

Supplemental Material

Factor Models for High-Dimensional Tensor Time Series

Rong Chen, Dan Yang and Cun-Hui Zhang

Rutgers University and The University of Hong Kong

Appendix A: Proofs

The following lemma is needed in both the proofs of Theorems 1 and 2.

Lemma 2. (i) Let $\mathbf{G} \in \mathbb{R}^{d_1 \times n}$ and $\mathbf{H} \in \mathbb{R}^{d_2 \times n}$ be two centered independent Gaussian matrices such that $\mathbb{E}(\mathbf{u}^\top \text{vec}(\mathbf{G}))^2 \leq \sigma^2 \forall \mathbf{u} \in \mathbb{R}^{d_1 n}$ and $\mathbb{E}(\mathbf{v}^\top \text{vec}(\mathbf{H}))^2 \leq \sigma^2 \forall \mathbf{v} \in \mathbb{R}^{d_2 n}$. Then,

$$\mathbb{E}[\|\mathbf{GH}^\top\|_S] \leq \sigma^2(\sqrt{d_1 d_2} + \sqrt{d_1 n} + \sqrt{d_2 n}).$$

(ii) Let $\mathbf{G}_i \in \mathbb{R}^{d_1 \times d_2}$, $\mathbf{H}_i \in \mathbb{R}^{d_3 \times d_4}$, $i = 1, \dots, n$, be independent centered Gaussian matrices such that $\mathbb{E}(\mathbf{u}^\top \text{vec}(\mathbf{G}_i))^2 \leq \sigma^2 \forall \mathbf{u} \in \mathbb{R}^{d_1 d_2}$ and $\mathbb{E}(\mathbf{v}^\top \text{vec}(\mathbf{H}_i))^2 \leq \sigma^2 \forall \mathbf{v} \in \mathbb{R}^{d_3 d_4}$. Then,

$$\mathbb{E}\left[\left\|\text{mat}_1\left(\sum_{i=1}^n \mathbf{G}_i \otimes \mathbf{H}_i\right)\right\|_S\right] \leq \sigma^2(\sqrt{d_1 n} + \sqrt{d_1 d_3 d_4} + \sqrt{nd_2 d_3 d_4}).$$

Proof. Assume $\sigma = 1$ without loss of generality.

(i) Independent of $\mathbf{G}$ and $\mathbf{H}$, let $\xi_j \in \mathbb{R}^{d_j}$ and $\zeta_j \in \mathbb{R}^n$, $j = 1, 2$, be independent standard Gaussian vectors. For $\|\mathbf{v}_1\|_2 = \|\mathbf{v}_2\|_2 = 1$ and $\|\mathbf{w}_1\|_2 \vee \|\mathbf{w}_2\|_2 \leq 1$, $\mathbb{E}(\mathbf{v}_1^\top \mathbf{H} \mathbf{w}_1 - \mathbf{v}_2^\top \mathbf{H} \mathbf{w}_2)^2 \leq \mathbb{E}((\mathbf{v}_1 - \mathbf{v}_2)^\top \xi_2 + (\mathbf{w}_1 - \mathbf{w}_2)^\top \zeta_2)^2$. Thus, by the Sudakov-Fernique inequality

$$\begin{aligned} \mathbb{E}[\|\mathbf{GH}^\top\|_S | \mathbf{G}] &\leq \|\mathbf{G}\|_S \mathbb{E}\left[\max_{\|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1} (\|\mathbf{G}\|_S^{-1} \mathbf{u}^\top \mathbf{G}) \mathbf{H}^\top \mathbf{v} \mid \mathbf{G}\right] \\ &\leq \|\mathbf{G}\|_S \mathbb{E}\left[\max_{\|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1} (\|\mathbf{G}\|_S^{-1} \mathbf{u}^\top \mathbf{G} \zeta_2 + \xi_2^\top \mathbf{v}) \mid \mathbf{G}\right] \\ &= \mathbb{E}\left[\max_{\|\mathbf{u}\|_2 = 1} \mathbf{u}^\top \mathbf{G} \zeta_2 \mid \mathbf{G}\right] + \|\mathbf{G}\|_S \sqrt{d_2}. \end{aligned}$$

Applying the same argument to $\mathbf{G}$, we have

$$\begin{aligned} \mathbb{E}[\|\mathbf{GH}^\top\|_S] &\leq \mathbb{E}\left[\max_{\|\mathbf{u}\|_2 = 1} \mathbf{u}^\top \mathbf{G} \zeta_2 + \|\mathbf{G}\|_S \sqrt{d_2}\right] \\ &\leq \mathbb{E}\left[\|\zeta_2\|_2 \max_{\|\mathbf{u}\|_2 = 1} (\mathbf{u}^\top \xi_1 + \zeta_1^\top \zeta_2 / \|\zeta_2\|_2) + (\sqrt{d_1} + \sqrt{n}) \sqrt{d_2}\right] \\ &\leq \sqrt{d_1 n} + (\sqrt{d_1} + \sqrt{n}) \sqrt{d_2}. \end{aligned}$$

(ii) We treat $(\mathbf{G}_1, \dots, \mathbf{G}_n) \in \mathbb{R}^{d_1 \times d_2 \times n}$ and $(\mathbf{H}_1, \dots, \mathbf{H}_n) \in \mathbb{R}^{d_3 \times d_4 \times n}$ as tensors. Let $\xi_i \in \mathbb{R}^{d_2}$ be additional independent standard Gaussian vectors. For $\mathbf{u} \in \mathbb{R}^{d_1}$ and $\mathbf{V} \in \mathbb{R}^{d_2 \times (d_3 d_4)}$,

$$\mathbb{E}\left[\left\|\text{mat}_1\left(\sum_{i=1}^n \mathbf{G}_i \otimes \mathbf{H}_i\right)\right\|_S\right]$$

$$\begin{aligned}
&= \mathbb{E} \left[\sup_{\|\mathbf{u}\|_2=1, \|\mathbf{V}\|_F=1} \mathbf{u}^\top \text{mat}_1(\mathbf{G}_1, \dots, \mathbf{G}_n) \text{vec}(\text{mat}_3(\mathbf{H}_1, \dots, \mathbf{H}_n) \mathbf{V}^\top) \right] \\
&\leq \sqrt{d_1} \mathbb{E} \left[\sup_{\|\mathbf{V}\|_F=1} \|\text{mat}_3(\mathbf{H}_1, \dots, \mathbf{H}_n) \mathbf{V}^\top\|_F \right] \\
&\quad + \mathbb{E} \left[\sup_{\|\mathbf{V}\|_F=1} (\text{vec}(\xi_1, \dots, \xi_n))^\top \text{vec}(\text{mat}_3(\mathbf{H}_1, \dots, \mathbf{H}_n) \mathbf{V}^\top) \right] \\
&= \sqrt{d_1} \mathbb{E} \left[\|\text{mat}_3(\mathbf{H}_1, \dots, \mathbf{H}_n)\|_S \right] + \mathbb{E} \left[\left(\sum_{j=1}^{d_2} \sum_{k=1}^{d_3 d_4} \left(\sum_{i=1}^n \xi_{i,j} \text{vec}(\mathbf{H}_i)_k \right)^2 \right)^{1/2} \right] \\
&\leq \sqrt{d_1} (\sqrt{n} + \sqrt{d_3 d_4}) + \sqrt{n d_2 d_3 d_4}.
\end{aligned}$$

Again, we apply the Sudakov-Fernique inequality twice above. $\square$

We first prove Theorem 2 as the analysis is simpler and facilitates the proof of Theorem 1.

Proof of Theorem 2. It suffices to consider $k = 1$ and $K = 2$ as the TIPUP (19) begins with mode- k matrix unfolding in (18). We observe $\mathbf{X}_t = \mathbf{M}_t + \mathbf{E}_t \in \mathbb{R}^{d_1 \times d_2}$ with $\mathbf{M}_t = \mathbf{A}_1 \mathbf{F}_t \mathbf{A}_2^\top$. With the $\Phi_{k,h}^{(\text{cano})*}$ in (32), we write (18) and its factor process version as $d_1 \times d_1$ matrices

$$\mathbf{V}_{1,h}^* = \sum_{t=h+1}^T \frac{\mathbf{X}_{t-h} \mathbf{X}_t^\top}{T-h}, \quad \boldsymbol{\Theta}_{1,h}^* = \sum_{t=h+1}^T \frac{\mathbf{M}_{t-h} \mathbf{M}_t^\top}{T-h} = \lambda^2 \mathbf{U}_1 \Phi_{1,h}^{(\text{cano})*} \mathbf{U}_1^\top.$$

Let $\boldsymbol{\Delta}_1^*$, $\boldsymbol{\Delta}_2^*$ and $\boldsymbol{\Delta}_3^*$ be respectively the three terms on the right-hand side below:

$$\mathbf{V}_{1,h}^* - \boldsymbol{\Theta}_{1,h}^* = \sum_{t=h+1}^T \frac{\mathbf{M}_{t-h} \mathbf{E}_t^\top}{T-h} + \sum_{t=h+1}^T \frac{\mathbf{E}_{t-h} \mathbf{M}_t^\top}{T-h} + \sum_{t=h+1}^T \frac{\mathbf{E}_{t-h} \mathbf{E}_t^\top}{T-h}. \quad (77)$$

Let $\sigma_* = \sigma \|\boldsymbol{\Theta}_{1,0}^*\|_S^{1/2} = \sigma \lambda \|\Phi_{1,0}^{(\text{cano})*}\|_S^{1/2}$. By Condition A, for any $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^{d_1}$ we have

$$\mathbb{E} \left\{ \mathbf{u}^\top \left(\sum_{t=h+1}^T \frac{\mathbf{M}_{t-h} \mathbf{E}_t^\top}{T^{1/2}} \right) \mathbf{v} \right\}^2 \leq \frac{\sigma^2 \|\mathbf{v}\|_2^2}{T} \left\| \sum_{t=h+1}^T \mathbf{M}_{t-h}^\top \mathbf{u} \right\|_2^2 \leq \sigma^2 \|\boldsymbol{\Theta}_{1,0}^*\|_S \|\mathbf{u}\|_2^2 \|\mathbf{v}\|_2^2.$$

Thus for vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ with $\|\mathbf{u}_i\|_2 = \|\mathbf{v}_i\|_2 = 1$,

$$\begin{aligned}
&\{(T-h)^2 / (T\sigma_*^2)\} \mathbb{E} (\mathbf{u}_1^\top \boldsymbol{\Delta}_1^* \mathbf{v}_1 - \mathbf{u}_2^\top \boldsymbol{\Delta}_1^* \mathbf{v}_2)^2 \\
&\leq \left(\|\mathbf{u}_1 - \mathbf{u}_2\|_2 \|\mathbf{v}_1\|_2 + \|\mathbf{u}_2\|_2 \|\mathbf{v}_1 - \mathbf{v}_2\|_2 \right)^2 \\
&\leq 2 \left(\|\mathbf{u}_1 - \mathbf{u}_2\|_2^2 + \|\mathbf{v}_1 - \mathbf{v}_2\|_2^2 \right) \\
&= 2 \mathbb{E} \{ (\mathbf{u}_1 - \mathbf{u}_2)^\top \boldsymbol{\xi} + (\mathbf{v}_1 - \mathbf{v}_2)^\top \boldsymbol{\zeta} \}^2
\end{aligned}$$

where $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ are iid $N(0, \mathbf{I}_{d_1})$ vectors. As $\boldsymbol{\Delta}_1^*$ is a $d_1 \times d_1$ Gaussian matrix under $\mathbb{E}$, the Sudakov-Fernique inequality yields

$$\left(\frac{T-h}{T^{1/2} \sigma_*} \right) \mathbb{E} \|\boldsymbol{\Delta}_1^*\|_S \leq \sqrt{2} \mathbb{E} \sup_{\|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1} |\mathbf{u}^\top \boldsymbol{\xi} + \mathbf{v}^\top \boldsymbol{\zeta}| = \sqrt{2} \mathbb{E} (\|\boldsymbol{\xi}\|_2 + \|\boldsymbol{\zeta}\|_2).$$

As $\mathbb{E}\|\boldsymbol{\xi}\|_2 = \mathbb{E}\|\boldsymbol{\zeta}\|_2 \leq \sqrt{d_1}$, it follows that for the first term on the right-hand side of (77)

$$\overline{\mathbb{E}}\|\boldsymbol{\Delta}_1^*\|_S \leq \frac{\sigma_*(8Td_1)^{1/2}}{T-h} = \frac{\sigma\lambda(8Td_1)^{1/2}}{T-h} \|\boldsymbol{\Phi}_{1,0}^{(\text{cano})*}\|_S^{1/2}. \quad (78)$$

Similarly, $\overline{\mathbb{E}}\|\boldsymbol{\Delta}_2^*\|_S \leq \sigma\lambda(8Td_1)^{1/2}(T-h)^{-1} \|\boldsymbol{\Phi}_{1,0}^{(\text{cano})*}\|_S^{1/2}$ for the second term.

Let $h \leq T/4$. For the third term $\boldsymbol{\Delta}_3^*$ on the right-hand side of (77), we split the sum into two terms over the index sets, $S_1 = \{(h, 2h] \cup (3h, 4h] \cup \dots\} \cap (h, T]$ and its complement S_2 in $(h, T]$, so that $(\mathbf{E}_{t-h}, t \in S_a)$ and $(\mathbf{E}_t, t \in S_a)$ are two independent $d_1 \times n_a$ centered Gaussian matrices for each $a = 1, 2$, with $n_1 + n_2 = (T-h)d_2$. Thus, by Lemma 2 (i),

$$\frac{\overline{\mathbb{E}}\|\boldsymbol{\Delta}_3^*\|_S}{\sigma^2} \leq \sum_{a=1}^2 \overline{\mathbb{E}} \left\| \sum_{t \in S_a} \frac{\mathbf{E}_{t-h} \mathbf{E}_t^\top}{\sigma^2(T-h)} \right\|_S \leq \sum_{a=1}^2 \frac{2\sqrt{d_1 n_a} + d_1}{T-h} \leq \frac{\sqrt{8d_1 d_2}}{\sqrt{T-h}} + \frac{2d_1}{T-h}. \quad (79)$$

By applying (78) and (79) to (77), we find that

$$\overline{\mathbb{E}} \left[\left\| \mathbf{V}_{1,h}^* / \lambda^2 - \mathbf{U}_1 \boldsymbol{\Phi}_{1,h}^{(\text{cano})*} \mathbf{U}_1^\top \right\|_S \right] \leq \frac{2\sigma(8Td_1)^{1/2}}{\lambda(T-h)} \|\boldsymbol{\Phi}_{1,0}^{(\text{cano})*}\|_S^{1/2} + \frac{\sigma^2}{\lambda^2} \left(\frac{\sqrt{8d_1 d_2}}{\sqrt{T-h}} + \frac{2d_1}{T-h} \right) = \Delta_{1,h}^*$$

with the $\Delta_{1,h}^*$ in (43). This is the first inequality in (44). The second inequality in (44) follows by Cauchy-Schwarz, $\|(\boldsymbol{\Delta}_1, \dots, \boldsymbol{\Delta}_{h_0})\|_S^2 \leq h_0 \sum_{h=1}^{h_0} \|\boldsymbol{\Delta}_h\|_S^2$. Finally, (45) follows from (44) with an application of the Wedin (1972) inequality.

Proof of Theorem 1. Again, it suffices to consider $k = 1$ and $K = 2$ as the TOPUP begins with mode- k matrix unfolding in (10). In this case, $\mathbf{X}_t = \mathbf{M}_t + \mathbf{E}_t \in \mathbb{R}^{d_1 \times d_2}$ with $\mathbf{M}_t = \mathbf{A}_1 \mathbf{F}_t \mathbf{A}_2^\top$, and $\mathbf{V}_{1,h}$ and its conditional expectation $\boldsymbol{\Theta}_{1,h} = \overline{\mathbb{E}}[\mathbf{V}_{1,h}]$ are given by

$$\mathbf{V}_{1,h} = \sum_{t=h+1}^T \frac{\mathbf{X}_{t-h} \otimes \mathbf{X}_t}{T-h}, \quad \boldsymbol{\Theta}_{1,h} = \sum_{t=h+1}^T \frac{\mathbf{M}_{t-h} \otimes \mathbf{M}_t}{T-h}.$$

Let $\boldsymbol{\Delta}_1$, $\boldsymbol{\Delta}_2$ and $\boldsymbol{\Delta}_3$ be respectively the three terms on the right-hand side below:

$$\begin{aligned} & \text{mat}_1(\mathbf{V}_{1,h}) - \lambda^2 \mathbf{U}_k \boldsymbol{\Phi}_{k,h}^{(\text{cano})} \mathbf{U}_{[2k] \setminus \{k\}}^\top \\ &= \sum_{t=h+1}^T \frac{\text{mat}_1(\mathbf{M}_{t-h} \otimes \mathbf{E}_t)}{T-h} + \sum_{t=h+1}^T \frac{\text{mat}_1(\mathbf{E}_{t-h} \otimes \mathbf{M}_t)}{T-h} + \sum_{t=h+1}^T \frac{\text{mat}_1(\mathbf{E}_{t-h} \otimes \mathbf{E}_t)}{T-h}. \end{aligned} \quad (80)$$

because $\text{mat}_1(\boldsymbol{\Theta}_{1,h}) = \lambda^2 \mathbf{U}_k \boldsymbol{\Phi}_{k,h}^{(\text{cano})} \mathbf{U}_{[2k] \setminus \{k\}}^\top$ as in (29) and (30).

For the first term $\boldsymbol{\Delta}_1$, we notice that $\mathbf{M}_{t-h} = \mathbf{M}_{t-h} \mathbf{P}_2$ for a fixed orthogonal projection of rank r_2 . Let $\mathbf{U}_2 \in \mathbb{R}^{d_2 \times r_2}$ with orthonormal columns and $\mathbf{U}_2 \mathbf{U}_2^\top = \mathbf{P}_2$. For $\mathcal{V} \in \mathbb{R}^{d_2 \times d_1 \times d_2}$, $\text{mat}_1(\mathbf{M}_{t-h} \otimes \mathbf{E}_t) \text{vec}(\mathcal{V}) = \text{mat}_1((\mathbf{M}_{t-h} \mathbf{U}_2) \otimes \mathbf{E}_t) \text{vec}(\mathcal{W})$ with $\mathcal{W} = \mathcal{V} \times_1 \mathbf{U}_2^\top \in \mathbb{R}^{r_2 \times d_1 \times d_2}$ satisfying $\|\text{vec}(\mathcal{W})\|_2 = \|\text{vec}(\mathcal{V})\|_2$, so that $\|\boldsymbol{\Delta}_1\|_S = \|\overline{\boldsymbol{\Delta}}_1\|_S$ with

$$\overline{\boldsymbol{\Delta}}_1 = \sum_{t=h+1}^T \frac{\text{mat}_1((\mathbf{M}_{t-h} \mathbf{U}_2) \otimes \mathbf{E}_t)}{T-h}.$$

By Condition A, for any $\mathbf{u} \in \mathbb{R}^{d_1}$ and $\mathcal{W} \in \mathbb{R}^{r_2 \times d_1 \times d_2}$

$$\begin{aligned}
& \mathbb{E} \left(T^{-1/2} \sum_t \mathbf{u}^T \text{mat}_1((\mathbf{M}_{t-h} \mathbf{U}_2) \otimes \mathbf{E}_t) \text{vec}(\mathcal{W}) \right)^2 \\
&= T^{-1} \sum_t \mathbb{E} \left(\sum_{i_1, j_1, i_2, j_2} u_{i_1} (\mathbf{M}_{t-h} \mathbf{U}_2)_{i_1, j_1} (\mathbf{E}_t)_{i_2, j_2} w_{j_1, i_2, j_2} \right)^2 \\
&\leq T^{-1} \sum_t \sigma^2 \sum_{i_2, j_2} \left(\sum_{i_1, j_1} u_{i_1} (\mathbf{M}_{t-h} \mathbf{U}_2)_{i_1, j_1} w_{j_1, i_2, j_2} \right)^2 \\
&\leq \sigma^2 T^{-1} \sum_t \sum_{i_2, j_2} \sum_{j_1} \left(\sum_{i_1} u_{i_1} (\mathbf{M}_{t-h} \mathbf{U}_2)_{i_1, j_1} \right)^2 \sum_{j_1} w_{j_1, i_2, j_2}^2 \\
&= \sigma^2 T^{-1} \sum_t \|\mathbf{M}_{t-h}^\top \mathbf{u}\|_2^2 \|\text{vec}(\mathcal{W})\|_2^2 \\
&= \sigma^2 \mathbf{u}^\top \boldsymbol{\Theta}_{1,0}^* \mathbf{u} \|\text{vec}(\mathcal{W})\|_2^2 \\
&\leq \sigma^2 \lambda^2 \|\boldsymbol{\Phi}_{1,0}^{(\text{cano})*}\|_S \|\mathbf{u}\|_2^2 \|\text{vec}(\mathcal{W})\|_2^2.
\end{aligned}$$

As in the derivation of (78), it follows that for $\|\mathbf{u}_i\|_2 = \|\text{vec}(\mathcal{W}_i)\|_2 = 1$,

$$\{(T-h)^2/(T\sigma_*^2)\} \mathbb{E} \{ \mathbf{u}_1^\top \overline{\boldsymbol{\Delta}}_1 \text{vec}(\mathcal{W}_1) - \mathbf{u}_2^\top \overline{\boldsymbol{\Delta}}_1 \text{vec}(\mathcal{W}_2) \}^2 \leq 2 \{ \|\mathbf{u}_1 - \mathbf{u}_2\|_2^2 + \|\text{vec}(\mathcal{W}_1 - \mathcal{W}_2)\|_2^2 \}.$$

As $\overline{\boldsymbol{\Delta}}_1$ is a Gaussian matrix under $\mathbb{E}$, the Sudakov-Fernique inequality yields

$$\mathbb{E} \|\boldsymbol{\Delta}_1\|_S = \mathbb{E} \|\overline{\boldsymbol{\Delta}}_1\|_S \leq \frac{\sigma \lambda (2T)^{1/2} (\sqrt{d_1} + \sqrt{r_2 d_1 d_2})}{T-h} \|\boldsymbol{\Phi}_{1,0}^{(\text{cano})*}\|_S^{1/2}. \quad (81)$$

For the second term $\boldsymbol{\Delta}_2$, $\text{mat}_1(\mathbf{E}_{t-h} \otimes \mathbf{M}_t) \text{vec}(\mathcal{V}) = \text{mat}_1(\mathbf{E}_{t-h} \otimes (\mathbf{U}_1^\top \mathbf{M}_t \mathbf{U}_2)) \text{vec}(\mathcal{W})$ with $\mathbf{U}_j \mathbf{U}_j^\top = \mathbf{P}_j$ and $\mathcal{W} = \mathcal{V} \times_2 \mathbf{U}_1^\top \times_3 \mathbf{U}_2^\top \in \mathbb{R}^{d_2 \times r_1 \times r_2}$, so that $\|\boldsymbol{\Delta}_2\|_S = \|\overline{\boldsymbol{\Delta}}_2\|_S$ with

$$\overline{\boldsymbol{\Delta}}_2 = \sum_{t=h+1}^T \frac{\text{mat}_1(\mathbf{E}_{t-h} \otimes (\mathbf{U}_1^\top \mathbf{M}_t \mathbf{U}_2))}{T-h} \in \mathbb{R}^{d_1 \times d_2 r_1 r_2}.$$

Moreover, as $\mathbf{F}_t^{(\text{cano})} = \lambda^{-1} \mathbf{U}_1^\top \mathbf{M}_t \mathbf{U}_2$ (6), for $\mathbf{u} \in \mathbb{R}^{d_1}$, $\mathcal{W} \in \mathbb{R}^{d_2 \times r_1 \times r_2}$,

$$\begin{aligned}
& \mathbb{E} \left(T^{-1/2} \sum_t \mathbf{u}^T \text{mat}_1(\mathbf{E}_{t-h} \otimes (\mathbf{U}_1^\top \mathbf{M}_t \mathbf{U}_2)) \text{vec}(\mathcal{W}) \right)^2 \\
&= T^{-1} \sum_t \mathbb{E} \left(\sum_{i_1, j_1, i_2, j_2} u_{i_1} (\mathbf{E}_{t-h})_{i_1, j_2} (\lambda \mathbf{F}_t^{(\text{cano})})_{i_2, j_2} w_{j_1, i_2, j_2} \right)^2 \\
&\leq (\lambda \sigma)^2 T^{-1} \sum_t \sum_{i_1, j_1} \left(\sum_{i_2, j_2} u_{i_1} (\mathbf{F}_t^{(\text{cano})})_{i_2, j_2} w_{j_1, i_2, j_2} \right)^2 \\
&= (\lambda \sigma)^2 \|\mathbf{u}\|_2^2 \sum_{j_1} T^{-1} \sum_t \text{trace} \left((\mathbf{F}_t^{(\text{cano})}) \mathcal{W}_{j_1}^{(2,3)} \right)^2 \\
&= (\lambda \sigma)^2 \|\mathbf{u}\|_2^2 \sum_{j_1} \text{vec}(\mathcal{W}_{j_1}^{(2,3)})^\top \boldsymbol{\Phi}_0^{(\text{cano})} \text{vec}(\mathcal{W}_{j_1}^{(2,3)})
\end{aligned}$$

$$\leq (\lambda\sigma)^2 \|\Phi_0^{(\text{cano})}\|_S \|\mathbf{u}\|_2^2 \|\text{vec}(\mathcal{W})\|_2^2$$

with the $\Phi_0^{(\text{cano})}$ in (31). Thus, as $\mathcal{W}$ is of dimension $d_2 \times r_1 \times r_2$, the derivation of (81) yields

$$\mathbb{E}\|\Delta_2\|_S = \mathbb{E}\|\overline{\Delta}_2\|_S \leq \frac{\sigma\lambda(2T)^{1/2}(\sqrt{d_1} + \sqrt{d_2 r_1 r_2})}{T-h} \|\Phi_0^{(\text{cano})}\|_S^{1/2}. \quad (82)$$

For the third term Δ_3 , we partition $(h, T]$ as $S_1 \cup S_2$ as in the derivation of (79), so that by Lemma 2 (ii) with $(d_1, d_2, d_3, d_4, n) = (d_1, d_2, d_1, d_2, |S_a|)$

$$\begin{aligned} \mathbb{E}\|\Delta_3\|_S &\leq \sum_{a=1}^2 \mathbb{E} \left\| \text{mat}_1 \left(\sum_{t \in S_a} \frac{\mathbf{E}_{t-h} \otimes \mathbf{E}_t}{T-h} \right) \right\|_S \\ &\leq \sum_{a=1}^2 \frac{\sigma^2 (\sqrt{d_1 |S_a|} + d_1 \sqrt{d_2} + d_2 \sqrt{|S_a| d_1})}{T-h} \\ &\leq \frac{\sigma^2 (1 + d_2) \sqrt{2d_1}}{\sqrt{T-h}} + \frac{2\sigma^2 d_1 \sqrt{d_2}}{T-h}. \end{aligned} \quad (83)$$

We obtain (36) by applying (81), (82) and (83) to the three terms in (80) and Cauchy-Schwarz. Finally (37) follows from (36) via Wedin (1972). $\square$

Proof of Corollaries 1 and 2. Because the population matrices are of full rank, (28) implies that the norms $\|\Phi_0^{(\text{cano})}\|_S$ and $\|\Phi_{k,0}^{(\text{cano})*}\|_S$ and singular values $\sigma_m(\Phi_{k,1:h_0}^{(\text{cano})})$ and $\sigma_m(\Phi_{k,1:h_0}^{(\text{cano})*})$ are all of constant order in the Fixed-Rank Factor Model. This gives (38) and (46) due to $d_k/d \leq 1 \leq r/r_k = O(1)$ and $r/d_k = O(1)$ in (35). For (39), we have $(\sigma/\lambda)^2 \sqrt{d_k d}/T \leq (\sigma/\lambda)^2 \sqrt{d/T}$ when $(\sigma/\lambda)^2 \sqrt{d/T} \leq 1$. We omit the rest of the proof as the same argument applies. $\square$

Proof of Lemma 1. By the definitions of the effective ranks of $\Phi_0^{(\text{cano})}$ and $\Phi_{k,0}^{(\text{cano})*}$ in (31) and (32), (61) holds. Let $\mathbf{u}_m$ and $\mathbf{v}_m = (\mathbf{v}_{m,1}^\top, \dots, \mathbf{v}_{m,h_0}^\top)^\top$ be the singular vectors of $\Phi_{k,1:h_0}^{(\text{cano})} \in \mathbb{R}^{r_k \times (r-k r h_0)}$ with $\mathbf{u}_m \in \mathbb{R}^{r_k}$ and $\mathbf{v}_{m,h} \in \mathbb{R}^{r-k}$. We identify $\mathbf{v}_{m,h}$ with $\mathbf{V}_{m,h} \in \mathbb{R}^{r-k \times r}$. By (29),

$$\begin{aligned} &\sum_{m=1}^{r_k} \sigma_m(\Phi_{k,1:h_0}^{(\text{cano})}) \\ &= \sum_{m=1}^{r_k} \mathbf{u}_m^\top (\Phi_{k,1}^{(\text{cano})}, \dots, \Phi_{k,h_0}^{(\text{cano})}) \mathbf{v}_m \\ &= \sum_{m=1}^{r_k} \sum_{h=1}^{h_0} \sum_{t=h+1}^T \frac{\mathbf{u}_m^\top \text{mat}_1(\text{mat}_k(\mathcal{F}_{t-h}^{(\text{cano})}) \otimes \text{mat}_k(\mathcal{F}_t^{(\text{cano})})) \mathbf{v}_{m,h}}{T-h} \\ &= \sum_{m=1}^{r_k} \sum_{h=1}^{h_0} \sum_{t=h+1}^T \frac{\mathbf{u}_m^\top \text{mat}_k(\mathcal{F}_{t-h}^{(\text{cano})}) (\mathbf{V}_{m,h} \text{vec}(\mathcal{F}_t^{(\text{cano})}))}{T-h} \\ &\leq \left(\sum_{m=1}^{r_k} \sum_{h=1}^{h_0} \sum_{t=h+1}^T \frac{\|\text{mat}_k^\top(\mathcal{F}_{t-h}^{(\text{cano})}) \mathbf{u}_m\|_2^2}{T-h} \right)^{1/2} \left(\sum_{m=1}^{r_k} \sum_{h=1}^{h_0} \sum_{t=h+1}^T \frac{\|\mathbf{V}_{m,h} \text{vec}(\mathcal{F}_t^{(\text{cano})})\|_2^2}{T-h} \right)^{1/2} \\ &= \left(\sum_{h=1}^{h_0} \text{trace} \left(\sum_{t=h+1}^T \frac{\text{mat}_k(\mathcal{F}_{t-h}^{(\text{cano})}) \text{mat}_k^\top(\mathcal{F}_{t-h}^{(\text{cano})})}{T-h} \right) \right)^{1/2} \end{aligned}$$

$$\begin{aligned}
& \times \left(\sum_{m=1}^{r_k} \sum_{h=1}^{h_0} \text{trace} \left(\mathbf{V}_{m,h}^\top \mathbf{V}_{m,h} \sum_{t=h+1}^T \frac{\text{vec}(\mathcal{F}_t^{(\text{cano})}) \text{vec}^\top(\mathcal{F}_t^{(\text{cano})})}{T-h} \right) \right)^{1/2} \\
& \leq h_0^{1/2} \left(\frac{\text{trace}(\boldsymbol{\Phi}_0^{(\text{cano})})}{1-h_0/T} \right)^{1/2} \left(\frac{\|\boldsymbol{\Phi}_0^{(\text{cano})}\|_S}{1-h_0/T} \right)^{1/2} \left(\text{trace} \left(\sum_{m=1}^{r_k} \sum_{h=1}^{h_0} \mathbf{V}_{m,h}^\top \mathbf{V}_{m,h} \right) \right)^{1/2} \\
& = \sqrt{h_0 r_k / r_0} \text{trace}(\boldsymbol{\Phi}_0^{(\text{cano})}) / (1-h_0/T).
\end{aligned}$$

This gives (62).

Let $\mathbf{u}_m$ and $\mathbf{v}_m = (\mathbf{v}_{m,1}^\top, \dots, \mathbf{v}_{m,h_0}^\top)^\top$ be the singular vectors of $\boldsymbol{\Phi}_{k,1:h_0}^{(\text{cano})*} \in \mathbb{R}^{r_k \times (r_k h_0)}$. By (32),

$$\begin{aligned}
\sum_{m=1}^{r_k} \sigma_m(\boldsymbol{\Phi}_{k,1:h_0}^{(\text{cano})*}) &= \sum_{m=1}^{r_k} \sum_{h=1}^{h_0} \sum_{t=h+1}^T \frac{\mathbf{u}_m^\top \text{mat}_k(\mathcal{F}_{t-h}^{(\text{cano})}) \text{mat}_k^\top(\mathcal{F}_t^{(\text{cano})}) \mathbf{v}_{m,h}}{T-h} \\
&\leq \left(h_0 \sum_{m=1}^{r_k} \frac{\mathbf{u}_m^\top \boldsymbol{\Phi}_{k,0}^{(\text{cano})*} \mathbf{u}_m}{1-h_0/T} \right)^{1/2} \left(\sum_{m=1}^{r_k} \sum_{h=1}^{h_0} \frac{\mathbf{v}_{m,h}^\top \boldsymbol{\Phi}_{k,0}^{(\text{cano})*} \mathbf{v}_{m,h}}{1-h_0/T} \right)^{1/2} \\
&\leq (h_0 \text{trace}(\boldsymbol{\Phi}_{k,0}^{(\text{cano})*}))^{1/2} (\|\boldsymbol{\Phi}_{k,0}^{(\text{cano})*}\|_S r_k)^{1/2} / (1-h_0/T) \\
&= \sqrt{h_0 r_k / r_{k,0}^*} \text{trace}(\boldsymbol{\Phi}_0^{(\text{cano})}) / (1-h_0/T).
\end{aligned}$$

This gives (63) and completes the proof. $\square$

Proof of Corollary 3. By (35), (53) and (54),

$$\begin{aligned}
& h_0^{1/2} \Delta_{k,h_0} / \sigma_{r_k}(\boldsymbol{\Phi}_{k,1:h_0}^{(\text{cano})}) \\
& \lesssim \frac{h_0^{1/2} d_{-k}^{1/2}}{\sigma_{r_k}(\boldsymbol{\Phi}_{k,1:h_0}^{(\text{cano})})} \left[\frac{\sigma d_k^{1/2}}{\lambda T^{1/2}} \sqrt{r/r_k} \left\{ \|\boldsymbol{\Phi}_{k,0}^{(\text{cano})*}\|_S^{1/2} + \|\boldsymbol{\Phi}_0^{(\text{cano})}\|_S^{1/2} \right\} + \frac{\sigma^2}{\lambda^2} (\sqrt{d/T} + d_k/T) \right] \\
& \lesssim d_{-k}^{1/2} d^{\delta_1 - \delta_0} \frac{\sqrt{r_0 r_k}}{r} \left\{ \frac{\sigma d_k^{1/2}}{\lambda T^{1/2}} \sqrt{r/r_k} \left\{ \sqrt{r/r_{k,0}^*} + \sqrt{r/r_0} \right\} + \frac{\sigma^2 d^{1/2}}{\lambda^2 \sqrt{T}} + \frac{\sigma^2 d_k}{\lambda^2 T} \right\},
\end{aligned}$$

which implies (64) as the third term above can be dropped due to $\|\widehat{\mathbf{P}}_k - \mathbf{P}_k\|_S \leq 1$. Similarly, (66) follows from (43), (53) and (56) as

$$\begin{aligned}
\frac{h_0^{1/2} \Delta_{k,h}^*}{\sigma_{r_k}(\boldsymbol{\Phi}_{k,1:h_0}^{(\text{cano})*})} &\asymp \frac{h_0^{1/2}}{\sigma_{r_k}(\boldsymbol{\Phi}_{k,1:h_0}^{(\text{cano})*})} \left[\frac{\sigma d_k^{1/2}}{\lambda T^{1/2}} \|\boldsymbol{\Phi}_{k,0}^{(\text{cano})*}\|_S^{1/2} + \frac{\sigma^2 \sqrt{d}}{\lambda^2 \sqrt{T}} + \frac{\sigma^2 d_k}{\lambda^2 T} \right] \\
&\asymp \frac{d^{\delta_1^* - \delta_0} \sqrt{r_k r_{k,0}^*}}{r} \left[\frac{\sigma d_k^{1/2} r^{1/2}}{\lambda T^{1/2} (r_{k,0}^*)^{1/2}} + \frac{\sigma^2 \sqrt{d}}{\lambda^2 \sqrt{T}} + \frac{\sigma^2 d_k}{\lambda^2 T} \right]
\end{aligned}$$

with $r_{k,0}^* \leq r_k$. The proof of (65) and (67) are similar and omitted. $\square$

Proof of Corollary 4. As $d^{-\delta_0}$ is assumed to be of the order of the empirical version of the SNR and λ^2 is scaled to $\sigma^2 d^{1-\delta_0}/r$, (61) gives (53) for $\text{trace}(\boldsymbol{\Phi}_0)$. By Corollary 3, it suffices to verify (54) and the bounds for the effective ranks. As $\mathcal{F}_t^{(\text{cano})} = \mathcal{F}_t \otimes_{k=1}^K (\mathbf{D}_k \mathbf{V}_k^\top)$, $\boldsymbol{\Phi}_{k,h}^{(\text{cano})} = (\mathbf{D}_k \mathbf{V}_k^\top) \boldsymbol{\Phi}_{k,h} \mathbf{B}_k^\top$, where $\mathbf{B}_k = \odot_{k \in [2K] \setminus \{k\}} (\mathbf{D}_k \mathbf{V}_k^\top) \in \mathbb{R}^{(r-kr) \times (r-kr)}$ with $\mathbf{D}_{k+K} = \mathbf{D}_k$ and $\mathbf{V}_{k+K} = \mathbf{V}_k$. Thus, by

the minimax principle for eigenvalues

$$\sigma_m^2(\Phi_{k,1:h_0}^{\text{cano}}) = \max_{\substack{H \subset \mathbb{R}^{r_k} \\ \dim(H)=m}} \min_{\substack{\mathbf{u} \in H \\ \|\mathbf{u}\|_2=1}} \mathbf{u}^\top \left(\sum_{h=1}^{h_0} (\mathbf{D}_k \mathbf{V}_k^\top) \Phi_{k,h} \mathbf{B}_k^\top \mathbf{B}_k \Phi_{k,h}^\top (\mathbf{V}_k \mathbf{D}_k^\top) \right) \mathbf{u}.$$

As $\mathbf{B}_k$ is of full rank with the smallest singular value $\prod_{j \in [2K] \setminus \{k\}} \sigma_{r_j}(\mathbf{D}_j) = \kappa_0^2 / \sigma_{r_k}(\mathbf{D}_k)$,

$$\begin{aligned} \sigma_{r_k}^2(\Phi_{k,1:h_0}^{\text{cano}}) &\geq \kappa_0^4 \sigma_{r_k}^{-2}(\mathbf{D}_k) \max_{\substack{H \subset \mathbb{R}^{r_k} \\ \dim(H)=r_k}} \min_{\substack{\mathbf{u} \in H \\ \|\mathbf{u}\|_2=1}} \mathbf{u}^\top (\mathbf{D}_k \mathbf{V}_k^\top) \left(\sum_{h=1}^{h_0} \Phi_{k,h} \Phi_{k,h}^\top \right) (\mathbf{V}_k \mathbf{D}_k^\top) \mathbf{u} \\ &\geq \kappa_0^4 \sigma_{r_k}^{-2}(\mathbf{D}_k) \sigma_{r_k}^2(\Phi_{k,1:h_0}) \max_{\substack{H \subset \mathbb{R}^{r_k} \\ \dim(H)=r_k}} \min_{\substack{\mathbf{u} \in H \\ \|\mathbf{u}\|_2=1}} \mathbf{u}^\top (\mathbf{D}_k \mathbf{V}_k^\top) (\mathbf{V}_k \mathbf{D}_k^\top) \mathbf{u} \\ &= \kappa_0^4 \sigma_{r_k}^2(\Phi_{k,1:h_0}), \end{aligned}$$

so that the condition on $\sigma_{r_k}^2(\Phi_{k,1:h_0})$ implies (54). Similarly, as $\|\mathbf{D}_k\|_S = 1$ and $\prod_{k=1}^K \sigma_{r_k}(\mathbf{D}_k) = \kappa_0$,

$$\left\{ \frac{\text{trace}(\Phi_0^{\text{cano}})}{\text{trace}(\Phi_0)}, \frac{\|\Phi_0^{\text{cano}}\|_S}{\|\Phi_0\|_S}, \frac{\text{trace}(\Phi_{k,0}^{\text{cano}*})}{\text{trace}(\Phi_{k,0}^*)}, \frac{\|\Phi_{k,0}^{\text{cano}*}\|_S}{\|\Phi_{k,0}^*\|_S} \right\} \subset [\kappa_0^2, 1],$$

which implies the bounds for the effective ranks. $\square$

Proof of Proposition 1. By the choice of λ and the condition $\mathbb{E}[\sum_{t=1}^T f_{it}^2/T] = 1$,

$$\sum_{i=1}^{r_1} \kappa_i^2 = \mathbb{E}[\text{trace}(\overline{\Sigma}_{\kappa f,0})] = r = r_1^K.$$

Let $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_{r_1})$ with $\mathbf{b}_i = \text{vec}(\otimes_{k=1}^K \mathbf{b}_{k,i}) \in \mathbb{R}^d$. It follows from (6) and (31) that

$$\left(\odot_{k=1}^K \mathbf{U}_k \right) \Phi_0^{\text{cano}} \left(\odot_{k=1}^K \mathbf{U}_k \right)^\top = (\lambda^2 T)^{-1} \sum_{t=1}^T \text{vec}(\mathcal{M}_t) \text{vec}(\mathcal{M}_t)^\top = \mathbf{B} \overline{\Sigma}_{\kappa f,0} \mathbf{B}^\top.$$

As both $\odot_{k=1}^K \mathbf{U}_k$ and $\mathbf{B}$ are orthonormal, $\sigma_m(\Phi_0^{\text{cano}}) = \sigma_m(\overline{\Sigma}_{\kappa f,0})$. Let $\mathbf{b}_{-k,i} = \text{vec}(\otimes_{j \neq k} \mathbf{b}_{j,i}) \in \mathbb{R}^{d-k}$, $\mathbf{v}_{k,i_1,i_2} = \text{vec}(\mathbf{b}_{-k,i_1} \otimes \mathbf{b}_{k,i_2} \otimes \mathbf{b}_{-k,i_2})$, and $\mathbf{U}_{[2K] \setminus \{k\}}^\top$ be as in (30). By (6), (29) and (69),

$$\mathbf{U}_k \Phi_{k,h}^{\text{cano}} \mathbf{U}_{[2K] \setminus \{k\}}^\top = \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_1} \kappa_{i_1} \kappa_{i_2} (\overline{\Sigma}_{f,h})_{i_1,i_2} \mathbf{b}_{k,i_1} \mathbf{v}_{k,i_1,i_2}^\top = \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_1} (\overline{\Sigma}_{\kappa f,h})_{i_1,i_2} \mathbf{b}_{k,i_1} \mathbf{v}_{k,i_1,i_2}^\top.$$

By (70), $\mathbf{b}_{-k,1}, \dots, \mathbf{b}_{-k,r_1}$ are orthonormal, so that $\mathbf{v}_{k,i_1,i_2}^\top \mathbf{v}_{k,i_3,i_4} = I\{i_1 = i_3, i_2 = i_4\}$. Thus, as $\mathbf{U}_k$ and $\mathbf{U}_{[2K] \setminus \{k\}}^\top$ are orthonormal,

$$\begin{aligned} \sigma_m^2(\Phi_{k,1:h_0}^{\text{cano}}) &= \sigma_m \left(\sum_{h=1}^{h_0} \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_1} \sum_{i_3=1}^{r_1} \sum_{i_4=1}^{r_1} (\overline{\Sigma}_{\kappa f,h})_{i_1,i_2} (\overline{\Sigma}_{\kappa f,h})_{i_3,i_4} \mathbf{b}_{k,i_1} \mathbf{v}_{k,i_1,i_2}^\top \mathbf{v}_{k,i_3,i_4} \mathbf{b}_{k,i_3}^\top \right) \\ &= \sigma_m \left(\sum_{h=1}^{h_0} \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_1} (\overline{\Sigma}_{\kappa f,h})_{i_1,i_2}^2 \mathbf{b}_{k,i_1} \mathbf{b}_{k,i_1}^\top \right) \end{aligned}$$

$$= \sigma_m \left(\sum_{h=1}^{h_0} \mathbf{B}_k \text{diag}(\overline{\boldsymbol{\Sigma}}_{\kappa f, h} \overline{\boldsymbol{\Sigma}}_{\kappa f, h}^\top) \mathbf{B}_k^\top \right).$$

In view of (32), the orthonormality of $\mathbf{b}_{-k,1}, \dots, \mathbf{b}_{-k,r_1}$ also gives

$$\mathbf{U}_k \boldsymbol{\Phi}_{k,h}^{(\text{cano})*} \mathbf{U}_k^\top = \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_1} (\overline{\boldsymbol{\Sigma}}_{\kappa f, h})_{i_1, i_2} \mathbf{b}_{k, i_1} \mathbf{b}_{-k, i_1}^\top \mathbf{b}_{-k, i_2} \mathbf{b}_{k, i_2}^\top = \mathbf{B}_k \text{diag}(\overline{\boldsymbol{\Sigma}}_{\kappa f, h}) \mathbf{B}_k^\top,$$

which gives (73) with $h = 0$ and

$$\sigma_m^2(\boldsymbol{\Phi}_{k,1:h_0}^{(\text{cano})*}) = \sigma_m \left(\sum_{h=1}^{h_0} \left(\mathbf{B}_k \text{diag}(\overline{\boldsymbol{\Sigma}}_{\kappa f, h}) \mathbf{B}_k^\top \right)^2 \right).$$

as $\mathbf{U}_k$ is orthonormal. □

Appendix B: Additional information on the import-export network example

In this appendix we provide the detailed data description of the import export data used in the example, as well as some additional figures.

The data is obtained from UN Comtrade Database at <https://comtrade.un.org>. In this study we use the monthly observations of 22 large economies in North America and Europe from January 2010 to December 2016. The countries used are Belgium (BE), Bulgaria (BU), Canada (CA), Denmark (DK), Finland (FI), France (FR), Germany (DE), Greece (GR), Hungary (HU), Iceland (IS), Ireland (IR), Italy (IT), Mexico (MX), Norway (NO), Poland (PO), Portugal (PT), Spain (ES), Sweden (SE), Switzerland (CH), Turkey (TR), United States (US) and United Kingdom (UK).

The trade data includes commodity classifier (2 digit Hamonized System codes). Following the classification shown at <https://www.foreign-trade.com/reference/hscod.htm>, we divide all products into 15 categories, including Animal & Animal Products (HS code 01-05), Vegetable Products (06-15), Foodstuffs (16-24), Mineral Products (25-27), Chemicals & Allied Industries (28-38), Plastics & Rubbers (39-40), Raw Hides, Skins, Leather & Furs (41-43), Wood & Wood Products (44-49), Textiles (50-63), Footwear & Headgear (64-67), Stone & Glass (68-71), Metals (72-83), Machinery & Electrical (84-85), Transportation (86-89), and Miscellaneous (90-97).

The following two figures are the network figures for condensed product groups 3 to 6. They can be interpreted similarly as Figure 7.

Figure 16 shows the normalized trading volumes among the hubs (factors) to show the variation in trading through time. Note that the scales are very different among the figures.

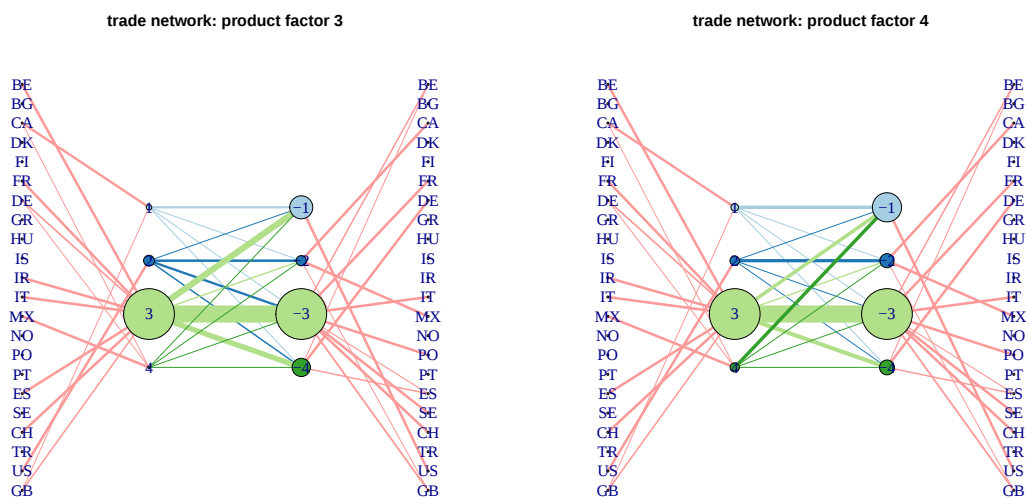


Figure 14: Trade network for condensed product group 3 (left) and group 4 (right). Export and import hubs on the left and right of the center network. Line width is proportional to total volume of trade between the hubs for the last three years (2015 to 2017). Vertex size is proportional to total volume of trades through the vertex. The line width between the countries and the hubs is proportional to the corresponding loading coefficients, for coefficients larger than 0.05 only.

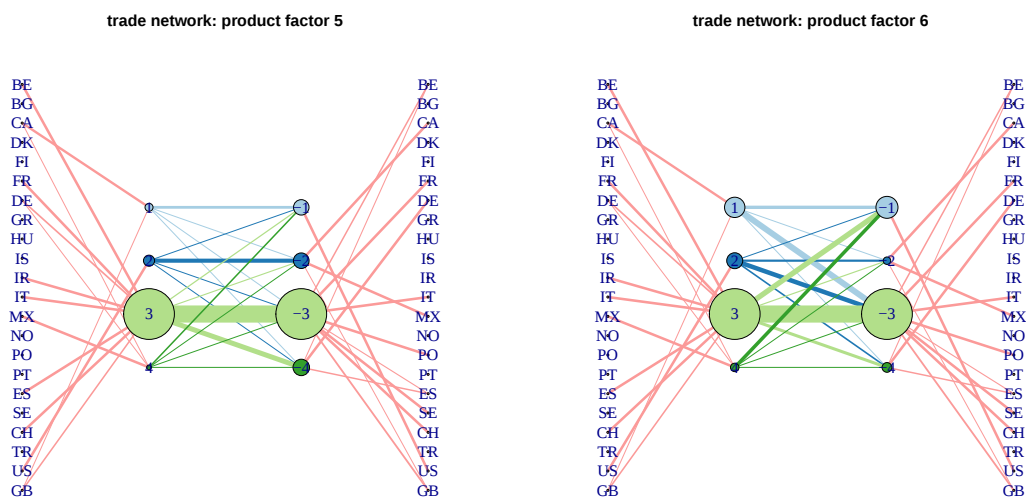


Figure 15: Trade network for condensed product group 5 (left) and group 6 (right). Export and import hubs on the left and right of the center network. Line width is proportional to total volume of trade between the hubs for the last three years (2015 to 2017). Vertex size is proportional to total volume of trades through the vertex. The line width between the countries and the hubs is proportional to the corresponding loading coefficients, for coefficients larger than 0.05 only.

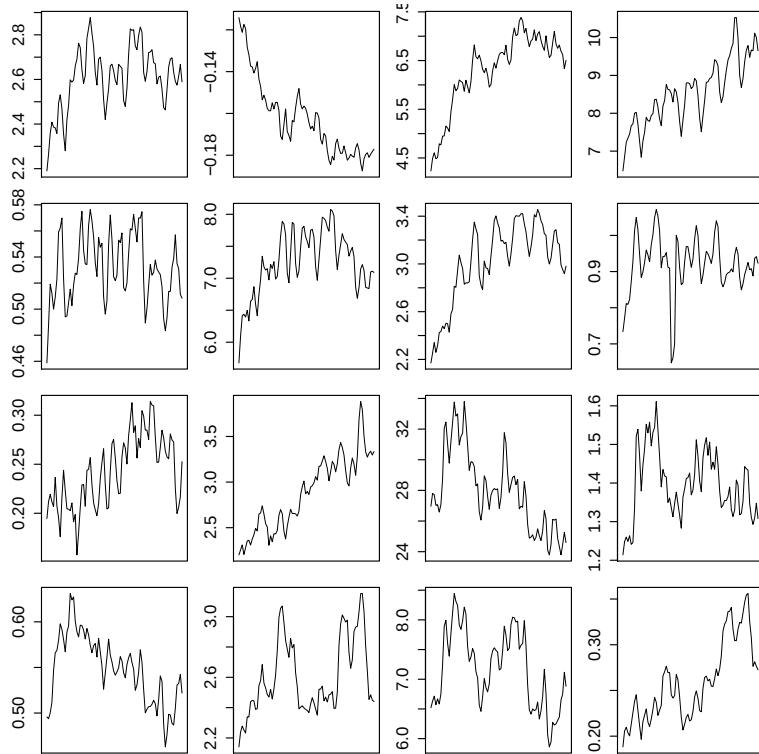


Figure 16: Trading volumes among the hubs (factors) of condensed product group 1. Rows are for export and columns for import.

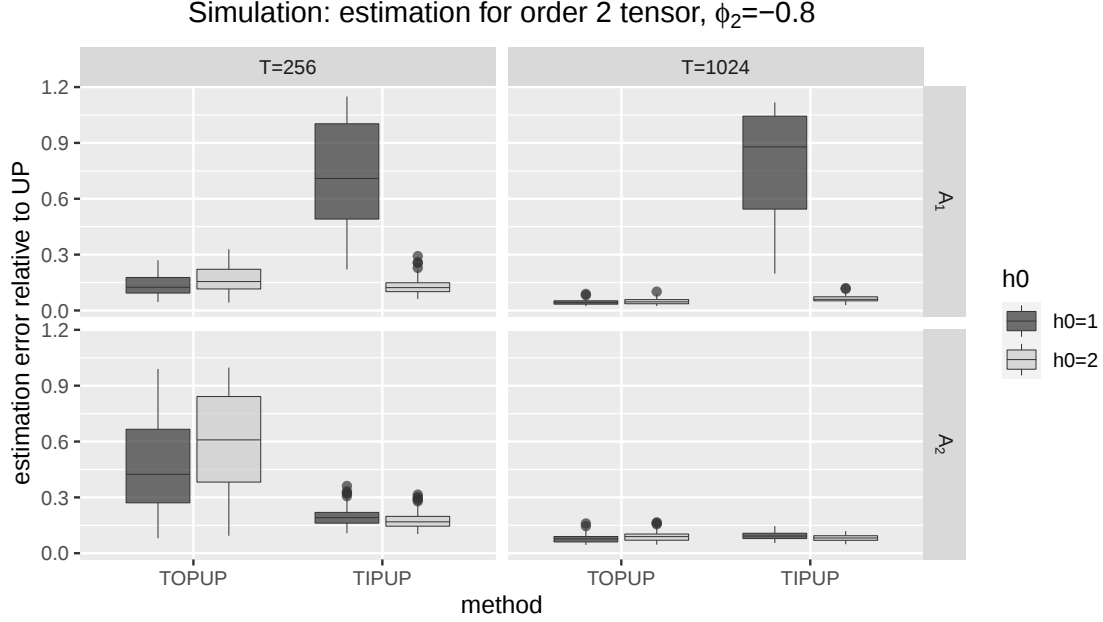


Figure 17: Finite sample comparison between TIPUP, TOPUP with different h_0 and UP under model (75) where $\phi_1 = 0.8$, $\phi_2 = -0.8$. The boxplots show the ratio of estimation error from TTPUP or TOPUP and the estimation error from UP. The top row is for estimation of the column space of the mode 1 loading matrix $\mathbf{A}_1$ and bottom for $\mathbf{A}_2$. The left column is for $T = 256$ and right for $T = 1024$.

Appendix C: Additional Simulation Results

C.1: More results from Example 2 in Section 6:

In Example 2 of Section 6, Figures 2 and 3 provide the estimation error comparison of the TOPUP, TIPUP and UP where boxplots of the logarithms of the estimation errors were provided. Figures 17 and 18 here show the relative estimation error, which is the ratio of the estimation error from TIPUP or TOPUP and the estimation error from UP. It is clear that the estimation improvements of TIPUP and TOPUP over UP are significant, especially when the sample size T is large. The figures also show the performance comparison between TIPUP and TOPUP, as well as that between using $h_0 = 1$ and $h_0 = 2$.

C.2: Simulation results of an order 3 tensor time series:

Here we present some simulation results under an order 3 tensor time series setting. Specifically, let

$$\mathcal{X}_t = 16\mathcal{F}_t \times_1 \mathbf{A}_1 \times_2 \mathbf{A}_2 \times_3 \mathbf{A}_3 + \mathcal{E}_t \quad (84)$$

where $\mathcal{F}_t$ is a $2 \times 2 \times 2$ factor, with three independent AR(1) processes $f_{111t} = 0.7f_{111t-1} +$

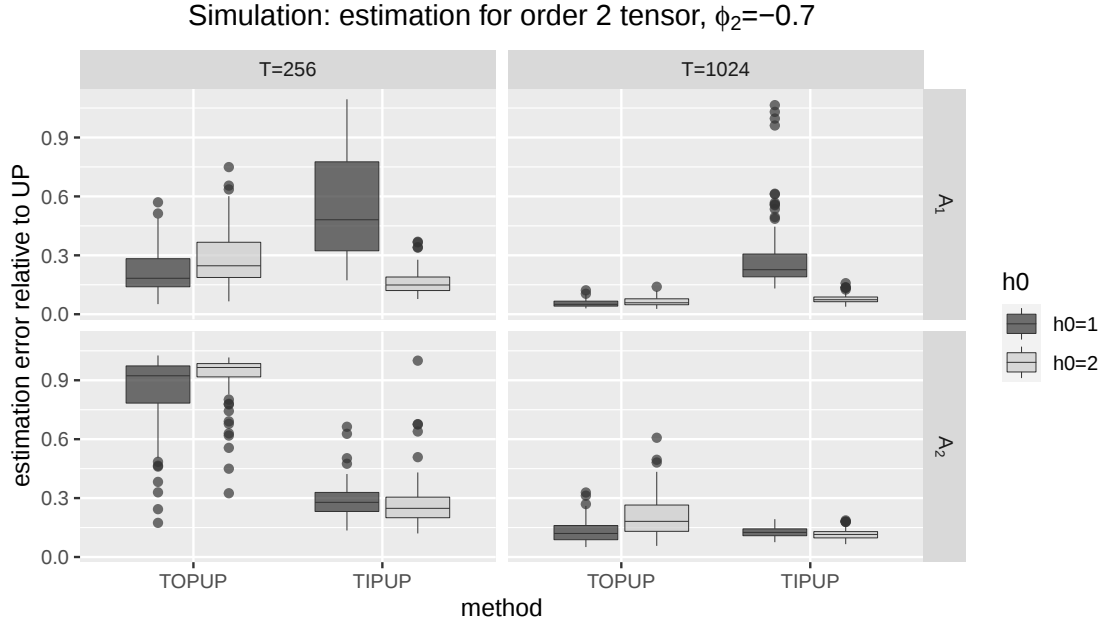


Figure 18: Finite sample comparison between TIPUP, TOPUP with different h_0 and UP under model (75) where $\phi_1 = 0.8$, $\phi_2 = -0.7$. The boxplots show the ratio of estimation error from TIPUP or TOPUP and the estimation error from UP. The top row is for estimation of the column space of the mode 1 loading matrix $\mathbf{A}_1$ and bottom for $\mathbf{A}_2$. The left column is for $T = 256$ and right for $T = 1024$.

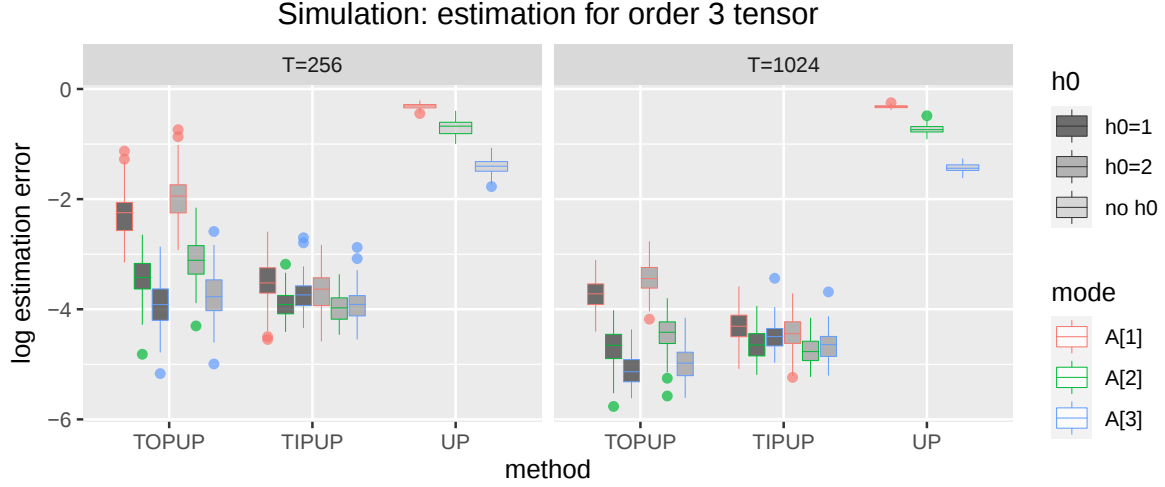


Figure 19: Finite sample comparison between TIPUP, TOPUP with different h_0 and UP under model (84) for order 3 tensor time series. The boxplots show the logarithms of the estimation errors. The three rows are for estimation of the three modes respectively. The left column is for $T = 256$ and right for $T = 1024$.

e_{111t} , $f_{211t} = 0.6f_{211t-1} + e_{211t}$, $f_{222t} = 0.8f_{222t-1} + e_{222t}$, one AR(2) process $f_{221t} = 0.5f_{221t-1} + 0.3f_{221t-2} + e_{221t}$, and four independent white noise for $f_{112t}, f_{121t}, f_{122t}, f_{212t}$. The noise is generated according to $\mathcal{E}_t = \mathcal{Z}_t \times_1 \Psi_1^{1/2} \times_2 \Psi_2^{1/2} \times_3 \Psi_3^{1/2}$, where $\mathcal{Z}_t \in \mathbb{R}^{16 \times 16 \times 16}$ are iid $N(0, 1)$, and Ψ_1, Ψ_2, Ψ_3 all have diagonal entries 1 with off diagonal entries 0.4, 0.3, 0.2, respectively. $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ are all orthonormal matrices obtained via QR decomposition of randomly generated 16×2 matrices with iid $N(0, 1)$ entries. The dimensions are $d_1 = d_2 = d_3 = 16$. We repeat the experiments 100 times.

Figure 19 show the boxplots of the logarithm of the estimation errors. In this setting, there is no signal cancellation, TOPUP and TIPUP with $h_0 = 1$ and $h_0 = 2$ uniformly dominate UP and the improvement is larger for larger sample size. The winner among TOPUP and TIPUP depends on the mode and sample size T . In general, the estimation of mode 1 is worse than the other two modes for both TOPUP and TIPUP with different h_0 . To understand the differences between different modes, we can resort to our theory partially: consider the signal strengths defined as $\lambda_k = \lambda_{\sigma_{r_k}}(\Phi_{k,1:h_0})/h_0^{1/2}$ as in (37) for TOPUP, and in parallel $\lambda_k^* = \lambda_{\sigma_{r_k}}(\Phi_{k,1:h_0}^*)/h_0^{1/2}$ in (45) for TIPUP. Table 7 provides the theoretical values of λ_k of TOPUP and λ_k^* of TIPUP for $k = 1, 2, 3$ and $h_0 = 1, 2$. One can see that for TOPUP with $h_0 = 1, 2$, $\lambda_1 < \lambda_2 < \lambda_3$, hence we expect the performance of TOPUP improves from mode 1 to mode 2 to mode 3, whereas for TIPUP with $h_0 = 1, 2$, we have $\lambda_1 < \lambda_2 \approx \lambda_3$, hence TIPUP's estimation of mode 1 is worse than the other two modes and mode 2 and mode 3 are on par. Both are consistent with the empirical performance shown in Figure 19.

	TOPUP λ_k		TIPUP λ_k^*	
Mode k	$h_0 = 1$	$h_0 = 2$	$h_0 = 1$	$h_0 = 2$
Mode 1	351.37	303.28	351.37	303.28
Mode 2	425.51	362.14	591.37	500.89
Mode 3	568.89	515.15	568.89	515.15

Table 7: Simulation study with order 3 tensor time series. The signal strengths of TOPUP and TIPUP respectively for $k = 1, 2, 3$ and $h_0 = 1, 2$.

C.3: A parametric bootstrap exercises

As an informal statistical inference exercises for the import-export transport network example in Section 7.1, a parametric bootstrap study is performed as follows. By using the estimates $\hat{\mathbf{A}}_1, \hat{\mathbf{A}}_2, \hat{\mathbf{A}}_3$ via the TIPUP, we obtain $\hat{\mathcal{M}}_t$ through projection and the residual $\hat{\mathcal{E}}_t = \mathcal{X}_t - \hat{\mathcal{M}}_t$. Assume the tensor $\mathcal{E}_t$ follows array normal distribution with covariance $\Sigma_1, \Sigma_2, \Sigma_3$, we can obtain their estimates $\hat{\Sigma}_1, \hat{\Sigma}_2, \hat{\Sigma}_3$ from the residuals $\hat{\mathcal{E}}_t$ by tensor unfolding. We then bootstrap $\mathcal{X}_t^{(b)} = \hat{\mathcal{M}}_t + \mathcal{E}_t^{(b)}$ for $b = 1, \dots, 100$, where $\mathcal{E}_t^{(b)}$ follows the array normal distribution with covariance $\hat{\Sigma}_1, \hat{\Sigma}_2, \hat{\Sigma}_3$. With each bootstrap sample $\mathcal{X}_{1:T}^{(b)}$, we implement the TIPUP (with $h_0 = 2$), TOPUP (with $h_0 = 2$), and UP, and obtain the output $\hat{\mathbf{A}}_1^{(b)}, \hat{\mathbf{A}}_2^{(b)}, \hat{\mathbf{A}}_3^{(b)}$ as estimates of $\hat{\mathbf{A}}_1, \hat{\mathbf{A}}_2, \hat{\mathbf{A}}_3$.

With 100 bootstrapped estimates, Figure 20 shows the boxplot of the logarithm of the bootstrap estimation errors, defined as $\|\mathbf{P}_k^{(b)} - \hat{\mathbf{P}}_k\|_S$, where $\hat{\mathbf{P}}_k = \hat{\mathbf{A}}_k(\hat{\mathbf{A}}_k^\top \hat{\mathbf{A}}_k)^{-1} \hat{\mathbf{A}}_k^\top$ is the projection matrix of $\hat{\mathbf{A}}_k$ and $\hat{\mathbf{P}}_k^{(b)}$ is the projection matrix of $\hat{\mathbf{A}}_k^{(b)}$. The figure shows that, for mode 1, the export mode, all three methods are similar. For mode 2, the import mode, TOPUP is slightly better than TIPUP, which is in turn slightly better than UP. For mode 3, the category mode, TOPUP and TIPUP both outperform UP significantly. For all three methods, the bootstrap variation for the estimation for mode 1 loading matrix $\hat{\mathbf{A}}_1$ is the smallest among the three modes. This can be better understood by comparing the signal strengths of TOPUP and TIPUP defined earlier in Section C.2. For TOPUP, the signal strengths are $\lambda_1^2 = 165.81$, $\lambda_2^2 = 21.74$, $\lambda_3^2 = 8.02$, respectively for three modes; for TIPUP, the signal strengths are $(\lambda_1^*)^2 = 50.39$, $(\lambda_2^*)^2 = 1.02$, $(\lambda_3^*)^2 = 0.28$ respectively. Since the signals for both TOPUP and TIPUP are the strongest along mode 1, the estimation of $\hat{\mathbf{A}}_1$ is the most reliable. Table 8 shows the 95% bootstrap upper confidence bound for the angular error of the estimates. That is, in 95% of the bootstrap samples, the largest canonical angle between the column spaces of $\hat{\mathbf{A}}_k$ and $\hat{\mathbf{A}}_k^{(b)}$ is smaller than the displayed number. The results give confidence regions for $\mathbf{P}_k$ in the same way as the standard parametric bootstrap confidence intervals.

C.4: A simulation study to verify the theoretical rates

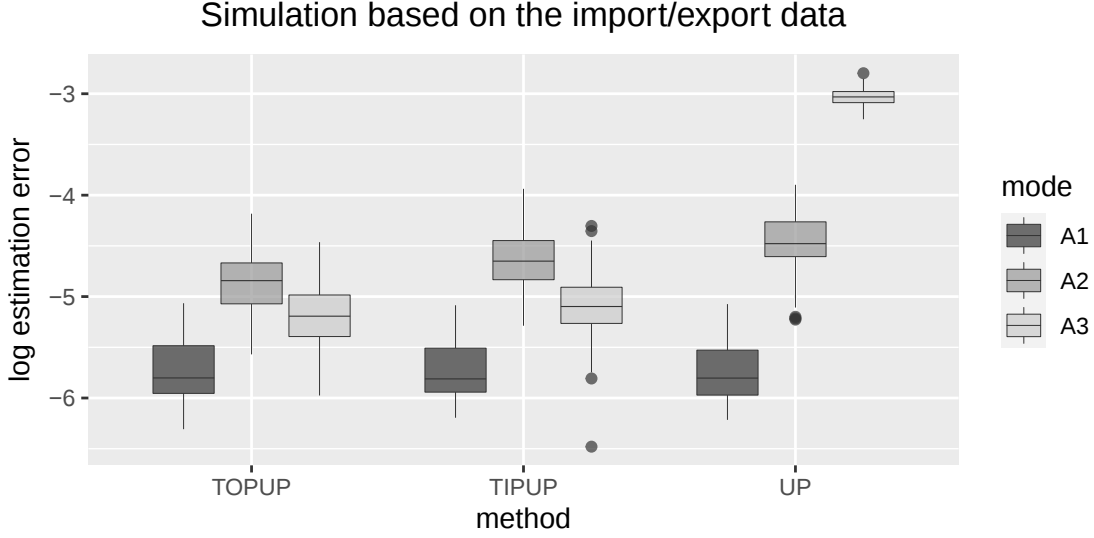


Figure 20: Parametric bootstrap results of the import/export example. Boxplots of the logarithms of the bootstrap estimation errors of the three modes using TOPUP, TIPUP and UP methods.

	TOPUP	TIPUP	UP
$\hat{P}_1$	0.0017π	0.0018π	0.0018π
$\hat{P}_2$	0.0036π	0.0049π	0.0053π
$\hat{P}_3$	0.0028π	0.0032π	0.0171π

Table 8: Parametric bootstrap upper confidence bound (95%) for the angular error in (26) for estimating P_k in the import/export example, $k = 1, 2, 3$.

We demonstrate the finite sample performance under a matrix factor model setting. We start with a simple one-factor model with $K = 2$. Let

$$\mathbf{X}_t = \lambda \mathbf{u}_1 f_t \mathbf{u}_2' + \mathbf{E}_t$$

where $\mathbf{X}_t$ and $\mathbf{E}_t$ are in $\mathbb{R}^{d_1 \times d_2}$, $\mathbf{u}_1 \in \mathbb{R}^{d_1 \times 1}$, $\mathbf{u}_2 \in \mathbb{R}^{d_2 \times 1}$ with $\|\mathbf{u}_1\|_2 = \|\mathbf{u}_2\|_2 = 1$, and the factor f_t is a univariate time series following $f_t \sim AR(1)$ with AR coefficient $\phi = .6$ and standard $N(0, 1)$ noise. The noise $\mathbf{E}_t$ is white, i.e. $\mathbf{E}_t \perp \mathbf{E}_{t+h}$, $h > 0$, following an array normal distribution (Hoff, 2011) such that $\mathbf{E}_t = \Psi_1^{1/2} \mathbf{Z}_t \Psi_2^{1/2}$ 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 elements in the $d_1 \times d_2$ matrix $\mathbf{Z}_t$ are i.i.d $N(0, 1)$. The elements of the loadings $\mathbf{u}_1$ and $\mathbf{u}_2$ are generated from i.i.d $N(0, 1)$, then normalized so $\|\mathbf{u}_1\|_2 = \|\mathbf{u}_2\|_2 = 1$. The sample size T , the dimensions d_1, d_2 and the factor strength λ are chosen to be $T = 2^\ell$ for $\ell = 1, \dots, 15$, $d_1 = d_2 = 2^\ell$ for $\ell = 1, \dots, 6$ and

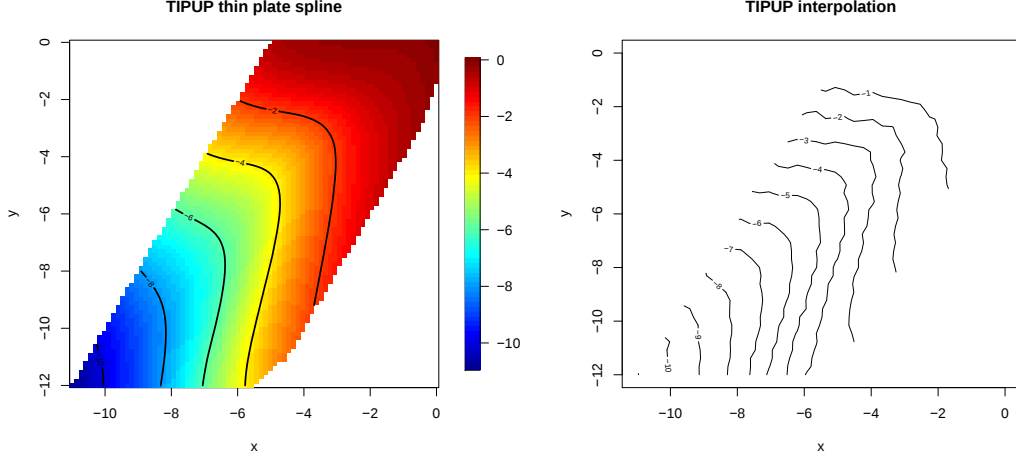


Figure 21: Estimated TIPUP risk in Example 1. The heatmap of the thin plate spline fit (left) and the interpolation contour (right) show the base-2 logarithm of the average of $|\sin(\text{angle}(\hat{\mathbf{u}}_1, \mathbf{u}_1))| = \|\mathbf{u}_1 \mathbf{u}_1^\top - \hat{\mathbf{u}}_1 \hat{\mathbf{u}}_1^\top\|_S$ via the TIPUP over 100 repetitions and over all possible combinations of d_1, d_2, T, λ for the same x and y coordinates, with $x = \log_2(d_1^{1/2}/(T^{1/2}\lambda))$ and $y = \log_2(d_1^{1/2}/(T^{1/2}\lambda^2))$ being the logarithms of the two components of the rate in (85).

$\lambda = 2^\ell$ for $\ell = -7, -6, \dots, 7$.

By Theorem 2 and (49), the error rate for estimating $\mathbf{u}_1$ via the TIPUP is

$$\frac{d_1^{1/2}}{T^{1/2}\lambda} + \frac{(d_1 d_2)^{1/2}}{T^{1/2}\lambda^2}. \quad (85)$$

Let $x = \log_2(d_1^{1/2}/(T^{1/2}\lambda))$, $y = \log_2((d_1 d_2)^{1/2}/(T^{1/2}\lambda^2))$, and z be the logarithm of the average of the corresponding estimation loss, $L = |\sin(\text{angle}(\hat{\mathbf{u}}_1, \mathbf{u}_1))| = \|\mathbf{u}_1 \mathbf{u}_1^\top - \hat{\mathbf{u}}_1 \hat{\mathbf{u}}_1^\top\|_S$ as in (49) and (26) over 100 simulation runs. A thin plate spline fit of (x, y, z) under different T, d_1, d_2, λ leads to the left panel of Figure 21 and its interpolation with the mean leads to the right panel.

The figures clearly confirm the theoretical results. Note that, when the two terms in (85) are of different rates (off 45 degree line in the figure), one of the rates would dominate hence the contour lines of the error rate should be either horizontal or vertical in the figure as (x, y) moves away from the 45 degree line. For negative AR coefficient $\phi = -0.6$ in the factor dynamics, the results are similar.

We also considered a parametric fit of the TIPUP error rate. Specifically, we fit the following model to the average loss L of estimating $\mathbf{u}_1$:

$$\log_2(L) \sim \log_2(c_1 2^{\nu_1} + c_6 2^{\nu_2}), \quad (86)$$

	c_1	c_2	c_3	c_5	c_5	c_6	c_7	c_8	c_9	c_{10}
Thm 2		0.50	0.00	-1.00	-0.50		0.50	0.50	-2.00	-0.50
fitted	1.19	0.51	-0.06	-0.73	-0.65	0.36	0.72	0.58	-1.92	-0.45

Table 9: Comparison between the theoretical rates and the simulated rates for the TIPUP.

	c_1	c_2	c_3	c_5	c_5	c_6	c_7	c_8	c_9	c_{10}
Thm 1		0.50	0.50	-1.00	-0.50		0.50	1.00	-2.00	-0.50
fit	0.83	0.51	-0.16	-0.66	-0.55	0.36	1.21	1.03	-2.51	-0.70

Table 10: Comparison between the theoretical rates and the simulated rates for the TOPUP.

where

$$\begin{aligned}\nu_1 &= c_2 \log_2 d_1 + c_3 \log_2 d_2 + c_4 \log_2 \lambda + c_5 \log_2 T \\ \nu_2 &= c_7 \log_2 d_1 + c_8 \log_2 d_2 + c_9 \log_2 \lambda + c_{10} \log_2 T.\end{aligned}$$

and compared the empirical fit with the theoretical results in Theorem 2. The results are shown in Table 9. They are reasonably close.

For TOPUP, the convergence rate based on Theorem 1 and (42) in this case is

$$\frac{(d_1 d_2)^{1/2}}{T^{1/2} \lambda} + \frac{(d_1 d_2^2)^{1/2}}{T^{1/2} \lambda^2}. \quad (87)$$

Similarly, let $x = \log_2 \left(\frac{(d_1 d_2)^{1/2}}{T^{1/2} \lambda} \right)$ and $y = \log_2 \left(\frac{(d_1 d_2^2)^{1/2}}{T^{1/2} \lambda^2} \right)$ and z be the logarithm of the average of corresponding estimation error over 100 runs. Figure 22 shows the results. The picture is not as clean as that of the TIPUP estimator, but it shows the trend.

Again, we fit the estimation error using (86) and compared it with the theoretical result in (87), shown in Table 10. There is some discrepancy though, possibly due to the limited range of the simulation setting. The results for multiple rank cases and three dimensional tensor time series show similar patterns.

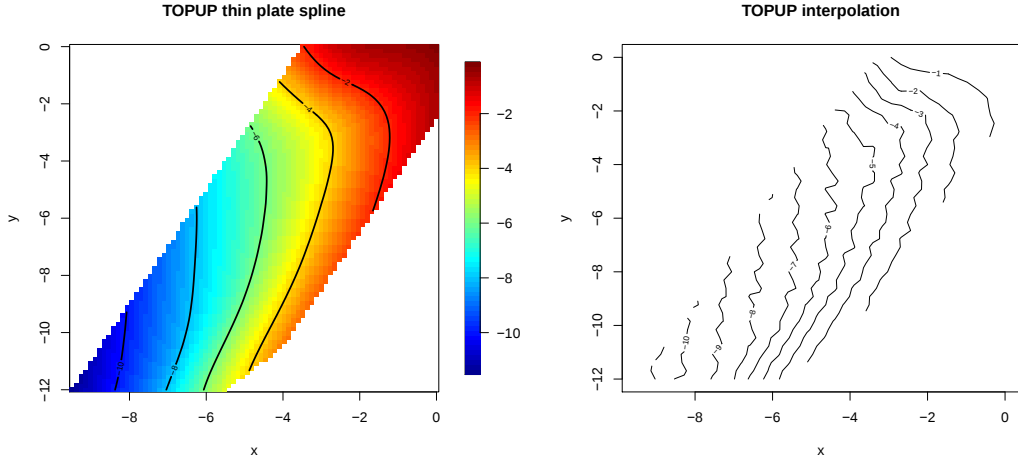


Figure 22: Estimated TOPUP risk in Example 1. The heatmap of the thin plate spline fit (left) and the interpolation contour (right) show the base-2 logarithm of the average of $|\sin(\text{angle}(\hat{\mathbf{u}}_1, \mathbf{u}_1))| = \|\mathbf{u}_1 \mathbf{u}_1^\top - \hat{\mathbf{u}}_1 \hat{\mathbf{u}}_1^\top\|_S$ via the TOPUP over 100 repetitions and over all possible combinations of d_1, d_2, T, λ for the same x and y coordinates, with $x = \log_2(d^{1/2}/(T^{1/2}\lambda))$ and $y = \log_2((dd_2)^{1/2}/(T^{1/2}\lambda^2))$ being the logarithms of the two components of the rate in (87).