Summary on basic time series studies

tensor data analysis with different data types

Haofan Zheng

Contents

1 High-dimentional α-PCA method

1.1 Overall Summary

This article considers the estimation and inference of the **low rank** components in high-dimentional matrixvariate models(tensor), and we propose an estimation method called α-PCA and it has some benefits with the high dimensions data favorably compared with other methods(traditional PCA, etc) based on the perfomance in the simulation.

1.2 Main model

The model is shown as the following:

$$\mathbf{Y}_t = \underbrace{\mathbf{R}\mathbf{F}_t\mathbf{C}^T}_{\text{signal part}} + \underbrace{\mathbf{E}_t}_{\text{noise part}} \tag{1}$$

$\mathbf{Y}_t$: $\mathbf{Y}_t \in \mathbb{R}^{p\times q}$, $1 \leq t \leq T$, observations,

$\mathbf{F}_t$: $\mathbf{F}_t \in \mathbb{R}^{k\times r}$, where $k \ll p$ and $r \ll q$ (**low rank**), latent matrix,

$\mathbf{E}_t$: $\mathbf{E}_t \in \mathbb{R}^{p\times q}$, noise matrix.

1.3 Main Statistics

An estimation procedure, namely α-PCA, aggregates the information in both first and second moments. Specifically, the two statistics are defined:

$$\widehat{\mathbf{M}}_R \triangleq \frac{1}{pq}\left((1+\alpha)\cdot \overline{\mathbf{Y}}\,\overline{\mathbf{Y}}^T + \frac{1}{T}\sum_{t=1}^{T}(\mathbf{Y}_t - \overline{\mathbf{Y}})(\mathbf{Y}_t - \overline{\mathbf{Y}})^T\right) \tag{2}$$

$$\widehat{\mathbf{M}}_C \triangleq \frac{1}{pq}\left((1+\alpha)\cdot \overline{\mathbf{Y}}^T\overline{\mathbf{Y}} + \frac{1}{T}\sum_{t=1}^{T}(\mathbf{Y}_t - \overline{\mathbf{Y}})^T(\mathbf{Y}_t - \overline{\mathbf{Y}})\right) \tag{3}$$

α : $\alpha \in [-1, +\infty)$, a hyperparameter,

$\overline{\mathbf{Y}} = \frac{1}{T}\sum_{i=1}^{T}\mathbf{Y}_t$, the sample mean.

Based on these two statistics, estimation of $\mathbf{R}$ and $\mathbf{C}$ can be obtained as $\sqrt{p}$ times the top k eigenvectors of $\widehat{\mathbf{M}}_R$ and $\sqrt{q}$ times the top q eigenvectors of $\widehat{\mathbf{M}}_C$ respectively, in descending order by corresponding eigenvalues.

1.4 Transformation

To simplify the estimator, we can transform the parameters, let the $\tilde{\alpha} = \sqrt{\alpha+1} - 1$, $\widetilde{\mathbf{Y}}_t \triangleq \mathbf{Y}_t + \tilde{\alpha}\overline{\mathbf{Y}}$, $\widetilde{\mathbf{F}}_t \triangleq \mathbf{F}_t + \tilde{\alpha}\overline{\mathbf{F}}$, and $\widetilde{\mathbf{E}}_t \triangleq \mathbf{E}_t + \tilde{\alpha}\overline{\mathbf{E}}$, Then we have

$$\widetilde{\mathbf{Y}}_t = \mathbf{R}\widetilde{\mathbf{F}}_t\mathbf{C}^T + \widetilde{\mathbf{E}}_t \tag{4}$$

The equation 2 and 3 can be rewritten as:

$$\widehat{\mathbf{M}}_R = \frac{1}{pqT}\sum_{t=1}^{T}\widetilde{\mathbf{Y}}_t\widetilde{\mathbf{Y}}_t^T, \text{ and } \widehat{\mathbf{M}}_C = \frac{1}{pqT}\sum_{t=1}^{T}\widetilde{\mathbf{Y}}_t^T\widetilde{\mathbf{Y}}_t \tag{5}$$

Same as in section 1.3, $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{C}}$ can be obtained as $\sqrt{p}$ times the top k eigenvectors of $\widehat{\mathbf{M}}_R$ and $\sqrt{q}$ times the top q eigenvectors of $\widehat{\mathbf{M}}_C$ respectively, in descending order by corresponding eigenvalues.

1.5 Interpretation

The estimator in Section 1.2 approximately minimized jointly the unexplained variation and bias

$$\begin{aligned} \underset{\mathbf{R},\mathbf{C},\{\mathbf{F}_t\}_{t=1}^T}{\textbf{minimize}} \quad & (1+\alpha)\underbrace{\frac{1}{pq}\left\|\overline{\mathbf{Y}} - \mathbf{R}\overline{\mathbf{F}}\mathbf{C}^T\right\|_F^2}_{\text{sample bias}} + \underbrace{\frac{1}{pqT}\sum_{t=1}^{T}\left\|\mathbf{Y}_t - \mathbf{R}\mathbf{F}\mathbf{C}^T\right\|_F^2}_{\text{sample variance}} \\ \textbf{subject to} \quad & \frac{1}{p}\mathbf{R}^T\mathbf{R} = \mathbf{I}\ ,\ \frac{1}{q}\mathbf{C}^T\mathbf{C} = \mathbf{I} \end{aligned} \tag{6}$$

The special case for $\alpha = -1$ corresponds to the least-square estimator.(***not convex***)

Projecting on $\mathbf{R}$:

$$\begin{aligned} \underset{\mathbb{R}}{\textbf{maximize}} \quad & Tr\left(\mathbb{E}\left[(1+\alpha)(\mathbf{R}^T\overline{\mathbf{Y}})(\mathbf{R}^T\overline{\mathbf{Y}})^T + (\mathbf{R}^T\mathbf{Y}_t - \mathbb{E}\left[R^T\mathbf{Y}_t\right])(\mathbf{R}^T\mathbf{Y}_t - \mathbb{E}\left[R^T\mathbf{Y}_t\right])^T\right]\right) \\ \textbf{subject to} \quad & \frac{1}{p}\mathbf{R}^T\mathbf{R} = \mathbf{I}, \frac{1}{q}\mathbf{C}^T\mathbf{C} = \mathbf{I} \end{aligned} \tag{7}$$

Where $\mathbf{M}_R \triangleq (1+\alpha)\mathbf{M}_R^{(1)} + \mathbf{M}_R^{(2)}, \mathbf{M}_R^{(1)} \triangleq \frac{1}{pq}\mathbb{E}\left[\overline{\mathbf{Y}}\overline{\mathbf{Y}}^T\right]$, and $\mathbf{M}_R^{(2)} \triangleq \frac{1}{pq}\mathbb{E}\left[\left(\mathbf{Y}_t - [\overline{\mathbf{Y}}]\right)\left(\mathbf{Y}_t - [\overline{\mathbf{Y}}]\right)^T\right]$
Then a solution by maximizing row and column variances respectively after projection is considered, projecting on $\mathbf{C}$ is similar.(***convex***)

1.6 Relative estimators

Based on the section 1.4,

$$\widehat{\mathbf{F}}_t = \frac{1}{pq}\widehat{\mathbf{R}}^T\widehat{\mathbf{Y}}_t\widehat{\mathbf{C}}, \text{ and the signal part } \widehat{\mathbf{S}}_t = \frac{1}{pq}\widehat{\mathbf{R}}\widehat{\mathbf{R}}^T\widehat{\mathbf{Y}}_t\widehat{\mathbf{C}}\widehat{\mathbf{C}}^T$$

Dimensions k and r are need to be determined:

1. the eigenvalue ratio-based estimator, proposed by Ahn and Horestein(2013)
2. the Scree plot which is standard in principal component analysis.

Let $\hat{\lambda}_1 \geq \hat{\lambda} \geq \cdots \geq \hat{\lambda}_k \geq 0$ be the ordered eigenvalues of $\widehat{\mathbf{M}}_R$. The ratio-based estimator for k is defined as follows:

$$\widehat{k} = \underset{1\leq j\leq k_{max}}{\arg\max} \frac{\widehat{\lambda}_j}{\widehat{\lambda}_{j+1}}$$

where k_{max} is the upper bound, usually taken as $\left\lceil \frac{p}{2} \right\rceil$ or $\left\lceil \frac{p}{3} \right\rceil$, according to Ahn and Horestein(2013), similarly for $\widehat{r}$ with respect to $\widehat{\mathbf{M}}_C$.

1.7 Theoretical Properties

1.7.1 Denotation and Definition

Before presenting the assumptions and theorems, some quantities need to be defined or denoted,

1. Let $\mathbf{V}_{R,pqT} \in \mathbb{R}^{k\times k}$ and $\mathbf{V}_{C,pqT} \in \mathbb{R}^{r\times r}$ be the diagonal matrices consisting of the first k and r largest eigenvalues of $\widehat{\mathbf{M}}_R$ and $\widehat{\mathbf{M}}_C$ in Section 1.4 in a decreasing order. By definition of estimators $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{C}}$,

$$\widehat{\mathbf{R}} = \frac{1}{pqT}\sum_{t=1}^{T}\widetilde{\mathbf{Y}}_t\widetilde{\mathbf{Y}}_t^T\widehat{\mathbf{R}}\mathbf{V}_{R,pqT}^{-1} \text{ and } \widehat{\mathbf{C}} = \frac{1}{pqT}\sum_{t=1}^{T}\widetilde{\mathbf{Y}}_t^T\widetilde{\mathbf{Y}}_t\widehat{\mathbf{C}}\mathbf{V}_{C,pqT}^{-1} \tag{8}$$

2. Define $\mathbf{H}_R \in \mathbb{R}^{k\times k}$ and $\mathbf{H}_C \in \mathbb{R}^{r\times r}$ as

$$\mathbf{H}_R \triangleq \frac{1}{pqT}\sum_{t=1}^{T}\widetilde{\mathbf{F}}_t\mathbf{C}^T\mathbf{C}\widetilde{\mathbf{F}}_t^T\mathbf{R}^T\widehat{\mathbf{R}}\mathbf{V}_{R,pqT}^{-1} \in \mathbb{R}^{k\times k} \tag{9}$$

$$\mathbf{H}_C \triangleq \frac{1}{pqT}\sum_{t=1}^{T}\widetilde{\mathbf{F}}_t\mathbf{R}^T\mathbf{R}\widetilde{\mathbf{F}}_t^T\mathbf{C}^T\widehat{\mathbf{C}}\mathbf{V}_{R,pqT}^{-1} \in \mathbb{R}^{r\times r} \tag{10}$$

($\boldsymbol{bounded\ as}$ $p, q, T \to \infty$)

3. Let $\mu_F = \mathbb{E}[\mathbf{F}_t]$ and

$$\Sigma_{FC} \triangleq \mathbb{E}\left[(\mathbf{F}_t - \mu_F)(\frac{\mathbf{C}^T\mathbf{C}}{q})(\mathbf{F}_t - \mu_F)^T\right] \text{ and } \Sigma_{FR} \triangleq \mathbb{E}\left[(\mathbf{F}_t - \mu_F)^T(\frac{\mathbf{R}^T\mathbf{R}}{p})(\mathbf{F}_t - \mu_F)\right] \tag{11}$$

then

$$\begin{aligned} \widetilde{\Sigma}_{FC} &\triangleq \frac{1}{q}\mathbb{E}\left[\widetilde{\mathbf{F}}_t\mathbf{C}^T\mathbf{C}\widetilde{\mathbf{F}}_t^T\right] = \Sigma_{FC} + (\alpha+1)\frac{1}{q}\mu_F\mathbf{C}^T\mathbf{C}\mu_F^T \\ \widetilde{\Sigma}_{FR} &\triangleq \frac{1}{q}\mathbb{E}\left[\widetilde{\mathbf{F}}_t\mathbf{R}^T\mathbf{R}\widetilde{\mathbf{F}}_t^T\right] = \Sigma_{FR} + (\alpha+1)\frac{1}{q}\mu_F\mathbf{R}^T\mathbf{R}\mu_F^T \end{aligned} \tag{12}$$

(***Matrix*** Σ ***can be interpreted as scaled row/column of*** $\mathbf{F}_t$)

1.7.2 Assumptions

Assumption 1. α-***mixing***. The vectorized factor VEC($\mathbf{F}_t$) and noise VEC($\mathbf{E}_t$) are α-mixing. Specifically, a vector process $\{\mathbf{x}_t, t = 0, \pm1, \pm2, \cdots\}$ is α-mixing if, for some $\gamma \geq 2$, the mixing coefficients satisfy the condition that

$$\sum_{h=1}^{+\infty} \alpha(h)^{1-\frac{2}{\gamma}} < \infty$$

where $\alpha(h) = \sup_{\tau} \sup_{A\in\mathcal{F}_{-\infty}^{\tau}, B\in\mathcal{F}_{\tau+h}^{\infty}} |P(A\cap B) - P(A)P(B)|$ and $\mathcal{F}_\tau^s$ is the σ-field generated by $\{\mathbf{x}_t : \tau \leq t \leq s\}$.(***only deal with temporal dependence***)

Assumption 2. *Factor and noise matrices.* There exists a positive constant $C < \infty$ such that for all N and T,

1. Factor matrix $\mathbf{F}_t$ is of fixed dimemsion $k \times r$ and $\mathbb{E}\|\mathbf{F}_t\|^4 \leq C$.

2. For all $i \in [p]$, $j \in [q]$ and $t \in [T]$, $\mathbb{E}[e_{t,ij}] = 0$ and $\mathbb{E}|e_{t,ij}|^8 \leq C$.[1]

3. Factor and noise are uncorrelated, that is, $\mathbb{E}[e_{t,ij}f_{s,lh}] = 0$ for any $t, s \in [T]$, $i \in [p]$, $j \in [q]$, $l \in [k]$, $h \in [r]$.

Assumption 3. *Loading matrix.* For each row of $\mathbf{R}$, $\|\mathbf{R}_{i\cdot}\| = \mathcal{O}(1)$, and, as $p, q \to \infty$, we have $\|p^{-1}\mathbf{R}^T\mathbf{R} - \Omega_R\| \to 0$ for some $k \times k$ positive definite matrix Ω_R. For each row of $\mathbf{C}$, $\|\mathbf{C}_{i\cdot}\| = \mathcal{O}(1)$,and, as $p, q \to \infty$, we have $\|p^{-1}\mathbf{C}^T\mathbf{C} - \Omega_R\| \to 0$ for some $r \times r$ positive definite matrix Ω_C.(***an extension of the pervasive assumption***[2])

[1] $[n] \triangleq \{1, \ldots, n\}$
[2] Stock and Watson 2002

Assumption 4. *Cross row(column) correlation of noise* $\mathbf{E}_t$.There exists some positive constant $C < \infty$ such that,

1. Let $\mathbf{U}_E = \mathbb{E}\left[\frac{1}{qT}\sum_{t=1}^{T}\mathbf{E}_t\mathbf{E}_t^T\right]$ and $\mathbf{V}_E = \mathbb{E}\left[\frac{1}{qT}\sum_{t=1}^{T}\mathbf{E}_t^T\mathbf{E}_t\right]$, we assume $\|\mathbf{U}_E\|_1 \leq C$ and $\|\mathbf{V}_E\|_1 \leq C$.
2. For all row $i \in [p]$ and $j \in [q]$ and $t \in [T]$, we assume $\sum_{l\in p, l\neq i}\sum_{h\in q, h\neq j}|\mathbb{E}\left[e_{t,ij}e_{t,lh}\right]| \leq C$.
3. For any row $i, l \in [p]$, any time $t \in [T]$, and any column $j \in [q]$,

$$\sum_{m\in[p]}\sum_{s\in[T]}\sum_{h\in[q],h\neq j}|cov\left[e_{t,ij}e_{t,lj}, e_{s,ih}e_{s,mh}\right]| \leq C$$

   Similar, for any column $j, h \in [q]$, any time $t \in [T]$, and any row $i \in [p]$,

$$\sum_{m\in[q]}\sum_{s\in[T]}\sum_{l\in[p],l\neq i}|cov\left[e_{t,ij}e_{t,ih}, e_{s,lj}e_{s,lm}\right]| \leq C$$

(***automatically hold when the errors $\mathbf{E}_t$ are i.i.d. over rows and columns for any t, C for weak correlation***)

Assumption 5. $\mathbf{E}_t$.There exists $m > 2$, $1 < a, b < \infty$, $\frac{1}{a} + \frac{1}{b} = 1$, such that, for some positives $C < \infty$,

1. For any $l \in [k]$, $i \in [p]$, and $t \in [T]$, $\mathbb{E}\left[|\frac{1}{\sqrt{q}}\sum_{j=1}^{q}e_{t,ij}|^{mb}\right] = \mathcal{O}(1)$, $\mathbb{E}\left[\|\frac{1}{\sqrt{q}}\sum_{j=1}^{q}\mathbf{C}_{j\cdot}e_{t,ij}\|^{mb}\right] = \mathcal{O}(1)$, and $\mathbb{E}\left[\|\mathbf{f}_{t,l\cdot}\|^{ma}\right] \leq C$
2. For any $h \in [r]$, $j \in [q]$, and $t \in [T]$, $\mathbb{E}\left[|\frac{1}{\sqrt{p}}\sum_{i=1}^{p}e_{t,ij}|^{mb}\right] = \mathcal{O}(1)$, $\mathbb{E}\left[\|\frac{1}{\sqrt{p}}\sum_{i=1}^{p}\mathbf{R}_{i\cdot}e_{t,ij}\|^{mb}\right] = \mathcal{O}(1)$, and $\mathbb{E}\left[\|\mathbf{f}_{t,\cdot h}\|^{ma}\right] \leq C$
3. For any and $t \in [T]$, $\mathbb{E}\left[|\frac{1}{\sqrt{pq}}\sum_{i=1}^{p}\sum_{j=1}^{q}e_{t,ij}|^{mb}\right] = \mathcal{O}(1)$, $\mathbb{E}\left[\|\frac{1}{\sqrt{pq}}\sum_{i=1}^{p}\sum_{j=1}^{q}\mathbf{R}_{i\cdot}\mathbf{C}_{j\cdot}^T e_{t,ij}\|^{mb}\right] = \mathcal{O}(1)$.

(***satisfied by Gaussian noise $\mathbf{E}_t$ with i.i.d. columns and rows***)

Assumption 6.

1.7.3 Theorems

Theorem 1. Under **??** 1–5, we have as k, r fixed and $p, q, T \to \infty$,

$$\frac{1}{p}\|\widehat{\mathbf{R}} - \mathbf{R}\mathbf{H}_R\|_F^2 = \mathcal{O}_p\left(\frac{1}{\min\{p, qT\}}\right)$$

$$\frac{1}{q}\|\widehat{\mathbf{C}} - \mathbf{C}\mathbf{H}_C\|_F^2 = \mathcal{O}_p\left(\frac{1}{\min\{p, qT\}}\right)$$

Consequently,

$$\frac{1}{p}\|\widehat{\mathbf{R}} - \mathbf{R}\mathbf{H}_R\|^2 = \mathcal{O}_p\left(\frac{1}{\min\{p, qT\}}\right)$$

$$\frac{1}{q}\|\widehat{\mathbf{C}} - \mathbf{C}\mathbf{H}_C\|^2 = \mathcal{O}_p\left(\frac{1}{\min\{p, qT\}}\right)$$

(***converge faster than the PCA for the vectorized model***)

1.8 Simulation

1.9 Application

2 High-Dimensional GLM with Binary Outcomes

2.1 Overall Summary

3 Ultra-High Dimensional GFM[3]

3.1 Overall Summary

4 Matrix-variate Logistic Regression with Measurement Error

5 A Likelihood-Based Approach for Multivariate Categorical Response Regression in High Dimensions

6 A likelihood-Based Approach for Semiparametric Regression with Panel Count Data

7 Time Series Latent Gaussian Count

8 Time Series Factor Models(tensor)

[3] Generalized Factor Model