



Comments on “Factor Models for High-Dimensional Tensor Time Series”

Jialin Ouyang and Ming Yuan

Columbia University, New York, NY

We would like to first congratulate Professors Chen, Yang and Zhang for making yet another important contribution to the development of theory and methodology for analyzing tensor or multilinear array data. As the data we collect becomes bigger and bigger, and encodes more and more complex relationships, it is natural to organize them in the form of multilinear arrays instead of the classical data matrices. How to exploit such structure in statistical analysis, however, can be rather challenging. This is especially the case for time series where temporal correlation adds another layer of complication. The elegant methodology developed in this timely contribution will certainly benefit a multitude of applications. To further understand the operating characteristics of the two proposed estimation procedures, TIPUP and TOPUP, we shall consider here the simplest case of rank-one matrix time series, and reflect upon the comparison between their performance. In doing so, we also gained insights into the effect of the hyper-parameter, namely the time lag h_0 .

In particular, consider

$$X_t = \lambda uv^\top f_t + E_t \in \mathbb{R}^{d \times d}, \quad (1)$$

where $u, v \in \mathbb{R}^d$, $\|u_1\|_2 = \|u_2\|_2 = 1$. For simplicity, we assume that $E_t \in \mathbb{R}^{d \times d}$ are independent Gaussian ensembles, where each entry of E_t follows the standard normal distribution, and that

$$\lim_{T \rightarrow \infty} \frac{1}{T-h} \sum_{t=h+1}^T f_{t-h} f_t = \gamma(h), \quad h = 0, 1, 2, \dots$$

1. TIPUP versus TOPUP

As shown in the article, under the rank-one model (1),

$$\begin{aligned} & \max\{\sin \angle(\hat{u}^{\text{TIPUP}}, u), \sin \angle(\hat{v}^{\text{TIPUP}}, v)\} \\ &= O_p\left(\frac{d}{\lambda\sqrt{T}} + \frac{\sqrt{d^3}}{\lambda^2\sqrt{T}}\right) \end{aligned}$$

and

$$\begin{aligned} & \max\{\sin \angle(\hat{u}^{\text{TOPUP}}, u), \sin \angle(\hat{v}^{\text{TOPUP}}, v)\} \\ &= O_p\left(\frac{\sqrt{d}}{\lambda\sqrt{T}} + \frac{d}{\lambda^2\sqrt{T}}\right) \end{aligned}$$

See, for example, Equations (42) and (49), and Appendix C.4. The aforementioned rates seemingly suggest that the difference in performance between TOPUP and TIPUP can be simply characterized by a multiplicative factor of $\sqrt{d}$. However, the empirical comparison between the two is sometimes more profound.

For illustration, we conducted a set of simulation studies where the core time series f_t follows a stationary AR(1) model:

$$f_t = \phi f_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1 - \phi^2).$$

We set $d = 20$, $T = 1000$, and $\phi = 0.5$. Two different levels of signal strength, $\lambda = 5$ or 30 , were considered. The empirical performance of TIPUP and TOPUP, based on 100 simulation runs, is given in Figure 1. Figure 1 shows the superiority of TOPUP over TIPUP, and the advantage is much more evident for $\lambda = 5$ than $\lambda = 30$. This seems to suggest a more subtle difference than predicted by the aforementioned rates of convergence. To better understand the operating characteristics, we took a close look at both estimating procedures and found that their estimation error under the rank-one model can be more precisely characterized, at least for sufficiently strong signal, as follows.

Denote by

$$\mathbf{y}_{h_0} = [\gamma(1), \gamma(2), \dots, \gamma(h_0)]^\top$$

and

$$\Gamma_{h_0} = \begin{bmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(h_0 - 1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(h_0 - 2) \\ & & \dots & \\ \gamma(h_0 - 1) & \gamma(h_0 - 2) & \dots & \gamma(0) \end{bmatrix}$$

Define

$$R(h_0) := \sqrt{\frac{d}{T}} \frac{\sqrt{\lambda^2 \mathbf{y}_{h_0}^\top \Gamma_{h_0} \mathbf{y}_{h_0} + \sum_{h=1}^{h_0} \gamma(h)^2}}{\lambda^2 \sum_{h=1}^{h_0} \gamma(h)^2},$$

and

$$R^*(h_0) := \sqrt{\frac{d}{T}} \frac{\sqrt{\lambda^2 \mathbf{y}_{h_0}^\top \Gamma_{h_0} \mathbf{y}_{h_0} + d \sum_{h=1}^{h_0} \gamma(h)^2}}{\lambda^2 \sum_{h=1}^{h_0} \gamma(h)^2},$$

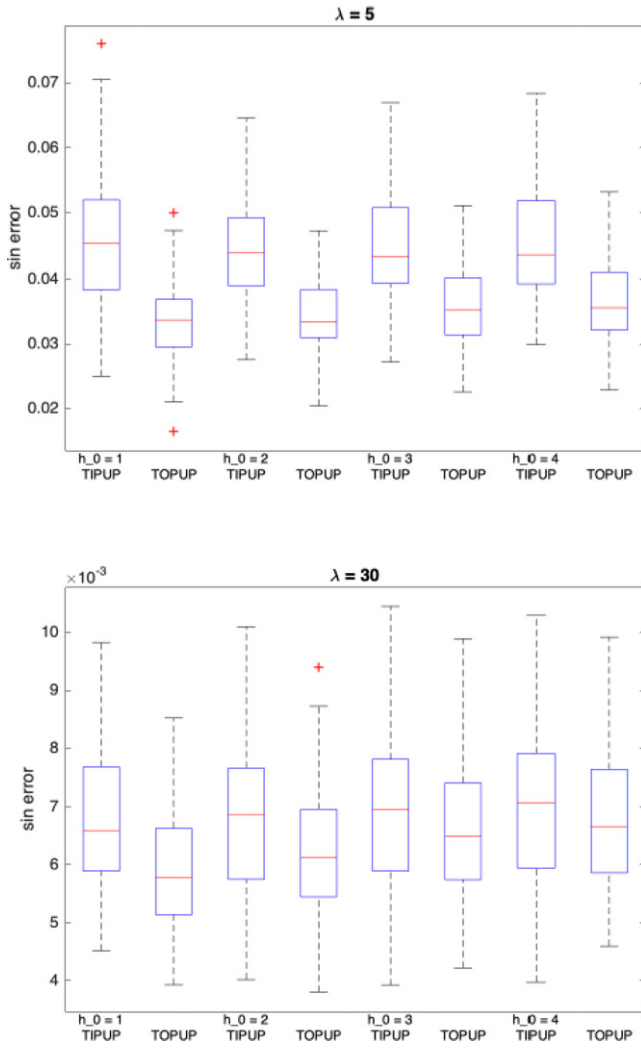


Figure 1. Comparison between TIPUP and TOPUP: each boxplot was summarized based on 100 simulation runs.

We have

Proposition 1. If $\lambda \gg d^{3/2}/T^{1/2} + d^{3/4}/T^{1/4} + d^{5/6}/T^{1/6}$, then $\max\{\sin \angle(\hat{u}^{\text{TOPUP}}, u), \sin \angle(\hat{v}^{\text{TOPUP}}, v)\} = [1 + o_p(1)]R(h_0)$.

If $\lambda \gg d^{1/2}/T^{1/4}$, then

$$\max\{\sin \angle(\hat{u}^{\text{TIPUP}}, u), \sin \angle(\hat{v}^{\text{TIPUP}}, v)\} = [1 + o_p(1)]R^*(h_0),$$

These bounds further confirm that the asymptotic error of TOPUP estimator is no greater than that of TIPUP. The difference is evident when $\lambda \lesssim \sqrt{d}$ and can be as much as a multiplicative factor of $\sqrt{d}$, that is, when $\lambda = o(1)$. However, when $\lambda \gg \sqrt{d}$, $\lambda^2 \mathbf{y}_{h_0}^\top \Gamma_{h_0} \mathbf{y}_{h_0}$ dominates $d \sum_{h=1}^{h_0} \gamma(h)^2$, so that $R^*(h_0) = R(h_0)(1 + o(1))$ and the two estimates are actually comparable!

We expanded the earlier numerical experiment to demonstrate the accuracy of the rates above: We set $d = 20$, $T = 2000$, and $\lambda = 4$ and run 100 simulations for each choice of h_0 . The results were presented in Figure 2. For reference, we also plotted $R^*(h) \sim h$ and $R(h) \sim h$ in the figure. In particular, $R^*(h_0)$ is minimized when $h_0 = 2$ whereas $R(h_0)$ at $h_0 = 1$.

2. Effect of h_0

Proposition 1 also provides insights to an important practical question for either TIPUP or TOPUP: does it matter which h_0 to choose? In general, we want to choose an h_0 to minimize $R(h_0)$ or $R^*(h_0)$ so that the estimate has the best rate of convergence. But how much difference does it make?

It is instructive to again consider the case when the core time series f_t follows an AR(1) model:

$$f_t = \phi f_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1 - \phi^2).$$

In this case, $\gamma(h) = \phi^{|h|}$ for $h = 0, \pm 1, \pm 2, \dots$ and thus

$$\begin{aligned} R(h_0) &= \sqrt{\frac{d}{T}} \frac{\sqrt{\lambda^2 \sum_{h=1}^{h_0} (2h-1)\phi^{2h} + \sum_{h=1}^{h_0} \phi^{2h}}}{\lambda^2 \sum_{h=1}^{h_0} \phi^{2h}} \\ &= \sqrt{\frac{d}{T}} \left(\sqrt{\frac{1}{\lambda} \cdot \frac{\phi^2 + \phi^4 - (2h_0+1)\phi^{2h_0+2} + (2h_0-1)\phi^{2h_0+4}}{(\phi^2 - \phi^{2h_0+2})^2}} + \frac{1}{\lambda^2} \cdot \frac{1 - \phi^2}{\phi^2 - \phi^{2h_0+2}} \right). \end{aligned}$$

and

$$\begin{aligned} R^*(h_0) &= \sqrt{\frac{d}{T}} \frac{\sqrt{\lambda^2 \sum_{h=1}^{h_0} (2h-1)\phi^{2h} + d \sum_{h=1}^{h_0} \phi^{2h}}}{\lambda^2 \sum_{h=1}^{h_0} \phi^{2h}} \\ &= \sqrt{\frac{d}{T}} \left(\sqrt{\frac{1}{\lambda} \cdot \frac{\phi^2 + \phi^4 - (2h_0+1)\phi^{2h_0+2} + (2h_0-1)\phi^{2h_0+4}}{(\phi^2 - \phi^{2h_0+2})^2}} + \frac{d}{\lambda^2} \cdot \frac{1 - \phi^2}{\phi^2 - \phi^{2h_0+2}} \right). \end{aligned}$$

When the signal is weak, that is, $\lambda = o(1)$, both are dominated by the second term on the rightmost hand side. As both terms are decreasing with h_0 suggesting that in this case, it is preferable to choose a large h_0 . In general, however, $R(h_0)$

or $R^*(h_0)$ may be optimized by a nontrivial choice of h_0 . To better comprehend their behavior, we focus on R^* and plotted in Figures 3 and 4 how it changes with different choices of ϕ and λ in the rank-one model:

The fact that h_0 may play a significant role in determining the performance of TIPUP or TOPUP suggests it is worthwhile to investigate more automatic and data-driven choices for it.

3. Summary

In this discussion, we further compare the two estimating procedures developed by Professors Chen, Yang and Zhang. By focusing on a simple rank-one model, we found that the com-

parison between TIPUP and TOPUP is more complex than it appears to be. In particular, TIPUP can even yield performance comparable to TOPUP in the presence of a strong signal. The exercise also reveals the effect of h_0 for both procedures and points to the importance of choosing h_0 at least in some settings. These somewhat surprising observations suggest that the operating characteristics of both estimating procedures may warrant further investigation.

Appendix A: Sketch of Proof for Proposition 1

Both TIPOP and TOPUP can be analyzed in identical fashion and we shall therefore, only discuss TOPUP. The main strategy is to decompose $\hat{u} \in \mathbb{R}^d$ into

$$\hat{u} = [1 + o_p(1)]u + \Delta_1 + \Delta_2,$$

and show that

$$\|\Delta_1\|_2 = [1 + o_p(1)]R(h_0), \quad (\text{A.1})$$

and

$$\|\Delta_2\|_2 = O_p\left(\frac{d^2}{\lambda^2 T} + \frac{d^3}{\lambda^4 T}\right). \quad (\text{A.2})$$

Under the assumption of the signal strength, we have $\|\Delta_2\|_2 = o_p[R(h_0)]$, which then completes the proof.

A.1. Decomposition of $\hat{u}$

Recall that $\hat{u}$ is the leading singular vector of

$$M := [\text{mat}_1(V_h), h = 1, \dots, h_0].$$

Under model (1),

$$\begin{aligned} V_h = & \frac{1}{T} \sum_{t=h+1}^T \lambda^2 uv^\top \otimes uv^\top f_{t-h} f_t + \frac{1}{T} \sum_{t=h+1}^T \lambda uv^\top f_{t-h} E_t \\ & + \frac{1}{T} \sum_{t=h+1}^T E_{t-h} \otimes \lambda uv^\top f_t + \frac{1}{T} \sum_{t=h+1}^T E_{t-h} \otimes E_t. \end{aligned} \quad (\text{A.3})$$

Denote $\hat{\gamma}(h) := \frac{1}{T} \sum_{t=h+1}^T f_{t-h} f_t$, then

$$\begin{aligned} \text{mat}_1(V_{1,h}) &= \lambda^2 u(v \odot u \odot v)^\top \hat{\gamma}(h) + \lambda u \odot v \odot \text{vec}\left(\frac{1}{T} \sum_{t=h+1}^T f_{t-h} E_t\right) \end{aligned}$$

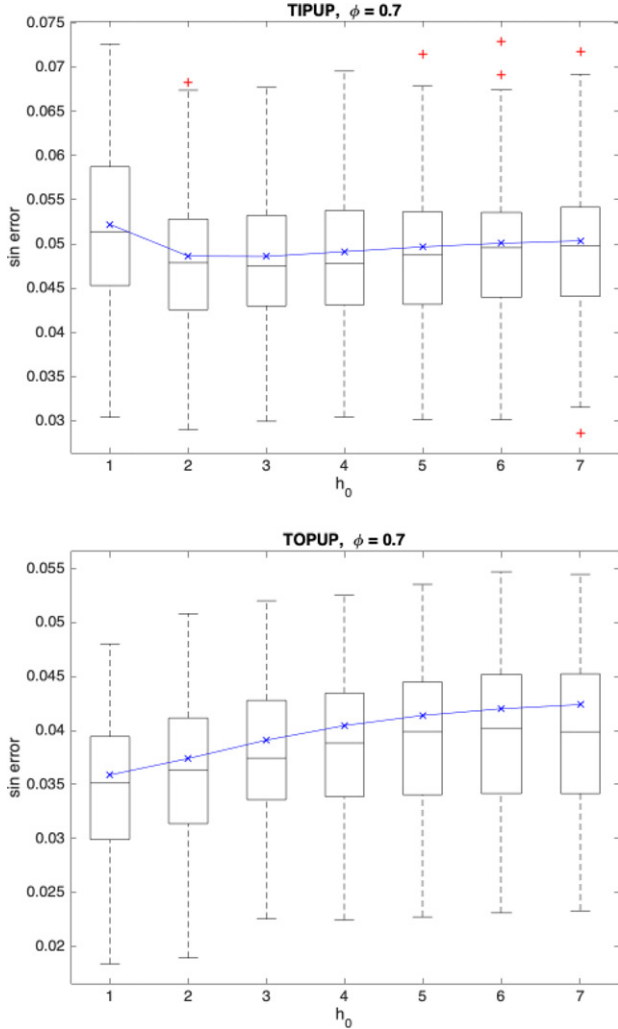


Figure 2. Choice of h_0 for TIPUP and TOPUP: in each panel the line and crosses correspond to $R^*(h_0)$ or $R(h_0)$ for each value of h_0 .

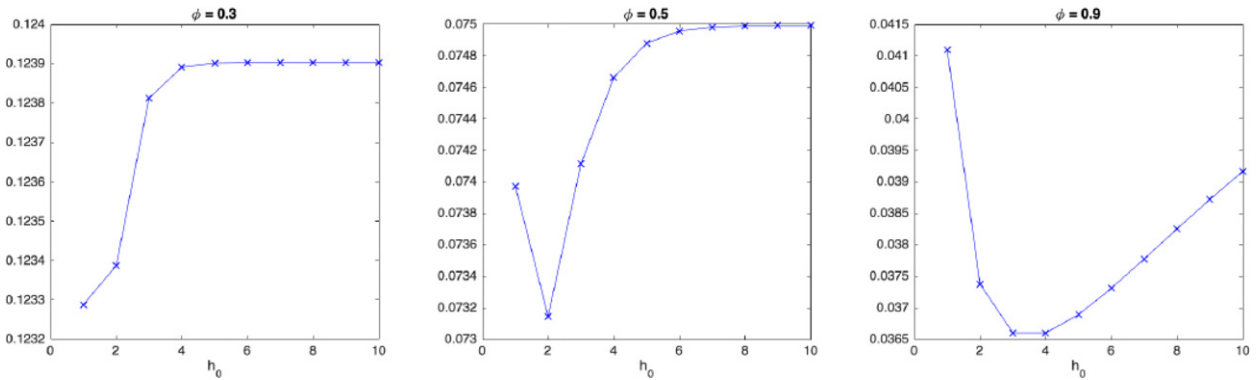


Figure 3. $R^*(h_0)$ for different ϕ : in each panel the line and crosses correspond to $R^*(h_0) \sim h_0$. In this example, we set $\lambda = 5$.

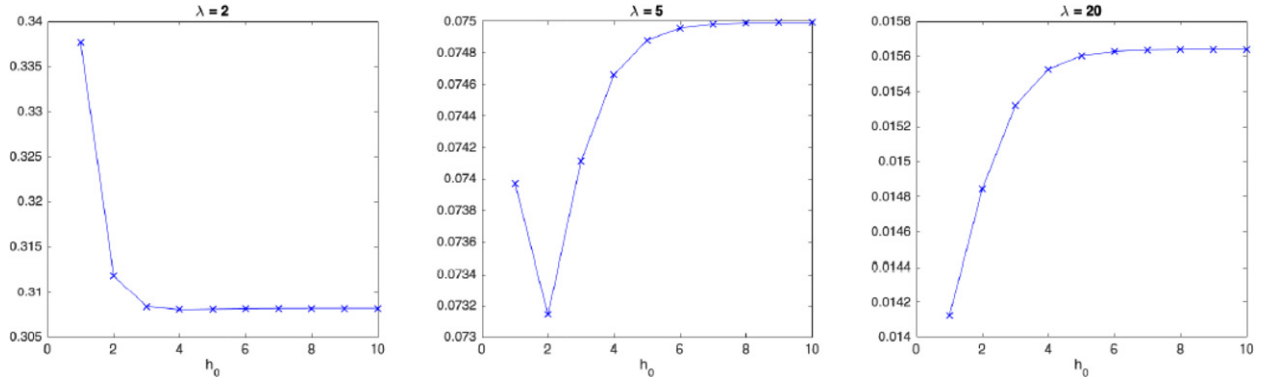


Figure 4. $R^*(h_0)$ for different λ : in each panel the line and crosses correspond to $R^*(h_0) \sim h_0$. In this example we set $\phi = 0.5$.

$$+ \lambda \left(\frac{1}{T} \sum_{t=h+1}^T E_{t-h} f_t \right) \odot u \odot v + \left(\frac{1}{T} \sum_{t=h+1}^T E_{t-h} \odot \text{vec}(E_t) \right),$$

$$:= M_{h,0} + M_{h,1} + M_{h,2} + M_{h,3}, \quad (\text{A.4})$$

where $\odot$ denotes the Kronecker product. Let

$$M_j := [M_{h,j}, h = 1, \dots, h_0], \quad j = 0, 1, 2, 3.$$

Observe that M_0 is rank-1 with singular value $\sqrt{\sum_{h=1}^{h_0} \hat{\gamma}(h)^2}$, and right singular vector

$$w := \frac{(v \odot u \odot v) \odot [\hat{\gamma}(1), \hat{\gamma}(2), \dots, \hat{\gamma}(h_0)]^\top}{\sqrt{\sum_{h=1}^{h_0} \hat{\gamma}(h)^2}}, \quad (\text{A.5})$$

and following from the proof of Theorem 1 in the article,

$$\|M_1\|_S + \|M_2\|_S + \|M_3\|_S = O_p \left(\frac{\lambda d}{T^{1/2}} + \frac{d^{3/2}}{T^{1/2}} \right). \quad (\text{A.6})$$

The leading right singular vector of $M = M_0 + M_1 + M_2 + M_3$ can be expressed as:

$$\hat{w} = [1 + o_p(1)]w + \Delta_w$$

where $\Delta_w \perp w$ and

$$\|\Delta_w\|_2 = O_p \left(\frac{d}{\lambda T^{1/2}} + \frac{d^{3/2}}{\lambda^2 T^{1/2}} \right). \quad (\text{A.7})$$

Since $\hat{u}$ is the leading singular vector of M , we get

$$\begin{aligned} \hat{u} &= \frac{M\hat{w}}{\|M\hat{w}\|_2} \\ &= \frac{(M_0 + M_1)\hat{w}}{\|M\hat{w}\|_2} + \frac{(M_2 + M_3)\hat{w}}{\|M\hat{w}\|_2} \\ &= \frac{M_0\hat{w} + M_1\hat{w}}{\|M\hat{w}\|_2} + \frac{[1 + o_p(1)](M_2 + M_3)w}{\|M\hat{w}\|_2} + \frac{(M_2 + M_3)\Delta_w}{\|M\hat{w}\|_2} \\ &:= \tilde{u} + \Delta_1 + \Delta_2 \end{aligned} \quad (\text{A.8})$$

the third equality follows from (A.7).

Observe that $M_0\Delta_w = 0$ (since $\Delta_w \perp w$), we have

$$M_0\hat{w} = [1 + o_p(1)]M_0w = u\lambda^2 \sqrt{\sum_{h=1}^{h_0} \hat{\gamma}(h)^2}.$$

Combine with bound (A.6), we immediately have

$$\|M\hat{w}\|_2 = [1 + o_p(1)]\lambda^2 \sqrt{\sum_{h=1}^{h_0} \hat{\gamma}(h)^2}. \quad (\text{A.9})$$

Since M_1 is also rank-1 with u as the left singular vector, applying bound (A.6) again, we have

$$\tilde{u} = [1 + o_p(1)]u. \quad (\text{A.10})$$

A.2. Bounding Δ_s

By definition,

$$\begin{aligned} \Delta_1 &= \frac{[1 + o_p(1)](M_2 + M_3)w}{\|M\hat{w}\|_2} \\ &= \frac{[1 + o_p(1)]}{\lambda^2 \sum_{h=1}^{h_0} \hat{\gamma}(h)^2} \sum_{h=1}^{h_0} \hat{\gamma}(h)(M_{2,h} + M_{3,h})(v \odot u \odot v) \\ &= \frac{[1 + o_p(1)]}{\lambda^2 \sum_{h=1}^{h_0} \hat{\gamma}(h)^2} \sum_{h=1}^{h_0} \hat{\gamma}(h) \\ &\quad \times \left[\lambda \left(\frac{1}{T} \sum_{t=h+1}^T f_t E_{t-h} v \right) + \left(\frac{1}{T} \sum_{t=h+1}^T E_{t-h} v u^\top E_t v \right) \right], \end{aligned}$$

let $\varepsilon_t = E_t v \in \mathbb{R}^d$, then

$$\begin{aligned} \Delta_1 &= \frac{[1 + o_p(1)]}{\lambda^2 \sum_{h=1}^{h_0} \hat{\gamma}(h)^2} \sum_{h=1}^{h_0} \hat{\gamma}(h) \\ &\quad \times \left[\lambda \left(\frac{1}{T} \sum_{t=h+1}^T f_t \varepsilon_{t-h} \right) + \left(\frac{1}{T} \sum_{t=h+1}^T \varepsilon_{t-h} (u^\top \varepsilon_t) \right) \right]. \end{aligned} \quad (\text{A.11})$$

Conditional on $\{f_t\}$, the entries of

$$y := \sum_{h=1}^{h_0} \hat{\gamma}(h) \left[\lambda \left(\frac{1}{\sqrt{T}} \sum_{t=h+1}^T f_t \varepsilon_{t-h} \right) + \left(\frac{1}{\sqrt{T}} \sum_{t=h+1}^T \varepsilon_{t-h} (u^\top \varepsilon_t) \right) \right]$$

has mean 0 and variance

$$\mathbb{E}y_j^2 = [1 + o(1)] \left(\lambda^2 \hat{\gamma}_{h_0}^\top \hat{\Gamma}_{h_0} \hat{\gamma}_{h_0} + \sum_{h=1}^{h_0} \hat{\gamma}(h)^2 \right), \quad j = 1, 2, \dots, d.$$

as $T \rightarrow \infty$. Equation (A.1) follows from the fact that $\text{var}(\sum_{j=1}^d y_j^2) = O(d^2/T) = o(d^2)$ and

$$\hat{\gamma}(h)^2 = [1 + o_p(1)]\gamma(h)^2.$$

Finally, (A.2) follows by combining (A.6), (A.7) and (A.9).

Funding

Research supported in part by NSF grants DMS-2015285 and DMS-2052955.