Autoregressive Models for Tensor-Valued Time Series

Zebang Li[1] and Han Xiao[2], Department of Statistics, Rutgers University

Objectives

Develop an tensor autoregressive model for time series.

- Introducing estimators based on projection, least squares, and maximum likelihood.
- Establishing asymptotics in fix and high dimensions.
- Establishing the model selection consistency.

Introduction

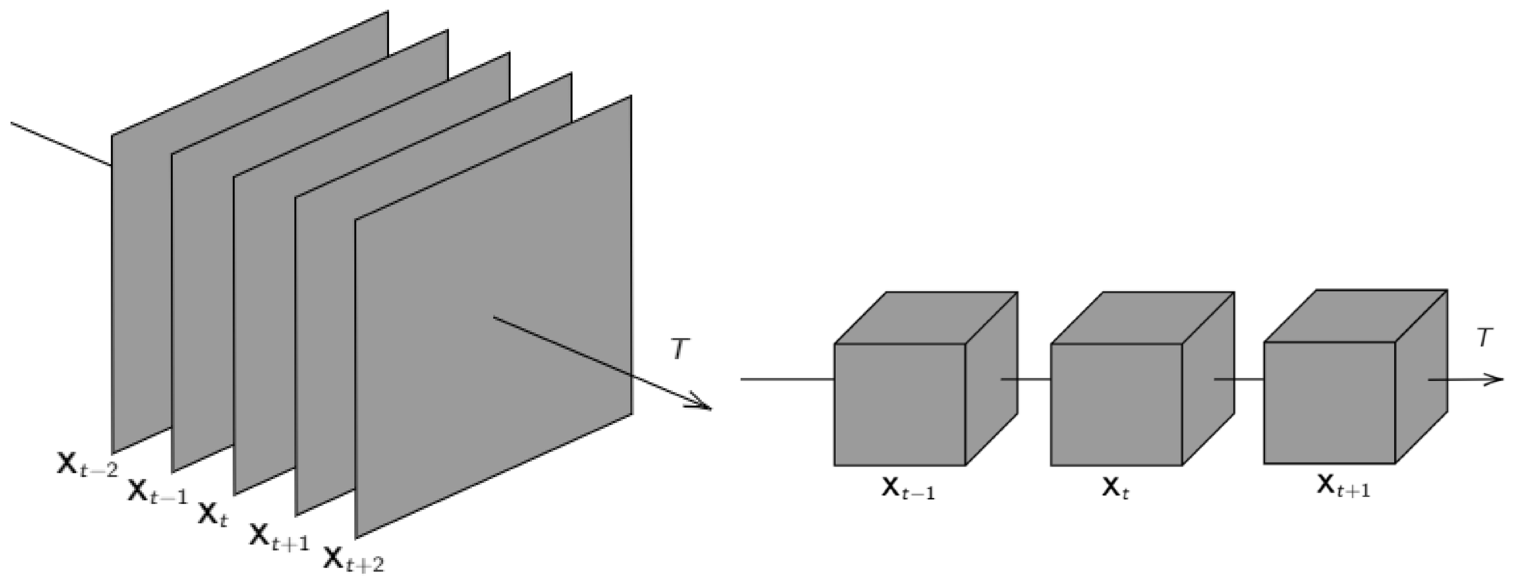

Figure 1:Matrix and Tensor-valued time series.

Our tensor autoregressive model:

- Generalizes the vector and matrix model.
- Preserves tensor structures and interpretations.
- Introduces a more flexible modeling framework with multiple terms and multiple lags.

Tensor Autogressive Models

Consider tensor time series $\{\mathcal{X}_t\}$, where at each time t, an order-K tensor $\mathcal{X}_t \in \mathbb{R}^{d_1\times d_2\times\cdots\times d_K}$ is observed. We introduce the tensor autoregressive model TenAR(1),

$$\mathcal{X}_t = \sum_{r=1}^{R} \mathcal{X}_{t-1} \times_1 \boldsymbol{A}_1^{(r)} \times_2 \cdots \times_K \boldsymbol{A}_K^{(r)} + \mathcal{E}_t,$$

where $\boldsymbol{A}_k^{(r)} \in \mathbb{R}^{d_k\times d_k}$ are coefficient matrices, and $\mathcal{E}_t \in \mathbb{R}^{d_1\times d_2\times\cdots\times d_K}$ is a tensor white noise. The model can be extended directly to TenAR(p),

$$\mathcal{X}_t = \sum_{i=1}^{p}\sum_{r=1}^{R_i} \mathcal{X}_{t-i} \times_1 \boldsymbol{A}_1^{(ir)} \times_2 \cdots \times_K \boldsymbol{A}_K^{(ir)} + \mathcal{E}_t.$$

[1] zl326@stat.rutgers.edu; [2] hxiao@stat.rutgers.edu

Estimation and Asymptotics

- Alternating least squares (LSE):

$$\min_{\boldsymbol{A}_1,\cdots,\boldsymbol{A}_K} \sum_{t=2}^{T} \|\mathcal{X}_t - \mathcal{X}_{t-1} \times_1 \boldsymbol{A}_1 \cdots \times_K \boldsymbol{A}_K\|_F^2.$$

- Gaussian MLE when $\mathrm{Cov}(\mathrm{vec}(\mathcal{E}_t))$ is separable:

$$\mathrm{Cov}(\mathrm{vec}(\mathcal{E}_t)) = \Sigma_K \otimes \Sigma_{K-1} \otimes \cdots \otimes \Sigma_1.$$

Theorem: Asymptotics

- Central Limit theorems are established.
- Convergence rates under high dimensionality: let $\Phi^{(i)} = \Sigma_{r=1}^{R_i} \boldsymbol{A}_K^{(ir)} \otimes \cdots \otimes \boldsymbol{A}_1^{(ir)}$, and $\hat{\Phi}^{(i)} = \Sigma_{r=1}^{R_i} \hat{\boldsymbol{A}}_K^{(ir)} \otimes \cdots \otimes \hat{\boldsymbol{A}}_1^{(ir)}$ be its corresponding estimator, constructed using the LSE.
- Assume the error tensors $\mathcal{E}_t$ are IID sub-Gaussian.
- Assume $d \log d/T \to 0$, where $d = d_1 d_2 \cdots d_K$.

$$\|\hat{\Phi}^{(i)} - \Phi^{(i)}\| = O_p(\sqrt{d/T}), \quad 1 \le i \le p.$$

Initialization of the Algorithm

- Initial values of the alternating algorithms (for both LSE and MLE) is critical to find the global minimum.
- Use projection method to initialize the $\boldsymbol{A}_K$ and Σ_k.
- Projection Method:

$$\min_{\boldsymbol{A}_1,\cdots,\boldsymbol{A}_K} \|\hat{\Phi} - \boldsymbol{A}_K \otimes \cdots \otimes \boldsymbol{A}_1\|_F^2$$

where $\hat{\Phi}$ is estimated coefficient matrix of VAR model.

$$\min_{\Sigma_1,\cdots,\Sigma_K} \|\hat{\Sigma} - \Sigma_K \otimes \cdots \otimes \Sigma_1\|_F^2$$

where $\hat{\Sigma}$ is residual sample covariance matrix.

Model Selection

For given order $\tilde{p}$ and K-ranks $\tilde{\boldsymbol{R}}_{\tilde{p}} = (\tilde{R}_1, \ldots, \tilde{R}_{\tilde{p}})'$, define the estimated K-ranks and the information criterion as

$$\hat{\boldsymbol{R}}_{\hat{p}} := \underset{\tilde{p}\le P_{\max}, \tilde{R}_1\le R_{\max},\ldots,\tilde{R}_{\tilde{p}}\le R_{\max}}{\arg\min} \mathrm{IC}(\tilde{\boldsymbol{R}}_{\tilde{p}}),$$

$$\mathrm{IC}(\tilde{\boldsymbol{R}}_{\tilde{p}}) := \frac{1}{2}\log\left(\frac{1}{dT}\sum_t \left\|\mathcal{X}_t - \sum_{i=1}^{\tilde{p}}\sum_{r=1}^{\tilde{R}_i} \mathcal{X}_{t-i} \times_1 \hat{\boldsymbol{A}}_1^{(ir)} \times_2 \cdots \times_K \hat{\boldsymbol{A}}_K^{(ir)}\right\|_F^2\right) + g(d,T)\sum_{i=1}^{\tilde{p}} \tilde{R}_i,$$

where $\hat{\boldsymbol{A}}_k^{(ir)}$ are estimates obtained under given K-ranks.

Theorem: Model Selection

- $g(d,T) \to 0$ and $\frac{T}{d} g(d,T) \to \infty$ as $T \to \infty$.
- Assume $d \log d/T \to 0$, where $d = d_1 d_2 \cdots d_K$.

Then we can select the right K-ranks consistently,

$$\lim_{T\to\infty} P(\hat{\boldsymbol{R}}_{\hat{p}} = \boldsymbol{R}_p) = 1.$$

Numerical Results

- Fama-French portfolio monthly returns: $2\times 4\times 4$ portfolio returns formed on Size, Book-to-Market Ratio and Operating Profitability.
- Manhattan taxi traffic data: number of rides moving among four zones (Midtown Center, Midtown East, Midtown North, Time Squares) between 8am and 3pm in a day, formed a $4\times 4\times 7$ tensor.

	LSE	MLE	iAR	VAR	MEAN	TOTAL
FF	36.63	36.49	36.91	38.76	36.96	37.48
Taxi	47.75	47.07	49.94	51.33	53.94	56.48

Table 1:Mean squared errors of rolling forecast. First row is for Fama-French data, from 2001 Jan to 2018 Dec; and second row is for NYC taxi traffic data on business-day, from 2018 Jan to 2019 Dec.

- LSE, MLE denotes TenAR(5) model with $\boldsymbol{R}_5 = (1,1,1,1,1)$ for taxi data; TenAR(1) with $R = 1$ for Fama-French data.
- iAR denotes individual AR(5) model for taxi data; individual AR(1) for Fama-French data.
- VAR denotes VAR(2) model for taxi data; VAR(1) for Fama-French data.
- MEAN: using sample means as the predictions.
- TOTAL denotes total sum of squares.