Multiple Forecasting based on Time Series PCA

Qiwei Yao

Department of Statistics, London School of Economics

q.yao@lse.ac.uk

Joint work with

Jinyuan Chang and Bin Guo

Southwestern University of Finance and Economics, Chengdu

Xinhao Qiao

London School of Economics

Conventional wisdom:

> When forecasting large number of time series, one will typically be better off from forecasting each time series separately, as the potential gain from looking into the cross correlations is not enough to offset the estimation error . . .

VAR and VARMA are overparametrized!

A way out: Time Series PCA (R-package: `PCA4TS`)

- Goal of study:

  a high-dim TS $\rightarrow$ several *uncorrelated* lower-dim TS

- Methodology: PCA for time series
  - Transformation via an eigenanalysis
  - Permutation
    - maximum cross correlations
    - FDR based on multiple tests

- Real data illustration
- Multiple volatility processes
- A bundle of curve time series: on-going ...

Goal: For $p \times 1$ weakly stationary time series $\mathbf{y}_t$, search for a contemporaneous linear transformation:

$$\mathbf{y}_t = \mathbf{A}\mathbf{x}_t, \quad \text{or} \quad \mathbf{x}_t = \mathbf{B}\mathbf{y}_t \quad (i.e.\ \mathbf{B} = \mathbf{A}^{-1})$$

such that

$$\mathbf{x}_t = \begin{pmatrix} \mathbf{x}_t^{(1)} \\ \vdots \\ \mathbf{x}_t^{(q)} \end{pmatrix}, \quad \mathrm{Cov}(\mathbf{x}_t^{(i)}, \mathbf{x}_s^{(j)}) = \mathbf{0} \quad \forall\, i \neq j \ \text{ and } \ t, s.$$

Hence, $\mathbf{x}_t^{(1)}, \cdots, \mathbf{x}_t^{(q)}$ can be forecasted separately!

Goal: For $p \times 1$ weakly stationary time series $\mathbf{y}_t$, search for a contemporaneous linear transformation:

$$\mathbf{y}_t = \mathbf{A}\mathbf{x}_t, \quad \text{or} \quad \mathbf{x}_t = \mathbf{B}\mathbf{y}_t \quad (i.e.\ \mathbf{B} = \mathbf{A}^{-1})$$

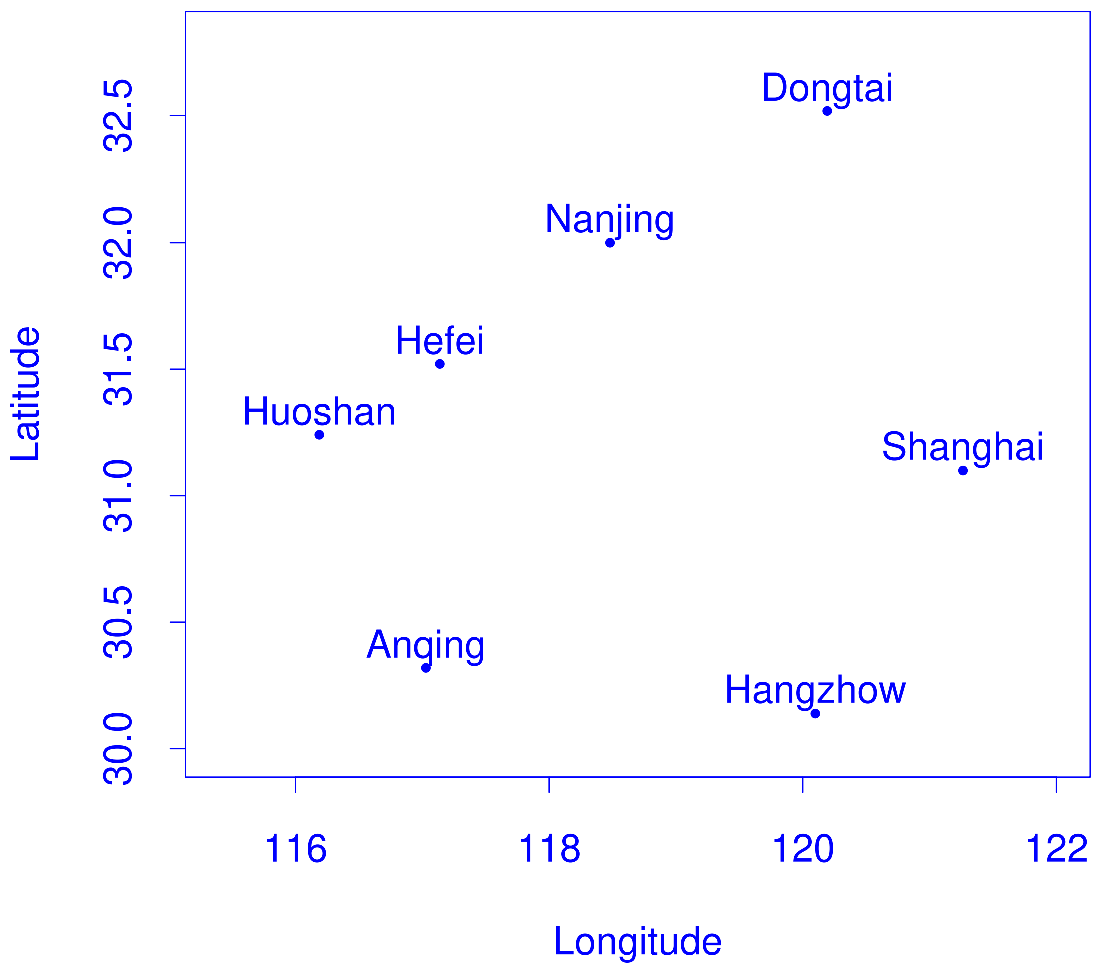
such that

$$\mathbf{x}_t = \begin{pmatrix} \mathbf{x}_t^{(1)} \\ \vdots \\ \mathbf{x}_t^{(q)} \end{pmatrix}, \quad \mathrm{Cov}(\mathbf{x}_t^{(i)}, \mathbf{x}_s^{(j)}) = \mathbf{0} \quad \forall\, i \neq j \ \text{ and } \ t, s.$$

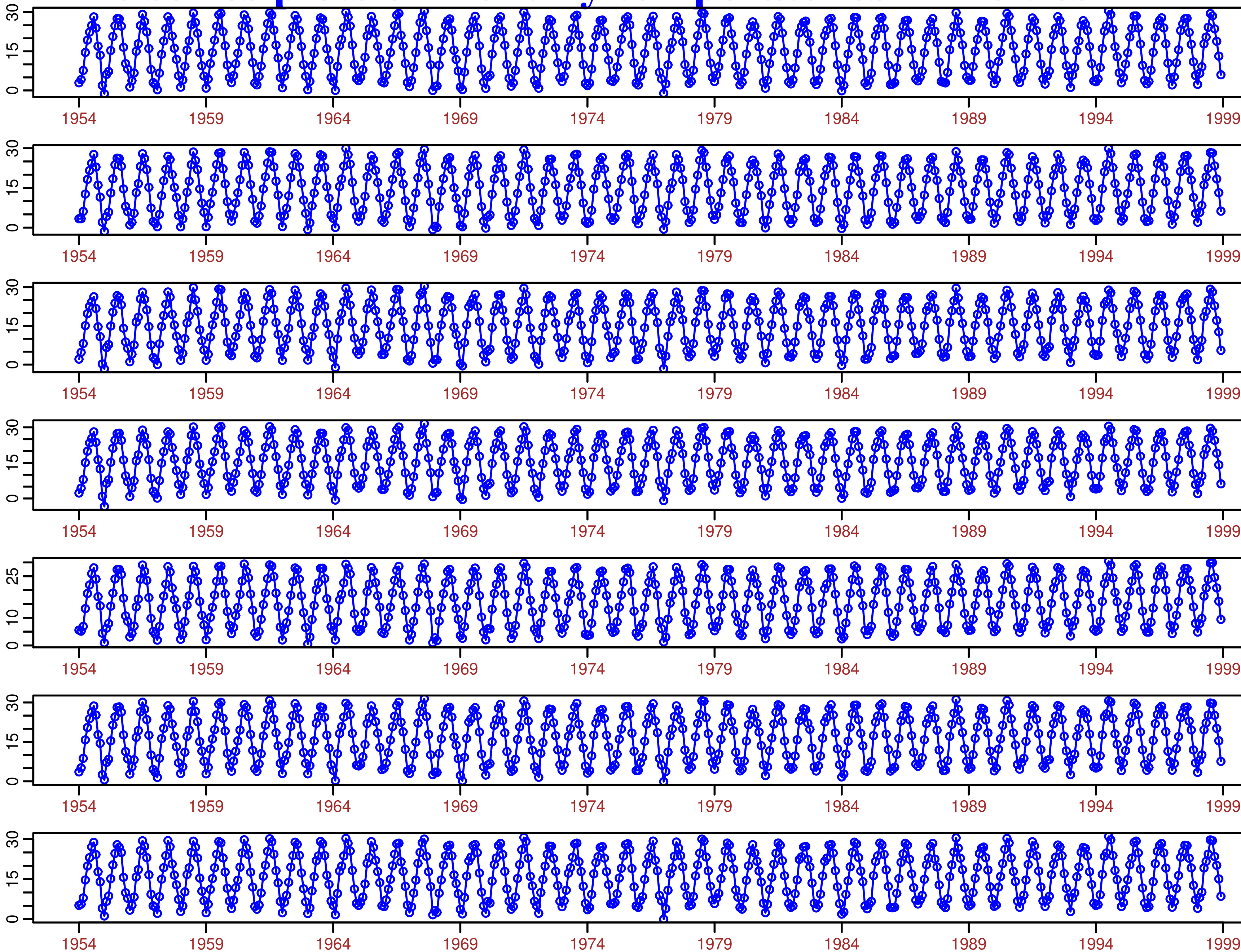
Hence, $\mathbf{x}_t^{(1)}, \cdots, \mathbf{x}_t^{(q)}$ can be forecasted separately!

- realistic?
- how to find B and $\mathbf{x}_t$?

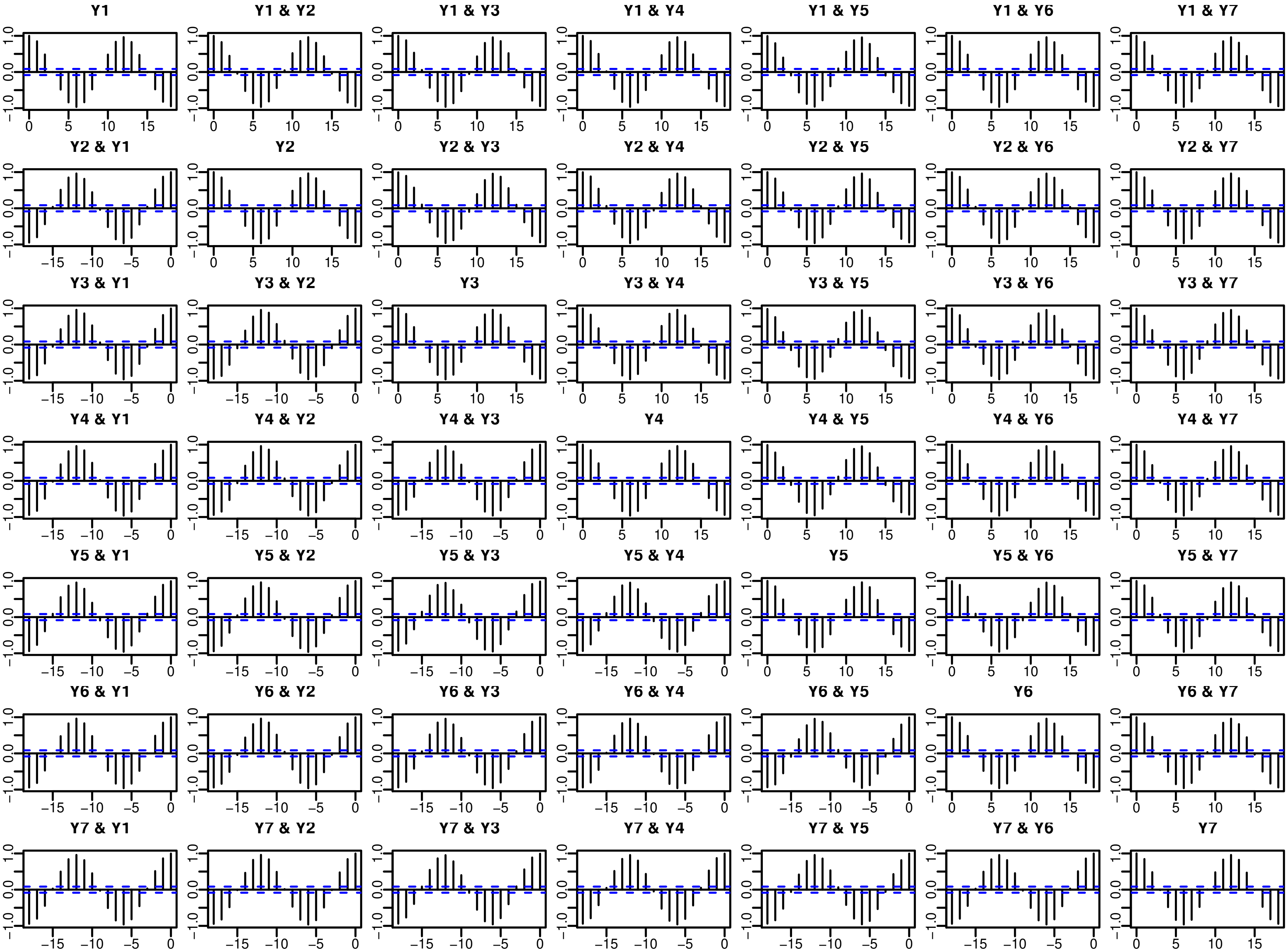
Example 1. Monthly temperatures in 1954 - 1998 in $p = 7$ cities in Eastern China ($n = 540$).



Time series plots of monthly temperatures in 7 cities



CCF of monthly temperatures in 7 cities

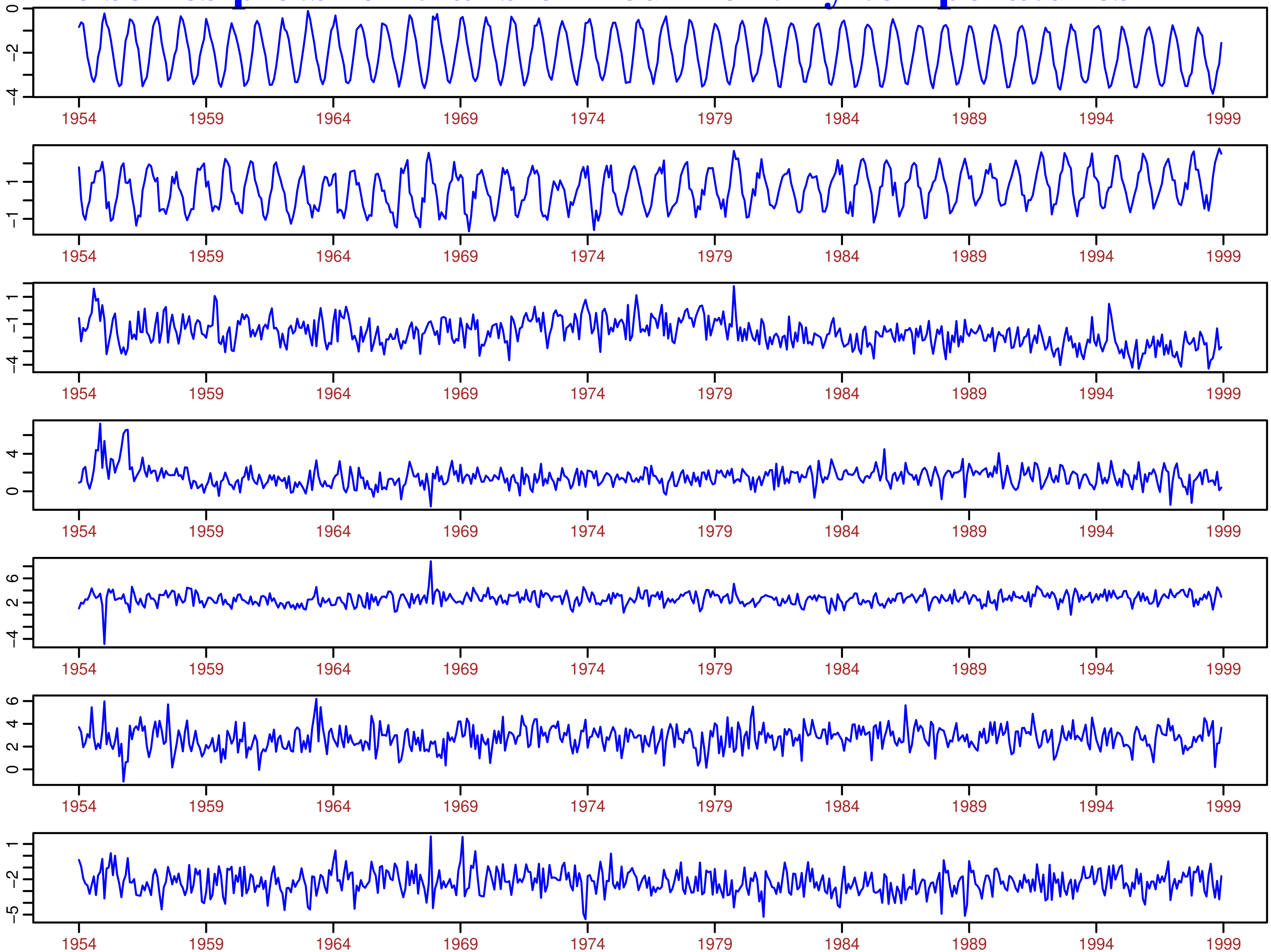


Transformation: $\widehat{\mathbf{x}}_t = \widehat{\mathbf{B}}\mathbf{y}_t$

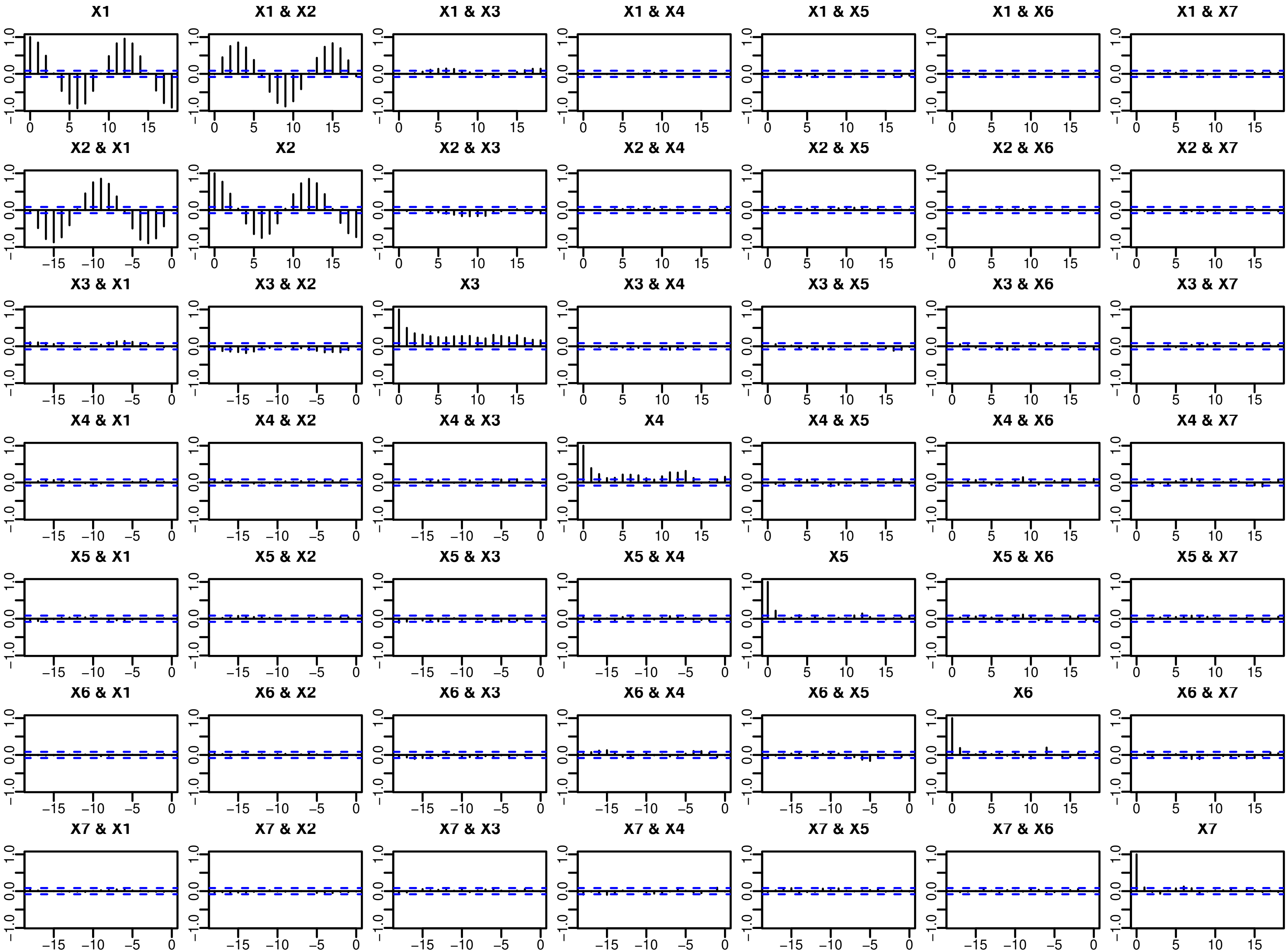
$$\widehat{\mathbf{B}} = \begin{pmatrix} 0.244 & -0.066 & 0.0187 & -0.050 & -0.313 & -0.154 & 0.200 \\ -0.703 & 0.324 & -0.617 & 0.189 & 0.633 & 0.499 & -0.323 \\ 0.375 & 1.544 & -1.615 & 0.170 & -2.266 & 0.126 & 1.596 \\ 3.025 & -1.381 & -0.787 & -1.691 & -0.212 & 1.188 & -0.165 \\ -0.197 & -1.820 & -1.416 & 3.269 & .301 & -1.438 & 1.299 \\ -0.584 & -0.354 & 0.847 & -1.262 & -0.218 & -0.151 & 1.831 \\ 1.869 & -0.742 & 0.034 & 0.501 & 0.492 & -2.533 & 0.339 \end{pmatrix}$$

$n = 540$, $p = 7$

Time series plots for transformed monthly temperatures



CCF for transformed monthly temperatures



Segmentation: $\{1,2,3\},\ \{4\},\ \{5\},\ \{6\},\ \{7\}$

Post-Sample Forecast

Forecasting based on segmentation: fit each subseries of $\mathbf{x}_t$ with a VAR model, forecast $\mathbf{x}_t$ based on the fitted models, the forecasts for $\mathbf{y}_t$ are obtained via $\mathbf{y}_t = \widehat{\mathbf{B}}^{-1}\mathbf{x}_t$.

Compare with the forecasts based on fitting a VAR and a restricted VAR (RVAR) directly to $\mathbf{y}_t$ (using `VAR` in *R*-package `vars`)

For each of the last 24 values (i.e. the monthly temperatures in 1997-1998), we use the data upto the previous month for the fittings. We calculate MSE of one-step-ahead forecasts, two-step-ahead forecasts (by plug-in) for each of 7 cities.

	One-step MSE	Two-step MSE
VAR	2.470 (0.416)	2.559 (0.385)
RVAR	2.530 (0.398)	2.615 (0.382)
via TS-PCA (5 groups)	2.221 (0.339)	2.203 (0.323)

Observations: $\mathbf{y}_1, \cdots, \mathbf{y}_n$ from a ${}_{p\times 1}$ weakly stationary TS

Assumption: $\mathbf{y}_t = \mathbf{A}\mathbf{x}_t$, and

$$\mathbf{x}_t = \begin{pmatrix} \mathbf{x}_t^{(1)} \\ \vdots \\ \mathbf{x}_t^{(q)} \end{pmatrix}, \quad \mathrm{Cov}(\mathbf{x}_t^{(i)}, \mathbf{x}_s^{(j)}) = \mathbf{0} \quad \forall\, i \neq j \text{ and } t, s.$$

Without loss of generality: $\mathrm{Var}(\mathbf{y}_t) = \mathrm{Var}(\mathbf{x}_t) = \mathbf{I}_p$, and thus

$$\mathbf{A}'\mathbf{A} = \mathbf{I}_p, \quad \textit{i.e. } \mathbf{A} \text{ is orthogonal}, \quad \mathbf{x}_t = \mathbf{A}'\mathbf{y}_t$$

Goal: estimate $\mathbf{A} = (\mathbf{A}_1, \cdots, \mathbf{A}_q)$, or more precisely, $\mathcal{M}(\mathbf{A}_1), \cdots,$

$\mathcal{M}(\mathbf{A}_q)$, as $\quad \widehat{\mathbf{x}}_t^{(j)} = \widehat{\mathbf{A}}_j'\mathbf{y}_t, \ j = 1, \cdots, q$.

Note. $(\mathbf{A}, \ \mathbf{x}_t)$ can be replaced by $(\mathbf{A}\mathbf{H}, \ \mathbf{H}'\mathbf{x}_t)$ for any $\mathbf{H} = \mathrm{diag}(\mathbf{H}_1, \cdots, \mathbf{H}_q)$ with $\mathbf{H}_j'\mathbf{H}_j = \mathbf{I}_{p_j}$

Step 1: Transformation via eigenanalysis

Notation: $\Sigma_y(k) = \text{Cov}(\mathbf{y}_{t+k}, \mathbf{y}_t)$, $\Sigma_x(k) = \text{Cov}(\mathbf{x}_{t+k}, \mathbf{x}_t)$,

$$\mathbf{W}_y = \sum_{k=0}^{k_0} \Sigma_y(k)\Sigma_y(k)' = \mathbf{I}_p + \sum_{k=1}^{k_0} \Sigma_y(k)\Sigma_y(k)',$$

$$\mathbf{W}_x = \sum_{k=0}^{k_0} \Sigma_x(k)\Sigma_x(k)' = \mathbf{I}_p + \sum_{k=1}^{k_0} \Sigma_x(k)\Sigma_x(k)'.$$

Step 1: Transformation via eigenanalysis

Notation: $\boldsymbol{\Sigma}_y(k) = \text{Cov}(\mathbf{y}_{t+k}, \mathbf{y}_t)$, $\boldsymbol{\Sigma}_x(k) = \text{Cov}(\mathbf{x}_{t+k}, \mathbf{x}_t)$,

$$\mathbf{W}_y = \sum_{k=0}^{k_0} \boldsymbol{\Sigma}_y(k)\boldsymbol{\Sigma}_y(k)' = \mathbf{I}_p + \sum_{k=1}^{k_0} \boldsymbol{\Sigma}_y(k)\boldsymbol{\Sigma}_y(k)',$$

$$\mathbf{W}_x = \sum_{k=0}^{k_0} \boldsymbol{\Sigma}_x(k)\boldsymbol{\Sigma}_x(k)' = \mathbf{I}_p + \sum_{k=1}^{k_0} \boldsymbol{\Sigma}_x(k)\boldsymbol{\Sigma}_x(k)'.$$

As $\boldsymbol{\Sigma}_y(k) = \mathbf{A}\boldsymbol{\Sigma}_x(k)\mathbf{A}'$, $\mathbf{W}_y = \mathbf{A}\mathbf{W}_x\mathbf{A}'$.

Step 1: Transformation via eigenanalysis

Notation: $\boldsymbol{\Sigma}_y(k) = \mathrm{Cov}(\mathbf{y}_{t+k}, \mathbf{y}_t)$, $\boldsymbol{\Sigma}_x(k) = \mathrm{Cov}(\mathbf{x}_{t+k}, \mathbf{x}_t)$,

$$\mathbf{W}_y = \sum_{k=0}^{k_0} \boldsymbol{\Sigma}_y(k)\boldsymbol{\Sigma}_y(k)' = \mathbf{I}_p + \sum_{k=1}^{k_0} \boldsymbol{\Sigma}_y(k)\boldsymbol{\Sigma}_y(k)',$$

$$\mathbf{W}_x = \sum_{k=0}^{k_0} \boldsymbol{\Sigma}_x(k)\boldsymbol{\Sigma}_x(k)' = \mathbf{I}_p + \sum_{k=1}^{k_0} \boldsymbol{\Sigma}_x(k)\boldsymbol{\Sigma}_x(k)'.$$

As $\boldsymbol{\Sigma}_y(k) = \mathbf{A}\boldsymbol{\Sigma}_x(k)\mathbf{A}'$, $\mathbf{W}_y = \mathbf{A}\mathbf{W}_x\mathbf{A}'$.

Eigenanalysis: $\mathbf{W}_x\boldsymbol{\Gamma}_x = \boldsymbol{\Gamma}_x\mathbf{D}$, columns of $\boldsymbol{\Gamma}_x$ are eigenvectors of $\mathbf{W}_x$ with the eigenvalues in diagonal matrix $\mathbf{D}$.

$$\mathbf{W}_y\mathbf{A}\boldsymbol{\Gamma}_x = \mathbf{A}\mathbf{W}_x\mathbf{A}'\mathbf{A}\boldsymbol{\Gamma}_x = \mathbf{A}\mathbf{W}_x\boldsymbol{\Gamma}_x = \mathbf{A}\boldsymbol{\Gamma}_x\mathbf{D}$$

Thus $\boldsymbol{\Gamma}_y = \mathbf{A}\boldsymbol{\Gamma}_x$, and $\boldsymbol{\Gamma}_y'\mathbf{y}_t = \boldsymbol{\Gamma}_x'\mathbf{A}'\mathbf{y}_t = \boldsymbol{\Gamma}_x'\mathbf{x}_t$.

$\mathbf{z}_t \equiv \mathbf{\Gamma}'_y \mathbf{y}_t (= \mathbf{\Gamma}'_x \mathbf{x}_t)$ is the required transformation upto a permutation

$\mathbf{z}_t \equiv \mathbf{\Gamma}_y' \mathbf{y}_t (= \mathbf{\Gamma}_x' \mathbf{x}_t)$ is the required transformation upto a permutation

Note $\mathbf{W}_x = \mathrm{diag}(\mathbf{W}_{x,1}, \cdots, \mathbf{W}_{x,q})$

Proposition 1. (i) $\mathbf{\Gamma}_x$ can be taken with the same block-diagonal structure as $\mathbf{W}_x$.

(ii) Any $\mathbf{\Gamma}_x$ is a column-permutation of $\mathbf{\Gamma}_x$ described in (i), provided $\lambda(\mathbf{W}_{x,i}) \neq \lambda(\mathbf{W}_{x,j})$ for any $i \neq j$.

$\mathbf{z}_t \equiv \mathbf{\Gamma}_y'\mathbf{y}_t (= \mathbf{\Gamma}_x'\mathbf{x}_t)$ is the required transformation upto a permutation

---

Note $\mathbf{W}_x = \mathrm{diag}(\mathbf{W}_{x,1}, \cdots, \mathbf{W}_{x,q})$

Proposition 1. (i) $\mathbf{\Gamma}_x$ can be taken with the same block-diagonal structure as $\mathbf{W}_x$.

(ii) Any $\mathbf{\Gamma}_x$ is a column-permutation of $\mathbf{\Gamma}_x$ described in (i), provided $\lambda(\mathbf{W}_{x,i}) \neq \lambda(\mathbf{W}_{x,j})$ for any $i \neq j$.

For $\mathbf{\Gamma}_x$ specified in (i), $\mathbf{\Gamma}_x'\mathbf{x}_t$ is segmented exactly the same as $\mathbf{x}_t$

$\mathbf{z}_t \equiv \mathbf{\Gamma}_y' \mathbf{y}_t (= \mathbf{\Gamma}_x' \mathbf{x}_t)$ is the required transformation upto a permutation

---

Note $\mathbf{W}_x = \mathrm{diag}(\mathbf{W}_{x,1}, \cdots, \mathbf{W}_{x,q})$

Proposition 1. (i) $\mathbf{\Gamma}_x$ can be taken with the same block-diagonal structure as $\mathbf{W}_x$.

(ii) Any $\mathbf{\Gamma}_x$ is a column-permutation of $\mathbf{\Gamma}_x$ described in (i), provided $\lambda(\mathbf{W}_{x,i}) \neq \lambda(\mathbf{W}_{x,j})$ for any $i \neq j$.

For $\mathbf{\Gamma}_x$ specified in (i), $\mathbf{\Gamma}_x' \mathbf{x}_t$ is segmented exactly the same as $\mathbf{x}_t$

Let

$$\widehat{\mathbf{W}}_y = \mathbf{I}_p + \sum_{k=1}^{k_0} \widehat{\mathbf{\Sigma}}_y(k) \widehat{\mathbf{\Sigma}}_y(k)', \qquad \widehat{\mathbf{W}}_y \widehat{\mathbf{\Gamma}}_y = \widehat{\mathbf{\Gamma}}_y \widehat{\mathbf{D}}.$$

Then $\widehat{\mathbf{z}}_t = \widehat{\mathbf{\Gamma}}_y' \mathbf{y}_t$ – require permute components of $\widehat{\mathbf{z}}_t$ to obtain $\widehat{\mathbf{x}}_t$

Permutation

Goal: put the connected components of $\widehat{\mathbf{z}}_t = \widehat{\mathbf{\Gamma}}_y' \mathbf{y}_t$ together.

Visual examination of CCF if p is not large!

Two component series of $\widehat{\mathbf{z}}_t$ is <u>connected</u> if the multiple null hypothesis

$$H_0 : \rho(k) = 0 \quad \text{for any } k = 0, \pm 1, \pm 2, \ldots, \pm m$$

is rejected, where $\rho(k)$ is cross correlation between two series at lag k.

Permutation is performed as follows:

i. Start with p groups: each containing one component of $\widehat{\mathbf{z}}_t$.

ii. Combine the two groups together if one connected pair are split over two groups.

iii. Repeat Step ii above until all connected components are within one group.

Permutation Method I: Max CCF

Put $\widehat{\mathbf{z}}_t = (\widehat{z}_{1,t}, \cdots, \widehat{z}_{p,t})'$.

Let $\widehat{\rho}_{i,j}(h)$ be the sample CCF of $(\widehat{z}_{i,t}, \widehat{z}_{j,t})$ at lag h, and

$$\widehat{L}_n(i,j) = \max_{|h| \le m} |\widehat{\rho}_{i,j}(h)|,$$

reject H_0 for the pair $(\widehat{z}_{i,t}, \widehat{z}_{j,t})$ for large values of $\widehat{L}_n(i,j)$.

Permutation Method I: Max CCF

Put $\widehat{\mathbf{z}}_t = (\widehat{z}_{1,t}, \cdots, \widehat{z}_{p,t})'$.

Let $\widehat{\rho}_{i,j}(h)$ be the sample CCF of $(\widehat{z}_{i,t}, \widehat{z}_{j,t})$ at lag h, and

$$\widehat{L}_n(i,j) = \max_{|h| \le m} |\widehat{\rho}_{i,j}(h)|,$$

reject H_0 for the pair $(\widehat{z}_{i,t}, \widehat{z}_{j,t})$ for large values of $\widehat{L}_n(i,j)$.

Line up $\widehat{L}_n(i,j)$, $1 \le i < j \le p$, in the descending order:

$$\widehat{L}_1 \ge \cdots \ge \widehat{L}_{p_0}, \qquad p_0 = p(p-1)/2.$$

Define

$$\widehat{r} = \arg \max_{1 \le j < c_0 p_0} \widehat{L}_j / \widehat{L}_{j+1}, \qquad c_0 \in (0,1).$$

Reject H_0 for the pairs corresponding to $\widehat{L}_1, \cdots, \widehat{L}_{\widehat{r}}$

Graph representation: Let vertexes $\widehat{V} = \{1, \cdots, p\}$ stand for the components of $\widehat{\mathbf{z}}_t = \widehat{\mathbf{\Gamma}}'\mathbf{y}_t$, and

$$\widehat{E}_n = \{\text{edge connecting } i \text{ and } j : \widehat{z}_{i,t}, \widehat{z}_{j,t} \text{ are connected}\}.$$

Let $V = \{1, \cdots, p\}$ stand for the components of $\mathbf{z}_t = \mathbf{\Gamma}'\mathbf{y}_t$, and

$$E = \{\text{edge connecting } i \text{ and } j : \max_{-m \le h \le m} |\mathrm{Corr}(z_{i,t+h}, z_{j,t})| > 0\,\}.$$

Graph representation: Let vertexes $\widehat{V} = \{1, \cdots, p\}$ stand for the components of $\widehat{\mathbf{z}}_t = \widehat{\mathbf{\Gamma}}' \mathbf{y}_t$, and

$$\widehat{E}_n = \{\text{edge connecting } i \text{ and } j : \widehat{z}_{i,t}, \widehat{z}_{j,t} \text{ are connected}\}.$$

Let $V = \{1, \cdots, p\}$ stand for the components of $\mathbf{z}_t = \mathbf{\Gamma}' \mathbf{y}_t$, and

$$E = \{\text{edge connecting } i \text{ and } j : \max_{-m \le h \le m} |\mathrm{Corr}(z_{i,t+h}, z_{j,t})| > 0\,\}.$$

Let $\varpi = \min_{i \neq j} \min |\lambda(\mathbf{W}_{x,i}) - \lambda(\mathbf{W}_{x,j})|$,

$$\max_{1 \le i < j \le p} \max_{|h| \le m} |\mathrm{Corr}(z_{i,t+h}, z_{j,t})| > 0, \qquad \min_{1 \le i < j \le p} \max_{|h| \le m} |\mathrm{Corr}(z_{i,t+h}, z_{j,t})| = 0,$$

$$\widehat{r} = \arg \max_{1 \le j < p_0} (\widehat{L}_j + \delta_n) / (\widehat{L}_{j+1} + \delta_n).$$

Proposition 2. As $n \to \infty$, $\delta_n \to 0$, $\frac{1}{\varpi} \| \widehat{\mathbf{W}}_y - \mathbf{W}_y \|_2 = o(\delta_n)$, and $\log p = o(n^{\alpha})$. Then $P(\widehat{E}_n = E) \to 1$.

Prewhitening

To make CCF for different pairs comparable, prewhiten each component series of $\widehat{\mathbf{z}}_t = \widehat{\mathbf{\Gamma}}_y^{\prime}\mathbf{y}_t$ separately first.

Prewhitening

To make CCF for different pairs comparable, prewhiten each component series of $\widehat{\mathbf{z}}_t = \widehat{\mathbf{\Gamma}}_y' \mathbf{y}_t$ separately first.

(i) If $\rho_{i,j}(h) = 0$, $\widehat{\rho}_{i,j}(h) \sim N(0, 1/n)$ asymptotically,

provided at least one of $x_{i,t}$ and $x_{j,t}$ is white noise.

(ii) For $h \neq k$, $\widehat{\rho}_{i,j}(h),\ \widehat{\rho}_{i,j}(k)$ are asymptotically independent,

and $\mathrm{Cov}\{\widehat{\rho}_{i,j}(h), \widehat{\rho}_{i,j}(k)\} = o_P(1/n)$, provided both $x_{i,t}$ and $x_{j,t}$ are white noise.

Brockwell & Davis (1996, Corollary 7.3.1).

Prewhitening

To make CCF for different pairs comparable, prewhiten each component series of $\widehat{\mathbf{z}}_t = \widehat{\mathbf{\Gamma}}'_y \mathbf{y}_t$ separately first.

(i) If $\rho_{i,j}(h) = 0$, $\widehat{\rho}_{i,j}(h) \sim N(0, 1/n)$ asymptotically,

provided at least one of $x_{i,t}$ and $x_{j,t}$ is white noise.

(ii) For $h \neq k$, $\widehat{\rho}_{i,j}(h)$, $\widehat{\rho}_{i,j}(k)$ are asymptotically independent,

and $\mathrm{Cov}\{\widehat{\rho}_{i,j}(h), \widehat{\rho}_{i,j}(k)\} = o_P(1/n)$, provided both $x_{i,t}$ and $x_{j,t}$ are white noise.

Brockwell & Davis (1996, Corollary 7.3.1).

In practice, we filter out the autocorrelation for each component series of $\widehat{\mathbf{z}}_t$ by fitting an AR with the order determined by AIC and not greater than 5.

Permutation Method II: FDR based on multiple tests

To fix the idea, let $\xi_t, \ \eta_t$ be two WN, $\rho(k) = \text{Corr}(\xi_{t+k}, \eta_t) = 0$,

$$\widehat{\rho}(k) = \sum_{t=1}^{n-k} (\xi_{t+k} - \bar{\xi})(\eta_t - \bar{\eta}) \Big/ \Big\{ \sum_{t=1}^{n} (\xi_t - \bar{\xi})^2 \sum_{t=1}^{n} (\eta_t - \bar{\eta})^2 \Big\}^{1/2}.$$

Asymptotically $\widehat{\rho}(k) \sim N(0, 1/n)$, $\widehat{\rho}(k)$ and $\widehat{\rho}(h)$ are independent.

Permutation Method II: FDR based on multiple tests

To fix the idea, let $\xi_t, \ \eta_t$ be two WN, $\rho(k) = \text{Corr}(\xi_{t+k}, \eta_t) = 0$,

$$\widehat{\rho}(k) = \sum_{t=1}^{n-k} (\xi_{t+k} - \bar{\xi})(\eta_t - \bar{\eta}) \Big/ \Big\{ \sum_{t=1}^{n} (\xi_t - \bar{\xi})^2 \sum_{t=1}^{n} (\eta_t - \bar{\eta})^2 \Big\}^{1/2}.$$

Asymptotically $\widehat{\rho}(k) \sim N(0, 1/n)$, $\widehat{\rho}(k)$ and $\widehat{\rho}(h)$ are independent.

The P-value for testing $\rho(k) = 0$ (simple) is $p_k = 2\Phi(-\sqrt{n}|\widehat{\rho}(k)|)$

Permutation Method II: FDR based on multiple tests

To fix the idea, let $\xi_t, \ \eta_t$ be two WN, $\rho(k) = \mathrm{Corr}(\xi_{t+k}, \eta_t) = 0$,

$$\widehat{\rho}(k) = \sum_{t=1}^{n-k} (\xi_{t+k} - \bar{\xi})(\eta_t - \bar{\eta}) \Big/ \Big\{ \sum_{t=1}^{n} (\xi_t - \bar{\xi})^2 \sum_{t=1}^{n} (\eta_t - \bar{\eta})^2 \Big\}^{1/2}.$$

Asymptotically $\widehat{\rho}(k) \sim N(0, 1/n)$, $\widehat{\rho}(k)$ and $\widehat{\rho}(h)$ are independent.

The P-value for testing $\rho(k) = 0$ (simple) is $p_k = 2\Phi(-\sqrt{n}|\widehat{\rho}(k)|)$

Let $p_{(1)} \le \cdots \le p_{(2m+1)}$ be the order statistics of $\{p_k, \ |k| \le m\}$.

Simes (1986): For $H_0 : \rho(k) = 0, \ \forall \, |k| \le m$, a multiple test

rejects H_0 at the level α if

$$p_{(j)} \le j\alpha/(2m+1) \quad \text{for at least one } 1 \le j \le 2m+1$$

Permutation Method II: FDR based on multiple tests

To fix the idea, let $\xi_t, \ \eta_t$ be two WN, $\rho(k) = \mathrm{Corr}(\xi_{t+k}, \eta_t) = 0$,

$$\widehat{\rho}(k) = \sum_{t=1}^{n-k} (\xi_{t+k} - \bar{\xi})(\eta_t - \bar{\eta}) \Big/ \Big\{ \sum_{t=1}^{n} (\xi_t - \bar{\xi})^2 \sum_{t=1}^{n} (\eta_t - \bar{\eta})^2 \Big\}^{1/2}.$$

Asymptotically $\widehat{\rho}(k) \sim N(0, 1/n)$, $\widehat{\rho}(k)$ and $\widehat{\rho}(h)$ are independent.

The P-value for testing $\rho(k) = 0$ (simple) is $p_k = 2\Phi(-\sqrt{n}|\widehat{\rho}(k)|)$

Let $p_{(1)} \le \cdots \le p_{(2m+1)}$ be the order statistics of $\{p_k, \ |k| \le m\}$.

Simes (1986): For $H_0 : \rho(k) = 0, \ \forall \, |k| \le m$, a multiple test rejects H_0 at the level α if

$$p_{(j)} \le j\alpha/(2m+1) \quad \text{for at least one } 1 \le j \le 2m+1$$

The P-value for the multiple test is

$$P = \min\{\alpha > 0 : p_{(j)} \le j\alpha/(2m+1) \text{ for some } 1 \le j \le 2m+1\}$$

$$= \min_{1 \le j \le 2m+1} p_{(j)} \, (2m+1)/j.$$

For each pair components of $\widehat{\mathbf{z}}_t = \widehat{\mathbf{\Gamma}}_y' \mathbf{y}_t$, we test multiple hypothesis H_0, obtaining P-value $P_{i,j}$ for $1 \leq i < j \leq q$.

Arranging those P-values in ascending order:

$$P_{(1)} \leq P_{(2)} \leq \cdots \leq P_{(p_0)}, \qquad p_0 = p(p-1)/2$$

FDR: For a given small $\beta \in (0,1)$, let

$$\widehat{d} = \max\{k : 1 \leq k \leq p_0,\ P_{(k)} \leq k\beta/p_0\},$$

and rejects the hypothesis H_0 for the $\widehat{d}$ pairs of the components of $\mathbf{z}_t$ corresponding to the P-values $P_{(1)}, \cdots, P_{(\widehat{d})}$

For each pair components of $\widehat{\mathbf{z}}_t = \widehat{\mathbf{\Gamma}}'_y \mathbf{y}_t$, we test multiple hypothesis H_0, obtaining P-value $P_{i,j}$ for $1 \leq i < j \leq q$.

Arranging those P-values in ascending order:

$$P_{(1)} \leq P_{(2)} \leq \cdots \leq P_{(p_0)}, \qquad p_0 = p(p-1)/2$$

FDR: For a given small $\beta \in (0,1)$, let

$$\widehat{d} = \max\{k : 1 \leq k \leq p_0,\ P_{(k)} \leq k\beta/p_0\},$$

and rejects the hypothesis H_0 for the $\widehat{d}$ pairs of the components of $\mathbf{z}_t$ corresponding to the P-values $P_{(1)}, \cdots, P_{(\widehat{d})}$

(i) The P-values p_k ($|k| < m$) are asymptotically independent

(ii) The P-values $P_{i,j}$ ($1 \leq < j \leq p$) are not independent, causing difficulties in choosing β in FDR.

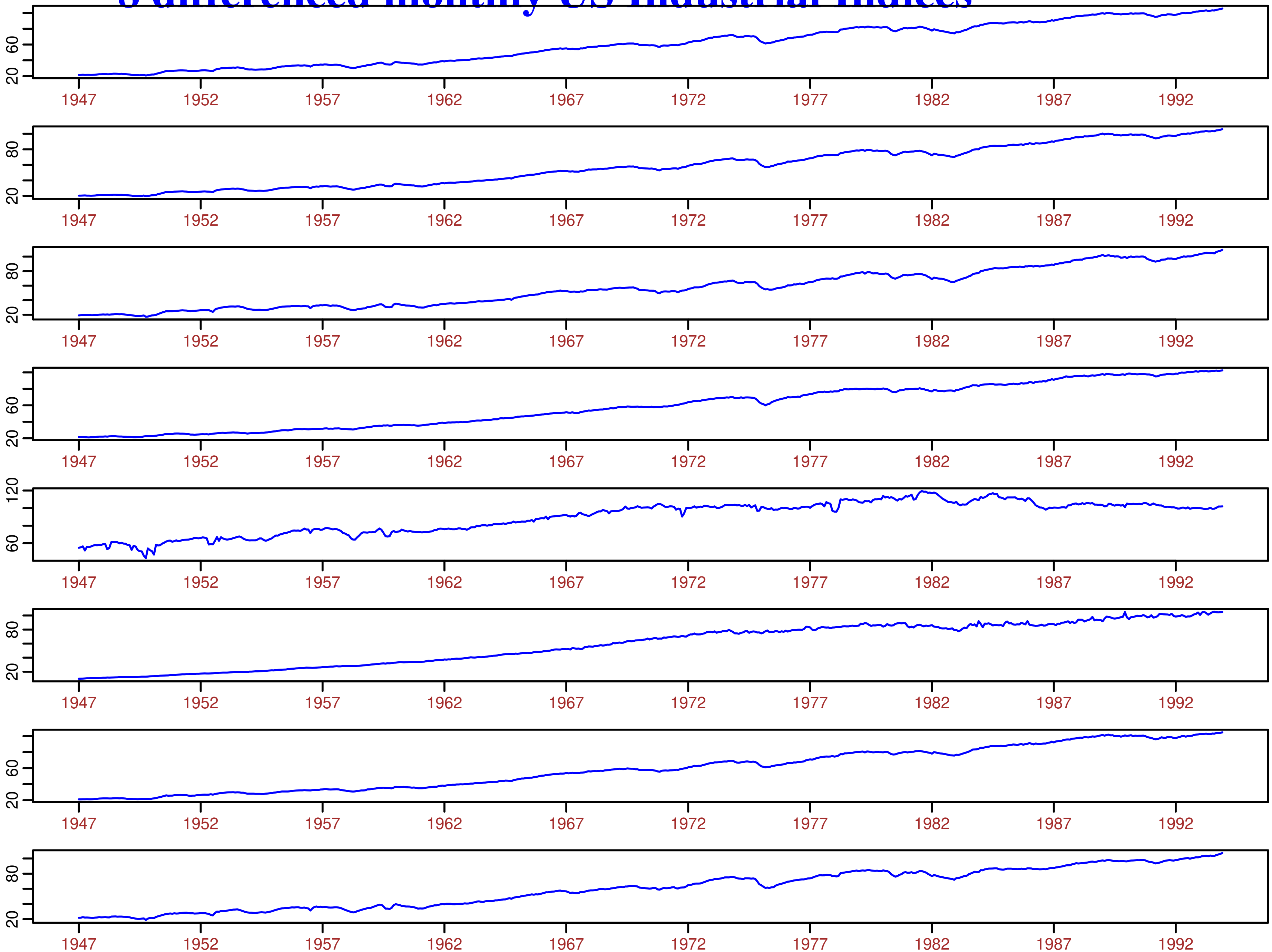
(iii) Ranking the pair components of $\widehat{\mathbf{z}}_t$ according to dependence strength.

Example 2. 8 monthly US Industrial Production indices in 1947-1993 published by the US Federal Reserve.

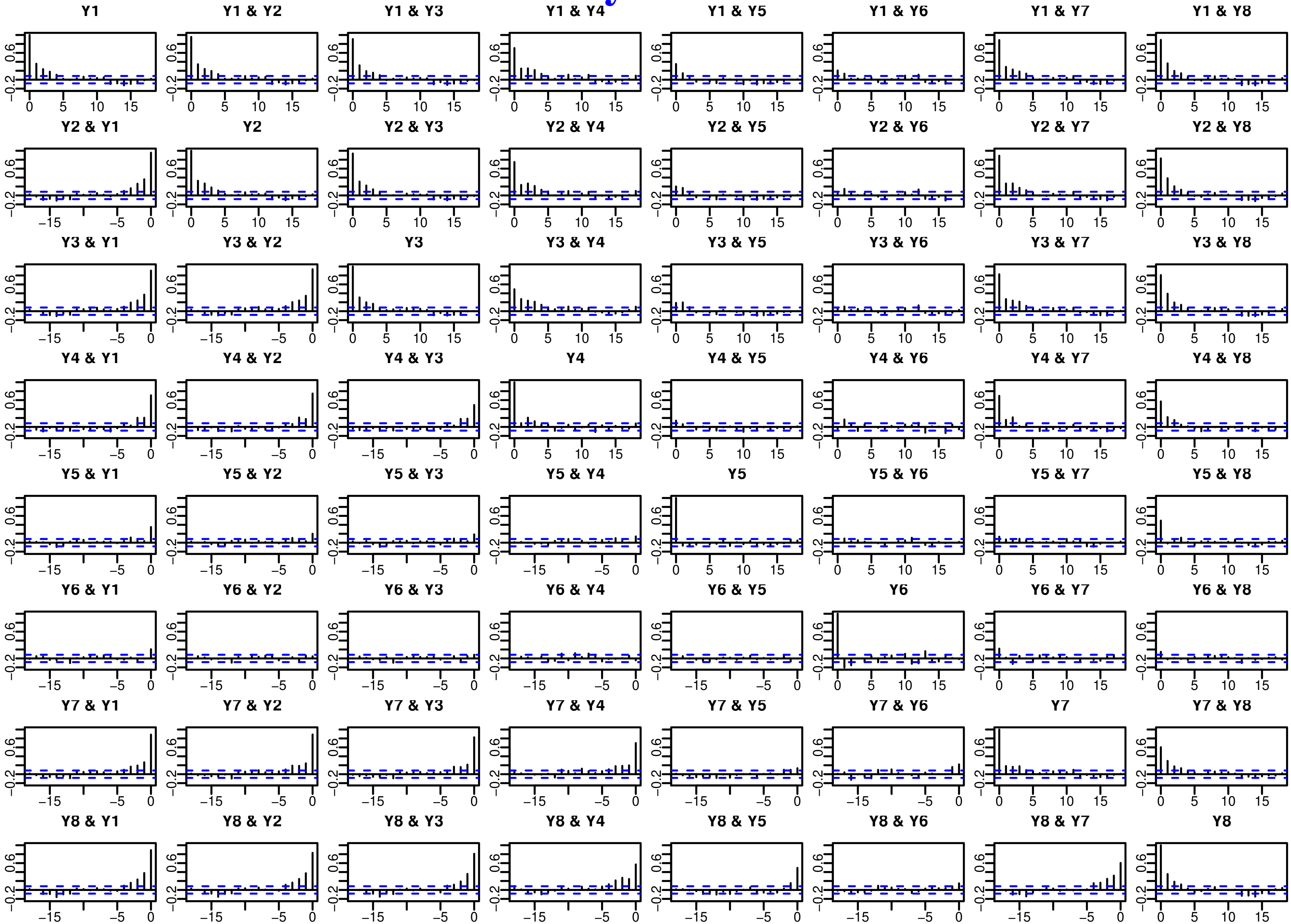
8 indices: *total index, manufacturing index, durable manufacturing, nondurable manufacturing, mining, utilities, products, materials*.

Nonstationary trends: difference each series

8 differenced monthly US Industrial Indices



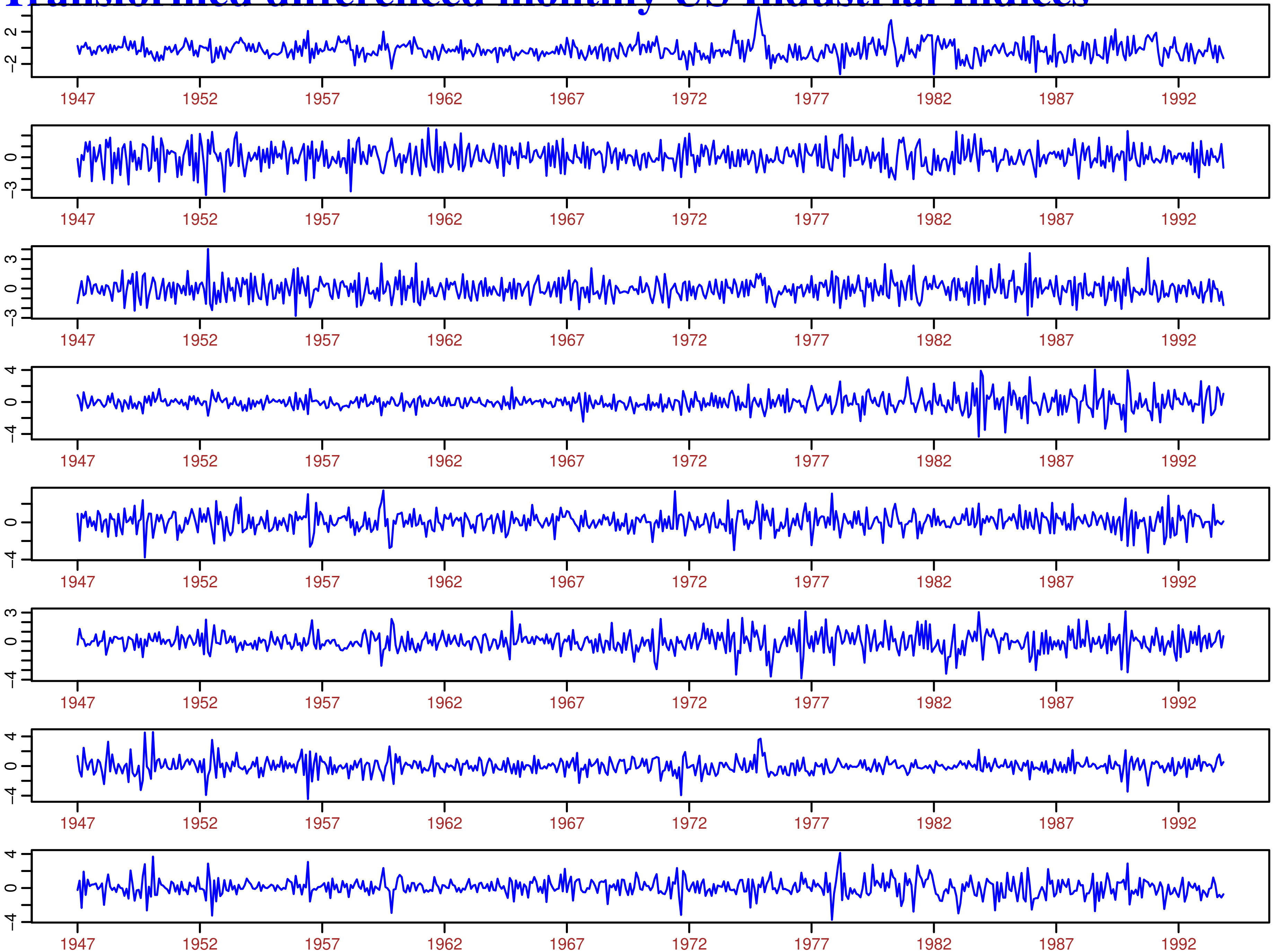
CCF of 8 differenced monthly US Industrial Indices



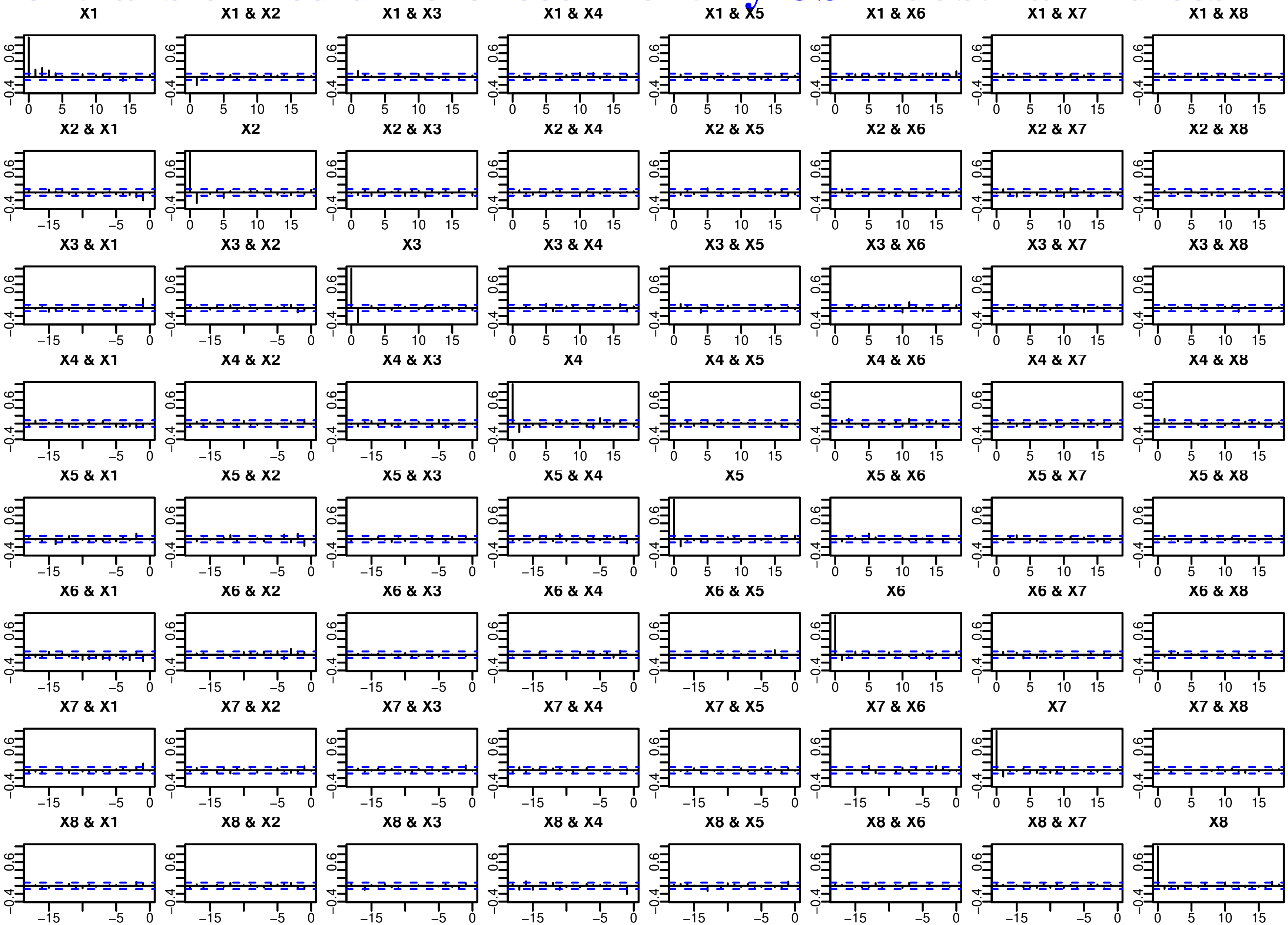
Transformation: $\widehat{\mathbf{x}}_t = \widehat{\mathbf{B}}\mathbf{y}_t$

$$\widehat{\mathbf{B}} = \begin{pmatrix} 5.012 & -1.154 & -0.472 & -0.880 & -0.082 & -0.247 & -2.69 & -1.463 \\ 10.391 & 8.022 & -3.981 & -3.142 & 0.186 & 0.019 & -6.949 & -4.203 \\ -6.247 & 11.879 & -4.8845 & -4.0436 & 0.289 & -0.011 & 2.557 & 0.243 \\ 1.162 & -6.219 & 3.163 & 1.725 & 0.074 & -0.823 & 0.646 & -0.010 \\ 6.172 & -4.116 & 2.958 & 1.887 & 0.010 & 0.111 & -2.542 & -3.961 \\ 0.868 & 1.023 & -2.946 & -4.615 & -0.271 & -0.354 & 3.972 & 1.902 \\ 3.455 & -2.744 & 5.557 & 3.165 & 0.753 & 0.725 & -2.331 & -1.777 \\ 0.902 & -2.933 & -1.750 & -0.123 & 0.191 & -0.265 & 3.759 & 0.987 \end{pmatrix}.$$

Transformed differenced monthly US Industrial Indices



CCF of transformed differenced monthly US Industrial Indices



Segmentation: $\{1, 2, 3\}$, $\{4, 8\}$, $\{5\}$, $\{6\}$, $\{7\}$

- Visual examining CCF: $\{1,2,3\},\{4,8\}$
- Permutation with max-CCF

  $1 \leq m \leq 20$: $\{1,3\}$
- Permutation with FDR

  $m = 20$ and $\beta \in [10^{-6},\ 0.01]$, or $m = 5$ and $\beta \in [10^{-6},\ 0.001]$:

  $\{1,3\}$

  $m = 5$ and $\beta = 0.005$: $\{1,2,3\},\{4,8\}$

  $m = 5$ and $\beta = 0.01$: $\{1,2,3,5,6,7\}$ and $\{4,8\}$.

  Two recommended groupings:

  seven groups: $\{1,3\}$

  five groups: $\{1,2,3\},\{4,8\}$

Post-sample forecast

Forecast 24 monthly indices in Jan 1992 – Dec 1993.

Using the segmentation: $\{1, 3\}$ and other six single element groups

Miss some small but significant cross correlations

	One-step MSE	Two-step MSE
VAR	$0.615_{(1.349)}$	$1.168_{(2.129)}$
RVAR	$0.606_{(1.293)}$	$1.159_{(2.285)}$
via TS-PCA	$0.588_{(1.341)}$	$1.154_{(2.312)}$

Example 3. Weekly notified measles cases in 7 cities in England

(i.e. London, Bristol, Liverpool, Manchester, Newcastle,

Birmingham and Sheffield) in 1948-1965, before the advent

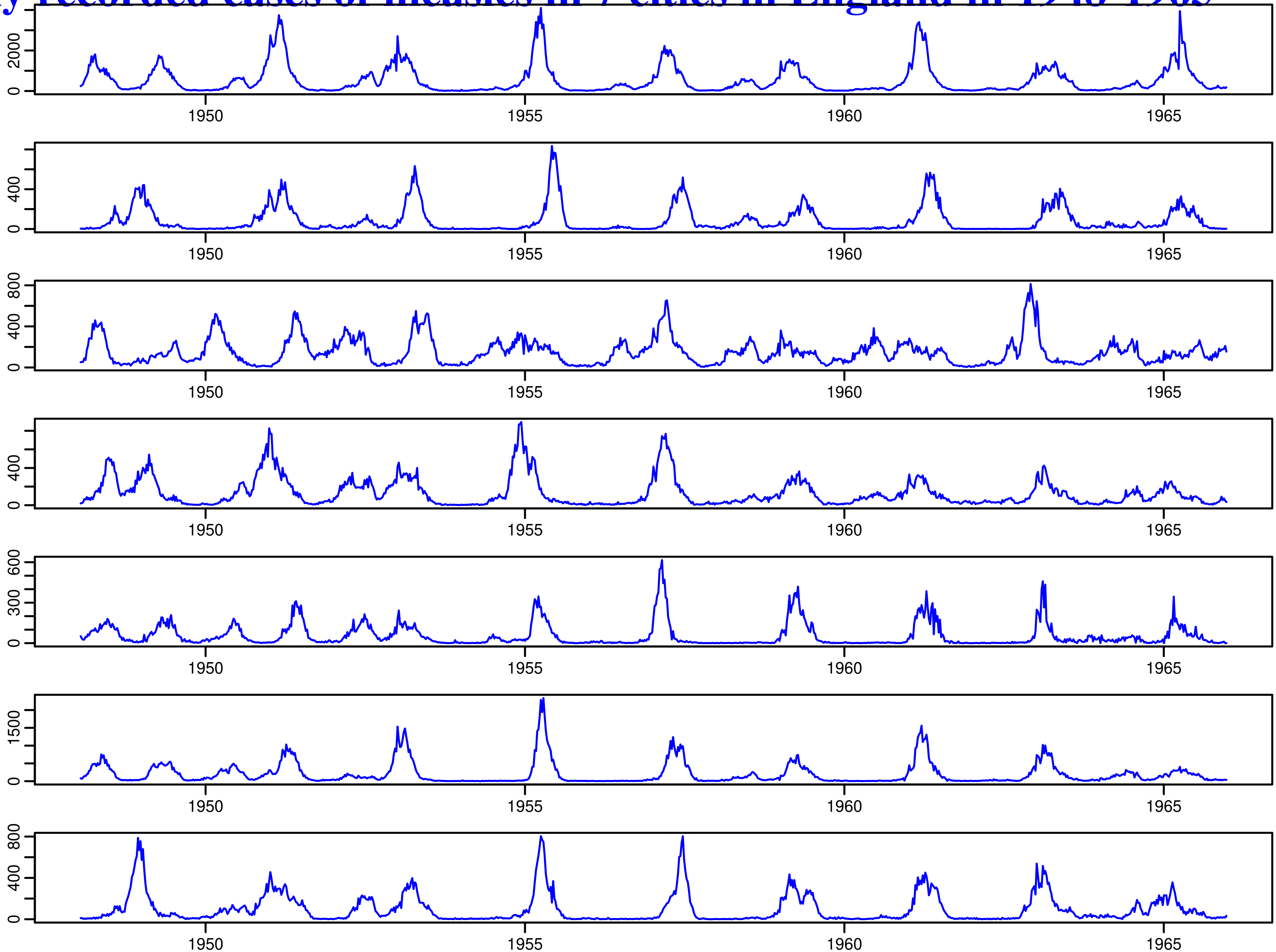
of vaccination.

All the 7 series show biennial cycles, which is the major driving

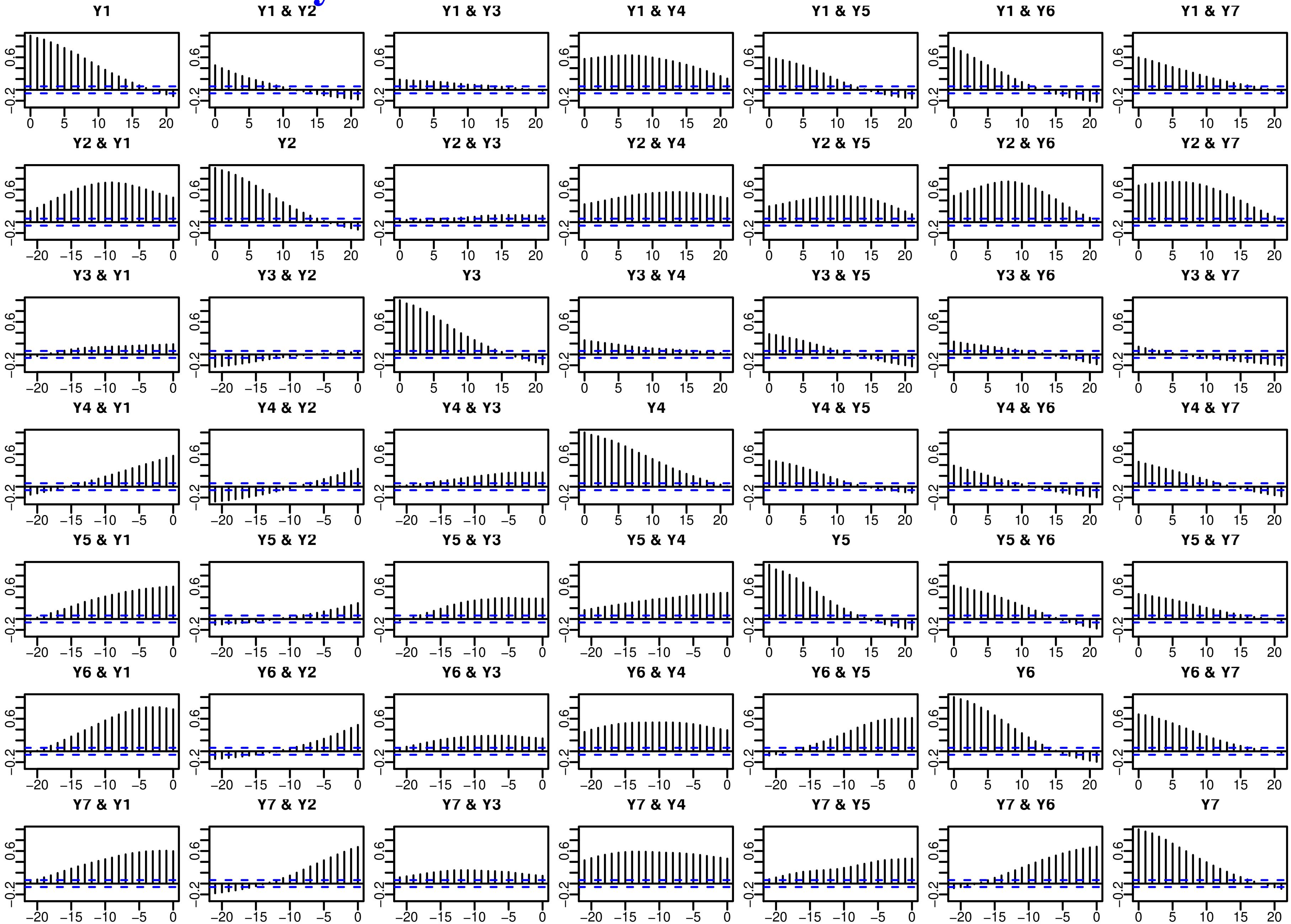
force for the cross correlations among different cities.

$n = 937$, $p = 7$.

eekly recorded cases of measles in 7 cities in England in 1948-1965



CCF of weekly recorded cases of measles in 7 cities

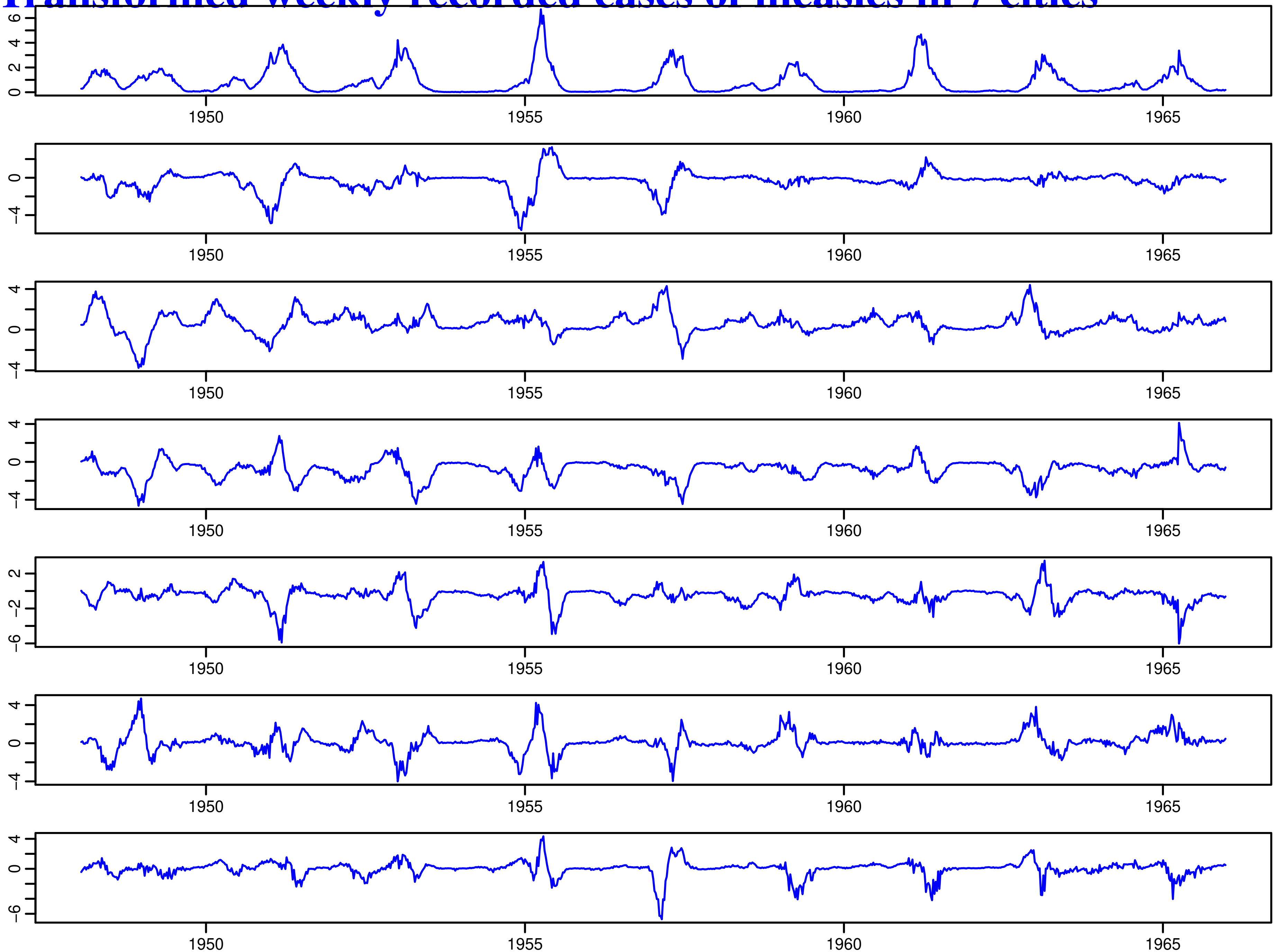


Transformation: $\widehat{\mathbf{x}}_t = \widehat{\mathbf{B}}\mathbf{y}_t$

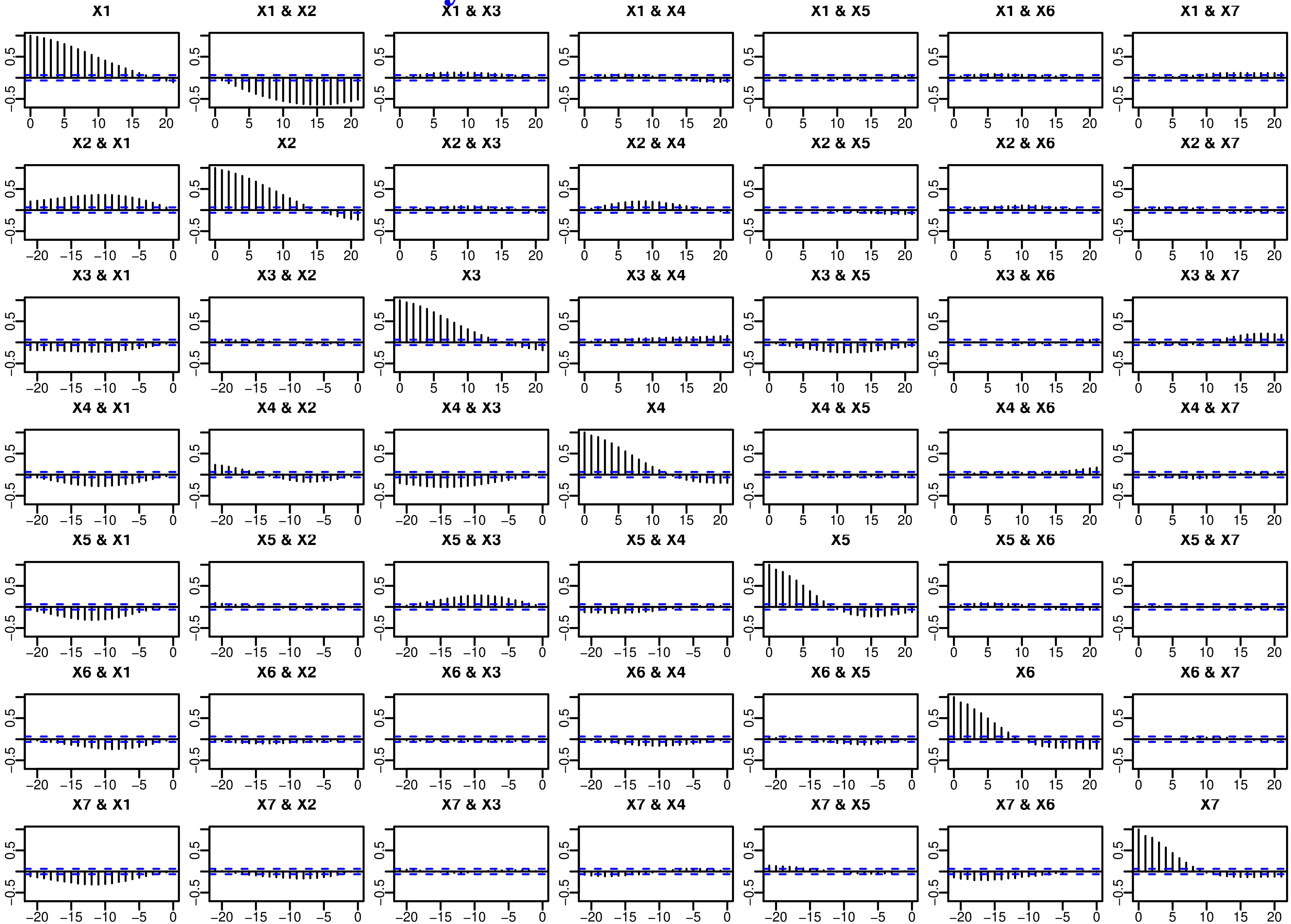
$$\widehat{\mathbf{B}} = \begin{pmatrix} -4.898e4 & 3.357e3 & -3.315e04 & -6.455e3 & 2.337e3 & 1.151e3 & -1.047e3 \\ 7.328e4 & 2.85e4 & -9.569e6 & -2.189e3 & 1.842e3 & 1.457e3 & 1.067e3 \\ -5.780e5 & 5.420e3 & -5.247e3 & 5.878e4 & -2.674e3 & -1.238e3 & 6.280e3 \\ -1.766e3 & 3.654e3 & 3.066e3 & 2.492e3 & 2.780e3 & 8.571e4 & 2.356e3 \\ -1.466e3 & -7.337e4 & -5.896e3 & 3.663e3 & 6.633e3 & 3.472e3 & -4.668e3 \\ -2.981e4 & -8.716e4 & 6.393e8 & -2.327e3 & 5.365e3 & -9.475e4 & 8.629e3 \\ -7.620e4 & -3.338e3 & 1.471e3 & 2.099e3 & -1.318e2 & 4.259e3 & 6.581e4 \end{pmatrix}$$

Code: $aek \;=\; a \times 10^{-k}$

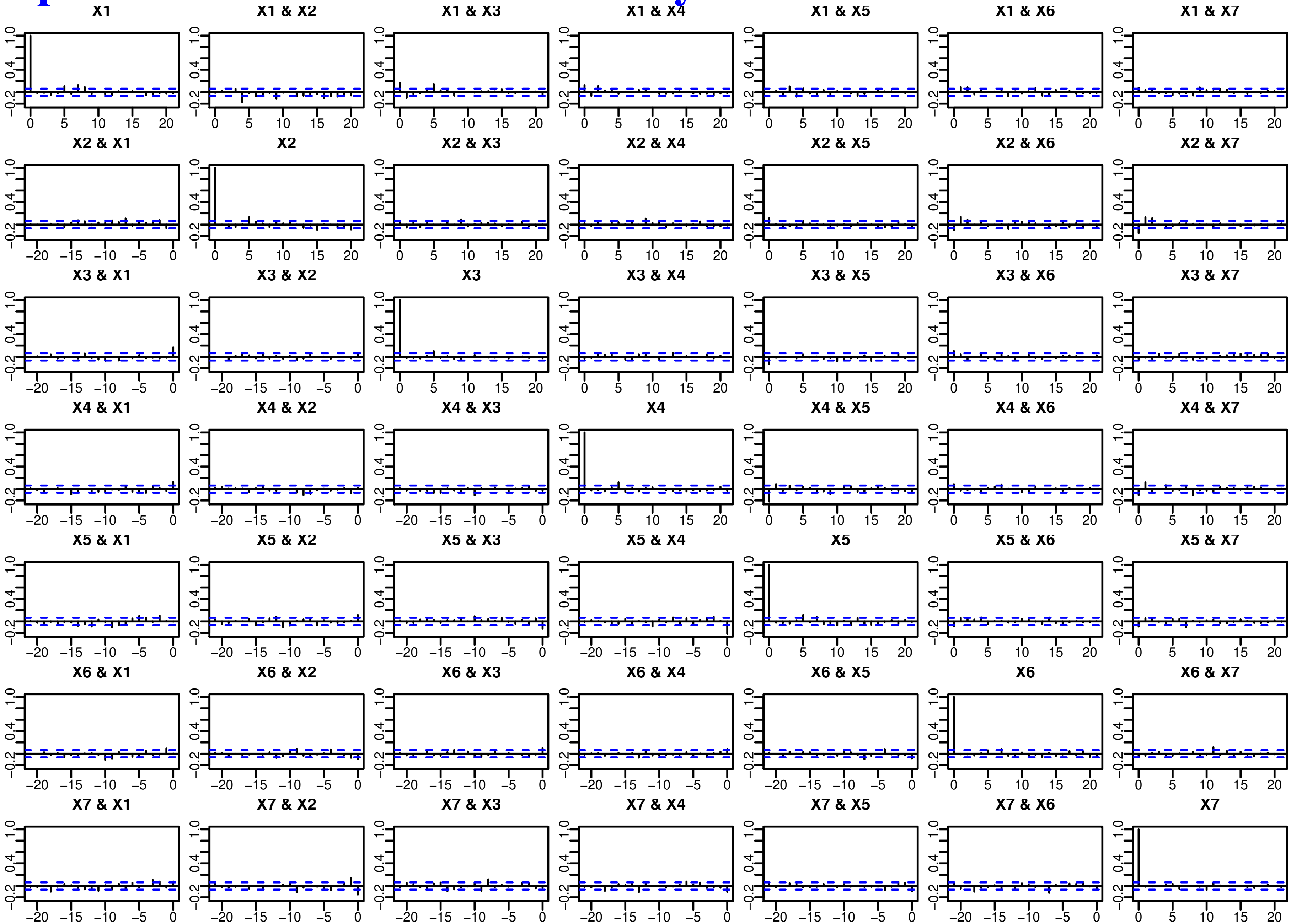
Transformed weekly recorded cases of measles in 7 cities



CCF of transformed weekly recorded cases of measles in 7 cities

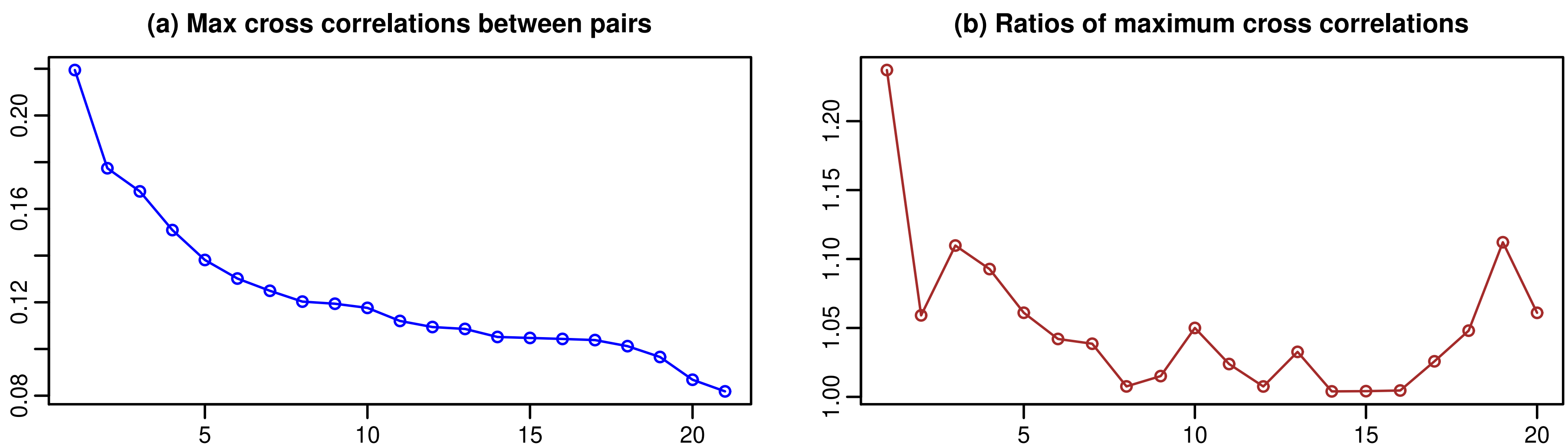


CF of prewhitened transformed weekly recorded cases of measles



Note. When none of the component series are WN, the confidence bounds $\pm 1.96/\sqrt{n} = .064$ could be misleading.

Segmentation assumption is invalid for this example!



(a) The maximum cross correlations, plotted in descending order, among each of the $\binom{7}{2} = 21$ pairs component series of the transformed and prewhitened measles series. The maximization was taken over the lags between -20 to 20. (b) The ratios of two successive correlations in (a).

The segmentations determined by different numbers of connected pairs for the transformed measles series from 7 cities in England.

No. of connected pairs	No. of groups	Segmentation
1	6	$\{4, 5\}, \{1\}, \{2\}, \{3\}, \{6\}, \{7\}$
2	5	$\{1, 2\}, \{4, 5\}, \{3\}, \{6\}, \{7\}$
3	4	$\{1, 2, 3\}, \{4, 5\}, \{6\}, \{7\}$
4	3	$\{1, 2, 3, 7\}, \{4, 5\}, \{6\}$
5	2	$\{1, 2, 3, 6, 7\}, \{4, 5\}$
6	1	$\{1, \cdots, 7\}$

Post-sample forecasting

Forecast the notified measles cases in the last 14 weeks of the period for all 7 cities

Using the segmentation with four groups: {1, 2, 3}, {4, 5}, {6} and {7}

	One-step MSE	Two-step MSE
Univariate AR	551.386(1322.345)	931.766(3115.508)
VAR	503.408(1124.213)	719.499(2249.986)
RVAR	574.582(1432.217)	846.141(2462.019)
via TS-PCA (4 groups)	472.106(1088.170)	654.843(1807.502)
via TS-PCA (7 groups)	510.825(1134.615)	798.374(2451.896)
via TS-PCA (3 groups)	497.025(1206.745)	677.125(1902.230)

Example 4. Daily sales of a clothing brand in 25 provinces in

China in 1 January 2008 – 16 December 2012.

$n = 1812,\ p = 25$

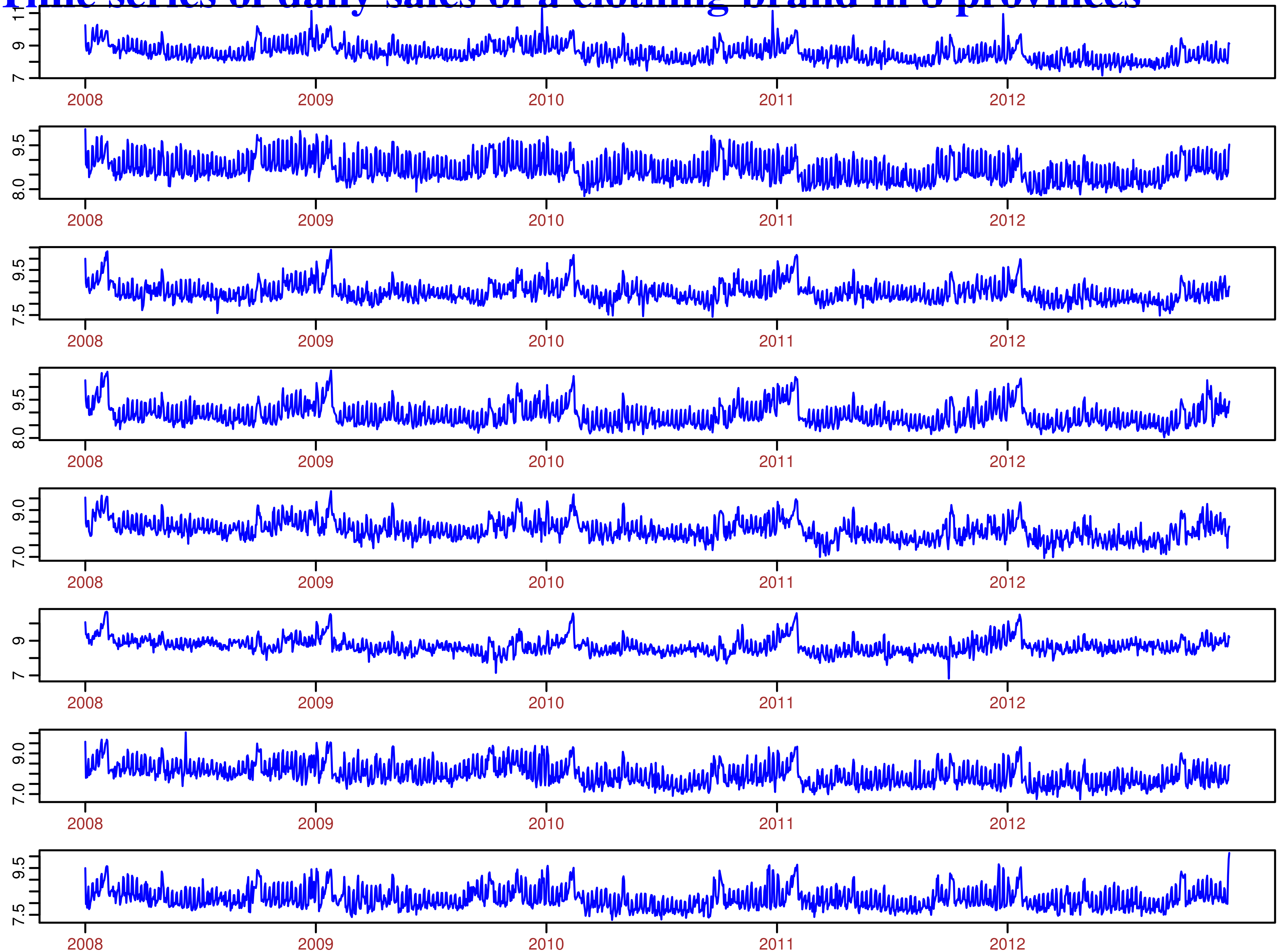
Annual pattern: peak in February

Strong periodicity component with the period 7.

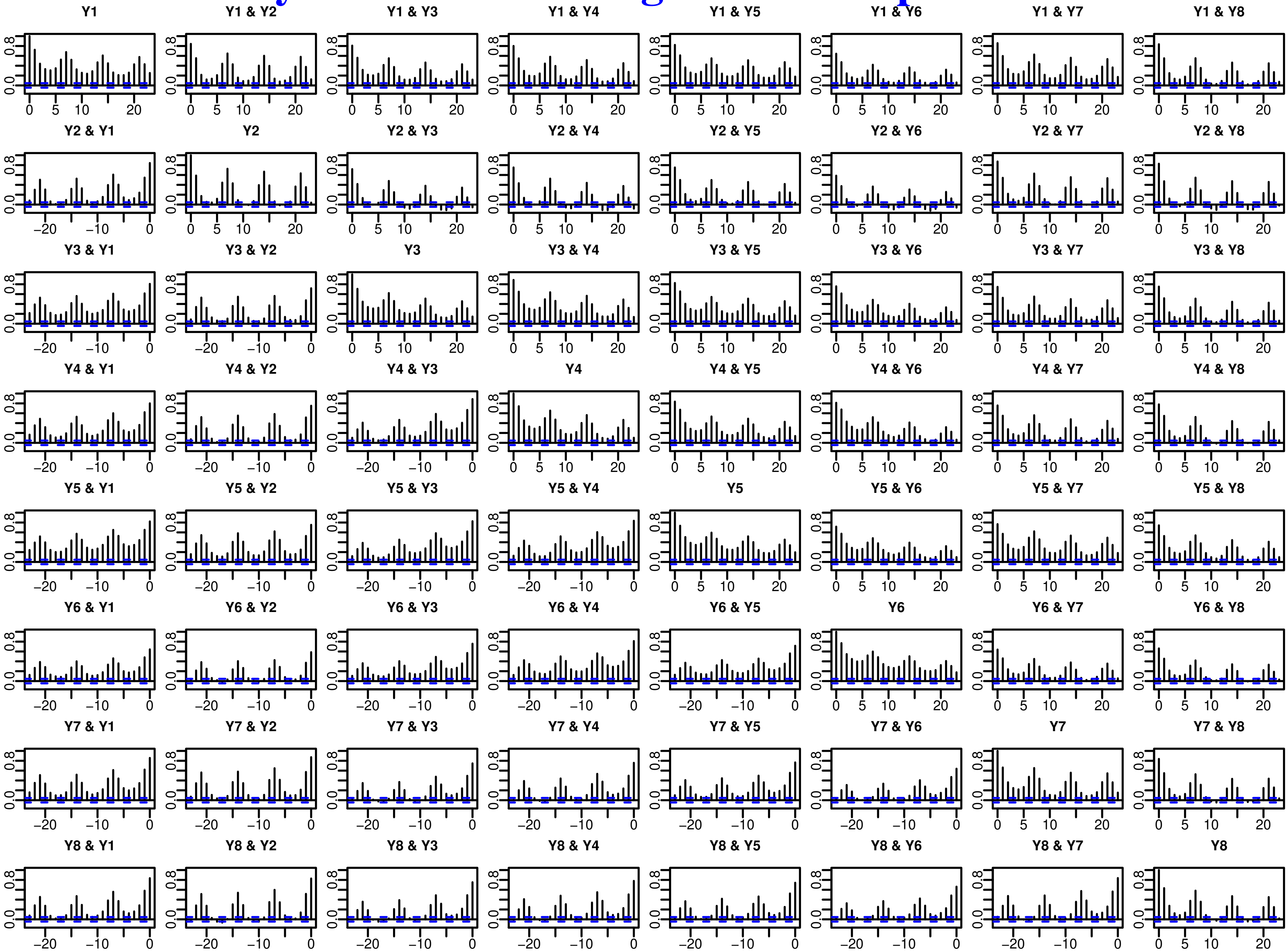
The 25 transformed series are segmented into 24 groups with $\{15, 16\}$ as one group.

Permutation is performed using the max-CCF with $14 \le m \le 30$.

Time series of daily sales of a clothing brand in 8 provinces

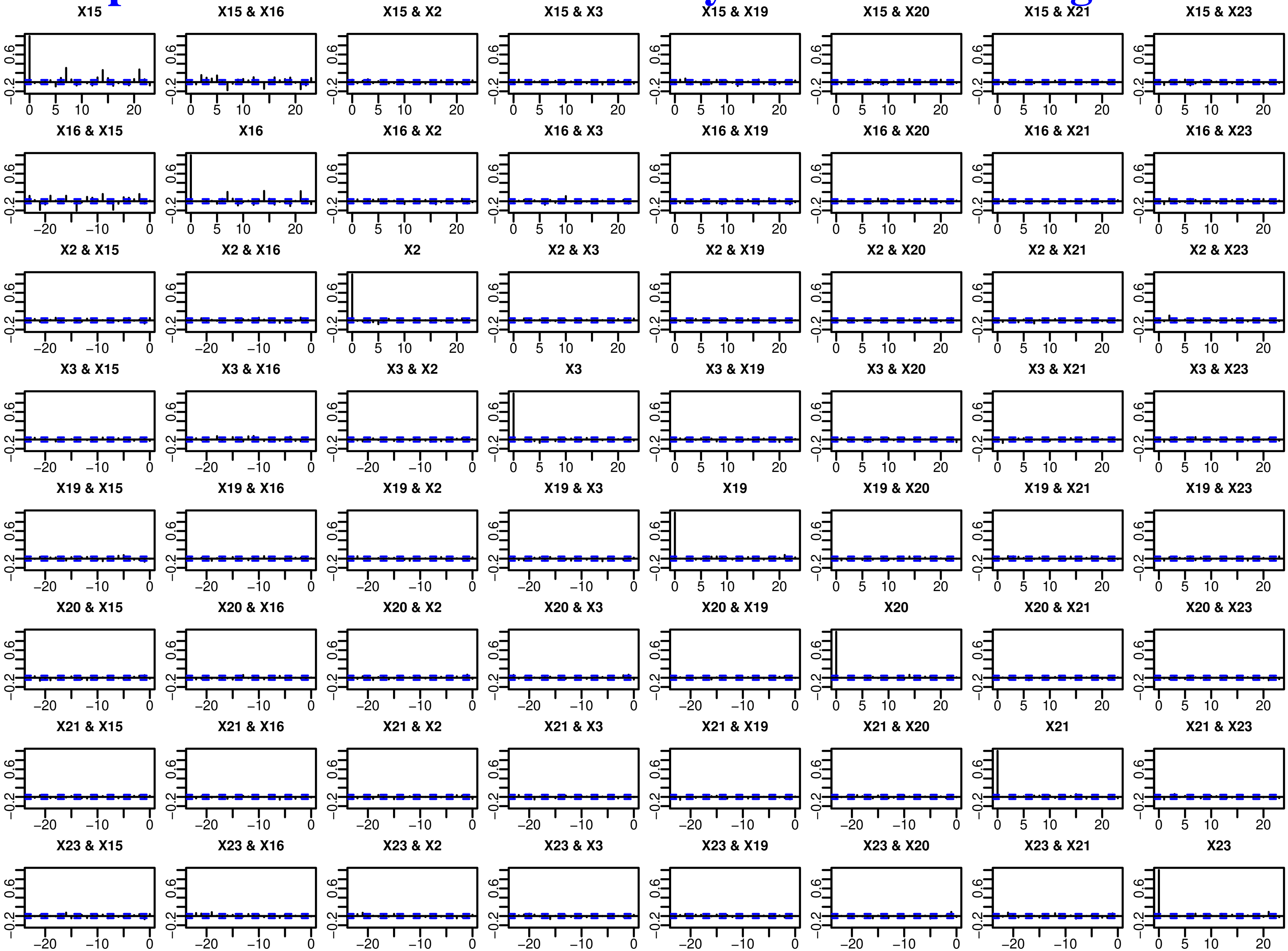


CCF of daily sales of a clothing brand in 8 provinces



CCF of 8 transformed daily sales of a clothing brand

CF of 8 prewhitened transformed daily sales of a clothing brand



Post-sample forecasting

Forecast the daily sales for the last two weeks in each of the 25 provinces.

Segmentation: 24 groups with $\{15, 16\}$ as one group

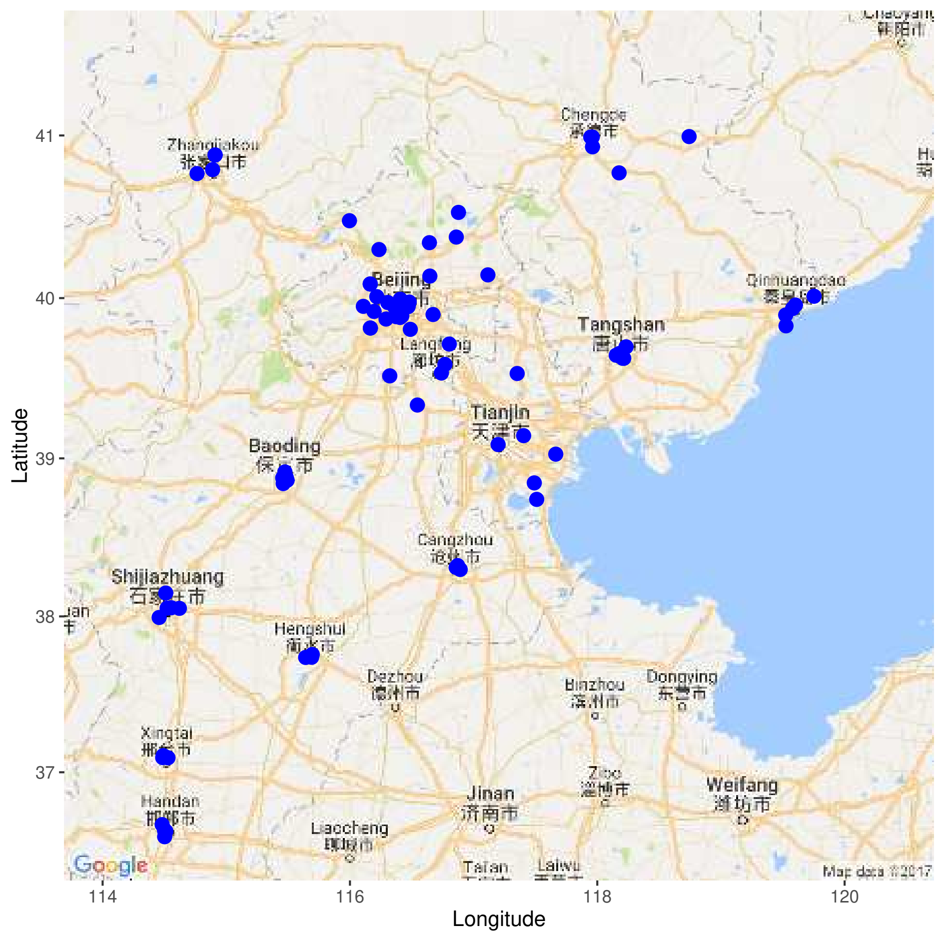
	One-step MSE	Two-step MSE
Univariate AR	0.208 (0.551)	0.194 (0.539)
VAR	0.295 (0.806)	0.301 (0.855)
RVAR	0.293 (0.820)	0.296 (0.863)
via TS-PCA (24 groups)	0.153 (0.134)	0.163 (0.124)
via TS-PCA (25 groups)	0.110 (0.084)	0.132 (0.091)
via TS-PCA (23 groups)	0.151 (0.133)	0.159 (0.121)

Cross correlations between provinces are useful information.

However they cannot be used via direct VAR!

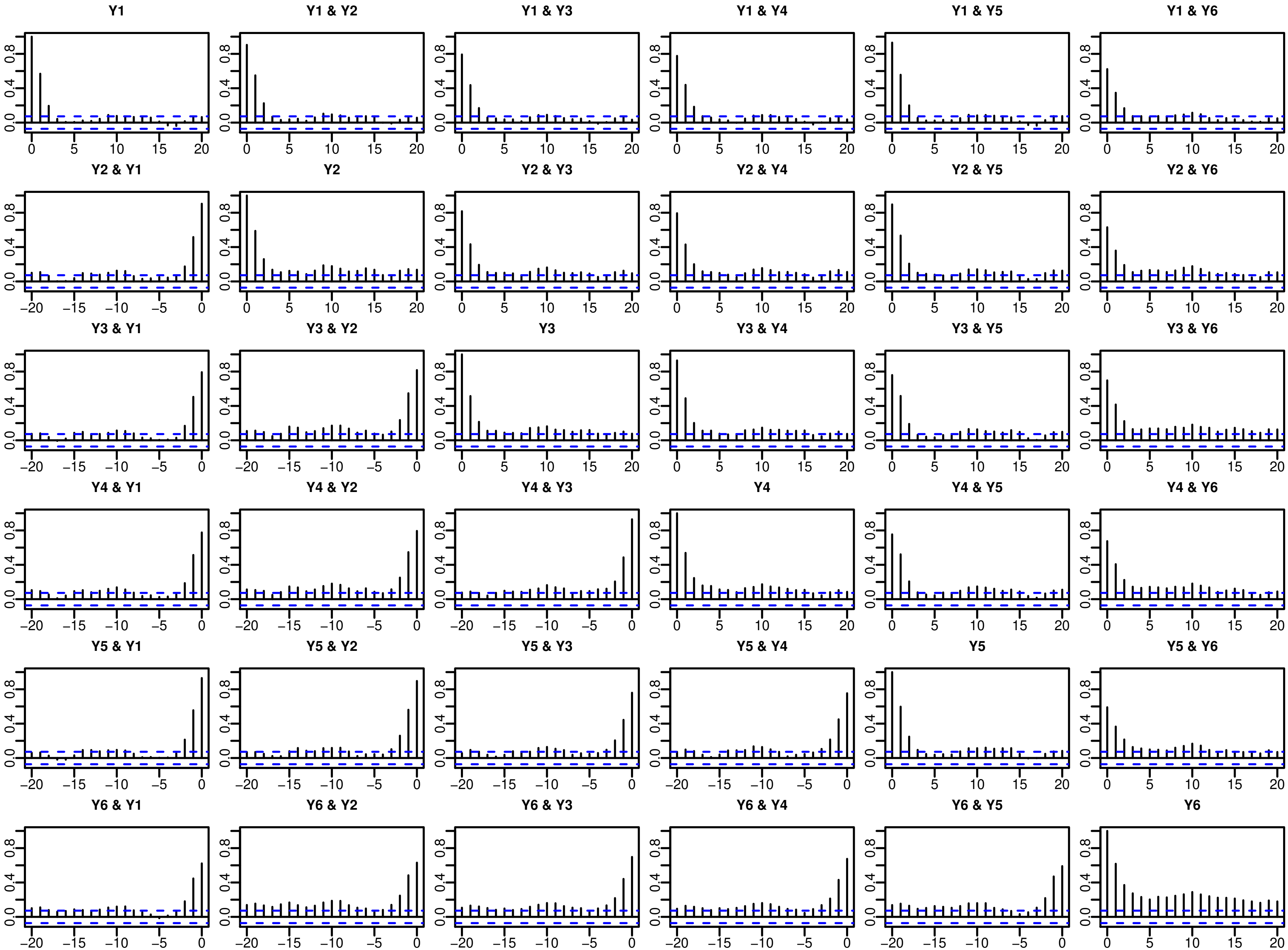
Example 5. Log daily $PM_{2.5}$ concentration readings at 84 monitoring stations in Beijing, Tianjin and Hebei in 1 Jan 2015 – 31 Dec 2016.

$PM_{2.5}$ consists of airborne particles with aerodynamic diameters smaller than 2.5μm.



$n = 731$ and $p = 84$

ross correlogram of log $PM_{2.5}$ from six randomly selected stations



The maximum cross correlation method in divides the 84 transformed time series into 83 groups

	One-step MSE	Two-step MSE
Univariate AR	0.525(0.204)	0.835(0.284)
via TS-PCA (83 groups)	0.485(0.185)	0.662(0.224)
via TS-PCA (84 groups)	0.484(0.184)	0.662(0.224)
via TS-PCA (50 groups)	0.492(0.187)	0.678(0.228)
via TS-PCA (70 groups)	0.474(0.180)	0.664(0.225)

Max group size is 8 for the segmentation with 50 groups, and is 4 for the segmentation with 70 groups.

Why is always better?

- $\mathbf{W}_y = \sum_k \mathbf{\Sigma}_y(k)\mathbf{\Sigma}_y(k)' = \mathbf{A}\mathbf{W}_x\mathbf{A}'$,

$$\mathrm{tr}(\mathbf{W}_y) = \sum_k \sum_{i,j=1}^{p} \rho_{ij}^{y}(k)^2 = \mathrm{tr}(\mathbf{W}_x) = \sum_k \sum_{i,j=1}^{p} \rho_{ij}^{x}(k)^2 = \sum_k \sum_{i=1}^{p} \rho_{ii}^{x}(k)^2$$

provided all components of $\mathbf{x}_t$ are uncorrelated across all time lags

Components of $\mathbf{x}_t$ are more predictable, due to stronger ACF!

- When TS-PCA does not exist, provide effective approximations by ignoring negligible, though significant, correlations

- When latent segmentation does not exist, use $\mathbf{z}_t = \boldsymbol{\Gamma}_y' \mathbf{y}_t$,

$$\mathbf{W}_y = \sum_k \boldsymbol{\Sigma}_y(k) \boldsymbol{\Sigma}_y(k)' = \boldsymbol{\Gamma}_y \mathbf{D} \boldsymbol{\Gamma}_y'.$$

Hence $\boldsymbol{\Sigma}_z(k) = \boldsymbol{\Gamma}_y' \boldsymbol{\Sigma}_y(k) \boldsymbol{\Gamma}_y$, and therefore

$$\mathbf{W}_z = \sum_k \boldsymbol{\Sigma}_z(k) \boldsymbol{\Sigma}_z(k)' = \boldsymbol{\Gamma}_y' \mathbf{W}_y \boldsymbol{\Gamma}_y = \mathbf{D},$$

i.e. $\mathbf{W}_z$ is a diagonal matrix.

Note. $\boldsymbol{\Sigma}_z(k)$ are unlikely to be diagonal, though the off-diagonal elements tend to be small.

- There are several tuning parameters in determining groupings by either max correlation method or multiple FDR.

  However both those methods rank all the pair components of the transformed serious according to their dependence strength.

  Different tuning parameters effectively determine how many those small (maybe still significant) correlations are ignored, which has little impact in practice.

- `R` package `PCA4TS` available from *CRAN*

Chang, J., Guo, B. and Yao, Q. (2018). Principal component analysis for second-order stationary vector time series.

Annals of Statistics, **46**, 2094-2124.

Segmenting multiple volatility processes

Let $\mathcal{F}_t = \sigma(\mathbf{y}_t, \mathbf{y}_{t-1}, \cdots)$,

$$E(\mathbf{y}_t|\mathcal{F}_{t-1}) = \mathbf{0}, \qquad \mathrm{Var}(\mathbf{y}_t|\mathcal{F}_{t-1}) = \mathbf{\Sigma}_y(t).$$

Assumption: $\mathbf{y}_t = \mathbf{A}\mathbf{x}_t$,
$\mathrm{Var}(\mathbf{x}_t|\mathcal{F}_{t-1}) = \mathrm{diag}\big(\mathbf{\Sigma}_1(t), \cdots, \mathbf{\Sigma}_q(t)\big)$.

Let $\mathrm{Var}(\mathbf{y}_t) = \mathrm{Var}(\mathbf{x}_t) = \mathbf{I}_p$, then $\mathbf{A}$ is orthogonal.

Let $\mathcal{B}_{t-1}$ be a π-class and $\sigma(\mathcal{B}_{t-1}) = \mathcal{F}_{t-1}$. put

$$\mathbf{W}_y = \sum_{B\in\mathcal{B}_{t-1}} [E\{\mathbf{y}_t\mathbf{y}_t' I(B)\}]^2, \quad \mathbf{W}_x = \sum_{B\in\mathcal{B}_{t-1}} [E\{\mathbf{x}_t\mathbf{x}_t' I(B)\}]^2.$$

For any $B \in \mathcal{B}_{t-1}$,

$$E\{\mathbf{x}_t\mathbf{x}_t' I(B)\} = E[I(B)E\{\mathbf{x}_t\mathbf{x}_t'|\mathcal{F}_{t-1}\}] = E[I(B)\mathrm{diag}(\mathbf{\Sigma}_1(t), \cdots, \mathbf{\Sigma}_q(t))]$$

is a block diagonal matrix, so is $\mathbf{W}_x$.

Since

$$\mathbf{W}_y = \mathbf{A}\mathbf{W}_x\mathbf{A}',$$

the proposed method continues to apply.

In practice we estimate $\mathbf{W}_y$ by

$$\widehat{\mathbf{W}}_y = \sum_{B \in \mathcal{B}} \sum_{k=1}^{k_0} \Big(\frac{1}{n-k} \sum_{t=k+1}^{n} \mathbf{y}_t \mathbf{y}_t' \, I(\mathbf{y}_{t-k} \in B) \Big)^2,$$

where $\mathcal{B}$ may consist of $\{\mathbf{u} \in R^p : \|\mathbf{u}\| \le \|\mathbf{y}_t\|\}$ for $t = 1, \cdots, n$.

Example 8. Consider the daily returns in 2 Jan 2002 – 10 July 2008 of six stocks: *Bank of America, Dell, JPMorgan, FedEx, McDonald and American International Group*.

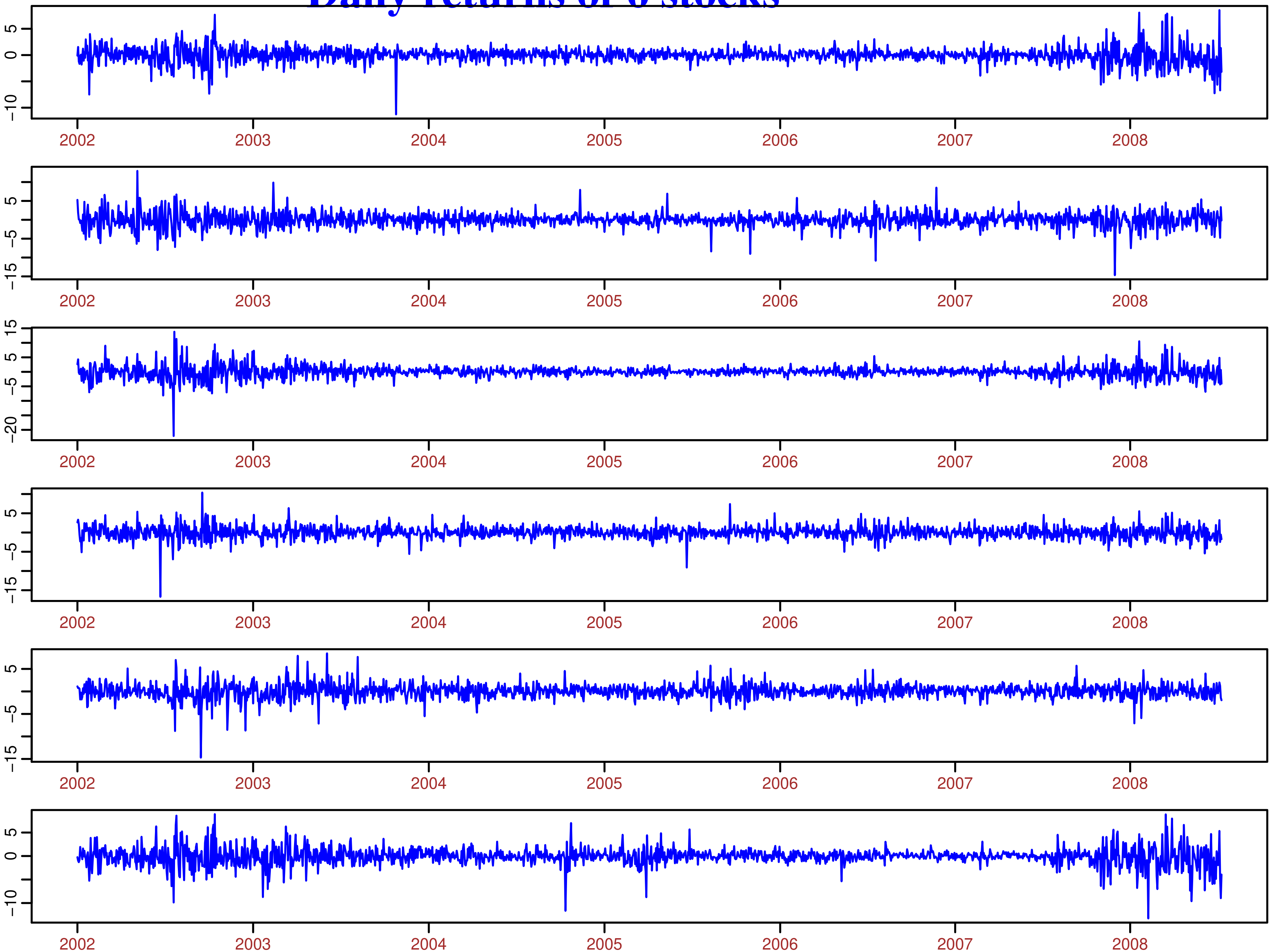
$n = 1642$, $p = 6$

$$\widehat{\mathbf{B}} = \begin{pmatrix} -0.227 & -0.093 & 0.031 & 0.550 & 0.348 & -0.041 \\ -0.203 & -0.562 & 0.201 & 0.073 & -0.059 & 0.158 \\ 0.022 & 0.054 & -0.068 & 0.436 & -0.549 & 0.005 \\ -0.583 & 0.096 & -0.129 & -0.068 & -0.012 & 0.668 \\ 0.804 & -0.099 & -0.409 & -0.033 & 0.008 & 0.233 \\ 0.144 & -0.012 & -0.582 & 0.131 & 0.098 & -0.028 \end{pmatrix}$$

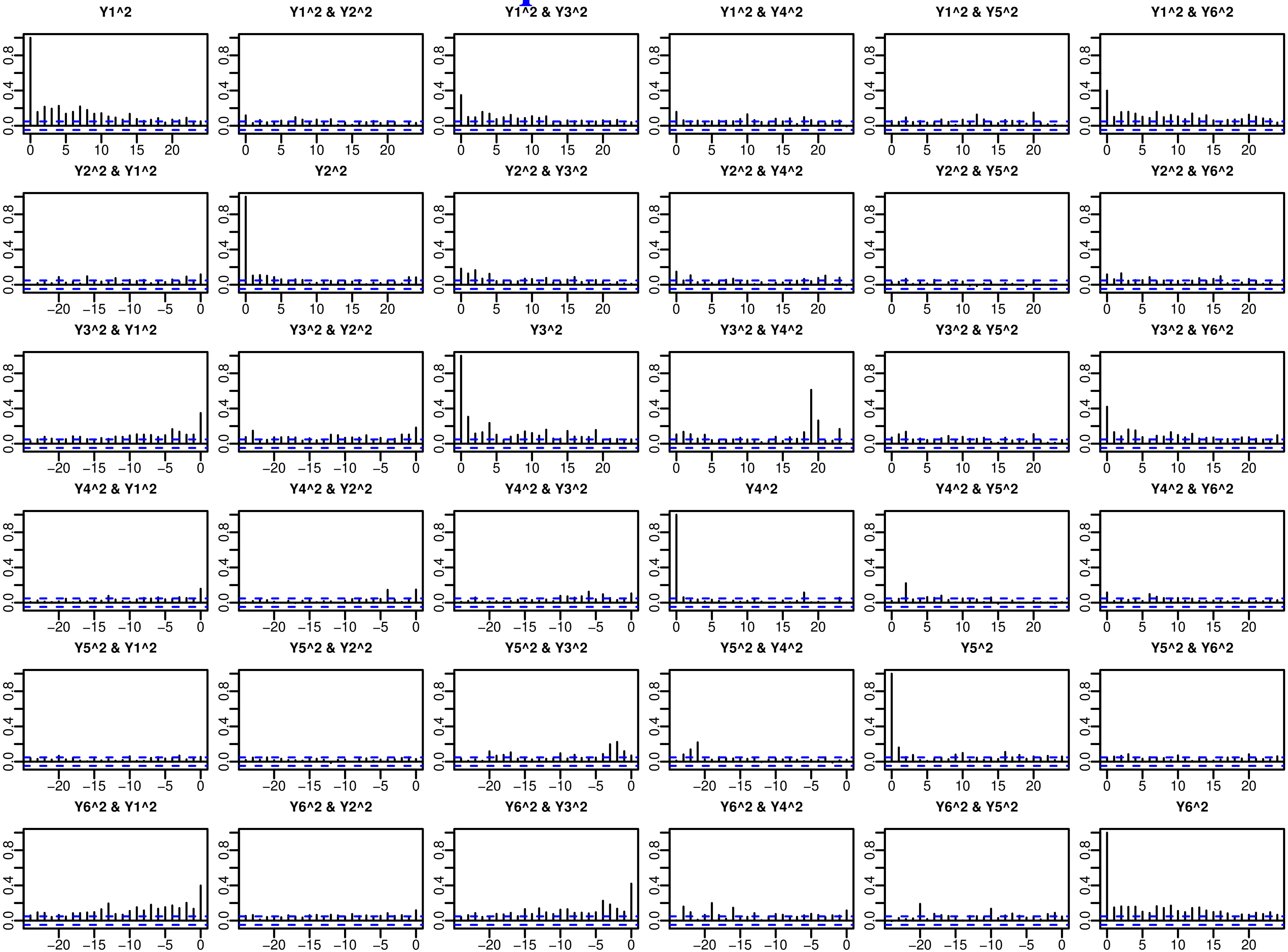
Segmentation for transformed series:

CUC of Fan, Wang & Y (2008)

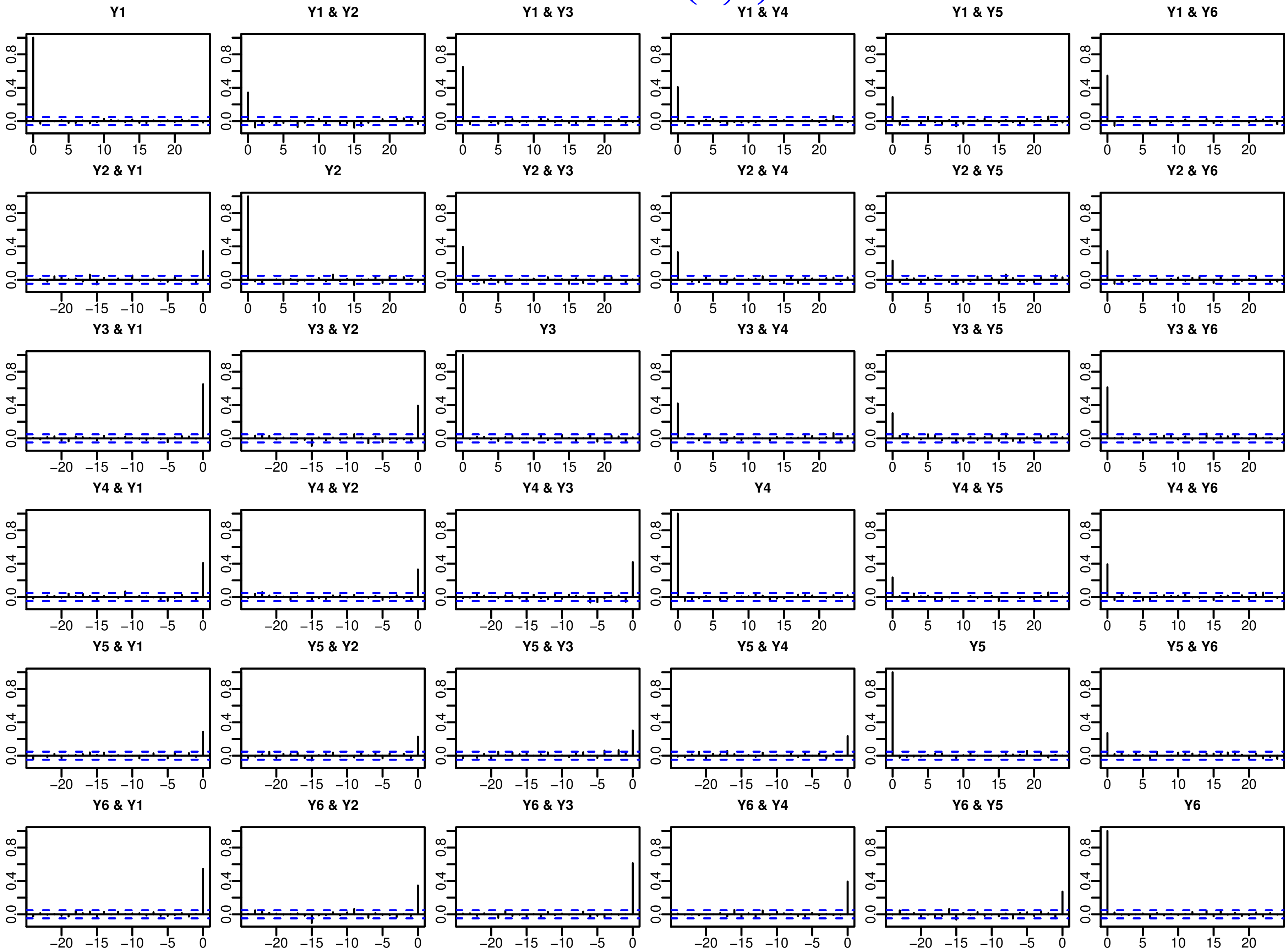
Daily returns of 6 stocks



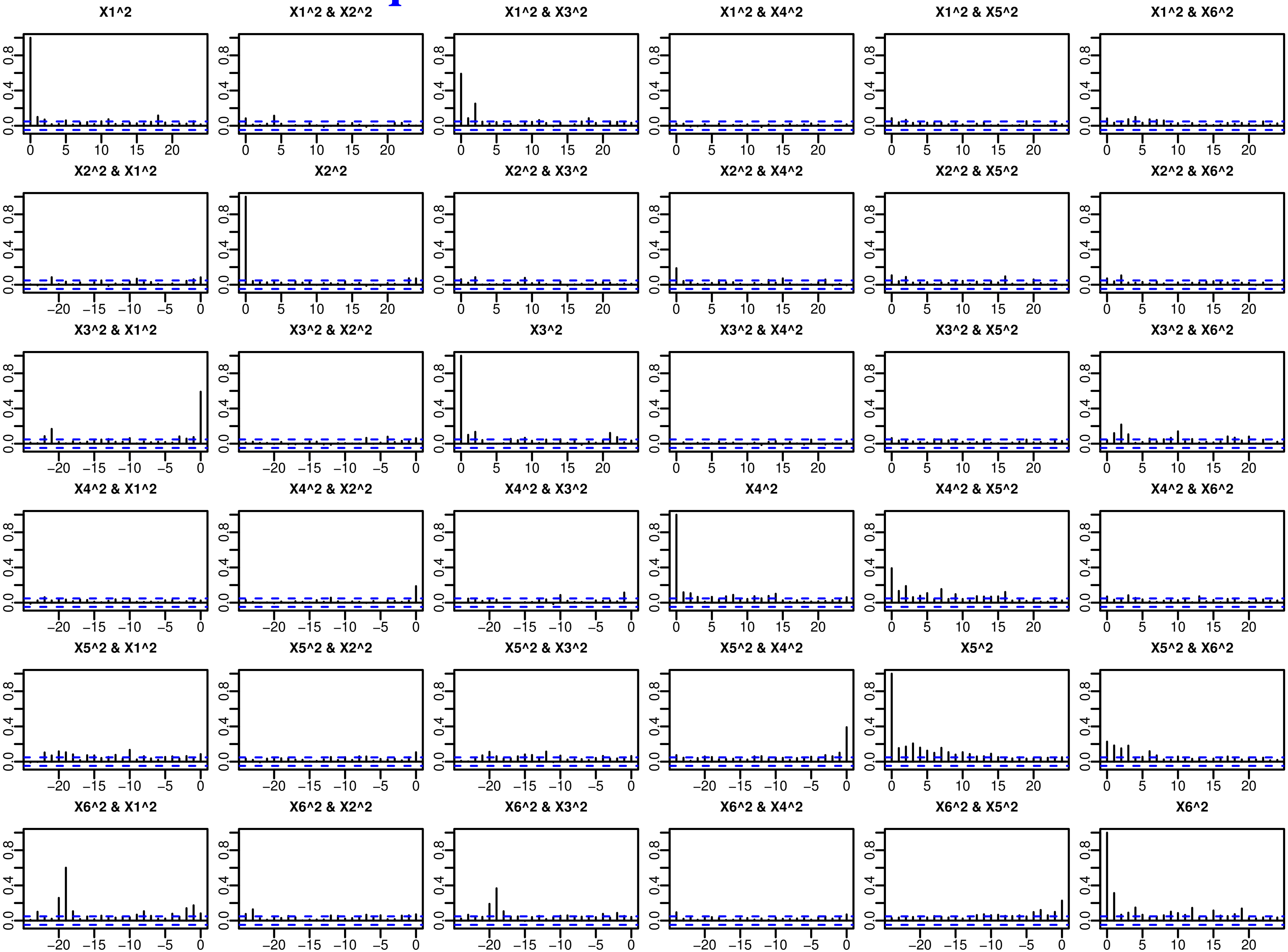
CCF of squared returns



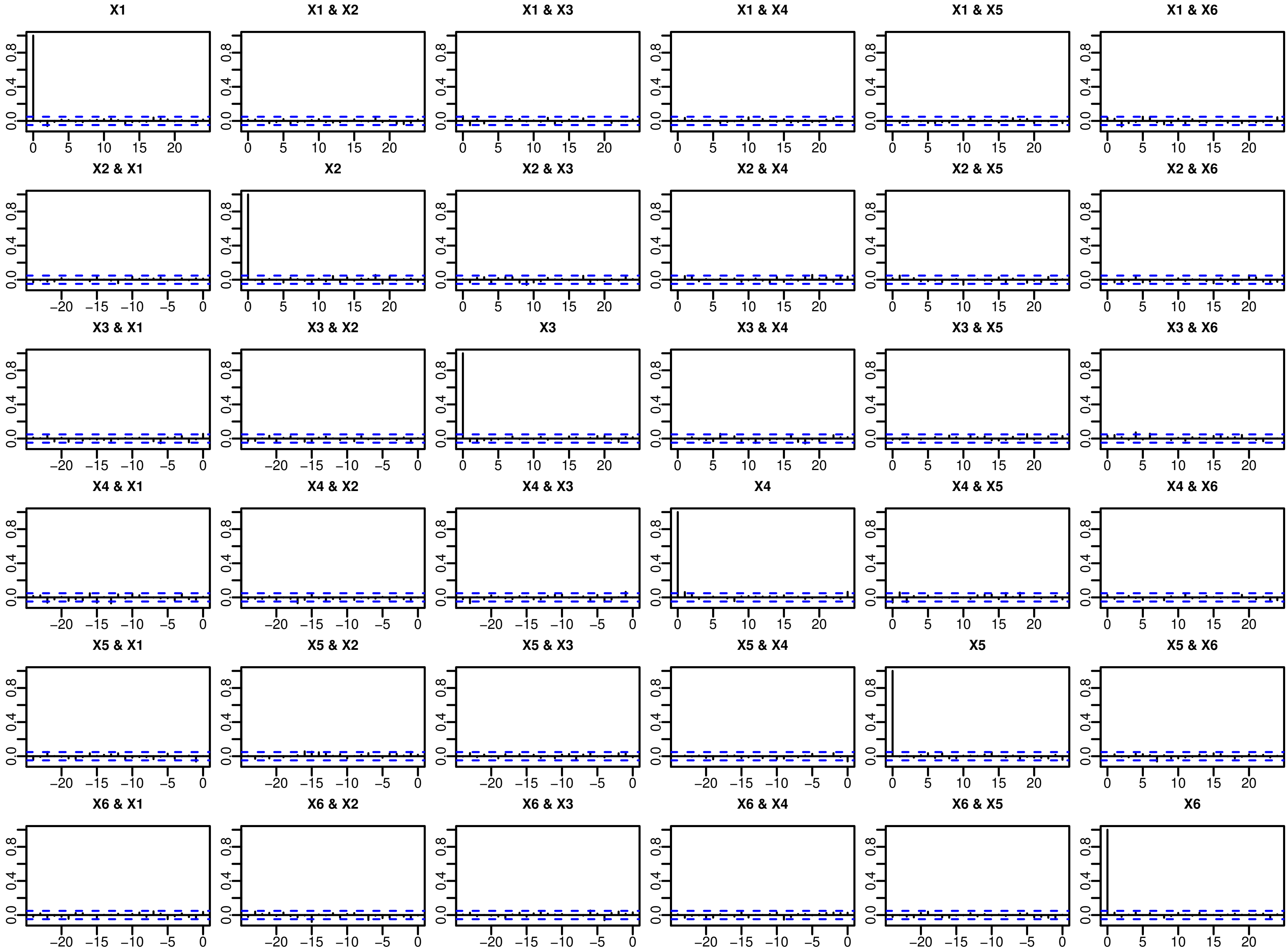
CCF of residuals from fitted GARCH(1,1) for each return series



CCF of squared transformed returns



CCF of residuals from transformed return series



A bundle of curve time series

Consider p curve time series

$$\mathbf{\mathcal{Y}}^t(\mathbf{u}) = \{\mathcal{Y}_1^t(u), \cdots, \mathcal{Y}_p^t(u)\}', \quad u \in \mathcal{I}, \ \ t = 0, \pm 1, \pm 2, \cdots .$$

Seek for a $p \times p$ orthogonal matrix, such that

$$\mathbf{\mathcal{X}}^t(u) \equiv \{\mathcal{X}_1^t(u), \ldots, \mathcal{X}_p^t(u)\}' = \mathbf{A}'\{\mathcal{Y}_1^t(u), \ldots, \mathcal{Y}_p^t(u)\}' = \mathbf{A}'\mathbf{\mathcal{Y}}^t(u),$$

$\mathcal{X}_1^t(\cdot), \ldots, \mathcal{X}_p^t(\cdot)$ are p uncorrlated curve time series across all time lags.

Then $p \times p$ matrix $\mathbf{\Sigma}_{x,k}(u_1, u_2)$ is a diagonal matrix for any $u_1, u_2 \in \mathcal{I}$, where

$$\begin{aligned}\mathbf{\Sigma}_{x,k}(u_1, u_2) &\equiv \mathrm{Cov}\{\mathbf{\mathcal{X}}^t(u_1), \mathbf{\mathcal{X}}^{t-k}(u_2)\} \\ &= \mathbf{A}'\mathrm{Cov}\{\mathbf{\mathcal{Y}}^t(u_1), \mathbf{\mathcal{Y}}^{t-k}(u_2)\}\mathbf{A} \equiv \mathbf{A}'\mathbf{\Sigma}_{y,k}(u_1, u_2)\mathbf{A}.\end{aligned}$$

Let

$$\mathbf{W}_z = \sum_{k=0}^{k_0} \int_{\mathcal{I}^2} \mathbf{\Sigma}_{z,k}(u_1, u_2)\{\mathbf{\Sigma}_{z,k}(u_1, u_2)\}' \, \mathrm{d}u_1 \mathrm{d}u_2,$$

$$\mathbf{W}_y = \sum_{k=0}^{k_0} \int_{\mathcal{I}^2} \mathbf{\Sigma}_{y,k}(u_1, u_2)\{\mathbf{\Sigma}_{y,k}(u_1, u_2)\}' \, \mathrm{d}u_1 \mathrm{d}u_2.$$

Then $\mathbf{W}_x = \mathbf{A}'\mathbf{W}_y\mathbf{A}$ is a diagonal matrix.

Replace $\mathbf{W}_y$ by its estimator, $\mathbf{A}$ can then be estimated by an eigen-analysis $\ldots$.