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
  - $p \to \infty$: 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 $\mathbf{y}_t$ as

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

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

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

Basic idea. The dynamical structure of each component of $\mathbf{y}_t$ 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

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

- $u_{kt} \sim \mathrm{WN}(0, 1)$, $k = 1, \cdots, r$, 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 $\boldsymbol{\xi}_{pt} = (\xi_{1t}, \cdots, \xi_{pt})^\tau$ and $\mathbf{y}_{pt} = (y_{1t}, \cdots, y_{pt})^\tau$.

Assumption: As $p \to \infty$, it holds almost surely on $[-\pi, \ \pi]$ that all the eigenvalues of spectral density matrices of $\boldsymbol{\xi}_{pt}$ are uniformly bounded, and only the r largest eigenvalues of $(\mathbf{y}_{pt} - \boldsymbol{\xi}_{pt})$ 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 $p \to \infty$, it holds almost surely on $[-\pi, \ \pi]$ that all the r largest eigenvalues of spectral density matrices of $\mathbf{y}_{pt}$ 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 variables) $p \to \infty$!

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

  i. Obtain an estimator $\widehat{\mathbf{\Sigma}}(\theta)$ for spectral density matrix of $\mathbf{y}_t$, $\theta \in [-\pi, \pi]$

  ii. Find eigenvalues and eigenvectors of $\widehat{\mathbf{\Sigma}}(\theta)$

  iii. Project $\mathbf{y}_t$ 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 $p \to \infty$!

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

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

$\mathbf{x}_t$: $r \times 1$ unobservable factors, $r\ (< p)$ unknown

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

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

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

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

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

$\mathbf{x}_t$: $r \times 1$ unobservable factors, $r\ (< p)$ unknown

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

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

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

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

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

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

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

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

$\mathbf{x}_t$: $r \times 1$ unobservable factors, $r \ (< p)$ unknown

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

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

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

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

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

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

The model is not new: Peña & Box (1987).

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 $\boldsymbol{\varepsilon}_t$. More significantly, allow correlation between $\boldsymbol{\varepsilon}_t$ 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 $\boldsymbol{\varepsilon}_t$. More significantly, allow correlation between $\boldsymbol{\varepsilon}_t$ 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 $p >> n$

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 $p >> n$

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

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

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

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

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

Reconciling to econometric models

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

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

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

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

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 $(\mathbf{A\Gamma}, \mathbf{\Gamma}'\mathbf{x}_t)$, where $\mathbf{\Gamma}$ is an $r \times r$ orthogonal matrix defined in Var$(\mathbf{x}_t) = \mathbf{\Gamma D \Gamma}'$.

Eigenvalues of 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 $\mathbf{z}_i = (z_{i1}, \cdots, z_{ir})'$ is $r \times 1$ random vector with uncorrelated components, and Var$(z_{i1}) > \cdots >$ Var(Z_{ir}). Furthermore $\mathcal{M}(\mathbf{A}_1) = \mathcal{M}(\mathbf{A}_2)$. Then $z_{1j} = \pm z_{2j}$ for $1 \le j \le r$.

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

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

Nonstationary factors

C1. $\boldsymbol{\varepsilon}_t \sim \mathrm{WN}(\boldsymbol{\mu}_{\boldsymbol{\varepsilon}}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})$, $\mathbf{c}'\mathbf{x}_t$ is not white noise for any constant $\mathbf{c} \in \mathcal{R}^p$. Furthermore $\mathbf{A}'\mathbf{A} = \mathbf{I}_r$.

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

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

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

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

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

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

Therefore

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

Search for mutually orthogonal directions $\mathbf{b}_1, \mathbf{b}_2, \cdots$ 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 $\mathbf{b}_1, \mathbf{b}_2, \cdots$ 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.

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

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

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 $\mathrm{MB} = 0$, i.e. the columns of B are the eigenvectors of 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 $\widehat{\mathbf{M}} = \sum_{k=1}^{k_0} \widehat{\mathbf{\Sigma}}_y(k)\widehat{\mathbf{\Sigma}}_y(k)'$, where $\widehat{\mathbf{\Sigma}}_y(k)$ denotes the sample covariance matrix of $\mathbf{y}_t$ 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 $\widehat{\mathbf{M}} = \sum_{k=1}^{k_0} \widehat{\mathbf{\Sigma}}_y(k)\widehat{\mathbf{\Sigma}}_y(k)'$, where $\widehat{\mathbf{\Sigma}}_y(k)$ denotes the sample covariance matrix of $\mathbf{y}_t$ 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 $\mathbf{M}$ 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 $\widehat{\lambda}_{r_0+1} > l_\alpha$.

Bootstrap to determine l_α:

1. Compute $\widehat{\mathbf{y}}_t$ with $\widehat{r} = r_0$. Let $\widehat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \widehat{\mathbf{y}}_t$.
2. Let $\mathbf{y}_t^* = \widehat{\mathbf{y}}_t + \boldsymbol{\varepsilon}_t^*$, where $\boldsymbol{\varepsilon}_t^*$ are drawn independently (with replacement) from $\{\widehat{\boldsymbol{\varepsilon}}_1, \cdots, \widehat{\boldsymbol{\varepsilon}}_n\}$.
3. Form the operator $\mathbf{M}^*$ in the same manner as $\widehat{\mathbf{M}}$ with $\{\mathbf{y}_t\}$ replaced by $\{\mathbf{y}_t^*\}$, compute the $(r_0 + 1)$-th largest eigenvalue $\lambda_{r_0+1}^*$ of $\mathbf{M}^*$.

$\mathcal{L}(\lambda_{r_0+1}^* | \{\mathbf{y}_t\})$ is taken as the distribution of $\widehat{\lambda}_{r_0+1}$ under H_0.

Asymptotics I: $n \to \infty$ 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 $\lambda_1 > \cdots > \lambda_r > 0$.

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

(i) $\widehat{\lambda}_j - \lambda_j = O_P(n^{-1/2})$ for $1 \le j \le r$,

(ii) $\widehat{\lambda}_{r+k} = O_P(n^{-1})$ for $1 \le k \le p - r$,

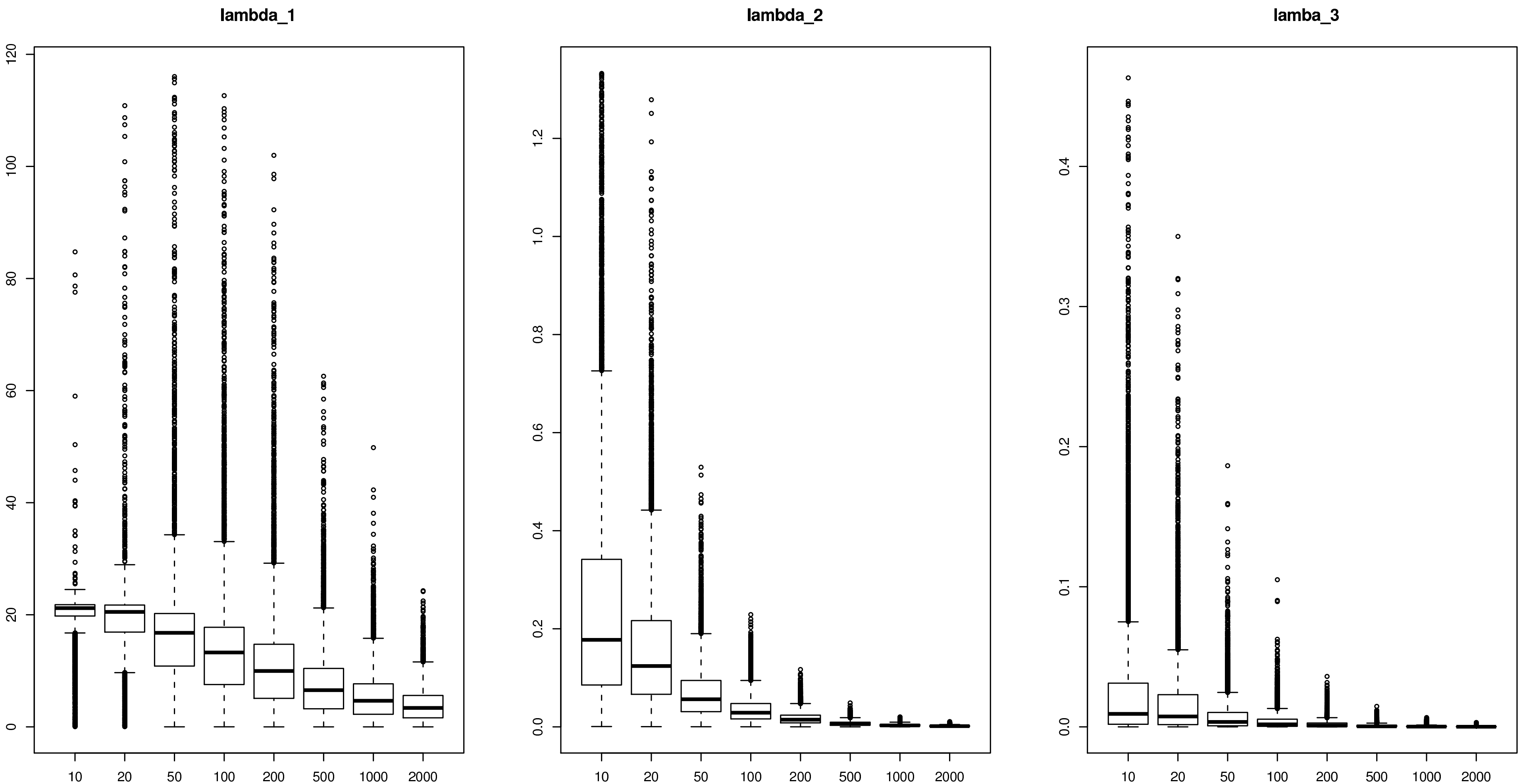
(iii) $D\{\mathcal{M}(\widehat{\mathbf{A}}),\ \mathcal{M}(\mathbf{A})\} = O_P(n^{-1/2})$ provided $\widehat{r} = r$ a.s.,

where

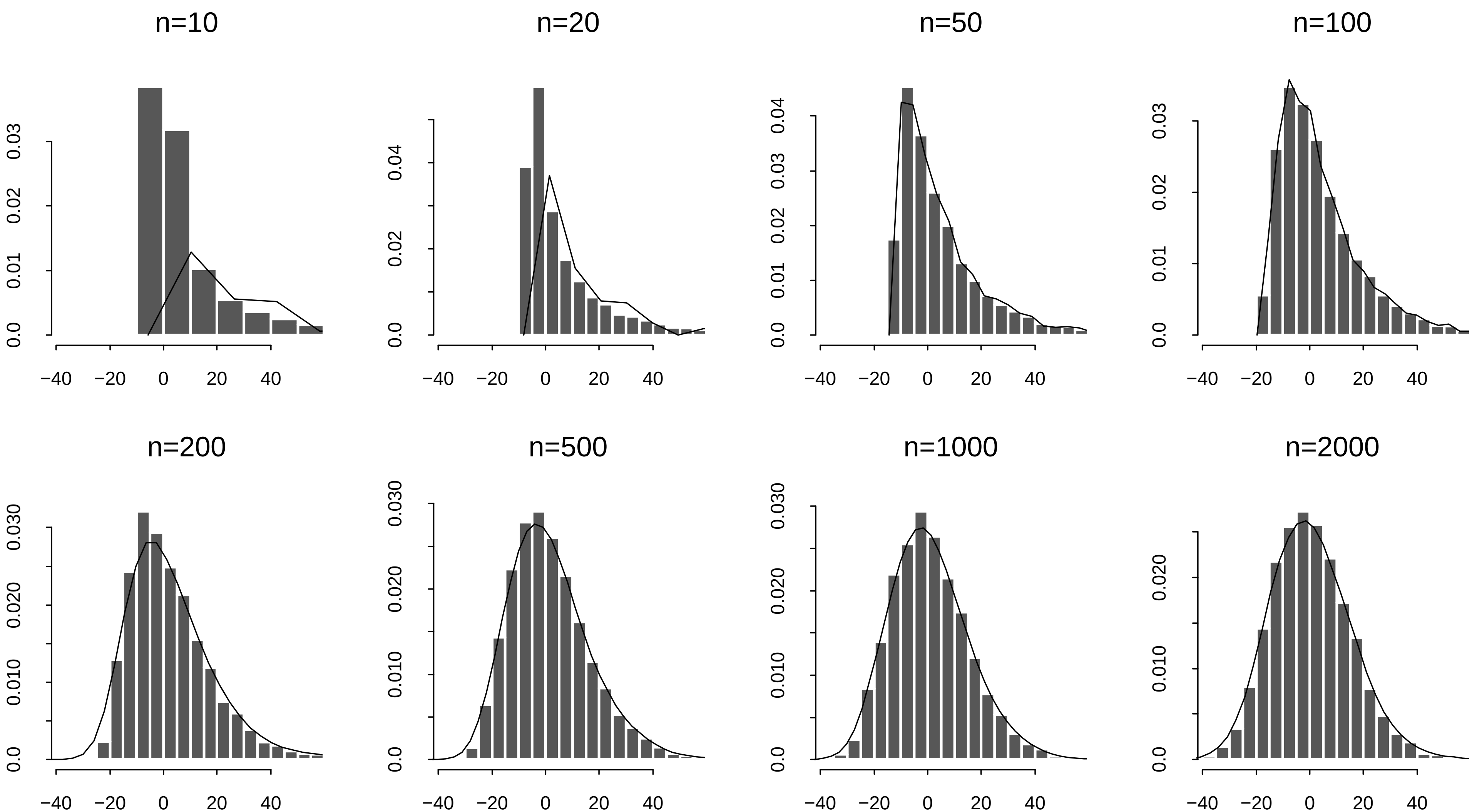
$$D\{\mathcal{M}(\widehat{\mathbf{A}}),\ \mathcal{M}(\mathbf{A})\} = 1 - \frac{1}{r}\,\text{tr}(\mathbf{A}\mathbf{A}^\tau \widehat{\mathbf{A}}\widehat{\mathbf{A}}^\tau).$$

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

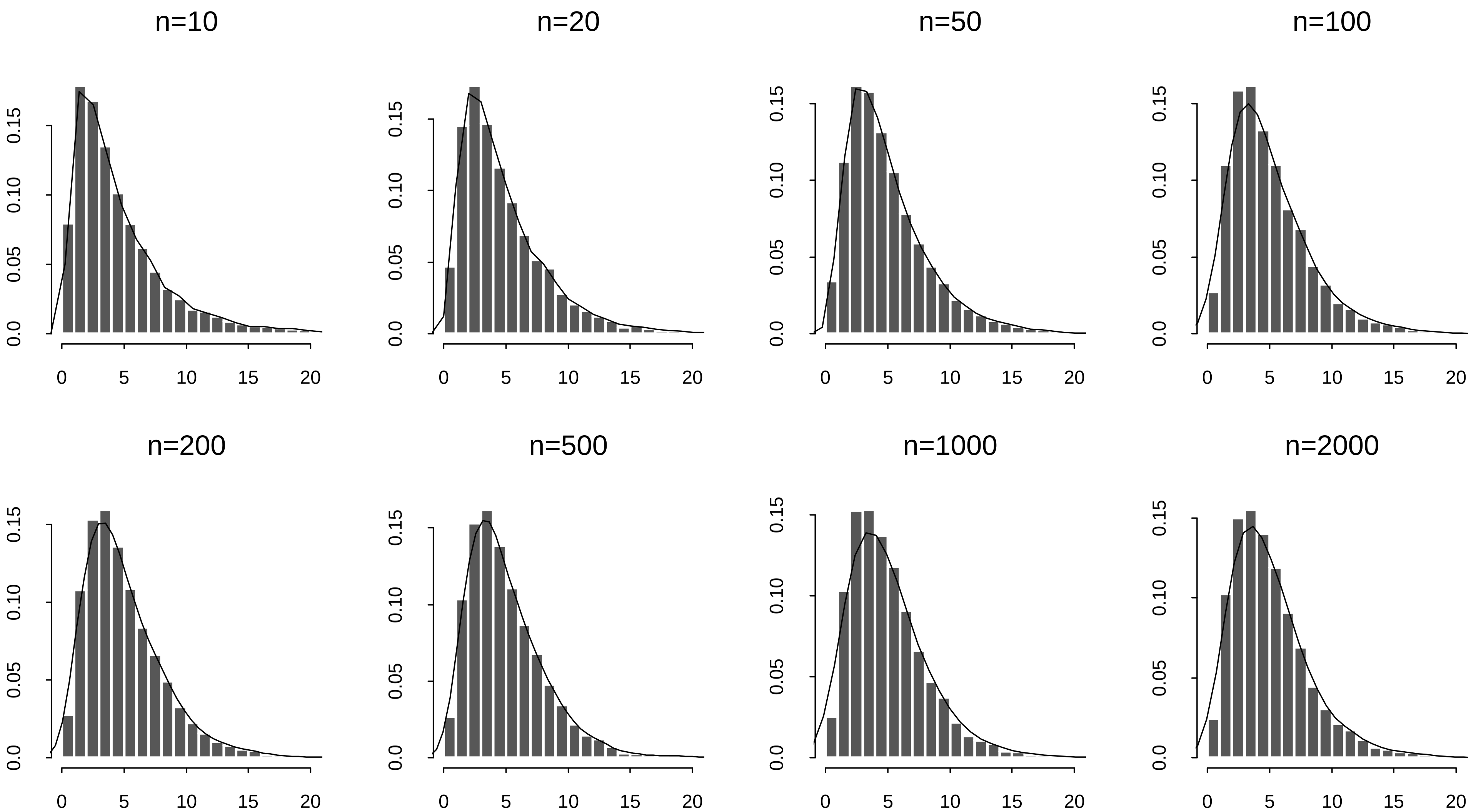
(Simulation replications: 10,000 times)



Histogram of $\sqrt{n}(\widehat{\lambda}_1 - \lambda_1)$



Histogram of $n\widehat{\lambda_2}$



Asymptotics II: $n \to \infty$, $p \to \infty$ 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) $\mathbf{A} = (\mathbf{a}_1, \cdots, \mathbf{a}_r)$, $||\mathbf{a}_i||^2 \asymp p^{1-\delta}, \ \ i = 1, \cdots, r, \ \ 0 \le \delta \le 1.$

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

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

where $a \asymp b$ represents $a = O(b)$ & $b = O(a)$, $\|\mathbf{A}\|^2 = \lambda_{\max}(\mathbf{A}\mathbf{A}^\tau)$ and $\|\mathbf{A}\|^2_{\min} = \min\{\lambda(\mathbf{A}\mathbf{A}^\tau) : \lambda(\mathbf{A}\mathbf{A}^\tau) > 0\}$.

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

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

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

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

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

Then under condition C1 and C2,

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

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 $\boldsymbol{\Sigma}_y$:

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

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

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

2. When the factors are of different levels of strengths, two (or multiple) step estimation procedure is preferred to estimate and to remove the stronger factors first.

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

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

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

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

$n = 200$	$\|\tilde{\mathbf{\Sigma}}_y - \mathbf{\Sigma}_y\|$	$\|\widehat{\mathbf{\Sigma}}_y - \mathbf{\Sigma}_y\|$
$p = 20$	$218(165)$	$218(165)$
$p = 180$	$1962(1500)$	$1963(1500)$
$p = 400$	$4102(3472)$	$4103(3471)$
$p = 1000$	$10797(6820)$	$10800(6818)$

Illustration With Real Data

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

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

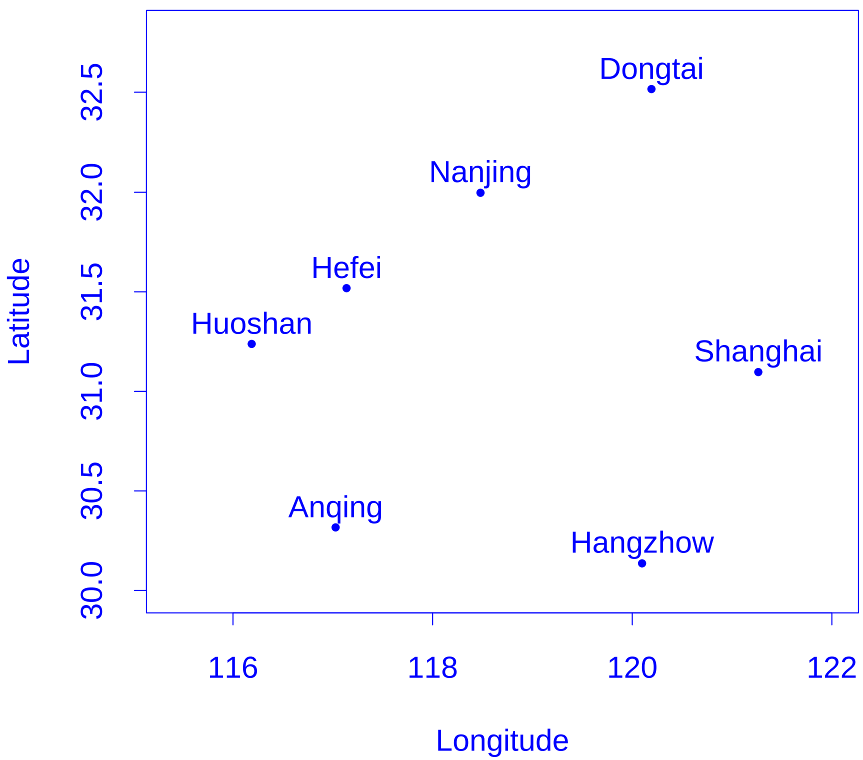
Example 2. Daily implied volatility surfaces for IBM, Microsoft and Dell call options in 2006

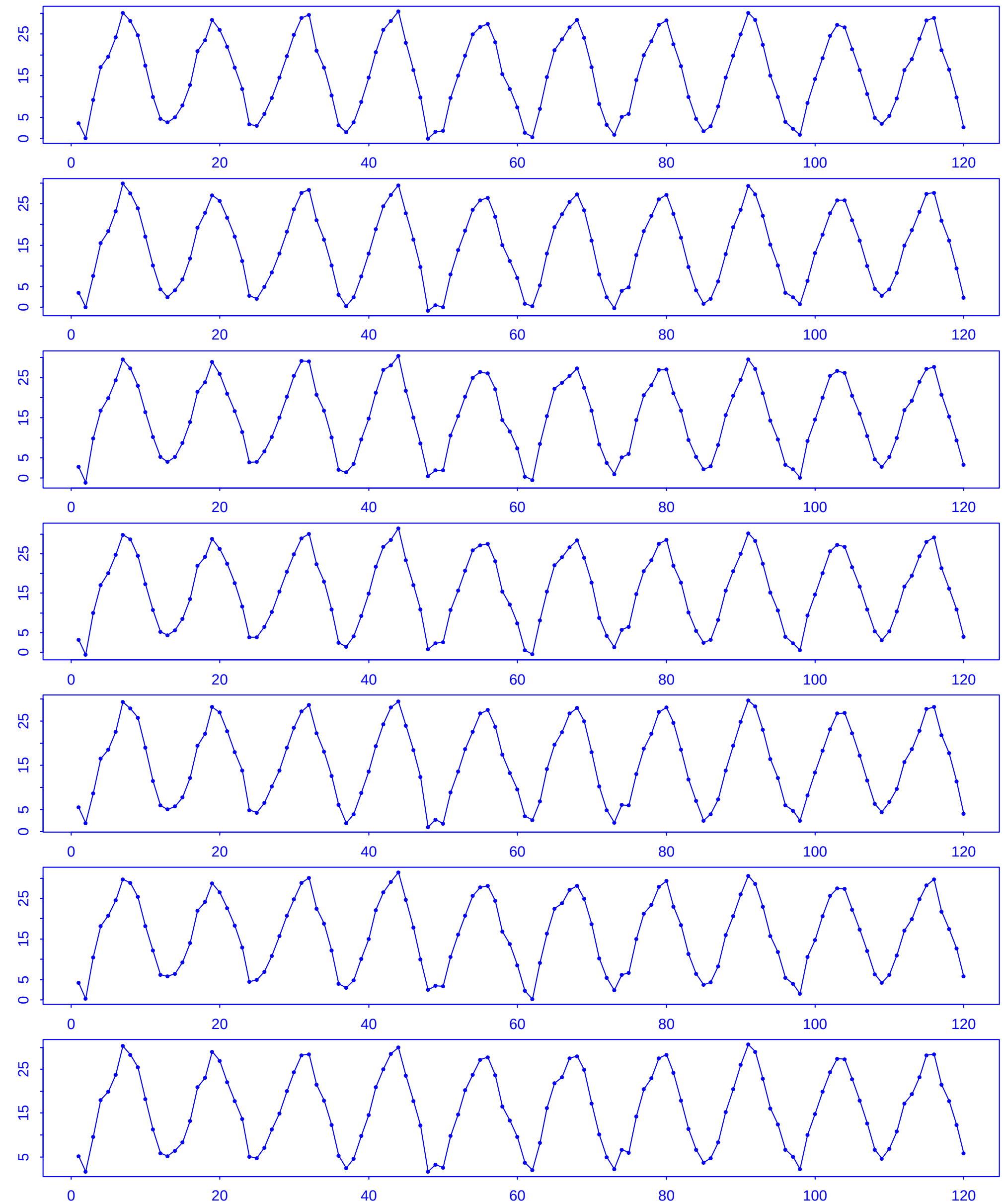
$$n = 100, \quad p = 130, \quad \widehat{r} = 1$$

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

$$n = 251, \quad p = \text{“}\infty\text{”}, \quad \widehat{r} = 2$$

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





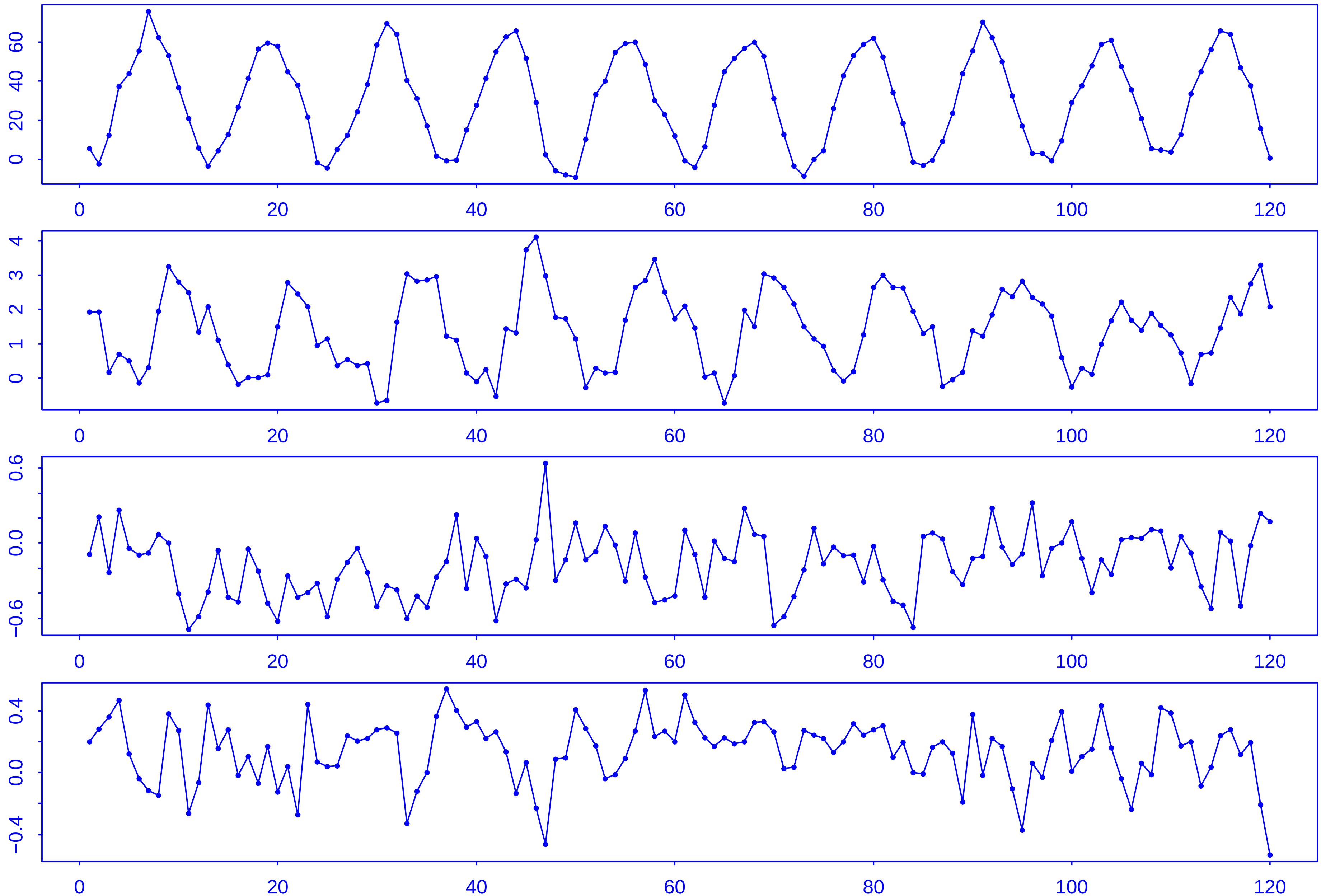
With $p = 12$, $\alpha = 1\%$, the fitted model is $\mathbf{y}_t = \widehat{\mathbf{A}}\mathbf{x}_t + \mathbf{e}_t$, $\widehat{r} = 4$,
$\mathbf{e}_t \sim \mathrm{WN}(\widehat{\boldsymbol{\mu}}_\varepsilon, \widehat{\boldsymbol{\Sigma}}_\varepsilon)$,

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

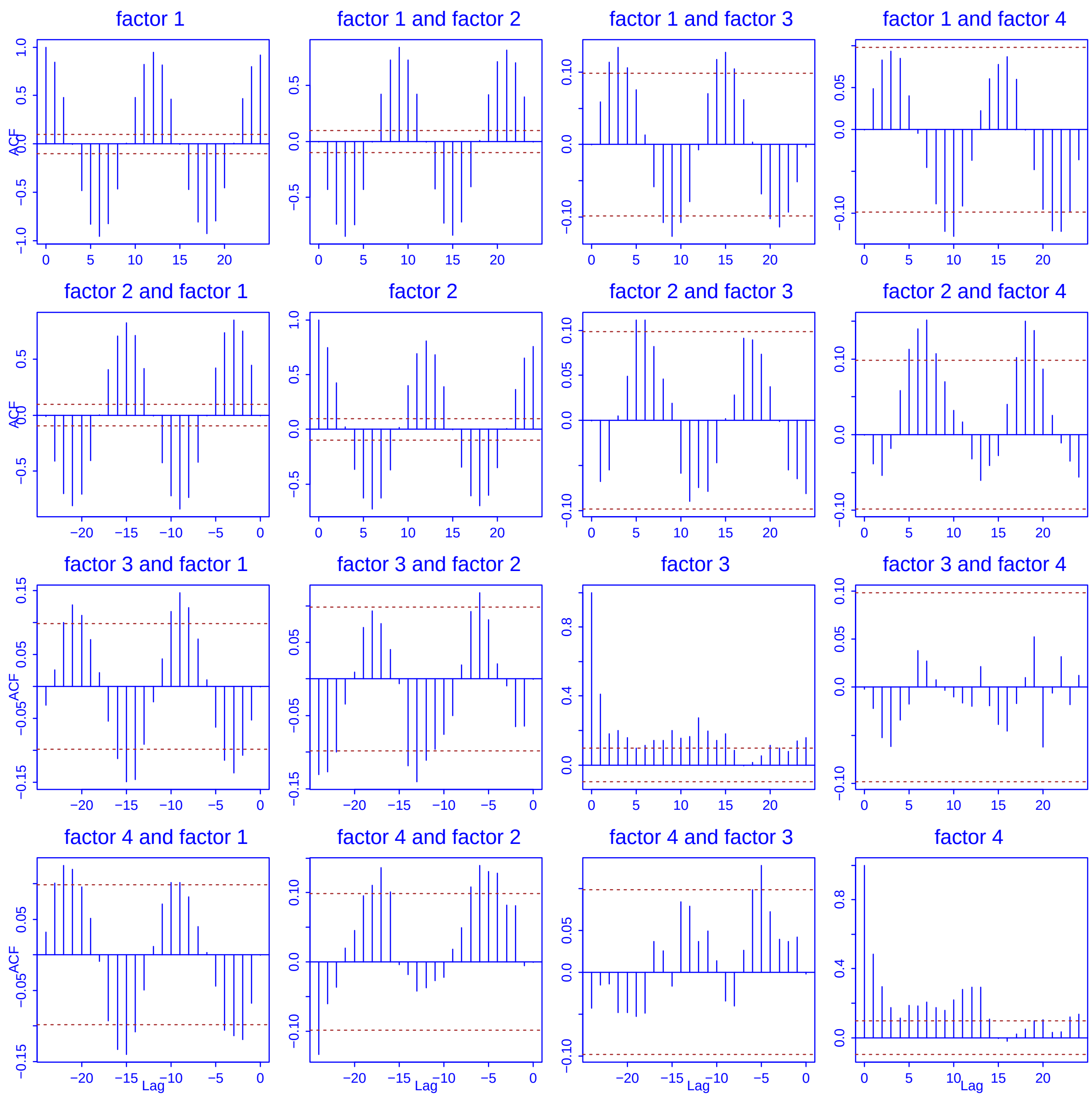
$$\widehat{\mathbf{A}} = \begin{pmatrix} .394 & .386 & .378 & .387 & .363 & .376 & .366 \\ -.086 & .225 & -.640 & -.271 & .658 & -.014 & .164 \\ .395 & .0638 & -.600 & .346 & -.494 & -.074 & .332 \\ .687 & -.585 & -.032 & -.306 & .173 & .206 & -.139 \end{pmatrix}^{\tau},$$

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

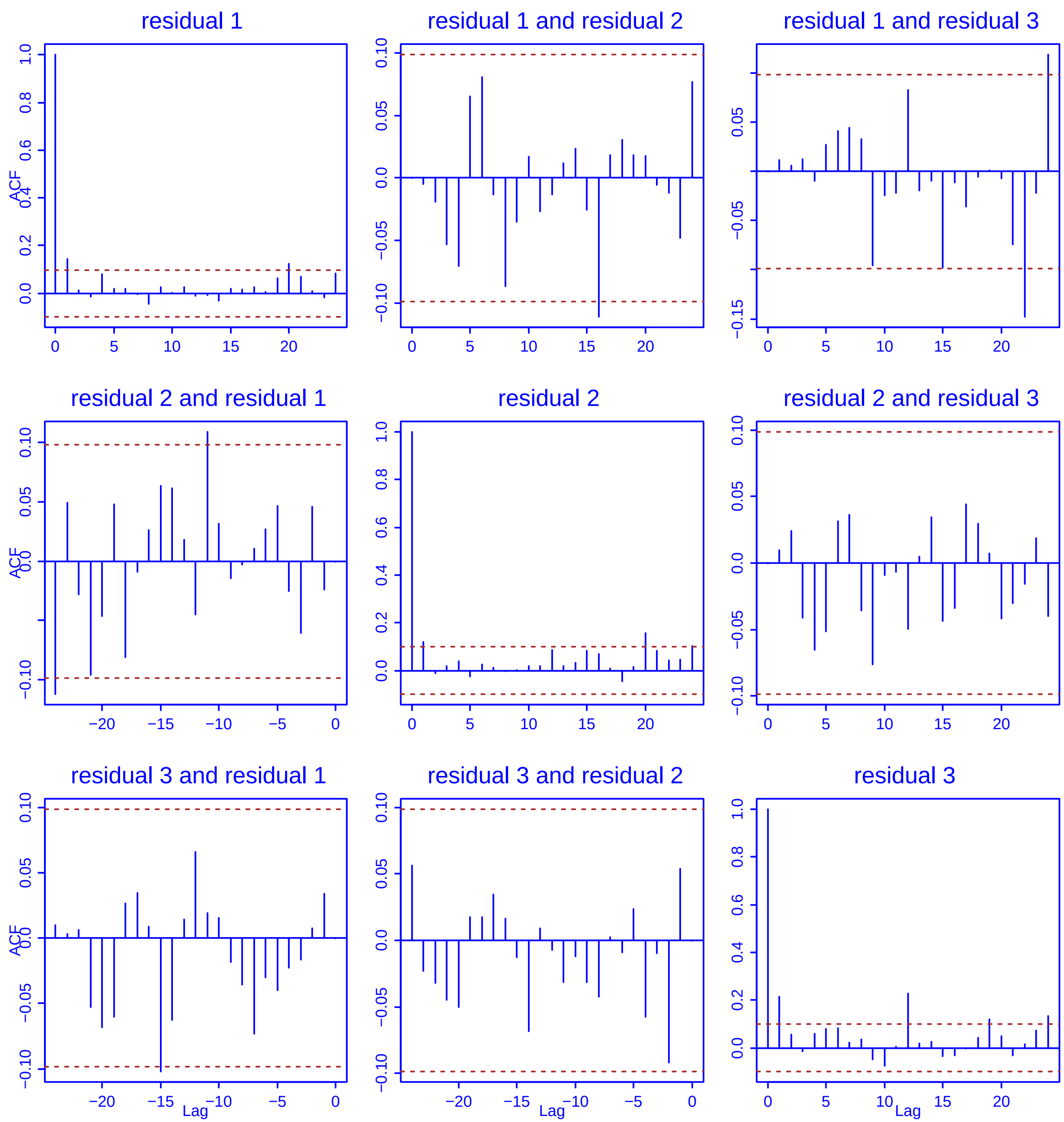
Time plots of the 4 estimated factors VAR(1)



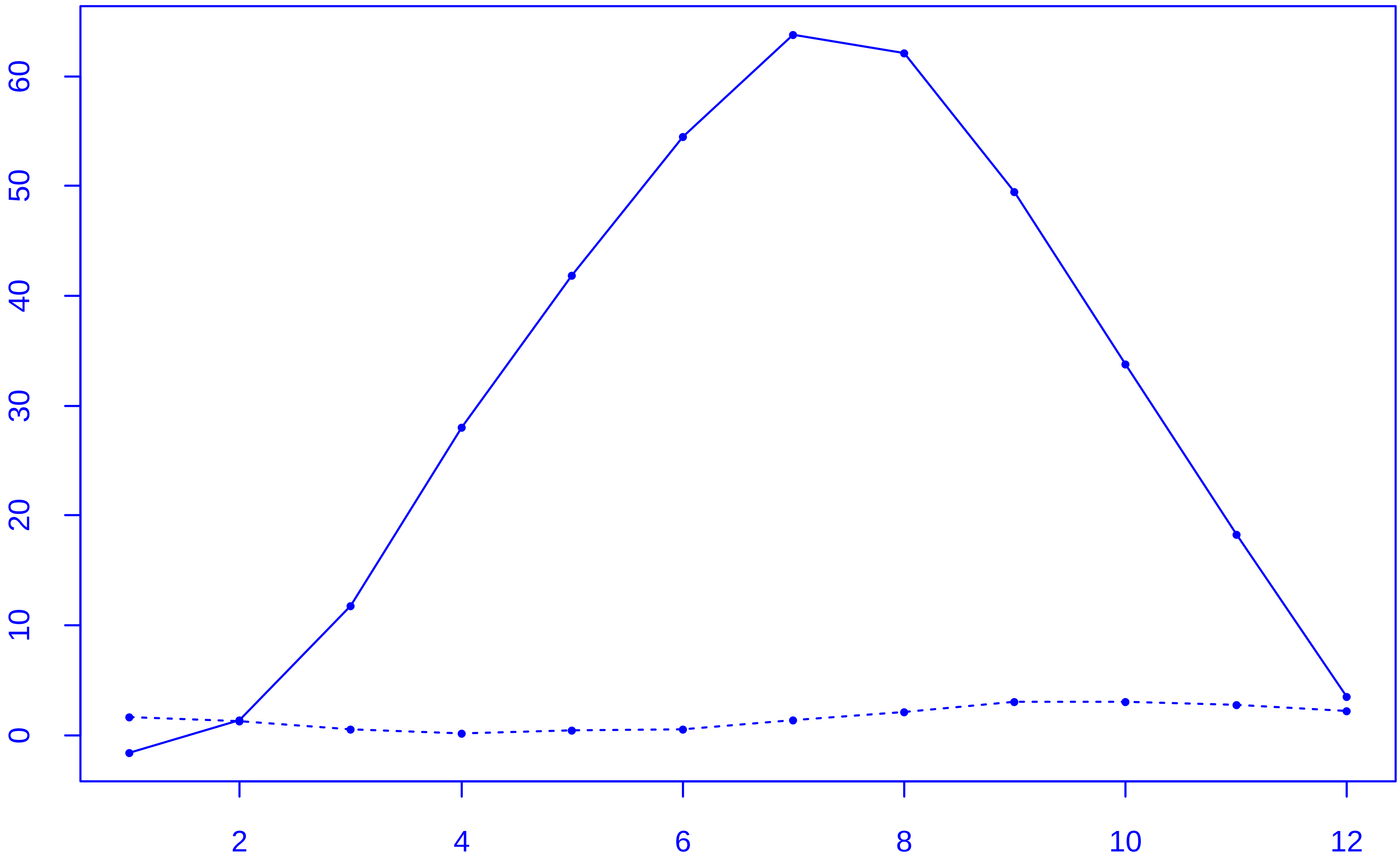
Sample cross-correlation of the 4 estimated factors



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



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



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

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

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

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

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

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

$\Rrightarrow$

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

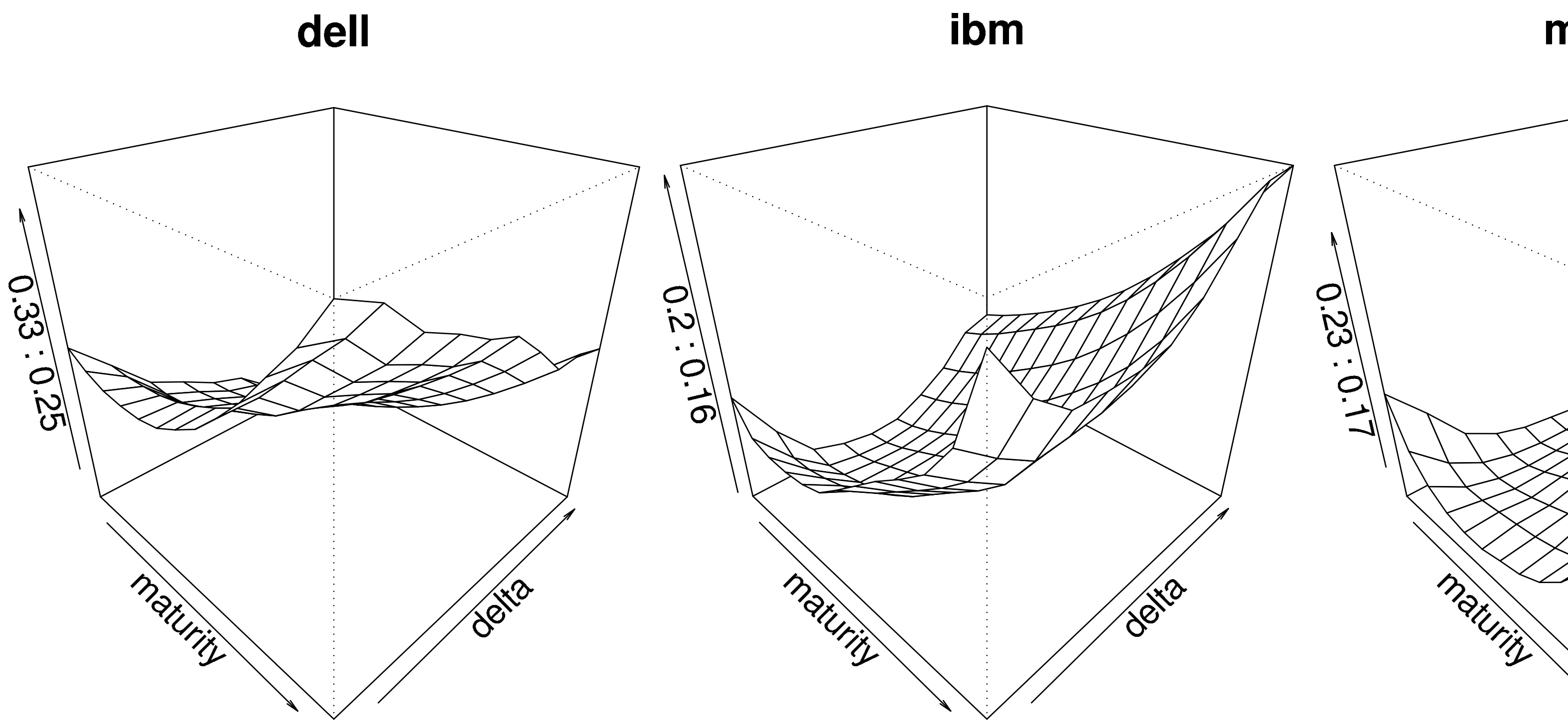
Source of Data: OptionMetrics at WRDS

Observations: for $t = 1, \cdots, 251$, 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 $u_1, \cdots, u_{10}$, and
- delta at 0.2, 0.25, 0.3, 0.35, 0.4, $\cdots$, 0.8, denoted by $v_1, \cdots, v_{13}$.

Total: $p = 10 \times 13 = 130$ time series with the length 251 each.

Take difference $\breve{w}_t(u_i, v_j) = w_t(u_i, v_j) - w_{t-1}(u_i, v_j)$, and vectorize the matrix $\big(\breve{w}_t(u_i, v_j)\big)$ into a 130×1 vector process $\mathbf{y}_t$.



Average daily implied volatility surfaces over 251 days.

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

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

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

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 $\mathbf{y}_{i+100}$:

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

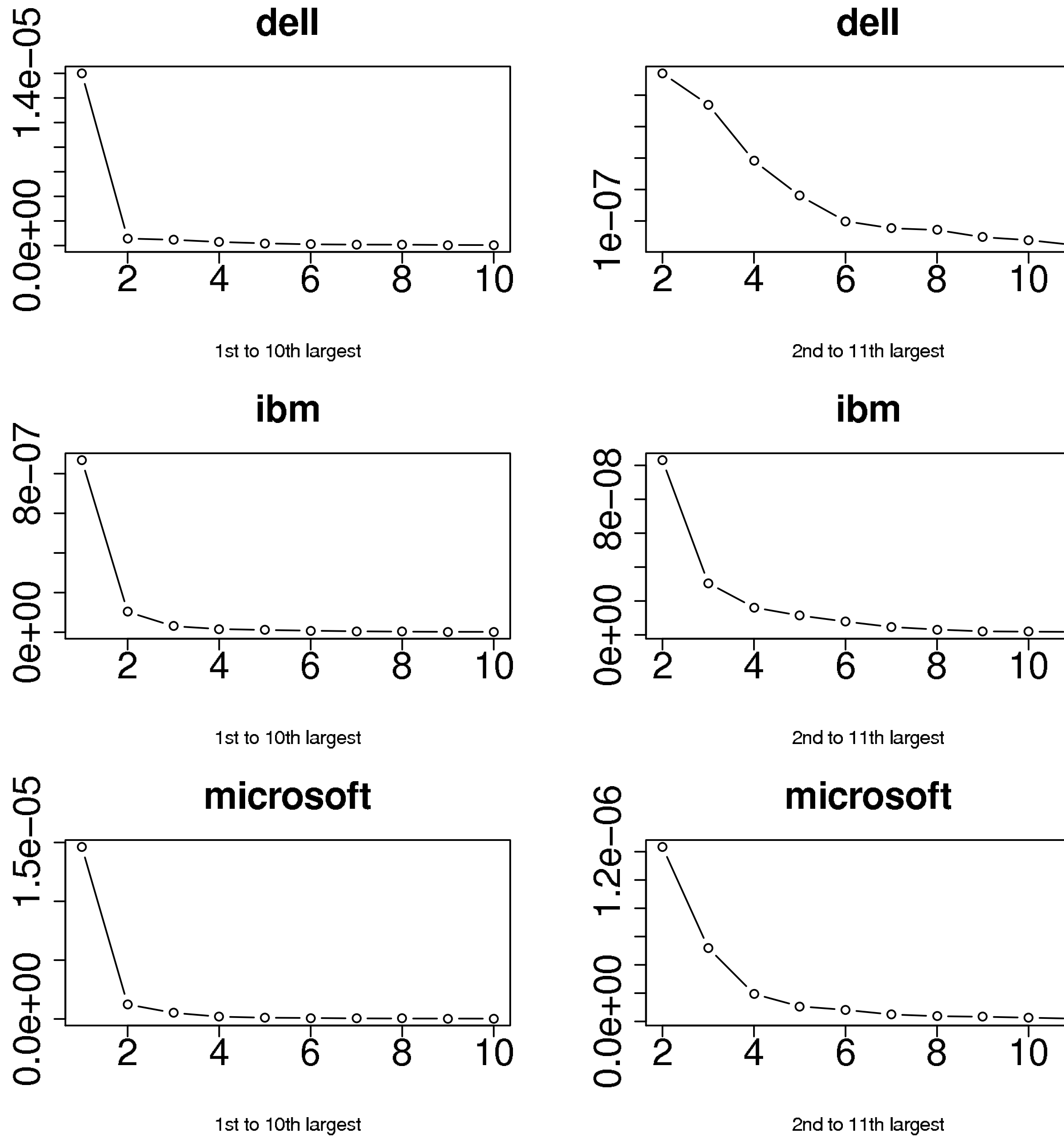
Put

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

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

3 panels on the left: 10 largest eigenvalues

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

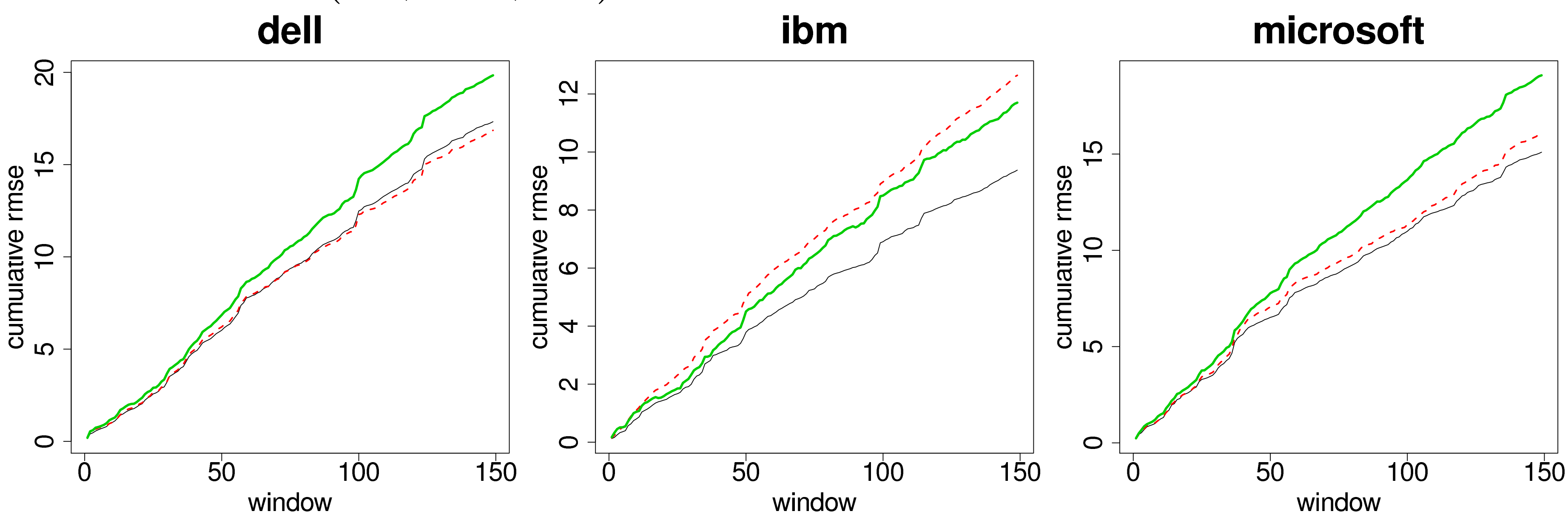


Benchmark prediction for $\mathbf{y}_{i+100}$: the previous value $\mathbf{y}_{i+99}$

Prediction based on Bai & Ng (2002) — factor-modelling based on the LSE: $(\widehat{\mathbf{A}}, \widehat{\mathbf{x}}_t)$ 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 $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$.



The cumulative RMSE over the 150 windows. Red dotted – Bai & Ng (2002), Green – benchmark, Black – our method.

Example 3. IBM stock intra-day prices in 2006

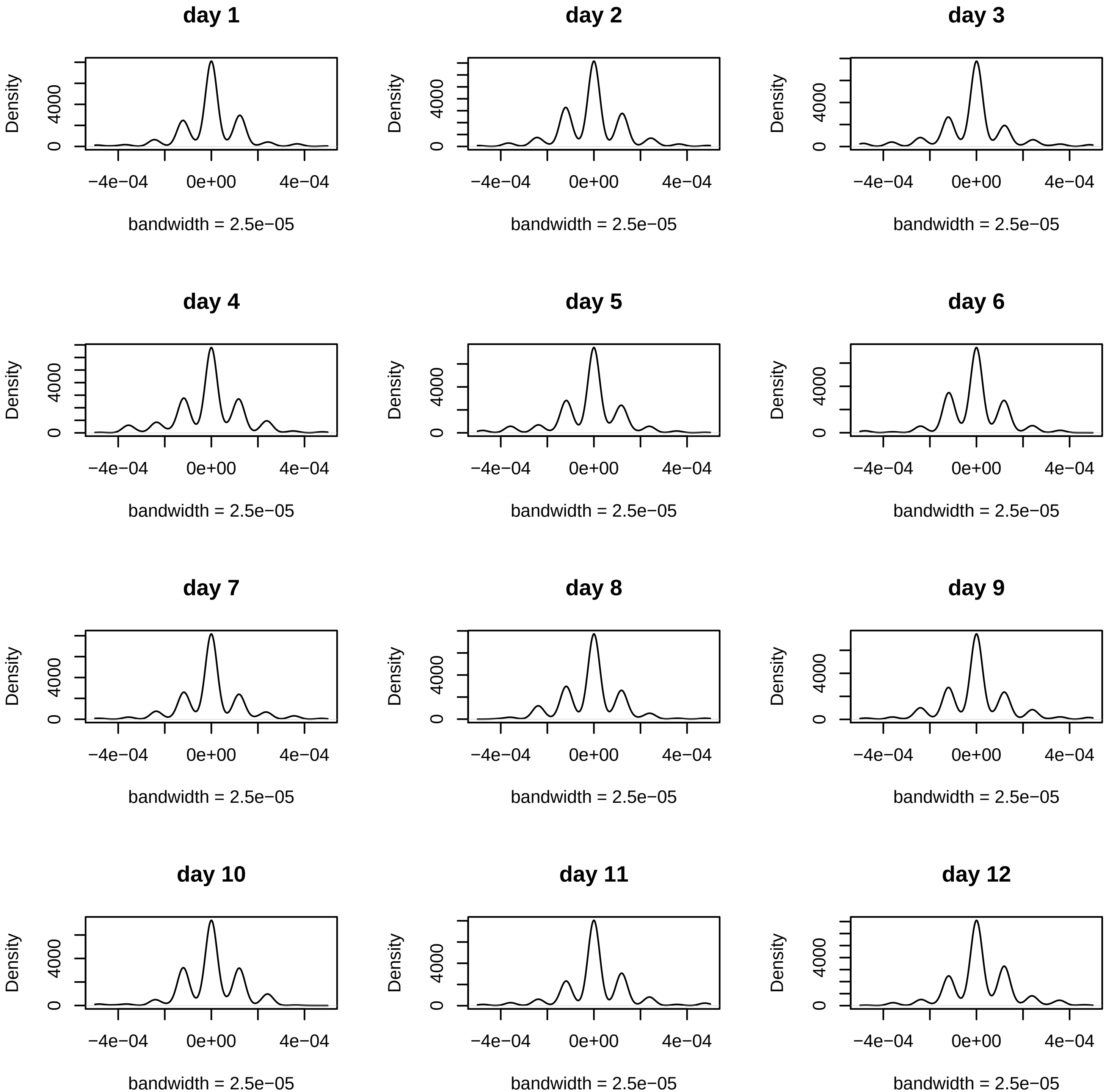
251 trading days, tick by tick prices collected in 9:30 — 16:00

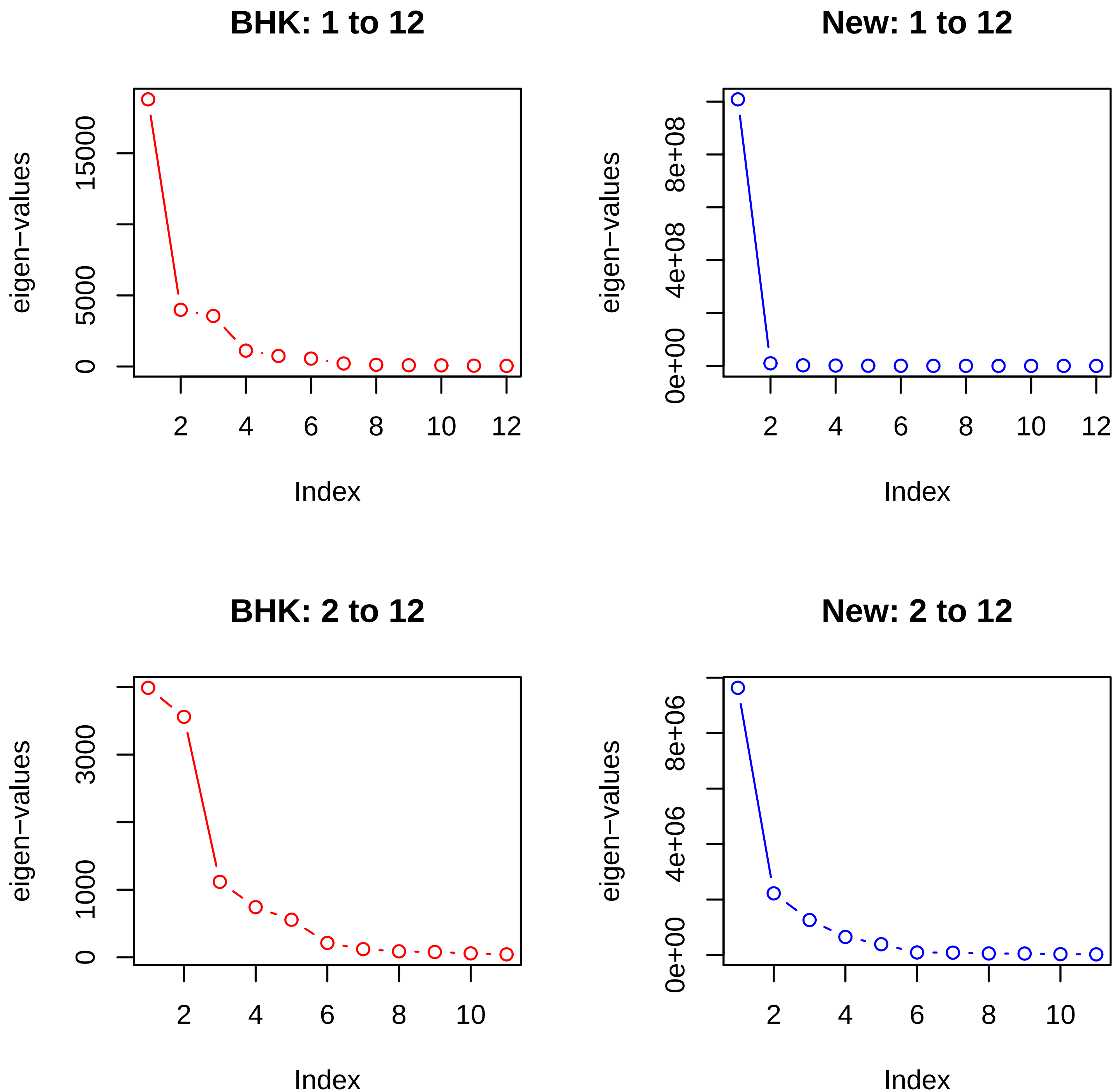
In total 2,786,649 observations (74MB)

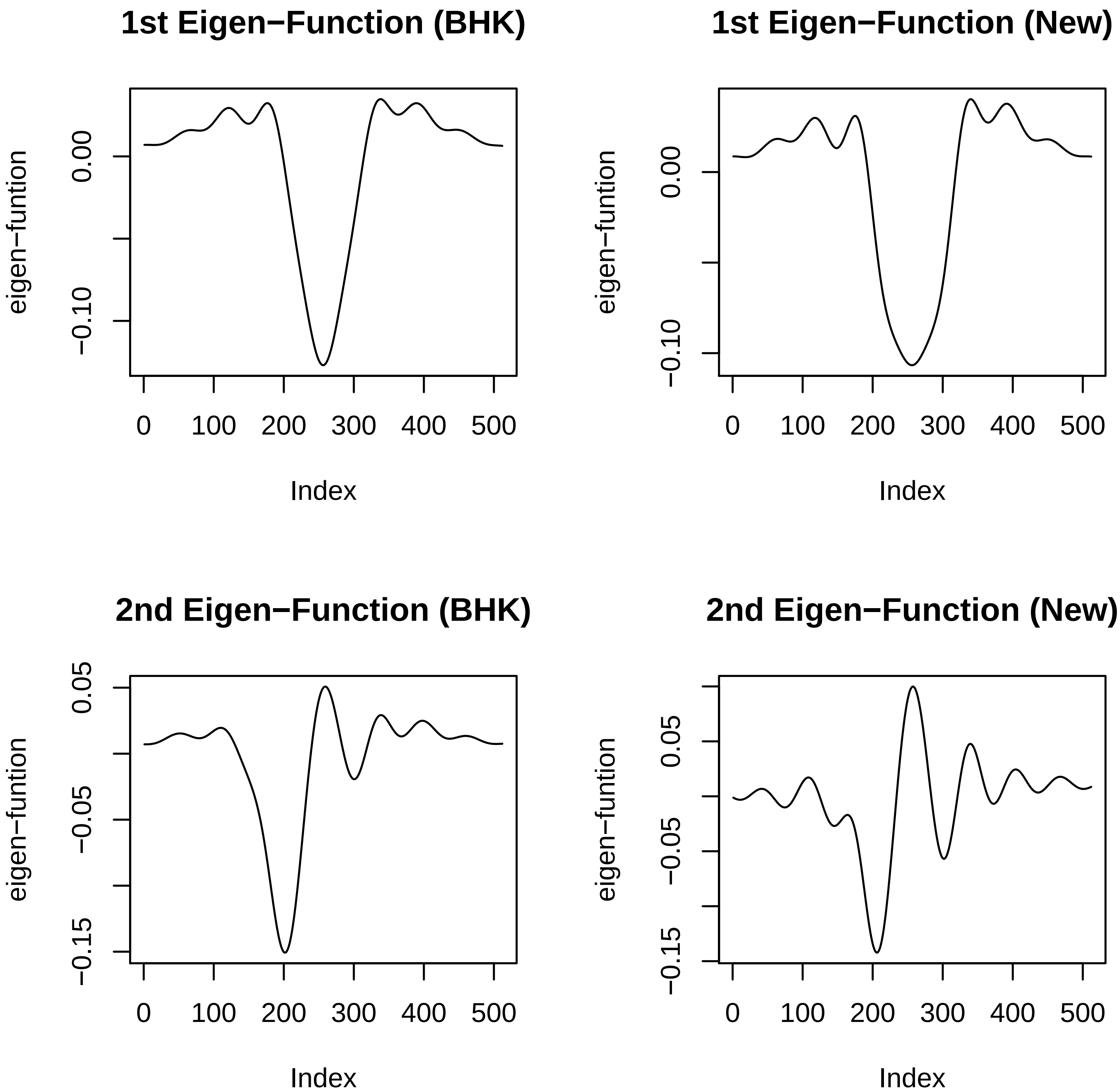
For each of 251 trading days, construct the pdf curve of one-minute log-return using the log-returns in 390 one-minute intervals: kernel density estimation with $h = 0.000025$

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

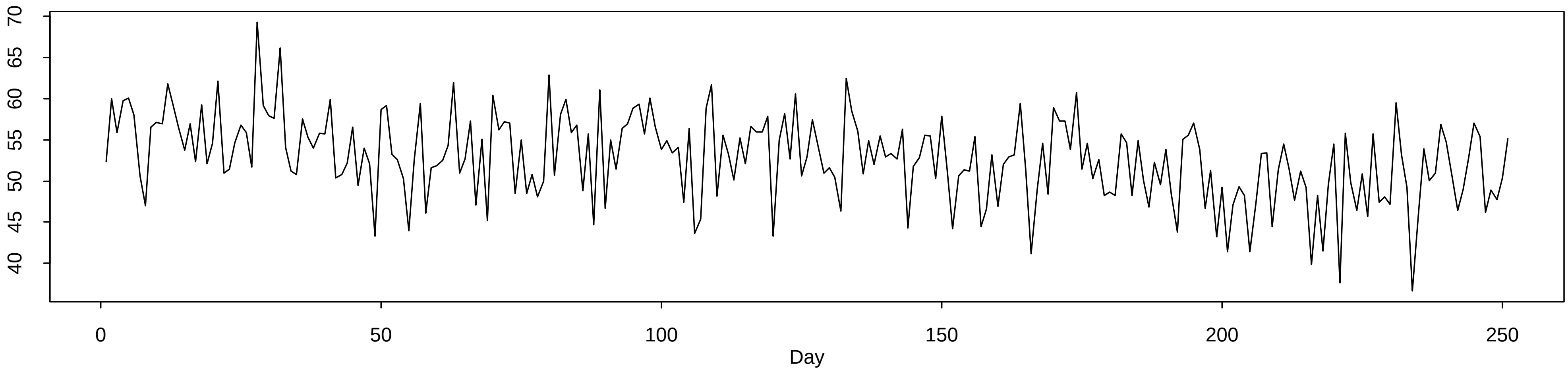
The white-noise test rejects $H_0 : r = 1$, but cannot reject $H_0 : r = 2$.

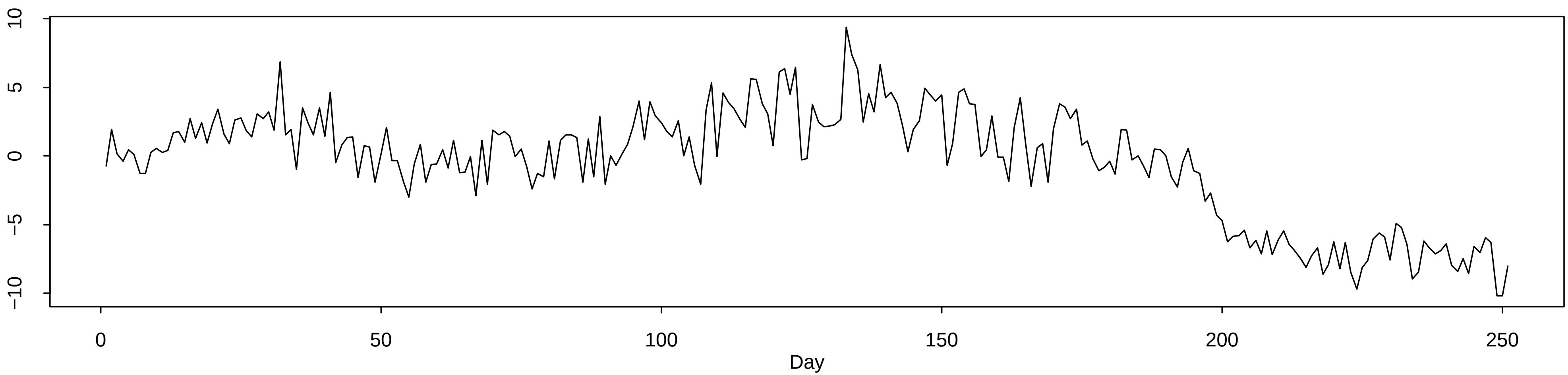

day 1
day 2
day 3
day 4
day 5
day 6
day 7
day 8
day 9
day 10
day 11
day 12
Density
0
4000
−4e−04
0e+00
4e−04
bandwidth = 2.5e−05

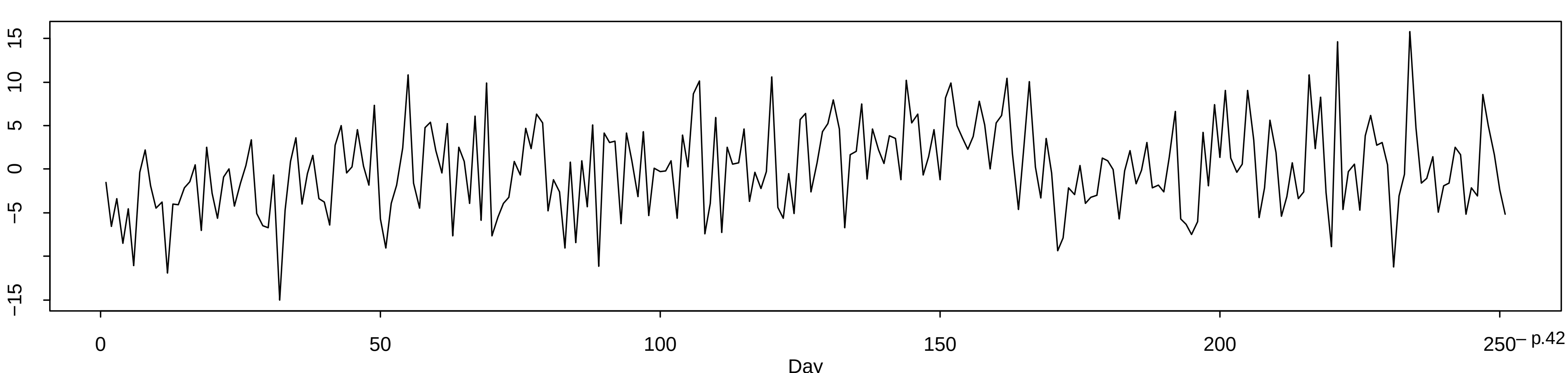

BHK: 1 to 12
eigen−values
0
5000
15000
2
4
6
8
10
12
Index
New: 1 to 12
eigen−values
0e+00
4e+08
8e+08
2
4
6
8
10
12
Index
BHK: 2 to 12
eigen−values
0
1000
3000
2
4
6
8
10
Index
New: 2 to 12
eigen−values
0e+00
4e+06
8e+06
2
4
6
8
10
Index


1st Eigen−Function (BHK)
1st Eigen−Function (New)
2nd Eigen−Function (BHK)
2nd Eigen−Function (New)
eigen−funtion
Index
0.00
−0.10
0.05
−0.05
−0.15
0
100
200
300
400
500

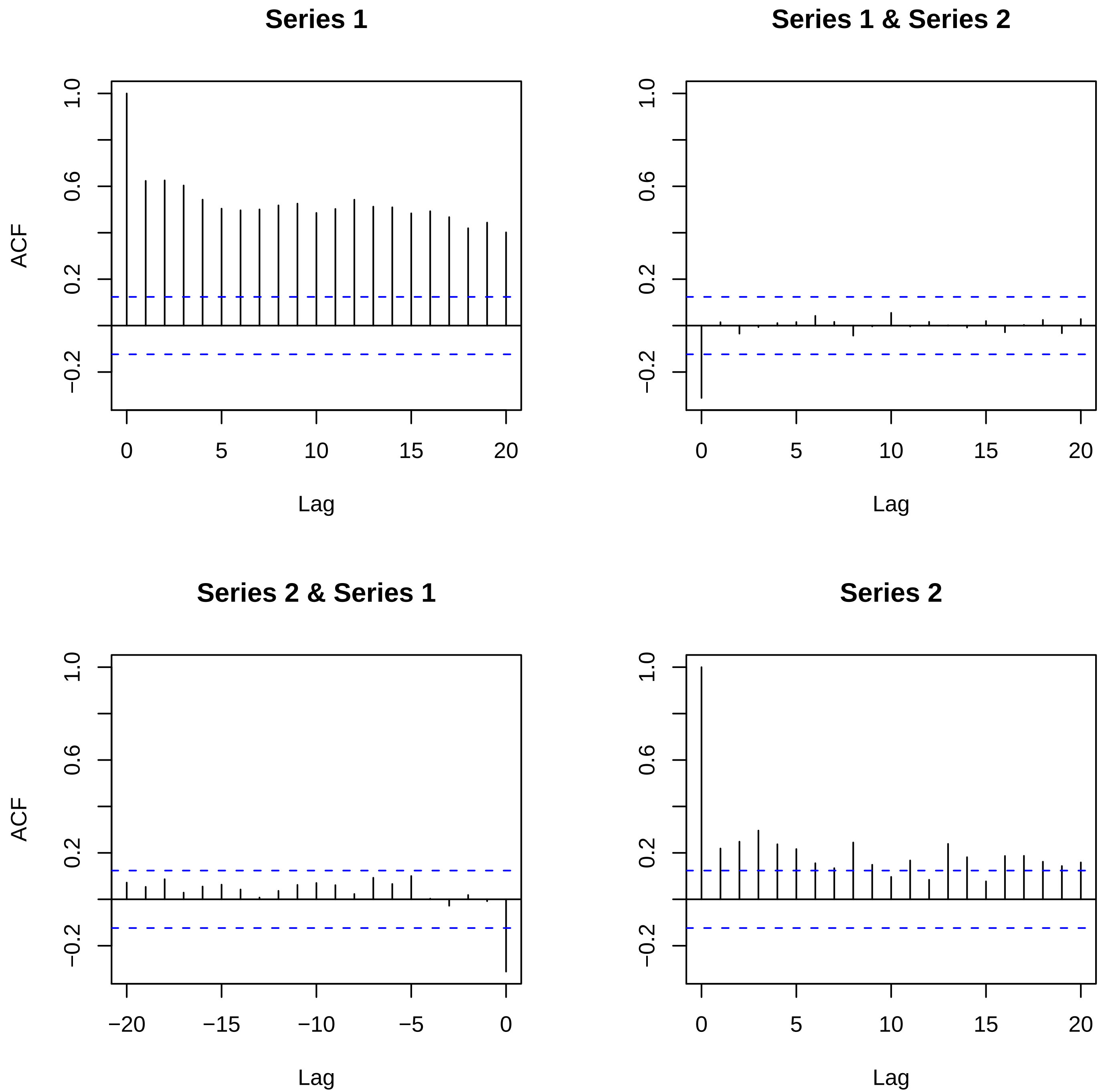
Time series plots of x_{t1} and x_{t2}



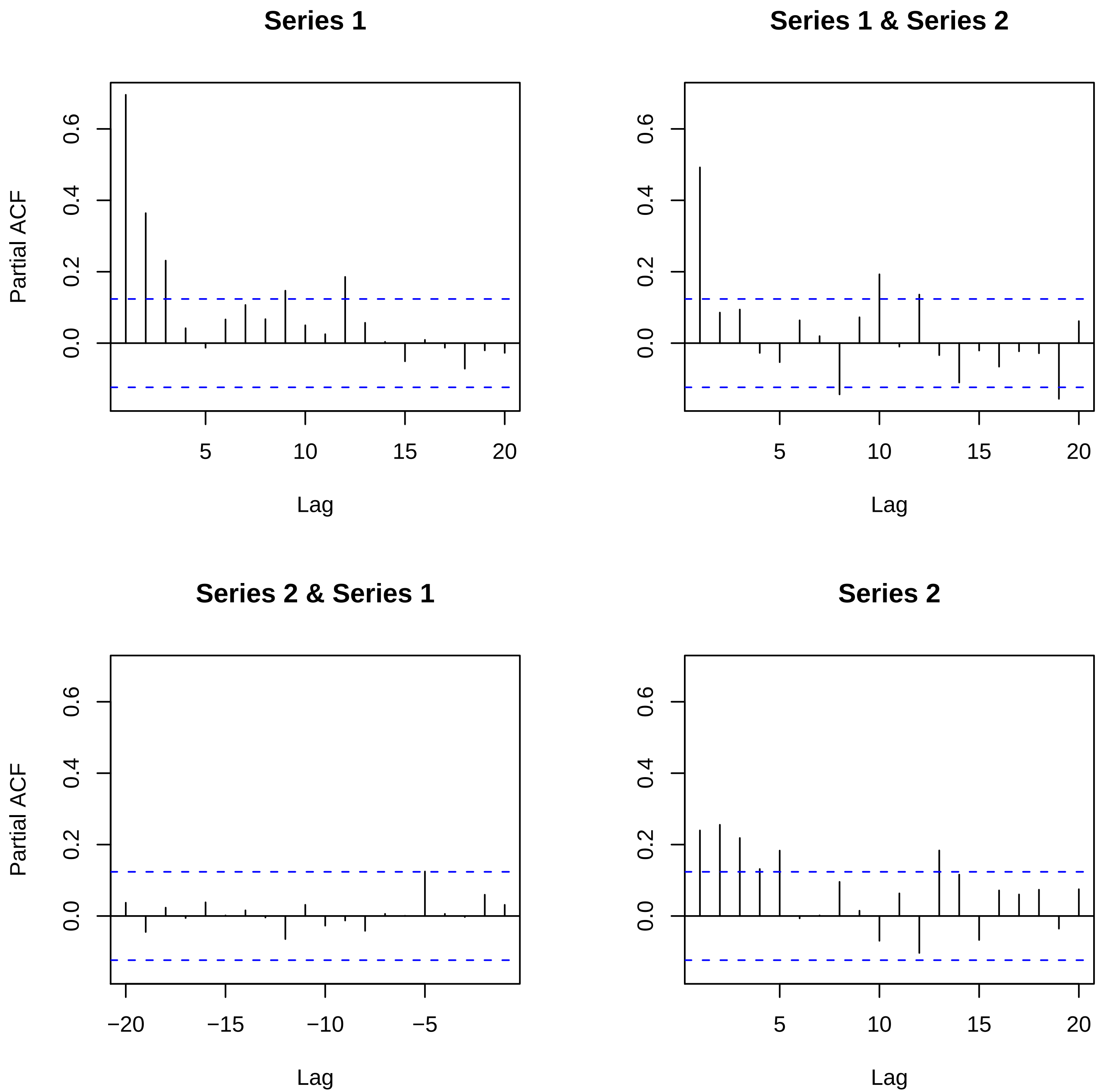




ACF of (x_{t1}, x_{t2})



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



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