Modelling Multiple Time Series via Common Factors

Jiazhu Pan and Qiwei Yao

Department of Statistics

London School of Economics

q.yao@lse.ac.uk

1. Models
2. Estimation method — An algorithm: expanding WN space
3. Illustration by simulation
4. Asymptotic properties
5. Illustration with real data sets

Let $\{\mathbf{Y}_t\}$ be a $d \times 1$ time series defined by

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t + \boldsymbol{\varepsilon}_t,$$

Let $\{\mathbf{Y}_t\}$ be a $d \times 1$ time series defined by

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t + \boldsymbol{\varepsilon}_t,$$

$\mathbf{X}_t$: $r \times 1$ unobservable factors, r $(\leq d)$ unknown

$\mathbf{A}$: $d \times r$ unknown constant factor loading matrix

$\{\boldsymbol{\varepsilon}_t\}$: vector WN($\boldsymbol{\mu}_\varepsilon, \boldsymbol{\Sigma}_\varepsilon$)

no linear combinations of $\mathbf{X}_t$ are WN.

Let $\{\mathbf{Y}_t\}$ be a $d \times 1$ time series defined by

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t + \boldsymbol{\varepsilon}_t,$$

$\mathbf{X}_t$: $r \times 1$ unobservable factors, $r \ (\leq d)$ unknown

$\mathbf{A}$: $d \times r$ unknown constant factor loading matrix

$\{\boldsymbol{\varepsilon}_t\}$: vector WN($\boldsymbol{\mu}_\varepsilon, \boldsymbol{\Sigma}_\varepsilon$)

no linear combinations of $\mathbf{X}_t$ are WN.

Lack of identification: $(\mathbf{A}, \mathbf{X}_t)$ may be replaced by $(\mathbf{A}\mathbf{H}, \mathbf{H}^{-1}\mathbf{X}_t)$ for any invertible $\mathbf{H}$.

Therefore, we assume $\mathbf{A}^\tau \mathbf{A} = \mathbf{I}_r$

But factor loading space $\mathcal{M}(\mathbf{A})$ is uniquely defined

Let $\{\mathbf{Y}_t\}$ be a $d \times 1$ time series defined by

$$\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t + \boldsymbol{\varepsilon}_t,$$

$\mathbf{X}_t$: $r \times 1$ unobservable factors, r $(\leq d)$ unknown

$\mathbf{A}$: $d \times r$ unknown constant factor loading matrix

$\{\boldsymbol{\varepsilon}_t\}$: vector WN($\boldsymbol{\mu}_\varepsilon, \boldsymbol{\Sigma}_\varepsilon$)

no linear combinations of $\mathbf{X}_t$ are WN.

Lack of identification: $(\mathbf{A}, \mathbf{X}_t)$ may be replaced by $(\mathbf{A}\mathbf{H}, \mathbf{H}^{-1}\mathbf{X}_t)$ for any invertible $\mathbf{H}$.

Therefore, we assume $\mathbf{A}^\tau \mathbf{A} = \mathbf{I}_r$

But factor loading space $\mathcal{M}(\mathbf{A})$ is uniquely defined

The model is not new, but it is effectively new: no model!

What is new?

- No distributional assumption on $\boldsymbol{\varepsilon}_t$. More significantly, allow correlation between $\boldsymbol{\varepsilon}_t$ and $\mathbf{X}_{t+k}$: the ACF of $\mathbf{Y}_t$ may be full-ranked.

$$
\begin{aligned}
\mathrm{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t+k}) &= \mathbf{A}\mathrm{Cov}(\mathbf{X}_t, \mathbf{X}_{t+k})\mathbf{A}^\tau + \mathbf{A}\mathrm{Cov}(\mathbf{X}_t, \boldsymbol{\varepsilon}_{t+k}) \\
&+ \mathrm{Cov}(\boldsymbol{\varepsilon}_t, \mathbf{X}_{t+k})\mathbf{A}^\tau, \qquad k \neq 0.
\end{aligned}
$$

What is new?

- No distributional assumption on $\boldsymbol{\varepsilon}_t$. More significantly, allow correlation between $\boldsymbol{\varepsilon}_t$ and $\mathbf{X}_{t+k}$: the ACF of $\mathbf{Y}_t$ may be full-ranked.

$$\begin{aligned} \mathrm{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t+k}) &= \mathbf{A}\mathrm{Cov}(\mathbf{X}_t, \mathbf{X}_{t+k})\mathbf{A}^\tau + \mathbf{A}\mathrm{Cov}(\mathbf{X}_t, \boldsymbol{\varepsilon}_{t+k}) \\ &+ \mathrm{Cov}(\boldsymbol{\varepsilon}_t, \mathbf{X}_{t+k})\mathbf{A}^\tau, \qquad k \neq 0. \end{aligned}$$

Therefore, if $\mathrm{Cov}(\boldsymbol{\varepsilon}_t, \mathbf{X}_{t+k}) = 0$ for all k, rk$\{\boldsymbol{\Gamma}_y(j)\} \leq r$ for all $j \neq 0$. Then $\mathbf{A}$ and r may be estimated via eigenanalysis (Peña and Box 1987).

What is new?

- No distributional assumption on $\boldsymbol{\varepsilon}_t$. More significantly, allow correlation between $\boldsymbol{\varepsilon}_t$ and $\mathbf{X}_{t+k}$: the ACF of $\mathbf{Y}_t$ may be full-ranked.

$$\begin{aligned}\mathrm{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t+k}) &= \mathbf{A}\mathrm{Cov}(\mathbf{X}_t, \mathbf{X}_{t+k})\mathbf{A}^\tau + \mathbf{A}\mathrm{Cov}(\mathbf{X}_t, \boldsymbol{\varepsilon}_{t+k}) \\ &+ \mathrm{Cov}(\boldsymbol{\varepsilon}_t, \mathbf{X}_{t+k})\mathbf{A}^\tau, \qquad k \neq 0.\end{aligned}$$

  Therefore, if $\mathrm{Cov}(\boldsymbol{\varepsilon}_t, \mathbf{X}_{t+k}) = 0$ for all k, rk$\{\boldsymbol{\Gamma}_y(j)\} \leq r$ for all $j \neq 0$. Then $\mathbf{A}$ and r may be estimated via eigenanalysis (Peña and Box 1987).

- A new estimation method: growing space $\mathcal{M}(\mathbf{A})^\perp$ by one dimension in each step

What is new?

- No distributional assumption on $\boldsymbol{\varepsilon}_t$. More significantly, allow correlation between $\boldsymbol{\varepsilon}_t$ and $\mathbf{X}_{t+k}$: the ACF of $\mathbf{Y}_t$ may be full-ranked.

$$\begin{aligned}\text{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t+k}) &= \mathbf{A}\text{Cov}(\mathbf{X}_t, \mathbf{X}_{t+k})\mathbf{A}^\tau + \mathbf{A}\text{Cov}(\mathbf{X}_t, \boldsymbol{\varepsilon}_{t+k}) \\ &+ \text{Cov}(\boldsymbol{\varepsilon}_t, \mathbf{X}_{t+k})\mathbf{A}^\tau, \qquad k \neq 0.\end{aligned}$$

  Therefore, if $\text{Cov}(\boldsymbol{\varepsilon}_t, \mathbf{X}_{t+k}) = 0$ for all k, rk$\{\boldsymbol{\Gamma}_y(j)\} \leq r$ for all $j \neq 0$. Then $\mathbf{A}$ and r may be estimated via eigenanalysis (Peña and Box 1987).

- A new estimation method: growing space $\mathcal{M}(\mathbf{A})^\perp$ by one dimension in each step

- Factor $\mathbf{X}_t$, and therefore also $\mathbf{Y}_t$, may be nonstationary, not necessarily driven by unit roots.

Let $\mathbf{B} = (\mathbf{b}_1, \cdots, \mathbf{b}_{d-r})$ be a $d \times (d-r)$ matrix such that

$(\mathbf{A}, \mathbf{B})$ is a $d \times d$ orthogonal matrix, i.e.

$$\mathbf{B}^\tau \mathbf{A} = 0, \qquad \mathbf{B}^\tau \mathbf{B} = \mathbf{I}_{d-r}.$$

Let $\mathbf{B} = (\mathbf{b}_1, \cdots, \mathbf{b}_{d-r})$ be a $d \times (d-r)$ matrix such that

$(\mathbf{A}, \mathbf{B})$ is a $d \times d$ orthogonal matrix, i.e.

$$\mathbf{B}^\tau \mathbf{A} = 0, \qquad \mathbf{B}^\tau \mathbf{B} = \mathbf{I}_{d-r}.$$

Since $\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t + \boldsymbol{\varepsilon}_t$,

$$\mathbf{B}^\tau \mathbf{Y}_t = \mathbf{B}^\tau \boldsymbol{\varepsilon}_t$$

i.e. $\{\mathbf{B}^\tau \mathbf{Y}_t,\ t = 0, \pm 1, \cdots\}$ is WN.

Let $\mathbf{B} = (\mathbf{b}_1, \cdots, \mathbf{b}_{d-r})$ be a $d \times (d-r)$ matrix such that

$(\mathbf{A}, \mathbf{B})$ is a $d \times d$ orthogonal matrix, i.e.

$$\mathbf{B}^\tau \mathbf{A} = 0, \qquad \mathbf{B}^\tau \mathbf{B} = \mathbf{I}_{d-r}.$$

Since $\mathbf{Y}_t = \mathbf{A}\mathbf{X}_t + \boldsymbol{\varepsilon}_t$,

$$\mathbf{B}^\tau \mathbf{Y}_t = \mathbf{B}^\tau \boldsymbol{\varepsilon}_t$$

i.e. $\{\mathbf{B}^\tau \mathbf{Y}_t,\ t = 0, \pm 1, \cdots\}$ is WN.

Therefore

$$\mathrm{Corr}(\mathbf{b}_i^\tau \mathbf{Y}_t, \mathbf{b}_j^\tau \mathbf{Y}_{t-k}) = 0 \quad \forall\, 1 \le i, j \le d-r \text{ and } 1 \le k \le p,$$

where $p \ge 1$ is an arbitrary integer.

Assuming $\mathbf{S}_0 \equiv \frac{1}{n}\sum_{t=1}^{n}(\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_t - \bar{\mathbf{Y}})^\tau = \mathbf{I}_d$

Assuming $\mathbf{S}_0 \equiv \frac{1}{n}\sum_{t=1}^{n}(\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_t - \bar{\mathbf{Y}})^\tau = \mathbf{I}_d$

An estimator for $\mathbf{B} = (\mathbf{b}_1, \cdots, \mathbf{b}_{d-r})$ is obtained by minimising

$$\Psi_n(\mathbf{B}) \equiv \sum_{k=1}^{p} ||\mathbf{B}^\tau \mathbf{S}_k \mathbf{B}||^2 = \sum_{k=1}^{p} \sum_{1\le i,j\le d-r} (\mathbf{b}_i^\tau \mathbf{S}_k \mathbf{b}_j)^2,$$

where $||\mathbf{H}|| = \{\text{tr}(\mathbf{H}^\tau \mathbf{H})\}^{1/2}$, and

$$\mathbf{S}_k = \frac{1}{n}\sum_{t=k+1}^{n}(\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_{t-k} - \bar{\mathbf{Y}})^\tau.$$

Assuming $\mathbf{S}_0 \equiv \frac{1}{n}\sum_{t=1}^{n}(\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_t - \bar{\mathbf{Y}})^\tau = \mathbf{I}_d$

An estimator for $\mathbf{B} = (\mathbf{b}_1, \cdots, \mathbf{b}_{d-r})$ is obtained by minimising

$$\Psi_n(\mathbf{B}) \equiv \sum_{k=1}^{p} ||\mathbf{B}^\tau \mathbf{S}_k \mathbf{B}||^2 = \sum_{k=1}^{p} \sum_{1\le i,j\le d-r} (\mathbf{b}_i^\tau \mathbf{S}_k \mathbf{b}_j)^2,$$

where $||\mathbf{H}|| = \{\text{tr}(\mathbf{H}^\tau \mathbf{H})\}^{1/2}$, and

$$\mathbf{S}_k = \frac{1}{n}\sum_{t=k+1}^{n}(\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_{t-k} - \bar{\mathbf{Y}})^\tau.$$

Remark. Without the above assumption, $\Psi_n(\mathbf{B})$ would be defined as $\sum_{k=1}^{p}\sum_{1\le i,j\le d-r}(\mathbf{b}_i^\tau \mathbf{S}_k \mathbf{b}_j)^2/\{\mathbf{b}_i^\tau \mathbf{S}_0 \mathbf{b}_i \, \mathbf{b}_j^\tau \mathbf{S}_0 \mathbf{b}_j\}$.

Tow problems:

- the number of factors r unknown
- nonlinear optimisation with $d(d-r)$ variables

Tow problems:

- the number of factors r unknown
- nonlinear optimisation with $d(d-r)$ variables

Algorithm:

reduce the $d(d-r)$-dim optimisation problem to several d- or lower-dimensional subproblems while determining r by the portmanteau tests for WN.

Put

$$\psi(\mathbf{b}) = \sum_{k=1}^{p} (\mathbf{b}^{\tau}\mathbf{S}_k\mathbf{b})^2, \quad \psi_m(\mathbf{b}) = \sum_{k=1}^{p} \sum_{i=1}^{m-1} \{(\mathbf{b}^{\tau}\mathbf{S}_k\widehat{\mathbf{b}}_i)^2 + (\widehat{\mathbf{b}}_i^{\tau}\mathbf{S}_k\mathbf{b})^2\}.$$

Step1. Let $\widehat{\mathbf{b}}_1 = \arg\min_{||\mathbf{b}||=1} \psi(\mathbf{b})$. Terminate with $\widehat{r} = d$, $\widehat{\mathbf{B}} = 0$ if

$$L_{p,1} \equiv n(n+2) \sum_{k=1}^{p} (\widehat{\mathbf{b}}_1^\tau \mathbf{S}_k \widehat{\mathbf{b}}_1)^2 / (n-k) > \chi^2_{p,\alpha}.$$

Otherwise proceed to Step 2.

Step1. Let $\widehat{\mathbf{b}}_1 = \arg\min_{||\mathbf{b}||=1} \psi(\mathbf{b})$. Terminate with $\widehat{r} = d$, $\widehat{\mathbf{B}} = 0$ if

$$L_{p,1} \equiv n(n+2) \sum_{k=1}^{p} (\widehat{\mathbf{b}}_1^{\tau} \mathbf{S}_k \widehat{\mathbf{b}}_1)^2 / (n-k) > \chi^2_{p,\alpha}.$$

Otherwise proceed to Step 2.

Step2. For $m = 2, \cdots, d$, let $\widehat{\mathbf{b}}_m = \arg\min\{\psi(\mathbf{b}) + \psi_m(\mathbf{b})\}$ subject to

$$||\mathbf{b}|| = 1, \quad \mathbf{b}^{\tau}\widehat{\mathbf{b}}_i = 0 \quad \text{for } i = 1, \cdots, m-1.$$

Terminate with $\widehat{r} = d - m + 1$ and $\widehat{\mathbf{B}} = (\widehat{\mathbf{b}}_1, \cdots, \widehat{\mathbf{b}}_{m-1})$ if

$$L_{p,m} \equiv n^2 \sum_{k=1}^{p} \frac{1}{n-k} \left[(\widehat{\mathbf{b}}_m^{\tau} \mathbf{S}_k \widehat{\mathbf{b}}_m)^2 + \sum_{j=1}^{m-1} \{ (\widehat{\mathbf{b}}_m^{\tau} \mathbf{S}_k \widehat{\mathbf{b}}_j)^2 + (\widehat{\mathbf{b}}_j^{\tau} \mathbf{S}_k \widehat{\mathbf{b}}_m)^2 \} \right]$$

is greater than $\chi^2_{p(2m-1),\alpha}$. $L^*_{p,m}$

Remarks

1. In the event that $L_{p,m} \leq \chi^2_{p(2m-1),\alpha}$ for all $1 \leq m \leq d$, define $\widehat{r} = 0$ and $\widehat{\mathbf{B}} = \mathbf{I}_d$.

Remarks

1. In the event that $L_{p,m} \leq \chi^2_{p(2m-1),\alpha}$ for all $1 \leq m \leq d$, define $\widehat{r} = 0$ and $\widehat{\mathbf{B}} = \mathbf{I}_d$.

2. The algorithm grows the dimension of $\mathcal{M}(\mathbf{B})$ by 1 each time until a newly selected direction $\widehat{\mathbf{b}}_m$ does not lead to a WN.

Remarks

1. In the event that $L_{p,m} \leq \chi^2_{p(2m-1),\alpha}$ for all $1 \leq m \leq d$, define $\widehat{r} = 0$ and $\widehat{\mathbf{B}} = \mathbf{I}_d$.

2. The algorithm grows the dimension of $\mathcal{M}(\mathbf{B})$ by 1 each time until a newly selected direction $\widehat{\mathbf{b}}_m$ does not lead to a WN.

3. Since $\widehat{\mathbf{B}}^\tau \widehat{\mathbf{B}} = \mathbf{I}_{d-\widehat{r}}$, we may let $\widehat{\mathbf{A}} = (\widehat{\mathbf{a}}_1, \cdots, \widehat{\mathbf{a}}_{\widehat{r}})$, where $\widehat{\mathbf{A}}^\tau \widehat{\mathbf{A}} = \mathbf{I}_{\widehat{r}}$, and

$$(\mathbf{I}_d - \widehat{\mathbf{B}}\widehat{\mathbf{B}}^\tau)\widehat{\mathbf{a}}_i = \widehat{\mathbf{a}}_i, \quad 1 \leq i \leq \widehat{r}.$$

4. Step 2 may be reduced to an unconstrained optimisation problem with $(d - m)$ free variables.

4. Step 2 may be reduced to an unconstrained optimisation problem with $(d-m)$ free variables.

If $||\mathbf{b}|| = 1$ and $\mathbf{b}^\tau \mathbf{B}_{m-1} \equiv \mathbf{b}^\tau(\widehat{\mathbf{b}}_1, \cdots, \widehat{\mathbf{b}}_{m-1}) = 0$, then

$$\mathbf{b} = \mathbf{D}_m \mathbf{u} \equiv (\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_{d-m+1})\mathbf{u},$$

where $||\mathbf{u}|| = 1$, $\mathbf{D}_m^\tau \mathbf{D}_m = \mathbf{I}_{d-m+1}$ and

$$(\mathbf{I}_d - \mathbf{B}_{m-1}\mathbf{B}_{m-1}^\tau)\boldsymbol{\gamma}_j = \boldsymbol{\gamma}_j, \quad 1 \le j \le d-m+1.$$

4. Step 2 may be reduced to an unconstrained optimisation problem with $(d-m)$ free variables.

If $||\mathbf{b}|| = 1$ and $\mathbf{b}^\tau \mathbf{B}_{m-1} \equiv \mathbf{b}^\tau(\widehat{\mathbf{b}}_1, \cdots, \widehat{\mathbf{b}}_{m-1}) = 0$, then

$$\mathbf{b} = \mathbf{D}_m \mathbf{u} \equiv (\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_{d-m+1})\mathbf{u},$$

where $||\mathbf{u}|| = 1$, $\mathbf{D}_m^\tau \mathbf{D}_m = \mathbf{I}_{d-m+1}$ and

$$(\mathbf{I}_d - \mathbf{B}_{m-1}\mathbf{B}_{m-1}^\tau)\boldsymbol{\gamma}_j = \boldsymbol{\gamma}_j, \quad 1 \le j \le d-m+1.$$

Unit vector $\mathbf{u}^\tau = (u_1, \cdots, u_k)$ may be expressed as

$$u_1 = \prod_{j=1}^{k-1} \cos\theta_j, \quad u_i = \sin\theta_{i-1} \prod_{j=i}^{k-1} \cos\theta_j, \; i = 2, \cdots, k-1,$$

and $u_k = \sin\theta_{k-1}$, depending on $\theta_1, \cdots, \theta_{k-1}$ only.

5. The univariate portmanteau test statistic $L_{p,1}$ has a non-standard normalised constant $n(n+2)$ to improve the finite sample performance (Ljung and Box 1978).

Li and McLeod (1981) proposed a multivariate version:

$$L^*_{p,m} = L_{p,m} + \frac{p(p+1)(2m-1)}{2n}.$$

5. The univariate portmanteau test statistic $L_{p,1}$ has a non-standard normalised constant $n(n+2)$ to improve the finite sample performance (Ljung and Box 1978).

Li and McLeod (1981) proposed a multivariate version:

$$L^*_{p,m} = L_{p,m} + \frac{p(p+1)(2m-1)}{2n}.$$

A univariate version:

$$L'_{p,m} = n(n+2) \sum_{k=1}^{p} \frac{(\widehat{\mathbf{b}}_m^\tau \mathbf{S}_k \widehat{\mathbf{b}}_m)^2}{n-k}.$$

5. The univariate portmanteau test statistic $L_{p,1}$ has a non-standard normalised constant $n(n+2)$ to improve the finite sample performance (Ljung and Box 1978).

Li and McLeod (1981) proposed a multivariate version:

$$L^*_{p,m} = L_{p,m} + \frac{p(p+1)(2m-1)}{2n}.$$

A univariate version:

$$L'_{p,m} = n(n+2)\sum_{k=1}^{p}\frac{(\widehat{\mathbf{b}}_m^\tau \mathbf{S}_k \widehat{\mathbf{b}}_m)^2}{n-k}.$$

6. Skip all the tests if r is known.

Modelling with estimated factors

Note $\widehat{\mathbf{A}}\widehat{\mathbf{A}}^\tau + \widehat{\mathbf{B}}\widehat{\mathbf{B}}^\tau = \mathbf{I}_d$. We may write

$$\mathbf{Y}_t = \widehat{\mathbf{A}}\boldsymbol{\xi}_t + \mathbf{e}_t,$$

$$\boldsymbol{\xi}_t = \widehat{\mathbf{A}}^\tau \mathbf{Y}_t = \widehat{\mathbf{A}}^\tau \mathbf{A}\mathbf{X}_t + \widehat{\mathbf{A}}^\tau \boldsymbol{\varepsilon}_t, \quad \mathbf{e}_t = \widehat{\mathbf{B}}\widehat{\mathbf{B}}^\tau \mathbf{Y}_t.$$

- $\mathbf{e}_t$ is treated as WN
- Model factors $\boldsymbol{\xi}_t$ by VARMA or other vector TS models.

Modelling with estimated factors

Note $\widehat{\mathbf{A}}\widehat{\mathbf{A}}^\tau + \widehat{\mathbf{B}}\widehat{\mathbf{B}}^\tau = \mathbf{I}_d$. We may write

$$\mathbf{Y}_t = \widehat{\mathbf{A}}\boldsymbol{\xi}_t + \mathbf{e}_t,$$

$$\boldsymbol{\xi}_t = \widehat{\mathbf{A}}^\tau \mathbf{Y}_t = \widehat{\mathbf{A}}^\tau \mathbf{A}\mathbf{X}_t + \widehat{\mathbf{A}}^\tau \boldsymbol{\varepsilon}_t, \quad \mathbf{e}_t = \widehat{\mathbf{B}}\widehat{\mathbf{B}}^\tau \mathbf{Y}_t.$$

- $\mathbf{e}_t$ is treated as WN
- Model factors $\boldsymbol{\xi}_t$ by VARMA or other vector TS models.

Replace $\widehat{\mathbf{A}}$ by $\widehat{\mathbf{A}}\mathbf{H}$ with appropriate orthogonal $\mathbf{H}$ such that $\boldsymbol{\xi}_t$ admits a simple model (Tiao and Tsay 1989), or replace $\boldsymbol{\xi}_t$ by their principal components. Note

$$\mathcal{M}(\widehat{\mathbf{A}}\mathbf{H}) = \mathcal{M}(\widehat{\mathbf{A}}).$$

Simulation

Basic model: $Y_{ti} = \begin{cases} X_{ti} + \varepsilon_{ti}, & 1 \le i \le 3 \\ \varepsilon_{ti}, & 4 \le i \le d. \end{cases}$

Case I: $\begin{cases} X_{t1} = 0.8X_{t-1,1} + e_{t1}, \\ X_{t2} = e_{t2} + 0.9e_{t-1,2} + 0.3e_{t-2,2}, \\ X_{t3} = -0.5X_{t-1,3} - \varepsilon_{t3} + 0.8\varepsilon_{t-1,3}. \end{cases}$

Case II: $\begin{cases} X_{t1} - 2t/n = 0.8(X_{t-1,1} - 2t/n) + e_{t1}, \\ X_{t2} = 3t/n, \\ X_{t3} = X_{t-1,3} + \sqrt{\frac{10}{n}}e_{t3}, \quad (\text{with } X_{0,3} \sim N(0,1)). \end{cases}$

All ε_{ti}, e_{ti} are independent $N(0,1)$.

True values: $r = 3$ and $\mathbf{A}^\tau = (\mathbf{I}_3, \mathbf{0})$

- set $n = 300, 600, 1000$ and $d = 5, 10, 20$
- in portmanteau tests: $\alpha = 5\%$ and $p = 15$
- simulation replication: 1000 times (for each settings)

Measure the estimation error for factor loading space:

$$D_1(\mathbf{A}, \widehat{\mathbf{A}}) = \Big([\text{tr}\{\widehat{\mathbf{A}}^\tau (I_d - \mathbf{A}\mathbf{A}^\tau)\widehat{\mathbf{A}}\} + \text{tr}(\widehat{\mathbf{B}}^\tau \mathbf{A}\mathbf{A}^\tau \widehat{\mathbf{B}})]/d\Big)^{1/2}.$$

Then

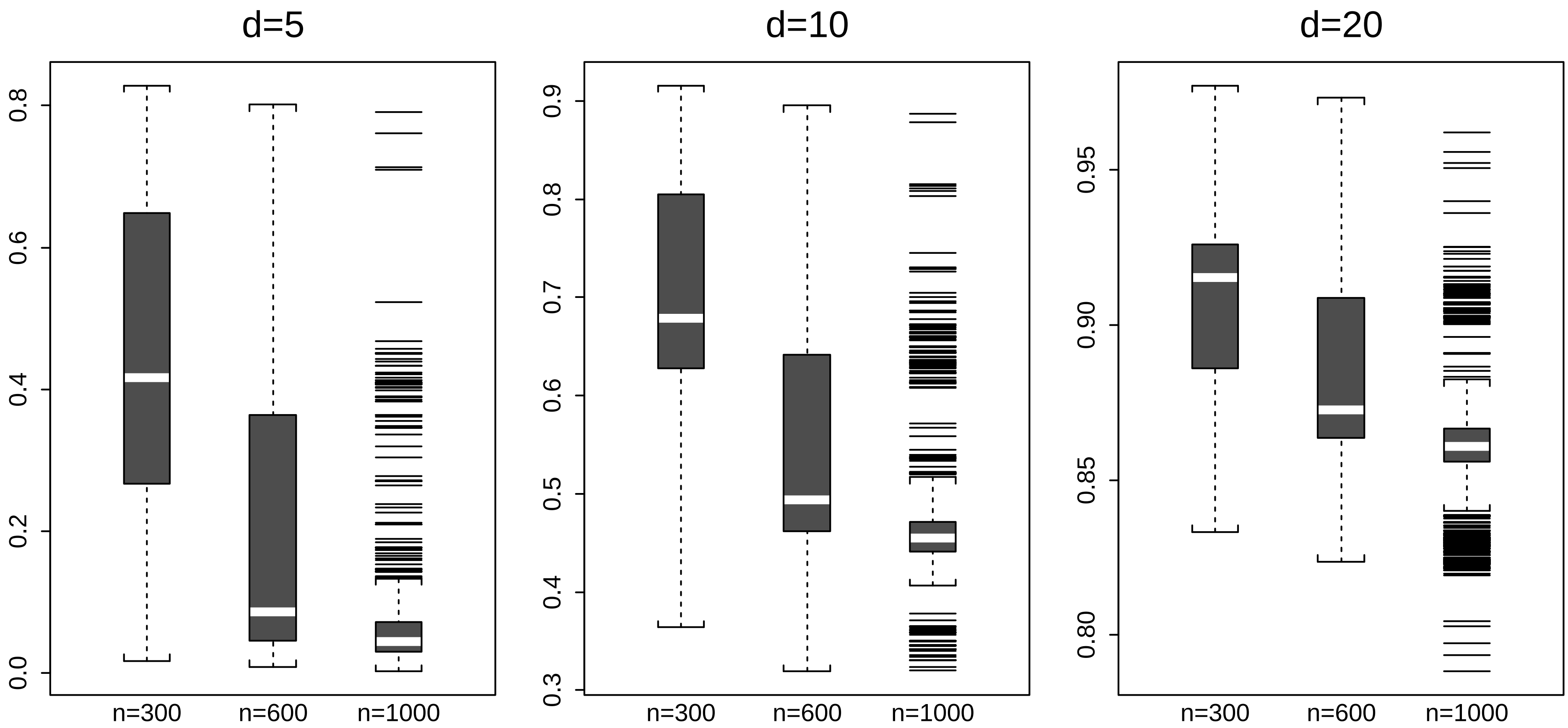
$$D_1(\mathbf{A}, \widehat{\mathbf{A}}) \in [0, 1]$$

$D_1(\mathbf{A}, \widehat{\mathbf{A}}) = 0$ iff $\mathcal{M}(\mathbf{A}) = \mathcal{M}(\widehat{\mathbf{A}})$

$D_1(\mathbf{A}, \widehat{\mathbf{A}}) = 1$ iff $\mathcal{M}(\mathbf{A}) = \mathcal{M}(\widehat{\mathbf{B}})$.

Case I: Relative frequency estimates of r

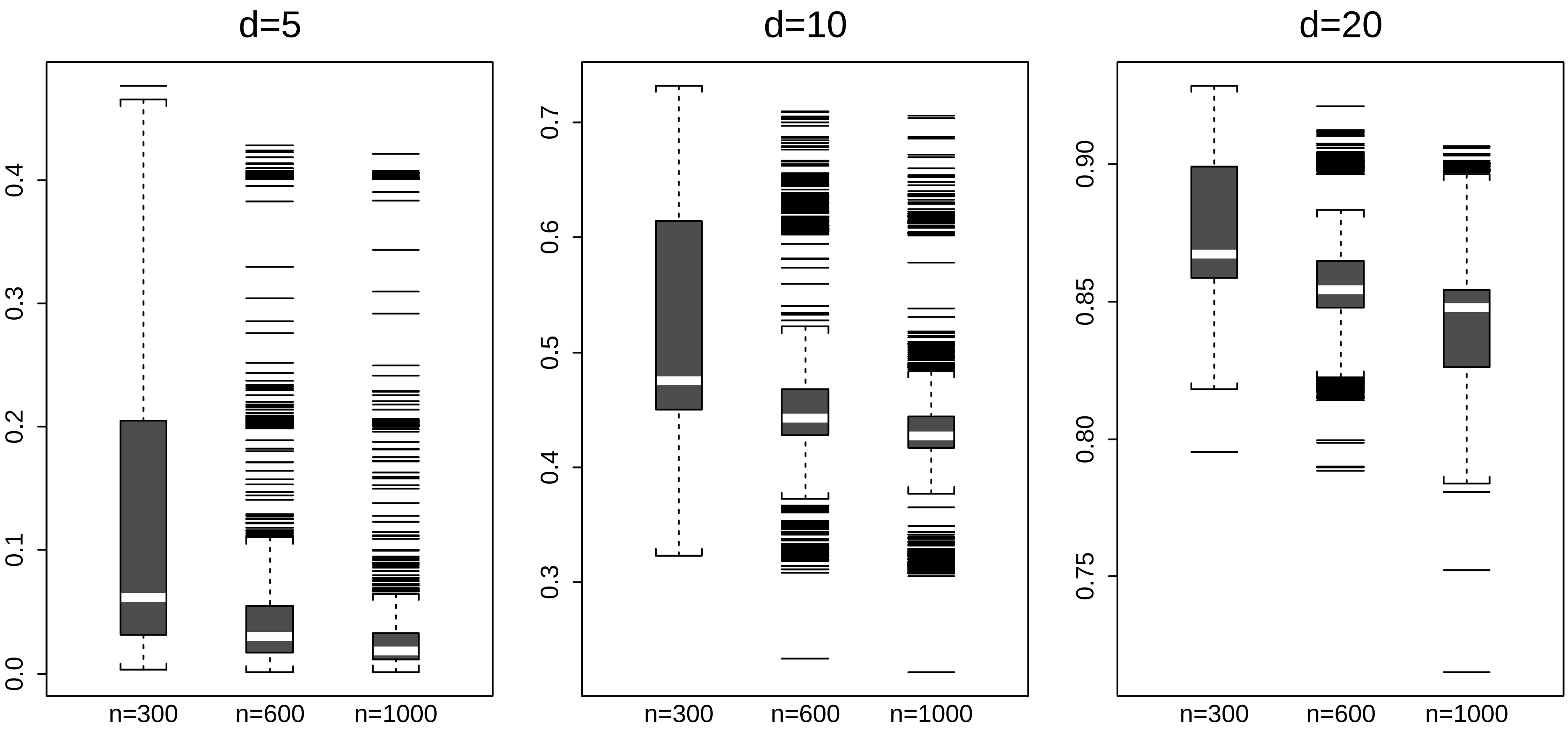
		$\widehat{r}$						
d	n	0	1	2	3	4	5	≥ 6
5	300	.000	.209	.444	.345	.002	.000	
	600	.000	.071	.286	.633	.010	.000	
	1000	.000	.004	.051	.933	.120	.000	
10	300	.000	.219	.524	.255	.002	.000	.000
	600	.000	.049	.290	.649	.012	.000	.000
	1000	.000	.007	.062	.898	.033	.000	.000
20	300	.000	.162	.543	.285	.010	.000	.000
	600	.000	.033	.305	.609	.053	.000	.000
	1000	.000	.004	.066	.822	.103	.005	.000



Case I: Boxplots of $D_1(\mathbf{A}, \widehat{\mathbf{A}})$

Case II: Relative frequency estimates of r

		$\widehat{r}$						
d	n	0	1	2	3	4	5	≥ 6
5	300	.000	.000	.255	.743	.002	.000	
	600	.000	.000	.083	.907	.010	.000	
	1000	.000	.000	.033	.945	.022	.000	
10	300	.000	.000	.283	.695	.022	.000	.000
	600	.000	.000	.103	.842	.054	.001	.000
	1000	.000	.000	.051	.871	.077	.001	.000
20	300	.000	.000	.258	.663	.076	.001	.002
	600	.000	.000	.035	.673	.278	.012	.002
	1000	.000	.000	.099	.733	.162	.006	.000



Case II: Boxplots of $D_1(\mathbf{A}, \widehat{\mathbf{A}})$

Estimation for non-stationary Case II is more accurate than that for stationary Case I, especially when $n = 300$ and 600.

Key: The quadratic forms of the sample covariance matrices

$$\mathbf{S}_k = \frac{1}{n} \sum_{t=k+1}^{n} (\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_{t-k} - \bar{\mathbf{Y}})^\tau, \quad k = 1, \cdots, p$$

are *significantly* non-zero in the directions in the factor loading space $\mathcal{M}(\mathbf{A})$.

Theoretical Properties

First, let r be known.

Recall

$$\widehat{\mathbf{B}} = \arg\min_{\mathbf{B}\in\mathcal{H}} \Psi_n(\mathbf{B}),$$

$$\mathcal{H} = \{\text{all } d \times r \text{ half orthogonal matrices}\},$$

$$\Psi_n(\mathbf{B}) = \sum_{k=1}^{p} ||\mathbf{B}^\tau \mathbf{S}_k \mathbf{B}||^2, \quad \Psi(\mathbf{B}) = \sum_{k=1}^{p} ||\mathbf{B}^\tau \mathbf{\Sigma}_k \mathbf{B}||^2.$$

Theoretical Properties

First, let r be known.

Recall

$$\widehat{\mathbf{B}} = \arg\min_{\mathbf{B}\in\mathcal{H}} \Psi_n(\mathbf{B}),$$

$$\mathcal{H} = \{\text{all } d \times r \text{ half orthogonal matrices}\},$$

$$\Psi_n(\mathbf{B}) = \sum_{k=1}^{p} ||\mathbf{B}^\tau \mathbf{S}_k \mathbf{B}||^2, \quad \Psi(\mathbf{B}) = \sum_{k=1}^{p} ||\mathbf{B}^\tau \mathbf{\Sigma}_k \mathbf{B}||^2.$$

C1. As $n \to \infty$, $\mathbf{S}_k \xrightarrow{P} \mathbf{\Sigma}_k$ for $k = 0, 1, \cdots, p$, and $\mathbf{\Sigma}_0 = \mathbf{I}_d$.

Remark. C1 is implied by ρ-mixing and $E\mathbf{S}_k \to \mathbf{\Sigma}_k$, and is also fulfilled by some deterministic processes. theorems

Lemma. Let $\{\mathbf{Y}_t\}$ be φ-mixing, and $E\mathbf{S}_k \to \mathbf{\Sigma}_k$. Suppose

$$\mathbf{Y}_t = \mathbf{U}_t + \mathbf{V}_t, \quad \mathrm{Cov}(\mathbf{U}_t, \mathbf{V}_t) = 0, \quad \sup_t E||U_t||^h < \infty \ (h > 2),$$

$$\frac{1}{n}\sum_{t=1}^{n} \mathbf{V}_t \xrightarrow{P} \mathbf{c}, \qquad \frac{1}{n}\sum_{t=1}^{n} E\mathbf{V}_t \to \mathbf{c}.$$

Then

(i) $\mathbf{S}_k \xrightarrow{P} \mathbf{\Sigma}_k$, and
(ii) $\mathbf{S}_k \xrightarrow{a.s.} \mathbf{\Sigma}_k$ provided $\frac{1}{n}\sum_{t=1}^{n} \mathbf{V}_t \xrightarrow{a.s.} \mathbf{c}$, and

$$\varphi(m) = \begin{cases} O(m^{-\frac{b}{2b-2}-\delta}), & \text{if } 1 < b < 2, \\ O(m^{-\frac{2}{b}-\delta}), & \text{if } b \geq 2, \end{cases}$$

where $\delta > 0$ is a constant.

For $\mathbf{H}_1, \mathbf{H}_2 \in \mathcal{H}$, define

$$D(\mathbf{H}_1, \mathbf{H}_2) = ||(\mathbf{I}_d - \mathbf{H}_1\mathbf{H}_1^\tau)\mathbf{H}_2|| = \sqrt{r - \text{tr}(\mathbf{H}_1\mathbf{H}_1^\tau\mathbf{H}_2\mathbf{H}_2^\tau)}.$$

Then $D(\mathbf{H}_1, \mathbf{H}_2) = 0$ iff $\mathcal{M}(\mathbf{H}_1) = \mathcal{M}(\mathbf{H}_2)$.

For $\mathbf{H}_1, \mathbf{H}_2 \in \mathcal{H}$, define

$$D(\mathbf{H}_1, \mathbf{H}_2) = ||(\mathbf{I}_d - \mathbf{H}_1 \mathbf{H}_1^\tau) \mathbf{H}_2|| = \sqrt{r - \text{tr}(\mathbf{H}_1 \mathbf{H}_1^\tau \mathbf{H}_2 \mathbf{H}_2^\tau)}.$$

Then $D(\mathbf{H}_1, \mathbf{H}_2) = 0$ iff $\mathcal{M}(\mathbf{H}_1) = \mathcal{M}(\mathbf{H}_2)$.

In fact, $D(\mathbf{H}_1, \mathbf{H}_2)$ is a well-defined distance on the quotient space $\mathcal{H}_D \equiv \mathcal{H}/D$, i.e.

$$D(\mathbf{H}_1, \mathbf{H}_2) \le D(\mathbf{H}_1, \mathbf{H}_3) + D(\mathbf{H}_3, \mathbf{H}_2).$$

For $\mathbf{H}_1, \mathbf{H}_2 \in \mathcal{H}$, define

$$D(\mathbf{H}_1, \mathbf{H}_2) = ||(\mathbf{I}_d - \mathbf{H}_1\mathbf{H}_1^\tau)\mathbf{H}_2|| = \sqrt{r - \text{tr}(\mathbf{H}_1\mathbf{H}_1^\tau\mathbf{H}_2\mathbf{H}_2^\tau)}.$$

Then $D(\mathbf{H}_1, \mathbf{H}_2) = 0$ iff $\mathcal{M}(\mathbf{H}_1) = \mathcal{M}(\mathbf{H}_2)$.

In fact, $D(\mathbf{H}_1, \mathbf{H}_2)$ is a well-defined distance on the quotient space $\mathcal{H}_D \equiv \mathcal{H}/D$, i.e.

$$D(\mathbf{H}_1, \mathbf{H}_2) \leq D(\mathbf{H}_1, \mathbf{H}_3) + D(\mathbf{H}_3, \mathbf{H}_2).$$

In fact both $\Psi_n(\cdot)$ and $\Psi(\cdot)$ are well-defined functions on the metric space $(\mathcal{H}_D, D)$.

For $\mathbf{H}_1, \mathbf{H}_2 \in \mathcal{H}$, define

$$D(\mathbf{H}_1, \mathbf{H}_2) = ||(\mathbf{I}_d - \mathbf{H}_1 \mathbf{H}_1^\tau)\mathbf{H}_2|| = \sqrt{r - \text{tr}(\mathbf{H}_1 \mathbf{H}_1^\tau \mathbf{H}_2 \mathbf{H}_2^\tau)}.$$

Then $D(\mathbf{H}_1, \mathbf{H}_2) = 0$ iff $\mathcal{M}(\mathbf{H}_1) = \mathcal{M}(\mathbf{H}_2)$.

In fact, $D(\mathbf{H}_1, \mathbf{H}_2)$ is a well-defined distance on the quotient space $\mathcal{H}_D \equiv \mathcal{H}/D$, i.e.

$$D(\mathbf{H}_1, \mathbf{H}_2) \le D(\mathbf{H}_1, \mathbf{H}_3) + D(\mathbf{H}_3, \mathbf{H}_2).$$

In fact both $\Psi_n(\cdot)$ and $\Psi(\cdot)$ are well-defined functions on the metric space $(\mathcal{H}_D, D)$.

C2. There exists a $\text{B}_0 \in \mathcal{H}_D$ which is the unique minimiser of $\Psi(\cdot)$.

theorems

Theorem 1. Under conditions C1 and C2, $D(\widehat{\mathbf{B}}, \mathbf{B}_0) \xrightarrow{P} 0$.

Theorem 1. Under conditions C1 and C2, $D(\widehat{\mathbf{B}}, \mathbf{B}_0) \xrightarrow{P} 0$.

Theorem 2. Under conditions C1 – C3,

$$\sup_{\mathbf{B}\in\mathcal{H}_D} |\Psi_n(\mathbf{B}) - \Psi(\mathbf{B})| = O_P(\frac{1}{\sqrt{n}}), \quad D(\widehat{\mathbf{B}}, \mathbf{B}_0) = O_P(n^{-\frac{1}{2c}}).$$

C3. It holds for any $\mathbf{B} \in \mathcal{H}_D$ that

$$\Psi(\mathbf{B}) - \Psi(\mathbf{B}_0) \geq a[D(\mathbf{B}, \mathbf{B}_0)]^c,$$

where $a, c > 0$ are some constants. Furthermore,

$$\sqrt{n}(E\mathbf{S}_k - \mathbf{\Sigma}_k) = O(1).$$

Theorem 1. Under conditions C1 and C2, $D(\widehat{\mathbf{B}}, \mathbf{B}_0) \xrightarrow{P} 0$.

Theorem 2. Under conditions C1 – C3,

$$\sup_{\mathbf{B} \in \mathcal{H}_D} |\Psi_n(\mathbf{B}) - \Psi(\mathbf{B})| = O_P(\frac{1}{\sqrt{n}}), \quad D(\widehat{\mathbf{B}}, \mathbf{B}_0) = O_P(n^{-\frac{1}{2c}}).$$

C3. It holds for any $\mathbf{B} \in \mathcal{H}_D$ that

$$\Psi(\mathbf{B}) - \Psi(\mathbf{B}_0) \geq a[D(\mathbf{B}, \mathbf{B}_0)]^c,$$

where $a, c > 0$ are some constants. Furthermore,

$$\sqrt{n}(E\mathbf{S}_k - \mathbf{\Sigma}_k) = O(1).$$

When r unknown?

Illustration With Real Data

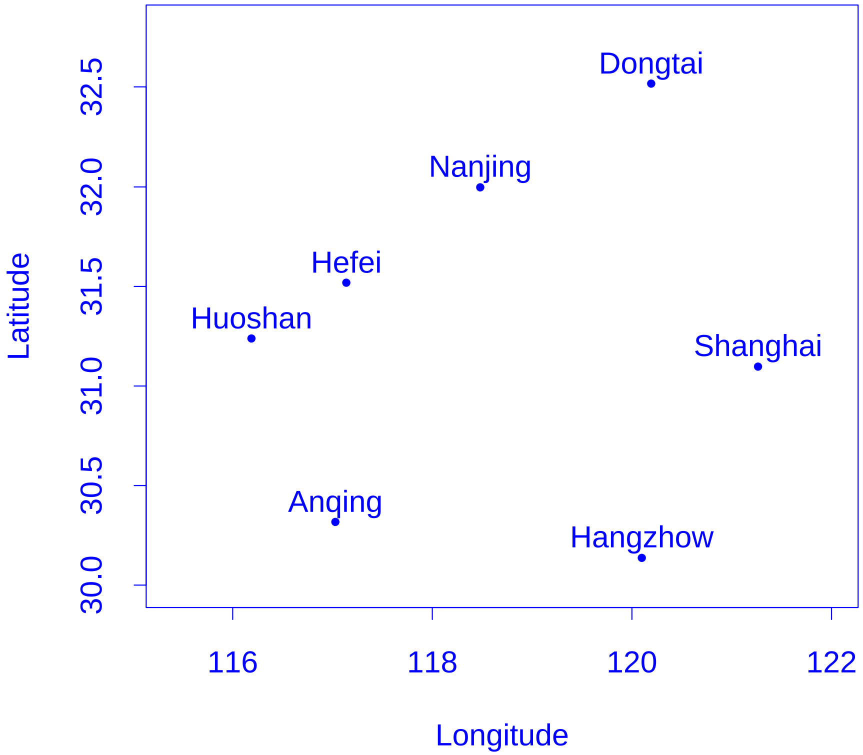
- *Easy example*: monthly temperature data from 7 cities in Eastern China in Jaunary 1954 — December 1986

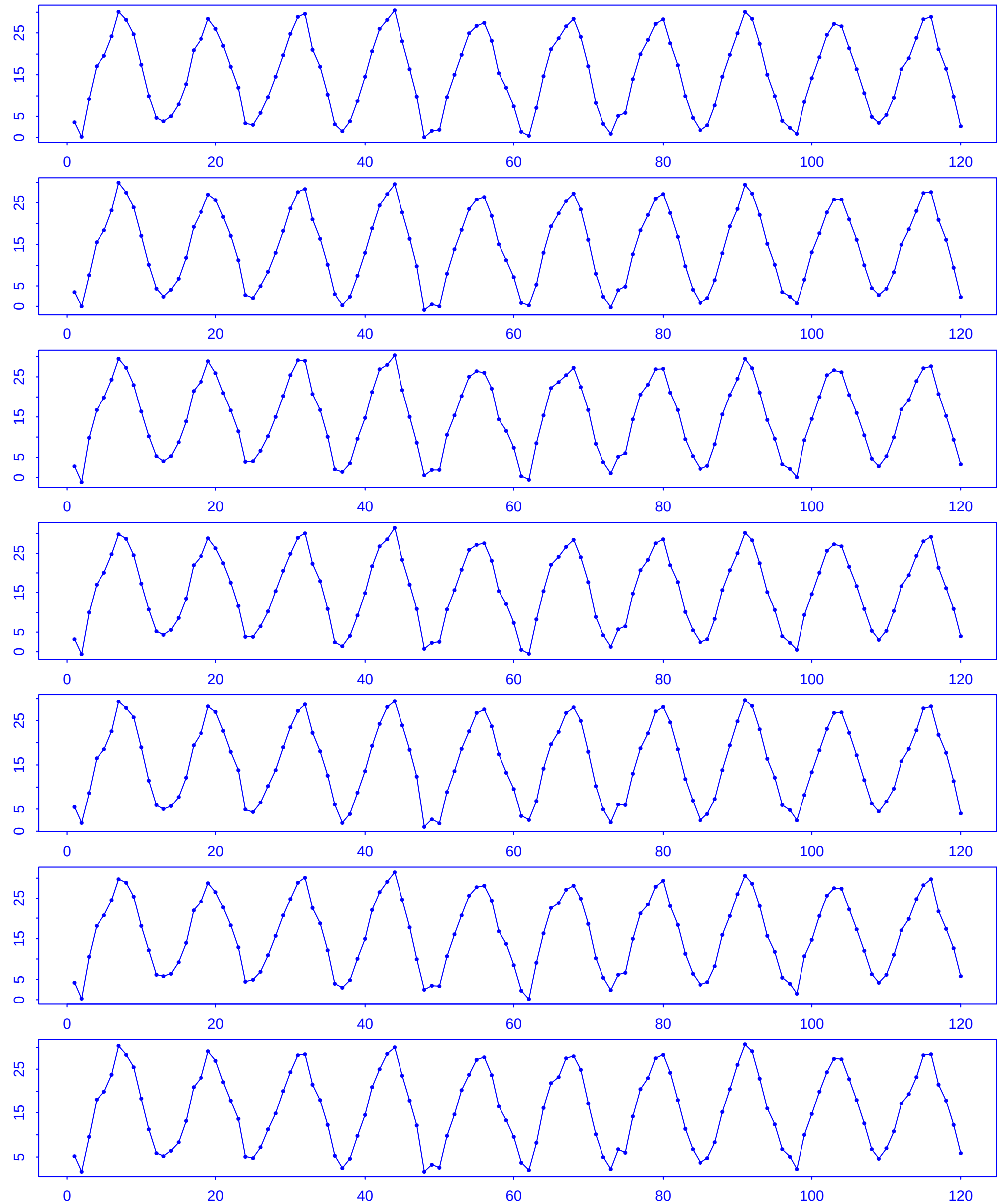
  $n = 396, \quad d = 7$

- *Less easy example*: weekly yields of the 3-month, 6-month and 12-month USA Treasury bills in 17 July 1959 – 12 August 1972

  $n = 700, \quad d = 3$

Time plots of the monthly temperature in 1959-1968 of Nanjing, Dongtai, Huoshan, Hefei, Shanghai, Anqing and Hangzhow.





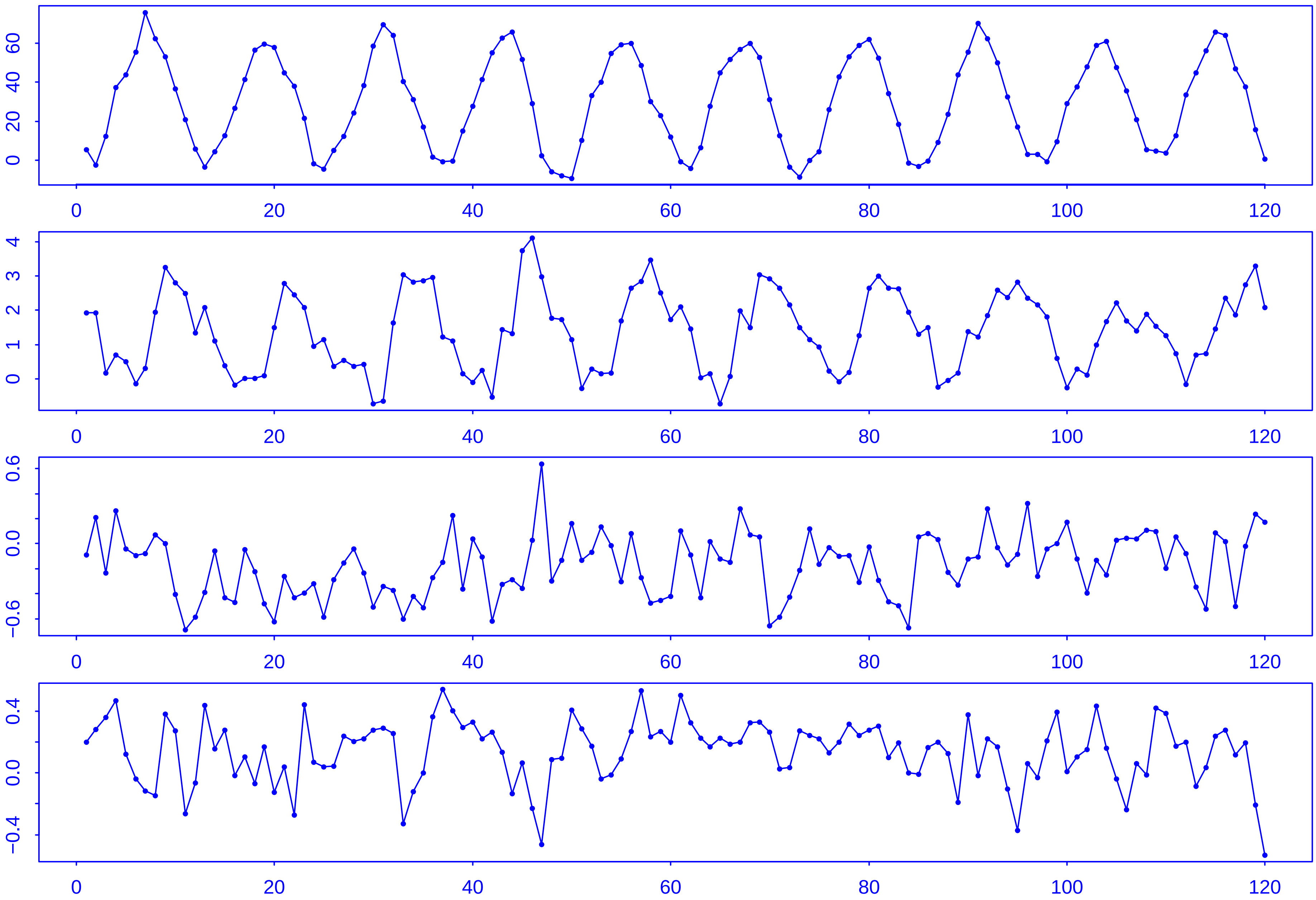
With $p = 12$, $\alpha = 1\%$, the fitted model is $\mathbf{Y}_t = \widehat{\mathbf{A}}\boldsymbol{\xi}_t + \mathbf{e}_t$, $\widehat{r} = 4$, $\mathbf{e}_t \sim \mathrm{WN}(\widehat{\boldsymbol{\mu}}_\varepsilon, \widehat{\boldsymbol{\Sigma}}_\varepsilon)$,

$$\widehat{\boldsymbol{\mu}}_e = \begin{pmatrix} 3.41 \\ 2.32 \\ 4.39 \\ 4.30 \\ 3.40 \\ 4.91 \\ 4.77 \end{pmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_e = \begin{pmatrix} 1.56 & & & & & & \\ 1.26 & 1.05 & & & & & \\ 1.71 & 1.34 & 1.91 & & & & \\ 1.90 & 1.49 & 2.10 & 2.33 & & & \\ 1.37 & 1.16 & 1.46 & 1.58 & 1.37 & & \\ 1.67 & 1.26 & 1.91 & 2.09 & 1.37 & 1.97 & \\ 1.41 & 1.14 & 1.58 & 1.67 & 1.39 & 1.56 & 1.53 \end{pmatrix}.$$

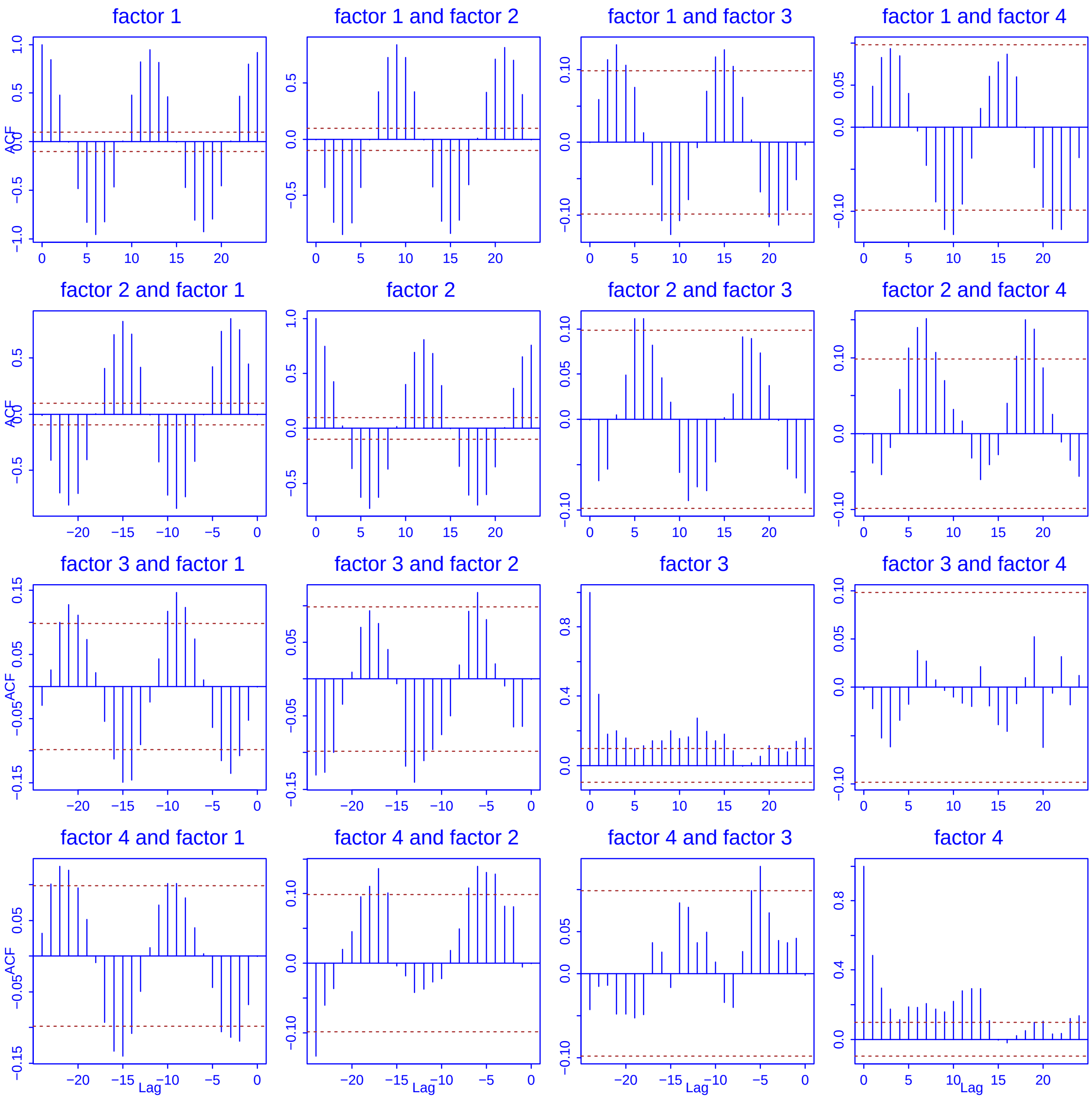
$$\widehat{\mathbf{A}} = \begin{pmatrix} .394 & .386 & .378 & .387 & .363 & .376 & .366 \\ -.086 & .225 & -.640 & -.271 & .658 & -.014 & .164 \\ .395 & .0638 & -.600 & .346 & -.494 & -.074 & .332 \\ .687 & -.585 & -.032 & -.306 & .173 & .206 & -.139 \end{pmatrix}^\tau,$$

$\boldsymbol{\xi}_t$ are PCAed factors: 1st PC accounts for 99% of TV of 4 factors, and 97.6% of the original 7 series.

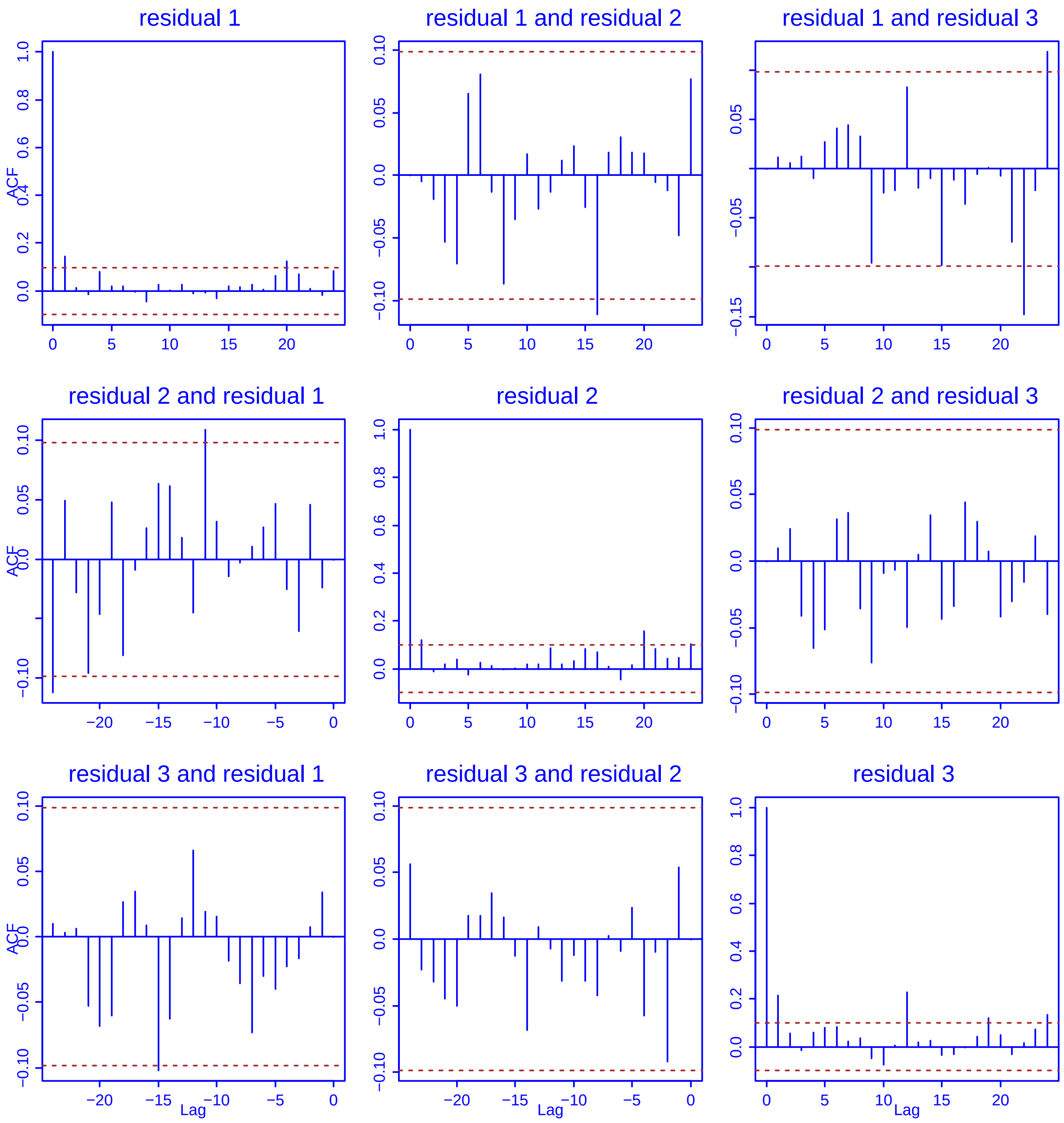
Time plots of the 4 estimated factors VAR(1)



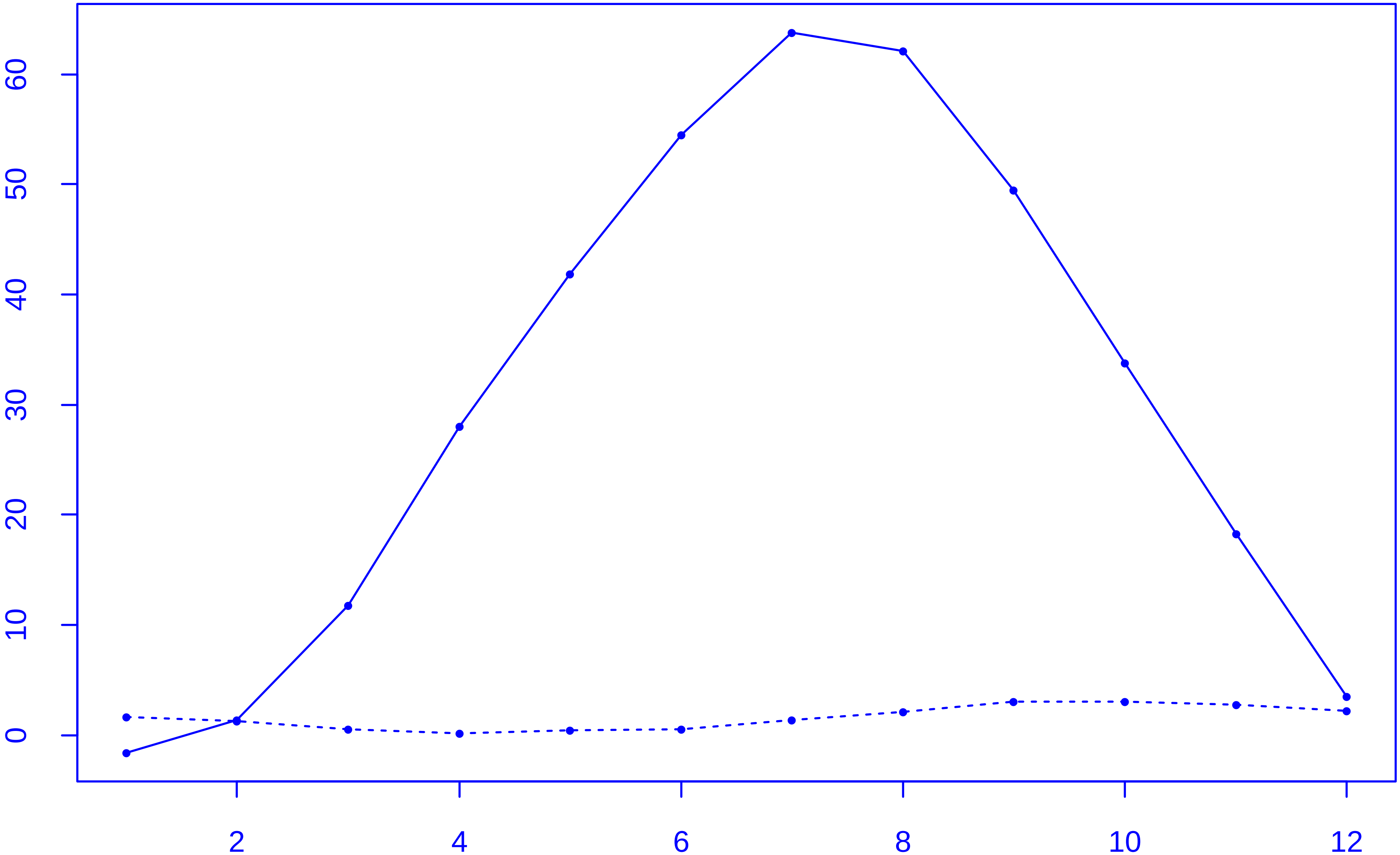
Sample cross-correlation of the 4 estimated factors



Sample cross-correlation of the 3 residuals (i.e. $\widehat{\mathbf{B}}^\tau \mathbf{Y}_t$)



Since the first two factors are dominated by periodic components, we remove them before fitting.



In the fitted factor model $\mathbf{Y}_t = \widehat{\mathbf{A}}\boldsymbol{\xi}_t + \mathbf{e}_t$, the AICC selected VAR(1) for the factor process:

$$\boldsymbol{\xi}_t - \boldsymbol{\alpha}_t = \widehat{\boldsymbol{\varphi}}_0 + \widehat{\boldsymbol{\Phi}}_1(\boldsymbol{\xi}_{t-1} - \boldsymbol{\alpha}_{t-1}) + \mathbf{u}_t,$$

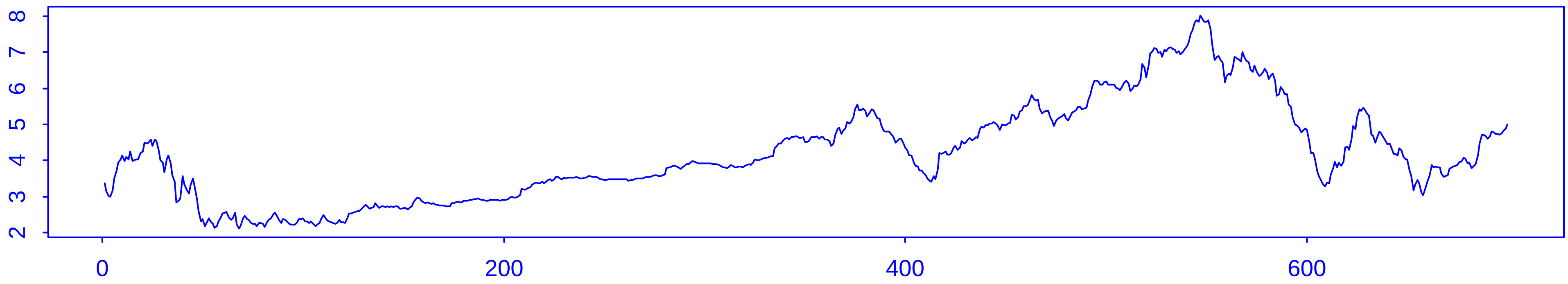
where $\boldsymbol{\alpha}_t^\tau = (p_{t1}, p_{t2}, 0, 0)$ is the periodic component, and

$$\widehat{\boldsymbol{\Phi}}_1 = \begin{pmatrix} .27 & -.31 & .72 & .40 \\ .01 & .36 & -.04 & .04 \\ .00 & -.01 & .42 & -.02 \\ -.00 & .03 & .03 & .48 \end{pmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_u = \begin{pmatrix} 14.24 & & & \\ -.17 & .23 & & \\ -.02 & .03 & .05 & \\ .042 & .01 & -.00 & .05 \end{pmatrix},$$

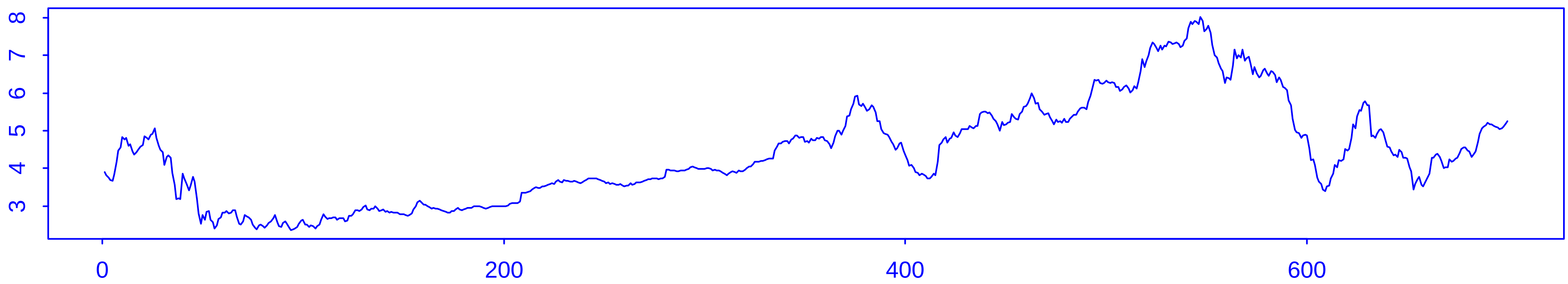
$\widehat{\boldsymbol{\varphi}}_0 = (.07, -.02, -.11, .10)^\tau$.

- Temperature dynamics in the 7 cities may be modelled in terms of 4 common factors
- The annual periodic fluctuations may be explained by a single common factor
- Removing the periodic components, the dynamics of the 4 common factors may be represented by an AR(1) model

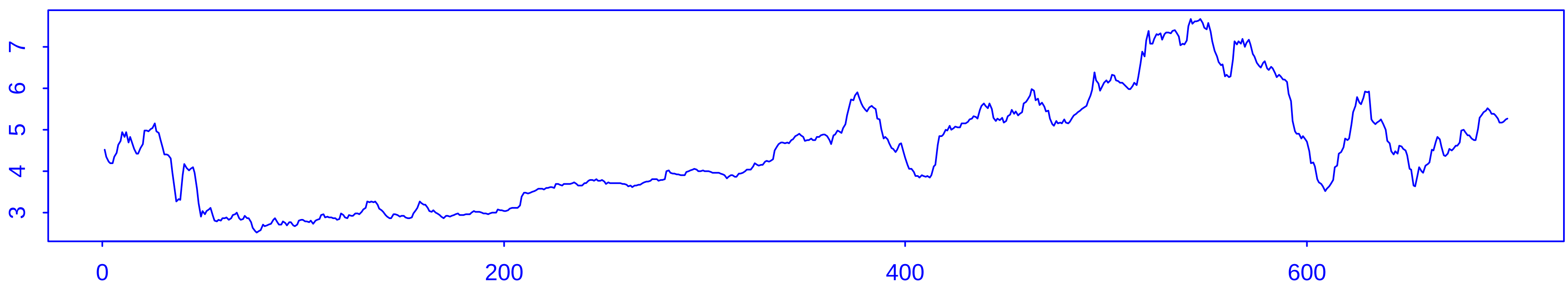
3−month Treasuary bills



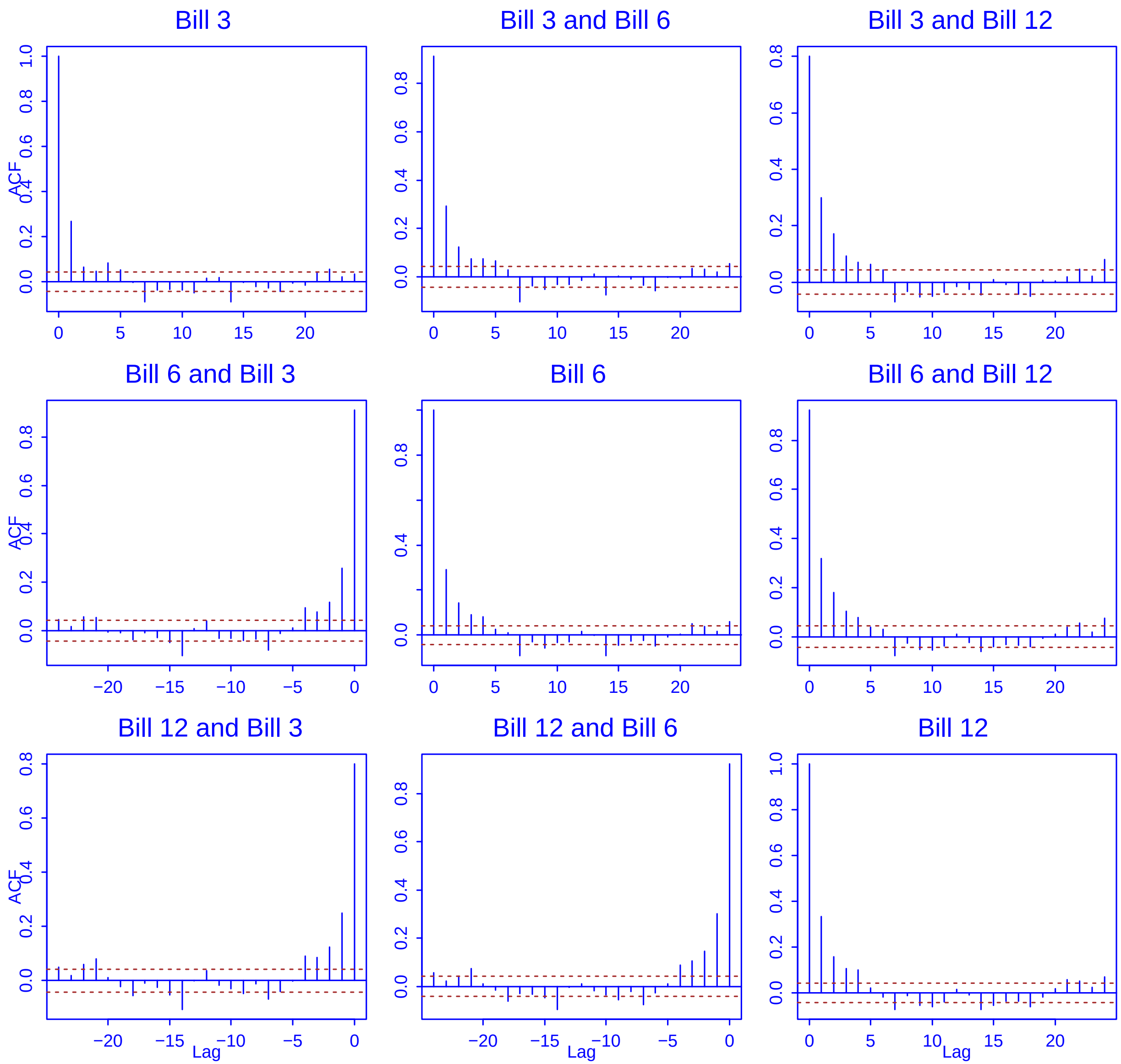
6−month Treasuary bills



12−month Treasuary bills



Sample cross-correlation of the differenced Treasury bills



With $p=15,\ \alpha=5\%$, $\widehat{r}=2$, $\mathbf{Y}_t=\widehat{\mathbf{A}}\boldsymbol{\xi}_t+\mathbf{e}_t$,

$$\widehat{\mathbf{A}}=\begin{pmatrix} .719 & -.547 \\ .452 & -.102 \\ .529 & .831 \end{pmatrix}, \quad \widehat{\boldsymbol{\mu}}_\varepsilon=\begin{pmatrix} .0006 \\ .0010 \\ .0007 \end{pmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_\varepsilon=\begin{pmatrix} .004 & & \\ .007 & .011 & \\ .005 & .008 & .005 \end{pmatrix}.$$

There exist little cross-correlation between the two factor series.

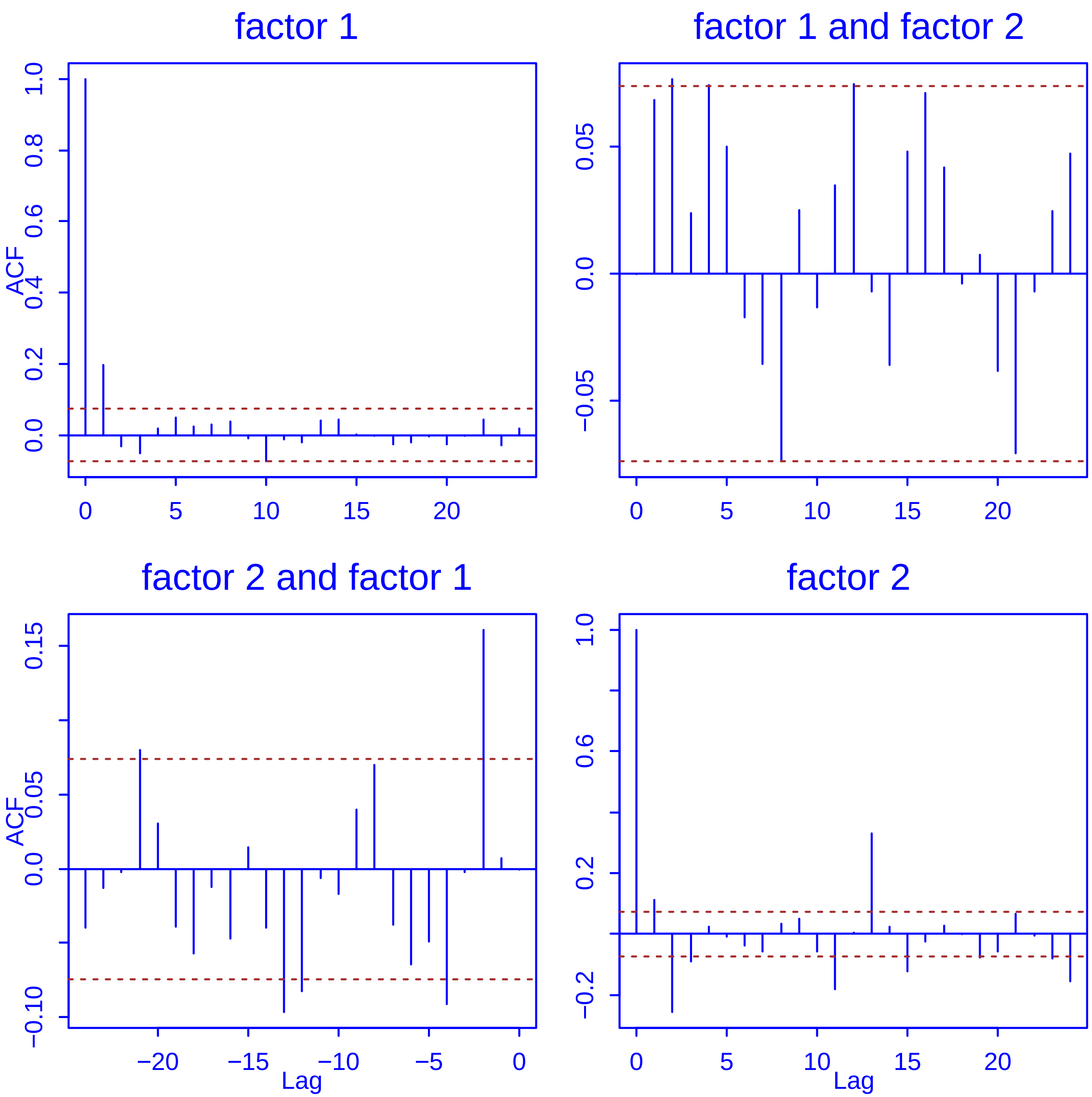
AICC models:

$$\xi_{t1}=1.64\xi_{t-1,1}-1.31\xi_{t-2,1}+.27\xi_{t-3,1}+u_{t1}-1.45u_{t-1,1}+.096u_{t-2,1},$$

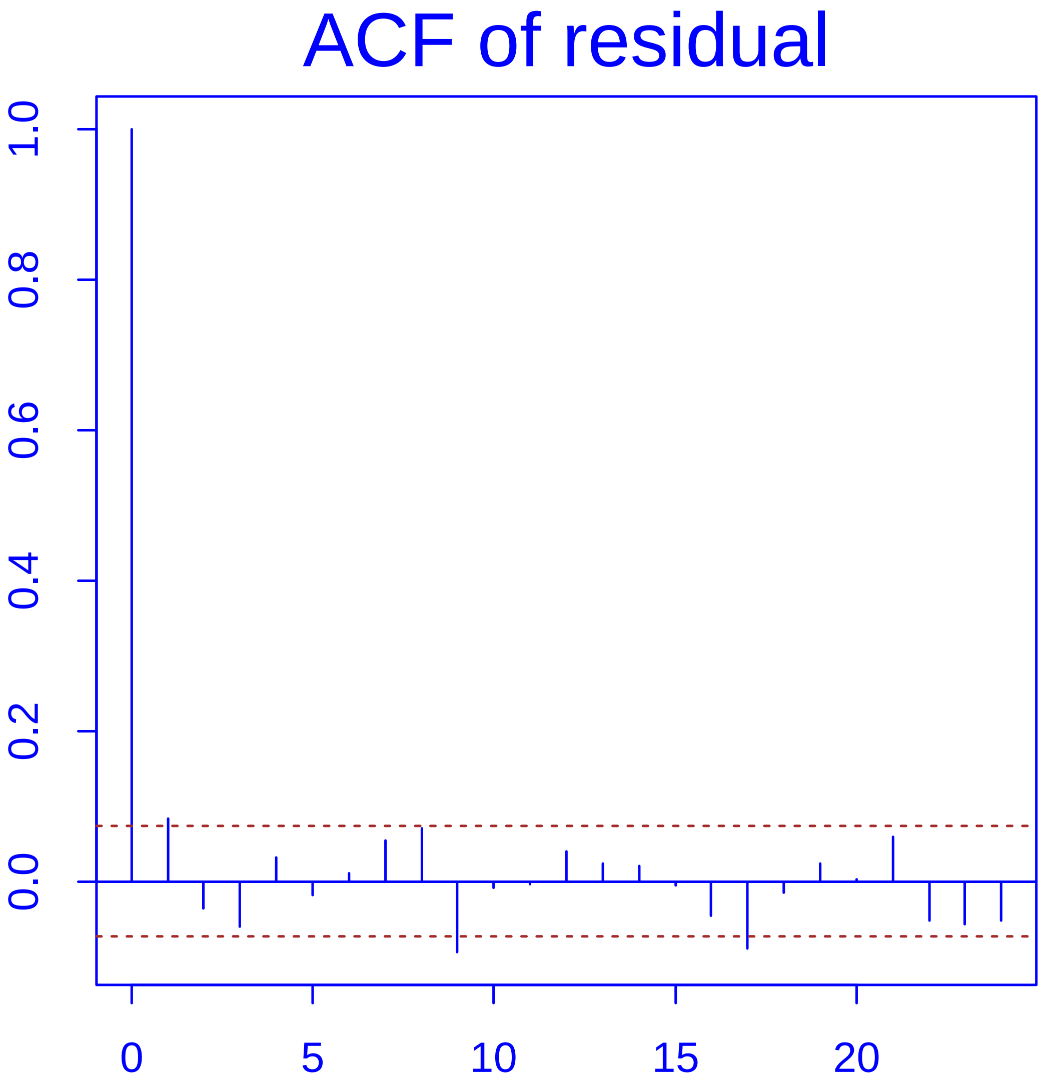
$$\begin{aligned}\xi_{t2} &= -0.04\xi_{t-7,2}-0.04\xi_{t-10,2}+0.74\xi_{t-13,2}+u_{t2}+0.09u_{t-1,2} \\ &\quad -0.20u_{t-2,2}-0.07u_{t-3,2}-0.04u_{t-5,2}-0.07u_{t-12,2}-0.49u_{t-13,2},\end{aligned}$$

where $u_{t1}\sim \mathrm{WN}(0,.017)$, $u_{t2}\sim \mathrm{WN}(0,.003)$

Sample cross-correlation functions of the 2 estimated factors



Sample ACF of $\widehat{\mathbf{B}}^{\tau}\mathbf{Y}_t$



Final Remarks

Factor models — a useful tool to reduce the dimensionality

Final Remarks

Factor models — a useful tool to reduce the dimensionality

- Still need to fit multi-dimensional factors
- Length of time series

Final Remarks

Factor models — a useful tool to reduce the dimensionality

- Still need to fit multi-dimensional factors
- Length of time series

A new algorithm for estimating conditional variance:

multivariate volatility models