

HAR-Itô models and high-dimensional vector HAR modeling for high-frequency data

Huiling Yuan¹, Kexin Lu², Yifeng Guo², Lan Li², and Guodong Li²

¹ *Donghua University, and* ² *University of Hong Kong*

Supplementary Material

This online supplementary material contains seven sections. Section S1 introduces tensor notation and the Tucker decomposition. Section S2 presents the technical proof of Proposition 1 in Section 2. Section S3 provides the proofs of Theorems 1 and 2 in Section 4, and Section S4 presents four auxiliary lemmas.

Lemma S1 establishes the covering number and discretization properties of low-multilinear-rank tensors. Lemma S2 establishes the restricted strong convexity (RSC) and restricted smoothness (RSM) conditions, and Lemma S3 derives the deviation bound; these results are used in the proofs of Theorems 1 and 2. Lemma S4 establishes the contractive projection property (CPP), which is used in the proof of Theorem 2.

Section S5 introduces the univariate HAR-Itô model and the corresponding proposition. Section S6 reports the selected ranks for the MLR-HAR and VHARI models in the real data analysis. Section S7 presents the definition and rolling estimates of the coefficient-based total spillover index for the VHAR, VHARI, and MLR-HAR models.

Throughout this paper, C denotes a generic positive constant whose value may vary from line to line.

S1 Tensor notations and Tucker decomposition

This section briefly introduces tensor notation and the Tucker decomposition, and a detailed review of tensor notations and operations can be found in Kolda and Bader (2009).

Tensors, also known as multidimensional arrays, are higher-order extensions of matrices, and a multidimensional array $\mathcal{A} \in \mathbb{R}^{p_1 \times \cdots \times p_K}$ is called a K -th-order tensor, where the order of a tensor is known as the way or mode. This paper concentrates on third-order tensors.

For a third-order tensor $\mathcal{A} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$, its element is denoted by $\mathcal{A}_{ijk}$ for $1 \leq i \leq p_1, 1 \leq j \leq p_2$, and $1 \leq k \leq p_3$, and the Frobenius norm is defined as $\|\mathcal{A}\|_F = \left(\sum_{i=1}^{p_1} \sum_{j=1}^{p_2} \sum_{k=1}^{p_3} X_{ijk}^2 \right)^{1/2}$.

Matricization, also known as unfolding or flattening, involves rearranging the elements of a higher-order tensor into a matrix format. Mode-1 matricization treats the first mode of the tensor as the rows of the matrix, while collapsing all other modes into the columns. Specifically, the element at position (i_1, i_2, i_3) in the tensor $\mathcal{A}$ corresponds to the (i_1, j) -th element in the matricized form $\mathcal{A}_{(1)} \in \mathbb{R}^{p_1 \times p_2 p_3}$, where $j = i_2 + (i_3 - 1)p_2$. Similarly, mode-2 and mode-3 matricizations can be defined. Matricizing tensors allows for establishing connections between matrix concepts and properties with those of tensors. The mode-1 multiplication, denoted as $\times_1$, between a tensor $\mathcal{A}$ and a matrix $\mathbf{B} \in \mathbb{R}^{q_1 \times p_1}$, is defined as follows

$$\mathcal{A} \times_1 \mathbf{B} = \left(\sum_{i=1}^{p_1} \mathcal{A}_{ijk} \mathbf{B}_{si} \right)_{1 \leq s \leq q_1, 1 \leq j \leq p_2, 1 \leq k \leq p_3}.$$

The mode- k multiplication, denoted as $\times_k$, where $k = 2, 3$, can be defined similarly.

The multilinear ranks of a tensor $\mathcal{A}$ are defined as (r_1, r_2, r_3) , where

$$r_1 = \text{rank}(\mathcal{A}_{(1)}), \quad r_2 = \text{rank}(\mathcal{A}_{(2)}), \quad \text{and} \quad r_3 = \text{rank}(\mathcal{A}_{(3)}).$$

Accordingly, there exists a Tucker decomposition (Tucker, 1966; De Lathauwer et al., 2000):

$$\mathcal{A} = \mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3 = \mathcal{G} \times_{i=1}^3 \mathbf{U}_i, \quad (\text{S1.1})$$

where $\mathbf{U}_i \in \mathbb{R}^{p_i \times r_i}$ for $i = 1, 2, 3$ represent the factor matrices, and $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ denotes the core tensor. Alternatively, this decomposition can be represented as $\mathcal{A} = \llbracket \mathcal{G}; \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3 \rrbracket$.

Note that the Tucker decomposition is not unique, since

$$\llbracket \mathcal{G}; \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3 \rrbracket = \llbracket \mathcal{G} \times_1 \mathbf{O}_1 \times_2 \mathbf{O}_2 \times_3 \mathbf{O}_3; \mathbf{U}_1 \mathbf{O}_1^{-1}, \mathbf{U}_2 \mathbf{O}_2^{-1}, \mathbf{U}_3 \mathbf{O}_3^{-1} \rrbracket.$$

for any invertible matrices $\mathbf{O}_i \in \mathbb{R}^{r_i \times r_i}$ with $1 \leq i \leq 3$. As a special case of the Tucker decomposition, the higher-order singular value decomposition (HOSVD) is defined by choosing $\mathbf{U}_1$, $\mathbf{U}_2$ and $\mathbf{U}_3$ as the tall matrices that consist of the top (r_1, r_2, r_3) left singular vectors of $\mathcal{A}_{(1)}$, $\mathcal{A}_{(2)}$ and $\mathcal{A}_{(3)}$ respectively, and then $\mathcal{G} = \mathcal{A} \times_1 \mathbf{U}_1^\top \times_2 \mathbf{U}_2^\top \times_3 \mathbf{U}_3^\top$. As a result, factor matrices $\mathbf{U}_1$, $\mathbf{U}_2$ and $\mathbf{U}_3$ are all orthonormal, and $\mathcal{G}$ possesses the all-orthogonal property, i.e., for each $1 \leq j \leq 3$, the rows of $\mathcal{G}_{(j)}$ are pairwise orthogonal. Note that the HOSVD is still not unique unless we impose more constraints (Wang et al., 2021).

The three ranks, r_1, r_2 and r_3 are not equal in general. In particular, when $r_3 = p_3$, the multilinear ranks of $\mathcal{A}$ are denoted by (r_1, r_2) instead, omitting the rank of mode-3 matricization. The multilinear ranks are also known as Tucker ranks, as they are closely related to the Tucker decomposition. There are many other tensor decomposition methods,

such as CP decomposition (Kolda and Bader, 2009), and the ranks of a tensor can be defined in many different ways.

S2 Proof of Proposition 1 in Section 2

Proof of Proposition 1. We define

$$R_i(k) = \int_{n-1}^n \frac{(n-t)^k}{k!} \sigma_{i,t}^2 dt = \frac{\sigma_{i,n-1}^2}{(k+1)!} + \int_{n-1}^n \frac{(n-t)^{k+1}}{(k+1)!} d\sigma_{i,t}^2, \quad 1 \leq i \leq N,$$

for all $k \geq 0$ and $n \geq P+1$. From Definition 1 and Itô's lemma, we have

$$\begin{aligned} d\sigma_{i,t}^2 &= \left(\omega_i + v_i(n-t) - \sigma_{i,n-1}^2 - v_i Z_{i,t}^2 + \sum_{j=1}^N \alpha_{i,j}^{(1)} \sigma_{j,t}^2 \right) dt + 2v_i(n-t) Z_{i,t} dZ_{i,t} \\ &\quad + \beta_i L_{i,t}^2 d\Lambda_{i,t} \end{aligned}$$

for $n-1 < t \leq n$, and

$$v_i \int_{n-1}^n \frac{(n-t)^{k+1}}{(k+1)!} Z_{i,t}^2 dt = \frac{v_i}{(k+3)!} + 2v_i \int_{n-1}^n \frac{(n-t)^{k+2}}{(k+2)!} Z_{i,t} dZ_{i,t}.$$

As a result, for any $k \geq 0$, let

$$\begin{aligned} R_i^{(1)}(k) &= 2v_i \int_{n-1}^n \frac{(k+1)(n-t)^{k+2}}{(k+2)!} Z_{i,t} dZ_{i,t}, & R_i^{(2)}(k) &= \beta_i \int_{n-1}^n \frac{(n-t)^{k+1}}{(k+1)!} L_{i,t}^2 d\Lambda_{i,t} \\ R_i^{(3)}(k) &= \frac{(k+3)w_i + (k+1)v_i}{(k+3)!}, & \text{and} & & R_i^{(4)}(k) &= \frac{(k+1)\sigma_{i,n-1}^2}{(k+2)!}, \end{aligned}$$

and then it holds that

$$\begin{aligned}
R_i(k) &= \frac{\omega_i}{(k+2)!} + \frac{(k+2)v_i}{(k+3)!} + \frac{(k+1)\sigma_{i,n-1}^2}{(k+2)!} \\
&\quad + 2v_i \int_{n-1}^n \frac{(n-t)^{k+2}}{(k+1)!} Z_{i,t} dZ_{i,t} - v_i \int_{n-1}^n \frac{(n-t)^{k+1}}{(k+1)!} Z_{i,t}^2 dt \\
&\quad + \beta_i \int_{n-1}^n \frac{(n-t)^{k+1}}{(k+1)!} L_{i,t}^2 d\Lambda_{i,t} + \sum_{j=1}^N \alpha_{i,j}^{(1)} R_j(k+1) \\
&= R_i^{(1)}(k) + R_i^{(2)}(k) + R_i^{(3)}(k) + R_i^{(4)}(k) + \sum_{j=1}^N \alpha_{i,j}^{(1)} R_j(k+1),
\end{aligned}$$

which can be rewritten into

$$\mathbf{R}(k) = \mathbf{R}^{(1)}(k) + \mathbf{R}^{(2)}(k) + \mathbf{R}^{(3)}(k) + \mathbf{R}^{(4)}(k) + \boldsymbol{\alpha}^{(1)} \mathbf{R}(k+1), \quad (\text{S2.1})$$

where $\mathbf{R}(k) = (R_1(k), R_2(k), \dots, R_N(k))^\top$ and $\mathbf{R}^{(j)}(k) = (R_1^{(j)}(k), R_2^{(j)}(k), \dots, R_N^{(j)}(k))^\top$ with $1 \leq j \leq 4$.

Due to the facts that $R_i(k) \leq \int_{n-1}^n \sigma_{i,t}^2 dt$ and the spectral radius of $\boldsymbol{\alpha}^{(1)}$ is less than one, it holds that $\boldsymbol{\alpha}^{(1)k} \mathbf{R}(k) \rightarrow 0$ almost surely as $k \rightarrow \infty$. Moreover, by iterating the formula at (S2.1), we have

$$\begin{aligned}
\mathbf{R}(0) &= \int_{n-1}^n \boldsymbol{\sigma}_t^2 dt = \sum_{k=0}^{\infty} \boldsymbol{\alpha}^{(1)k} \mathbf{R}^{(1)}(k) + \sum_{k=0}^{\infty} \boldsymbol{\alpha}^{(1)k} \mathbf{R}^{(2)}(k) + \sum_{k=0}^{\infty} \boldsymbol{\alpha}^{(1)k} \mathbf{R}^{(3)}(k) \\
&\quad + \sum_{k=0}^{\infty} \boldsymbol{\alpha}^{(1)k} \mathbf{R}^{(4)}(k)
\end{aligned} \quad (\text{S2.2})$$

with probability one. It can be verified that

$$\begin{aligned}
\sum_{k=0}^{\infty} \boldsymbol{\alpha}^{(1)k} \mathbf{R}_1(k) &= \sum_{k=0}^{\infty} \boldsymbol{\alpha}^{(1)k} \left[2v_i \int_{n-1}^n \frac{(k+1)(n-t)^{k+2}}{(k+2)!} Z_{i,t} dZ_{i,t} \right]_{i=1, \dots, N}^{\top} \\
&= 2\mathbf{V} \boldsymbol{\alpha}^{(1)-2} \int_{n-1}^n \left\{ ((n-t)\boldsymbol{\alpha}^{(1)} - \mathbf{I}_N) \mathbf{e}^{(n-t)\boldsymbol{\alpha}^{(1)}} + \mathbf{I}_N \right\} \mathbf{Z}_t d\mathbf{Z}_t \\
&:= \boldsymbol{\varepsilon}_n^c, \\
\sum_{k=0}^{\infty} \boldsymbol{\alpha}^{(1)k} \mathbf{R}_2(k) &= \sum_{k=0}^{\infty} \boldsymbol{\alpha}^{(1)k} \int_{n-1}^n \left[\beta_i \frac{(n-t)^{k+1}}{(k+1)!} L_{i,t}^2 d\Lambda_{i,t} \right]_{i=1, \dots, N}^{\top} \\
&= \boldsymbol{\varrho}_2 \boldsymbol{\beta} \boldsymbol{\omega}_L \boldsymbol{\lambda} + \boldsymbol{\varepsilon}_n^J, \\
\sum_{k=0}^{\infty} \boldsymbol{\alpha}^{(1)k} \mathbf{R}_3(k) &= \boldsymbol{\varrho}_2 \boldsymbol{\omega} + (\boldsymbol{\varrho}_2 - 2\boldsymbol{\varrho}_3) \mathbf{V} \mathbf{1}_N \quad \text{and} \\
\sum_{k=0}^{\infty} \boldsymbol{\alpha}^{(1)k} \mathbf{R}_4(k) &= (\boldsymbol{\varrho}_1 - \boldsymbol{\varrho}_2) \boldsymbol{\sigma}_{n-1}^2,
\end{aligned} \tag{S2.3}$$

where

$$\boldsymbol{\varepsilon}_n^J = \boldsymbol{\beta} \boldsymbol{\alpha}^{(1)-1} \left\{ \int_{n-1}^n \left(\mathbf{e}^{\boldsymbol{\alpha}^{(1)}(n-t)} - \mathbf{I}_N \right) [\mathbf{M}_t d\boldsymbol{\Lambda}_t + \boldsymbol{\omega}_L (d\boldsymbol{\Lambda}_t - \boldsymbol{\lambda} dt)] \right\}.$$

Note that $\boldsymbol{\sigma}_t^2$ at integer time point n has the form of

$$\boldsymbol{\sigma}_n^2 = \boldsymbol{\omega} + \sum_{l=1}^P \boldsymbol{\alpha}^{(l)} \mathbf{y}_{n-l+1} + \boldsymbol{\beta} \mathbf{J}_n,$$

and we then have the autoregressive form at (2.3).

Note that random variables $\boldsymbol{\varepsilon}_n^c$ and $\boldsymbol{\varepsilon}_n^J$ are both independent of the σ -field $\mathcal{F}_{n-1}$, and hence $\{\boldsymbol{\varepsilon}_n^c\}$ and $\{\boldsymbol{\varepsilon}_n^J\}$ are martingale difference sequences (Song et al., 2021), i.e. $E(\boldsymbol{\varepsilon}_n^c) = E(\boldsymbol{\varepsilon}_n^J) = 0$. We further show that both $\{\boldsymbol{\varepsilon}_n^c\}$ and $\{\boldsymbol{\varepsilon}_n^J\}$ are *i.i.d.* sequences with mean zero and finite variance matrices.

Let $\boldsymbol{\varrho}_4(t) = ((n-t)\boldsymbol{\alpha}^{(1)} - \mathbf{I}_N) \mathbf{e}^{(n-t)\boldsymbol{\alpha}^{(1)}} + \mathbf{I}_N$, and then $|\boldsymbol{\varrho}_4(t)| < 2\mathbf{e} + \mathbf{I}_N$ for all $n-1 < t \leq n$. Moreover, $\mathbf{Z}_t = \int_{[t]}^t d\mathbf{W}_t = \mathbf{W}_t - \mathbf{W}_{n-1}$, $d\mathbf{Z}_t = d\mathbf{W}_t$ and $(d\mathbf{Z}_t)(d\mathbf{Z}_t)^{\top} = \mathbf{I}_N dt$. As a

result, by Itô isometry,

$$\begin{aligned} E(\boldsymbol{\varepsilon}_n^c)^2 &= E \left(2\mathbf{V}\boldsymbol{\alpha}^{(1)-2} \int_{n-1}^n \boldsymbol{\varrho}_4(t) \mathbf{Z}_t d\mathbf{Z}_t \right)^2 = 4\mathbf{V}^2 \boldsymbol{\alpha}^{(1)-4} E \int_{n-1}^n [\boldsymbol{\varrho}_4(t)]^2 \mathbf{Z}_t (d\mathbf{Z}_t) (d\mathbf{Z}_t)^\top \mathbf{Z}_t^\top \\ &< 4\mathbf{V}^2 \boldsymbol{\alpha}^{(1)-4} (2\mathbf{e} + \mathbf{I}_N)^2 E \int_{n-1}^n \mathbf{Z}_t \mathbf{Z}_t^\top dt. \end{aligned} \quad (\text{S2.4})$$

Due to Itô lemma and definition of $\mathbf{Z}_t$, it holds that $d\mathbf{Z}_t \mathbf{Z}_t^\top = \mathbf{Z}_t d\mathbf{Z}_t^\top + d\mathbf{Z}_t (\mathbf{Z}_t^\top) + \mathbf{I}_N dt$,

$\mathbf{Z}_{n-1} = \mathbf{0}$, both $\{\int_{n-1}^t \mathbf{Z}_s d\mathbf{Z}_s^\top\}$ and $\{\int_{n-1}^t d\mathbf{Z}_s (\mathbf{Z}_s^\top)\}$ are martingale difference sequences. Thus,

$$\mathbf{Z}_t \mathbf{Z}_t^\top = \mathbf{Z}_t \mathbf{Z}_t^\top - \mathbf{Z}_{n-1} (\mathbf{Z}_{n-1})^\top = \int_{n-1}^t (\mathbf{Z}_s d\mathbf{Z}_s^\top + d\mathbf{Z}_s (\mathbf{Z}_s^\top) + \mathbf{I}_N ds), \text{ and}$$

$$E \int_{n-1}^n \mathbf{Z}_t \mathbf{Z}_t^\top dt = E \left(\int_{n-1}^n \int_{n-1}^t \mathbf{Z}_s d\mathbf{Z}_s^\top dt \right) + E \left(\int_{n-1}^n \int_{n-1}^t (d\mathbf{Z}_s^\top) (\mathbf{Z}_s) dt \right) + 0.5\mathbf{I}_N = 0.5\mathbf{I}_N,$$

which, together with (S2.4), implies that $\text{var}(\boldsymbol{\varepsilon}_n) < 2\mathbf{V}^2 \boldsymbol{\alpha}^{(1)-4} (2\mathbf{e} + \mathbf{I}_N)^2 < \infty$.

We next consider $\boldsymbol{\varepsilon}_n^J$. Let $\boldsymbol{\varrho}_5(t) = \mathbf{e} \boldsymbol{\alpha}^{(1)(n-t)} - \mathbf{I}_N$, and then $|\boldsymbol{\varrho}_5(t)| < \mathbf{e} + \mathbf{I}_N$ for all $n-1 < t \leq n$. From (S2.3) and the fact that $E(\boldsymbol{\varepsilon}_n^J) = \mathbf{0}$, it holds that

$$\text{var}(\boldsymbol{\varepsilon}_n^J) = \beta^2 \boldsymbol{\alpha}^{(-1)-2} E \left(\int_{n-1}^n \boldsymbol{\varrho}_5(t) \mathbf{L}_t^2 d\Lambda_t \right) \left(\int_{n-1}^n \boldsymbol{\varrho}_5(t) \mathbf{L}_t^2 d\Lambda_t \right)^\top - (\boldsymbol{\varrho}_2 \beta \boldsymbol{\omega}_L \boldsymbol{\lambda}) (\boldsymbol{\varrho}_2 \beta \boldsymbol{\omega}_L \boldsymbol{\lambda})^\top. \quad (\text{S2.5})$$

Note that $\{\mathbf{L}_t\}$ are independent of the Poisson process $\{\Lambda_t\}$ and, by the partitioning method,

we have

$$\begin{aligned} E \left(\int_{n-1}^n \boldsymbol{\varrho}_5(t) \mathbf{L}_t^2 d\Lambda_t \right)^2 &= E \left(\sum_{n-1 < t \leq n} \boldsymbol{\varrho}_5(t) \mathbf{L}_t^2 \Delta \Lambda_t \right) \left(\sum_{n-1 < t \leq n} \boldsymbol{\varrho}_5(t) \mathbf{L}_t^2 \Delta \Lambda_t \right)^\top \\ &< (\mathbf{e} + \mathbf{I}_N)^2 E \left(\sum_{n-1 < t \leq n} \mathbf{L}_t^2 \Delta \Lambda_t \right) \left(\sum_{n-1 < t \leq n} \mathbf{L}_t^2 \Delta \Lambda_t \right)^\top \\ &= (\mathbf{e} + \mathbf{I}_N)^2 E(\mathbf{L}_t^2 \mathbf{L}_t^{2\top}) E \left(\sum_{n-1 < t \leq n} \Delta(\text{diag}(\Lambda_t^2)) \right) \\ &= (\mathbf{e} + \mathbf{I}_N)^2 (\text{diag}(\lambda_i^2 + \lambda_i)) E(\mathbf{L}_t^2 (\mathbf{L}_t^2)^\top), \quad 1 \leq i \leq N, \end{aligned}$$

which, together with (S2.5) and the fact that $E(\mathbf{L}_t^2(\mathbf{L}_t^2)^\top) < \infty$, implies that $\text{var}(\boldsymbol{\epsilon}_n^J) < \infty$.

□

S3 Proofs of Theorems 1 and 2 in Section 4

Proof of Theorem 1. For simplicity, denote the multilinear low-rank estimator $\hat{\mathcal{A}}_{\text{MLR}}$ by $\hat{\mathcal{A}}$, and let $\boldsymbol{\Delta} = \hat{\mathcal{A}} - \mathcal{A}^*$, where $\mathcal{A}^*$ is the true parameter tensor. The loss function has the form of

$$L(\mathcal{A}) = \frac{1}{T} \sum_{n=1}^T \|\tilde{\mathbf{y}}_n - \mathcal{A}_{(1)} \tilde{\mathbf{x}}_n\|_2^2,$$

where $\tilde{\mathbf{x}}_n = (\tilde{\mathbf{y}}_{n-1}^\top, \dots, \tilde{\mathbf{y}}_{n-P}^\top)^\top \in \mathbb{R}^{NP \times 1}$. Due to the optimality of $\hat{\mathcal{A}}$, it holds that

$$\frac{1}{T} \sum_{n=1}^T \|\tilde{\mathbf{y}}_n - \hat{\mathcal{A}}_{(1)} \tilde{\mathbf{x}}_n\|_2^2 \leq \frac{1}{T} \sum_{n=1}^T \|\tilde{\mathbf{y}}_n - \mathcal{A}_{(1)}^* \tilde{\mathbf{x}}_n\|_2^2,$$

which implies that

$$\frac{1}{T} \sum_{n=1}^T \|\boldsymbol{\Delta}_{(1)} \tilde{\mathbf{x}}_n\|_2^2 \leq \frac{2}{T} \sum_{n=1}^T \langle \boldsymbol{\epsilon}_n, \boldsymbol{\Delta}_{(1)} \tilde{\mathbf{x}}_n \rangle \leq 2 \langle \frac{1}{T} \sum_{n=1}^T \boldsymbol{\epsilon}_n \circ \tilde{\mathbf{X}}_n, \boldsymbol{\Delta} \rangle, \quad (\text{S3.1})$$

where $\tilde{\mathbf{X}}_n = (\tilde{\mathbf{y}}_{n-1}, \dots, \tilde{\mathbf{y}}_{n-P}) \in \mathbb{R}^{N \times P}$, $\sum_{n=1}^T \langle \boldsymbol{\epsilon}_n, \boldsymbol{\Delta}_{(1)} \tilde{\mathbf{x}}_n \rangle = \langle \sum_{n=1}^T \boldsymbol{\epsilon}_n \circ \tilde{\mathbf{X}}_n, \boldsymbol{\Delta} \rangle$, and $\circ$ denotes the outer product.

Denote the set of tensors

$$\mathcal{S}(r_1, r_2, r_3) = \{\mathcal{A} \in \mathbb{R}^{N \times N \times P} : \|\mathcal{A}\|_{\text{F}} = 1, \text{rank}_i(\mathcal{A}_{(i)}) \leq r_i, 1 \leq i \leq 3\}.$$

Note that the Tucker ranks of both $\hat{\mathcal{A}}$ and $\mathcal{A}^*$ are (r_1, r_2, r_3) , and hence the Tucker ranks of $\boldsymbol{\Delta}$ are at most $(2r_1, 2r_2, 2r_3)$. As a result, from (S3.1),

$$\frac{1}{T} \sum_{n=1}^T \|\boldsymbol{\Delta}_{(1)} \tilde{\mathbf{x}}_n\|_2^2 \leq 2 \|\boldsymbol{\Delta}\|_{\text{F}} \sup_{\boldsymbol{\Delta} \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \langle \frac{1}{T} \sum_{n=1}^T \boldsymbol{\epsilon}_n \circ \tilde{\mathbf{X}}_n, \boldsymbol{\Delta} \rangle,$$

and therefore we can derive the estimation error bound by applying Lemmas S2 and Lemma S3 with δ being $\delta + 1/2$ for simplicity. The prediction error bound can also be established from the above inequality, estimation error bound and Lemma S3. $\square$

Proof of Theorem 2. For a fixed $1 \leq k \leq K$, define a linear space,

$$\mathcal{A} = \{\alpha_1 \hat{\mathcal{A}}_k + \alpha_2 \mathcal{A}, \quad \alpha_1, \alpha_2 \in \mathbb{R}\},$$

and denote by $(\mathcal{B})_{\mathcal{A}}$ the projection of $\mathcal{B} \in \mathbb{R}^{N \times N \times P}$ onto the space $\mathcal{A}$, where the dependence of $\mathcal{A}$ on k is suppressed for simplicity. Since $\mathcal{A} \in \Theta(r_1, r_2, r_3)$ and $\hat{\mathcal{A}}_k \in \Theta(r'_1, r'_2, r'_3)$, it holds that $\mathcal{A} \subset \Theta(r_1 + r'_1, r_2 + r'_2, r_3 + r'_3)$.

Note that $\tilde{\mathcal{A}}_k = \hat{\mathcal{A}}_{k-1} - \eta \nabla L(\hat{\mathcal{A}}_{k-1})$, and $r'_i \geq \left(\sqrt[3]{1 + \frac{\kappa_{L,A}}{24\kappa_{U,A}}} - 1\right)^{-2} r_i$ with $1 \leq i \leq 3$.

From Lemma S4, we have

$$\|\hat{\mathcal{A}}_k - (\tilde{\mathcal{A}}_k)_{\mathcal{A}}\|_{\text{F}} \leq \left[\prod_{i=1}^3 \left(\sqrt{\frac{r'_i}{r_i}} + 1\right) - 1\right] \|\mathcal{A} - (\tilde{\mathcal{A}}_k)_{\mathcal{A}}\|_{\text{F}} \leq \frac{\kappa_{L,A}}{24\kappa_{U,A}} \|\mathcal{A} - (\tilde{\mathcal{A}}_k)_{\mathcal{A}}\|_{\text{F}},$$

which, together with the fact that $1 < 1 + \frac{\kappa_{L,A}}{24\kappa_{U,A}} < 2$, implies that

$$\begin{aligned} \|\hat{\mathcal{A}}_k - \mathcal{A}\|_{\text{F}} &\leq \|\hat{\mathcal{A}}_k - (\tilde{\mathcal{A}}_k)_{\mathcal{A}}\|_{\text{F}} + \|\mathcal{A} - (\tilde{\mathcal{A}}_k)_{\mathcal{A}}\|_{\text{F}} \leq \left(1 + \frac{\kappa_{L,A}}{24\kappa_{U,A}}\right) \|\mathcal{A} - (\tilde{\mathcal{A}}_k)_{\mathcal{A}}\|_{\text{F}} \\ &\leq \left(1 + \frac{\kappa_{L,A}}{24\kappa_{U,A}}\right) \|(\mathcal{A} - \hat{\mathcal{A}}_{k-1} - \eta[\nabla L(\mathcal{A}) - \nabla L(\hat{\mathcal{A}}_{k-1})])_{\mathcal{A}}\|_{\text{F}} + 2\eta \|(\nabla L(\mathcal{A}))_{\mathcal{A}}\|_{\text{F}} \\ &:= A_1 + A_2, \end{aligned} \tag{S3.2}$$

where $A_1 = \left(1 + \frac{\kappa_{L,A}}{24\kappa_{U,A}}\right) \|(\mathcal{A} - \hat{\mathcal{A}}_{k-1} - \eta[\nabla L(\mathcal{A}) - \nabla L(\hat{\mathcal{A}}_{k-1})])_{\mathcal{A}}\|_{\text{F}}$, and $A_2 = 2\eta \|(\nabla L(\mathcal{A}))_{\mathcal{A}}\|_{\text{F}}$.

We first handle the term of A_1 . Let $\mathbf{H} = T^{-1} \sum_{n=1}^T (\tilde{\mathbf{x}}_n \tilde{\mathbf{x}}_n^{\top} \otimes \mathbf{I}_N)$, and it holds that $T^{-1} \sum_{n=1}^T \|\Delta_{(1)} \tilde{\mathbf{x}}_n\|_2^2 = T^{-1} \sum_{n=1}^T \|(\tilde{\mathbf{x}}_n^{\top} \otimes \mathbf{I}_N) \text{vec}(\Delta)\|_2^2 = \text{vec}(\Delta)^{\top} \mathbf{H} \text{vec}(\Delta)$. Then, from

Lemma S2 and for all $\Delta \in \Theta(r_1 + r'_1, r_2 + r'_2, r_3 + r'_3)$,

$$\frac{1}{8}\kappa_{L,A}\|\Delta\|_F^2 \leq \text{vec}(\Delta)^\top \mathbf{H} \text{vec}(\Delta) \leq \frac{8}{3}\kappa_{U,A}\|\Delta\|_F^2$$

with a probability at least

$$1 - 2\exp(-CT(\kappa_{L,A}/\kappa_{U,B})^2 \min\{\lambda^{-2}, \lambda^{-4}\}) - \exp(-Cd'_{\mathcal{M}}). \quad (\text{S3.3})$$

Note that $\eta = 2/(3\kappa_{U,A})$, and $\mathbf{H}$ is the Hessian matrix of the loss function $L(\mathbf{B})$ with respect to $\text{vec}(\mathbf{B})$. It holds that $\text{vec}(\nabla L(\mathcal{A}) - \nabla L(\hat{\mathcal{A}}_{k-1})) = \mathbf{H} \text{vec}(\mathcal{A} - \hat{\mathcal{A}}_{k-1})$, and

$$\begin{aligned} A_1 &= (1 + \frac{\kappa_{L,A}}{24\kappa_{U,A}}) \|((\mathbf{I} - \eta\mathbf{H})\text{vec}(\mathcal{A} - \hat{\mathcal{A}}_{k-1}))_{\mathcal{A}}\|_2 \\ &\leq (1 + \frac{\kappa_{L,A}}{24\kappa_{U,A}})(1 - \frac{\kappa_{L,A}}{12\kappa_{U,A}}) \|\hat{\mathcal{A}}_{k-1} - \mathcal{A}\|_F \\ &\leq (1 - \frac{\kappa_{L,A}}{24\kappa_{U,A}}) \|\hat{\mathcal{A}}_{k-1} - \mathcal{A}\|_F, \end{aligned} \quad (\text{S3.4})$$

with the probability at (S3.3), where $(\text{vec}(\mathbf{B}))_{\mathcal{A}} = (\mathbf{B})_{\mathcal{A}}$, and the first inequality is by Lemma 4 of Chen et al. (2019).

We next handle the term of A_2 and, by Lemma 5 in Chen et al. (2019),

$$A_2 \leq \frac{32}{\kappa_{L,A}} \|(\nabla L(\mathcal{A}))_{\mathcal{A}}\|_F = \frac{32}{\kappa_{L,A}} \sup_{\mathcal{S} \in \mathcal{A}, \|\mathcal{S}\|_F=1} \langle \nabla L(\mathcal{A}), \mathcal{S} \rangle \leq \frac{32}{\kappa_{L,A}} \xi,$$

where $\mathcal{S}(r_1 + r'_1, r_2 + r'_2, r_3 + r'_3) = \Theta(r_1 + r'_1, r_2 + r'_2, r_3 + r'_3) \cap \{\|\mathcal{S}\|_F = 1\}$ and

$$\begin{aligned} \xi &= \sup_{\mathcal{S} \in \mathcal{S}(r_1+r'_1, r_2+r'_2, r_3+r'_3)} \langle \nabla L(\mathcal{A}), \mathcal{S} \rangle = \sup_{\mathcal{S} \in \mathcal{S}(r_1+r'_1, r_2+r'_2, r_3+r'_3)} \left\langle \frac{2}{T} \sum_{n=1}^T \epsilon_n \circ \tilde{\mathbf{X}}_n, \mathcal{S} \right\rangle \\ &\leq C \left[(\lambda^2 \sqrt{\lambda_{\max}(\Sigma_{\varepsilon}) \kappa_{U,A}} + \lambda \sqrt{\kappa_{U,B}}) \sqrt{\frac{d'_{\mathcal{M}}}{T}} + \frac{T^{1+2\delta}}{m^{1/4}} \right] \end{aligned}$$

with a probability at least $1 - \exp(-Cd'_{\mathcal{M}}) - 2\exp(-CT(\kappa_{L,A}/\kappa_{U,B})^2 \min\{\lambda^{-2}, \lambda^{-4}\})$. This, together with (S3.2) and (S3.4), completes the proof. $\square$

S4 Auxiliary Lemmas for the Proof of Theorem 1 and 2

Lemma S1. (*Covering number and discretization of low-multilinear-rank tensors*). Suppose that $\bar{\mathcal{S}}(r_1, r_2, r_3)$ is an ϵ -net of the set $\mathcal{S}(r_1, r_2, r_3) := \{\mathcal{A} \in \mathbb{R}^{N \times N \times P} : \|\mathcal{A}\|_{\text{F}} = 1, \text{rank}_i(\mathcal{A}_{(i)}) \leq r_i, 1 \leq i \leq 3\}$.

(i) The cardinality of $\bar{\mathcal{S}}(r_1, r_2, r_3)$ satisfies

$$|\bar{\mathcal{S}}(r_1, r_2, r_3)| \leq (12/\epsilon)^{(r_1 r_2 r_3 + N r_1 + N r_2 + P r_3)}.$$

(ii) For any tensor $\mathcal{N} \in \mathbb{R}^{N \times N \times P}$ and matrix $\mathbf{Z} \in \mathbb{R}^{NP \times T}$, it holds that,

$$\begin{aligned} \sup_{\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \langle \mathcal{N}, \Delta \rangle &\leq (1 - 2\sqrt{2}\epsilon)^{-1} \max_{\bar{\Delta} \in \bar{\mathcal{S}}(2r_1, 2r_2, 2r_3)} \langle \mathcal{N}, \bar{\Delta} \rangle, \quad \text{and} \\ \sup_{\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \|\Delta_{(1)} \mathbf{Z}\|_{\text{F}} &\leq (1 - 2\sqrt{2}\epsilon)^{-1} \max_{\bar{\Delta} \in \bar{\mathcal{S}}(2r_1, 2r_2, 2r_3)} \|\bar{\Delta}_{(1)} \mathbf{Z}\|_{\text{F}}. \end{aligned}$$

Proof of Lemma S1. Result at (i) is from Lemma A.1 of Wang et al. (2020), and we here prove (ii) only.

Consider an ϵ -net $\bar{\mathcal{S}}(2r_1, 2r_2, 2r_3)$ for $\mathcal{S}(2r_1, 2r_2, 2r_3)$. Then for any tensor $\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)$, there exists a $\bar{\Delta} \in \bar{\mathcal{S}}(2r_1, 2r_2, 2r_3)$ such that $\|\Delta - \bar{\Delta}\|_{\text{F}} \leq \epsilon$. Since the rank of $\bar{\mathcal{W}} = \Delta - \bar{\Delta}$ are at most $(4r_1, 4r_2, 4r_3)$, we can split the HOSVD of $\bar{\mathcal{W}}$ into 8 parts such that $\bar{\mathcal{W}} = \sum_{i=1}^8 \bar{\mathcal{W}}_i$, where $\text{rank}_j(\bar{\mathcal{W}}_i) \leq 2r_j$ for $1 \leq i \leq 8$ and $1 \leq j \leq 3$, and $\langle \bar{\mathcal{W}}_j, \bar{\mathcal{W}}_k \rangle = 0$ for any $j \neq k$. Then for any $\mathcal{N} \in \mathbb{R}^{N \times N \times P}$, we have

$$\langle \mathcal{N}, \Delta \rangle = \langle \mathcal{N}, \bar{\Delta} \rangle + \sum_{i=1}^8 \langle \mathcal{N}, \bar{\mathcal{W}}_i \rangle = \langle \mathcal{N}, \bar{\Delta} \rangle + \sum_{i=1}^8 \langle \mathcal{N}, \bar{\mathcal{W}}_i / \|\bar{\mathcal{W}}_i\|_{\text{F}} \rangle \|\bar{\mathcal{W}}_i\|_{\text{F}}, \quad (\text{S4.1})$$

where $\bar{\mathcal{W}}_i / \|\bar{\mathcal{W}}_i\|_{\text{F}} \in \mathcal{S}(2r_1, 2r_2, 2r_3)$, and $\langle \mathcal{N}, \bar{\mathcal{W}}_i / \|\bar{\mathcal{W}}_i\|_{\text{F}} \rangle \leq \sup_{\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \langle \mathcal{N}, \Delta \rangle$.

Note that $\|\bar{\mathbf{W}}\|_{\text{F}}^2 = \sum_{i=1}^8 \|\bar{\mathbf{W}}_i\|_{\text{F}}^2$, and it holds that $\sum_{i=1}^8 \|\bar{\mathbf{W}}_i\|_{\text{F}} \leq 2\sqrt{2}\|\bar{\mathbf{W}}\|_{\text{F}} \leq 2\sqrt{2}\epsilon$, which, together with (S4.1), implies that

$$\begin{aligned} \gamma &:= \sup_{\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \langle \mathbf{N}, \Delta \rangle \leq \max_{\bar{\Delta} \in \bar{\mathcal{S}}(2r_1, 2r_2, 2r_3)} \langle \mathbf{N}, \bar{\Delta} \rangle + 2\sqrt{2}\gamma\epsilon, \quad \text{or} \\ \gamma &= \sup_{\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \langle \mathbf{N}, \Delta \rangle \leq (1 - 2\sqrt{2}\epsilon)^{-1} \max_{\bar{\Delta} \in \bar{\mathcal{S}}(2r_1, 2r_2, 2r_3)} \langle \mathbf{N}, \bar{\Delta} \rangle. \end{aligned}$$

For matrix $\mathbf{Z} \in \mathbb{R}^{NP \times T}$, it holds that

$$\begin{aligned} \|\Delta_{(1)}\mathbf{Z}\|_{\text{F}} &\leq \|\bar{\Delta}_{(1)}\mathbf{Z}\|_{\text{F}} + \sum_{i=1}^8 \|(\bar{\mathbf{W}}_i)_{(1)}\mathbf{Z}\|_{\text{F}} \\ &= \|\bar{\Delta}_{(1)}\mathbf{Z}\|_{\text{F}} + \sum_{i=1}^8 \|\bar{\mathbf{W}}_i\|_{\text{F}} \|(\bar{\mathbf{W}}_i)_{(1)}/\|\bar{\mathbf{W}}_i\|_{\text{F}} \mathbf{Z}\|_{\text{F}} \\ &\leq \|\bar{\Delta}_{(1)}\mathbf{Z}\|_{\text{F}} + \sum_{i=1}^8 \|\bar{\mathbf{W}}_i\|_{\text{F}} \sup_{\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \|\Delta_{(1)}\mathbf{Z}\|_{\text{F}} \\ &\leq \|\bar{\Delta}_{(1)}\mathbf{Z}\|_{\text{F}} + 2\sqrt{2}\epsilon \sup_{\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \|\Delta_{(1)}\mathbf{Z}\|_{\text{F}}, \end{aligned}$$

and we accomplish the proof by taking supremum on both sides. $\square$

Lemma S2. (*Restricted strong convexity and smoothness*). Suppose that (3.2) and Assumptions 1 and 2 hold. If $T \gtrsim \max(\lambda^2, \lambda^4)(\kappa_{U,B}/\kappa_{L,A})^2 d_{\mathcal{M}}$, $m^{1/4} \gtrsim T^{2\delta}$, $m^{1/4}T^{2\delta-1} \gtrsim N^2 P \exp(d_{\mathcal{M}})$ for some $\delta > 1/2$, then

$$\frac{1}{8}\kappa_{L,A} \|\Delta\|_{\text{F}}^2 \leq \frac{1}{T} \sum_{n=1}^T \|\Delta_{(1)}\tilde{\mathbf{x}}_n\|_2^2 \leq \frac{8}{3}\kappa_{U,A} \|\Delta\|_{\text{F}}^2,$$

for all $\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)$ with probability at least

$$1 - 2\exp(-CT(\kappa_{L,A}/\kappa_{U,B})^2 \min\{\lambda^{-2}, \lambda^{-4}\}) - \exp(-Cd_{\mathcal{M}}),$$

where κ , $\kappa_{L,A}$, $\kappa_{U,B}$ and $d_{\mathcal{M}}$ are defined in Theorem 1.

Proof of Lemma S2. Denote $R_T(\mathbf{\Delta}) = \sum_{n=1}^T \|\mathbf{\Delta}_{(1)} \tilde{\mathbf{x}}_n\|_2^2$, and it holds that

$$\begin{aligned} R_T(\mathbf{\Delta}) &= \sum_{n=1}^T \|\mathbf{\Delta}_{(1)}(\mathbf{x}_n + \mathbf{\vartheta}_n)\|_2^2 \\ &= \sum_{n=1}^T \mathbf{x}_n^\top \mathbf{\Delta}_{(1)}^\top \mathbf{\Delta}_{(1)} \mathbf{x}_n + 2 \sum_{n=1}^T \mathbf{x}_n^\top \mathbf{\Delta}_{(1)}^\top \mathbf{\Delta}_{(1)} \mathbf{\vartheta}_n + \sum_{n=1}^T \mathbf{\vartheta}_n^\top \mathbf{\Delta}_{(1)}^\top \mathbf{\Delta}_{(1)} \mathbf{\vartheta}_n \\ &:= R_1 + R_2 + R_3, \end{aligned} \quad (\text{S4.2})$$

where $\mathbf{\vartheta}_n = \tilde{\mathbf{x}}_n - \mathbf{x}_n = (\boldsymbol{\eta}_{n-1}^\top, \dots, \boldsymbol{\eta}_{n-P}^\top)^\top \in \mathbb{R}^{NP \times 1}$. By the spectral measure of ARMA processes in Basu and Michailidis (2015), we have $\lambda_{\min}\{\mathbb{E}(\mathbf{x}_n \mathbf{x}_n^\top)\} \geq \lambda_{\min}(\boldsymbol{\Sigma}_\varepsilon)/\mu_{\max}(\mathcal{A}^*) = \kappa_{L,A}$ and $\lambda_{\max}\{\mathbb{E}(\mathbf{x}_n \mathbf{x}_n^\top)\} \leq \lambda_{\max}(\boldsymbol{\Sigma}_\varepsilon)/\mu_{\min}(\mathcal{A}^*) = \kappa_{U,A}$, and it then holds that

$$T\kappa_{L,A} \leq \mathbb{E}(R_1) = \mathbb{E} \left(\sum_{n=1}^T \text{vec}(\mathbf{\Delta})^\top (\mathbf{I}_N \otimes \mathbf{x}_n \mathbf{x}_n^\top) \text{vec}(\mathbf{\Delta}) \right) \leq T\kappa_{U,A}, \quad (\text{S4.3})$$

as $\|\mathbf{\Delta}\|_F = 1$. Furthermore, $R_T(\mathbf{\Delta}) \geq \mathbb{E}(R_1) - \sup_{\mathbf{\Delta} \in \mathcal{S}} \{|R_1 - \mathbb{E}(R_1)| + |R_2| + |R_3|\}$, and $R_T(\mathbf{\Delta}) \leq \mathbb{E}(R_1) + \sup_{\mathbf{\Delta} \in \mathcal{S}} \{|R_1 - \mathbb{E}(R_1)| + |R_2| + |R_3|\}$, where $\mathcal{S} = \mathcal{S}(2r_1, 2r_2, 2r_3)$. We next first bound $|R_1 - \mathbb{E}(R_1)| + |R_2| + |R_3|$ for each fixed $\mathbf{\Delta} \in \mathcal{S}(2r_1, 2r_2, 2r_3)$.

Consider the term of $R_1 - \mathbb{E}(R_1)$. Note that $\mathbf{x}_n = (\mathbf{y}_{n-1}^\top, \dots, \mathbf{y}_{n-P}^\top)^\top \in \mathbb{R}^{NP \times 1}$, and $\mathcal{A}_{(1)} = (\mathbf{A}_1, \dots, \mathbf{A}_P)$ with each $\mathbf{A}_i$ being an N -by- N matrix. From (3.3), we have $\mathbf{y}_n = \mathcal{A}_{(1)} \mathbf{x}_n + \boldsymbol{\varepsilon}_n$. It can be further rewritten into an VAR(1) form, $\mathbf{x}_n = \mathbf{B} \mathbf{x}_{n-1} + \mathbf{e}_n$, and hence the VMA representation of $\mathbf{x}_n = \sum_{j=0}^{\infty} \mathbf{B}^j \mathbf{e}_{n-j}$, or $\mathbf{z} = \mathbf{P} \mathbf{e}$, where

$$\mathbf{B} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{P-1} & \mathbf{A}_P \\ \mathbf{I}_N & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_N & \mathbf{0} \end{pmatrix}, \quad \mathbf{P} = \begin{pmatrix} \mathbf{I}_{NP} & \mathbf{B} & \mathbf{B}^2 & \cdots & \mathbf{B}^{T-1} & \cdots \\ \mathbf{0} & \mathbf{I}_{NP} & \mathbf{B} & \cdots & \mathbf{B}^{T-2} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{NP} & \cdots & \mathbf{B}^{T-3} & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_{NP} & \cdots \end{pmatrix},$$

$\mathbf{e}_n = (\boldsymbol{\varepsilon}_n^\top, \dots, \mathbf{0})^\top \in \mathbb{R}^{NP \times 1}$, $\mathbf{e} = (\mathbf{e}_{T-1}^\top, \dots, \mathbf{e}_P^\top, \dots)^\top$, and $\mathbf{z} = (\mathbf{x}_T^\top, \dots, \mathbf{x}_1^\top)^\top \in \mathbb{R}^{NPT \times 1}$.

Moreover, by Assumption 2, the error term has the form of $\mathbf{e} = \bar{\Sigma} \bar{\xi}$, where $\bar{\xi}_n = (\xi_n^\top, \dots, \mathbf{0})^\top \in \mathbb{R}^{NP \times 1}$, $\bar{\xi} = (\bar{\xi}_{T-1}^\top, \dots, \bar{\xi}_P^\top, \dots)^\top$,

$$\bar{\Sigma}_\varepsilon = \begin{pmatrix} \Sigma_\varepsilon & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Sigma_\varepsilon & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \Sigma_\varepsilon \end{pmatrix} \in \mathbb{R}^{NP \times NP} \quad \text{and} \quad \bar{\Sigma} = \begin{pmatrix} \bar{\Sigma}_\varepsilon^{1/2} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \bar{\Sigma}_\varepsilon^{1/2} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \bar{\Sigma}_\varepsilon^{1/2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Let $\kappa_{U,B} = \lambda_{\max}(\Sigma_\varepsilon)/\mu_{\min}(\mathbf{B}^*)$, $\Sigma_\Delta = \bar{\Sigma} \mathbf{P}^\top (\mathbf{I}_T \otimes \Delta_{(1)}^\top \Delta_{(1)}) \mathbf{P} \bar{\Sigma}$, and then

$$\mathbf{R}_1 = \sum_{n=1}^T \mathbf{x}_n^\top \Delta_{(1)}^\top \Delta_{(1)} \mathbf{x}_n = \bar{\xi}^\top \bar{\Sigma} \mathbf{P}^\top (\mathbf{I}_T \otimes \Delta_{(1)}^\top \Delta_{(1)}) \mathbf{P} \bar{\Sigma} \bar{\xi} = \bar{\xi}^\top \Sigma_\Delta \bar{\xi}.$$

Note that $\lambda_{\max}(\mathbf{P} \mathbf{P}^\top) = 1/\mu_{\min}(\mathbf{B}^*)$, then $\|\Sigma_\Delta\|_{\text{op}} \leq \kappa_{U,B}$ and

$$\|\Sigma_\Delta\|_{\text{F}} \leq \|\bar{\Sigma}\|_{\text{op}}^2 \|\mathbf{P}\|_{\text{op}} \|\mathbf{P}^\top\|_{\text{op}} \|\mathbf{I}_T \otimes \Delta_{(1)}^\top \Delta_{(1)}\|_{\text{F}} = \sqrt{T} \kappa_{U,B} \leq \sqrt{T} \kappa_{U,B}.$$

For any $t > 0$, by Hanson-Wright inequality for sub-Gaussian random variables, we can

bound $\mathbf{R}_1 - \mathbb{E}(\mathbf{R}_1)$ below,

$$\begin{aligned} \mathbb{P}[|\mathbf{R}_1 - \mathbb{E}(\mathbf{R}_1)| \geq t] &\leq 2 \exp \left(-C \min \left(\frac{t}{\lambda^2 \|\Sigma_\Delta\|_{\text{op}}}, \frac{t^2}{\lambda^4 \|\Sigma_\Delta\|_{\text{F}}^2} \right) \right) \\ &\leq 2 \exp \left(-C \min \left(\frac{t}{\lambda^2 \kappa_{U,B}}, \frac{t^2}{\lambda^4 T \kappa_{U,B}^2} \right) \right). \end{aligned} \tag{S4.4}$$

We next consider the term of $\mathbf{R}_2$, which has the form of

$$\mathbf{R}_2 = 2\mathbf{z}^\top (\mathbf{I}_T \otimes \Delta_{(1)}^\top \Delta_{(1)}) \mathbf{c} = 2\bar{\xi}^\top \bar{\Sigma} \mathbf{P}^\top (\mathbf{I}_T \otimes \Delta_{(1)}^\top \Delta_{(1)}) \mathbf{c} = 2\mathbf{m}_1^\top \bar{\xi},$$

where $\mathbf{m}_1 = \bar{\Sigma} \mathbf{P}^\top (\mathbf{I}_T \otimes \Delta_{(1)}^\top \Delta_{(1)}) \mathbf{c}$, $\mathbf{c} = (\boldsymbol{\vartheta}_T^\top, \dots, \boldsymbol{\vartheta}_1^\top)^\top \in \mathbb{R}^{NPT \times 1}$. Let $\eta_{\max} = \max_{1 \leq n \leq T}$

$\max_{1 \leq i \leq N} |\eta_{i,n}|$, and it holds that $\|\mathbf{c}\|_2^2 \leq NPT \eta_{\max}^2$ and $\|\mathbf{m}_1\|_2^2 \leq NPT \eta_{\max}^2 \kappa_{U,B}$. Moreover,

by the inequality at (3.2) and Markov inequality, we have

$$\mathbb{P} \left\{ \eta_{\max} \geq \frac{x}{m^{1/4}} \right\} \leq \frac{CNT}{x^2}, \quad (\text{S4.5})$$

which can be used to control the rate of $\eta_{\max}$ by varying the value of x . By the sub-Gaussian condition of $\bar{\boldsymbol{\xi}}$ at Assumption 2 and letting $x = m^{1/8}T^\delta/N^{1/2}P^{1/2}$ in (S4.5),

$$\mathbb{P} [2 |\mathbf{m}_1^\top \bar{\boldsymbol{\xi}}| \geq t] \leq 2 \exp \left(-\frac{Cm^{1/4}t^2}{\lambda^2 T^{1+2\delta} \kappa_{U,B}} \right) + \frac{CN^2P}{m^{1/4}T^{2\delta-1}} \quad (\text{S4.6})$$

for any $t > 0$.

Finally, for the third term of R_3 at (S4.2), it holds that

$$\sup_{\boldsymbol{\Delta} \in \mathcal{S}(2r_1, 2r_2, 2r_3)} |R_3| = \sup_{\boldsymbol{\Delta} \in \mathcal{S}(2r_1, 2r_2, 2r_3)} |\mathbf{c}^\top (\mathbf{I}_T \otimes \boldsymbol{\Delta}_{(1)}^\top \boldsymbol{\Delta}_{(1)}) \mathbf{c}| \leq NPT\eta_{\max}^2,$$

and, by letting $x = m^{1/8}T^\delta/N^{1/2}P^{1/2}$, we have

$$\mathbb{P} \left[\sup_{\boldsymbol{\Delta} \in \mathcal{S}(2r_1, 2r_2, 2r_3)} |R_3| \geq \frac{T^{1+2\delta}}{m^{1/4}} \right] \leq \frac{CN^2P}{m^{1/4}T^{2\delta-1}}. \quad (\text{S4.7})$$

Thus, combining (S4.4), (S4.6) and (S4.7), we have

$$\begin{aligned} & \mathbb{P} \left[|\mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3 - \mathbb{E}(\mathbf{R}_1)| \geq t_1 + t_2 + \frac{T^{1+2\delta}}{m^{1/4}} \right] \\ & \leq \mathbb{P} [|\mathbf{R}_1 - \mathbb{E}(\mathbf{R}_1)| \geq t_1] + \mathbb{P} [|\mathbf{R}_2| \geq t_2] + \mathbb{P} \left[|\mathbf{R}_3| \geq \frac{T^{1+2\delta}}{m^{1/4}} \right] \\ & \leq 2 \exp \left(-C \min \left(\frac{t_1}{\lambda^2 \kappa_{U,B}}, \frac{t_1^2}{\lambda^4 T \kappa_{U,B}^2} \right) \right) + 2 \exp \left(-\frac{Cm^{1/4}t^2}{\lambda^2 T^{1+2\delta} \kappa_{U,B}} \right) + \frac{CN^2P}{m^{1/4}T^{2\delta-1}}. \end{aligned}$$

Let $t_1 = t_2 = T\kappa_{L,A}/6$ and, from (S4.3), it holds that

$$\begin{aligned} & \mathbb{P} [0.5\kappa_{L,A} \leq T^{-1}R_T(\boldsymbol{\Delta}) \leq 1.5\kappa_{U,A}] \\ & \geq 1 - 2 \exp \left(-C \min \left(\frac{T\kappa_{L,A}}{\lambda^2 \kappa_{U,B}}, \frac{T\kappa_{L,A}^2}{\lambda^4 \kappa_{U,B}^2} \right) \right) - 2 \exp \left(-\frac{C\kappa_{L,A}^2 m^{1/4}}{T^{2\delta-1} \lambda^2 \kappa_{U,B}} \right) - \frac{CN^2P}{m^{1/4}T^{2\delta-1}}, \end{aligned} \quad (\text{S4.8})$$

as $m^{1/4} \geq 6T^{2\delta}/\kappa_{L,A}$. Let $\bar{\mathcal{S}}$ to be an ϵ -covering net of $S(2r_1, 2r_2, 2r_3)$. To construct the union bound, we rewrite $R_T(\Delta)$ as $R_T(\Delta) = \|\Delta_{(1)}\tilde{\mathbf{X}}\|_{\text{F}}^2$, where $\tilde{\mathbf{X}} = (\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_T) \in \mathbb{R}^{NP \times T}$.

Define the event

$$\mathcal{E}(\epsilon) = \left\{ \forall \Delta \in \bar{\mathcal{S}}(2r_1, 2r_2, 2r_3) : \sqrt{0.5\kappa_{L,A}} \leq \frac{1}{\sqrt{T}} \|\Delta_{(1)}\tilde{\mathbf{X}}\|_{\text{F}} \leq \sqrt{1.5\kappa_{U,A}} \right\}.$$

Then, by the pointwise bound in (S4.8) and the covering number in Lemma S1 (i),

$$\begin{aligned} \mathbb{P}[\mathcal{E}^c(\epsilon)] &\leq 2 \exp \left(Cd_{\mathcal{M}} - C \min \left(\frac{T\kappa_{L,A}}{\lambda^2\kappa_{U,B}}, \frac{T\kappa_{L,A}^2}{\lambda^4\kappa_{U,B}^2} \right) \right) \\ &\quad + 2 \exp \left(Cd_{\mathcal{M}} - \frac{C\kappa_{L,A}^2 m^{1/4}}{T^{2\delta-1}\lambda^2\kappa_{U,B}} \right) + \exp(Cd_{\mathcal{M}}) \frac{CN^2P}{m^{1/4}T^{2\delta-1}}. \end{aligned}$$

Note that, by Lemma S1 (ii),

$$\begin{aligned} \mathcal{E}(\epsilon) &\subset \left\{ \max_{\Delta \in \bar{\mathcal{S}}(2r_1, 2r_2, 2r_3)} \frac{1}{\sqrt{T}} \|\Delta_{(1)}\tilde{\mathbf{X}}\|_{\text{F}} \leq \sqrt{1.5\kappa_{U,A}} \right\} \\ &\subset \left\{ \sup_{\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \frac{1}{\sqrt{T}} \|\Delta_{(1)}\tilde{\mathbf{X}}\|_{\text{F}} \leq \frac{\sqrt{1.5\kappa_{U,A}}}{1 - 2\sqrt{2}\epsilon} \right\}. \end{aligned}$$

Moreover, similarly to Lemma S1(ii), we can show that

$$\begin{aligned} \|\Delta_{(1)}\mathbf{Z}\|_{\text{F}} &\geq \|\bar{\Delta}_{(1)}\mathbf{Z}\|_{\text{F}} - \sum_{i=1}^8 \|(\bar{\mathbf{W}}_i)_{(1)}\mathbf{Z}\|_{\text{F}} \\ &\geq \|\bar{\Delta}_{(1)}\mathbf{Z}\|_{\text{F}} - \sum_{i=1}^8 \|\bar{\mathbf{W}}_i\|_{\text{F}} \sup_{\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \|\Delta_{(1)}\mathbf{Z}\|_{\text{F}} \\ &\geq \|\bar{\Delta}_{(1)}\mathbf{Z}\|_{\text{F}} - 2\sqrt{2}\epsilon \sup_{\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \|\Delta_{(1)}\mathbf{Z}\|_{\text{F}}, \end{aligned}$$

where the last inequality is due to $\sum_{i=1}^8 \|\bar{\mathbf{W}}_i\|_{\text{F}} \leq 2\sqrt{2}\epsilon$. Taking infimum on both sides, if

$0 \leq \epsilon \leq \frac{1}{4\sqrt{2}}$, we have

$$\begin{aligned} \inf_{\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \frac{1}{\sqrt{T}} \|\Delta_{(1)} \mathbf{Z}\|_F &\geq \min_{\Delta \in \bar{\mathcal{S}}(2r_1, 2r_2, 2r_3)} \frac{1}{\sqrt{T}} \|\Delta_{(1)} \mathbf{Z}\|_F \\ &\quad - 2\sqrt{2}\epsilon \frac{1}{\sqrt{T}} \sup_{\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \|\Delta_{(1)} \mathbf{Z}\|_F \\ &\geq \sqrt{0.5\kappa_{L,A}} - 2\sqrt{2}\epsilon \frac{\sqrt{1.5\kappa_{U,A}}}{1 - 2\sqrt{2}\epsilon} \geq \sqrt{0.5\kappa_{L,A}} - 4\epsilon\sqrt{3\kappa_{U,A}}. \end{aligned}$$

When ϵ is chosen to be $\frac{1}{8}\sqrt{\frac{\kappa_{L,A}}{6\kappa_{U,A}}}$,

$$\mathcal{E}(\epsilon) \subset \left\{ \inf_{\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \frac{1}{\sqrt{T}} \|\Delta_{(1)} \tilde{\mathbf{X}}\|_F \geq \frac{\sqrt{0.5\kappa_{L,A}}}{2} \right\},$$

As a result, with the above choice of ϵ ,

$$\mathcal{E}(\epsilon) \subset \left\{ \frac{\kappa_{L,A}}{8} \leq \inf_{\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \frac{1}{T} \|\Delta_{(1)} \tilde{\mathbf{X}}\|_F^2 \leq \sup_{\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \frac{1}{T} \|\Delta_{(1)} \tilde{\mathbf{X}}\|_F^2 \leq \frac{8\kappa_{U,A}}{3} \right\}.$$

Given the conditions that $T \gtrsim (\kappa_{U,B}/\kappa_{L,A})^2 \max(\lambda^2, \lambda^4) d_{\mathcal{M}}$, $m^{1/4} \gtrsim T^{2\delta}$, $m^{1/4} T^{2\delta-1} \gtrsim N^2 P \exp(d_{\mathcal{M}})$ for some $\delta > 1/2$, we have that for all $\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)$,

$$\begin{aligned} \mathbb{P} \left[\frac{\kappa_{L,A}}{8} \leq \frac{1}{T} \sum_{n=1}^T \|\Delta_{(1)} \tilde{\mathbf{x}}_n\|_2^2 \leq \frac{8\kappa_{U,A}}{3} \right] &\geq 1 - 2 \exp(-CT(\kappa_{L,A}/\kappa_{U,B})^2 \min(\lambda^{-2}, \lambda^{-4})) \\ &\quad - \exp(-Cd_{\mathcal{M}}). \end{aligned}$$

This accomplishes the proof. $\square$

Lemma S3. Suppose that (3.2) and Assumptions 1 and 2 hold. If sample size $T \gtrsim \max(\lambda^2, \lambda^4) (\kappa_{U,B}/\kappa_{L,A})^2 d_{\mathcal{M}}$, $m^{1/4} \gtrsim T^{2\delta}$ and $m^{1/4} T^{2\delta-1} \gtrsim N^2 P^{1/2} \exp(d_{\mathcal{M}})$ for some $\delta > 1/2$, then

$$\sup_{\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \left\langle \frac{1}{T} \sum_{n=1}^T \epsilon_n \circ \tilde{\mathbf{X}}_n, \Delta \right\rangle \leq C(\kappa^2 \sqrt{\lambda_{\max}(\Sigma_{\epsilon}) \kappa_{U,A}} + \kappa \sqrt{\kappa_{U,B}}) \sqrt{\frac{d_{\mathcal{M}}}{T}} + \frac{T^{2\delta}}{m^{1/4}}$$

with probability at least

$$1 - \exp(-Cd_{\mathcal{M}}) - 2 \exp(-CT(\kappa_{L,A}/\kappa_{U,B})^2 \min\{\lambda^{-2}, \lambda^{-4}\}),$$

where λ , $\kappa_{L,A}$, $\kappa_{U,B}$ and $d_{\mathcal{M}}$ are defined in Theorem 1.

Proof of Lemma S3. To separate the influence of model error and estimation error, we let

$\mathbf{v}_n = \boldsymbol{\eta}_n - \sum_{j=1}^P \mathbf{A}_j \boldsymbol{\eta}_{n-j} \in \mathbb{R}^{N \times 1}$, $\mathbf{X}_n = (\mathbf{y}_{n-1}, \dots, \mathbf{y}_{n-P}) \in \mathbb{R}^{N \times P}$, and $\boldsymbol{\Upsilon}_n = (\boldsymbol{\eta}_{n-1}, \dots, \boldsymbol{\eta}_{n-P}) \in \mathbb{R}^{N \times P}$. Note that $\boldsymbol{\epsilon}_n = \boldsymbol{\epsilon}_n + \mathbf{v}_n$ and $\tilde{\mathbf{X}}_n = \mathbf{X}_n + \boldsymbol{\Upsilon}_n$. As a result,

$$\begin{aligned} & \sup_{\boldsymbol{\Delta} \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \left\langle \frac{1}{T} \sum_{n=1}^T \boldsymbol{\epsilon}_n \circ \tilde{\mathbf{X}}_n, \boldsymbol{\Delta} \right\rangle \\ & \leq \sup_{\boldsymbol{\Delta} \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \left\langle \frac{1}{T} \sum_{n=1}^T \boldsymbol{\epsilon}_n \circ \tilde{\mathbf{X}}_n, \boldsymbol{\Delta} \right\rangle \\ & \quad + \sup_{\boldsymbol{\Delta} \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \left\langle \frac{1}{T} \sum_{n=1}^T \mathbf{v}_n \circ \mathbf{X}_n, \boldsymbol{\Delta} \right\rangle + \sup_{\boldsymbol{\Delta} \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \left\langle \frac{1}{T} \sum_{n=1}^T \mathbf{v}_n \circ \boldsymbol{\Upsilon}_n, \boldsymbol{\Delta} \right\rangle. \end{aligned} \tag{S4.9}$$

We shall provide upper bounds for the three terms one by one.

For the first term of (S4.9), it is easily verified that $\langle \boldsymbol{\epsilon}_n \circ \tilde{\mathbf{X}}_n, \boldsymbol{\Delta} \rangle = \langle \boldsymbol{\epsilon}_n, \boldsymbol{\Delta}_{(1)} \tilde{\mathbf{x}}_n \rangle$. Denote $S_t(\boldsymbol{\Delta}) = \sum_{n=1}^t \langle \boldsymbol{\epsilon}_n, \boldsymbol{\Delta}_{(1)} \tilde{\mathbf{x}}_n \rangle$ and $R_t(\boldsymbol{\Delta}) = \sum_{n=1}^t \|\boldsymbol{\Delta}_{(1)} \tilde{\mathbf{x}}_n\|_2^2$, for $1 \leq t \leq T$. By the Chernoff bound of errors, for any $\alpha > 0$, $\beta > 0$ and $c > 0$, there exists $\eta > 0$,

$$\begin{aligned} & \mathbb{P} \left[\{S_T(\boldsymbol{\Delta}) \geq \alpha\} \cap \{R_T(\boldsymbol{\Delta}) \leq \beta\} \right] \\ & = \inf_{\eta > 0} \mathbb{P} \left[\{\exp(\eta S_T(\boldsymbol{\Delta})) \geq \exp(\eta \alpha)\} \cap \{R_T(\boldsymbol{\Delta}) \leq \beta\} \right] \\ & = \inf_{\eta > 0} \mathbb{P} [\exp(\eta S_T(\boldsymbol{\Delta})) \mathbb{I}(R_T(\boldsymbol{\Delta}) \leq \beta) \geq \exp(\eta \alpha)] \\ & \leq \inf_{\eta > 0} \exp(-\eta \alpha) \mathbb{E} [\exp(\eta S_T(\boldsymbol{\Delta})) \mathbb{I}(R_T(\boldsymbol{\Delta}) \leq \beta)] \\ & = \inf_{\eta > 0} \exp(-\eta \alpha + c\eta^2 \beta) \mathbb{E} [\exp(\eta S_T(\boldsymbol{\Delta}) - c\eta^2 \beta) \mathbb{I}(R_T(\boldsymbol{\Delta}) \leq \beta)] \\ & \leq \inf_{\eta > 0} \exp(-\eta \alpha + c\eta^2 \beta) \mathbb{E} [\exp(\eta S_T(\boldsymbol{\Delta}) - c\eta^2 R_T(\boldsymbol{\Delta}))]. \end{aligned} \tag{S4.10}$$

By the tower property for conditional expectations, we have

$$\begin{aligned}
& \mathbb{E} \left[\exp \left(\eta S_T(\Delta) - c\eta^2 R_T(\Delta) \right) \right] \\
&= \mathbb{E} \left[\mathbb{E} \left[\exp \left(\eta S_T(\Delta) - c\eta^2 R_T(\Delta) \right) \mid \mathcal{F}_{T-1} \right] \right] \\
&= \mathbb{E} \left[\exp \left(\eta S_{T-1}(\Delta) - c\eta^2 R_{T-1}(\Delta) \right) \mathbb{E} \left[\exp \left(\eta \langle \varepsilon_T, \Delta_{(1)} \tilde{\mathbf{x}}_T \rangle - c\eta^2 \|\Delta_{(1)} \tilde{\mathbf{x}}_T\|_2^2 \right) \mid \mathcal{F}_{T-1} \right] \right].
\end{aligned}$$

With the sub-Gaussianity condition in Assumption 2, then $\langle \varepsilon_T, \Delta_{(1)} \tilde{\mathbf{x}}_T \rangle = \langle \xi_T, \Sigma_\varepsilon^{1/2} \Delta_{(1)} \tilde{\mathbf{x}}_T \rangle$, and $\mathbb{E} \left[\exp(\eta \langle \varepsilon_T, \Delta_{(1)} \tilde{\mathbf{x}}_T \rangle) \right] \leq \exp \left(\eta^2 \lambda^2 \lambda_{\max}(\Sigma_\varepsilon) \|\Delta_{(1)} \tilde{\mathbf{x}}_T\|_2^2 / 2 \right)$. Since $\tilde{\mathbf{x}}_n$ is $\mathcal{F}_{n-1}$ -measurable, ε_n is $\mathcal{F}_n$ -measurable and $\varepsilon_n | \mathcal{F}_{n-1}$ is mean-zero, let $c = \lambda^2 \lambda_{\max}(\Sigma_\varepsilon) / 2$, and the following inequalities can be easily deduced,

$$\begin{aligned}
& \mathbb{E} \left[\exp(\eta S_T(\Delta) - \eta^2 \lambda^2 \lambda_{\max}(\Sigma_\varepsilon) R_T(\Delta) / 2) \right] \\
& \leq \mathbb{E} \left[\exp(\eta S_{T-1}(\Delta) - \eta^2 \lambda^2 \lambda_{\max}(\Sigma_\varepsilon) R_{T-1}(\Delta) / 2) \right] \\
& \leq \dots \leq \mathbb{E} \left[\exp(\eta S_1(\Delta) - \eta^2 \lambda^2 \lambda_{\max}(\Sigma_\varepsilon) R_1(\Delta) / 2) \right] \leq 1.
\end{aligned}$$

As a result, for any $\alpha > 0$ and $\beta > 0$, we can have the following inequality of (S4.10),

$$\begin{aligned}
& \mathbb{P}[\{S_T(\Delta) \geq \alpha\} \cap \{R_T(\Delta) \leq \beta\}] \\
& \leq \inf_{\eta > 0} \exp(-\eta\alpha + \eta^2 \lambda^2 \lambda_{\max}(\Sigma_\varepsilon) \beta / 2) \\
& = \exp \left(-\frac{\alpha^2}{2\lambda^2 \lambda_{\max}(\Sigma_\varepsilon) \beta} \right).
\end{aligned} \tag{S4.11}$$

Moreover, according to Lemma S2, the following bounds for $R_T(\Delta)$ hold that

$$\frac{T}{8} \kappa_{L,A} \leq R_T(\Delta) \leq \frac{8T}{3} \kappa_{U,A} \tag{S4.12}$$

with probability at least $1 - 2 \exp(-CT(\kappa_{L,A}/\kappa_{U,B})^2 \min\{\lambda^{-2}, \lambda^{-4}\}) - \exp(-Cd_{\mathcal{M}})$.

By Lemma S1 (ii), for any $x > 0$,

$$\begin{aligned} & \mathbb{P} \left[\sup_{\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \left\langle \frac{1}{T} \sum_{n=1}^T \boldsymbol{\varepsilon}_n \circ \tilde{\mathbf{X}}_n, \Delta \right\rangle \geq x \right] \\ & \leq \mathbb{P} \left[\max_{\Delta \in \bar{\mathcal{S}}(2r_1, 2r_2, 2r_3)} \left\langle \frac{1}{T} \sum_{n=1}^T \boldsymbol{\varepsilon}_n \circ \tilde{\mathbf{X}}_n, \Delta \right\rangle \geq (1 - 2\sqrt{2}\epsilon)x \right] \\ & \leq |\bar{\mathcal{S}}(2r_1, 2r_2, 2r_3)| \cdot \mathbb{P} \left[\left\langle \frac{1}{T} \sum_{n=1}^T \boldsymbol{\varepsilon}_n \circ \tilde{\mathbf{X}}_n, \Delta \right\rangle \geq (1 - 2\sqrt{2}\epsilon)x \right], \end{aligned} \quad (\text{S4.13})$$

which, together with (S4.11) and (S4.12), implies that

$$\begin{aligned} & \mathbb{P} \left[\left\langle \frac{1}{T} \sum_{n=1}^T \boldsymbol{\varepsilon}_n \circ \tilde{\mathbf{X}}_n, \Delta \right\rangle \geq (1 - 2\sqrt{2}\epsilon)x \right] \\ & \leq \mathbb{P}[\{S_T(\Delta) \geq T(1 - 2\sqrt{2}\epsilon)x\} \cap \{R_T(\Delta) \leq CT\lambda^2\kappa_{U,A}\}] \\ & \quad + \mathbb{P}[R_T(\Delta) \geq CT\lambda^2\kappa_{U,A}] \\ & \leq \exp \left[-\frac{(1 - 2\sqrt{2}\epsilon)^2 T x^2}{2C\lambda^4 \lambda_{\max}(\Sigma_\epsilon) \kappa_{U,A}} \right] + 2 \exp(-CT(\kappa_{L,A}/\kappa_{U,B})^2 \min\{\lambda^{-2}, \lambda^{-4}\}) + \exp(-Cd_{\mathcal{M}}) \end{aligned}$$

for any $x > 0$. Note that, from Lemma S1, $|\bar{\mathcal{S}}(r_1, r_2, r_3)| \leq (12/\epsilon)^{d_{\mathcal{M}}}$. By letting $\epsilon = 1/10$,

and $x = C\sqrt{d_{\mathcal{M}}\lambda^4 \lambda_{\max}(\Sigma_\epsilon) \kappa_{U,A}/T}$, we then have

$$\begin{aligned} & \mathbb{P} \left[\sup_{\Delta \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \left\langle \frac{1}{T} \sum_{t=1}^T \boldsymbol{\varepsilon}_n \circ \tilde{\mathbf{X}}_n, \Delta \right\rangle \geq C\lambda^2 \sqrt{\lambda_{\max}(\Sigma_\epsilon) \kappa_{U,A}} \sqrt{\frac{d_{\mathcal{M}}}{T}} \right] \\ & \leq \exp(-Cd_{\mathcal{M}}) + 2 \exp(-CT(\kappa_{L,A}/\kappa_{U,B})^2 \min\{\lambda^{-2}, \lambda^{-4}\}). \end{aligned} \quad (\text{S4.14})$$

For the second term in (S4.9), since $\mathbf{x}_n = (\mathbf{y}_{n-1}^\top, \dots, \mathbf{y}_{n-P}^\top)^\top \in \mathbb{R}^{NP \times 1}$, it can be verified that $\langle \mathbf{v}_n \circ \mathbf{X}_n, \Delta \rangle = \langle \mathbf{v}_n, \Delta_{(1)} \mathbf{x}_n \rangle$. Denote $\mathbf{z} = (\mathbf{x}_T^\top, \dots, \mathbf{x}_1^\top)^\top \in \mathbb{R}^{NPT \times 1}$, and $\mathbf{v} = (\mathbf{v}_T^\top, \dots, \mathbf{v}_1^\top)^\top \in \mathbb{R}^{NT \times 1}$, and it holds that

$$\sum_{n=1}^T \langle \mathbf{v}_n, \Delta_{(1)} \mathbf{x}_n \rangle = \mathbf{v}^\top (\mathbf{I}_{T-P} \otimes \Delta_{(1)}) \mathbf{z} := \mathbf{m}_2^\top \bar{\boldsymbol{\xi}},$$

where $\mathbf{m}_2 = \bar{\Sigma} \mathbf{P}^\top (\mathbf{I}_T \otimes \Delta_{(1)}) \mathbf{v}$, $\mathbf{z} = \mathbf{P} \mathbf{e}$, and $\mathbf{e} = \bar{\Sigma} \bar{\boldsymbol{\xi}}$. Note that $\|\mathbf{v}\|_2^2 \leq CNT \eta_{\max}^2$ and

$\|\mathbf{m}_2\|_2^2 \leq CNT\eta_{\max}^2\kappa_{U,B}$, where $\eta_{\max} = \max_{1 \leq n \leq T} \max_{1 \leq i \leq N} |\eta_{n,i}|$. Refer to the proof of Lemma S2 for more details.

With the sub-Gaussianity condition of $\boldsymbol{\xi}$ at Assumption 2, similar to the proof of Lemma S2, we can obtain that

$$\mathbb{P}[|\mathbf{m}_2^\top \bar{\boldsymbol{\xi}}| \geq t | \eta_{\max}] \leq \exp\left(-\frac{Ct^2}{\lambda^2\kappa_{U,B}NT\eta_{\max}^2}\right) \quad \text{for all } t > 0$$

and, by letting $x = m^{1/8}T^\delta/N^{1/2}$ in (S4.5),

$$\mathbb{P}[|\mathbf{m}_2^\top \bar{\boldsymbol{\xi}}| \geq t] \leq \exp\left(-\frac{Cm^{1/4}t^2}{\lambda^2\kappa_{U,B}T^{1+2\delta}}\right) + \frac{CN^2}{m^{1/4}T^{2\delta-1}}.$$

Therefore, by a method similar to (S4.13),

$$\begin{aligned} & \mathbb{P}\left[\sup_{\boldsymbol{\Delta} \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \left\langle \frac{1}{T} \sum_{n=1}^T \mathbf{v}_n \circ \mathbf{X}_t, \boldsymbol{\Delta} \right\rangle \geq t\right] \\ & \leq \mathbb{P}\left[\max_{\boldsymbol{\Delta} \in \bar{\mathcal{S}}(2r_1, 2r_2, 2r_3)} \left\langle \frac{1}{T} \sum_{n=1}^T \mathbf{v}_n \circ \mathbf{X}_t, \boldsymbol{\Delta} \right\rangle \geq (1 - 2\sqrt{2}\epsilon)t\right] \\ & \leq |\bar{\mathcal{S}}(2r_1, 2r_2, 2r_3)| \mathbb{P}\left[\left\langle \sum_{n=1}^T \mathbf{v}_n \circ \mathbf{X}_t, \boldsymbol{\Delta} \right\rangle \geq (1 - 2\sqrt{2}\epsilon)Tt\right] \\ & \leq \exp(Cd_{\mathcal{M}}) \frac{CN^2}{m^{1/4}T^{2\delta-1}} + \exp\left(Cd_{\mathcal{M}} - \frac{Cm^{1/4}t^2}{\lambda^2\kappa_{U,B}T^{2\delta-1}}\right). \end{aligned}$$

By letting $t = C\sqrt{\lambda^2\kappa_{U,B}T^{2\delta-1}d_{\mathcal{M}}/m^{1/4}}$ and with the conditions that $m^{1/4} \gtrsim T^{2\delta}$ and $m^{1/4}T^{2\delta-1} \gtrsim N^2 \exp(d_{\mathcal{M}})$ for some $\delta > 1/2$, we have

$$\mathbb{P}\left[\sup_{\boldsymbol{\Delta} \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \left\langle \frac{1}{T} \sum_{n=1}^T \mathbf{v}_n \circ \mathbf{X}_t, \boldsymbol{\Delta} \right\rangle \geq C\lambda\sqrt{\kappa_{U,B}}\sqrt{\frac{d_{\mathcal{M}}}{T}}\right] \leq \exp(-Cd_{\mathcal{M}}). \quad (\text{S4.15})$$

Finally for the third term at (S4.9), denote $\boldsymbol{\vartheta}_n = (\boldsymbol{\eta}_{n-1}^\top, \dots, \boldsymbol{\eta}_{n-P}^\top)^\top \in \mathbb{R}^{NP \times 1}$, $\mathbf{v}_n = \boldsymbol{\eta}_n - \sum_{j=1}^P \mathbf{A}_j \boldsymbol{\eta}_{n-j} \in \mathbb{R}^{N \times 1}$, $\boldsymbol{\Upsilon}_n = (\boldsymbol{\eta}_{n-1}, \dots, \boldsymbol{\eta}_{n-P}) \in \mathbb{R}^{N \times P}$, and it holds that $\langle \mathbf{v}_n \circ \boldsymbol{\Upsilon}_n, \boldsymbol{\Delta} \rangle =$

$\langle \mathbf{v}_n, \mathbf{\Delta}_{(1)} \mathbf{v}_n \rangle$. Let $\mathbf{c} = (\mathbf{v}_T^\top, \dots, \mathbf{v}_1^\top)^\top$, and

$$\sum_{n=1}^T \langle \mathbf{v}_n, \mathbf{\Delta}_{(1)} \mathbf{v}_n \rangle = \mathbf{v}^\top (\mathbf{I}_T \otimes \mathbf{\Delta}_{(1)}) \mathbf{c},$$

then, we have

$$|\mathbf{v}^\top (\mathbf{I}_T \otimes \mathbf{\Delta}_{(1)}) \mathbf{c}| \leq CNP^{1/2} T \eta_{\max}^2,$$

as $\|\mathbf{v}\|_2 \leq C\sqrt{NT} \eta_{\max}$ and $\|\mathbf{c}\|_2 \leq \sqrt{NTP} \eta_{\max}$. Thus, letting $x = m^{1/8} T^\delta / N^{1/2} P^{1/4}$ at

(S4.5), we have

$$\mathbb{P} \left[\sup_{\mathbf{\Delta} \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \sum_{n=1}^T \langle \mathbf{v}_n, \mathbf{\Delta}_{(1)} \mathbf{v}_n \rangle \geq \frac{T^{1+2\delta}}{m^{1/4}} \right] \leq \frac{CN^2 P^{1/2}}{m^{1/4} T^{2\delta-1}}. \quad (\text{S4.16})$$

Combining (S4.14), (S4.15), and (S4.16), with $m^{1/4} T^{2\delta-1} \gtrsim N^2 P^{1/2} \exp(d_{\mathcal{M}})$ the complete deviation bound can be obtained,

$$\begin{aligned} & \mathbb{P} \left[\sup_{\mathbf{\Delta} \in \mathcal{S}(2r_1, 2r_2, 2r_3)} \frac{1}{T} \sum_{n=1}^T \langle \mathbf{\epsilon}_n, \mathbf{\Delta}_{(1)} \tilde{\mathbf{x}}_n \rangle \geq C(\kappa^2 \sqrt{\lambda_{\max}(\mathbf{\Sigma}_\epsilon) \kappa_{U,A}} + \lambda \sqrt{\kappa_{U,B}}) \sqrt{\frac{d_{\mathcal{M}}}{T}} + \frac{T^{2\delta}}{m^{1/4}} \right] \\ & \leq \exp(-Cd_{\mathcal{M}}) + 2 \exp(-CT(\kappa_{L,A}/\kappa_{U,B})^2 \min\{\lambda^{-2}, \lambda^{-4}\}). \end{aligned}$$

We hence complete the proof. $\square$

Lemma S4 (Lemma 3 of Chen et al. (2019)). *Consider three Tucker ranks $(r_1^{(j)}, r_2^{(j)}, r_3^{(j)})$ for $0 \leq j \leq 2$, and assume that $r_i^{(1)} < r_i^{(2)} < r_i^{(0)}$ for $0 \leq i \leq 3$. For $\mathbf{X} \in \mathbf{\Theta}(r_1^{(0)}, r_2^{(0)}, r_3^{(0)})$, it holds that*

$$\|P_{\mathbf{\Theta}(r_1^{(2)}, r_2^{(2)}, r_3^{(2)})}(\mathbf{X}) - \mathbf{X}\|_{\text{F}} \leq [\Pi_{i=1}^3 (\beta_i + 1) - 1] \|\mathbf{Y} - \mathbf{X}\|_{\text{F}} \quad (\text{S4.17})$$

for any $\mathbf{Y} \in \mathbf{\Theta}(r_1^{(1)}, r_2^{(1)}, r_3^{(1)})$, where $\beta_i = \sqrt{(r_i^{(0)} - r_i^{(2)}) / (r_i^{(0)} - r_i^{(1)})}$.

S5 Univariate HAR-Itô model

When $N = 1$, the VHAR-Itô model reduces to the univariate HAR-Itô model. In this section, we present the univariate HAR-Itô model.

Let X_t be the log price of an asset at time $t \in \mathbb{R}_+$, where $\mathbb{R}_+ = [0, \infty)$. Denote by $\mathcal{F}_t$ the collection of all information up to time t , and then $\{\mathcal{F}_t, t \geq 0\}$ is a filtration. To account for jump components in financial industry, we consider a jump diffusion model,

$$dX_t = \mu_t dt + \sigma_t dB_t + L_t d\Lambda_t, \quad (\text{S5.1})$$

where μ_t is a drift term, σ_t is an instantaneous volatility process, B_t is a standard Brownian motion, Λ_t is a standard Poisson process with constant intensity $\lambda > 0$, the jump sizes $\{L_t, t \geq 0\}$ are *i.i.d.* with finite fourth moment, the processes of σ_t , B_t and Λ_t are all adapted to the filtration $\{\mathcal{F}_t\}$, and the jump sizes are independent of the three processes. Moreover, let $Z_t = \int_{[t]}^t dW_t$, where W_t is another standard Brownian motion with respect to $\mathcal{F}_t$ such that $dB_t dW_t = \rho dt$ with $|\rho| < 1$ almost surely, and ρ can be seen as the leverage effect between the log price and the instantaneous volatility. We then define the HAR-Itô model below.

Definition S1. The log prices $\{X_t, t \in \mathbb{R}_+\}$ at (S5.1) are said to follow an univariate HAR-Itô model if the instantaneous volatility satisfies

$$\begin{aligned} \sigma_t^2 = & \sigma_{[t]}^2 + (t - [t]) \left(\omega - \sigma_{[t]}^2 \right) + \alpha_1 \int_{[t]}^t \sigma_s^2 ds + \sum_{l=2}^P \alpha_l \int_{[t]-l+1}^{[t]-l+2} \sigma_s^2 ds \\ & + \beta \int_{[t]}^t L_s^2 d\Lambda_s + v([t] + 1 - t) Z_t^2, \end{aligned} \quad (\text{S5.2})$$

where $[t]$ denotes the integer part of t , i.e. $[t] = n - 1$ for $n - 1 < t \leq n$, P is a positive

integer, the parameters of ω , β , v , α_1 are all positive, and $\alpha_l \geq 0$ with $2 \leq l \leq P$.

Denote $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_P)^\top$ and $\boldsymbol{\theta} = (\lambda, \rho, \omega, \beta, v, \boldsymbol{\alpha}^\top)^\top$, and the above definition relies on the parameter vector $\boldsymbol{\theta}$. In addition, the instantaneous volatility σ_t^2 at (S5.2) is a process with continuous time and is defined for all $t \in \mathbb{R}_+$. Consider the integer time points only, and then the process has the form of

$$\sigma_n^2 = \omega + \sum_{l=1}^P \alpha_l y_{n-l+1} + \beta J_n,$$

with probability one, where

$$y_n = IV_n = \int_{n-1}^n \sigma_s^2 ds \quad \text{and} \quad J_n = \int_{n-1}^n L_s^2 d\Lambda_s$$

are the integrated volatility and jump variation, respectively. Moreover, from (S5.2), the daily volatility (or integrated volatility) in the past will contribute to the spot volatility (or instantaneous volatility) by the introduction of α_l with $2 \leq l \leq P$.

As in Song et al. (2021), we denote $\omega_L = \mathbb{E}(L_t^2)$ and $M_t = L_t^2 - \omega_L$. Moreover, let

$$\omega^g = \varrho_1 \omega + \varrho_2 \beta \omega_L \lambda + (\varrho_2 - 2\varrho_3) v, \quad \beta^g = (\varrho_1 - \varrho_2) \beta \quad \text{and} \quad \alpha_l^g = (\varrho_1 - \varrho_2) \alpha_l$$

for all $1 \leq l \leq P$, where $\varrho_1 = \alpha_1^{-1}(e^{\alpha_1} - 1)$, $\varrho_2 = \alpha_1^{-2}(e^{\alpha_1} - 1 - \alpha_1)$ and $\varrho_3 = \alpha_1^{-3}(e^{\alpha_1} - 1 - \alpha_1 - \alpha_1^2/2)$.

Proposition S1. *Suppose that $n \geq P + 1$ and $0 < \alpha_1 < 1$. Then the integrated volatility for univariate HAR-Itô models at Definition S1 satisfies*

$$y_n = \omega^g + \sum_{j=1}^P \alpha_j^g y_{n-j} + \beta^g J_{n-1} + \varepsilon_n \tag{S5.3}$$

with probability one, where $\varepsilon_n = \varepsilon_n^c + \varepsilon_n^J$,

$$\begin{aligned}\varepsilon_n^c &= 2v\alpha_1^{-2} \int_{n-1}^n \left\{ \alpha_1 (n-t - \alpha_1^{-1}) e^{\alpha_1(n-t)} + 1 \right\} Z_t dZ_t, \\ \varepsilon_n^J &= \beta\alpha_1^{-1} \left\{ \int_{n-1}^n (e^{\alpha_1(n-t)} - 1) [M_t d\Lambda_t + \omega_L(d\Lambda_t - \lambda dt)] \right\},\end{aligned}$$

and ε_n^c and ε_n^J are innovation terms from the continuous diffusion process and the jump component, respectively. Moreover, both $\{\varepsilon_n^c\}$ and $\{\varepsilon_n^J\}$ are i.i.d. sequences with mean zero and finite variance.

When there is no jump component, i.e. $L_t = 0$ with probability one for all t , the iterative form at (S5.3) can be further simplified into

$$y_n = \omega^g + \sum_{j=1}^P \alpha_j^g y_{n-j} + \varepsilon_n, \quad (\text{S5.4})$$

and the integrated volatility y_n can be estimated consistently by the realized volatility, denoted by RV_n . Let $IV_n = y_n$, $IV_n^{(w)} = 5^{-1} \sum_{l=1}^5 y_{n+1-l}$ and $IV_n^{(m)} = 22^{-1} \sum_{l=1}^{22} y_{n+1-l}$ be the daily, weekly and monthly integrated volatilities, respectively, and we consider a special case of (S5.4) below,

$$IV_n = \beta_0 + \beta_d IV_{n-1} + \beta_w IV_{n-1}^{(w)} + \beta_m IV_{n-1}^{(m)} + \varepsilon_n, \quad (\text{S5.5})$$

where $P = 22$, $\omega^g = \beta_0$, $(\alpha_1^g, \dots, \alpha_P^g) = (\beta_d, \beta_w, \beta_m) \mathbf{U}_C^\top$, and

$$\mathbf{U}_C^\top = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 & 0 & \cdots & 0 \\ 1/22 & 1/22 & 1/22 & 1/22 & 1/22 & 1/22 & \cdots & 1/22 \end{pmatrix} \in \mathbb{R}^{3 \times 22}.$$

Corsi (2009) made use of model (S5.5) to propose the HAR model for predicting the realized volatility, and the proposed HAR-Itô model at Definition S1 and Proposition S1 provides a

probabilistic support for this popular model. On the other hand, the cascade structure at (S5.5) reveals that there usually exists a low-dimensional structure among the parameters α_j 's and α_j^g 's along the direction of lags.

S6 A Table for the selected ranks of the MLR-HAR and VHARI models in Real data analysis

Table S1 gives the results of the selected ranks for the MLR-HAR and VHARI models.

Table S1: Selected ranks of the VHARI and MLR-HAR models with 150 stocks and 229 stocks during the first (2017.01 – 2019.12), second (2020.01 – 2024.12) and full period (2017.01 – 2024.12)

		Panel A: RV			Panel B: MedRV		
N	period	VHARI	MLR-HAR		VHARI	MLR-HAR	
		$P = 22$	$P = 22$	$P = 66$	$P = 22$	$P = 22$	$P = 66$
150	first	$r = 4$	$(r_1, r_2, r_3) = (6, 6, 2)$	$(r_1, r_2, r_3) = (6, 6, 1)$	$r = 3$	$(r_1, r_2, r_3) = (6, 6, 1)$	$(r_1, r_2, r_3) = (6, 6, 1)$
	second	$r = 4$	$(r_1, r_2, r_3) = (7, 7, 1)$	$(r_1, r_2, r_3) = (6, 6, 1)$	$r = 4$	$(r_1, r_2, r_3) = (7, 7, 1)$	$(r_1, r_2, r_3) = (6, 6, 1)$
	full	$r = 5$	$(r_1, r_2, r_3) = (3, 3, 1)$	$(r_1, r_2, r_3) = (4, 4, 1)$	$r = 5$	$(r_1, r_2, r_3) = (5, 5, 2)$	$(r_1, r_2, r_3) = (5, 5, 1)$
229	first	$r = 4$	$(r_1, r_2, r_3) = (5, 5, 1)$	$(r_1, r_2, r_3) = (5, 5, 1)$	$r = 4$	$(r_1, r_2, r_3) = (5, 5, 1)$	$(r_1, r_2, r_3) = (5, 5, 1)$
	second	$r = 4$	$(r_1, r_2, r_3) = (6, 6, 1)$	$(r_1, r_2, r_3) = (6, 6, 1)$	$r = 4$	$(r_1, r_2, r_3) = (6, 6, 1)$	$(r_1, r_2, r_3) = (5, 5, 1)$
	full	$r = 4$	$(r_1, r_2, r_3) = (5, 5, 1)$	$(r_1, r_2, r_3) = (3, 3, 1)$	$r = 3$	$(r_1, r_2, r_3) = (4, 4, 1)$	$(r_1, r_2, r_3) = (5, 5, 1)$

S7 Coefficient-Based Total Spillover Index from the VHAR, VHARI, and MLR-HAR models in Real data analysis

The section examines volatility spillovers among $N = 150$ and $N = 229$ stocks across three periods. To quantify dynamic connectedness, we construct a coefficient-based spillover index from the estimated lag coefficient matrices. Specifically, for each asset pair i and j , we define

$$S_{ij} = \sum_{p=1}^P \left(\hat{A}_{ij,p} \right)^2, \quad \tilde{S}_{ij} = \frac{S_{ij}}{\sum_k \sum_l S_{kl}}, \quad \text{TSI} = 1 - \sum_i \tilde{S}_{ii},$$

where S_{ij} measures the total lagged transmission intensity from Assets j to i . The resulting TSI measures the proportion of cross-asset dependence in total system dynamics and can be interpreted as a time-varying index of structural connectedness. Unlike the generalized forecast error variance decomposition (FEVD) of Diebold and Yilmaz (2012, 2014), which is based on multi-step forecast error variance shares from a moving-average representation, our index is constructed directly from the estimated autoregressive coefficients. This avoids horizon selection, truncation, and covariance matrix inversion, thereby enhancing numerical stability in high-dimensional regularized settings.

Using the rolling-window procedure described in Section 6 of the main paper, we track the evolution of TSI under the three forecasting settings: evaluation in 2019 with $T = 502$, and in 2024 with $T = 1006$ or $T = 1760$. Figures S1 and S2 present the rolling estimates for $N = 150$ and $N = 229$ stocks. Compared with VHAR, MLR-HAR(22) and VHARI generally exhibit smoother short-term variation in TSI, although the levels and trends differ across settings. These patterns provide a descriptive comparison of cross-asset connectedness

reflected in the estimated lag coefficients.

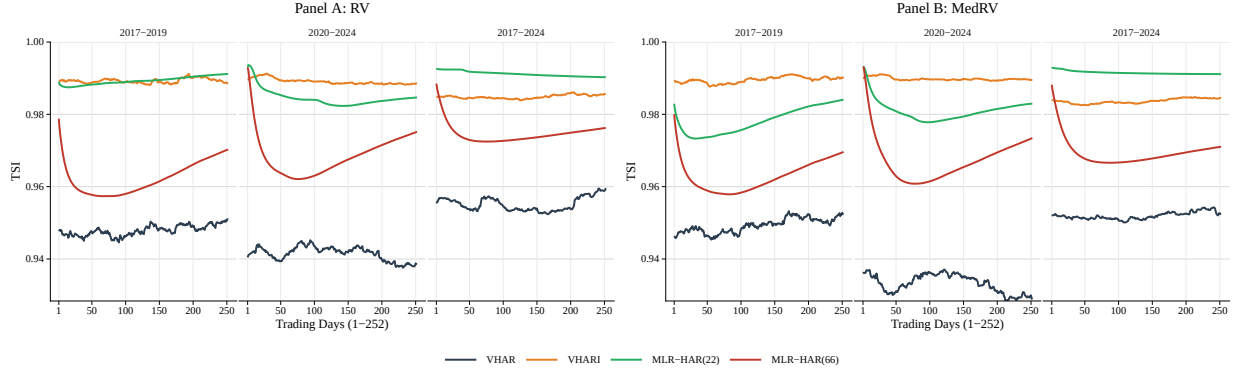


Figure S1: Plots of TSIs for $N = 150$ stocks from VHAR, VHARI, MLR-HAR(22), and MLR-HAR(66). From left to right, the groups use data from 2017–2019, 2020–2024, and 2017–2024, with the final year of each period used for rolling evaluation. Panel A uses RV and Panel B uses medRV.

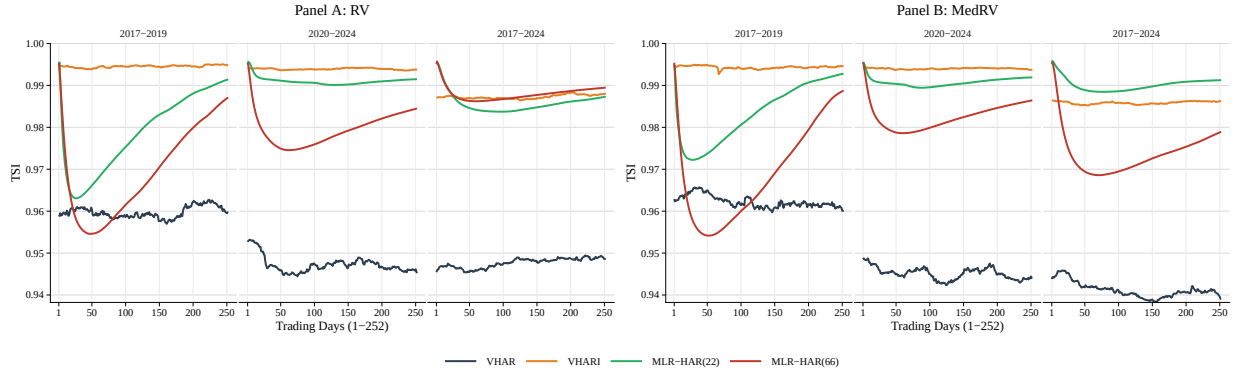


Figure S2: Plots of TSIs for $N = 229$ stocks from VHAR, VHARI, MLR-HAR(22), and MLR-HAR(66). From left to right, the groups use data from 2017–2019, 2020–2024, and 2017–2024, with the final year of each period used for rolling evaluation. Panel A uses RV and Panel B uses medRV.

References

- Basu, S. and G. Michailidis (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics* 43, 1535–1567.
- Chen, H., G. Raskutti, and M. Yuan (2019). Non-convex projected gradient descent for generalized low-rank tensor regression. *The Journal of Machine Learning Research* 20, 172–208.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7, 174–196.
- De Lathauwer, L., B. De Moor, and J. Vandewalle (2000). A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications* 21, 1253–1278.
- Diebold, F. X. and K. Yilmaz (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting* 28, 57–66.
- Diebold, F. X. and K. Yilmaz (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics* 182, 119–134.
- Kolda, T. G. and B. W. Bader (2009). Tensor decompositions and applications. *SIAM Review* 51, 455–500.
- Song, X., D. Kim, H. Yuan, X. Cui, Z. Lu, Y. Zhou, and Y. Wang (2021). Volatility analysis with realized GARCH-Itô models. *Journal of Econometrics* 222, 393–410.

- Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika* 31, 279–311.
- Wang, D., F. Huang, J. Zhao, G. Li, and G. Tian (2020). Compact autoregressive network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Volume 34, pp. 6145–6152.
- Wang, D., Y. Zheng, H. Lian, and G. Li (2021). High-dimensional vector autoregressive time series modeling via tensor decomposition. *Journal of the American Statistical Association* To appear.