the form of $a_{ij}\boldsymbol{B}$ where a_{ij} is the (i,j)-th element of $\boldsymbol{A}$. Then the trace of the block is $a_{ij}\mathrm{trace}(\boldsymbol{B})$, proportional of a_{ij}, hence, providing an estimate of $\boldsymbol{A}$. We observe the following.

First, the partial trace operator provides another neat way of writing the TIPUP estimator which is described using matrix inner product (after tensor unfolding) in (18) and partial inner product in (21) in our paper. Indeed, the TIPUP_k statistic in (19) can be written as the partial trace operator applied to the sample cross-covariance matrix $\widehat{\boldsymbol{\Sigma}}_h = (T-h)^{-1}\sum_{t=h+1}^{T}\mathrm{vec}(\mathcal{X}_t)\mathrm{vec}(\mathcal{X}_{t-h})^T$.

This can be more easily seen in the matrix case ($K = 2$). Under the factor model, the observed time series elements can be written as

$$(\boldsymbol{X}_t)_{i_1,i_2,t} = x_{i_1,i_2,t} = \sum_{a_1=1}^{r_1}\sum_{a_2=1}^{r_2} A_{1,i_1,a_1}A_{2,i_2,a_2}\boldsymbol{F}_{a_1,a_2,t}.$$

For $K = 2$, $k = 1$ and a fixed h, TIPUP_1 has elements

$$\begin{aligned}(\mathrm{TIPUP}_{1,h})_{i_1,j_1} &= \left(\sum_{t=h+1}^{T}\frac{\boldsymbol{X}_{t-h}\boldsymbol{X}_t^\top}{T-h}\right)_{i_1,j_1}\\ &= \sum_{t=h+1}^{T}\sum_{i_2=1}^{d_2}\frac{x_{i_1,i_2,t-h}x_{j_1,i_2,t}}{T-h}\\ &= \sum_{t=h+1}^{T}\frac{\mathrm{trace}(\boldsymbol{x}_{i_1,\cdot,t}\boldsymbol{x}_{j_1,\cdot,t}^\top)}{T-h},\end{aligned}$$

where $\boldsymbol{x}_{i_1,\cdot,t} = (x_{i_1,1,t},\ldots,x_{i_1,d_2,t})^\top$. Hence, the (i_1,j_1)-th element of $\mathrm{TIPUP}_{1,h}$ is the trace of the (i_i,j_1)-th block of $(T-h)^{-1}\sum_{t=1}^{T}\mathrm{vec}(\boldsymbol{X}_t)\,\mathrm{vec}^\top(\boldsymbol{X}_{t-h})$, or

$$\mathrm{TIPUP}_{1,h} = \mathrm{PTR}_1(\widehat{\boldsymbol{\Sigma}}_h).$$

In fact for any $0 \le h < T$,

$$\begin{aligned}&(\mathbb{E}[\mathrm{TIPUP}_{1,h}])_{i_1,j_1}\\ &= \sum_{a_1=1}^{r_1}\sum_{b_1=1}^{r_1} A_{1,i_1,a_1}A_{1,j_1,b_1}\Bigg(\sum_{i_2=1}^{d_2}\sum_{a_2=1}^{r_2}\sum_{b_2=1}^{r_2} A_{2,i_2,a_2}A_{2,i_2,b_2}\\ &\quad \mathbb{E}\left[\sum_{t=h+1}^{T}\frac{F_{a_1,a_2,t-h}F_{b_1,b_2,t}}{T-h}\right]\Bigg) + I_{\{h=0\}}\sum_{i_2=1}^{d_2}\sigma_{i_1,i_2,j_1,i_2},\end{aligned}$$

where $\sigma_{i_1,i_2,j_1,j_2} = \mathbb{E}[\varepsilon_{i_1,i_2,t}\varepsilon_{j_1,j_2,t}]$. In the special case considered in TL, where $h = 0$ and

$$\mathbb{E}\left[\sum_{t=1}^{T}F_{a_1,a_2,t}F_{b_1,b_2,t}/T\right] = I\{a_1 = b_1, a_2 = b_2\},$$
$$\sigma_{i_1,i_2,j_1,j_2} = \sigma^2 I\{i_1 = j_1, i_2 = j_2\},$$

we have

$$\begin{aligned}&(\mathbb{E}[\mathrm{TIPUP}_{1,0}])_{i_1,j_1}\\ &= \sum_{a_1=1}^{r_1} A_{1,i_1,a_1}A_{1,j_1,a_1}\left(\sum_{i_2=1}^{d_2}\sum_{a_2=1}^{r_2} A_{2,i_2,a_2}A_{2,i_2,a_2}\right) + d_2\sigma^2 I_{\{i_1=j_1\}}\\ &= \left(\boldsymbol{A}_1\boldsymbol{A}_1^\top\mathrm{trace}(\boldsymbol{A}_2\boldsymbol{A}_2^\top) + d_2\sigma^2\boldsymbol{I}_{d_1}\right)_{i_1,j_1}.\end{aligned}$$

We appreciate TLs calculation which provides a simple way of expressing the population consistency of our methods.

Second, when $h = 0$, $\mathrm{cov}(\mathrm{vec}(\mathcal{F}_t)) = \boldsymbol{I}$ and $\mathrm{cov}(\mathrm{vec}(\mathcal{E}_t)) = \sigma^2\boldsymbol{I}$, TL provides an elegant formula for the estimation of $\boldsymbol{A}_k\boldsymbol{A}_k^T$. For more general $\mathrm{cov}(\mathrm{vec}(\mathcal{F}_t))$ and $\mathrm{cov}(\mathrm{vec}(\mathcal{E}_t))$, we use $h > 1$ to utilize the whiteness assumption of $\mathcal{E}$. In this case, $\mathrm{TIPUP}_{k,h} = \mathrm{PTR}_k(\widehat{\boldsymbol{\Sigma}}_h)$ is in the form of $\boldsymbol{A}_k[\star]\boldsymbol{A}_k^T$ and we use SVD to estimate the column space of $\boldsymbol{A}_k$. Under the assumption of $\mathrm{cov}(\mathrm{vec}(\mathcal{E}_t)) = \sigma^2\boldsymbol{I}$, $\mathrm{TIPUP}_{k,0} = \mathrm{PTR}_k(\widehat{\boldsymbol{\Sigma}}_0)$ is a spike covariance matrix and the column space of $\boldsymbol{A}_k$ can be estimated with SVD as well, especially if $\boldsymbol{A}_k$ is of low rank.

TL also raised the important question on how to incorporate covariates in the tensor factor models. When a covariate is global (having impact on all elements in the tensor), a regression term may be included in the model, and the dimension of its large coefficient tensor may be reduced through certain low rank tensor decomposition assumptions. When a covariate is associated with a specific tensor mode (e.g., GDP of each country in the import–export example), it may be incorporated in the loading matrix by assuming the loading is a function of the covariate, as proposed and studied by Connor, Hagmann, and Linton (2012) and Fan, Liao, and Wang (2016) in the vector factor model case. This is an important research topic as many applications involve covariates. Various model settings, estimation procedures and their theoretical properties all require detailed investigation.

3. Phase Transition of the Convergence Rates

Professors Yuan and Ouyang (YO) carefully studied the characteristics of the two proposed procedures, TIPUP and TOPUP, via analyzing the simplest case of rank-one matrix time series. Their assumptions of rank-one matrix time series model in (1) is a special case of our One-Factor Model in (7) in the paper, where for the ease of proof, YO simplified the model to the matrix case $K = 2$ with equal dimensions $d_1 = d_2$ and rank one $r_1 = r_2 = 1$, and he further assumed that $\mathcal{E}$ has independent standard normal elements.

With our notations and under our more general assumptions with $\sigma = 1$, for the case of rank-one matrix time series, our results, Equations (42) and (49), show that

$$\sin\angle(\hat{u}^{\mathrm{TOPUP}}, u) = O_p\left(\frac{d_1^{1/2}d_2^{1/2}}{\lambda\sqrt{T}} + \frac{d_2 d_1^{1/2}}{\lambda^2\sqrt{T}}\right), \quad (2)$$

$$\sin\angle(\hat{u}^{\mathrm{TIPUP}}, u) = O_p\left(\frac{d_1^{1/2}}{\lambda\sqrt{T}} + \frac{d_1^{1/2}d_2^{1/2}}{\lambda^2\sqrt{T}}\right). \quad (3)$$

In our notation, YO's result in their Proposition 1 can be written as follows,

$$\sin\angle(\hat{u}^{\mathrm{TOPUP}}, u) = O_p\left(\frac{d_1^{1/2}}{\lambda\sqrt{T}} + \frac{d_1^{1/2}}{\lambda^2\sqrt{T}}\right), \quad (4)$$

$$\sin\angle(\hat{u}^{\mathrm{TIPUP}}, u) = O_p\left(\frac{d_1^{1/2}}{\lambda\sqrt{T}} + \frac{d_1^{1/2}d_2^{1/2}}{\lambda^2\sqrt{T}}\right). \quad (5)$$

Note that $d = \prod_{k=1}^{K} d_k$ in our paper and $d = d_1 = d_2$ in YO's discussion. Equation (4) holds under the extra condition on signal-to-noise ratio in their Proposition 1 that

$$\lambda \gg d_1^{1/2}d_2/T^{1/2} + d_1^{1/4}d_2^{1/2}/T^{1/4} + d_1^{1/6}d_2^{2/3}/T^{1/6} \quad (6)$$

Equation (2) holds under no condition, but it is only meaningful when

$$\lambda \gg (d_1 d_2)^{1/2}/T^{1/2} + d_1^{1/4}d_2^{1/2}/T^{1/4}.$$

Comparing these equations, the rate obtained for TIPUP by YO and us are the same, as seen in (3) and (5). YO's rate for TOPUP (4) is faster than ours (2) under stronger condition on the signal-to-noise ratio. Therefore, our theoretical rates, (2) and (3), show that the worst case error bound for TOPUP is slower than that of TIPUP in the more general regime, while YO's results, (4) and (5) show the contrary when signal is stronger. Furthermore, YO's simulation results in Figure 1 in their discussion corroborate their theoretical claim that TOPUP is faster than TIPUP.

Although YO only proved Proposition 1 under the case of $K = 2$ with $d_1 = d_2$, $r_1 = r_2 = 1$ and iid Gaussian noise, their strategy sheds light and can improve our theoretical rates in more general settings. The thrust of this argument can be found in the original proof of Wedin (1972), which leads to the following lemma.

Lemma 1. Let M be a matrix of rank r with the smallest nonzero singular value λ_r, U and V be, respectively, the left and right singular matrices of M associated with the nonzero singular values, and $U_\perp$ and $V_\perp$ be the orthonormal complements of U and V. Let $\hat{M} = M + \Delta$ be a noisy version of M, and $\hat{U}$ and $\hat{V}$ be, respectively, the left and right singular matrices associated with the top r singular values of $\hat{M}$. Let $\|\cdot\|$ be a matrix norm satisfying $\|ABC\| \le \|A\|_{\mathrm{S}}\|C\|_{\mathrm{S}}\|B\|$, $\epsilon_1 = \|\Delta\|_{\mathrm{S}}$, $\epsilon_1' = \|U^\top\Delta V_\perp\|$ and $\epsilon_2 = \|U_\perp^\top\Delta V\|$. Then,

$$\|U_\perp^\top\hat{U}\| \le \frac{\epsilon_1'\epsilon_1 + \epsilon_2(\lambda_r - \epsilon_1)}{(\lambda_r - \epsilon_1)^2 - \epsilon_1^2}. \quad (7)$$

In particular, for the spectral norm $\|\cdot\| = \|\cdot\|_{\mathrm{S}}$ and $\mathrm{error}_j = \epsilon_j/\lambda_r$ satisfying $\mathrm{error}_1 = o(1)$,

$$\left\|\hat{U}\hat{U}^\top - UU^\top\right\|_{\mathrm{S}} = (1 + o(1))\left(\mathrm{error}_2 + \mathrm{error}_1^2\right). \quad (8)$$

Here is an explanation. Let $\lambda_r(A)$ be the r-th singular value of matrix A. Wedin (1972) used the decomposition $U_\perp^\top\hat{U}\hat{U}^\top\hat{M}\hat{V} = U_\perp^\top\hat{M}\hat{V} = U_\perp^\top M V_\perp V_\perp^\top\hat{V} + U_\perp^\top\Delta\hat{V}$ and its transposed version to obtain two constraints on the errors $\|U_\perp^\top\hat{U}\|$ and $\|V_\perp^\top\hat{V}\|$ and solved the linear programming by taking the maximum of both sides of the two constraints. In our case, we are interested in upper bounds in terms of $U^\top\Delta V_\perp$ and $U_\perp^\top\Delta V$, so that a slightly different decomposition $U_\perp^\top\hat{U}\hat{U}^\top\hat{M}\hat{V} = U_\perp^\top\hat{M}V_\perp V_\perp^\top\hat{V} + U_\perp^\top\Delta VV^\top\hat{V}$ and its transposed version are used to obtain the following constraints on the errors,

$$\|U_\perp^\top\hat{U}\| \le \left(\|U_\perp^\top\hat{M}V_\perp\|_{\mathrm{S}}\|V_\perp^\top\hat{V}\| + \|U_\perp^\top\Delta V\|\right)/\lambda_r(\hat{U}^\top\hat{M}\hat{V}),$$
$$\|V_\perp^\top\hat{V}\| \le \left(\|U_\perp^\top\hat{M}V_\perp\|_{\mathrm{S}}\|U_\perp^\top\hat{U}\| + \|V_\perp^\top\Delta^\top U\|\right)/\lambda_r(\hat{U}^\top\hat{M}\hat{V}).$$

The lemma follows by inserting the second inequality into the first.

It is clear that (8) is sharper than the bound $(1 + o(1))\mathrm{error}_1$ from a direct application of Wedin's inequality when $\mathrm{error}_2 \ll \mathrm{error}_1 = o(1)$, but the condition for consistency is still $\mathrm{error}_1 = o(1)$.

Lemma 1 is a variation of Theorem 1 of Cai and Zhang (2018) but it is in a form directly relevant to YO's discussion and its simple derivation comes directly from Wedin's proof.

Note that our original Corollary 1 essentially proved that with nondiverging ranks, $\mathrm{error}_1 = O_p\big(d_1^{1/2}d_2^{1/2}/(\lambda\sqrt{T}) + d_2 d_1^{1/2}/(\lambda^2\sqrt{T})\big)$, which is the rate in (2). Using the same proof technique as to bound error_1 and set $V = \odot_{j\in[2K]\setminus\{k\}}U_j$ in Lemma 1, one can show that $\mathrm{error}_2 = O_p\big(d_1^{1/2}/(\lambda\sqrt{T}) + d_1^{1/2}/(\lambda^2\sqrt{T})\big)$, which is the rate in (4). The proof in bounding error_2 works for tensor case with general ranks and general covariance structure of the noise tensor. In summary, in the fixed rank model

$$\sin\angle(\hat{u}^{\mathrm{TOPUP}}, u) = \mathrm{error}_2 + \mathrm{error}_1^2 \quad (9)$$
$$= O_p\left(\left(\frac{d_1^{1/2}}{\lambda\sqrt{T}} + \frac{d_1^{1/2}}{\lambda^2\sqrt{T}}\right) + \left(\frac{d_1^{1/2}d_2^{1/2}}{\lambda\sqrt{T}} + \frac{d_2 d_1^{1/2}}{\lambda^2\sqrt{T}}\right)^2\right),$$

under no extra signal-to-noise ratio condition. It can be seen that $\mathrm{error}_2 \ll \mathrm{error}_1$ in (9), whose fundamental reason embedded in the proof is that when a special low-rank Kronecker product V is multiplied with the noise Δ, the high dimension $d_1 \times (d_2 d_1 d_2)$ is reduced to $d_1 \times (r_2 r_1 r_2)$ due to the nature of the outer-product of TOPUP. When $\mathrm{error}_1 = o_p(1)$, we have $\mathrm{error}_1^2 \ll \mathrm{error}_1$. Therefore, when $\mathrm{error}_1 = o_p(1)$, such strategy proves a faster convergence rate of TOPUP than our previous result in the general regime. When $\mathrm{error}_1^2 \ll \mathrm{error}_2$, which is reflected in YO's signal-to-noise ratio requirement (6), the rate is dominated by the term error_2, which is YO's convergence rate (4).

We need to comment that such strategy is not applicable to TIPUP, as for TIPUP $\text{error}_2 \asymp \text{error}_1$, whose fundamental reason is that the inner product is taken before SVD and the dimension reduction from $d_1 \times d_1$ to $d_1 \times r_1$ does not reduce the spectrum norm of the noise when multiplying the $d_1 \times d_1$ noise matrix for TIPUP with a V of rank r_1.

Now comparing the TOPUP rate (9) with the TIPUP rate (3), it is not guaranteed that TOPUP is faster than TIPUP, which only happens when the second term in (9), that is, error_1^2, is less than the TIPUP rate.

To verify these findings, we perform further simulation study with similar but broader setting than YO's. To be specific, we simulate data from

$$X_t = \lambda u v^\top f_t + E_t,$$

where $d_1 = d_2 = 16$, $T = 256$ or 1024, $\|u\|_2 = \|v\|_2 = 1$, f_t follows an AR(1) model with AR coefficient $\phi = 0.5$ and standard normal innovations. The noise E_t is white, following matrix normal distribution $N(0, \Psi, \Psi)$, where Ψ has diagonal entries being 1 and off diagonal entries being ρ. We varied the correlation ρ among 0, 0.05, and 0.1. Figures 1–3 present estimation errors corresponding to the different ρ's, respectively, under various T, λ and h_0 combinations. In Figure 1, $\rho = 0$, which satisfies YO's assumption, it can be seen that TOPUP indeed behaves better than TIPUP. In Figure 3, $\rho = 0.1$, which is relatively far from the independence assumption, TIPUP uniformly dominates TOPUP. In Figure 2, $\rho = 0.05$, with the intermediate choice of correlation, TIPUP is better than TOPUP under weaker signal with smaller λ while TOPUP performs

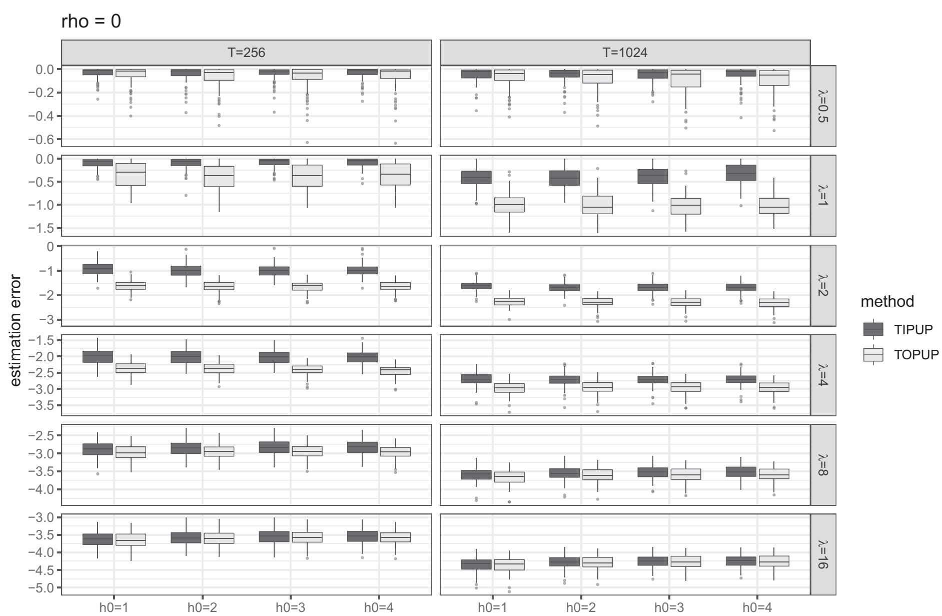

Figure 1. Simulation result for $\rho = 0$. The boxplot of the logarithm of the estimation errors from TIPUP and TOPUP with multiple choices of h_0 for the column space of A_1, that is, u. Two columns correspond to different sample sizes T and six rows correspond to different signal strengths λ.

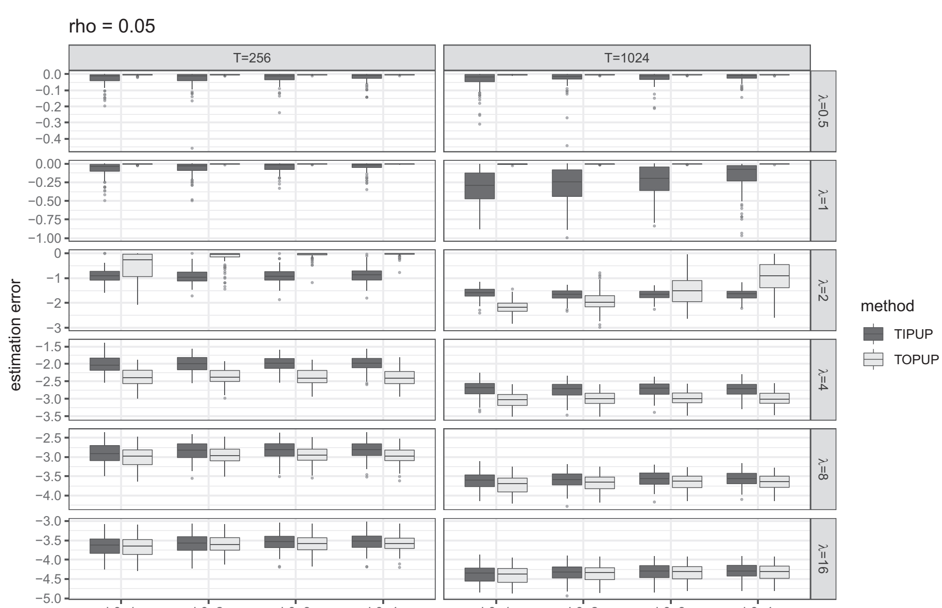

Figure 2. Simulation result for $\rho = 0.05$. The boxplot of the logarithm of the estimation errors from TIPUP and TOPUP with multiple choices of h_0 for the column space of A_1, that is, u. Two columns correspond to different sample sizes T and six rows correspond to different signal strengths λ.

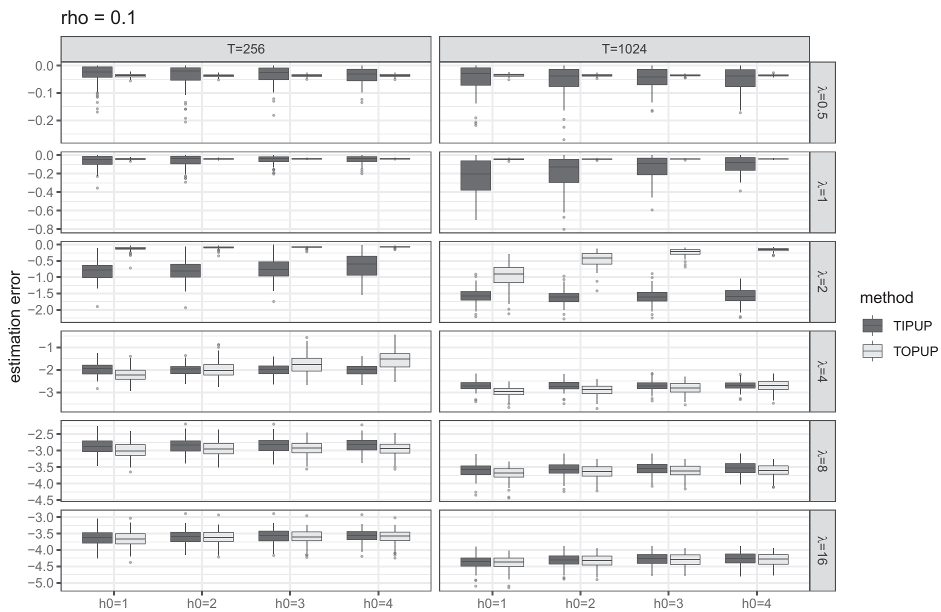

Figure 3. Simulation result for $\rho = 0.1$. The boxplot of the logarithm of the estimation errors from TIPUP and TOPUP with multiple choices of h_0 for the column space of A_1, that is, u. Two columns correspond to different sample sizes T and six rows correspond to different signal strengths λ.

better under stronger signal with larger λ. So in practice, neither TIPUP nor TOPUP dominates the other because their relative finite sample performance depends on the signal-to-noise ratio, whether error_1^2 is less than the rate of TIPUP, and the constants in front of the rates of convergence, which may become important in finite sample situations.

Indeed, we appreciate YO's results, including their enlightening proof and interesting simulations. Their analysis in the simplified setting helped us to obtain a faster convergence rate for TOPUP in our more general settings and establish a better understanding of TOPUP.

YO also provided an AR(1) example with detailed analysis on the impact of h_0 and shows it is an important aspect in practice. We totally agree with their comment. A larger h_0 increases the signal strength, but also adds more noise in the TOPUP and TIPUP statistics. Developing a data-driven approach to determine the optimal h_0 is indeed useful and important. As h_0 does not change model complexity, a simple objective function such as residual sum of squares or certain statistics measuring the "whiteness" in the residuals can be used.

ORCID

Rong Chen http://orcid.org/0000-0003-0793-4546
Dan Yang http://orcid.org/0000-0001-9490-7284

References

Cai, T. T., and Zhang, A. (2018), "Rate-Optimal Perturbation Bounds for Singular Subspaces with Applications to High-Dimensional Statistics," *The Annals of Statistics*, 46, 60–89. [130]

Connor, G., Hagmann, M., and Linton, O. (2012), "Efficient Semiparametric Estimation of the Fama-French Model and Extensions," *Econometrica*, 80, 713–754. [129]

Fan, J., Liao, Y., and Wang, W. (2016), "Projected Principal Component Analysis in Factor Models," *The Annals of Statistics*, 44, 219–254. [129]

Wedin, P.-Å. (1972), "Perturbation Bounds in Connection with Singular Value Decomposition," *BIT Numerical Mathematics*, 12, 99–111. [130]