ST418 Week 11

Kaixin Liu[1]
[1]PhD Student in Statistics, LSE

Apr 2, 2026

Table of Contents

Table of Contents

Granger causality

Let $\mathbf{w}_t = (\mathbf{x}_t^\top, \mathbf{y}_t^\top, \mathbf{z}_t^\top)^\top$. Let $\mathcal{F}_t$ be the information up to time t, and let $\mathcal{F}_{-x,t}$ be the same information with the history of $\{\mathbf{x}_t\}$ removed.

Definition (Granger causality)

For an h-step forecast, define

$$\mathbf{z}_t(h) = \mathbb{E}(\mathbf{z}_{t+h} \mid \mathcal{F}_t), \qquad \mathbf{z}_{-x,t}(h) = \mathbb{E}(\mathbf{z}_{t+h} \mid \mathcal{F}_{-x,t}).$$

We say that $\{\mathbf{x}_t\}$ **Granger-causes** $\{\mathbf{z}_t\}$ conditional on $\{\mathbf{y}_t\}$ if

$$\mathrm{MSPE}\big(\mathbf{z}_t(h)\big) < \mathrm{MSPE}\big(\mathbf{z}_{-x,t}(h)\big).$$

Claim: If the best forecast $\mathbf{z}_t(h)$ depends explicitly on $\{\mathbf{x}_s : s \leq t\}$, then $\{\mathbf{x}_t\}$ Granger-causes $\{\mathbf{z}_t\}$ conditional on $\{\mathbf{y}_t\}$.
Interpretation: $\{\mathbf{x}_t\}$ contains predictive information for $\{\mathbf{z}_t\}$ beyond what is already in $\{\mathbf{y}_t\}$ and the past of $\{\mathbf{z}_t\}$.

Table of Contents

Latent vector factor model

Motivation: when p is large and T is moderate, unrestricted covariance or VARMA estimation becomes unstable. A low-dimensional factor structure regularises the problem.

Vector factor model

$$\mathbf{r}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\epsilon}_t, \qquad t = 1, \ldots, T,$$

where $\mathbf{f}_t \in \mathbb{R}^K$ is latent, $\mathbf{B} \in \mathbb{R}^{p \times K}$ is the loading matrix, and $\boldsymbol{\epsilon}_t$ is the idiosyncratic noise.

If $\mathrm{Cov}(\mathbf{f}_t, \boldsymbol{\epsilon}_t) = \mathbf{0}$, then

$$\boldsymbol{\Sigma} := \mathrm{Var}(\mathbf{r}_t) = \mathbf{B}\boldsymbol{\Sigma}_f\mathbf{B}^\top + \boldsymbol{\Sigma}_\varepsilon.$$

- **Strict factor model:** $\boldsymbol{\Sigma}_\varepsilon$ is diagonal.
- **Approximate factor model:** $\boldsymbol{\Sigma}_\varepsilon$ may have small/sparse off-diagonal entries.

Estimation of latent vector factor model

Assume K is fixed, $\mathbf{B}^\top \mathbf{B} = \mathbf{I}_K$, and $\bar{\mathbf{f}} = \mathbf{0}$. Estimate $(\boldsymbol{\alpha}, \mathbf{B}, \mathbf{f}_1, \ldots, \mathbf{f}_T)$ by least squares:

$$\min_{\boldsymbol{\alpha}, \mathbf{B}, \mathbf{f}_t \colon \mathbf{B}^\top \mathbf{B} = \mathbf{I}_K} \frac{1}{T} \sum_{t=1}^{T} \|\mathbf{r}_t - \boldsymbol{\alpha} - \mathbf{B}\mathbf{f}_t\|^2.$$

Least-squares result:

$$\widehat{\boldsymbol{\alpha}} = \bar{\mathbf{r}}, \qquad \widehat{\mathbf{f}}_t = \widehat{\mathbf{B}}^\top (\mathbf{r}_t - \bar{\mathbf{r}}),$$

where $\widehat{\mathbf{B}}$ contains the K eigenvectors corresponding to the K largest eigenvalues of

$$\widehat{\boldsymbol{\Sigma}}_r = \frac{1}{T} \sum_{t=1}^{T} (\mathbf{r}_t - \bar{\mathbf{r}})(\mathbf{r}_t - \bar{\mathbf{r}})^\top.$$

Hence estimation reduces to **PCA**. For the strict factor model,

$$\widehat{\boldsymbol{\Sigma}}_\varepsilon = \mathrm{diag}\left(\frac{1}{T} \sum_{t=1}^{T} \widehat{\boldsymbol{\epsilon}}_t \widehat{\boldsymbol{\epsilon}}_t^\top \right), \quad \widehat{\boldsymbol{\epsilon}}_t = \mathbf{r}_t - \widehat{\boldsymbol{\alpha}} - \widehat{\mathbf{B}}\widehat{\mathbf{f}}_t.$$

Why the need to restrict $\boldsymbol{\Sigma}_\varepsilon$

- A general covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ has $p(p+1)/2$ free parameters.
- Under

$$\boldsymbol{\Sigma} = \mathbf{B}\boldsymbol{\Sigma}_f\mathbf{B}^\top + \boldsymbol{\Sigma}_\varepsilon, \qquad \mathrm{Cov}(\mathbf{f}_t, \boldsymbol{\epsilon}_t) = \mathbf{0},$$

  the number of unknowns becomes

$$m_p = pK + \frac{K(K+1)}{2} + m_p(\boldsymbol{\Sigma}_\varepsilon).$$

- If $\boldsymbol{\Sigma}_\varepsilon$ is diagonal, then $m_p(\boldsymbol{\Sigma}_\varepsilon) = p$ and

$$m_p = (p + K/2)(K+1).$$

- Example: for $p = 100$ and $K = 1$, we go from 5050 parameters to 201.
- If $\boldsymbol{\Sigma}_\varepsilon$ is sparse, then $m_p(\boldsymbol{\Sigma}_\varepsilon)$ can still be of order p.

Without a restriction on $\boldsymbol{\Sigma}_\varepsilon$, the factor model does not achieve meaningful regularisation.

Table of Contents

Matrix factor model

Motivation: when data are naturally arranged in a matrix, stacking ignores the row/column structure and turns dimensions (p, q) into pq.

Matrix factor model

$$\mathbf{Y}_t = \mathbf{C}_t + \mathbf{E}_t = \mathbf{A}\mathbf{F}_t\mathbf{B}^\top + \mathbf{E}_t, \qquad t = 1, \ldots, T,$$

where

$$\mathbf{A} \in \mathbb{R}^{p \times k_r}, \qquad \mathbf{B} \in \mathbb{R}^{q \times k_c}, \qquad \mathbf{F}_t \in \mathbb{R}^{k_r \times k_c},$$

with $k_r \ll p$ and $k_c \ll q$.

- $\mathbf{A}$: row loading matrix.
- $\mathbf{B}$: column loading matrix.
- $\mathbf{F}_t$: low-dimensional factor matrix.

The common dynamics are carried jointly by row effects, column effects, and the core factor matrix $\mathbf{F}_t$.

Estimation of latent matrix factor model: HOSVD

Write

$$\mathbf{Y} = [\mathbf{Y}_1, \dots, \mathbf{Y}_T] \in \mathbb{R}^{p \times qT}, \qquad \mathbf{Y}_v = [\mathbf{Y}_1^\top, \dots, \mathbf{Y}_T^\top] \in \mathbb{R}^{q \times pT}.$$

Without loss of generality, assume

$$\mathbf{A}^\top \mathbf{A} = \mathbf{I}_{k_r}, \qquad \mathbf{B}^\top \mathbf{B} = \mathbf{I}_{k_c}.$$

In the noise-free case,

$$\mathbf{Y} = \mathbf{A}[\mathbf{F}_1 \mathbf{B}^\top, \dots, \mathbf{F}_T \mathbf{B}^\top], \qquad \mathbf{Y}_v = \mathbf{B}[\mathbf{F}_1^\top \mathbf{A}^\top, \dots, \mathbf{F}_T^\top \mathbf{A}^\top].$$

Hence col($\mathbf{A}$) and col($\mathbf{B}$) are the leading left-singular subspaces of $\mathbf{Y}$ and $\mathbf{Y}_v$.

HOSVD estimator

1. $\widehat{\mathbf{A}}$: first k_r left singular vectors of $\mathbf{Y}$.
2. $\widehat{\mathbf{B}}$: first k_c left singular vectors of $\mathbf{Y}_v$.
3. $\widehat{\mathbf{F}}_t = \widehat{\mathbf{A}}^\top \mathbf{Y}_t \widehat{\mathbf{B}}$.

Estimation of latent matrix factor model: HOOI

HOOI estimator

Start from $\widehat{\mathbf{A}}^{(0)}$ from HOSVD. For $i \geq 1$:

1. Form
$$\mathbf{P}_{i,v} = [\mathbf{Y}_1^\top \widehat{\mathbf{A}}^{(i-1)}, \ldots, \mathbf{Y}_T^\top \widehat{\mathbf{A}}^{(i-1)}]$$
and take $\widehat{\mathbf{B}}^{(i)}$ as its first k_c left singular vectors.
2. Form
$$\mathbf{P}_i = [\mathbf{Y}_1 \widehat{\mathbf{B}}^{(i)}, \ldots, \mathbf{Y}_T \widehat{\mathbf{B}}^{(i)}]$$
and take $\widehat{\mathbf{A}}^{(i)}$ as its first k_r left singular vectors.

After convergence,
$$\widehat{\mathbf{F}}_t = \widehat{\mathbf{A}}^\top \mathbf{Y}_t \widehat{\mathbf{B}}.$$

Projection reduces the effective noise dimension before re-estimating the other loading space.

Matrix factor model or just factor model

Let $\mathbf{y}_t = \mathbf{vec}(\mathbf{Y}_t)$. A general vector factor model for the stacked data is

$$\mathbf{y}_t = \mathbf{C}\mathbf{f}_t + \epsilon_t,$$

with unrestricted loading matrix $\mathbf{C}$.
If the matrix factor model holds, then

$$\mathbf{y}_t = \mathbf{vec}(\mathbf{A}\mathbf{F}_t\mathbf{B}^\top) + \mathbf{vec}(\mathbf{E}_t) = (\mathbf{B} \otimes \mathbf{A})\mathbf{vec}(\mathbf{F}_t) + \mathbf{vec}(\mathbf{E}_t).$$

Hence

$$\mathbf{f}_t = \mathbf{vec}(\mathbf{F}_t), \qquad \mathbf{C} = \mathbf{B} \otimes \mathbf{A}, \qquad r = k_r k_c.$$

Key point

A matrix factor model is a vector factor model with a **Kronecker product restriction** on the loading matrix.

Gain: fewer parameters and interpretable row/column effects. Cost: less flexibility than an unrestricted vector factor model.