High-Dimensional Vector Autoregressive Time Series Modeling via Tensor Decomposition

Ben-Zheng Li
UESTC

2025.3.14

Outline

Outline

向量自回归 (Vector Autoregression, VAR)

数学表达: 给定 N 维时间序列 $\{y_t\}_{t=1}^{T}$，VAR(p) 模型的形式为:

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \varepsilon_t, \tag{1}$$

其中:

- $y_t \in \mathbb{R}^N$ 是 t 时刻的观测向量，
- $A_1, A_2, \ldots, A_p \in \mathbb{R}^{N\times N}$ 是待求解的参数矩阵，
- $\varepsilon_t \in \mathbb{R}^N$ 是独立同分布、均值为零、协方差矩阵有限的噪声项。

向量自回归 (Vector Autoregression, VAR)

数学表达：给定 N 维时间序列 $\{y_t\}_{t=1}^{T}$，VAR(p) 模型的形式为：

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \varepsilon_t, \tag{1}$$

其中：

- $y_t \in \mathbb{R}^N$ 是 t 时刻的观测向量，
- $A_1, A_2, \ldots, A_p \in \mathbb{R}^{N\times N}$ 是待求解的参数矩阵，
- $\varepsilon_t \in \mathbb{R}^N$ 是独立同分布、均值为零、协方差矩阵有限的噪声项。

作用：

- 分析时间序列内部的动态行为；
- 基于历史数据进行多步预测，广泛用于经济、金融、环境科学等领域。

如何求解系数矩阵?

以 $y_1, y_2, \cdots, y_7$ 对应的 2 阶自回归模型为例。

$$\begin{cases} y_7 = A_1 y_6 + A_2 y_5 + \varepsilon_7, \\ y_6 = A_1 y_5 + A_2 y_4 + \varepsilon_6, \\ y_5 = A_1 y_4 + A_2 y_3 + \varepsilon_5, \\ y_4 = A_1 y_3 + A_2 y_2 + \varepsilon_4, \\ y_3 = A_1 y_2 + A_2 y_1 + \varepsilon_3. \end{cases}$$

如何求解系数矩阵?

以 $y_1, y_2, \cdots, y_7$ 对应的 2 阶自回归模型为例。

$$
\begin{cases}
y_7 = A_1 y_6 + A_2 y_5 + \varepsilon_7, \\
y_6 = A_1 y_5 + A_2 y_4 + \varepsilon_6, \\
y_5 = A_1 y_4 + A_2 y_3 + \varepsilon_5, \\
y_4 = A_1 y_3 + A_2 y_2 + \varepsilon_4, \\
y_3 = A_1 y_2 + A_2 y_1 + \varepsilon_3.
\end{cases}
\rightarrow
\begin{cases}
\varepsilon_7 = y_7 - A_1 y_6 - A_2 y_5, \\
\varepsilon_6 = y_6 - A_1 y_5 - A_2 y_4, \\
\varepsilon_5 = y_5 - A_1 y_4 - A_2 y_3, \\
\varepsilon_4 = y_4 - A_1 y_3 - A_2 y_2, \\
\varepsilon_3 = y_3 - A_1 y_2 - A_2 y_1.
\end{cases}
$$

如何求解系数矩阵?

以 $y_1, y_2, \cdots, y_7$ 对应的 2 阶自回归模型为例。

$$\begin{cases} y_7 = A_1 y_6 + A_2 y_5 + \varepsilon_7, \\ y_6 = A_1 y_5 + A_2 y_4 + \varepsilon_6, \\ y_5 = A_1 y_4 + A_2 y_3 + \varepsilon_5, \\ y_4 = A_1 y_3 + A_2 y_2 + \varepsilon_4, \\ y_3 = A_1 y_2 + A_2 y_1 + \varepsilon_3. \end{cases} \rightarrow \begin{cases} \varepsilon_7 = y_7 - A_1 y_6 - A_2 y_5, \\ \varepsilon_6 = y_6 - A_1 y_5 - A_2 y_4, \\ \varepsilon_5 = y_5 - A_1 y_4 - A_2 y_3, \\ \varepsilon_4 = y_4 - A_1 y_3 - A_2 y_2, \\ \varepsilon_3 = y_3 - A_1 y_2 - A_2 y_1. \end{cases}$$

最小二乘估计:

$$\min_{A_1,\ldots,A_p} \sum_{t=p+1}^{T} \left\| y_t - \sum_{\ell=1}^{p} A_\ell y_{t-\ell} \right\|_2^2 \tag{2}$$

上述模型面临的几个关键性挑战

(1) 参数量大: 对于 “高而窄” 的矩阵（即 $N \gg T$），其 **参数数量**为 pN^2，远大于观测数量 NT。
(2) 需要平稳性假设
(3) 对缺失值/异常值敏感
(4) 递归求解会积累误差，不适合长期预测

Outline

Tensor Decomposition-Based VAR

传统 VAR(p):

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t,\ \ A_1, \cdots, A_p \in \mathbb{R}^{N\times N}. \tag{3}$$

Tensor Decomposition-Based VAR

传统 VAR(p):

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t,\ \ A_1, \cdots, A_p \in \mathbb{R}^{N\times N}. \tag{3}$$

将 $A_1, \cdots, A_p$ 排成一个张量 $\mathcal{A} \in \mathbb{R}^{N\times N\times p}$, 则

$$\mathcal{A}_{(1)} = (A_1, A_2, \cdots, A_p) \in \mathbb{R}^{N\times pN}.$$

Tensor Decomposition-Based VAR

传统 VAR(p):

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t, \ \ A_1, \cdots, A_p \in \mathbb{R}^{N \times N}. \tag{3}$$

将 $A_1, \cdots, A_p$ 排成一个张量 $\mathcal{A} \in \mathbb{R}^{N \times N \times p}$, 则

$$\mathcal{A}_{(1)} = (A_1, A_2, \cdots, A_p) \in \mathbb{R}^{N \times pN}.$$

方程 (3) 可以重写为

$$y_t = (A_1, A_2, \cdots, A_p)(y_{t-1}^\top, \ldots, y_{t-p}^\top)^\top + \varepsilon_t \triangleq \mathcal{A}_{(1)} x_t + \varepsilon_t, \tag{4}$$

其中 $x_t = (y_{t-1}^\top, \ldots, y_{t-p}^\top)^\top \in \mathbb{R}^{pN}$.

Tensor Decomposition-Based VAR

利用 **Tucker 分解**对张量 $\mathcal{A} \in \mathbb{R}^{N\times N\times p}$ 进行降秩:

$$\mathcal{A} = \mathcal{G} \times_1 U_1 \times_2 U_2 \times_3 U_3,$$

其中 $\mathcal{G} \in \mathbb{R}^{r_1\times r_2\times r_3}$ $(r_1, r_2 < N, r_3 < p)$ 是核心张量, $U_1 \in \mathbb{R}^{N\times r_1}$ 是响应因子, $U_2 \in \mathbb{R}^{N\times r_2}$ 是预测因子, $U_3 \in \mathbb{R}^{p\times r_3}$ 是滞后因子.

Tensor Decomposition-Based VAR

利用 **Tucker 分解**对张量 $\mathcal{A} \in \mathbb{R}^{N\times N\times p}$ 进行降秩:

$$\mathcal{A} = \mathcal{G} \times_1 U_1 \times_2 U_2 \times_3 U_3,$$

其中 $\mathcal{G} \in \mathbb{R}^{r_1\times r_2\times r_3}$ $(r_1, r_2 < N, r_3 < p)$ 是核心张量, $U_1 \in \mathbb{R}^{N\times r_1}$ 是响应因子, $U_2 \in \mathbb{R}^{N\times r_2}$ 是预测因子, $U_3 \in \mathbb{R}^{p\times r_3}$ 是滞后因子.

- 传统 VAR(p) 的参数为 pN^2.
- TVAR(p) 参数量为 $r_1r_2r_3 + r_1N + r_2N + r_3p$.
- 若 U_1, U_2, U_3 为半正交矩阵 (HOSVD 分解)，则参数量进一步降为

$$r_1r_2r_3 + r_1(N - r_1) + r_2(N - r_2) + r_3(p - r_3).$$

静态因子模型（SFM）和动态因子模型（DFM）

因子模型核心思想：用少数因子解释高维数据的主要变化。
静态因子模型（SFM）：

$$y_t = \Lambda f_t + e_t.$$

其中：

$$y_t \in \mathbb{R}^N\text{(高维时间序列)}, \qquad f_t \in \mathbb{R}^r\text{(低维因子，}r \ll N\text{)},$$

$$\Lambda \in \mathbb{R}^{N\times r}\text{(因子载荷矩阵)}, \qquad e_t\text{(误差项)。}$$

特点：仅静态数据的**降维**，没考虑时序依赖性，不适用于时序预测。

静态因子模型（SFM）和动态因子模型（DFM）

因子模型核心思想：用少数因子解释高维数据的主要变化。
静态因子模型（SFM）：

$$y_t = \Lambda f_t + e_t.$$

其中：

$$y_t \in \mathbb{R}^N(\text{高维时间序列}), \quad f_t \in \mathbb{R}^r(\text{低维因子，} r \ll N),$$

$$\Lambda \in \mathbb{R}^{N\times r}(\text{因子载荷矩阵}), \quad e_t(\text{误差项})。$$

特点：仅静态数据的**降维**，没考虑时序依赖性，不适用于时序预测。
动态因子模型（DFM）：

$$y_t = \Lambda f_t + e_t, \quad f_t = A f_{t-1} + \nu_t.$$

特点：结合 **SFM（降维）+ VAR（时间建模）**，可用于时序预测。

TVAR（张量自回归）vs. DFM（动态因子模型）

DFM 公式：

$$y_t = \Lambda f_t + e_t, \quad f_t = A f_{t-1} + \nu_t. \tag{5}$$

TVAR 公式：

$$y_t = (\mathcal{G} \times_1 U_1 \times_2 U_2 \times_3 U_3)_{(1)} x_t + \epsilon_t.$$

由 U_1 是半正交矩阵，可得：

$$U_1' y_t = \mathcal{G}_{(1)} (U_3 \otimes U_2)' x_t + U_1' \epsilon_t. \tag{6}$$

	DFM	TVAR
降维方式	Λf_t (单个维度降维)	U_1, U_2, U_3 (多维度降维)
时间建模	因子自回归 $f_t = A f_{t-1} + \nu_t$	通过 U_3 隐式学习时间依赖
参数量	$O(N r_1 + r_1^2)$	$O(r_1 r_2 r_3 + r_1 N + r_2 N + r_3 p)$ (当 $r_2, r_3 << N$ 时较少)

Sparse Higher-Order Reduced-Rank VAR (SHORR)

TVAR 目标函数：

$$\arg\min_{\mathcal{G},U_1,U_2,U_3} \sum_{t=p+1}^{T} \left\| y_t - (\mathcal{G} \times_1 U_1 \times_2 U_2 \times_3 U_3)_{(1)} x_t \right\|_2^2 \tag{7}$$

SHORR 目标函数：

$$\arg\min_{\mathcal{G},U_1,U_2,U_3} \sum_{t=p+1}^{T} \left\| y_t - (\mathcal{G} \times_1 U_1 \times_2 U_2 \times_3 U_3)_{(1)} x_t \right\|_2^2 + \lambda \| U_3 \otimes U_2 \otimes U_1 \|_1 \tag{8}$$

效果：

- 通过 L1 正则化，使得因子载荷矩阵 U_i 具有稀疏性。
- 变量选择能力增强，提高模型的可解释性和计算效率。

Outline

TVAR 的渐近正态性

平稳性条件：自回归对应的特征方程

$$\mathcal{A}(z) = I_N - A_1 z - \cdots - A_p z^p \tag{9}$$

所有根必须在单位圆外，即 $|z| > 1$.

TVAR 的渐近正态性

平稳性条件：自回归对应的特征方程

$$\mathcal{A}(z) = I_N - A_1 z - \cdots - A_p z^p \tag{9}$$

所有根必须在单位圆外，即 $|z| > 1$.

示例：一阶自回归 (AR(1)) 过程

$$x_t = a_1 x_{t-1} + \varepsilon_t \ \text{的特征根为}\ z = a_1^{-1} \tag{10}$$

情况 1： $|a_1| < 1$ 如 $a_1 = 0.5$，则 $x_t = 0.5x_{t-1} + \varepsilon_t$（平稳）

情况 2： $|a_1| > 1$ 如 $a_1 = 1.2$，则 $x_t = 1.2x_{t-1} + \varepsilon_t$（不平稳）

TVAR 的渐近正态性

MLR 估计量的定义：

$$\hat{\mathcal{A}}_{\text{MLR}} = \arg\min_{\mathcal{G},U_1,U_2,U_3} \frac{1}{T}\sum_{t=1}^{T} \|y_t - (\mathcal{G} \times_1 U_1 \times_2 U_2 \times_3 U_3)_{(1)} x_t\|_2^2. \quad (11)$$

定理 1（渐近正态性）： 若时间序列 $\{y_t\}$ 由模型 (5) 生成，且：

- $\mathbb{E}\|\varepsilon_t\|_2^4 < \infty$，
- N, P 固定，
- (r_1, r_2, r_3) 已知，

则在**平稳性假设**下，估计量 $\hat{\mathcal{A}}_{\text{MLR}}$ 满足渐近正态分布：

$$\sqrt{T}\left\{\text{vec}(\hat{\mathcal{A}}_{\text{MLR}(1)}) - \text{vec}(\mathcal{A}_{(1)})\right\} \xrightarrow{d} N(0, \Sigma_{\text{MLR}}), \quad (12)$$

其中：

$$\Sigma_{\text{MLR}} = H(H^\top J H)^\dagger H^\top. \quad (13)$$

$\dagger$ 表示 Moore-Penrose 伪逆。

SHORR 的误差界

4 个假设条件：
假设 1（高斯误差）： 误差项 ε_t 服从均值为 0、协方差矩阵 Σ_ε 正定的独立同分布高斯分布：$\varepsilon_t \sim \mathcal{N}(0, \Sigma_\varepsilon)$.
假设 2（稀疏性）： 因子矩阵 U_i 的每一列最多具有 s_i 个非零元素，$i = 1, 2, 3$。
假设 3（正交性约束）： 因子矩阵 U 满足：

$$U \in \mathbb{R}^{p_i \times r_i}, \quad U^\top U = I_{r_i}, \quad U_{ij}^2 \geq \nu > 0 \text{ 或} U_{ij} = 0.$$

其中 $p_1 = p_2 = N, p_3 = P$，ν 是下界。
假设 4（相对谱间隙）： 张量 $\mathcal{A}$ 的三个 mode 展开矩阵的奇异值满足

$$\sigma_{j-1}^2(\mathcal{A}_{(i)}) - \sigma_j^2(\mathcal{A}_{(i)}) \geq \delta \sigma_{j-1}^2(\mathcal{A}_{(i)}),$$

其中 $\delta > 0$ 控制奇异值的间隔大小。

SHORR 的误差界

定理 2（误差界）：设**平稳性假设**和**假设条件 1-4** 成立，且秩参数 (r_1, r_2, r_3) 已知。若正则化参数 λ 和样本数 T 满足：

$$\lambda \gtrsim \mathcal{M}\sqrt{\frac{\log(N^2P)}{T}}, \quad T \gtrsim \log(N^2P)+\mathcal{M}^2 d\min[\log(NP), \log(cNP/d)],$$

则估计量的误差界满足：

1. Frobenius 误差界：

$$\|\hat{\mathcal{A}}_{\text{SHORR}} - \mathcal{A}\|_F \leq C_1\tau\frac{\sqrt{S}\lambda}{\alpha}. \tag{14}$$

2. 预测误差界：

$$T^{-1}\sum_{t=1}^{T}\|(\hat{\mathcal{A}}_{\text{SHORR}} - \mathcal{A})_{(1)}x_t\|_2^2 \leq C_2\tau^2\frac{S\lambda^2}{\alpha}. \tag{15}$$

3. 概率保证：

$$1 - C\exp(-c\log(N^2P)) - C\exp\{-cd\min[\log(NP), \log(cNP/d)]\}.$$

Outline

数值验证 MLR 和 SHORR 的理论性质

实验目的：评估 MLR 和 SHORR 的估计误差和方差，比较不同方法的估计性能

比较方法：

(1) 传统最小二乘回归（OLS）：

$$\min_{A} \sum_{t} \|y_t - Ax_t\|_2^2$$

(2) 低维自回归模型（ARRR）：

$$\min_{A} \sum_{t} \|y_t - Ax_t\|_2^2, \quad \text{s.t. } \operatorname{rank}(A) \leq r$$

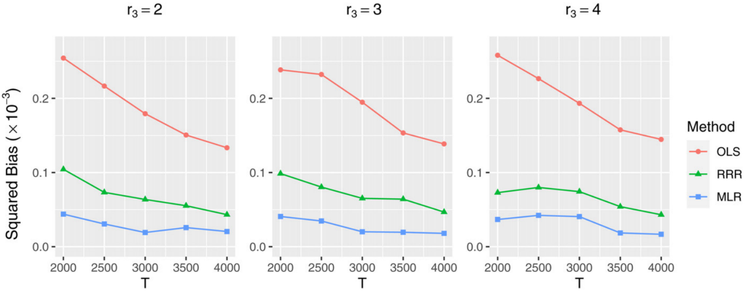
(3) 本文的 MLR 方法：

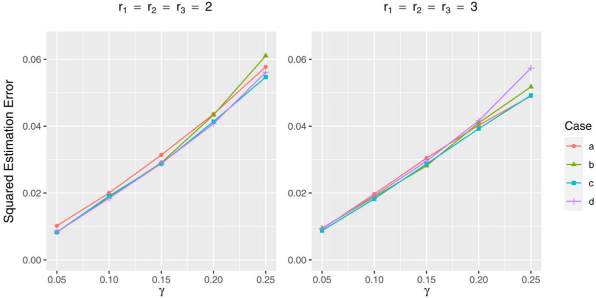
$$\min_{\mathcal{G}, U_1, U_2, U_3} \sum_{t} \|y_t - (\mathcal{G} \times_1 U_1 \times_2 U_2 \times_3 U_3)_{(1)} x_t\|_2^2$$

(4) 本文的 SHORR 方法：

$$\min_{\mathcal{G}, U_1, U_2, U_3} \sum_{t} \|y_t - (\mathcal{G} \times_1 U_1 \times_2 U_2 \times_3 U_3)_{(1)} x_t\|_2^2 + \lambda \|U_3 \otimes U_2 \otimes U_1\|_1$$

数值验证 MLR 和 SHORR 的理论性质





对比基于稀疏/低秩的 VAR 方法

1. Lasso (L1 正则化回归)

$$\min_A \sum_t \|y_t - Ax_t\|_2^2 + \lambda \|A\|_1.$$

2. 核范数 (Nuclear Norm, NN)

$$\min_A \sum_t \|y_t - Ax_t\|_2^2 + \lambda \|A\|_*.$$

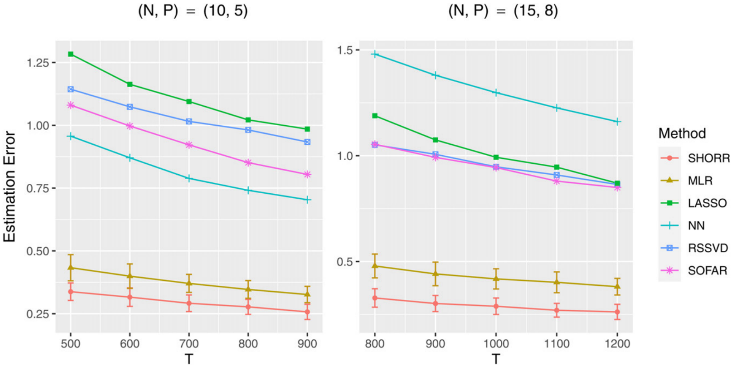
3. 稀疏奇异值分解回归 (RSSVD)

$$\min_{S,U,V} \sum_t \|y_t - USVx_t\|_2^2 + \lambda \|U\|_1 + \lambda \|V\|_1 \quad \text{s.t. } U^T U = I, V^T V = I.$$

4. 稀疏正交因子回归 (SOFAR)

$$\min_{B,U} \sum_t \|y_t - UBx_t\|_2^2 + \lambda \|U\|_1, \quad \text{s.t. } U^T U = I.$$

对比结果



(N, P) = (10, 5)
(N, P) = (15, 8)
Estimation Error
1.25
1.00
0.75
0.50
0.25
500
600
700
800
900
T
1.5
1.0
0.5
800
900
1000
1100
1200
T
Method
SHORR
MLR
LASSO
NN
RSSVD
SOFAR