Principal Component Analysis for Time Series: Segmentation via Contemporaneous Linear Transformation

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

Simulation

Segmenting multiple volatility processes
**Goal:** For $p \times 1$ weakly stationary time series $y_t$, search for a *contemporaneous* linear transformation:

$$y_t = Ax_t, \quad \text{or} \quad x_t = By_t \quad (\text{i.e. } B = A^{-1})$$

such that

$$x_t = \begin{pmatrix} x_t^{(1)} \\ \vdots \\ x_t^{(q)} \end{pmatrix}, \quad \text{Cov}(x_t^{(i)}, x_s^{(j)}) = 0 \quad \forall \ i \neq j \quad \text{and} \quad t, s.$$

Hence, $x_t^{(1)}, \cdots, x_t^{(q)}$ can be modelled separately!
**Goal:** For $p \times 1$ weakly stationary time series $y_t$, search for a contemporaneous linear transformation:

$$y_t = Ax_t, \quad \text{or} \quad x_t = By_t \quad (i.e. \ B = A^{-1})$$

such that

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

Hence, $x_t^{(1)}, \ldots, x_t^{(q)}$ can be modelled separately!

- realistic?  
- how to find $B$ and $x_t$?
Observations: \( y_1, \cdots, y_n \) from a \( p \times 1 \) weakly stationary TS

Assumption: \( y_t = A x_t \), and

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

Without loss of generality: \( \text{Var}(y_t) = \text{Var}(x_t) = I_p \), and thus

\[
A' A = I_p, \quad i.e. \ A \text{ is orthogonal}, \quad x_t = A'y_t
\]

Goal: estimate \( A = (A_1, \cdots, A_q) \), or more precisely, \( M(A_1), \cdots, M(A_q) \), as

\[
\hat{x}^{(j)}_t = \hat{A}'_j y_t, \quad j = 1, \cdots, q.
\]

Note. \((A, x_t)\) can be replaced by \((AH, H'x_t)\) for any \( H = \text{diag}(H_1, \cdots, H_q) \) with \( H'_j H_j = I_{p_j} \)
Step 1: Transformation via eigenanalysis

Notation: \( \Sigma_y(k) = \text{Cov}(y_{t+k}, y_t), \quad \Sigma_x(k) = \text{Cov}(x_{t+k}, x_t), \)

\[ W_y = \sum_{k=0}^{k_0} \Sigma_y(k) \Sigma_y(k)' = I_p + \sum_{k=1}^{k_0} \Sigma_y(k) \Sigma_y(k)', \]

\[ W_x = \sum_{k=0}^{k_0} \Sigma_x(k) \Sigma_x(k)' = I_p + \sum_{k=1}^{k_0} \Sigma_x(k) \Sigma_x(k)' . \]
Step 1: Transformation via eigenanalysis

Notation: \( \Sigma_y(k) = \text{Cov}(y_{t+k}, y_t), \quad \Sigma_x(k) = \text{Cov}(x_{t+k}, x_t), \)

\[
W_y = \sum_{k=0}^{k_0} \Sigma_y(k) \Sigma_y(k)' = I_p + \sum_{k=1}^{k_0} \Sigma_y(k) \Sigma_y(k)',
\]

\[
W_x = \sum_{k=0}^{k_0} \Sigma_x(k) \Sigma_x(k)' = I_p + \sum_{k=1}^{k_0} \Sigma_x(k) \Sigma_x(k)'.
\]

As \( \Sigma_y(k) = A \Sigma_x(k) A' \), \( W_y = AW_x A' \).
Step 1: Transformation via eigenanalysis

Notation:  \( \Sigma_y(k) = \text{Cov}(y_{t+k}, y_t) \),  \( \Sigma_x(k) = \text{Cov}(x_{t+k}, x_t) \),

\[
W_y = \sum_{k=0}^{k_0} \Sigma_y(k) \Sigma_y(k)' = I_p + \sum_{k=1}^{k_0} \Sigma_y(k) \Sigma_y(k)',
\]

\[
W_x = \sum_{k=0}^{k_0} \Sigma_x(k) \Sigma_x(k)' = I_p + \sum_{k=1}^{k_0} \Sigma_x(k) \Sigma_x(k)'.
\]

As \( \Sigma_y(k) = A \Sigma_x(k) A' \), \( W_y = AW_x A' \).

Eigenanalysis: \( W_x \Gamma_x = \Gamma_x D \), columns of \( \Gamma_x \) are eigenvectors of \( W_x \) with the eigenvalues in diagonal matrix \( D \).

\[
W_y A \Gamma_x = AW_x A' A \Gamma_x = AW_x \Gamma_x = A \Gamma_x D
\]

Thus \( \Gamma_y = A \Gamma_x \), and \( \Gamma_y y_t = \Gamma_x A' y_t = \Gamma_x x_t \).
\( z_t \equiv \Gamma'_y y_t (= \Gamma'_x x_t) \) is the required transformation upto a permutation.
$z_t \equiv \Gamma'_y y_t (= \Gamma'_x x_t)$ is the required transformation up to a permutation.

Note $W_x = \text{diag}(W_{x,1}, \cdots, W_{x,q})$

**Proposition 1.** (i) $\Gamma_x$ can be taken with the same block-diagonal structure as $W_x$.

(ii) Any $\Gamma_x$ is a column-permutation of $\Gamma_x$ described in (i), provided

$$\lambda(W_{x,i}) \neq \lambda(W_{x,j}) \text{ for any } i \neq j.$$
\[ z_t \equiv \Gamma'_y y_t (= \Gamma'_x x_t) \] is the required transformation upto a permutation

Note \( W_x = \text{diag}(W_{x,1}, \cdots, W_{x,q}) \)

**Proposition 1.** (i) \( \Gamma_x \) can be taken with the same block-diagonal structure as \( W_x \).

(ii) Any \( \Gamma_x \) is a column-permutation of \( \Gamma_x \) described in (i), provided

\[ \lambda(W_{x,i}) \neq \lambda(W_{x,j}) \text{ for any } i \neq j. \]

For \( \Gamma_x \) specified in (i), \( \Gamma'_x x_t \) is segmented exactly the same as \( x_t \).
\( z_t \equiv \Gamma'_{y} y_t (= \Gamma'_{x} x_t) \) is the required transformation upto a permutation

Note \( W_x = \text{diag}(W_{x,1}, \cdots, W_{x,q}) \)

**Proposition 1.** (i) \( \Gamma_x \) can be taken with the same block-diagonal structure as \( W_x \).

(ii) Any \( \Gamma_x \) is a column-permutation of \( \Gamma_x \) described in (i), provided \( \lambda(W_{x,i}) \neq \lambda(W_{x,j}) \) for any \( i \neq j \).

For \( \Gamma_x \) specified in (i), \( \Gamma'_{x} x_t \) is segmented exactly the same as \( x_t \)

Let
\[
\hat{W}_y = I_p + \sum_{k=1}^{k_0} \hat{\Sigma}_y(k) \hat{\Sigma}_y(k)', \quad \hat{W}_y \hat{\Gamma}_y = \hat{\Gamma}_y \hat{D}.
\]

Then \( \hat{z}_t = \hat{\Gamma}'_y y_t \) – require permute components of \( \hat{z}_t \) to obtain \( \hat{x}_t \)
**Permutation**

**Goal:** put the connected components of $\hat{z}_t = \hat{\Gamma}_y y_t$ together.

Visual examination of CCF if $p$ is not large!

Two component series of $\hat{z}_t$ 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 $\hat{z}_t$.

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.

– p.7
Permutation Method I: Max CCF

Put $\hat{z}_t = (\hat{z}_{1,t}, \ldots, \hat{z}_{p,t})'$.

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

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

reject $H_0$ for the pair $(\hat{z}_{i,t}, \hat{z}_{j,t})$ for large values of $\hat{L}_n(i,j)$. 
Permutation Method I: Max CCF

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

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

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

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

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

$$\hat{L}_1 \geq \cdots \geq \hat{L}_{p_0}, \quad p_0 = \frac{p(p - 1)}{2}.$$

Define

$$\hat{r} = \arg \max_{1 \leq j < c_0 p_0} \frac{\hat{L}_j}{\hat{L}_{j+1}}, \quad c_0 \in (0, 1).$$

Reject $H_0$ for the pairs corresponding to $\hat{L}_1, \cdots, \hat{L}_{\hat{r}}$. 
Graph representation: Let vertexes $\hat{V} = \{1, \cdots , p\}$ stand for the components of $\hat{z}_t = \hat{\Gamma}'y_t$, and $\hat{E}_n = \{\text{edge connecting } i \text{ and } j : \hat{z}_{i,t}, \hat{z}_{j,t} \text{ are connected}\}$.

Let $V = \{1, \cdots , p\}$ stand for the components of $z_t = \Gamma'y_t$, and $E = \{\text{edge connecting } i \text{ and } j : \max_{-m \leq h \leq m} |\text{Corr}(z_{i,t+h}, z_{j,t})| > 0 \}$. 
Graph representation: Let vertexes $\hat{V} = \{1, \cdots, p\}$ stand for the components of $\hat{z}_t = \hat{\Gamma}'y_t$, and

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

Let $V = \{1, \cdots, p\}$ stand for the components of $z_t = \Gamma'y_t$, and

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

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

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

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

Proposition 2. As $n \to \infty$, $\delta_n \to 0$, $\frac{1}{\varpi}\|\hat{W}_y - W_y\|_2 = o(\delta_n)$, and

$$\log p = o(n^{\alpha}). \text{ Then } P(\hat{E}_n = E) \to 1.$$
Prewhitening

To make CCF for different pairs comparable, prewhiten each component series of $\hat{z}_t = \hat{\Gamma}_y y_t$ separately first.
Prewhitening
To make CCF for different pairs comparable, prewhiten each component series of \( \hat{z}_t = \hat{\Gamma}'_y y_t \) separately first.

(i) If \( \rho_{i,j}(h) = 0 \), \( \hat{\rho}_{i,j}(h) \sim N(0, 1/n) \) asymptotically, provided at least one of \( x_{i,t} \) and \( x_{j,t} \) is white noise.

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

and \( \text{Cov}\{\hat{\rho}_{i,j}(h), \hat{\rho}_{i,j}(k)\} = o_P(1/n) \), 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 \( \hat{z}_t = \hat{\Gamma}' y_t \) separately first.

(i) If \( \rho_{i,j}(h) = 0 \), \( \hat{\rho}_{i,j}(h) \sim N(0, 1/n) \) asymptotically, provided at least one of \( x_{i,t} \) and \( x_{j,t} \) is white noise.

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

and \( \text{Cov}\{\hat{\rho}_{i,j}(h), \hat{\rho}_{i,j}(k)\} = o_P(1/n) \), 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 \( \hat{z}_t \) by fitting an AR with the order determined by AIC and not greater than 5.
To fix the idea, let $\xi_t$, $\eta_t$ be two WN, $\rho(k) = \text{Corr}(\xi_{t+k}, \eta_t) = 0$,

\[
\hat{\rho}(k) = \sum_{t=1}^{n-k} (\xi_{t+k} - \bar{\xi})(\eta_t - \bar{\eta}) / \left\{ \sum_{t=1}^{n} (\xi_t - \bar{\xi})^2 \sum_{t=1}^{n} (\eta_t - \bar{\eta})^2 \right\}^{1/2}.
\]

Asymptotically $\hat{\rho}(k) \sim \mathcal{N}(0, 1/n)$, $\hat{\rho}(k)$ and $\hat{\rho}(h)$ are independent.
Permutation Method II: FDR based on multiple tests

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

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

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

The $P$-value for testing $\rho(k) = 0$ (simple) is $p_k = 2\Phi(-\sqrt{n}|\hat{\rho}(k)|)$
Permutation Method II: FDR based on multiple tests

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

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

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

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

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

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

$$p(j) \leq j\alpha/(2m + 1)$$

for at least one $1 \leq j \leq 2m + 1$
Permutation Method II: FDR based on multiple tests

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

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

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

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

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

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

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

The $P$-value for the multiple test is

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

$$= \min_{1 \leq j \leq 2m+1} p(j) (2m + 1)/j.$$
For each pair components of \( \hat{z}_t = \hat{\Gamma}' y_t \), 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:

\[
P_{(1)} \leq P_{(2)} \leq \cdots \leq P_{(p_0)}, \quad p_0 = p(p - 1)/2
\]

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

\[
\hat{d} = \max\{k : 1 \leq k \leq p_0, \ P_{(k)} \leq k\beta/p_0\},
\]

and rejects the hypothesis \( H_0 \) for the \( \hat{d} \) pairs of the components of \( z_t \) corresponding to the \( P \)-values \( P_{(1)}, \cdots, P_{(\hat{d})} \).
For each pair components of $\hat{z}_t = \hat{\Gamma}'_y y_t$, 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:

$$P(1) \leq P(2) \leq \cdots \leq P(p_0), \quad p_0 = p(p - 1)/2$$

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

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

and rejects the hypothesis $H_0$ for the $\hat{d}$ pairs of the components of $z_t$ corresponding to the $P$-values $P(1), \cdots, P(\hat{d})$

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

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

(iii) Ranking the pair components of $\hat{z}_t$ according to dependence strength.
Real data examples

CCF of monthly temperatures in 7 cities
Transformation: $\hat{x}_t = \hat{B}y_t$

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

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

$n = 540, p = 7$

Used $k_0 = 5$ (in defining $W_y$), hardly changed for $2 \leq k_0 \leq 36$. 

- p.16
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
Post-Sample Forecast

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

Compare with the forecasts based on fitting a VAR and a restricted VAR (RVAR) directly to $y_t$ (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 up to 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.

<table>
<thead>
<tr>
<th></th>
<th>One-step MSE</th>
<th>Two-step MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAR</td>
<td>2.470 (0.416)</td>
<td>2.559 (0.385)</td>
</tr>
<tr>
<td>RVAR</td>
<td>2.530 (0.398)</td>
<td>2.615 (0.382)</td>
</tr>
<tr>
<td>Segmentation (5 groups)</td>
<td>2.221 (0.339)</td>
<td>2.203 (0.323)</td>
</tr>
<tr>
<td>Segmentation (6 groups)</td>
<td>2.417 (0.348)</td>
<td>2.419 (0.326)</td>
</tr>
<tr>
<td>Segmentation (4 groups)</td>
<td>2.421 (0.343)</td>
<td>2.422 (0.325)</td>
</tr>
</tbody>
</table>

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

Nonstationary trends: difference each series
CCF of 8 differenced monthly US Industrial Indices
Transformation: \( \hat{x}_t = \hat{B}y_t \)

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

\[ 1 \leq m \leq 20: \quad \{1, 3\} \]

• Permutation with FDR

\[ m = 20 \text{ and } \beta \in [10^{-6}, 0.01], \text{ or } m = 5 \text{ and } \beta \in [10^{-6}, 0.001]: \]
\[ \{1, 3\} \]

\[ m = 5 \text{ and } \beta = 0.005: \quad \{1, 2, 3\}, \{4, 8\} \]

\[ m = 5 \text{ and } \beta = 0.01: \quad \{1, 2, 3, 5, 6, 7\} \text{ and } \{4, 8\}. \]

Two recommended groupings:

seven groups: \( \{1, 3\} \)

five groups: \( \{1, 2, 3\}, \{4, 8\} \)
**Post-sample forecast**


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

Miss some small but significant cross correlations

<table>
<thead>
<tr>
<th></th>
<th>One-step MSE</th>
<th>Two-step MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAR</td>
<td>(0.615_{(1.349)})</td>
<td>(1.168_{(2.129)})</td>
</tr>
<tr>
<td>RVAR</td>
<td>(0.606_{(1.293)})</td>
<td>(1.159_{(2.285)})</td>
</tr>
<tr>
<td>Segmentation</td>
<td>(0.588_{(1.341)})</td>
<td>(1.154_{(2.312)})</td>
</tr>
</tbody>
</table>
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.

\[ n = 937, \quad p = 7. \]
Weekly recorded cases of measles in 7 cities in England in 1948-1965
CCF of weekly recorded cases of measles in 7 cities
Transformation: \( \hat{x}_t = \hat{B}y_t \)

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

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 $\pm 1.96/\sqrt{n} = .064$ 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.

<table>
<thead>
<tr>
<th>No. of connected pairs</th>
<th>No. of groups</th>
<th>Segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>{4, 5}, {1}, {2}, {3}, {6}, {7}</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>{1, 2}, {4, 5}, {3}, {6}, {7}</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>{1, 2, 3}, {4, 5}, {6}, {7}</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>{1, 2, 3, 7}, {4, 5}, {6}</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>{1, 2, 3, 6, 7}, {4, 5}</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>{1, \ldots, 7}</td>
</tr>
</tbody>
</table>
**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\}

<table>
<thead>
<tr>
<th></th>
<th>One-step MSE</th>
<th>Two-step MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Univariate AR</td>
<td>551.386(1322.345)</td>
<td>931.766(3115.508)</td>
</tr>
<tr>
<td>VAR</td>
<td>503.408(1124.213)</td>
<td>719.499(2249.986)</td>
</tr>
<tr>
<td>RVAR</td>
<td>574.582(1432.217)</td>
<td>846.141(2462.019)</td>
</tr>
<tr>
<td>Segmentation (4 groups)</td>
<td>472.106(1088.170)</td>
<td>654.843(1807.502)</td>
</tr>
<tr>
<td>Segmentation (7 groups)</td>
<td>510.825(1134.615)</td>
<td>798.374(2451.896)</td>
</tr>
<tr>
<td>Segmentation (3 groups)</td>
<td>497.025(1206.745)</td>
<td>677.125(1902.230)</td>
</tr>
</tbody>
</table>
Example 4. Daily impressions of 32 keywords for an air ticket booking website powered by Baidu (www.baidu.com) over 130 days. 

\[ n = 130, \quad p = 32 \]

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 \( 1 \leq m \leq 15 \)), 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)

\[ n = 1812, \ p = 25 \]

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

<table>
<thead>
<tr>
<th></th>
<th>One-step MSE</th>
<th>Two-step MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Univariate AR</td>
<td>0.208 (0.551)</td>
<td>0.194 (0.539)</td>
</tr>
<tr>
<td>VAR</td>
<td>0.295 (0.806)</td>
<td>0.301 (0.855)</td>
</tr>
<tr>
<td>RVAR</td>
<td>0.293 (0.820)</td>
<td>0.296 (0.863)</td>
</tr>
<tr>
<td>Segmentation (24 groups)</td>
<td>0.153 (0.134)</td>
<td>0.163 (0.124)</td>
</tr>
<tr>
<td>Segmentation (25 groups)</td>
<td>0.110 (0.084)</td>
<td>0.132 (0.091)</td>
</tr>
<tr>
<td>Segmentation (23 groups)</td>
<td>0.151 (0.133)</td>
<td>0.159 (0.121)</td>
</tr>
</tbody>
</table>

Cross correlations between provinces are useful information.

However they cannot be used via direct VAR!
Example 5. Log daily $\text{PM}_{2.5}$ concentration readings at 84 monitoring stations in Beijing, Tianjin and Hebei in 1 Jan 2015 – 31 Dec 2016.

$\text{PM}_{2.5}$ consists of airborne particles with aerodynamic diameters smaller than $2.5 \mu\text{m}$.

$n = 731$ and $p = 84$
Cross 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.

<table>
<thead>
<tr>
<th></th>
<th>One-step MSE</th>
<th>Two-step MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Univariate AR</td>
<td>0.525(0.204)</td>
<td>0.835(0.284)</td>
</tr>
<tr>
<td>Segmentation (83 groups)</td>
<td>0.485(0.185)</td>
<td>0.662(0.224)</td>
</tr>
<tr>
<td>Segmentation (84 groups)</td>
<td>0.484(0.184)</td>
<td>0.662(0.224)</td>
</tr>
<tr>
<td>Segmentation (50 groups)</td>
<td>0.492(0.187)</td>
<td>0.678(0.228)</td>
</tr>
<tr>
<td>Segmentation (70 groups)</td>
<td>0.474(0.180)</td>
<td>0.664(0.225)</td>
</tr>
</tbody>
</table>

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

For \( p \times r \) matrices \( H_1 \) and \( H_2 \) and \( H_1' H_1 = H_2' H_2 = I_r \), let

\[
D(\mathcal{M}(H_1), \mathcal{M}(H_2)) = \sqrt{1 - \frac{1}{r} \text{tr}(H_1 H_1' H_2 H_2')}.
\]

Then \( D(\mathcal{M}(H_1), \mathcal{M}(H_2)) \in [0, 1] \).

It equals 0 iff \( \mathcal{M}(H_1) = \mathcal{M}(H_2) \), and 1 iff \( \mathcal{M}(H_1) \perp \mathcal{M}(H_2) \).

\( y_t \) is assumed to be weakly stationary and \( \alpha \)-mixing, i.e.

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

where \( \mathcal{F}_{i}^{j} = \sigma(y_t : i \leq t \leq j) \).
\( p \) fixed

C1. For some constant \( \gamma > 2 \),

\[
\sup_t \max_{1 \leq i \leq p} E(|y_{i,t} - E y_{i,t}|^{2\gamma}) < \infty.
\]

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

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

\[
\max_{1 \leq j \leq q} \mathbb{D}(\mathcal{M}(\hat{A}_j), \mathcal{M}(A_j)) = O_p(n^{-1/2}).
\]
\[ p = o(n^c) \text{ for some } c > 0 \]

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

\[
\max_{1 \leq j \leq p} \sum_{i=1}^{p} |a_{i,j}|^{\iota} \leq s_1 \quad \text{and} \quad \max_{1 \leq i \leq p} \sum_{j=1}^{p} |a_{i,j}|^{\iota} \leq s_2,
\]

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

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

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

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

Remark. As \( \Sigma_y(k) = \mathbf{A}\Sigma_x(k)\mathbf{A}' \), the sparsity of \( \mathbf{A} \) and the maximum block size of \( \Sigma_x(k) \) provide a measure for the sparsity of \( \Sigma_y(k) \).
Let $S_{\text{max}}$ be the maximum block size in $\Sigma_x(k)$, and

$$
\rho_j = \min_{1 \leq i \leq q} \min_{i \neq j} |\lambda(W_{x,i}) - \lambda(W_{x,j})|, \quad j = 1, \ldots, q,
$$

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

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

$$
\tilde{\Sigma}_y(k) = \left(\hat{\sigma}_{i,j}(k) I\{|\hat{\sigma}_{i,j}(k)| \geq u\}\right), \quad u = Mp^{2/l} n^{-1/2}.
$$

**Theorem 2.** Let conditions C3-C5 hold, $\min_j \rho_j > 0$ and $p = o(n^{1/4})$. Then there exists an $\hat{\Gamma} = (\hat{\Gamma}_1, \ldots, \hat{\Gamma}_q)$ of which the columns are a permutation of the columns of $\hat{\Gamma}_y$, such that

$$
\max_{1 \leq j \leq q} \rho_j D(M(\hat{A}_j), M(A_j))
$$

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

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

(ii) The smaller $\iota, s_1, s_2$ and $S_{\text{max}}$ are, more sparse $\Sigma_y(k)$ 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 $y_t$ and exponential decay rates for $\alpha$-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

\[
\hat{\mathbf{A}} = (\hat{\mathbf{A}}_1, \cdots, \hat{\mathbf{A}}_q), \quad \hat{\mathbf{A}}_j \text{ is } p \times \hat{p}_j, \quad \sum_j \hat{p}_j = p.
\]

Correct segmentation: \( \hat{q} = q \), \( \hat{p}_j = p_j \), and

\[
D(\mathcal{M}(\mathbf{A}_j), \mathcal{M}(\hat{\mathbf{A}}_j)) = \min_{1 \leq i \leq q} D(\mathcal{M}(\mathbf{A}_j), \mathcal{M}(\hat{\mathbf{A}}_i)), \quad j = 1, \cdots, q.
\]

for \( H_1' H_1 = I_{r_1} \) and \( H_2' H_2 = I_{r_2} \),

\[
d(\mathcal{M}(H_1), \mathcal{M}(H_2)) = \left\{ 1 - \frac{1}{\min(r_1, r_2)} \text{tr}(H_1 H_1' H_2 H_2') \right\}^{1/2}.
\]

Incomplete segmentation: \( \hat{q} < q \), and each \( \mathcal{M}(\hat{\mathbf{A}}_j) \) is an estimator for the linear space spanned by one, or more than one \( \mathbf{A}_i \).
No. of replications: 500 times for each setting.

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

Example 6. Let $p = 6$, $y_t = Ax_t$, and

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

where

$$
\eta_{t}^{(1)} = 0.5\eta_{t-1}^{(1)} + 0.3\eta_{t-2}^{(1)} + e_t^{(1)} - 0.9e_{t-1}^{(1)} + 0.3e_{t-2}^{(1)} + 1.2e_{t-3}^{(1)} + 1.3e_{t-4}^{(1)},
$$
$$
\eta_{t}^{(2)} = -0.4\eta_{t-1}^{(2)} + 0.5\eta_{t-2}^{(2)} + e_t^{(2)} + e_{t-1}^{(2)} - 0.8e_{t-2}^{(2)} + 1.5e_{t-3}^{(2)},
$$
$$
\eta_{t}^{(3)} = 0.9\eta_{t-1}^{(3)} + e_t^{(3)},
$$

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

$$
q = 3, \quad \text{i.e. 3 segmented subseries with 3, 2, 1 elements.}
$$
Relative frequencies of correct and incomplete segmentations

<table>
<thead>
<tr>
<th>$n$</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>1000</th>
<th>1500</th>
<th>2000</th>
<th>2500</th>
<th>3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>.350</td>
<td>.564</td>
<td>.660</td>
<td>.712</td>
<td>.800</td>
<td>.896</td>
<td>.906</td>
<td>.920</td>
<td>.932</td>
<td>.945</td>
</tr>
<tr>
<td>Incomplete</td>
<td>.508</td>
<td>.386</td>
<td>.326</td>
<td>.282</td>
<td>.192</td>
<td>.104</td>
<td>.094</td>
<td>.080</td>
<td>.068</td>
<td>.055</td>
</tr>
</tbody>
</table>

Boxplots of $\frac{1}{3} \sum_{1 \leq i \leq 3} D(M(A_i), M(\hat{A}_i))$ (with correct segmentations only)
CCF of $y_t$ (one instance)
CCF of $\hat{x}_t$ (one instance)

Segmentation: $\{1, 4, 6\}$, $\{2, 5\}$, $\{3\}$
Example 7. Let $p = 20$, $q = 5$, \((p_1, \ldots, p_5) = (6, 5, 4, 3, 2)\)

\(x_t\) is defined similarly as in Example 6.

Relative frequencies of correct and incomplete segmentations

<table>
<thead>
<tr>
<th>(n)</th>
<th>400</th>
<th>500</th>
<th>1000</th>
<th>1500</th>
<th>2000</th>
<th>2500</th>
<th>3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>0.058</td>
<td>0.118</td>
<td>0.492</td>
<td>0.726</td>
<td>0.862</td>
<td>0.902</td>
<td>0.940</td>
</tr>
<tr>
<td>Incomplete</td>
<td>0.516</td>
<td>0.672</td>
<td>0.460</td>
<td>0.258</td>
<td>0.130</td>
<td>0.096</td>
<td>0.060</td>
</tr>
</tbody>
</table>

Boxplots of $\frac{1}{5} \sum_{1 \leq i \leq 5} D(M(A_i), \hat{M}(\hat{A}_i))$ (with correct segmentations only)
Segmenting multiple volatility processes

Let \( F_t = \sigma(y_t, y_{t-1}, \cdots) \),

\[
E(y_t|F_{t-1}) = 0, \quad \text{Var}(y_t|F_{t-1}) = \Sigma_y(t).
\]

**Assumption:** \( y_t = Ax_t \),
\[\text{Var}(x_t|F_{t-1}) = \text{diag}(\Sigma_1(t), \cdots, \Sigma_q(t))\].

Let \( \text{Var}(y_t) = \text{Var}(x_t) = I_p \), then \( A \) is orthogonal.

Let \( B_{t-1} \) be a \( \pi \)-class and \( \sigma(B_{t-1}) = F_{t-1} \). put

\[
W_y = \sum_{B \in B_{t-1}} [E\{y_t y'_t I(B)\}]^2, \quad W_x = \sum_{B \in B_{t-1}} [E\{x_t x'_t I(B)\}]^2.
\]

For any \( B \in B_{t-1} \),

\[
E\{x_t x'_t I(B)\} = E[I(B)E\{x_t x'_t|F_{t-1}\}] = E[I(B)\text{diag}(\Sigma_1(t), \cdots, \Sigma_q(t))]
\]

is a block diagonal matrix, so is \( W_x \).
Since 

$$W_y = AW_x A',$$

the proposed method continues to apply.

In practice we estimate $W_y$ by 

$$\hat{W}_y = \sum_{B \in \mathcal{B}} \sum_{k=1}^{k_0} \left( \frac{1}{n-k} \sum_{t=k+1}^{n} y_t y'_t I(y_{t-k} \in B) \right)^2,$$

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

\[ n = 1642, \ p = 6 \]

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

Segmentation for transformed series:

Daily returns of 6 stocks
CCF of squared returns

Y1^2
Y1^2 & Y2^2
Y1^2 & Y3^2
Y1^2 & Y4^2
Y1^2 & Y5^2
Y1^2 & Y6^2
Y2^2
Y2^2 & Y3^2
Y2^2 & Y4^2
Y2^2 & Y5^2
Y2^2 & Y6^2
Y3^2
Y3^2 & Y4^2
Y3^2 & Y5^2
Y3^2 & Y6^2
Y4^2
Y4^2 & Y5^2
Y4^2 & Y6^2
Y5^2
Y5^2 & Y6^2
Y6^2

− p.67
CCF of residuals from fitted GARCH(1,1) for each return series
CCF of squared transformed returns

- $X_1^2$
- $X_1^2$ & $X_2^2$
- $X_1^2$ & $X_3^2$
- $X_1^2$ & $X_4^2$
- $X_1^2$ & $X_5^2$
- $X_1^2$ & $X_6^2$

- $X_2^2$
- $X_2^2$ & $X_1^2$
- $X_2^2$ & $X_3^2$
- $X_2^2$ & $X_4^2$
- $X_2^2$ & $X_5^2$
- $X_2^2$ & $X_6^2$

- $X_3^2$
- $X_3^2$ & $X_1^2$
- $X_3^2$ & $X_2^2$
- $X_3^2$ & $X_4^2$
- $X_3^2$ & $X_5^2$
- $X_3^2$ & $X_6^2$

- $X_4^2$
- $X_4^2$ & $X_1^2$
- $X_4^2$ & $X_2^2$
- $X_4^2$ & $X_3^2$
- $X_4^2$ & $X_5^2$
- $X_4^2$ & $X_6^2$

- $X_5^2$
- $X_5^2$ & $X_1^2$
- $X_5^2$ & $X_2^2$
- $X_5^2$ & $X_3^2$
- $X_5^2$ & $X_4^2$
- $X_5^2$ & $X_6^2$

- $X_6^2$
- $X_6^2$ & $X_1^2$
- $X_6^2$ & $X_2^2$
- $X_6^2$ & $X_3^2$
- $X_6^2$ & $X_4^2$
- $X_6^2$ & $X_5^2$

- $X_7^2$

- $X_8^2$

- $X_9^2$

- $X_10^2$

- $X_11^2$

- $X_12^2$
CCF of residuals from transformed return series

X1

X1 & X2

X1 & X3

X1 & X4

X1 & X5

X1 & X6

X2

X2 & X1

X2 & X2

X2 & X3

X2 & X4

X2 & X5

X2 & X6

X3

X3 & X1

X3 & X2

X3 & X3

X3 & X4

X3 & X5

X3 & X6

X4

X4 & X1

X4 & X2

X4 & X3

X4 & X4

X4 & X5

X4 & X6

X5

X5 & X1

X5 & X2

X5 & X3

X5 & X4

X5 & X5

X5 & X6

X6

X6 & X1

X6 & X2

X6 & X3

X6 & X4

X6 & X5

X6 & X6

- p.70
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? \[ W_y = \sum_k \Sigma_y(k) \Sigma_y(k)' = AW_x A', \]

\[
\text{tr}(W_y) = \sum_k \sum_{i,j=1}^p \rho_{ij}^y(k)^2 = \text{tr}(W_x) = \sum_k \sum_{i,j=1}^p \rho_{ij}^x(k)^2 = \sum_k \sum_i \rho_{ii}^x(k)^2
\]

provided all components of \( x_t \) are uncorrelated across all time lags

Components of \( x_t \) are more predictable, due to stronger ACF!
• When latent segmentation does not exist, use $z_t = \Gamma'_y y_t$,

$$W_y = \sum_k \Sigma_y(k) \Sigma_y(k)' = \Gamma_y D \Gamma'_y.$$  

Hence $\Sigma_z(k) = \Gamma'_y \Sigma_y(k) \Gamma_y$, and therefore

$$W_z = \sum_k \Sigma_z(k) \Sigma_z(k)' = \Gamma'_y W_y \Gamma_y = D,$$

i.e. $W_z$ is a diagonal matrix.

**Note.** $\Sigma_z(k)$ 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.

• Improve the test for vector white noise: Chang, Yao and Zhou (2017).

• R package **PCA4TS** available from **CRAN**