

Factor Models for Multiple Time Series

Qiwei Yao

Department of Statistics, London School of Economics

q.yao@lse.ac.uk

Joint work with

Neil Bathia, University of Melbourne

Clifford Lam, London School of Mathematics

Jiazhu Pan, University of Strathclyde

Flavio Ziegelmann, Federal University of Rio Grande do Sul

- Econometric factor models: a brief survey
- Statistical factor models: identification
- Estimation
 - expanding white noise space: non-stationary factors
 - eigenanalysis: stationary cases
- Asymptotic properties (stationary cases only in this talk)
 - fixed p : fast convergence rate for zero-eigenvalues
 - $p \rightarrow \infty$: convergence rates independent of p
- Illustration with real data sets
 - temperature data
 - implied volatility surfaces
 - densities of intraday returns

Econometric modelling: represent a $p \times 1$ time series y_t as

$$y_t = \mathbf{f}_t + \boldsymbol{\xi}_t,$$

both \mathbf{f}_t and $\boldsymbol{\xi}_t$ are unobservable, and

- \mathbf{f}_t : driven by r **common factors**, and $r \ll p$
- $\boldsymbol{\xi}_t$: **idiosyncratic components**

Basic idea. The dynamical structure of each component of y_t is driven by the r common factors plus one or a few idiosyncratic components.

Practical motivation: asset pricing models, yield curves, portfolio risk management, derivative pricing, macroeconomic behaviour and forecasting, consumer theory etc.

Sargent & Sims (1977) and Geweke (1977): dynamic-factor models

Chamberlain & Rothschild (1983): *approximate* and *static* factor models

Forni, Hallin, Lippi & Reichlin (2002 –): generalized dynamic factor models – combining the above two together

$$y_{it} = b_{i1}(L)u_{1t} + \cdots + b_{ir}(L)u_{rt} + \xi_{it}, \quad i = 1, 2, \cdots, t = 0, \pm 1, \cdots,$$

- $u_{kt} \sim \text{WN}(0, 1)$, $k = 1, \cdots, r$, are **common (dynamic) factors**, and are uncorrelated with each other,
- ξ_{it} are stationary in t , are **idiosyncratic noise**, and $\{u_{kt}\}$ and $\{\xi_{it}\}$ are uncorrelated.

Only y_{it} are observable.

Let $\xi_{pt} = (\xi_{1t}, \dots, \xi_{pt})^\top$ and $y_{pt} = (y_{1t}, \dots, y_{pt})^\top$.

Assumption: As $p \rightarrow \infty$, it holds almost surely on $[-\pi, \pi]$ that all the eigenvalues of spectral density matrices of ξ_{pt} are uniformly bounded, and only the r largest eigenvalues of $(y_{pt} - \xi_{pt})$ converge to ∞ .

Intuition: The r common factors affect the dynamics of most component series, while each idiosyncratic noise only affects the dynamics of a few component series.

Characteristics result: As $p \rightarrow \infty$, it holds almost surely on $[-\pi, \pi]$ that all the r largest eigenvalues of spectral density matrices of y_{pt} converge to ∞ , and the $(r + 1)$ -th largest eigenvalue is uniformly bounded.

The model is asymptotically identifiable, when the number of time series (i.e. the number of cross-sectional .057172.14.68(637

- Estimation for GDFM when r is given — **Dynamic principle component analysis** (Brillinger 1981):

- i. Obtain an estimator $\hat{\Sigma}(\theta)$ for spectral density matrix of y_t , $\theta \in [-\pi, \pi]$
- ii. Find eigenvalues and eigenvectors of $\hat{\Sigma}(\theta)$
- iii. Project y_t onto the space spanned by the r eigenvectors corresponding to the r largest eigenvalues:

the projection is defined as the mean square limit of a Fourier sequence, and

each component of the projection is a sum of r uncorrelated MA processes.

- Determine r : **only identifiable when $p \rightarrow \infty$!**

'There is no way a slowly diverging sequence can be told from an eventually bounded sequence' (Forni et al. 2000).

Let $\{\mathbf{y}_t\}$ be a $p \times 1$ time series defined by

$$\mathbf{y}_t = \mathbf{A}\mathbf{x}_t + \boldsymbol{\varepsilon}_t,$$

Let $\{y_t\}$ be a $p \times 1$ time series defined by

$$y_t = \mathbf{A}x_t + \varepsilon_t,$$

x_t : $r \times 1$ unobservable **factors**, $r (< p)$ unknown

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

$\{\varepsilon_t\}$: vector $WN(\mu_\varepsilon, \Sigma_\varepsilon)$

no linear combinations of x_t are WN.

Let $\{y_t\}$ be a $p \times 1$ time series defined by

$$y_t = \mathbf{A}x_t + \varepsilon_t,$$

x_t : $r \times 1$ unobservable **factors**, $r (< p)$ unknown

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

$\{\varepsilon_t\}$: vector $WN(\mu_\varepsilon, \Sigma_\varepsilon)$

no linear combinations of x_t are WN.

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

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

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

Let $\{y_t\}$ be a $p \times 1$ time series defined by

$$y_t = \mathbf{A}x_t + \varepsilon_t,$$

x_t : $r \times 1$ unobservable **factors**, $r (< p)$ unknown

\mathbf{A} : $p \times r$ unknown constant

What is new?

- No distributional assumption on ε_t . More significantly, allow correlation between ε_t and \mathbf{x}_{t+k} .

What is new?

- No distributional assumption on ε_t . More significantly, allow correlation between ε_t and \mathbf{x}_{t+k} .
- Factor \mathbf{x}_t , and therefore also \mathbf{y}_t , may be nonstationary, not necessarily driven by unit roots.

What is new?

- No distributional assumption on ε_t . More significantly, allow correlation between ε_t and \mathbf{x}_{t+k} .
- Factor \mathbf{x}_t , and therefore also \mathbf{y}_t , may be nonstationary, not necessarily driven by unit roots.
- A new estimation method for stationary cases: an eigenanalysis, applicable when $p \gg n$

What is new?

- No distributional assumption on ε_t . More significantly, allow correlation between ε_t and \mathbf{x}_{t+k} .
- Factor \mathbf{x}_t , and therefore also \mathbf{y}_t , may be nonstationary, not necessarily driven by unit roots.
- A new estimation method for stationary cases: an eigenanalysis, applicable when $p \gg n$

Key: estimate \mathbf{A} , or more precisely, $\mathcal{M}(\mathbf{A})$.

With available an estimator $\hat{\mathbf{A}}$, a natural estimator for factor and the associated residuals are

$$\hat{\mathbf{x}}_t = \hat{\mathbf{A}}^\tau \mathbf{y}_t, \quad \hat{\varepsilon}_t = (\mathbf{I}_p - \hat{\mathbf{A}}\hat{\mathbf{A}}^\tau)\mathbf{y}_t.$$

By modelling the lower-dimensional $\hat{\mathbf{x}}_t$, we obtain the dynamical model for \mathbf{y}_t :

$$\hat{\mathbf{y}}_t = \hat{\mathbf{A}}\hat{\mathbf{x}}_t.$$

Reconciling to econometric models

‘Common factors’ & ‘idiosyncratic noise’: conceptually appealing,
only identifiable when $p \rightarrow \infty$.

Goal: identify those components of \mathbf{x}_t , each of them affects most (or a few) components of \mathbf{y}_t .

Put $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_r)$ and $\mathbf{x}_t = (x_{t1}, \dots, x_{tr})'$. Then

$$\mathbf{y}_t = \mathbf{a}_1 x_{t1} + \dots + \mathbf{a}_r x_{tr} + \boldsymbol{\varepsilon}_t.$$

Hence the number of non-zero coefficients of \mathbf{a}_j is the number of components of \mathbf{y}_t which are affected by the factor x_{tj} .

To avoid the correlation among the components of \mathbf{x}_t , apply PCA to \mathbf{x}_t , i.e. replace $(\mathbf{A}, \mathbf{x}_t)$ by $(\mathbf{A}\boldsymbol{\Gamma}, \boldsymbol{\Gamma}'\mathbf{x}_t)$, where $\boldsymbol{\Gamma}$ is an $r \times r$ orthogonal matrix defined in $\text{Var}(\mathbf{x}_t) = \boldsymbol{\Gamma}\mathbf{D}\boldsymbol{\Gamma}'$.

Eigenvalues of $\text{Var}(\mathbf{x}_t)$ are different, this representation is unique.

Lemma 1. Let $\mathbf{A}_1 \mathbf{z}_1 = \mathbf{A}_2 \mathbf{z}_2$, where, for $i = 1, 2$, \mathbf{A}_i is $p \times r$ matrix, $\mathbf{A}_i' \mathbf{A}_i = \mathbf{I}_r$, and $\mathbf{z}_i = (z_{i1}, \dots, z_{ir})'$ is $r \times 1$ random vector with uncorrelated components, and $\text{Var}(z_{i1}) > \dots > \text{Var}(z_{ir})$. Furthermore $\mathcal{M}(\mathbf{A}_1) = \mathcal{M}(\mathbf{A}_2)$. Then $z_{1j} = \pm z_{2j}$ for $1 \leq j \leq r$.

In practice, we use the PCA-ed factor $\hat{\mathbf{x}}_t$.

The number of non-zero elements of the j -th column of $\hat{\mathbf{A}}$ is the number of the components of \mathbf{y}_t whose dynamics depends on the j -th factor \hat{x}_{tj} .

Nonstationary factors

C1. $\varepsilon_t \sim \text{WN}(\mu_\varepsilon, \Sigma_\varepsilon)$, $\mathbf{c}'\mathbf{x}_t$ is not white noise for any constant $\mathbf{c} \in \mathcal{R}^p$. Furthermore $\mathbf{A}'\mathbf{A} = \mathbf{I}_r$.

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

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

$$\mathbf{B}^\top \mathbf{A} = \mathbf{0}, \quad \mathbf{B}^\top \mathbf{B} = \mathbf{I}_{p-r}.$$

Since $\mathbf{y}_t = \mathbf{A}\mathbf{x}_t + \varepsilon_t$,

$$\mathbf{B}^\top \mathbf{y}_t = \mathbf{B}^\top \varepsilon_t$$

i.e. $\{\mathbf{B}^\top \mathbf{y}_t, t = 0, \pm 1, \dots\}$ is WN.

Therefore

$$\text{Corr}(\mathbf{b}_i^\top \mathbf{y}_t, \mathbf{b}_j^\top \mathbf{y}_{t-k}) = 0 \quad \forall 1 \leq i, j \leq p-r \text{ and } k \geq 1.$$

Search for mutually orthogonal directions $\mathbf{b}_1, \mathbf{b}_2, \dots$ one by one such that the projection of \mathbf{y}_t on each of those directions is a white noise.

Stop the search when such a direction is no longer available, and take $p - k$ as the estimated value of r , where k is the number of directions obtained in the search.

Search for mutually orthogonal directions b_1, b_2, \dots one by one such that the projection of y_t on each of those directions is a white noise.

Stop the search when such a direction is no longer available, and take $p - k$ as the estimated value of r , where k is the number of directions obtained in the search.

See Pan and Yao (2008) for further details, and also some (preliminary) asymptotic results.

Stationary models

C2. \mathbf{x}_t is stationary, and $\text{Cov}(\mathbf{x}_t, \boldsymbol{\varepsilon}_{t+k}) = 0$ for any $k \geq 0$.

Put $\boldsymbol{\Sigma}_y(k) = \text{Cov}(\mathbf{y}_{t+k}, \mathbf{y}_t)$, $\boldsymbol{\Sigma}_x(k) = \text{Cov}(\mathbf{x}_{t+k}, \mathbf{x}_t)$,
 $\boldsymbol{\Sigma}_{x\varepsilon}(k) = \text{Cov}(\mathbf{x}_{t+k}, \boldsymbol{\varepsilon}_t)$. By $\mathbf{y}_t = \mathbf{A}\mathbf{x}_t + \boldsymbol{\varepsilon}_t$,

$$\boldsymbol{\Sigma}_y(k) = \mathbf{A}\boldsymbol{\Sigma}_x(k)\mathbf{A}' + \mathbf{A}\boldsymbol{\Sigma}_{x\varepsilon}(k), \quad k \geq 1.$$

For a prescribed integer $k_0 \geq 1$, define

$$\mathbf{M} = \sum_{k=1}^{k_0} \boldsymbol{\Sigma}_y(k)\boldsymbol{\Sigma}_y(k)'$$

Then $\mathbf{M}\mathbf{B} = 0$, i.e. the columns of \mathbf{B} are the eigenvectors of \mathbf{M} corresponding to zero-eigenvalues.

Hence under C1 and C2,

Hence under C1 and C2,

The factor loading space $\mathcal{M}(\mathbf{A})$ are spanned by the eigenvectors of \mathbf{M} corresponding to its non-zero eigenvalues, and the number of the non-zero eigenvalues is r .

Hence under C1 and C2,

The factor loading space $\mathcal{M}(\mathbf{A})$ are spanned by the eigenvectors of \mathbf{M} corresponding to its non-zero eigenvalues, and the number of the non-zero eigenvalues is r .

Let $\widehat{\mathbf{M}} = \sum_{k=1}^{k_0} \widehat{\Sigma}_y(k) \widehat{\Sigma}_y(k)'$, where $\widehat{\Sigma}_y(k)$ denotes the sample covariance matrix of y_t at lag k .

Hence under C1 and C2,

The factor loading space $\mathcal{M}(\mathbf{A})$ are spanned by the eigenvectors of \mathbf{M} corresponding to its non-zero eigenvalues, and the number of the non-zero eigenvalues is r .

Let $\widehat{\mathbf{M}} = \sum_{k=1}^{k_0} \widehat{\Sigma}_y(k) \widehat{\Sigma}_y(k)'$, where $\widehat{\Sigma}_y(k)$ denotes the sample covariance matrix of y_t at lag k .

\widehat{r} : No. of non-zero eigenvalues of $\widehat{\mathbf{M}}$,

$\widehat{\mathbf{A}}$: its columns are the \widehat{r} orthonormal eigenvectors of $\widehat{\mathbf{M}}$ corresponding to its \widehat{r} largest eigenvalues.

Bootstrap test for r

Note that $r = r_0$ iff the $(r_0 + 1)$ -th largest eigenvalue of M is 0 and the r_0 -th largest eigenvalue is nonzero.

Consider the testing for $H_0 : \lambda_{r_0+1} = 0$,

We reject H_0 if $\hat{\lambda}_{r_0+1} > l_\alpha$.

Bootstrap to determine l_α :

1. Compute \hat{y}_t with $\hat{r} = r_0$. Let $\hat{\varepsilon}_t = y_t - \hat{y}_t$.
2. Let $y_t^* = \hat{y}_t + \varepsilon_t^*$, where ε_t^* are drawn independently (with replacement) from $\{$

Asymptotics I: $n \rightarrow \infty$ and p fixed

- (i) \mathbf{y}_t is strictly stationary, $E\|\mathbf{y}_t\|^{4+\delta} < \infty$ for some $\delta > 0$.
- (ii) \mathbf{y}_t is α -mixing satisfying $\sum_j \alpha(j)^{\frac{\delta}{2+\delta}} < \infty$.
- (iii) \mathbf{M} has r non-zero eigenvalues $\lambda_1 > \dots > \lambda_r > 0$.

Then under condition C1 and C2, the following assertions hold.

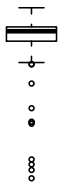
- (i) $\hat{\lambda}_j - \lambda_j = O_P(n^{-1/2})$ for $1 \leq j \leq r$,
- (ii) $\hat{\lambda}_{r+k} = O_P(n^{-1})$ for $1 \leq k \leq p - r$,
- (iii) $D\{\mathcal{M}(\hat{\mathbf{A}}), \mathcal{M}(\mathbf{A})\} = O_P(n^{-1/2})$ provided $\hat{r} = r$ a.s.,

where

$$D\{\mathcal{M}(\hat{\mathbf{A}}), \mathcal{M}(\mathbf{A})\} = 1 - \frac{1}{r} \text{tr}(\mathbf{A}\mathbf{A}^\tau \hat{\mathbf{A}}\hat{\mathbf{A}}^\tau).$$

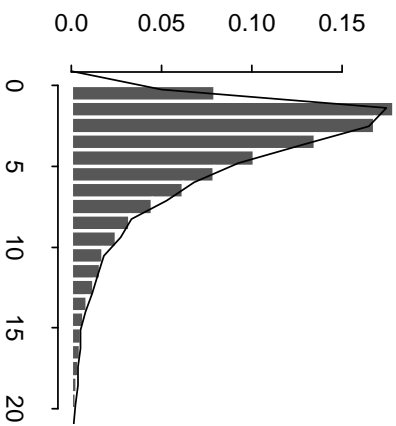
Numerical illustration: $\lambda_1 = 1.884$, $\lambda_2 = \lambda_3 = \lambda_4 = 0$ ($p = 4, r = 1$)

(Simulation replications: 10,000 times)

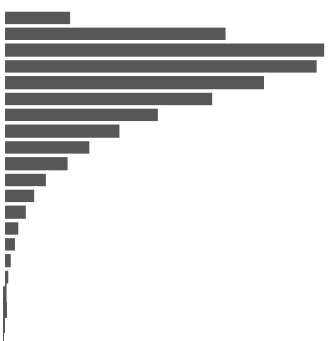
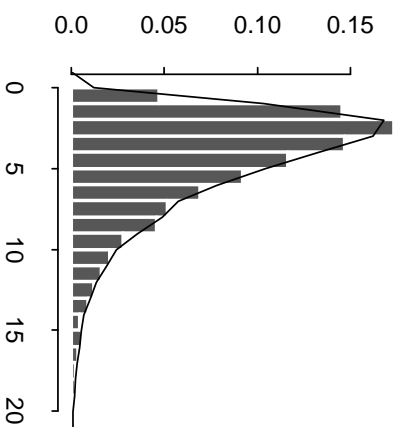


Histogram of $n\hat{\lambda}_2$

n=10



n=20



Asymptotics II: $n \rightarrow \infty, p \rightarrow \infty$ and r fixed

Recall model: $\mathbf{y}_t = \mathbf{A}\mathbf{x}_t + \boldsymbol{\varepsilon}_t$, and \mathbf{A} is $p \times r$

1. Assumptions on **Strength of factors**:

(i) $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_r)$, $\|\mathbf{a}_i\|^2 \asymp p^{1-\delta}$, $i = 1, \dots, r$, $0 \leq \delta \leq 1$.

(ii) For $k = 0, 1, \dots, k_0$, $\boldsymbol{\Sigma}_x(k) \equiv \text{Cov}(\mathbf{x}_{t+k}, \mathbf{x}_t)$ is full-ranked, and $\boldsymbol{\Sigma}_{x,\varepsilon}(k) \equiv \text{Cov}(\mathbf{x}_{t+k}, \boldsymbol{\varepsilon}_t) = O(1)$ elementwisely.

We call

- factors are strong if $\delta = 0$,
- factors are weak if $\delta > 0$.

Standardization ' $\mathbf{A}^\tau \mathbf{A} = \mathbf{I}_r$ ' + (i, ii) imply:

$$\|\boldsymbol{\Sigma}_x(k)\| \asymp p^{1-\delta} \asymp \|\boldsymbol{\Sigma}_x(k)\|_{\min}, \quad \|\boldsymbol{\Sigma}_{x,\varepsilon}(k)\| = O(p^{1-\delta/2}),$$

where $a \asymp b$ represents $a = O(b)$ & $b = O(a)$, $\|\mathbf{A}\|^2 = \lambda_{\max}(\mathbf{A}\mathbf{A}^\tau)$

and $\|\mathbf{A}\|_{\min}^2 = \min\{\lambda(\mathbf{A}\mathbf{A}^\tau) : \lambda(\mathbf{A}\mathbf{A}^\tau) > 0\}$.

2. For $k = 0, 1, \dots, k_0$, $\|\Sigma_{x,\epsilon}(k)\| = o(p^{1-\delta})$, and it holds elementwisely that

$$\tilde{\Sigma}_x(k) - \Sigma_x(k) = O_P(n^{-l_x}), \quad \tilde{\Sigma}_\epsilon(k) - \Sigma_\epsilon(k) = O_P(n^{-l_\epsilon}),$$

$$\tilde{\Sigma}_{x,\epsilon}(k) - \Sigma_{x,\epsilon}(k) = O_P(n^{-l_{x\epsilon}}) = \tilde{\Sigma}_{\epsilon,x}(k)$$

for some constants $0 < l_x, l_{x\epsilon}, l_\epsilon \leq 1/2$, and $\tilde{\Sigma}$ denotes the sample version of Σ .

3. \mathbf{M} has r different non-zero eigenvalues.

Then under condition C1 and C2,

$$\|\hat{\mathbf{A}} - \mathbf{A}\| = O_P(h_n) = O_P(n^{-l_x} + p^{\delta/2}n^{-l_{x\epsilon}} + p^\delta n^{-l_\epsilon}),$$

provided $h_n = o(1)$.

Remark. When all factors are strong (i.e. $\delta = 0$), the convergence rate h_n is independent of the dimension p .

Our asymptotic theory also shows:

1. Factor model-based estimator for Σ_y :

$$\hat{\Sigma}_y = \hat{\mathbf{A}}\hat{\Sigma}_x\hat{\mathbf{A}}^\tau + \hat{\Sigma}_\epsilon, \quad \text{where} \quad \hat{\Sigma}_x = \hat{\mathbf{A}}^\tau(\tilde{\Sigma}_y - \hat{\Sigma}_\epsilon)\hat{\mathbf{A}},$$

cannot improve over the sample covariance estimator $\tilde{\Sigma}_y$.

But the convergence rate for $\|\hat{\Sigma}_y^{-1} - \Sigma_y^{-1}\|$ is independent of p when all the factors are strong.

Simulation with $r = 1$ and $\delta = 0$ (only one strong factor):

$$x_t = 0.9x_{t-1} + N(0, 4),$$

$\varepsilon_{tj} \sim_{iid} N(0, 4)$, and the i -th element of \mathbf{A} is $2 \cos(2\pi i/p)$.

$n = 200$	$\ \hat{\mathbf{A}} - \mathbf{A}\ $	$\ \tilde{\Sigma}_y^{-1} - \Sigma_y^{-1}\ $	$\ \hat{\Sigma}_y^{-1} - \Sigma_y^{-1}\ $
$p = 20$.022(.005)	.24(.03)	.009(.002)
$p = 180$.023(.004)	79.8(29.8)	.007(.001)
$p = 400$.022(.004)	-	.007(.001)
$p = 1000$.023(.004)	-	.007(.001)

$$n = \mathbf{\Sigma} \mathbf{0} \mathbf{\Sigma}$$

Illustration With Real Data

Example 1. The monthly temperature data from 7 cities in Eastern China in January 1954 — December 1986

$$n = 396, \quad p = 7, \quad \hat{r} = 4$$

Example 2. Daily implied volatility surfaces for IBM, Microsoft and Dell call options in 2006

$$n = 100, \quad p = 130, \quad \hat{r} = 1$$

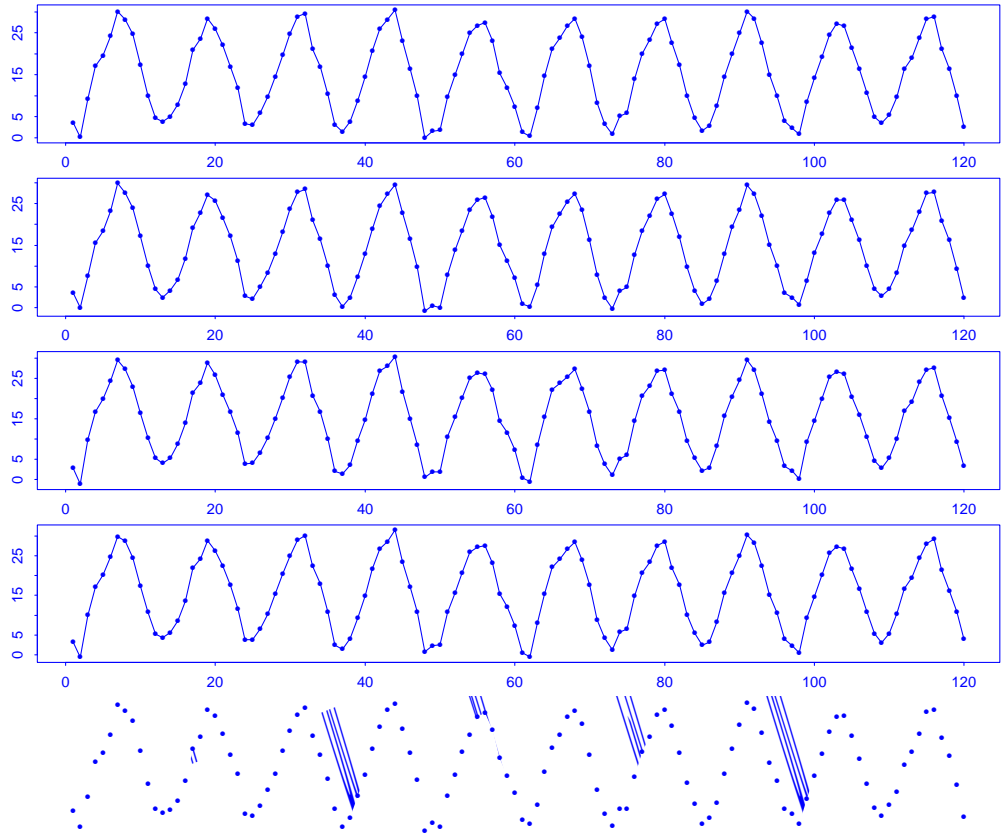
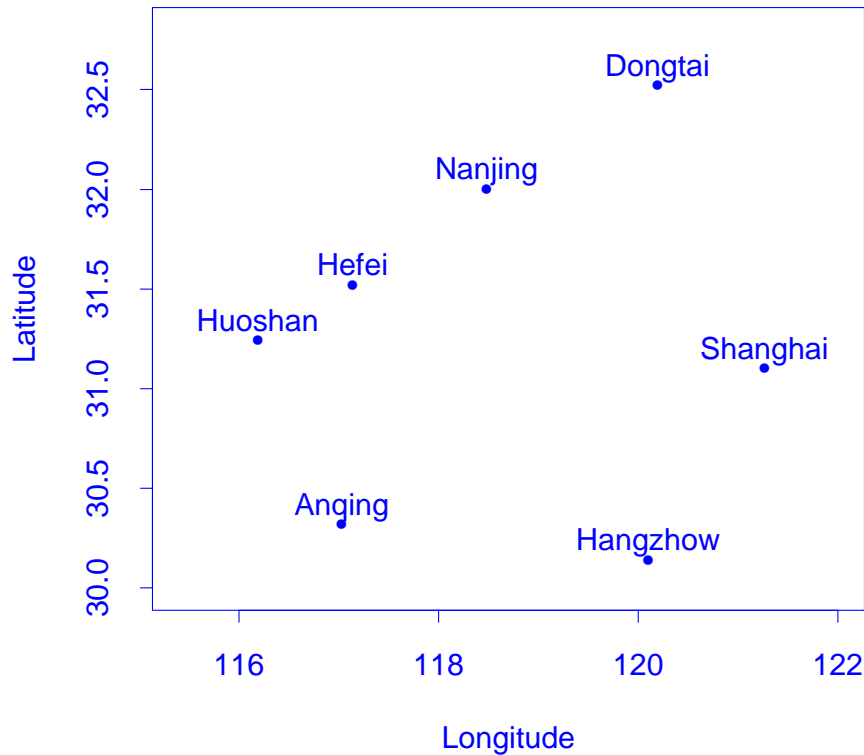


Example 3. Daily densities of one-minute returns of IBM stock price in 2006

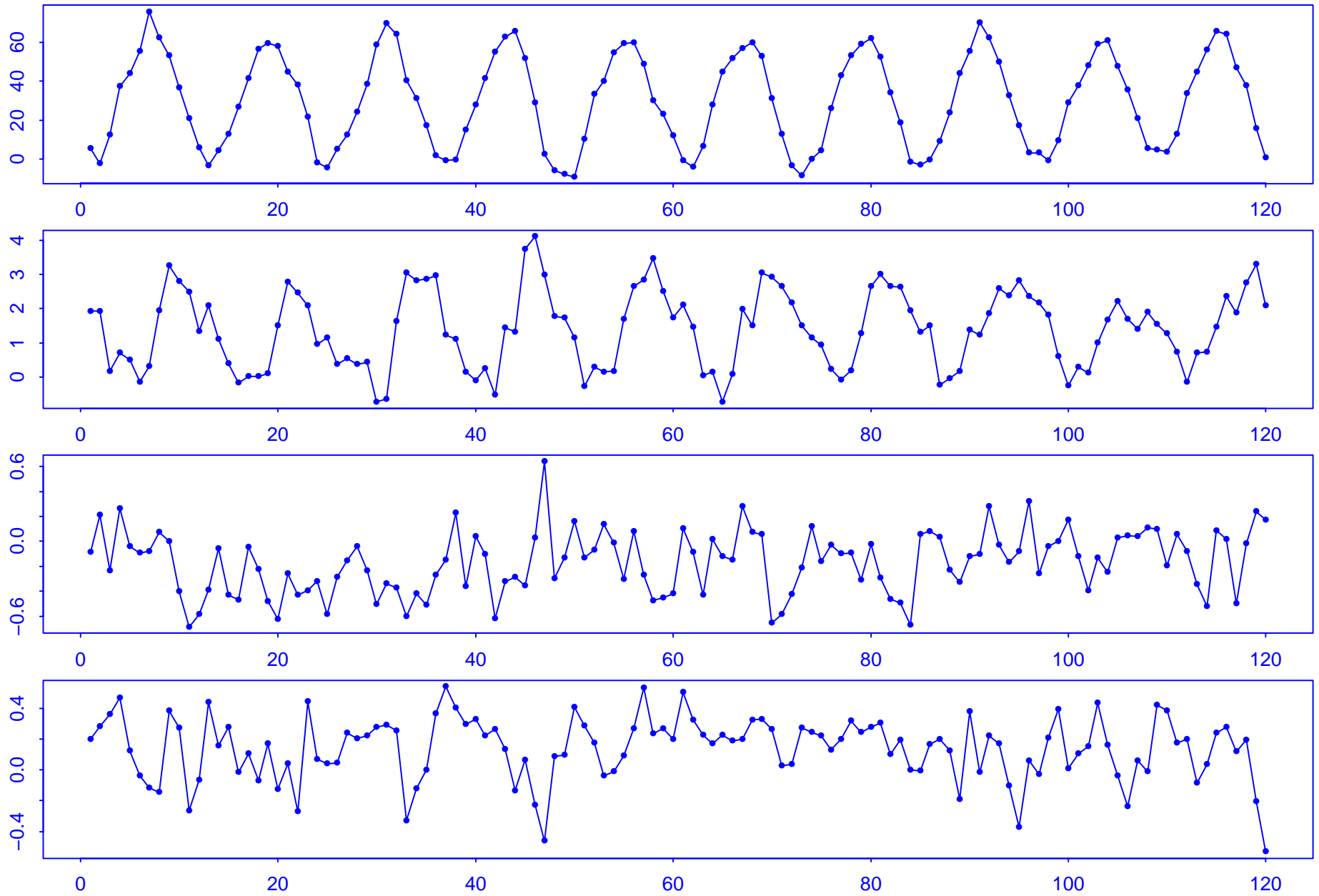
$$n = 251, \quad p = \infty, \quad \hat{r} = 2$$



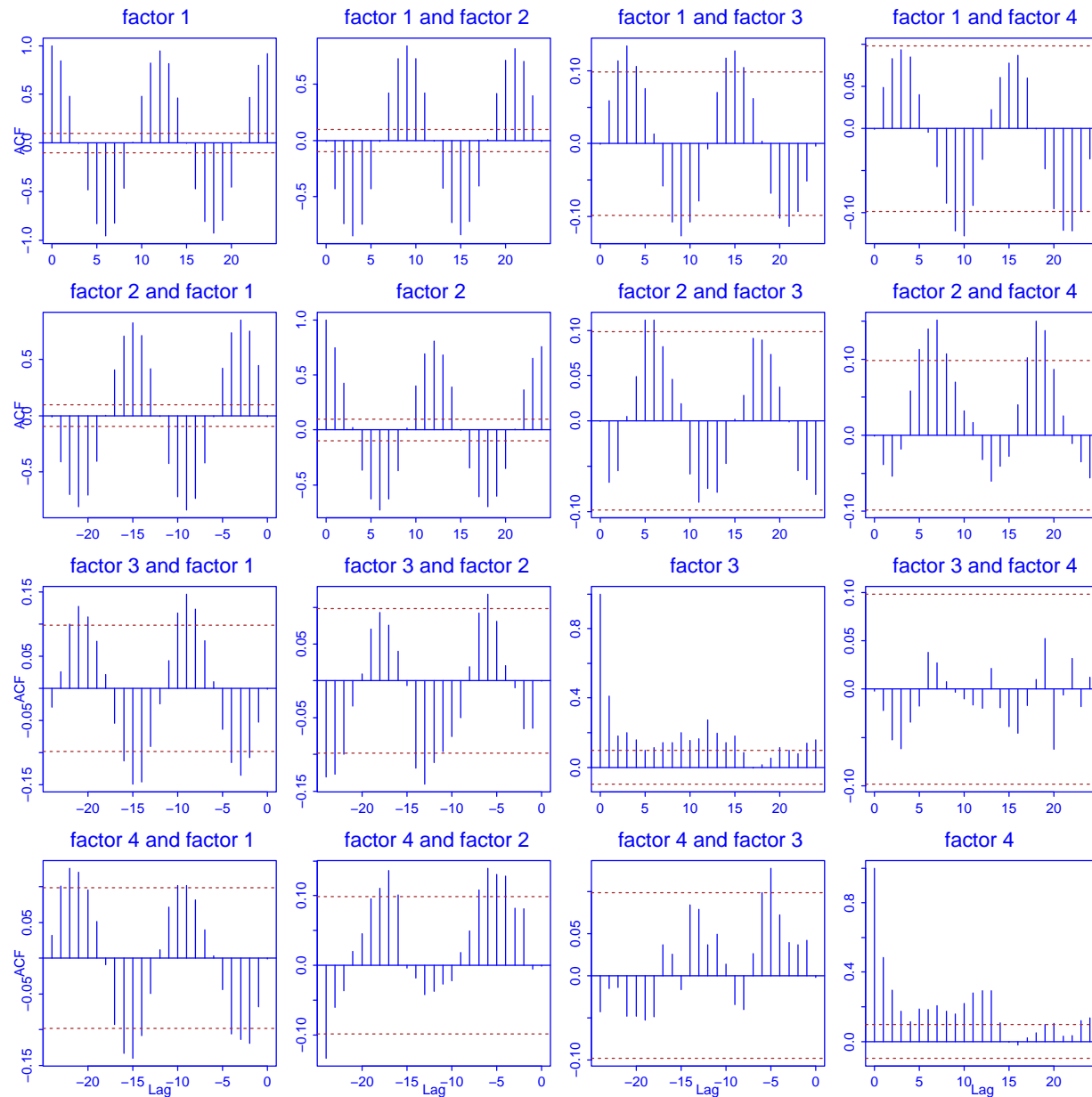
Time plots of the monthly temperature in 1959-1968 of Nanjing, Dongtai, Huoshan, Hefei, Shanghai, Anqing and Hangzhou.



Time plots of the 4 estimated factors VAR(1)



Sample cross-correlation of the 4 estimated factors

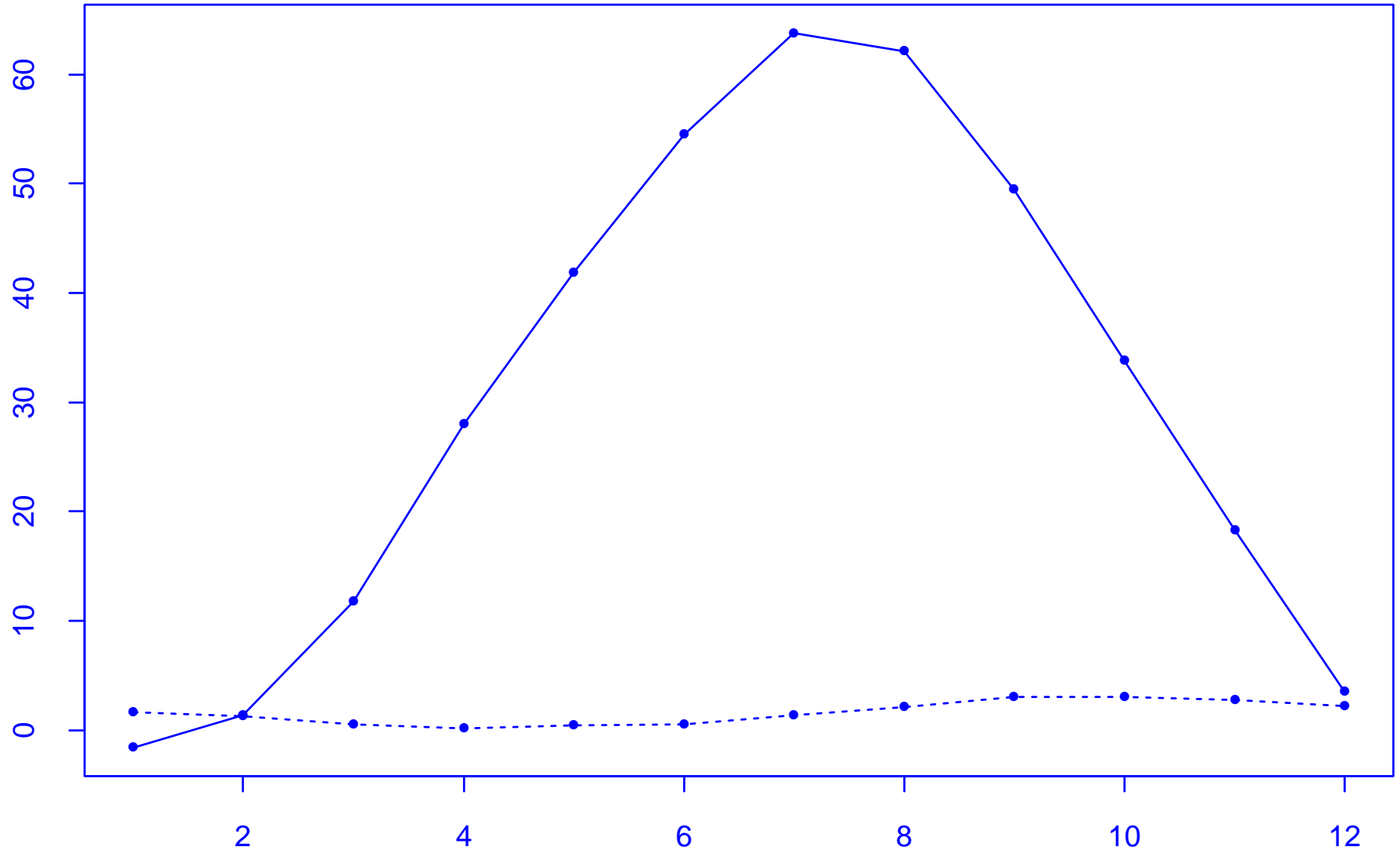


Sample cross-correlation of the 3 residuals (i.e. $\hat{B}^\tau y_t$)



1

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



In the fitted factor model $y_t = \hat{\mathbf{A}}\mathbf{x}_t + e_t$, the AICC selected **VAR(1)** for the

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



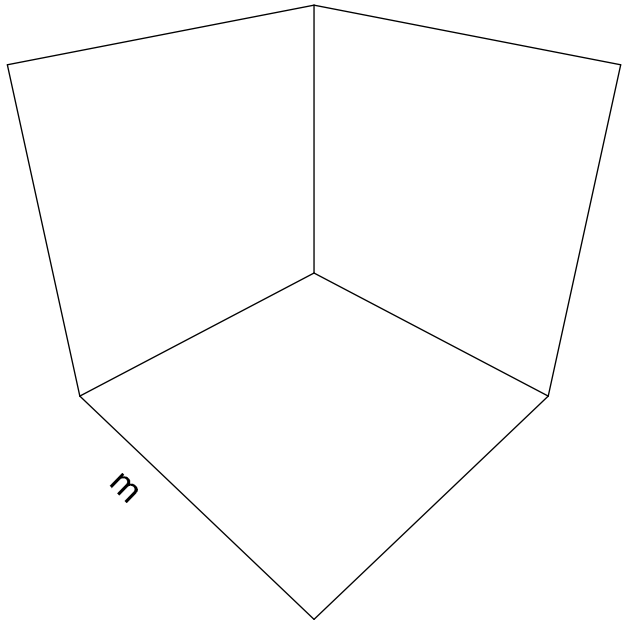
Example 2. Implied volatility surfaces of IBM, Microsoft and Dell stocks in 2006 (i.e. 251 trading days).

Source of Data: OptionMetrics at WRDS

Observations: for $t = 1, \dots, 251$, implied volatility $w_t(u_i, v_j)$ computed from call options at

- time to maturity at 30, 60, 91, 122, 152, 182, 273, 365, 547 & 730 calendar days, denoted by u_1, \dots, u_{10} , and
- delta at 0.2, 0.25, 0.3, 0.35, 0.4, \dots , 0.8, denoted by v_1, \dots, v_{13} .

Total: $p = 10 \times 13 = 130$ time series with the len



Fitting a factor model on each of the rolling windows of length 100 days:

$$\mathbf{y}_i, \mathbf{y}_{i+1}, \dots, \mathbf{y}_{i+99}, \quad i = 1, \dots, 150.$$

The estimated number of factors for all 3 stocks across different windows is always $\hat{r} = 1$.

Based on a fitted AR model to the estimated factor process, we predict the next value x_{i+100} , denoted by \check{x}_{i+100} . It leads to the one-step ahead prediction for \mathbf{y}_{i+100} :

$$\check{\mathbf{y}}_{i+100} = \hat{\mathbf{A}}\check{x}_{i+100}.$$

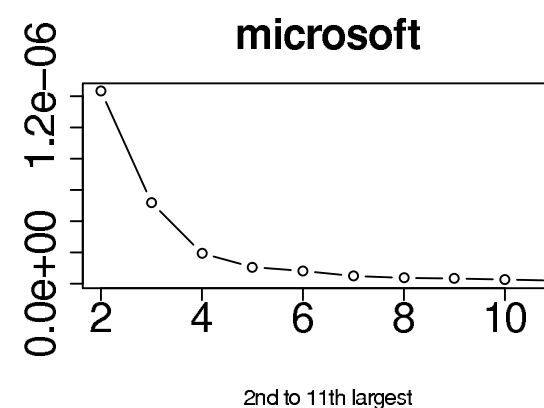
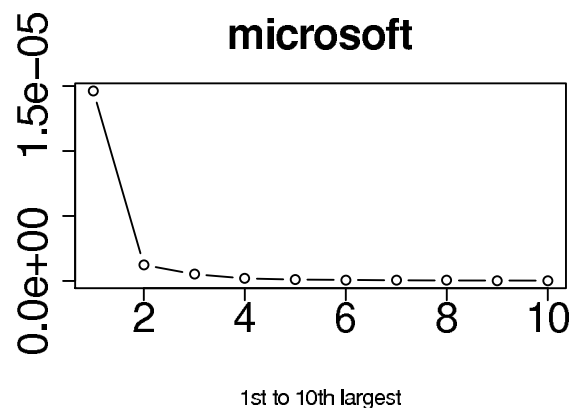
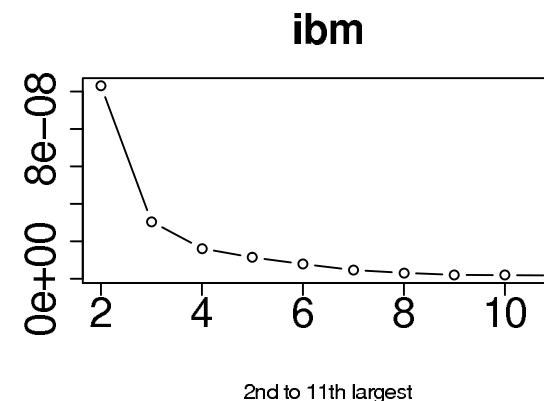
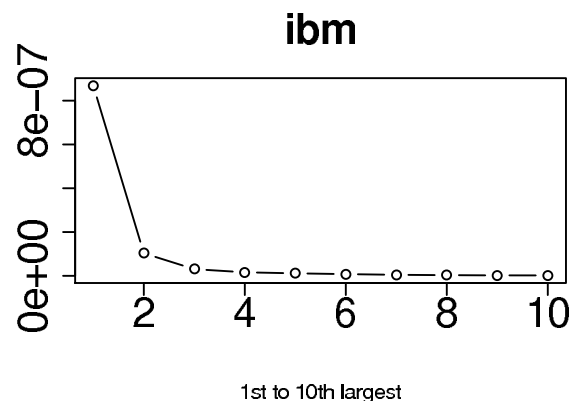
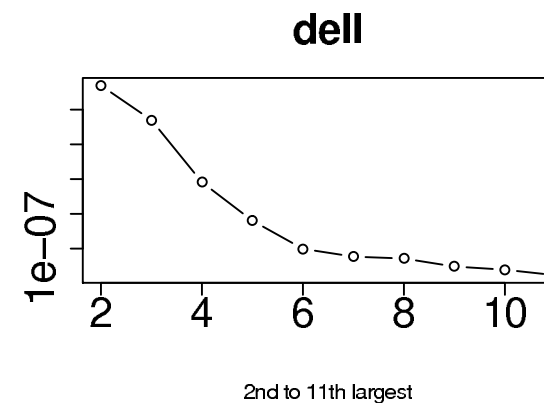
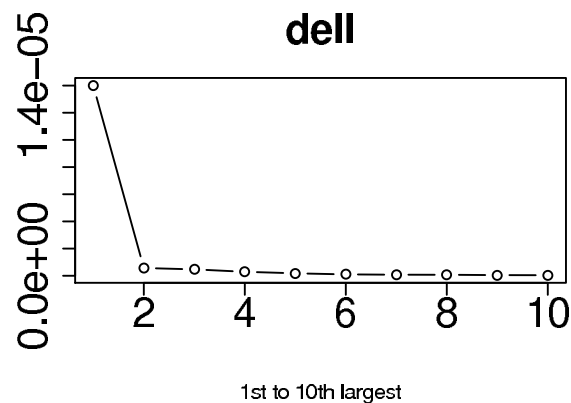
Put

$$\text{RMSE}_i = \frac{1}{\sqrt{p}} \|\check{\mathbf{y}}_{i+100} - \mathbf{y}_{i+100}\|, \quad i = 1, \dots, 150.$$

Average of the ordered eigenvalues of \widehat{M} over the 150 rolling windows.

3 panels on the left: 10 largest eigenvalues

3 panels on the right: 2nd–11th largest eigenvalues.

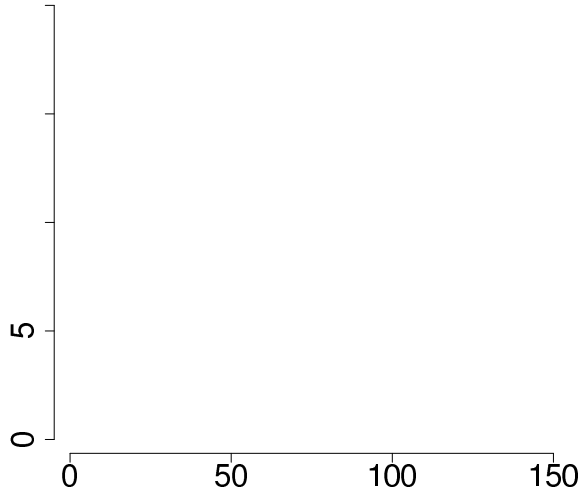


Benchmark prediction for y_{i+100} : the previous value y_{i+99}

Prediction based on Bai & Ng (2002) — factor-modelling based on the LSE: $(\hat{\mathbf{A}}, \hat{\mathbf{x}}_t)$ is the solution of

$$\min_{\mathbf{A}, \mathbf{x}_t} \sum_{t=1}^n \|\mathbf{y}_t - \mathbf{A}\mathbf{x}_t\|^2, \quad \text{subject to } \mathbf{A}^\tau \mathbf{A} / p = \mathbf{I}_r \text{ and } \mathbf{X}^\tau \mathbf{X} / n = \mathbf{I}_r,$$

where $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$.



Example 3. IBM stock intra-day prices in 2006

251 trading days, tick by tick prices collected in 9:30 — 16:00

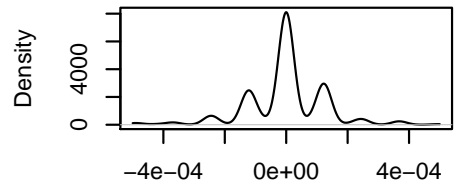
In total 2,786,649 observations (74MB)

For each of 251 trading days, construct the pdf curve of one-minute log-return using the log-returns in 390 one-minute intervals: kernel density estimation with $h = 0.000025$

Treating the 251 pdfs as a high-dimensional time series, apply the proposed procedure.

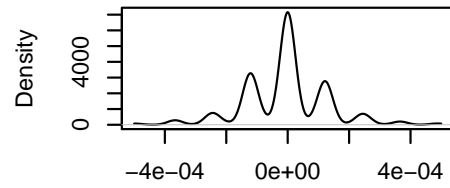
The white-noise test rejects $H_0 : r = 1$, but cannot reject $H_0 : r = 2$.

day 1



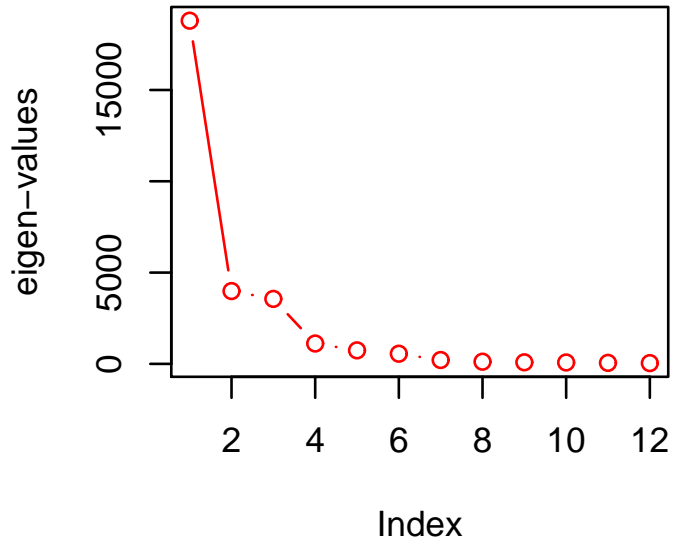
bandwidth = $2.5e-05$

day 2

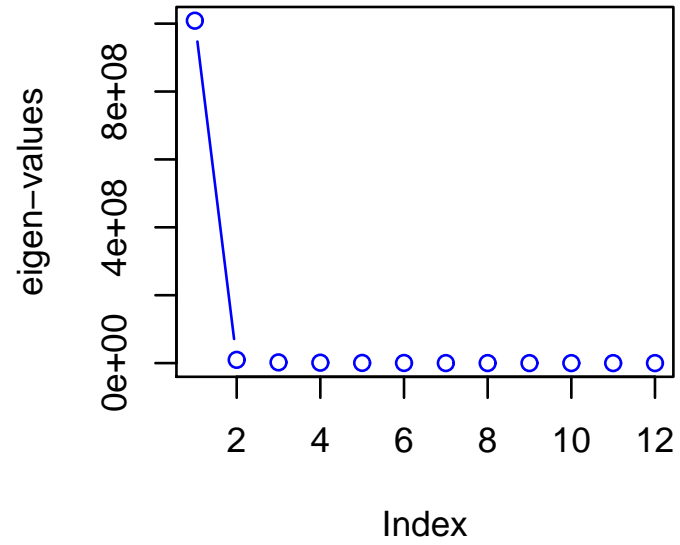


bandwidth = $2.5e-05$

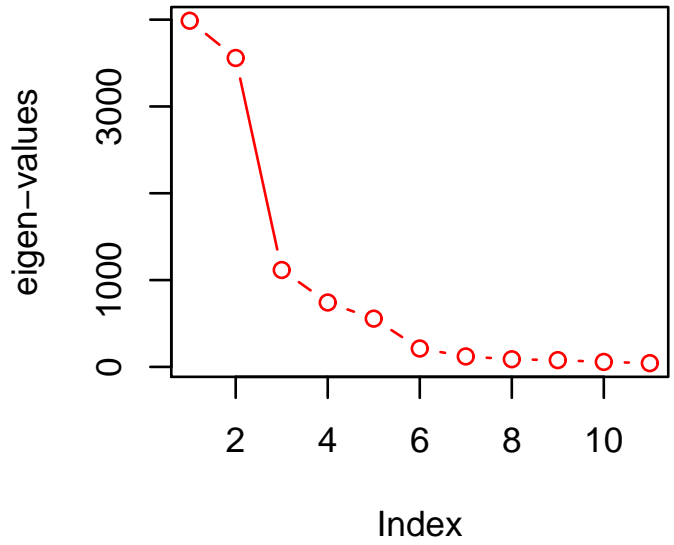
BHK: 1 to 12



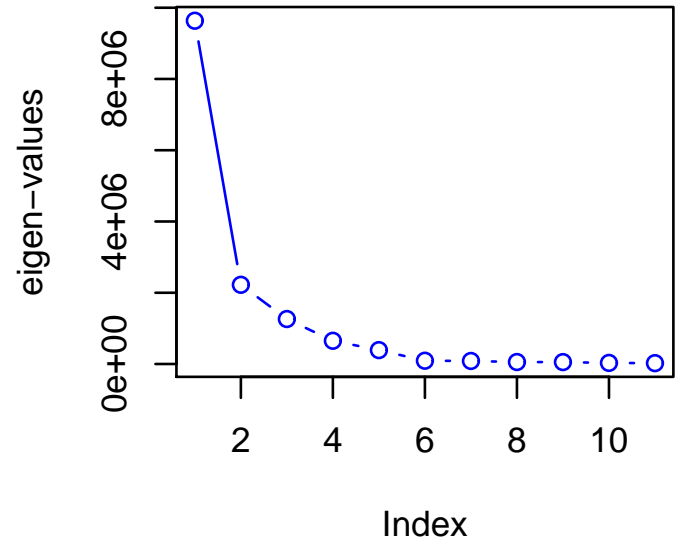
New: 1 to 12



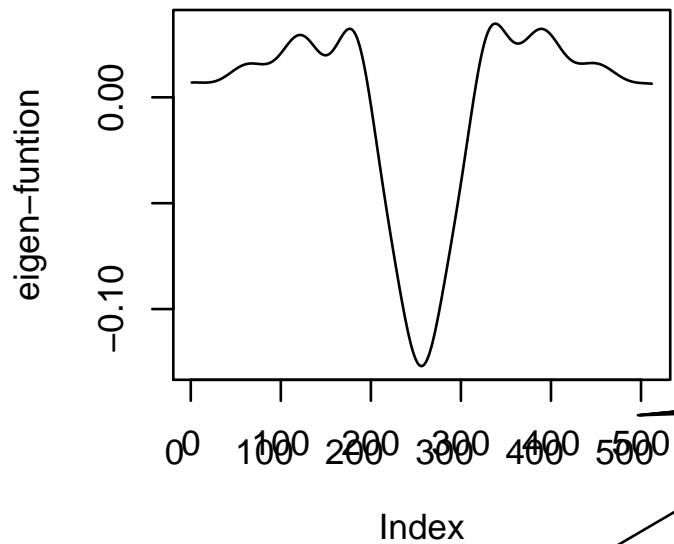
BHK: 2 to 12



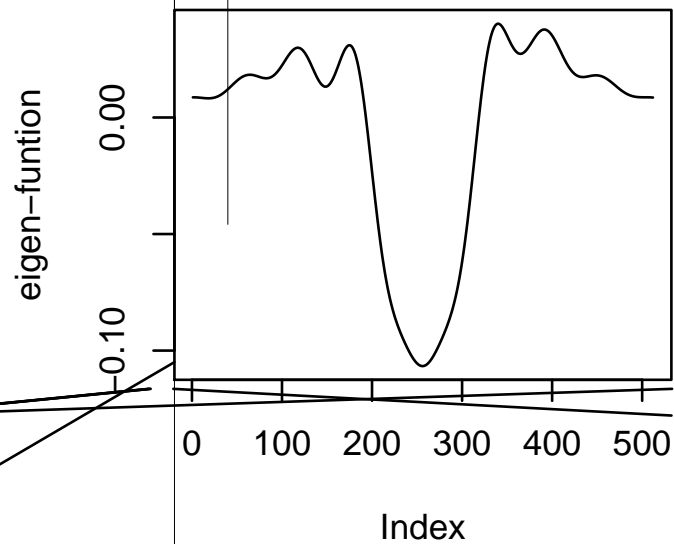
New: 2 to 12



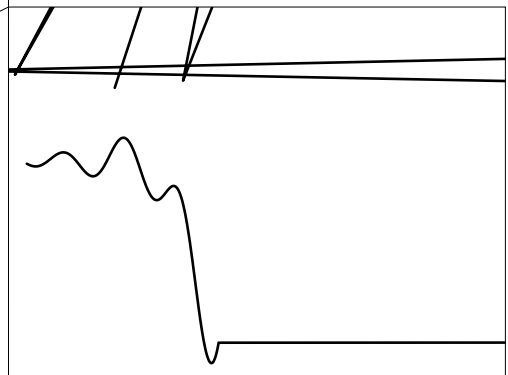
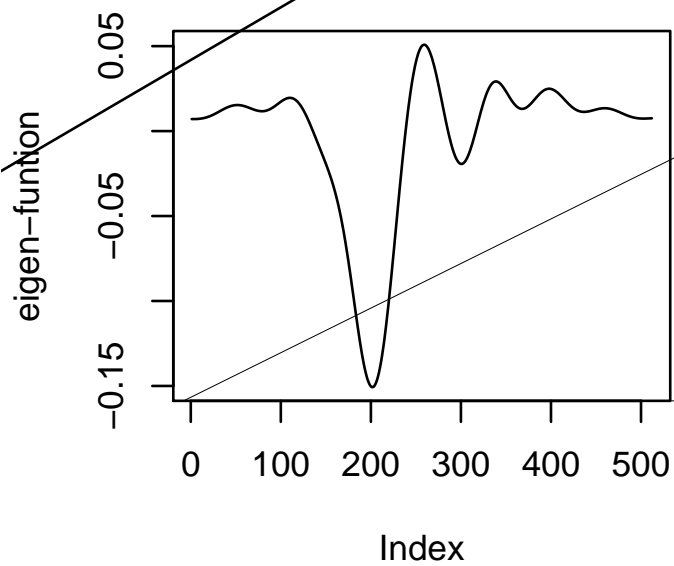
1st Eigen-Function (BHK)



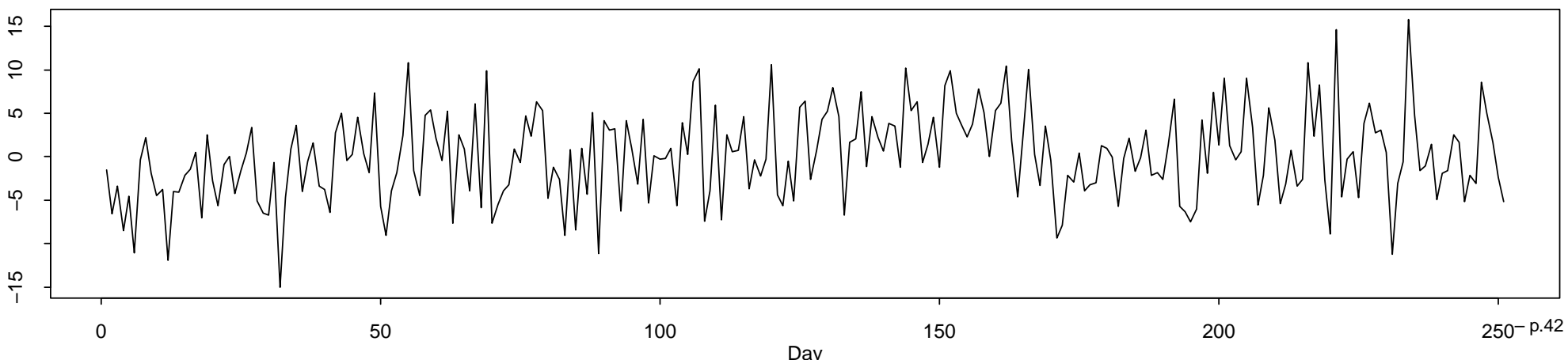
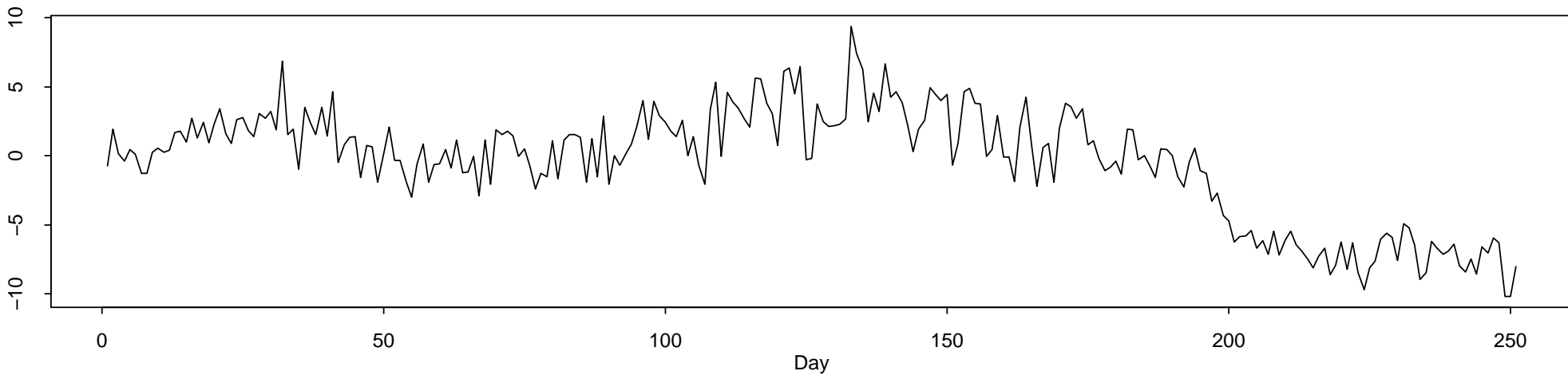
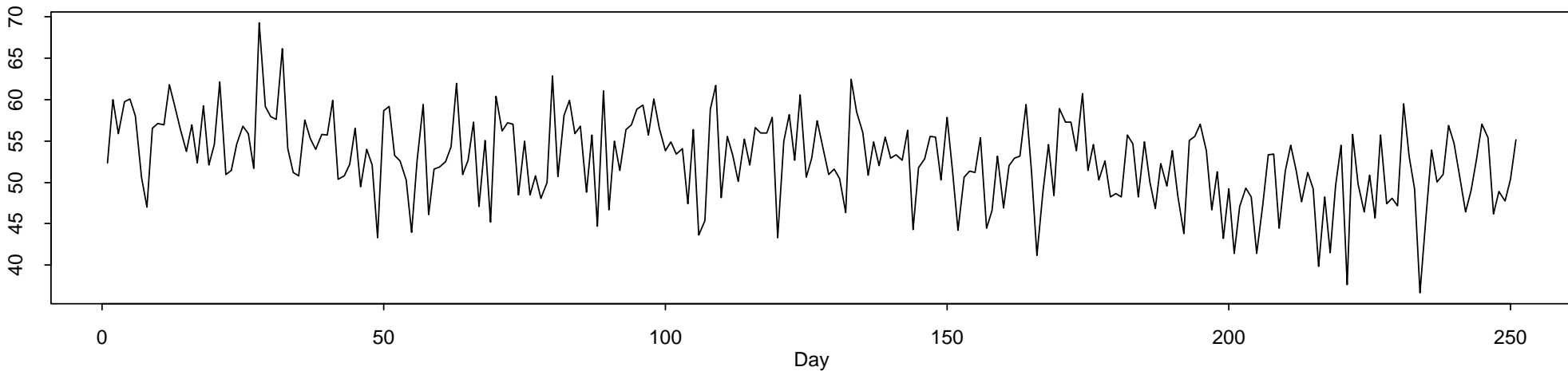
1st Eigen-Function (New)



2nd Eigen-Function (BHK)

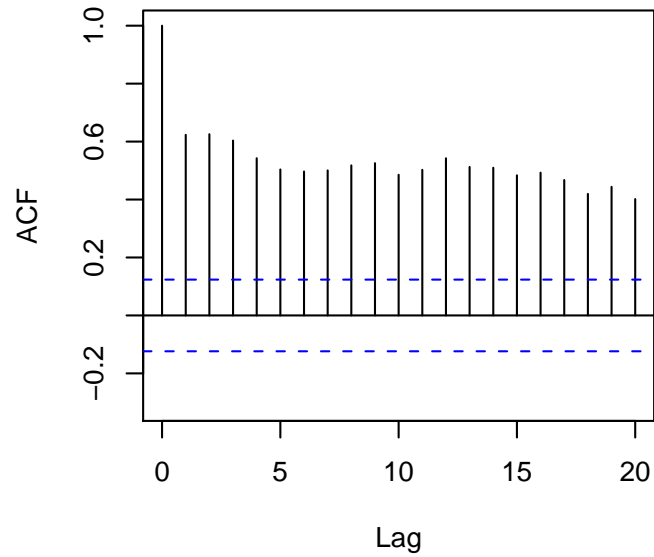


Time series plots of x_{t1} and x_{t2}

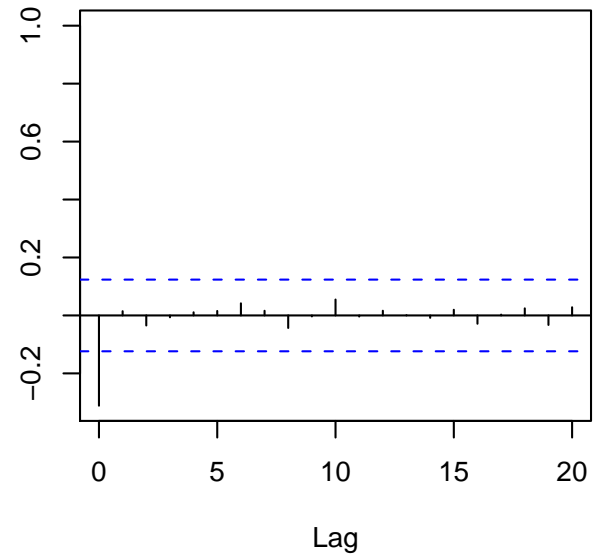


ACF of (x_{t1}, x_{t2})

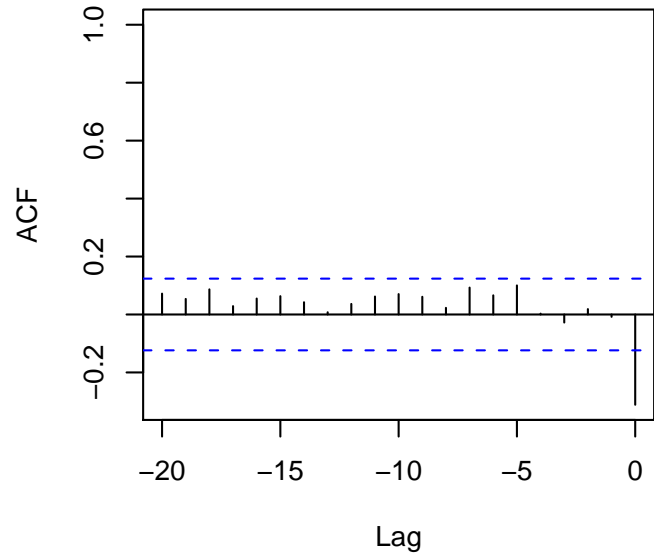
Series 1



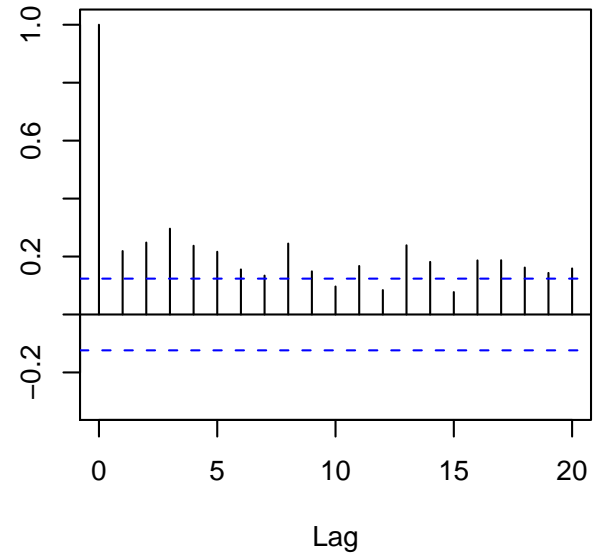
Series 1 & Series 2



Series 2 & Series 1

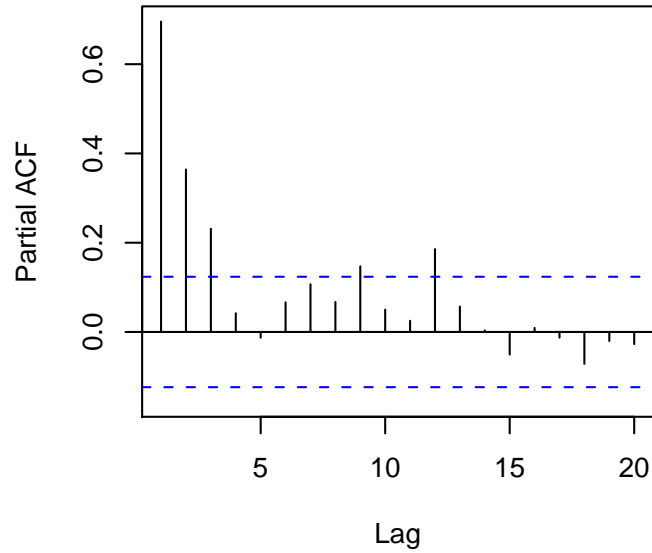


Series 2

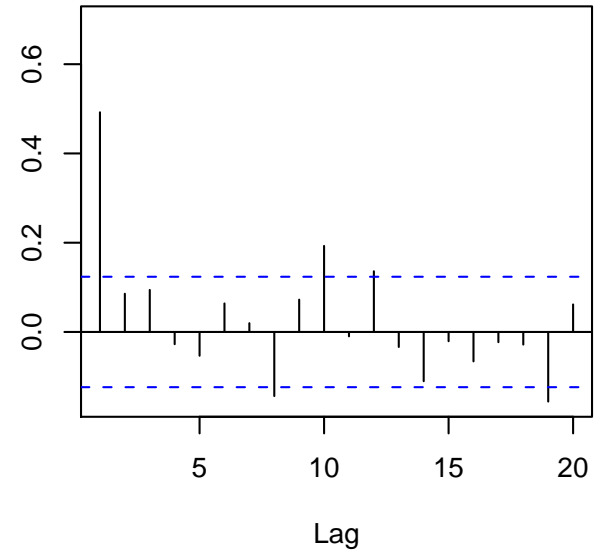


PACF of (x_{t1}, x_{t2})

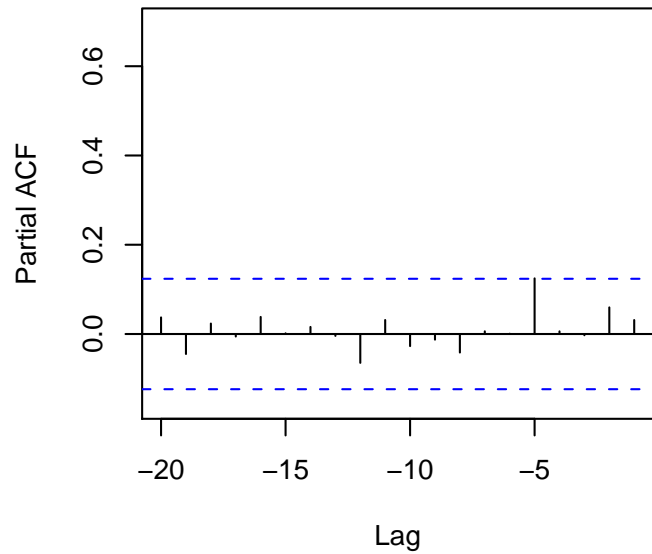
Series 1



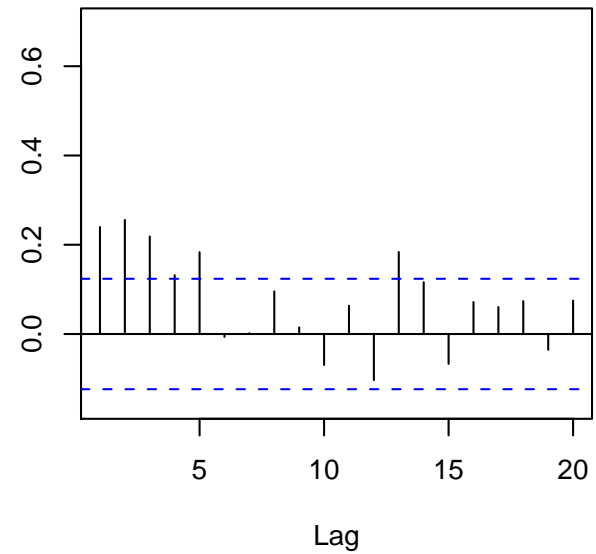
Series 1 & Series 2



Series 2 & Series 1



Series 2



Fitting time series $\mathbf{x}_t = (x_{t1}, x_{t2})'$

Since there is little cross correlation between the two component series, we fit them separately.

For $\{x_{t1}\}$, AIC selected ARMA(1,1) with AIC=4556.76:

$$x_{t+1,1} = 0.985x_{t1} + \varepsilon_{t+1,1} - 0.787\varepsilon_{t,1}.$$

For $\{x_{t2}\}$, AIC selected ARMA(1,1) with AIC=4323.1:

$$x_{t+1,2} = 0.982x_{t2} + \varepsilon_{t+1,2} - 0.885\varepsilon_{t,2}.$$

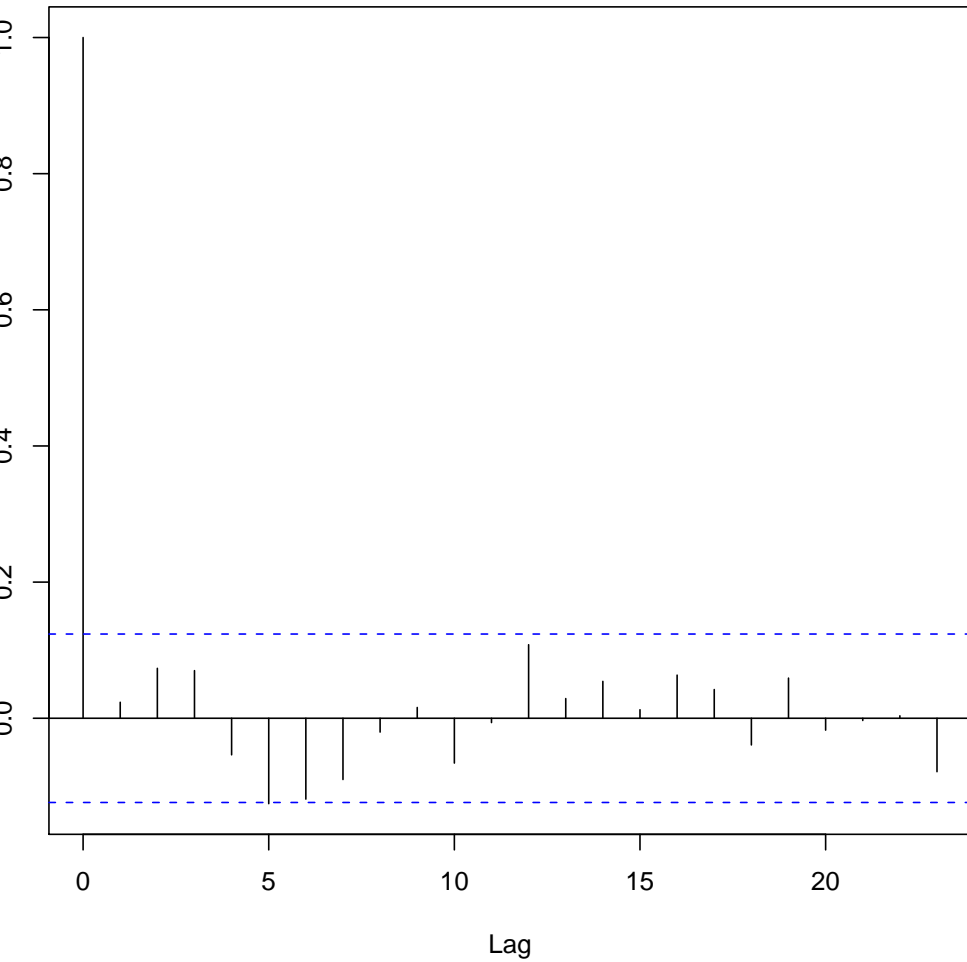
Allowing nonstationarity — ARIMA(1,1,1):

$$x_{t+1,1} - x_{t1} = 0.062(x_{t1} - x_{t-1,1}) + \varepsilon_{t+1,1} - 0.847\varepsilon_{t,1}, \quad (\text{AIC} = 4537.13)$$

$$x_{t+1,2} - x_{t2} = 0.046(x_{t2} - x_{t-1,2}) + \varepsilon_{t+1,2} - 0.889\varepsilon_{t,2}, \quad (\text{AIC} = 4306.08)$$

ACF of the residuals from the fitted ARMA(1,1) models

Series residuals(arima(xi[, 1], order = c(1, 0, 1)))



Series residuals(arima(xi[, 2], order = c(1, 0, 1)))

