Segmenting Multiple Time Series by Contemporaneous Linear Transformation: PCA for time series

Qiwei Yao

Department of Statistics, London School of Economics

q.yao@lse.ac.uk

Joint work with

Jinyuan Chang   University of Melbourne

Bin Guo   Peking University

- 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
- Asymptotic properties in 3 settings:

  p fixed, $\quad p = o(n^c)$, $\quad \log p = o(n^c)$

- Simulation
- Segmenting multiple volatility processes

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 modelled 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 modelled separately!

- realistic?
- how to find B and x_t?

RealData

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

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

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

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

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

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

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 $\widehat{\mathbf{z}}_t = \widehat{\boldsymbol{\Gamma}}_y' \mathbf{y}_t$ together.

Visual examination of CCF if p is not large!

Two component series of $\widehat{\mathbf{z}}_t$ is connected 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 the 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

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

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

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

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

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 $\widehat{\mathbf{z}}_t = \widehat{\mathbf{\Gamma}}_y' \mathbf{y}_t$ 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 $\text{Cov}\{\widehat{\rho}_{i,j}(h), \widehat{\rho}_{i,j}(k)\} = o_P(1/n)$, 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, $\rho(k) = \mathrm{Corr}(\xi_{t+k}, \eta_t) = 0$,

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

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

Permutation Method II: FDR based on multiple tests

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

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

Asymptotically $\widehat{\rho}(k) \sim N(0, 1/n)$, and $\widehat{\rho}(k), \widehat{\rho}(h)$ 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 $\widehat{\rho}(k) \sim N(0, 1/n)$, and $\widehat{\rho}(k), \widehat{\rho}(h)$ 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 $\widehat{\rho}(k) \sim N(0, 1/n)$, and $\widehat{\rho}(k), \widehat{\rho}(h)$ 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 $1 \le i < j \le q$.

Arranging those P-values in ascending order:

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

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

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

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 $1 \le i < j \le q$.

Arranging those P-values in ascending order:

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

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

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

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

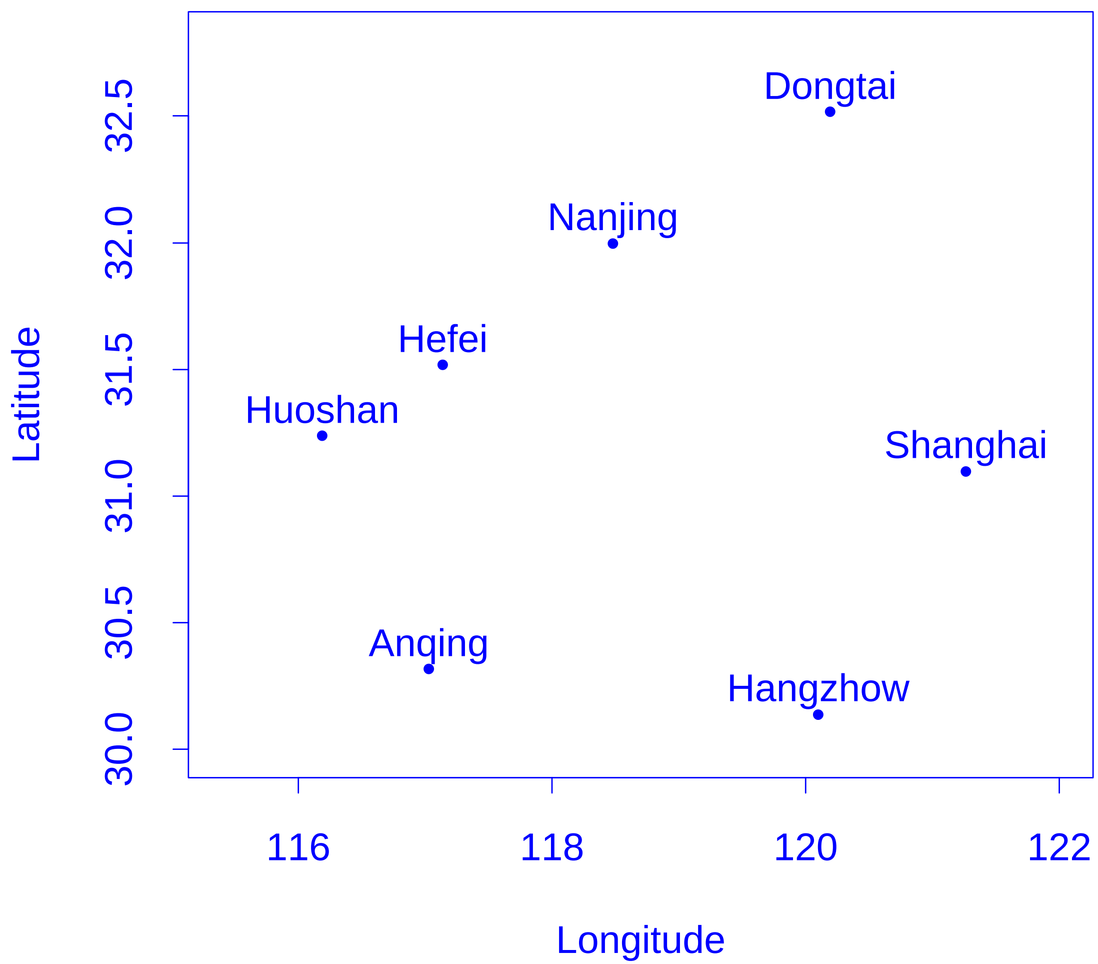
(i) The P-values p_k ($|k| < m$) are asymptotically independent

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

(iii) Ranking the pairwise dependences among the components of $\widehat{\mathbf{z}}_t$.

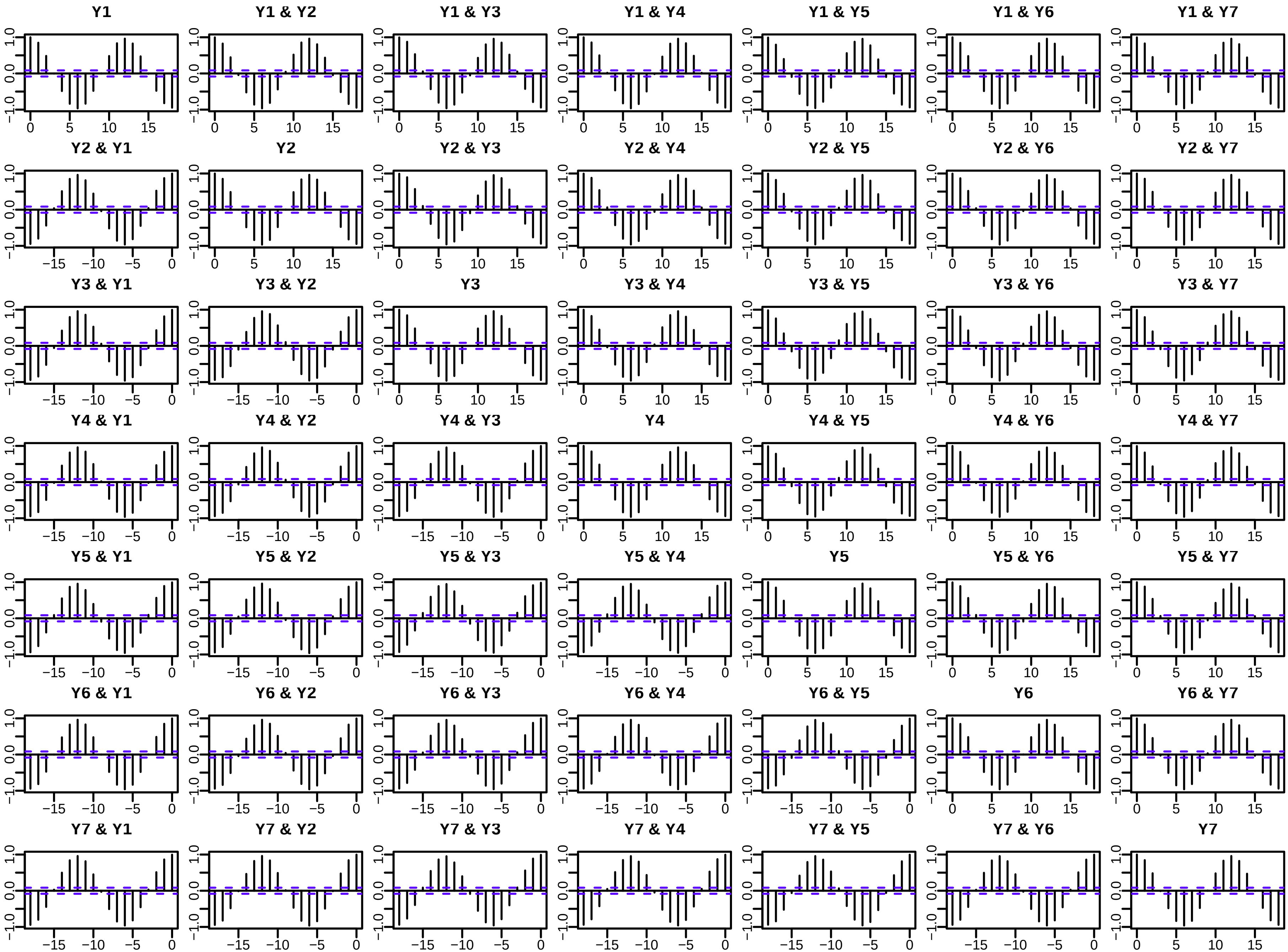
Real data examples

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



Time series plots of monthly temperatures in 7 cities

CCF of monthly temperatures in 7 cities



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

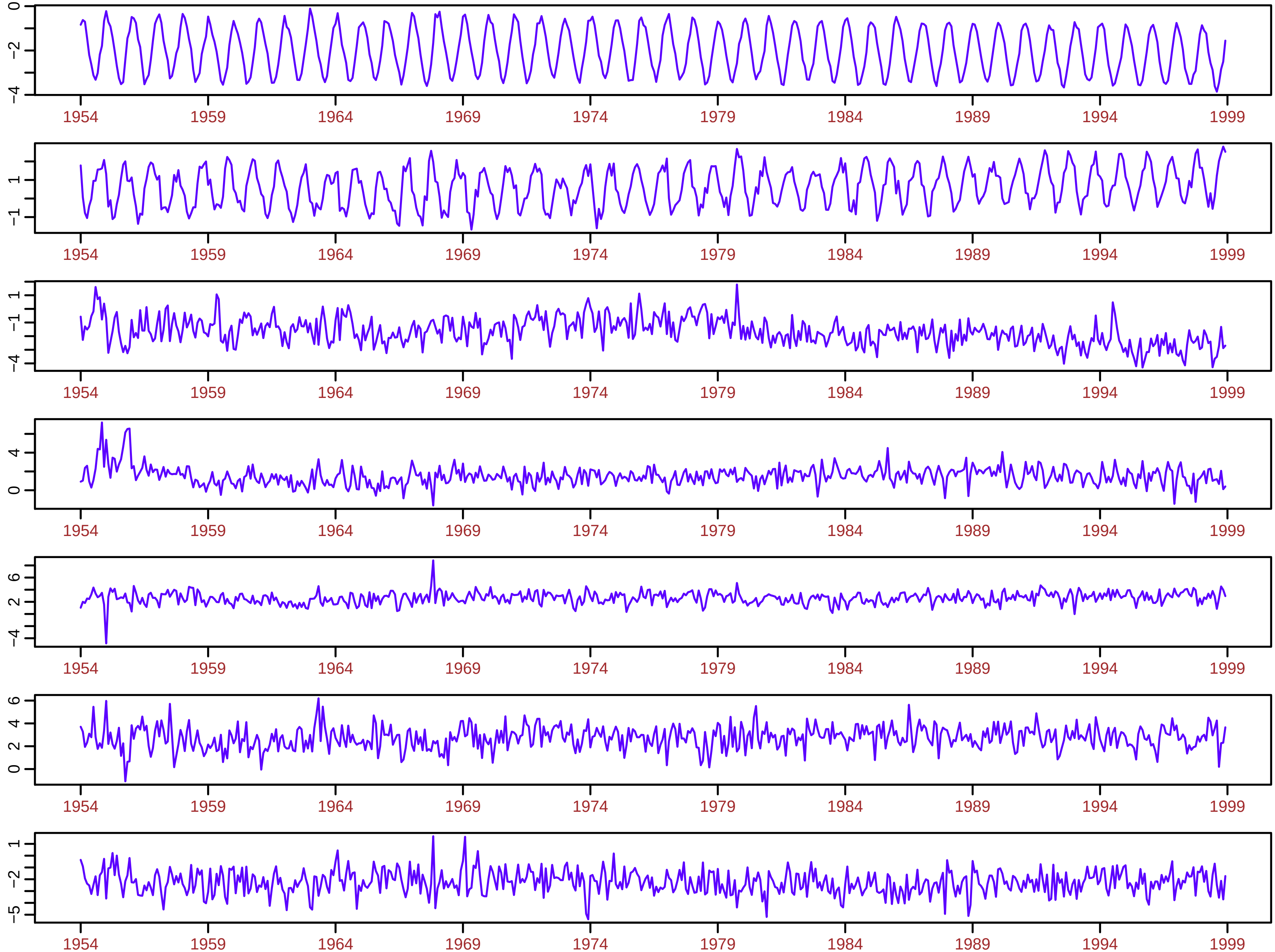
$$\widehat{\mathbf{B}} = \begin{pmatrix} 0.244 & -0.066 & 0.0187 & -0.050 & -0.313 & -0.154 & 0.200 \\ -0.703 & 0.324 & -0.617 & 0.189 & 0.633 & 0.499 & -0.323 \\ 0.375 & 1.544 & -1.615 & 0.170 & -2.266 & 0.126 & 1.596 \\ 3.025 & -1.381 & -0.787 & -1.691 & -0.212 & 1.188 & -0.165 \\ -0.197 & -1.820 &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

Note. $\widehat{\mathbf{B}} = \widehat{\mathbf{\Gamma}}_y' \widehat{\mathbf{\Sigma}}_y(0)^{-1/2}$.

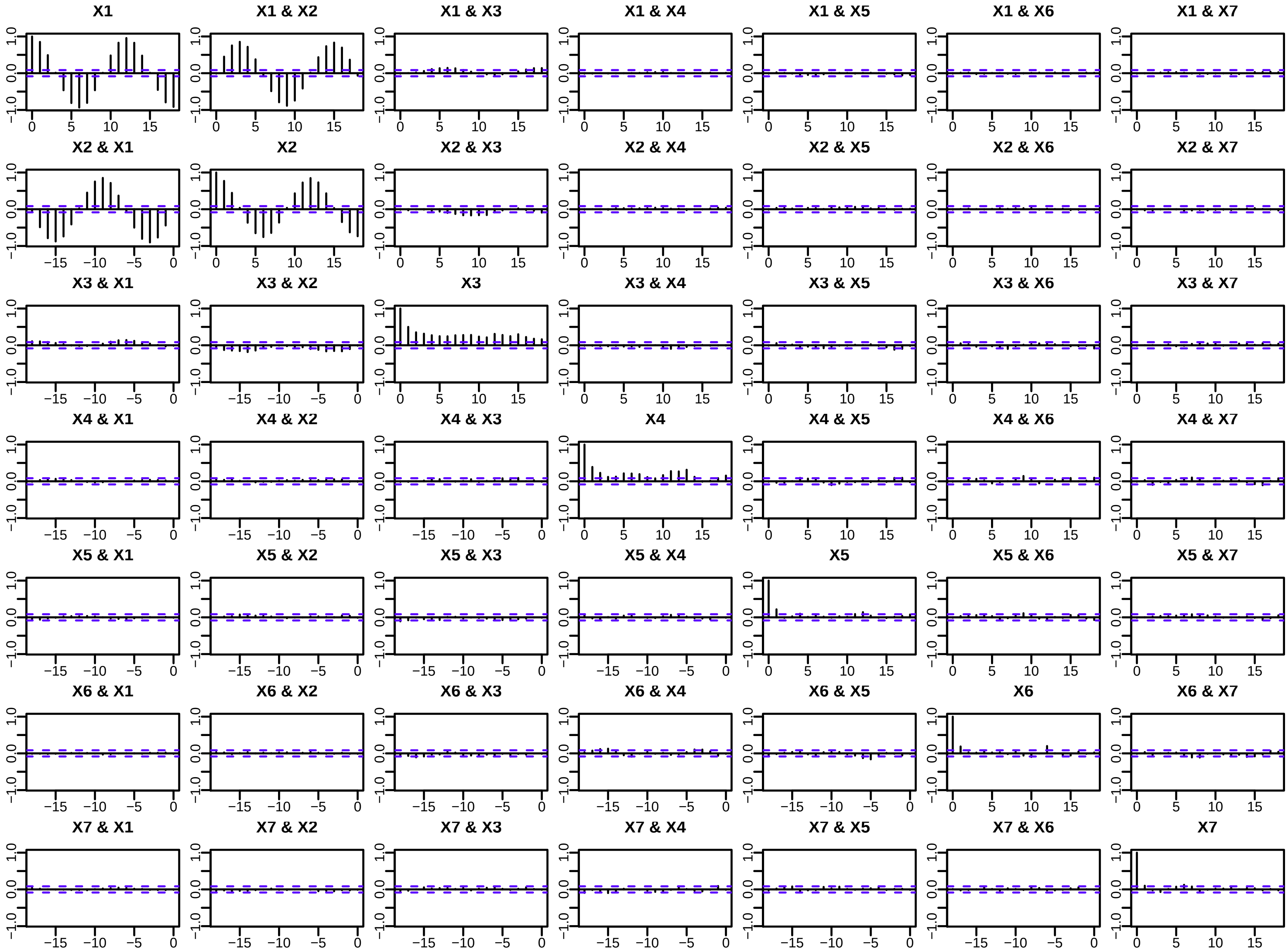
$n = 540$, $p = 7$

Used $k_0 = 5$ (in defining $\mathbf{W}_y$), hardly changed for $2 \leq k_0 \leq 36$.

Time series plots for transformed monthly temperatures



CCF for transformed monthly temperatures



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

- Visual examination of CCF: {1, 2, 3}, {4}, {5}, {6} and {7}
- Permutation based on max-CCF: the same grouping with $2 \leq m \leq 30$
- Permutation based on FDR: the same grouping with $2 \leq m \leq 30$ and $\beta \in [0.001\%, 1\%]$.

7-dim time series $\rightarrow$ 5 uncorrelated time series

Methodology

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 $\mathbf{y}_t$.

Implementation: 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.

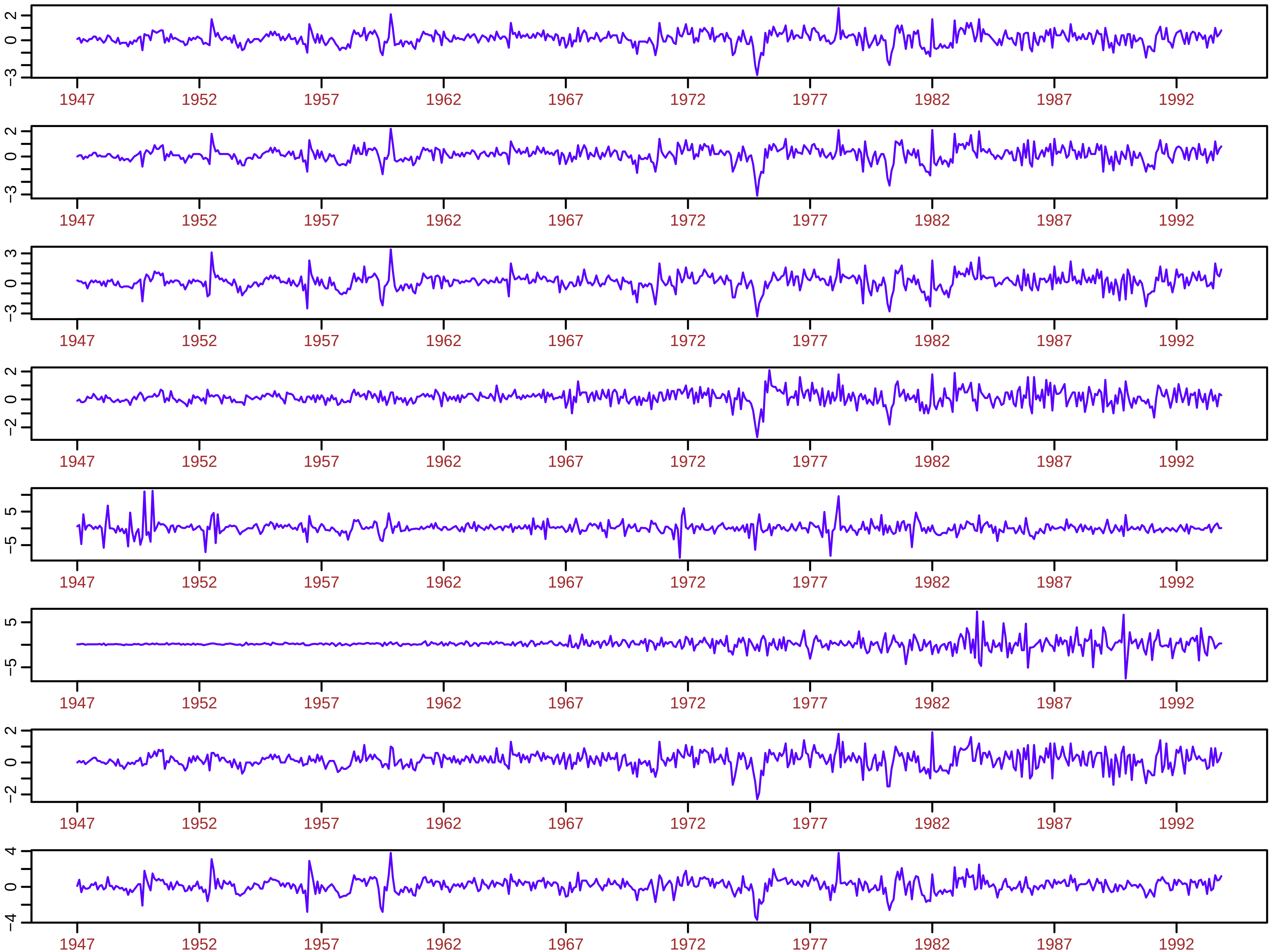
	One-step MSE	Two-step MSE
VAR	$1.669_{(2.355)}$	$1.815_{(2.372)}$
RVAR	$1.677_{(2.324)}$	$1.829_{(2.398)}$
Segmentation	$1.381_{(1.888)}$	$1.543_{(1.874)}$

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

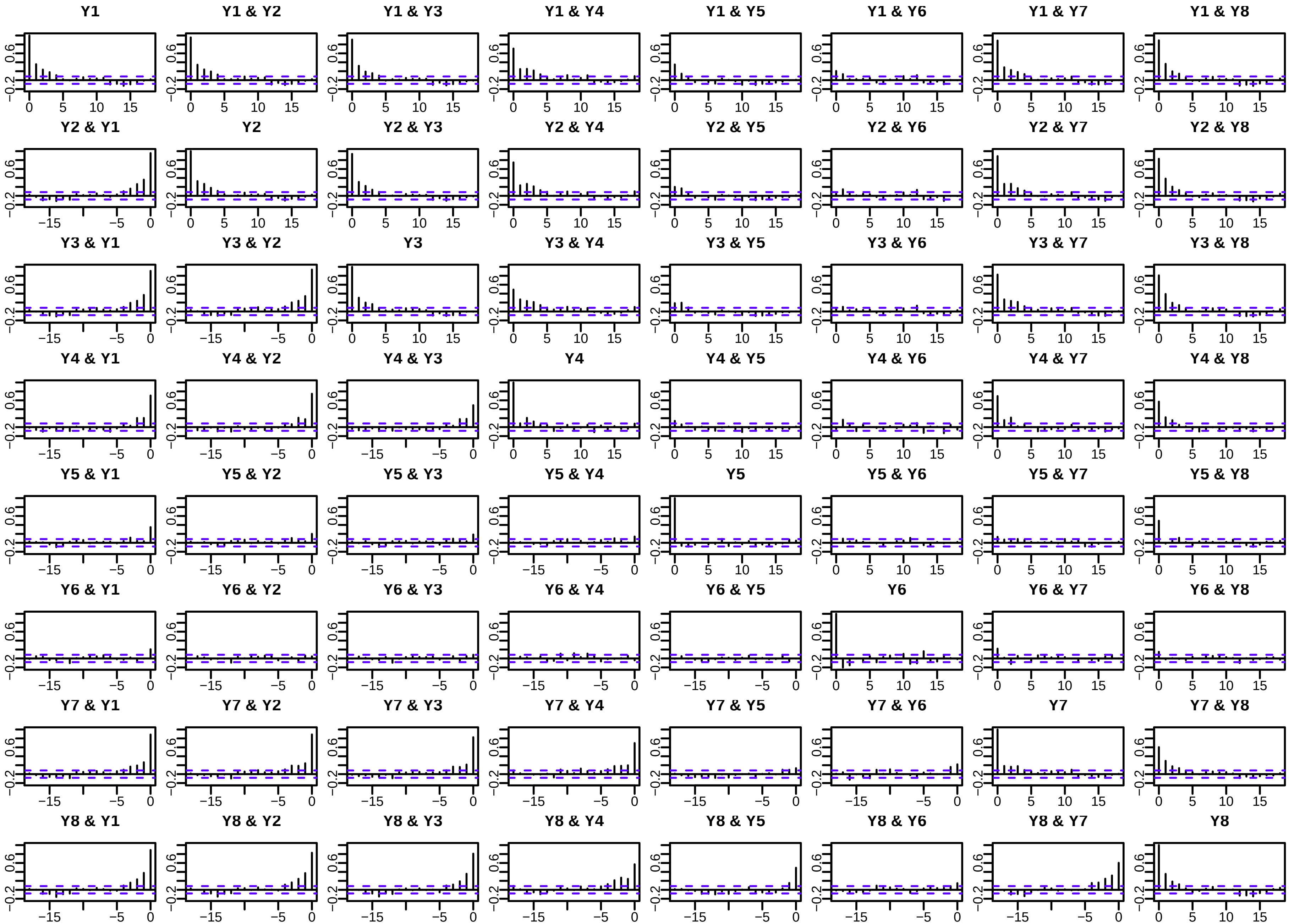
8 indices: *total index, manufacturing index, durable manufacturing, nondurable manufacturing, mining, utilities, products, materials*.

Nonstationary trends: difference each series

8 differenced monthly US Industrial Indices



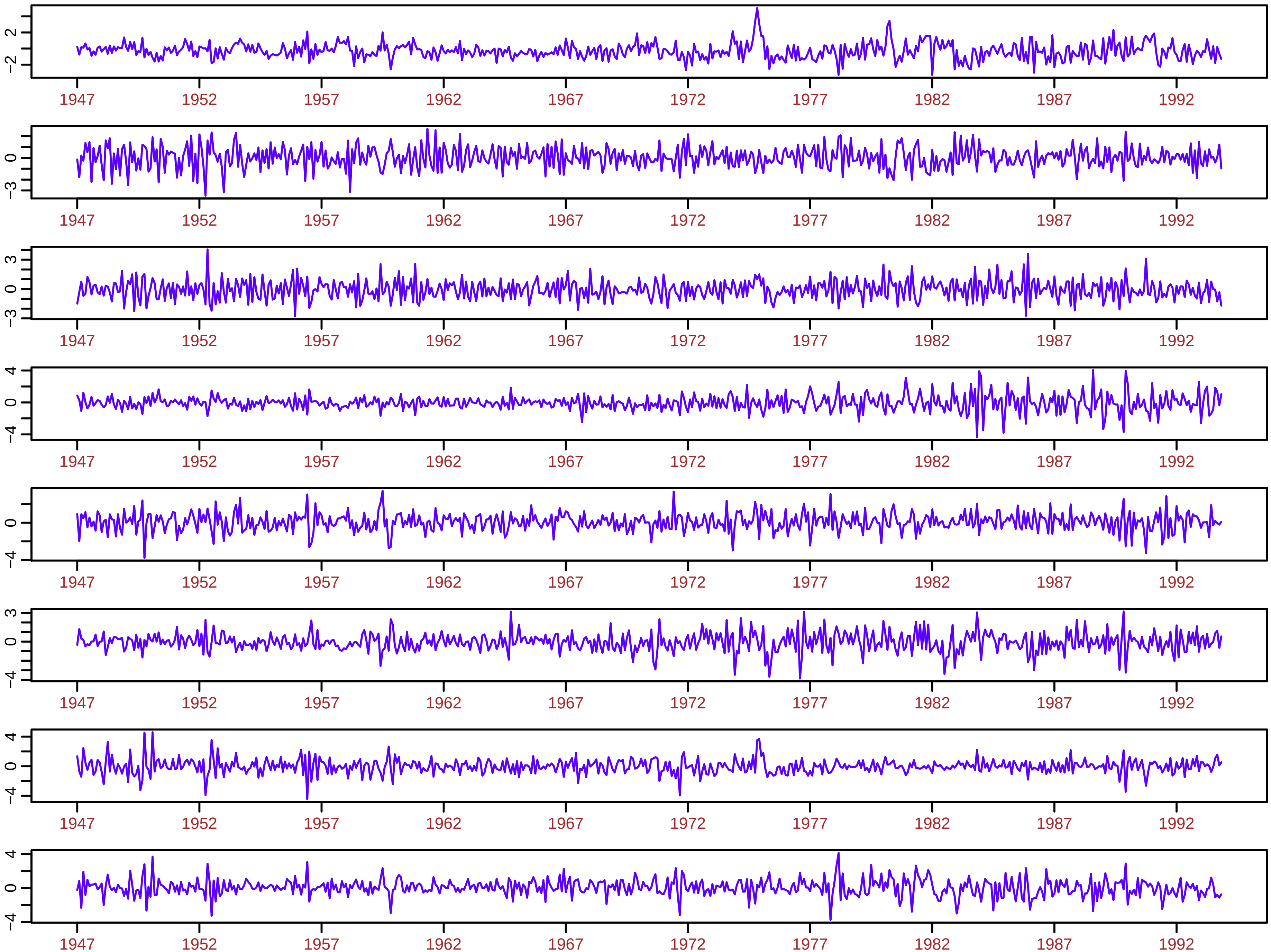
CCF of 8 differenced monthly US Industrial Indices



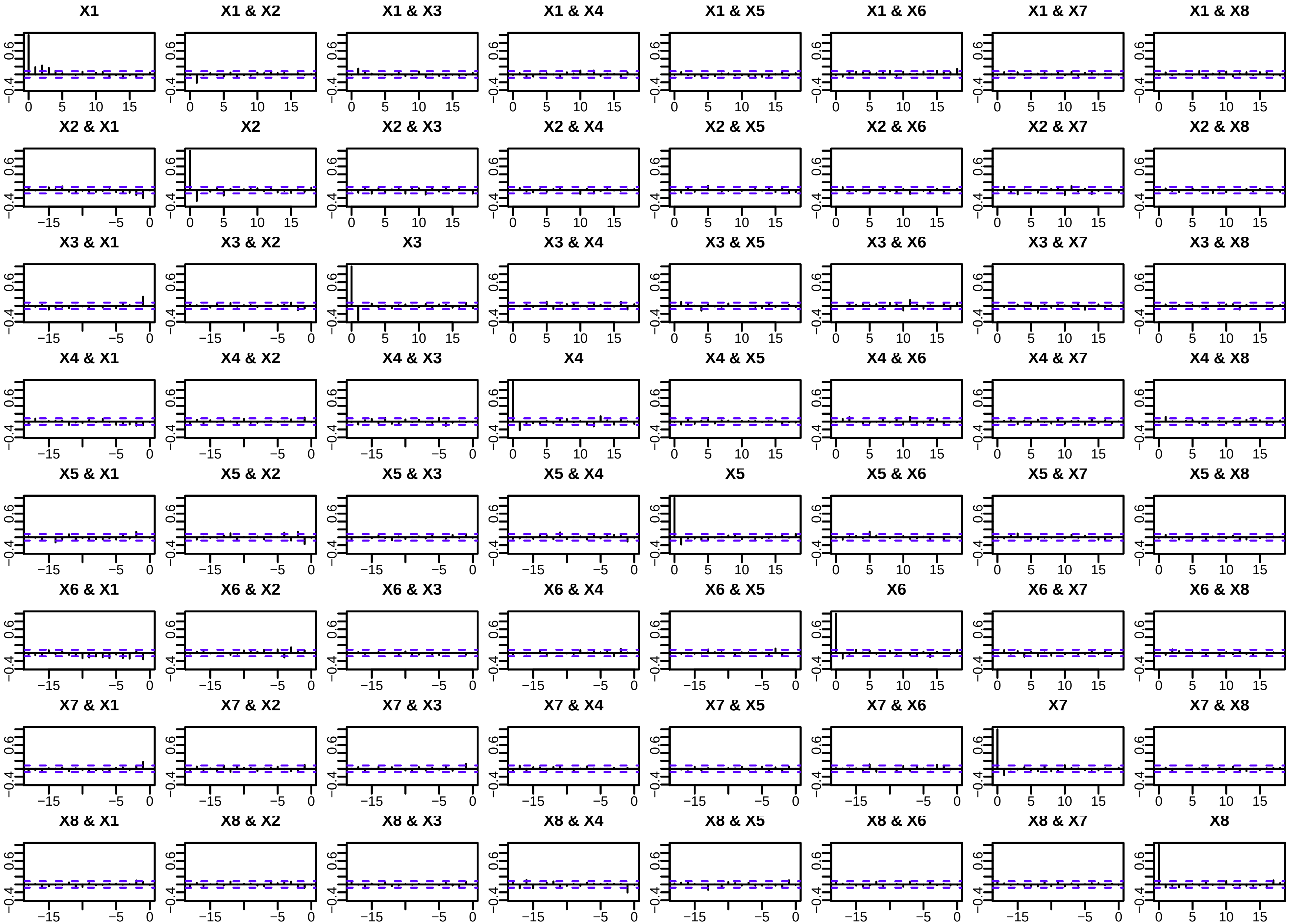
Transformation: $\widehat{\mathbf{x}}_t = \widehat{\mathbf{B}}\mathbf{y}_t$

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

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

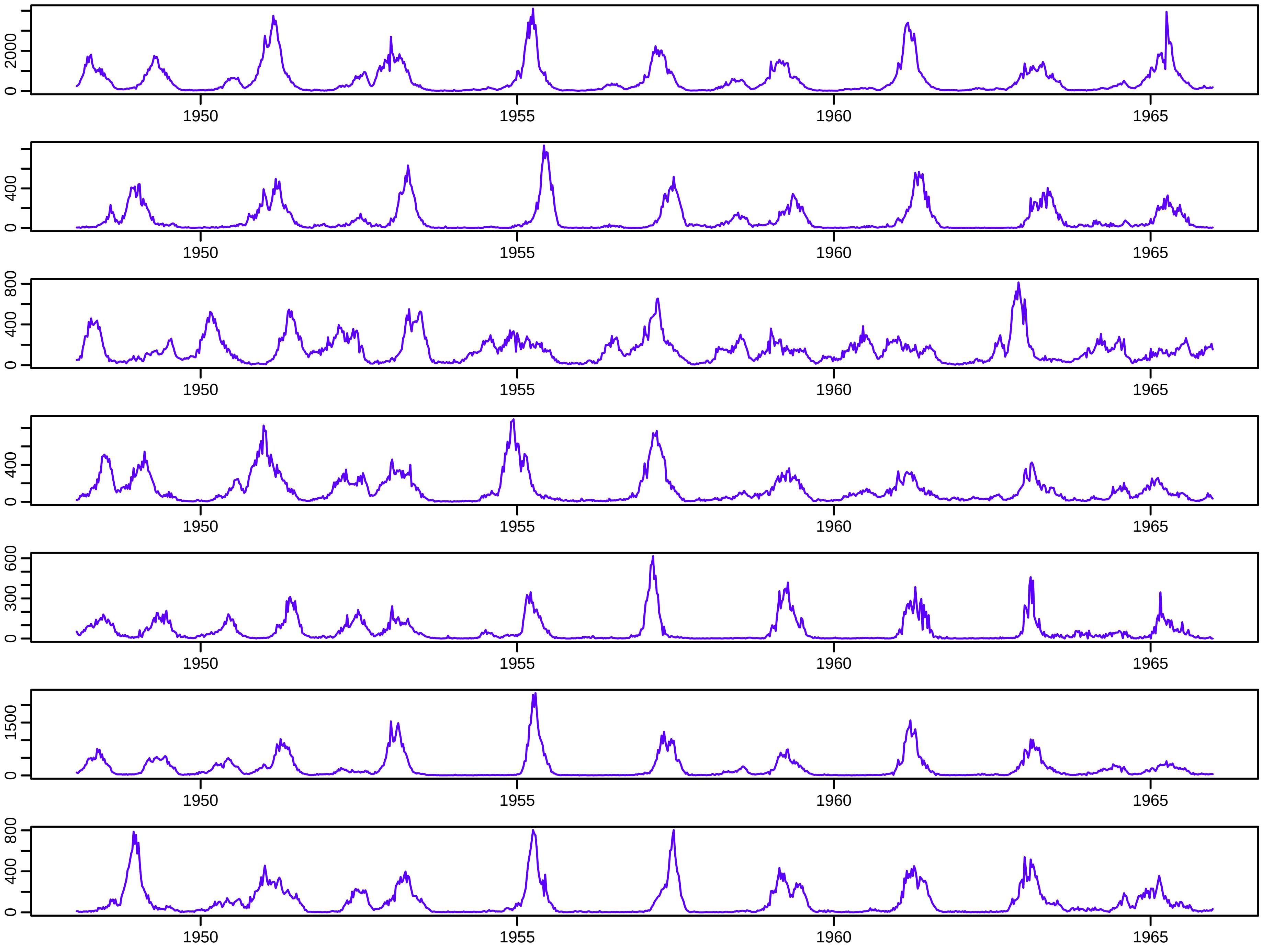
	One-step MSE	Two-step MSE
VAR	$0.615_{(1.349)}$	$1.168_{(2.129)}$
RVAR	$0.606_{(1.293)}$	$1.159_{(2.285)}$
Segmentation	$0.588_{(1.341)}$	$1.154_{(2.312)}$

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.

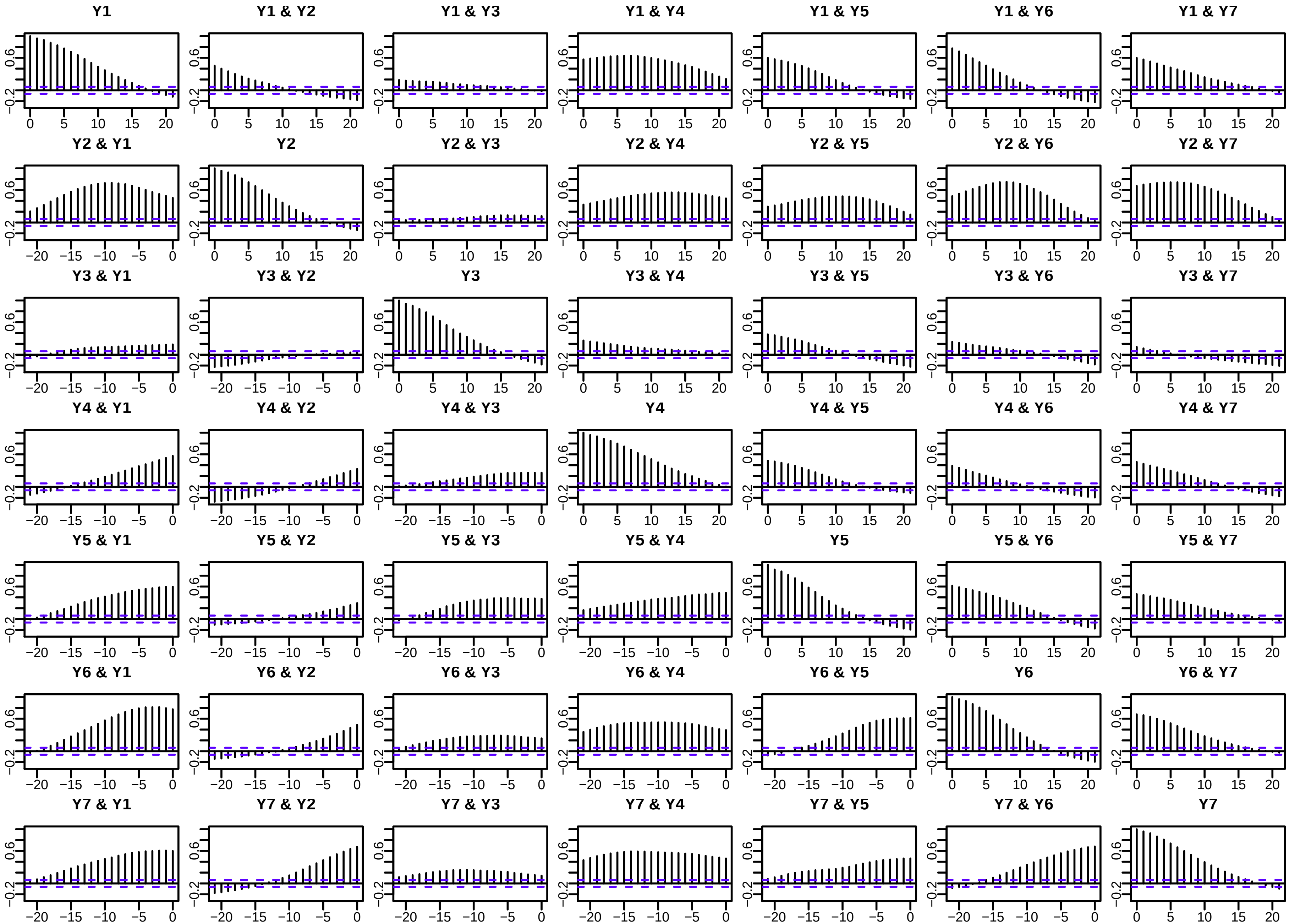
All the 7 series show biennial cycles, which is the major driving force for the cross correlations among different cities.

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

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



CCF of weekly recorded cases of measles in 7 cities

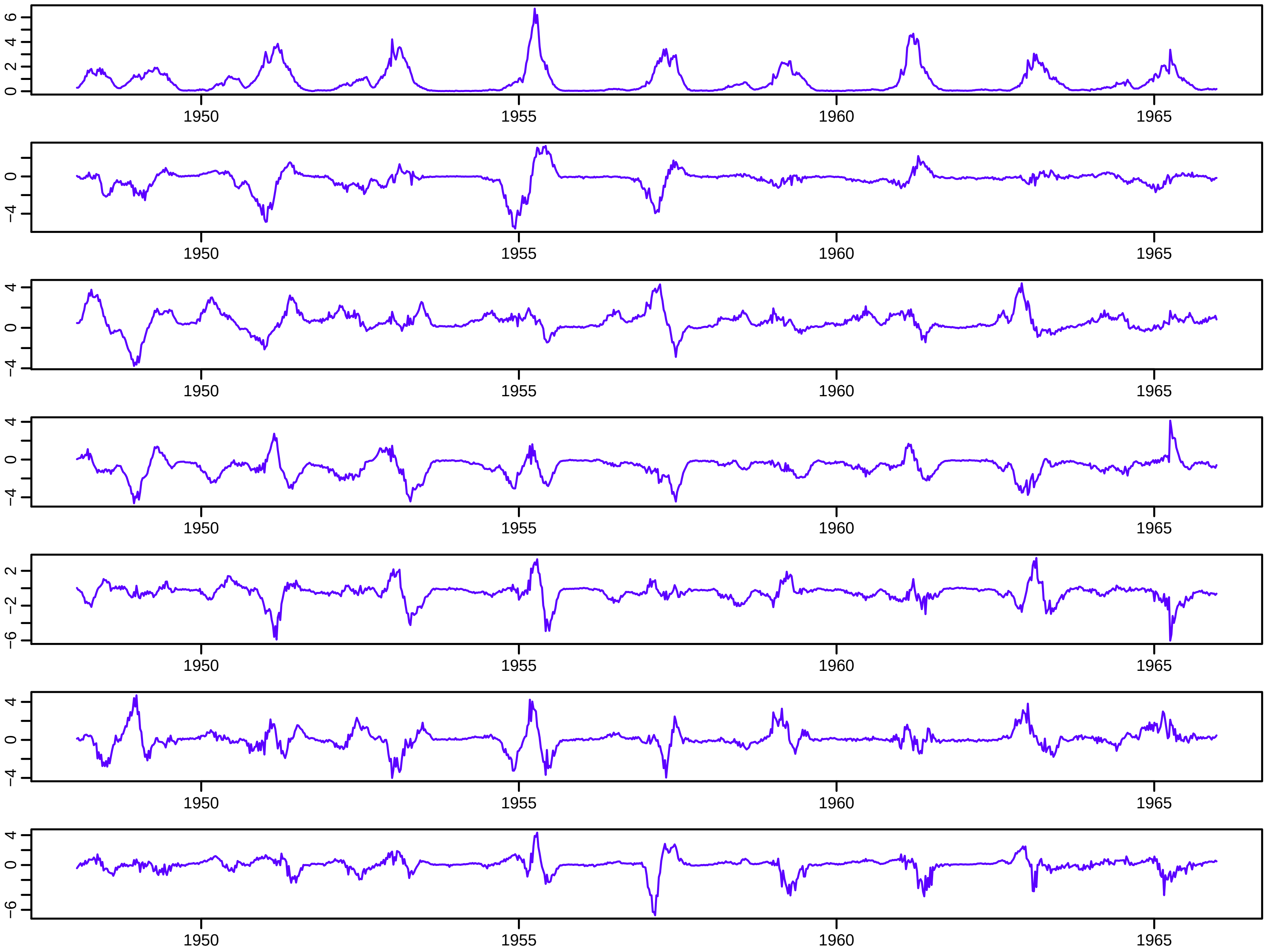


Transformation: $\widehat{\mathbf{x}}_t = \widehat{\mathbf{B}}\mathbf{y}_t$

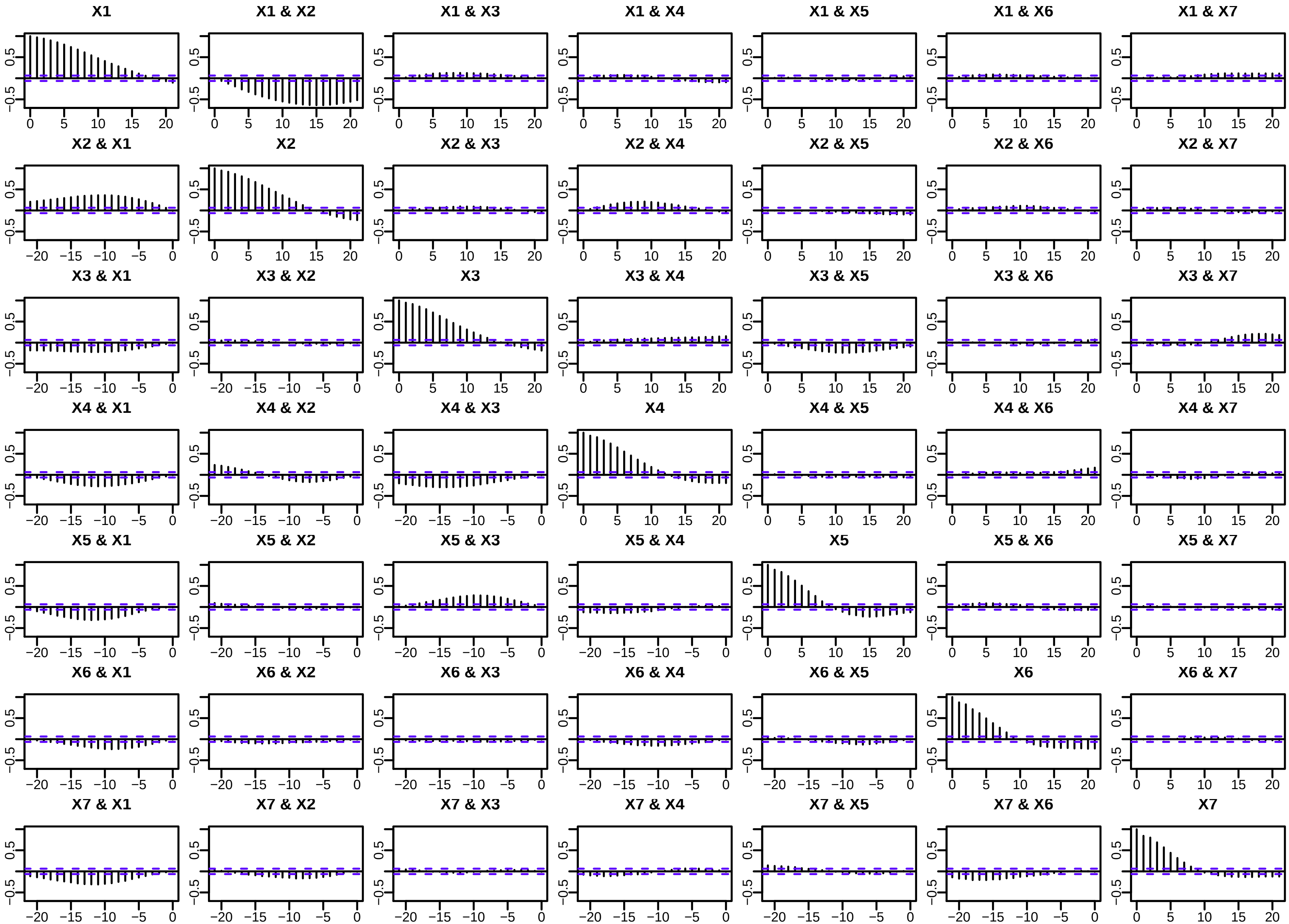
$$\widehat{\mathbf{B}} = \begin{pmatrix} -4.898e4 & 3.357e3 & -3.315e04 & -6.455e3 & 2.337e3 & 1.151e3 & -1.047e3 \\ 7.328e4 & 2.85e4 & -9.569e6 & -2.189e3 & 1.842e3 & 1.457e3 & 1.067e3 \\ -5.780e5 & 5.420e3 & -5.247e3 & 5.878e4 & -2.674e3 & -1.238e3 & 6.280e3 \\ -1.766e3 & 3.654e3 & 3.066e3 & 2.492e3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

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

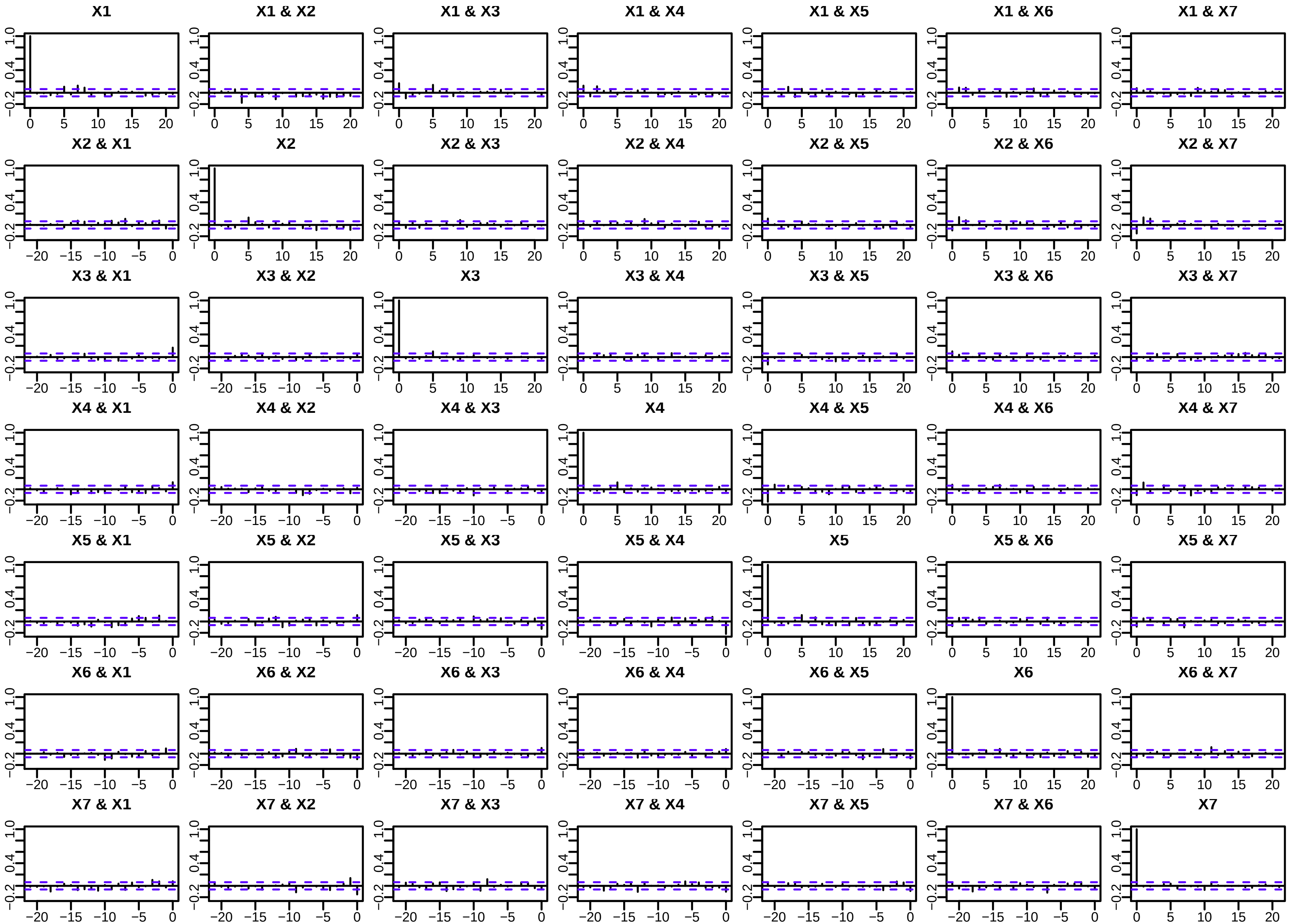
Transformed weekly recorded cases of measles in 7 cities



CCF of transformed weekly recorded cases of measles in 7 cities

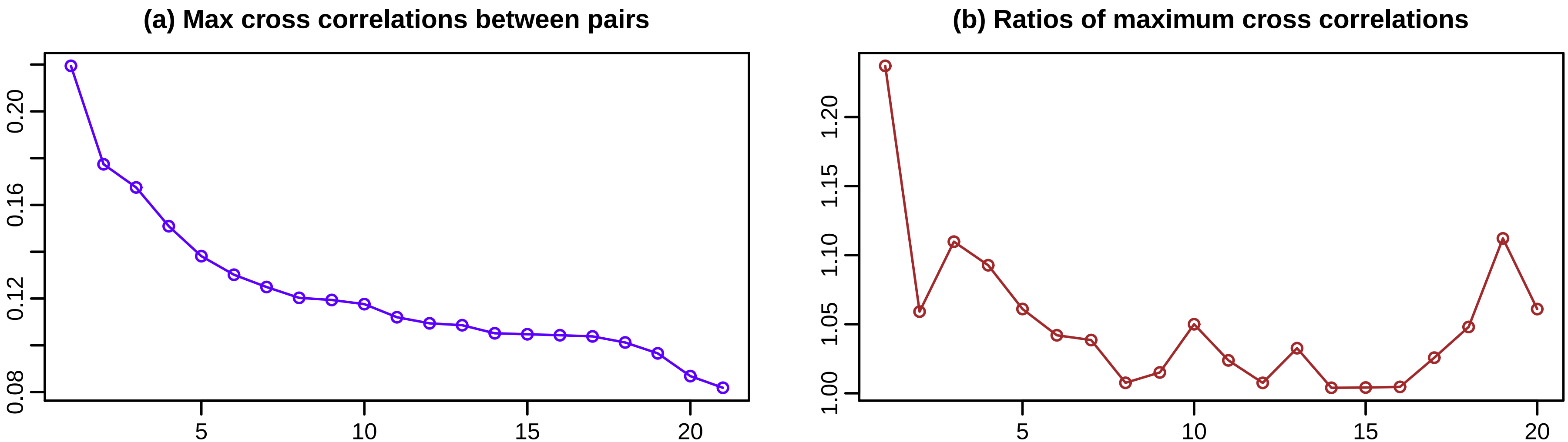


CCF of prewhitened transformed weekly recorded cases of measles



Note. When none of the component series are WN, the confidence bounds $\pm 1.96/\sqrt{n}$ 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
VAR	$503.408_{(1124.213)}$	$719.499_{(2249.986)}$
RVAR	$574.582_{(1432.217)}$	$846.141_{(2462.019)}$
Segmentation	$472.106_{(1088.17)}$	$654.843_{(1807.502)}$

Example 4. Daily impressions of 32 keywords for an air ticket booking website powered by Baidu (www.baidu.com) over 130 days. $\Longrightarrow$

$$n = 130, \quad p = 32$$

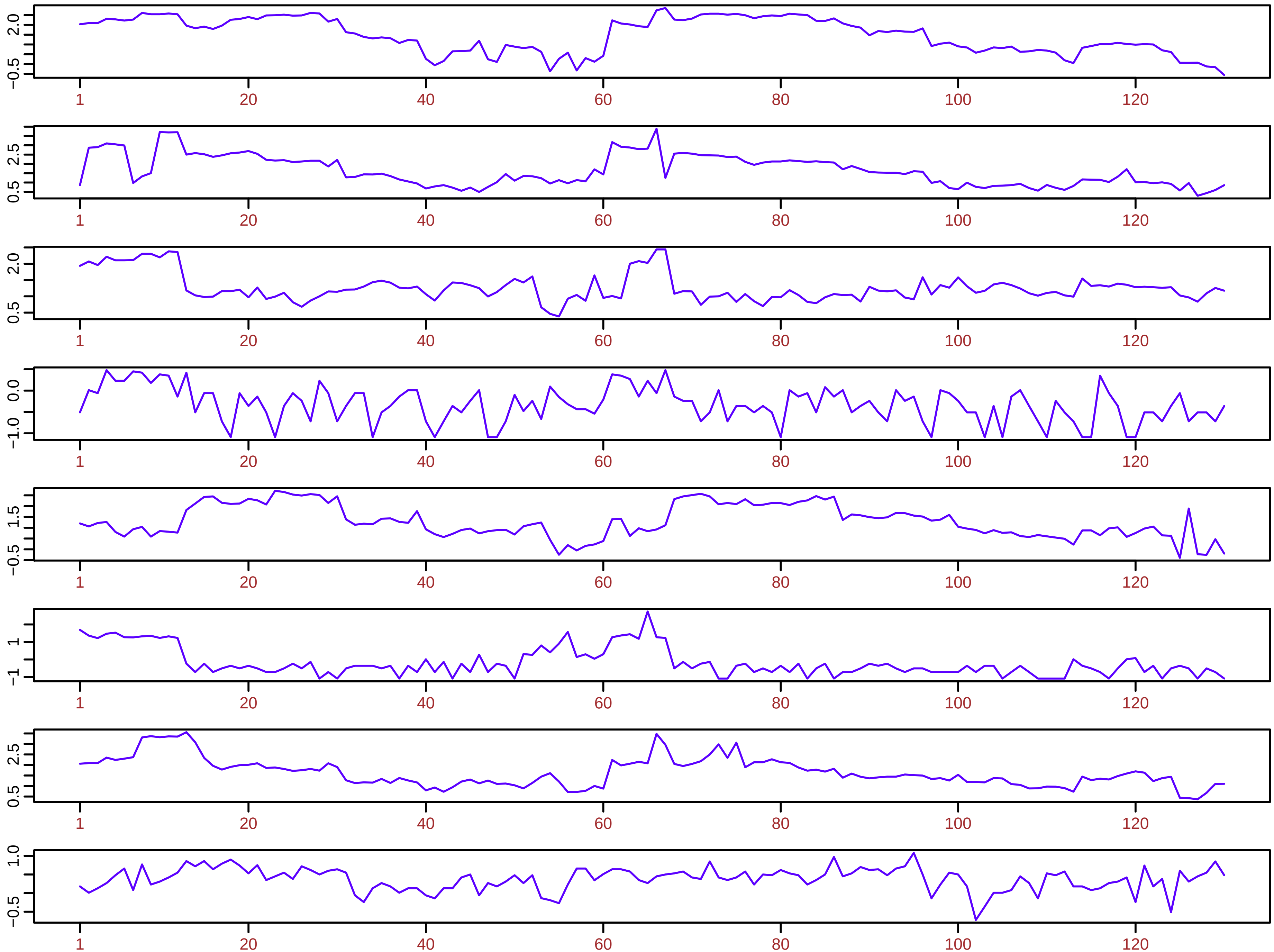
Data have been coded to protect the confidentiality.

Advertisements are accessed by keywords search from a search engine.

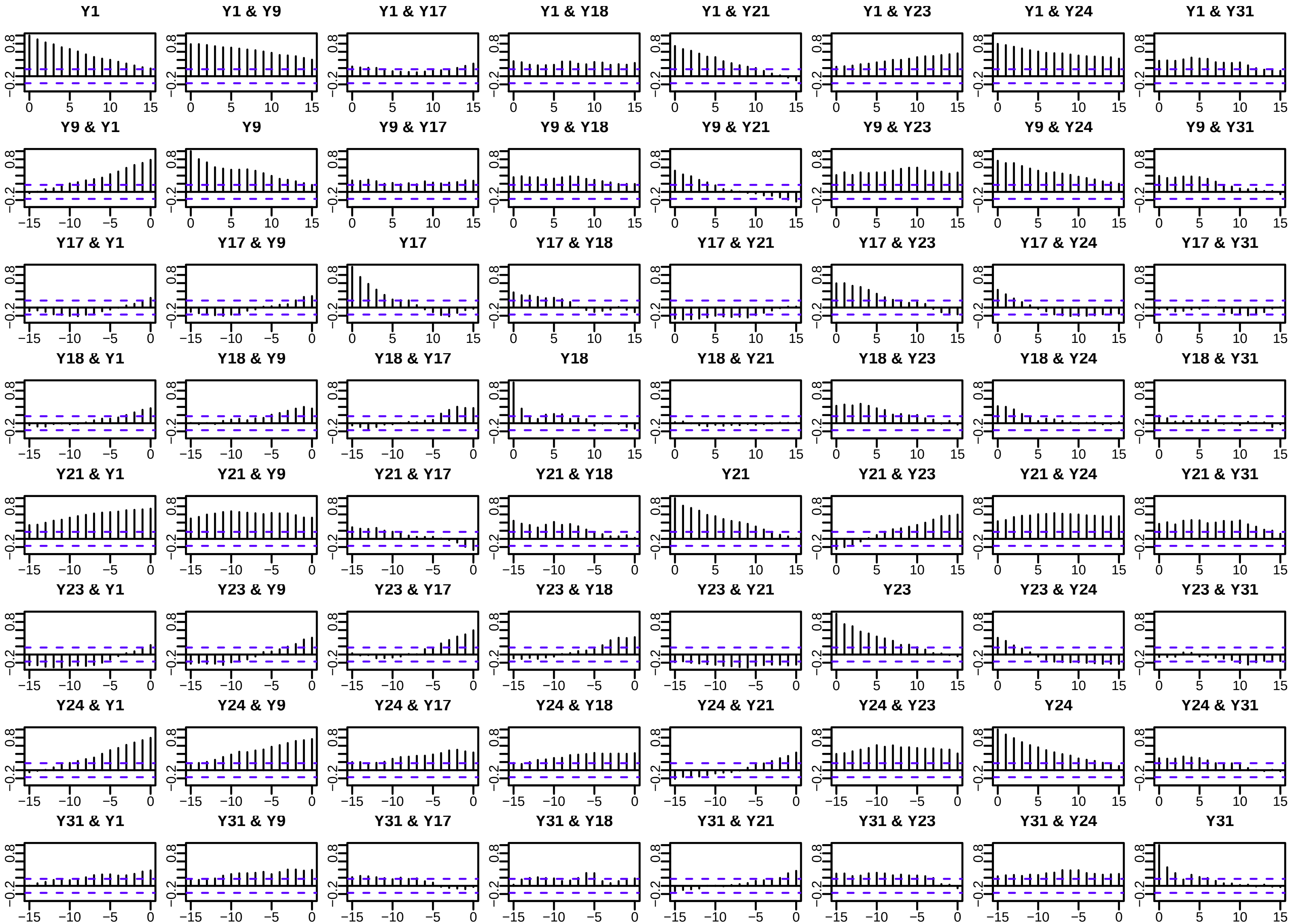
Each display of an advertisement is counted as one impression.

Applying the proposed transformation with $m = 10$ (or $1 \le m \le 15$), the transformed 32 time series are segmented into 31 groups with the only non-single-element group $\{10, 13\}$.

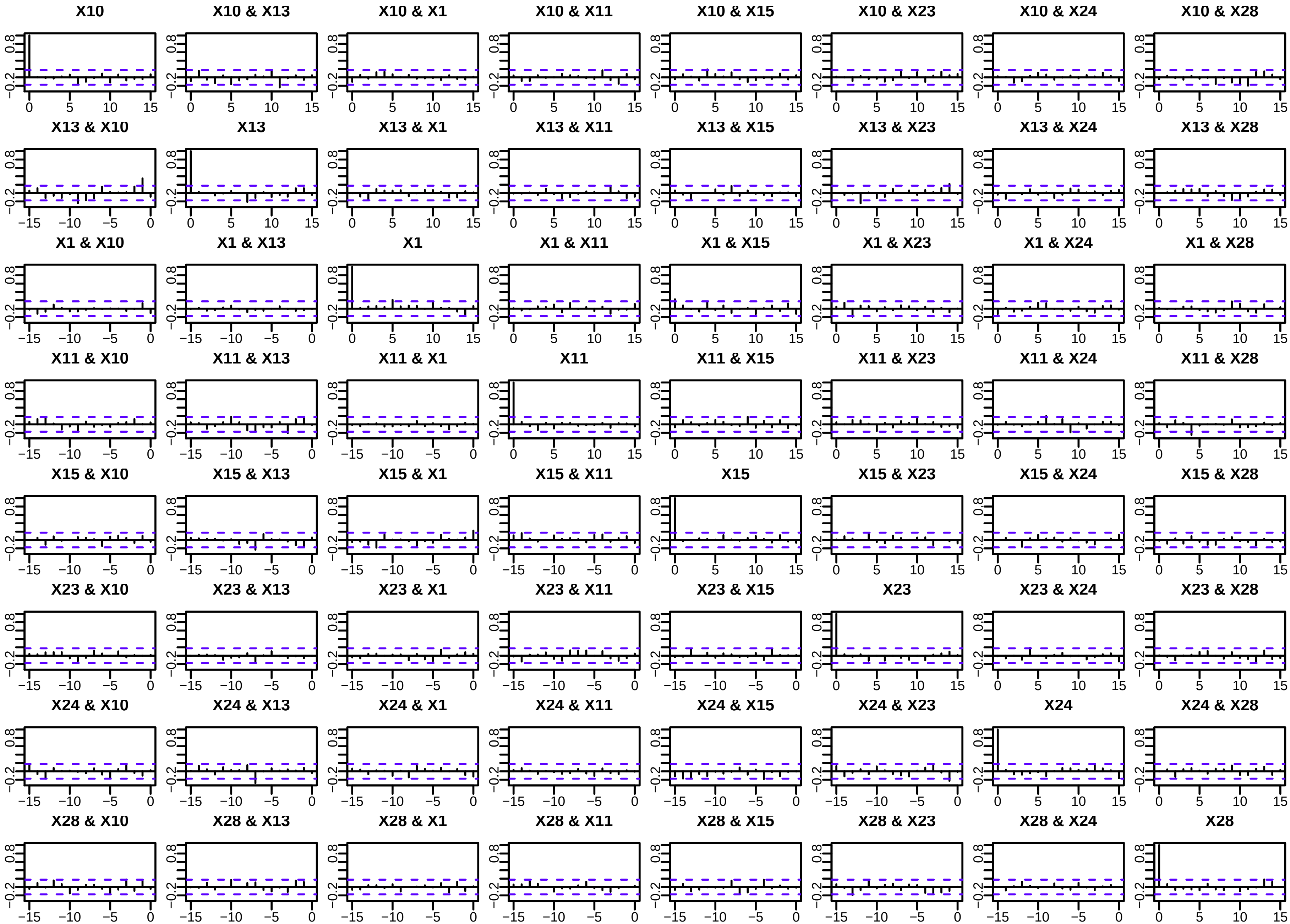
Time series of 8 randomly selected daily impressions



CCF of 8 randomly selected daily impressions



CF of 8 prewhitened transformed daily impressions (6 randomly selected



Example 5. Daily sales of a clothing brand in 25 provinces in China in 1 January 2008 – 16 December 2012.

$n = 1812, \ p = 25$

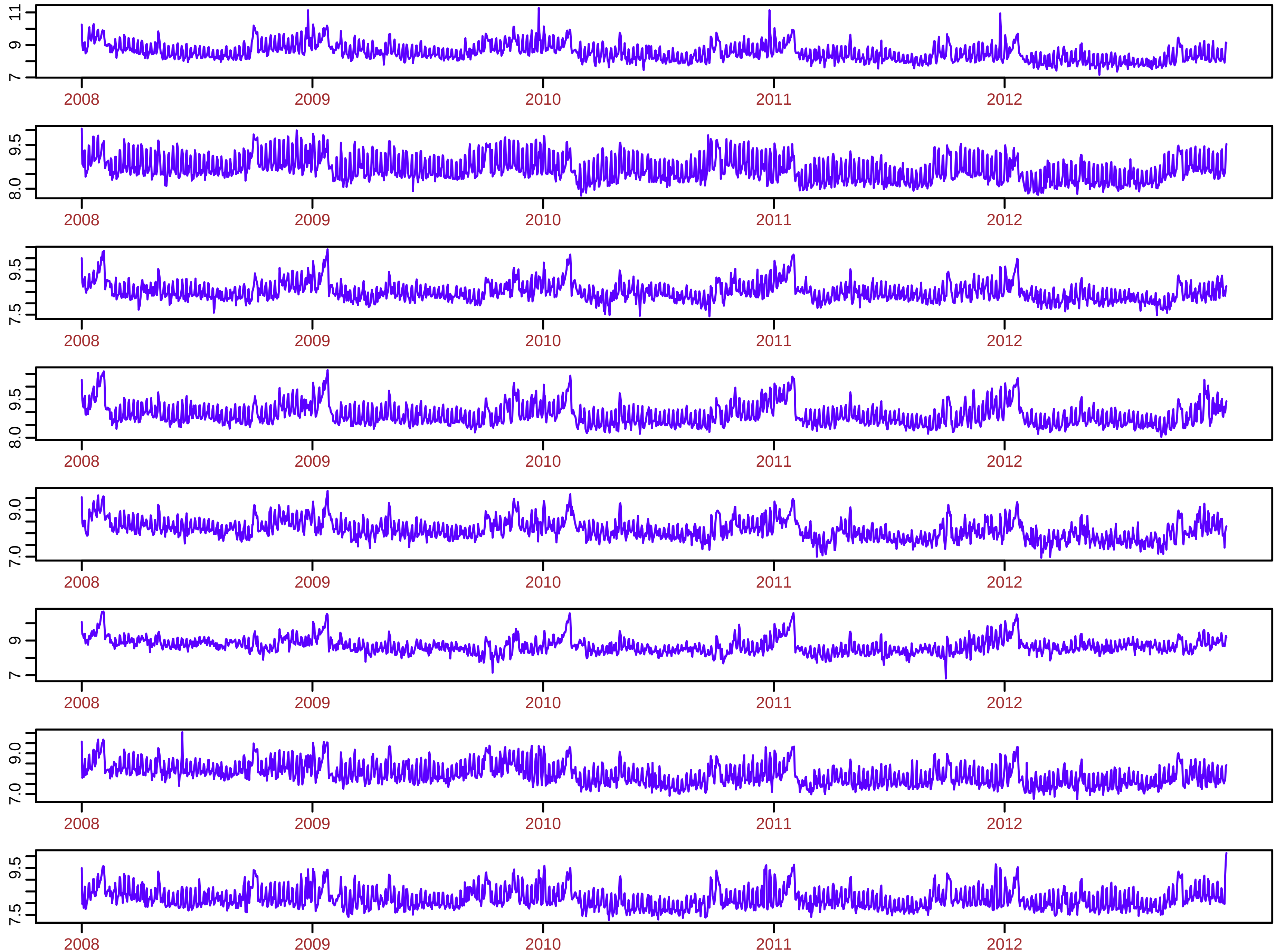
Annual pattern: peak in February

Strong periodicity component with the period 7.

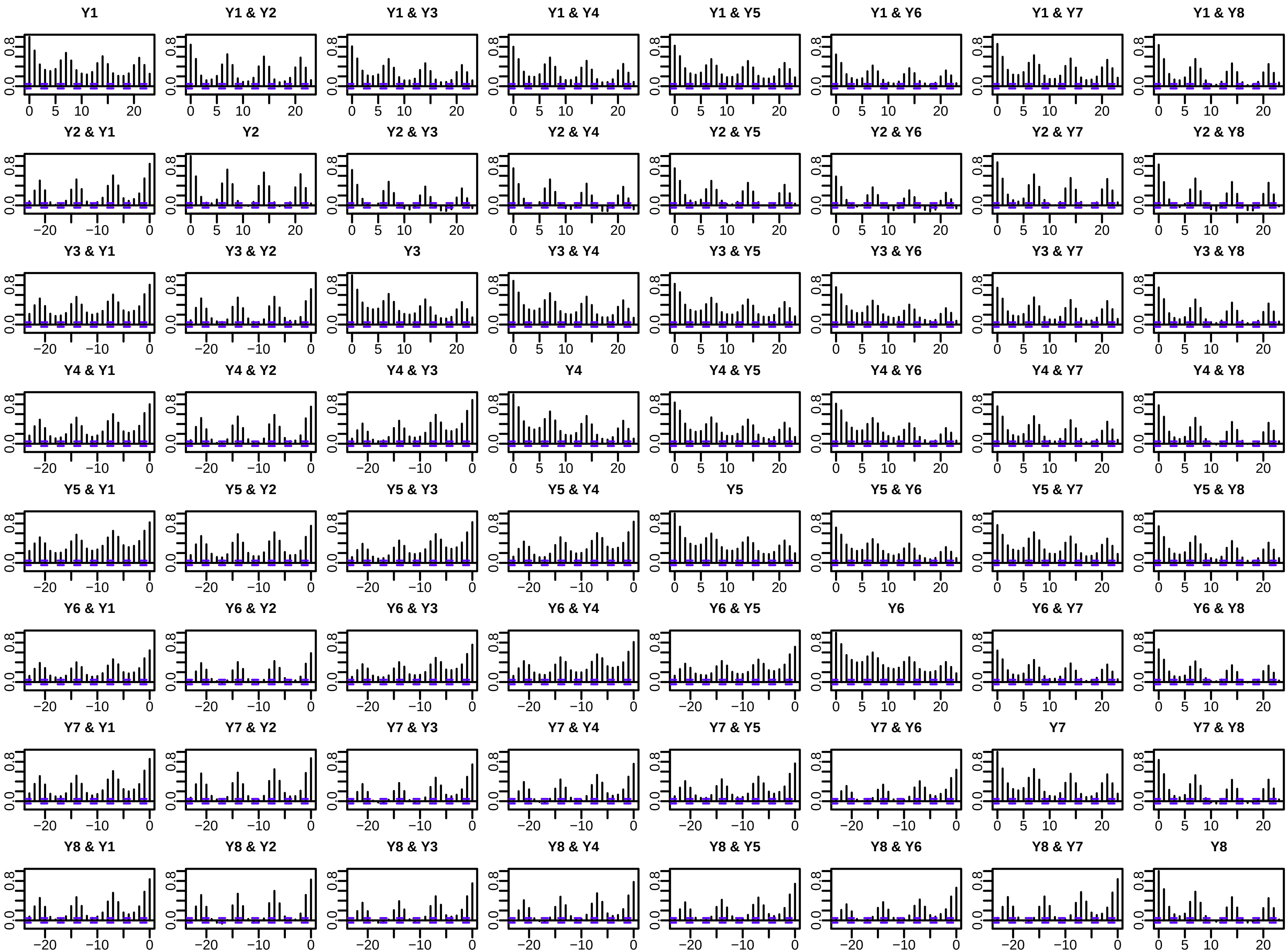
The 25 transformed series are segmented into 24 groups with $\{15, 16\}$ as one group.

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

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

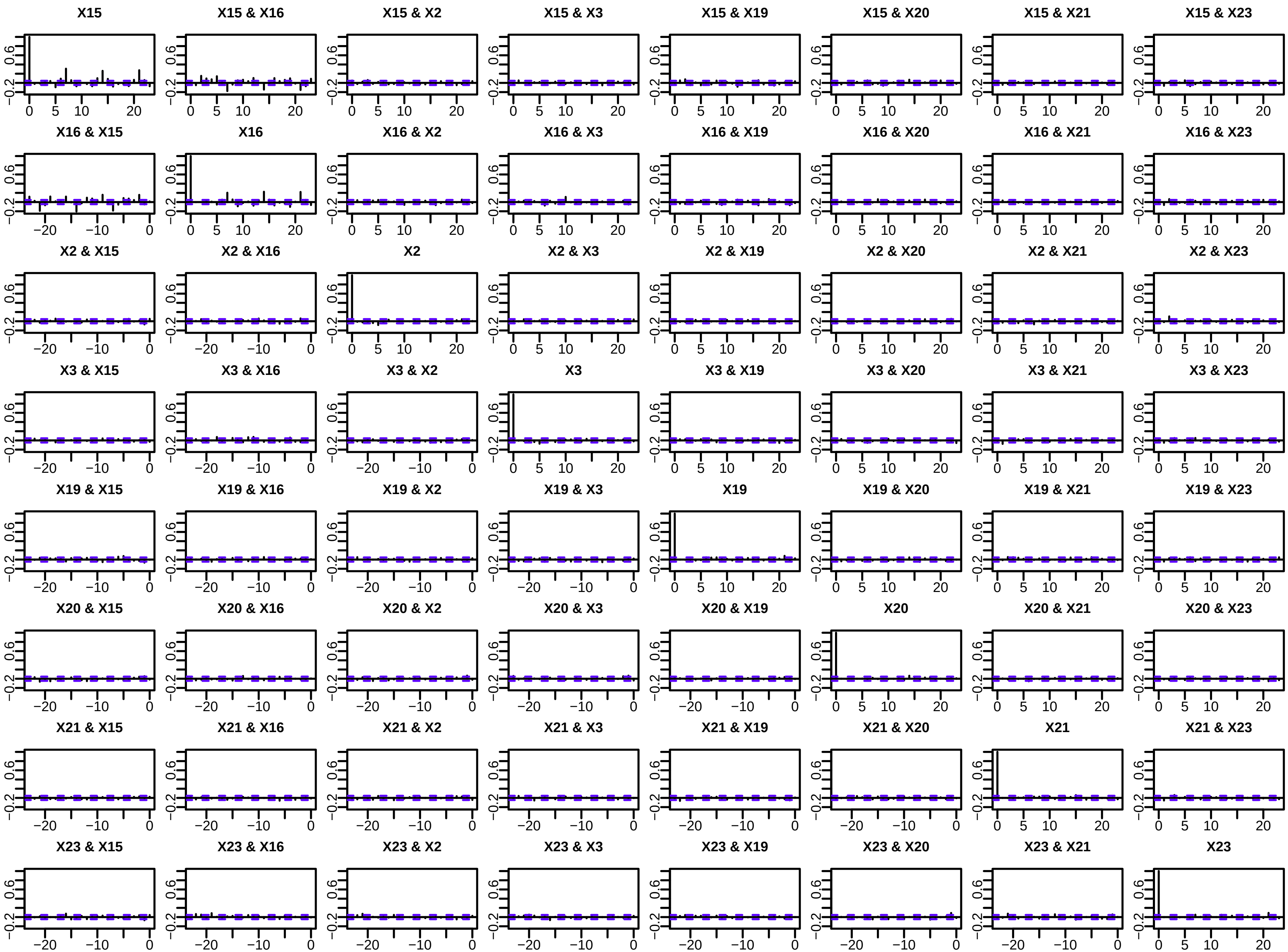


CCF of daily sales of a clothing brand in 8 provinces



CCF of 8 transformed daily sales of a clothing brand

CCF 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	$0.208_{(1.255)}$	$0.194_{(1.250)}$
VAR	$0.295_{(2.830)}$	$0.301_{(2.930)}$
RVAR	$0.293_{(2.817)}$	$0.296_{(2.962)}$
Segmentation	$0.153_{(0.361)}$	$0.163_{(0.341)}$

Cross correlations between provinces are useful information.

However they cannot be used via direct VAR!

Asymptotic theory

For $p \times r$ matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ and $\mathbf{H}_1'\mathbf{H}_1 = \mathbf{H}_2'\mathbf{H}_2 = \mathbf{I}_r$, let

$$D(\mathcal{M}(\mathbf{H}_1), \mathcal{M}(\mathbf{H}_2)) = \sqrt{1 - \frac{1}{r}\text{tr}(\mathbf{H}_1\mathbf{H}_1'\mathbf{H}_2\mathbf{H}_2')}.$$

Then $D(\mathcal{M}(\mathbf{H}_1), \mathcal{M}(\mathbf{H}_2)) \in [0, 1]$.

It equals 0 iff $\mathcal{M}(\mathbf{H}_1) = \mathcal{M}(\mathbf{H}_2)$, and 1 iff $\mathcal{M}(\mathbf{H}_1) \perp \mathcal{M}(\mathbf{H}_2)$.

$\mathbf{y}_t$ is assumed to be weakly stationary and α-mixing, i.e.

$$\alpha_k \equiv \sup_i \sup_{A \in \mathcal{F}_{-\infty}^i,\, B \in \mathcal{F}_{i+k}^\infty} |P(A \cap B) - P(A)P(B)| \to 0, \quad \text{as } k \to \infty,$$

where $\mathcal{F}_i^j = \sigma(\mathbf{y}_t : \; i \le t \le j)$.

p **fixed**

C1. For some constant $\gamma > 2$,

$$\sup_t \max_{1\le i\le p} E(|y_{i,t} - Ey_{i,t}|^{2\gamma}) < \infty.$$

C2. $\sum_{k=1}^{\infty} \alpha_k^{1-2/\gamma} < \infty$.

Theorem 1. Let conditions C1, C2 hold, ϖ be positive and p be fixed. Then there exists an $\widehat{\mathbf{A}} = (\widehat{\mathbf{A}}_1, \cdots, \widehat{\mathbf{A}}_q)$ of which the columns are a permutation of the columns of $\widehat{\mathbf{\Gamma}}_y$, such that

$$\max_{1\le j\le q} D(\mathcal{M}(\widehat{\mathbf{A}}_j), \mathcal{M}(\mathbf{A}_j)) = O_p(n^{-1/2}).$$

$p = o(n^c)$ **for some** $c > 0$

C3. (Sparsity of $\mathbf{A} = (a_{i,j})$) For some constant $\iota \in [0, 1)$,

$$\max_{1 \le j \le p} \sum_{i=1}^{p} |a_{i,j}|^{\iota} \le s_1 \qquad \text{and} \qquad \max_{1 \le i \le p} \sum_{j=1}^{p} |a_{i,j}|^{\iota} \le s_2,$$

where s_1 and s_2 are positive constants which may diverge together with p.

C4. For some positive constants $l > 2$ and $\tau > 0$,

$$\sup_t \max_{1 \le i \le p} P(|y_{i,t} - \mu_i| > x) = O(x^{-2(l+\tau)}) \text{ as } x \to \infty.$$

C5. $\alpha_k = O(k^{-l(l+\tau)/(2\tau)})$ as $k \to \infty$.

Remark. As $\boldsymbol{\Sigma}_y(k) = \mathbf{A}\boldsymbol{\Sigma}_x(k)\mathbf{A}'$, the sparsity of $\mathbf{A}$ and the maximum block size of $\boldsymbol{\Sigma}_x(k)$ provide a measure for the sparsity of $\boldsymbol{\Sigma}_y(k)$.

Let $S_{\max}$ be the maximum block size in $\mathbf{\Sigma}_x(k)$, and

$$\rho_j = \min_{\substack{1\le i\le q\\ i\ne j}} \min |\lambda(\mathbf{W}_{x,i}) - \lambda(\mathbf{W}_{x,j})|, \qquad j = 1, \cdots, q,$$

$$\delta = s_1 s_2 \max_{k=1,\ldots,k_0} \|\mathbf{\Sigma}_x(k)\|_\infty^\iota, \quad \kappa_1 = \min_{k=1,\ldots,k_0} \|\mathbf{\Sigma}_x(k)\|_2, \quad \kappa_2 = \max_{k=1,\ldots,k_0} \|\mathbf{\Sigma}_x(k)\|_2.$$

Thresholding CCVF: Let $\widehat{\Sigma}_y(k) = (\widehat{\sigma}_{i,j}(k))$ be CCVF. Define

$$\widetilde{\Sigma}_y(k) = \Big(\widehat{\sigma}_{i,j}(k) I\{|\widehat{\sigma}_{i,j}(k)| \ge u\}\Big), \qquad u = Mp^{2/l}n^{-1/2}.$$

Theorem 2. Let conditions C3-C5 hold, $\min_j \rho_j > 0$ and $p = o(n^{l/4})$. Then there exists an $\widehat{\mathbf{A}} = (\widehat{\mathbf{A}}_1, \cdots, \widehat{\mathbf{A}}_q)$ of which the columns are a permutation of the columns of $\widehat{\mathbf{\Gamma}}_y$, such that

$$\max_{1\le j\le q} \rho_j D(\mathcal{M}(\widehat{\mathbf{A}}_j), \mathcal{M}(\mathbf{A}_j))$$

$$= \begin{cases} O_p\big\{\kappa_2 (p^{4/l}n^{-1})^{(1-\iota)/2} S_{\max}\delta\big\}, & \kappa_1^{-1}(p^{4/l}n^{-1})^{(1-\iota)/2} S_{\max}\delta = O(1) \\ O_p\big\{(p^{4/l}n^{-1})^{1-\iota} S_{\max}^2 \delta^2\big\}, & \kappa_2 (p^{4/l}n^{-1})^{-(1-\iota)/2} S_{\max}^{-1}\delta^{-1} = O(1). \end{cases}$$

Remarks

(i) Theorem 2 gives the uniform convergence rate for $\rho_j D(\mathcal{M}(\widehat{\mathbf{A}}_j), \mathcal{M}(\mathbf{A}_j))$. The smaller ρ_j is, more difficult the estimation for $\mathcal{M}(\mathbf{A}_j)$ is.

(ii) The smaller ι, s_1, s_2 and $S_{\max}$ are, more sparse $\mathbf{\Sigma}_y(k)$ is, and the faster the convergences are.

(iii) Similar results can be obtained for the cases with $\log p = o(n^c)$ by assuming the sub-Gaussianity for $\mathbf{y}_t$ and exponential decay rates for α-mixing coefficients.

Simulation

Let $\mathbf{A} = (\mathbf{A}_1, \cdots, \mathbf{A}_q)$, $\mathbf{A}_j$ is $p \times p_j$, $\sum_j p_j = p$, and

$\widehat{\mathbf{A}} = (\widehat{\mathbf{A}}_1, \cdots, \widehat{\mathbf{A}}_{\widehat{q}})$, $\widehat{\mathbf{A}}_j$ is $p \times \widehat{p}_j$, $\sum_j \widehat{p}_j = p$.

Correct segmentation: $\widehat{q} = q$, $\widehat{p}_j = p_j$, and

$$D(\mathcal{M}(\mathbf{A}_j), \mathcal{M}(\widehat{\mathbf{A}}_j)) = \min_{1 \le i \le q} D(\mathcal{M}(\mathbf{A}_j), \mathcal{M}(\widehat{\mathbf{A}}_i)), \quad j = 1, \cdots, q.$$

for $\mathbf{H}_1'\mathbf{H}_1 = \mathbf{I}_{r_1}$ and $\mathbf{H}_2'\mathbf{H}_2 = \mathbf{I}_{r_2}$,

$$d(\mathcal{M}(\mathbf{H}_1), \mathcal{M}(\mathbf{H}_2)) = \left\{1 - \frac{1}{\min(r_1, r_2)} \text{tr}(\mathbf{H}_1\mathbf{H}_1'\mathbf{H}_2\mathbf{H}_2')\right\}^{1/2}.$$

Incomplete segmentation: $\widehat{q} < q$, and each $\mathcal{M}(\widehat{\mathbf{A}}_j)$ is an estimator for

the linear space spanned by *one*, or *more than one* $\mathbf{A}_i$

volatility

No. of replications: 500 times for each setting.

Let $\mathbf{A} = (a_{ij})$, $a_{ij} \sim U(-3, 3)$ independently.

Example 6. Let $p = 6$, $\mathbf{y}_t = \mathbf{A}\mathbf{x}_t$, and

$$x_{i,t} = \eta^{(1)}_{t+i-1} \ \text{ for } \ i = 1, 2, 3, \ \ x_{j,t} = \eta^{(2)}_{t+j-4} \ \text{ for } \ j = 4, 5, \ \ x_{6,t} = \eta^{(3)}_t.$$

where

$$\eta^{(1)}_t = 0.5\eta^{(1)}_{t-1} + 0.3\eta^{(1)}_{t-2} + e^{(1)}_t - 0.9e^{(1)}_{t-1} + 0.3e^{(1)}_{t-2} + 1.2e^{(1)}_{t-3} + 1.3e^{(1)}_{t-4},$$

$$\eta^{(2)}_t = -0.4\eta^{(2)}_{t-1} + 0.5\eta^{(2)}_{t-2} + e^{(2)}_t + e^{(2)}_{t-1} - 0.8e^{(2)}_{t-2} + 1.5e^{(2)}_{t-3},$$

$$\eta^{(3)}_t = 0.9\eta^{(3)}_{t-1} + e^{(3)}_t,$$

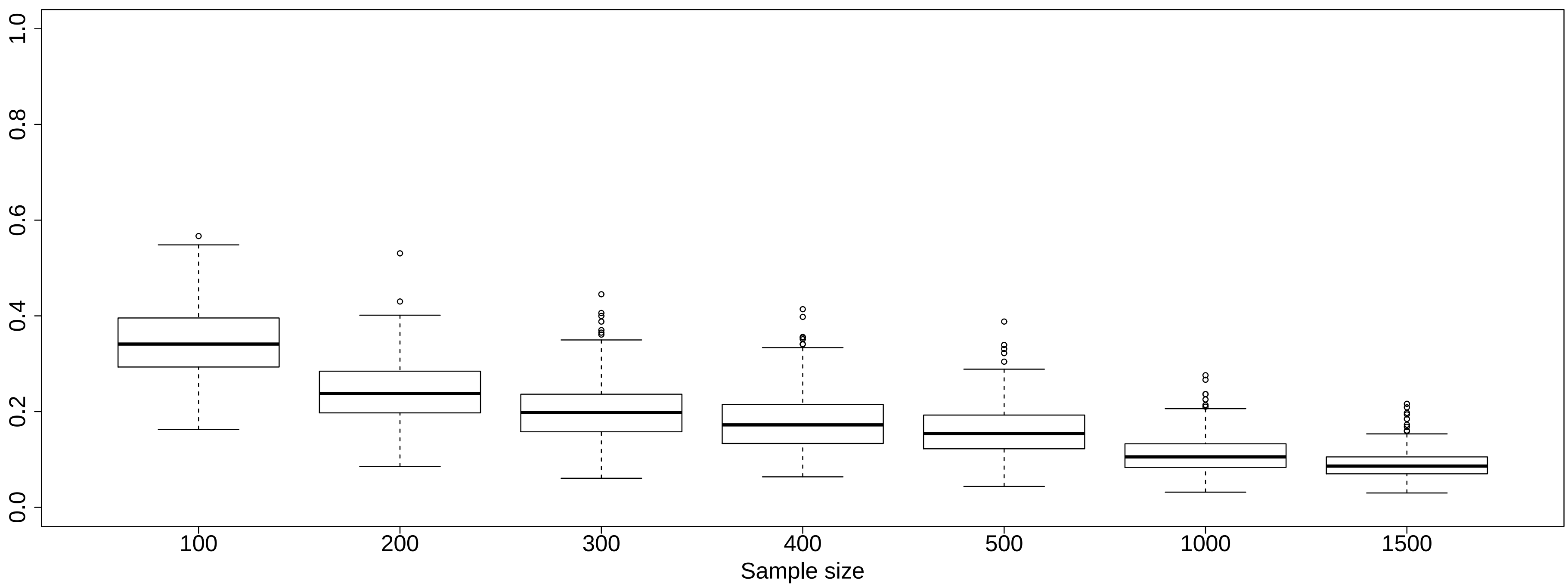
and $e^{(1)}_t, \ e^{(2)}_t, \ e^{(3)}_t$ are indep $N(0, 1)$. Thus

$q = 3$, i.e. 3 segmented subseries with $3, 2, 1$ elements.

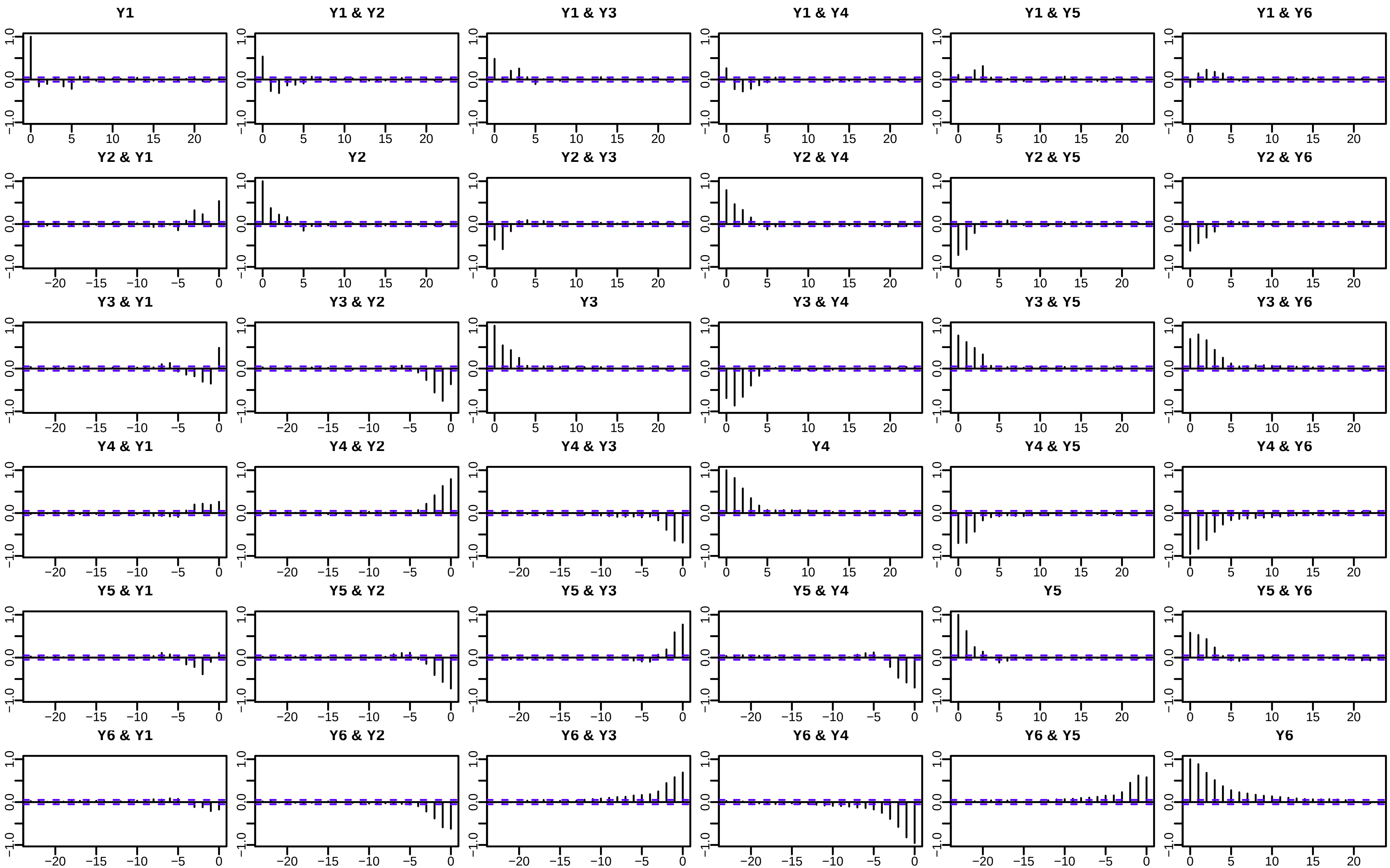
Relative frequencies of correct and incomplete segmentations

n	100	200	300	400	500	1000	1500	2000	2500	3000
Correct	.350	.564	.660	.712	.800	.896	.906	.920	.932	.945
Incomplete	.508	.386	.326	.282	.192	.104	.094	.080	.068	.055

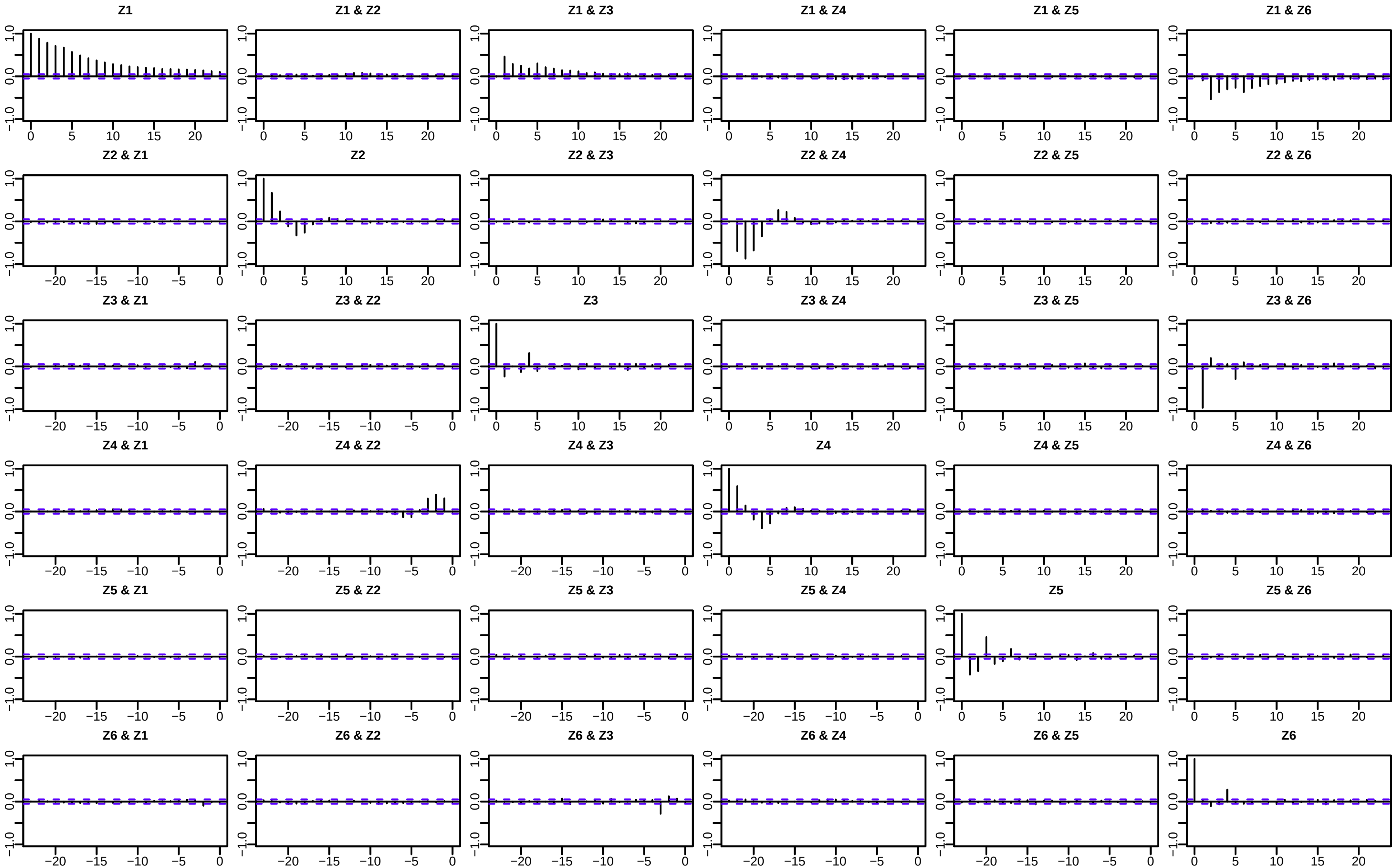
Boxplots of $\frac{1}{3}\sum_{1\le i\le 3} D(\mathcal{M}(\mathbf{A}_i), \mathcal{M}(\widehat{\mathbf{A}}_i))$ (with correct segmentations only)



CCF of y_t (one instance)



CCF of $\widehat{\mathbf{x}}_t$ (one instance)



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

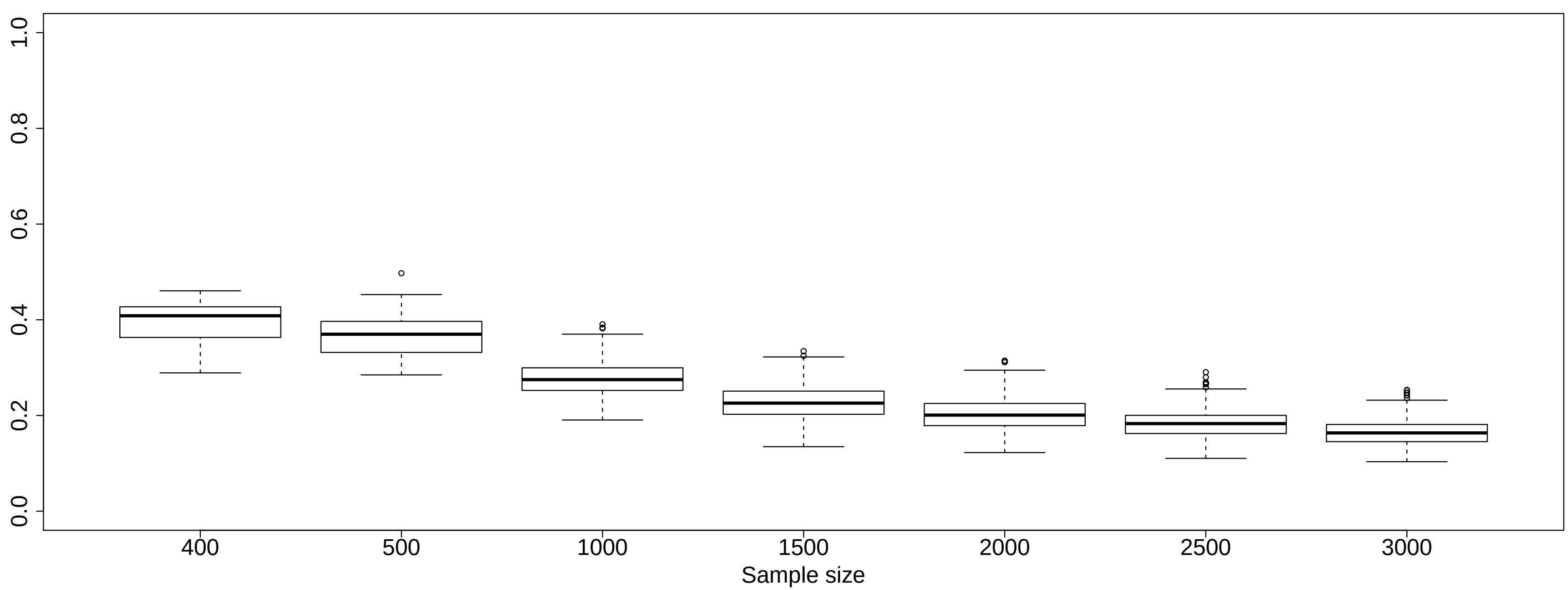
Example 7. Let $p = 20,\ q = 5,\ (p_1, \cdots, p_5) = (6, 5, 4, 3, 2)$

$\mathbf{x}_t$ is defined similarly as in Example 6.

Relative frequencies of correct and incomplete segmentations

n	400	500	1000	1500	2000	2500	3000
Correct	0.058	0.118	0.492	0.726	0.862	0.902	0.940
Incomplete	0.516	0.672	0.460	0.258	0.130	0.096	0.060

Boxplots of $\frac{1}{5}\sum_{1\le i\le 5} D(\mathcal{M}(\mathbf{A}_i), \mathcal{M}(\widehat{\mathbf{A}}_i))$ (with correct segmentations only)



Segmenting multiple volatility processes

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

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

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

Let $\text{Var}(\mathbf{y}_t) = \text{Var}(\mathbf{x}_t) = \mathbf{I}_p$, 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 $B \in \mathcal{B}_{t-1}$,

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

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*.

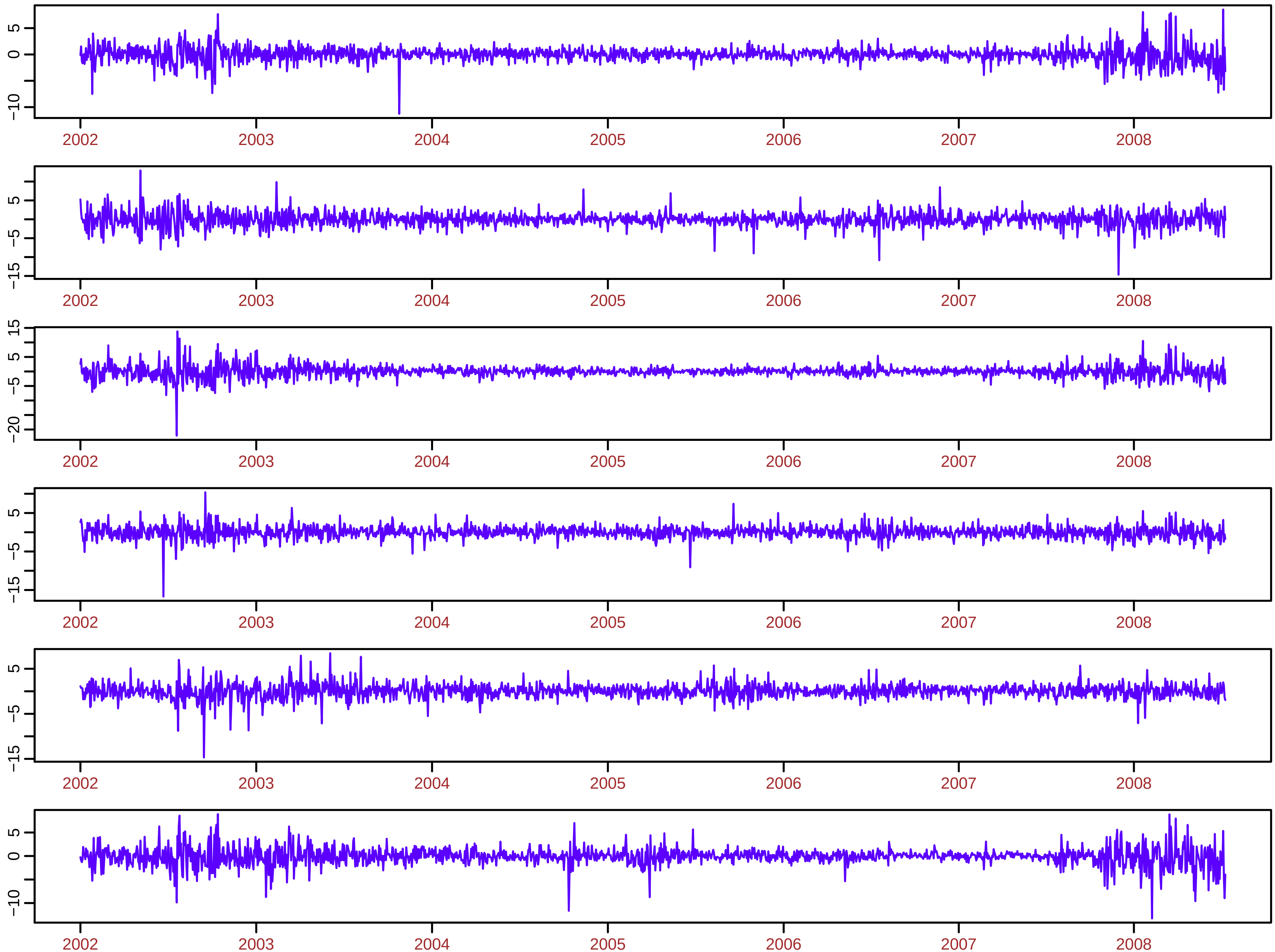
$n = 1642$, $p = 6$

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

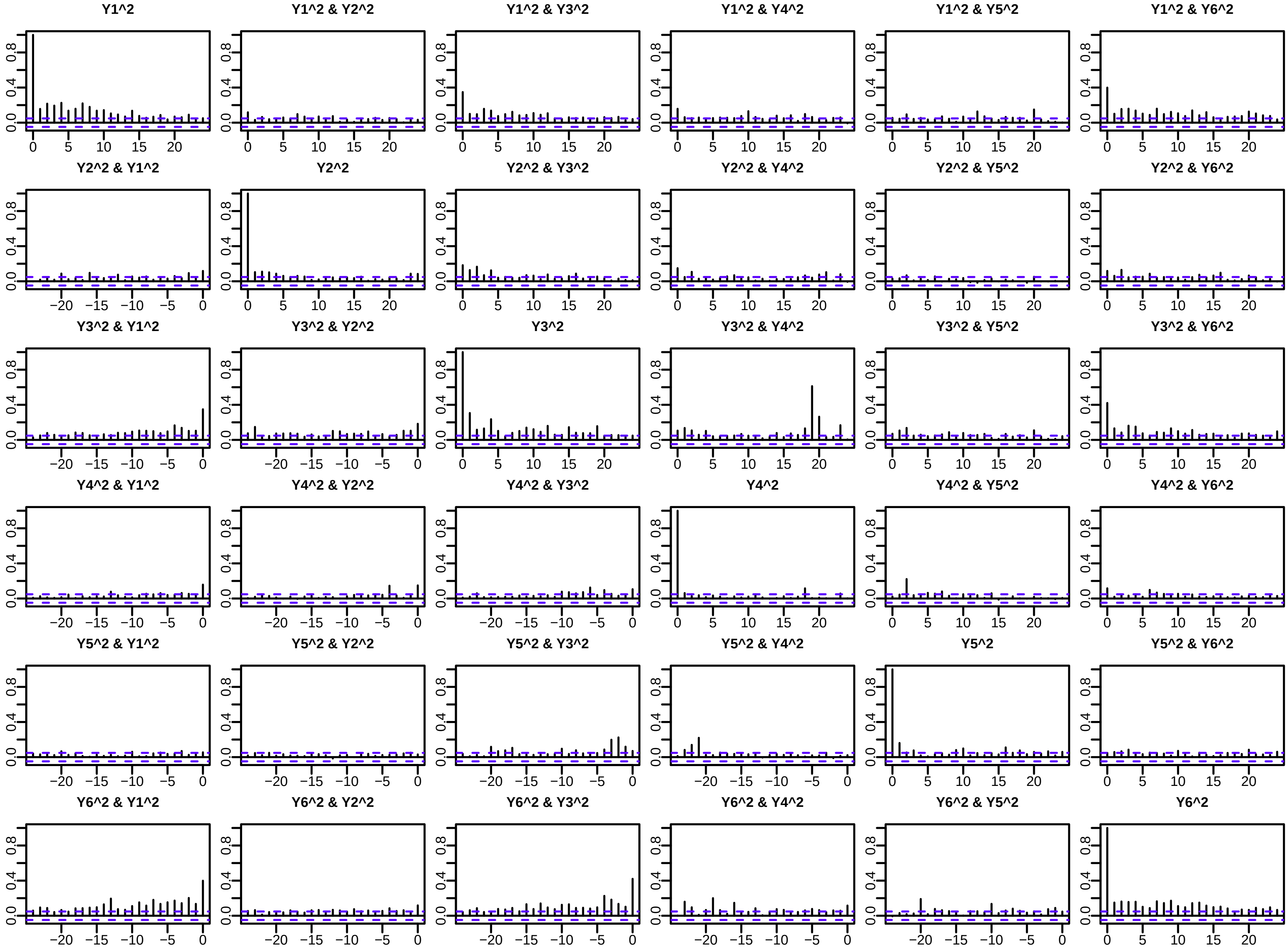
Segmentation for transformed series:

CUC of Fan, Wang & Y (2008)

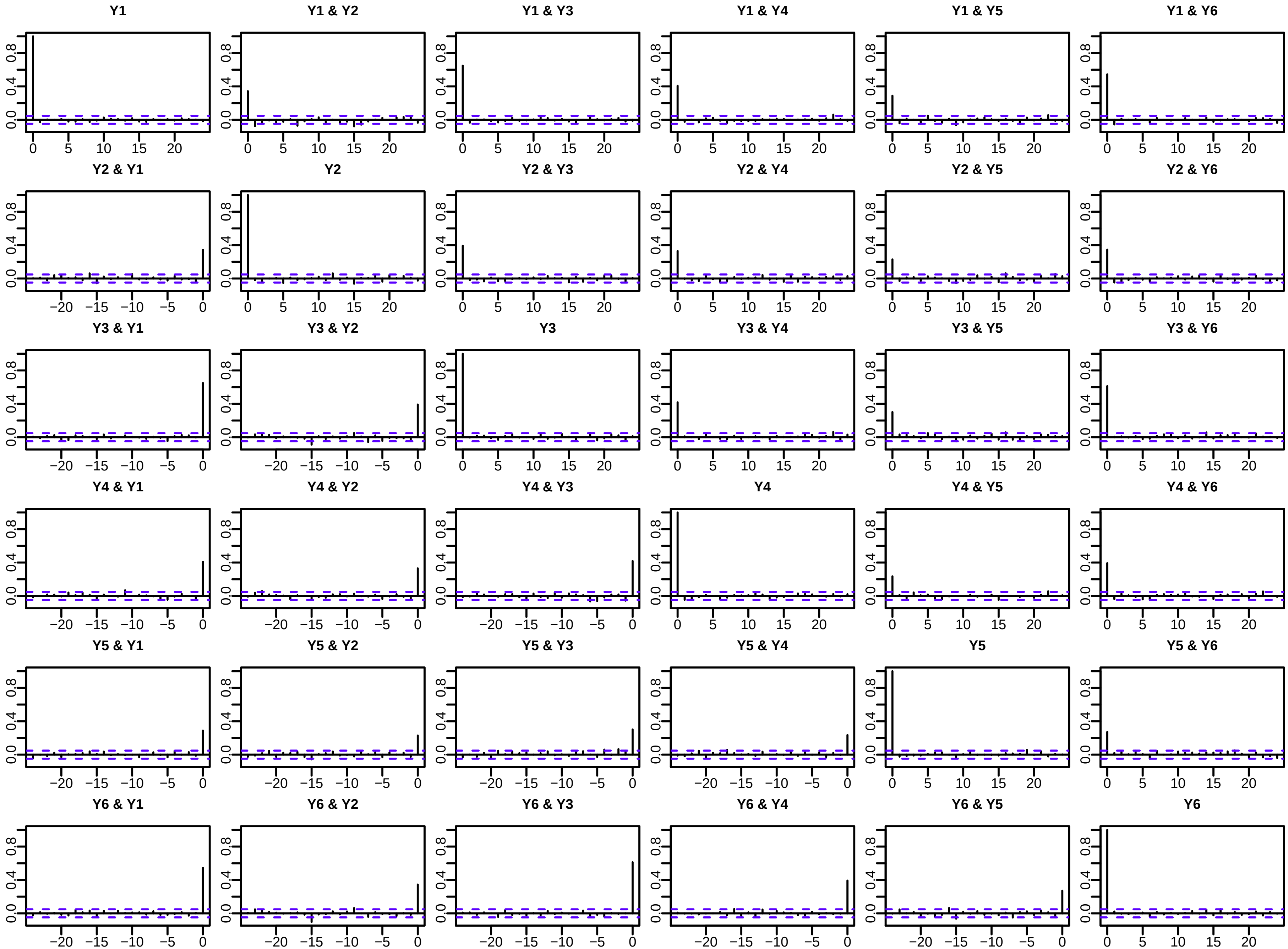
Daily returns of 6 stocks



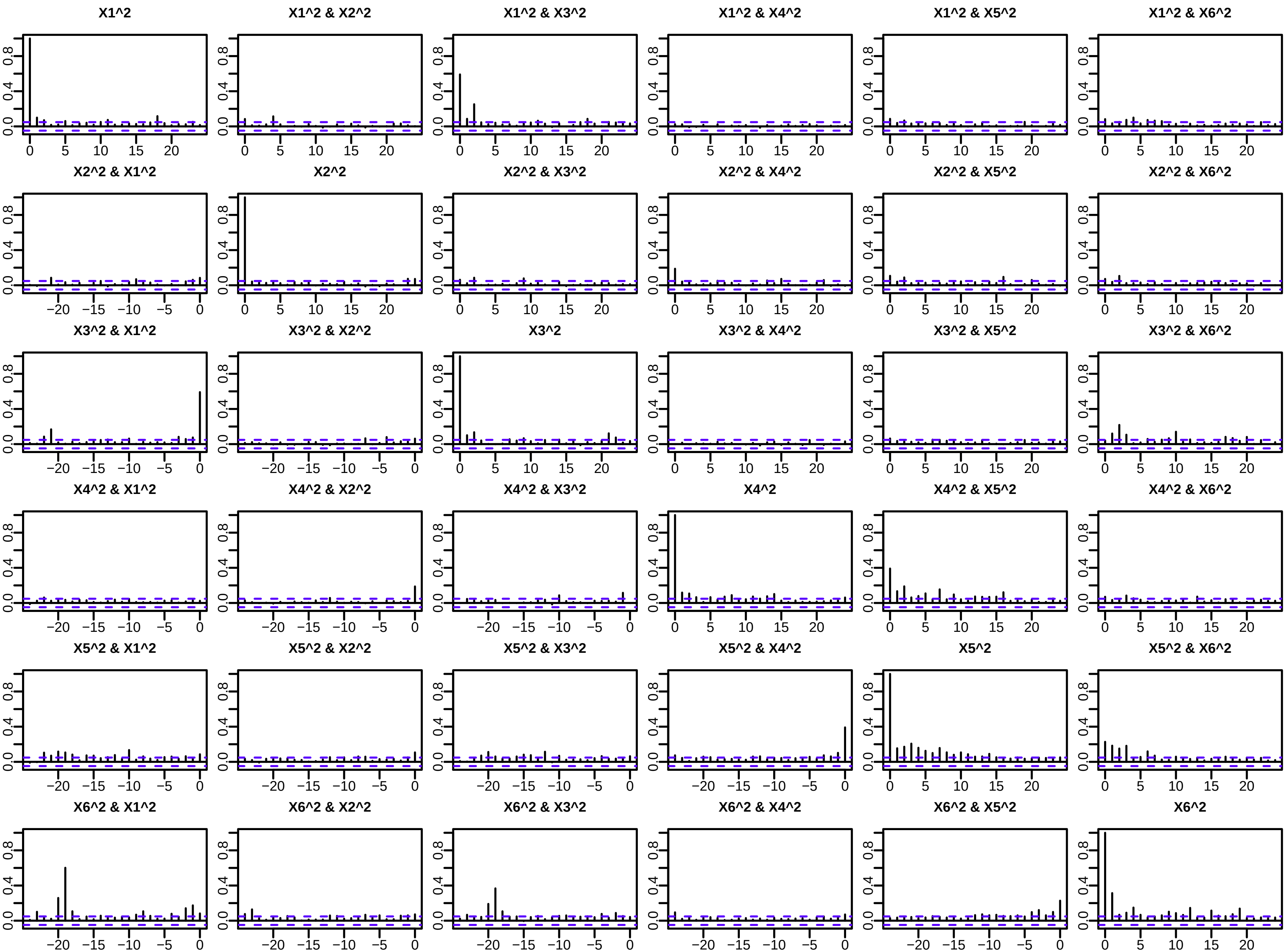
CCF of squared returns



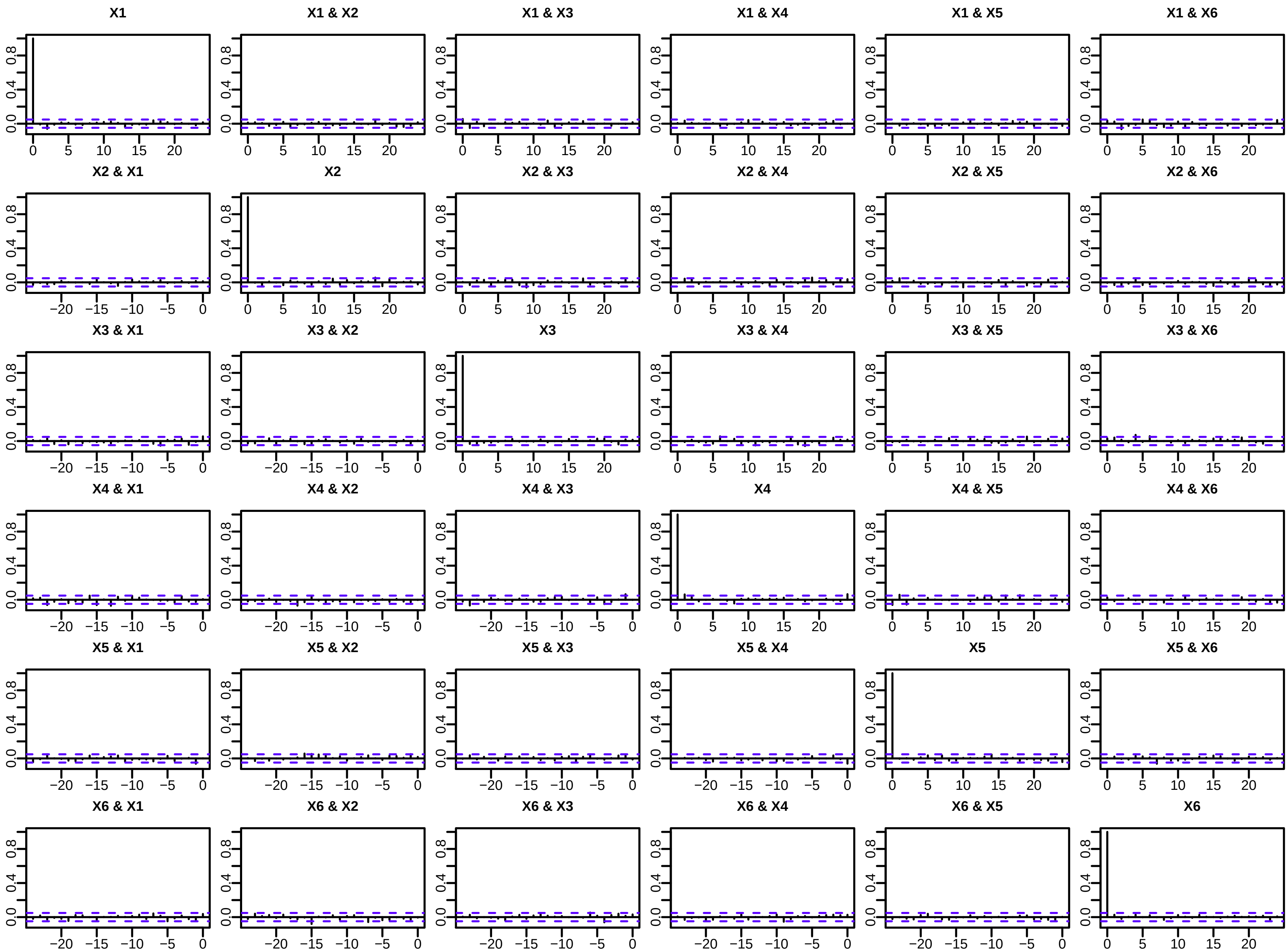
CCF of residuals from fitted GARCH(1,1) for each return series



CCF of squared transformed returns



CCF of residuals from transformed return series



Conclusions

- A new version of PCA for time series: finding a latent segmentation via a contemporaneous linear transformation
- When the segmentation does not exist, provide effective approximations by ignoring negligible, though significant, correlations
- Why works? $\mathbf{W}_y = \sum_k \boldsymbol{\Sigma}_y(k)\boldsymbol{\Sigma}_y(k)' = \mathbf{A}\mathbf{W}_x\mathbf{A}'$,

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

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

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

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

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

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

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

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

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