Cointegration

Qiwei Yao

Department of Statistics

London School of Economics

q.yao@lse.ac.uk

Fan & Yao (2015). The Elements of Financial Econometrics, §4.3

Zhang, Robinson and Yao (2015). Identifying Cointegration by Eigenanalysis.

- Unit roots and cointegration
- Engle-Granger method and error correction models
- Johansen's likelihood inference

- Identification by eigenanalysis
- Asymptotic properties
- Numerical illustration
- Extension to fractional integrated orders

Cointegration: each individual time series is not stationary, but some linear combinations are.

Reflecting *co-movements* in either the same or opposite directions.

Box and Tiao (1977, *Biometrika*, p.355-)

Granger (1981, 1983), Engle and Granger (1987)

$\mathrm{I}(k)$ processes: $\nabla^k X_t$ is stationary (ARMA).

A vector time series $\mathbf{X}_t$ is said to be cointegrated with order (k, h) ($k \geq h \geq 1$), denoted as $\mathbf{X}_t \sim \mathrm{CI}(k, h)$, if

(i) all component series of $\mathbf{X}_t$ are $\mathrm{I}(k)$, and,

(ii) there exists a non-zero vector $\boldsymbol{\beta}$ such that $\boldsymbol{\beta}'\mathbf{X}_t \sim \mathrm{I}(k - h)$.

CI$(1, 1)$ is the most frequently used cointegration model.

ECM: Engle and Granger Method

Let X_t and Y_t be I(1). Engle & Granger (1987) suggests to run regression

$$Y_t = \alpha + \beta X_t + Z_t.$$

As both X_t and Y_t non-stationary, the LSE $(\widehat{\alpha}, \widehat{\beta})$ are not asymptotically normal. So t- or F-tests do not apply.

Apply the *cointegration augmented Dickey-Fuller* (CADF) test to

$$\widehat{Z}_t \equiv Y_t - \widehat{\alpha} - \widehat{\beta} X_t,$$

the cointegration is identified if the unit-root hypothesis is rejected. Then fit ECM:

$$\nabla Y_t = a_0 + a_1 \widehat{Z}_{t-1} + a_2 \nabla X_{t-1} + a_3 \nabla Y_{t-1} + \varepsilon_{t1},$$

$$\nabla X_t = b_0 + b_1 \widehat{Z}_{t-1} + b_2 \nabla Y_{t-1} + b_3 \nabla X_{t-1} + \varepsilon_{t2},$$

Now all terms in the models are stationary.

Spurious linear regression

If Y_t and X_t are two independent random walks, $\beta = 0$ in $Y_t = \alpha + \beta X_t + Z_t$.

However, LSE $\widehat{\beta}$ converges in distribution to a function of a Brownian motion, and usual t-statistic for testing $H_0 : \beta = 0$ diverge at rate $T^{1/2}$, hence H_0 is rejected (Phillips, 1986).

In practice spurious regression phenomenon is encountered when we run regression for two I(1) processes.

Apply a unit-root test to $\widehat{Z}_t$ first, and only proceed to fitting ECM when the unit-root hypothesis for $\widehat{Z}_t$ is rejected.

When Z_t is I(0) and both Y_t and X_t are I(1), $\widehat{\beta}$ is n-consistent, the t-test statistic is often not normal and may be asymmetric. One alternative is to estimate β by $\nabla Y_t = \beta \nabla X_t + \nabla Z_t$.

A simple ECM: A drunk and his/her dog

Drunk's position: $X_0 = 0$, $X_t = X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \mathrm{IID}(0,1)$.

Dog's position: $Y_t = X_t + Z_t$, $Z_t \sim \mathrm{IID}(0,\sigma^2)$, Z_t, X_t are indep.

Both $X_t,\ Y_t$ are I(1), X_t is a random walk and Y_t is not, and Z_t is an error correction term.

Goal: 1-step-ahead prediction for dog's position Y_t

Since Y_{t-1} is known at time $t-1$, only need to predict

$$\nabla Y_t = Y_t - Y_{t-1} = \nabla X_t + \nabla Z_t = \nabla Z_t + \varepsilon_t.$$

Note $\varepsilon_t \sim \mathrm{IID}(0,1)$ is unpredictable, $\nabla Z_t \sim \mathrm{MA}(1)$,

$$\mathrm{Corr}(\nabla Y_t, \nabla Y_{t-1}) = \frac{\mathrm{Cov}(\nabla Z_t, \nabla Z_{t-1})}{\mathrm{Var}(\nabla Y_t)} = \frac{-\sigma^2}{(1+2\sigma^2)}.$$

$$\mathrm{Corr}(\nabla Y_t, Z_{t-1}) = \frac{\mathrm{Cov}(\nabla Z_t, Z_{t-1})}{\{\mathrm{Var}(\nabla Y_t)\mathrm{Var}(Z_{t-1})\}^{1/2}} = \frac{-\sigma}{\sqrt{(1+2\sigma^2)}} = \sqrt{2+\frac{1}{\sigma^2}}\,\mathrm{Corr}(\nabla Y_t, \nabla Y_{t-1}).$$

A drunk and his/her dog: Numerical illustration

To generate two paths for drunk and dog:

```
> n=200; e=rnorm(n)
> X=cumsum(e) # Drunk's random walk path
> Y=X+rnorm(n, 0, 1/2) # Dog's path, sigma=1/2
```

To extract the error-correction terms $\widehat{Z}_t$:

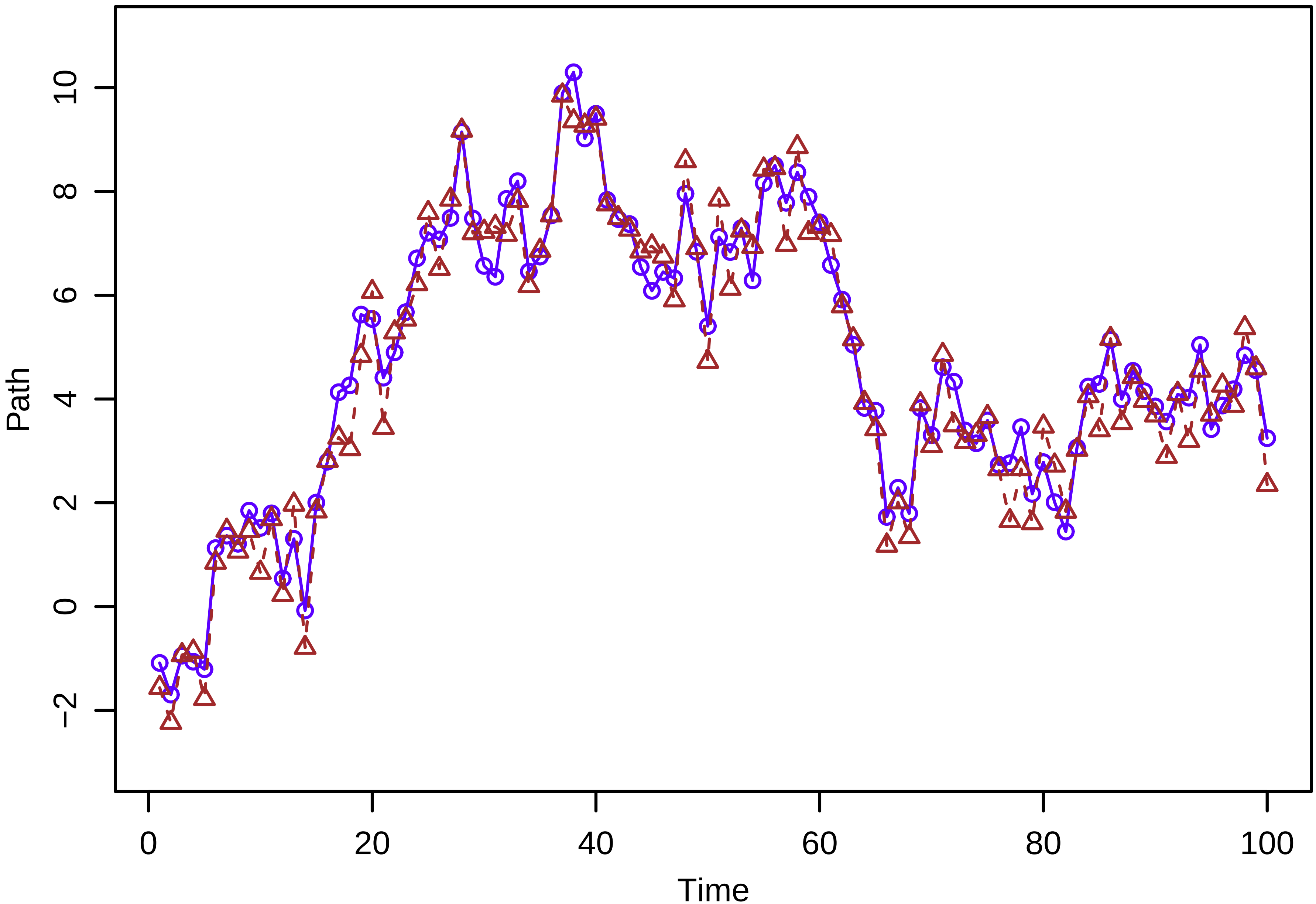
```
> t=lm(Y~X); Z1=residuals(t)
> hatZ = Z1[-c(1,n)] # leave out the 1st and last eleme
```

The fitted cointegration: $\widehat{Z}_t = Y_t - 0.0285 - 0.9987X_t$

To line up the differenced data together,

```
> dX=diff(X)
> dY=diff(Y)
> YXZ=data.frame(embed(cbind(dY,dX),2), hatZ)
> colnames(YXZ)=c("dY0","dX0","dY1", "dX1", "hatZ")
> attach(YXZ)
```

A part of druck's path & Dog's path



To fit an ECM for ∇Y_t,

```
> ecm=lm(dY0~hatZ+dY1+dX1); summary(ecm)
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.09331    0.07557   1.235 0.218378
hatZ         -0.84048    0.25003  -3.361 0.000934 ***
dY1          -0.26102    0.16458  -1.586 0.114375
dX1           0.19575    0.18111   1.081 0.281093
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
Residual standard error: 1.059 on 194 degrees of freedom
Multiple R-squared:  0.2075,    Adjusted R-squared:  0.1952
F-statistic: 16.93 on 3 and 194 DF,  p-value: 8.294e-10
```

The re-fitted ECM: $\widehat{\nabla Y_t} = -1.123\widehat{Z}_{t-1}$ with $R^2 = 19.30\%$.

∇Y_{t-1}, ∇X_{t-1} and the intercept are not significant!

```
> anova(ecm)
Response: dY0
           Df  Sum Sq Mean Sq F value    Pr(>F)
hatZ        1  53.279  53.279 47.5093 7.484e-11 ***
dY1         1   2.364   2.364  2.1080    0.1482
dX1         1   1.310   1.310  1.1683    0.2811
Residuals 194 217.559   1.121

> lm(formula = dY0 ~ dY1)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.10931    0.08187   1.335 0.183380
dY1         -0.24110    0.06943  -3.472 0.000634 ***
---
Residual standard error: 1.149 on 196 degrees of freedom
Multiple R-squared: 0.05796,  Adjusted R-squared: 0.05315
F-statistic: 12.06 on 1 and 196 DF, p-value: 0.0006345
```

Johansen's likelihood method

Setting: A $d \times 1$ process $\mathbf{X}_t \sim \mathrm{CI}(1,1)$, for a $d \times r$ matrix

$$\mathbf{V} = (\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_r) \text{ with rank}(\mathbf{V}) = r < d$$

such that $\boldsymbol{\beta}_j' \mathbf{X}_t \sim I(0)$ for $j = 1, \cdots, r$.

The ECM implied by the Granger representation theorem:

$$\nabla \mathbf{X}_t = \mathbf{W}\mathbf{V}'\mathbf{X}_{t-1} + \sum_{i=1}^{p} \mathbf{A}_i \nabla \mathbf{X}_{t-i} + \boldsymbol{\varepsilon}_t,$$

where both $\mathbf{W}$ and $\mathbf{V}$ are $d \times r$ with ranks r, $\boldsymbol{\varepsilon}_t \sim \mathrm{WN}(0, \boldsymbol{\Sigma}_\varepsilon)$.

Let $\mathbf{Z}_{t0} = \nabla \mathbf{X}_t,\ \mathbf{Z}_{t1} = \mathbf{X}_t,\ \mathbf{Z}_{t2} = (\nabla \mathbf{X}_{t-1}', \cdots, \nabla \mathbf{X}_{t-p}')',\ \mathbf{A} = (\mathbf{A}_1, \cdots, \mathbf{A}_p)$. Then the ECM is

$$\mathbf{Z}_{t0} = \mathbf{W}\mathbf{V}'\mathbf{Z}_{t1} + \mathbf{A}\mathbf{Z}_{t2} + \boldsymbol{\varepsilon}_t.$$

Johansen's method: 4 steps, focusing on $(\mathbf{V}$, $\mathbf{W})$ and r.

(i) Auxiliary regression

Run two regressions: $\mathbf{Z}_{t0} = \widehat{\mathbf{H}}_0 \mathbf{Z}_{t2} + \mathbf{R}_{t0}, \;\; \mathbf{Z}_{t1} = \widehat{\mathbf{H}}_1 \mathbf{Z}_{t2} + \mathbf{R}_{t1},$

$$\widehat{\mathbf{H}}_i = \mathbf{W}_{i2} \mathbf{W}_{22}^{-1}, \quad \mathbf{W}_{ij} = \frac{1}{n} \sum_{t=1}^{n} \mathbf{Z}_{ti} \mathbf{Z}_{tj}'.$$

Then it holds that

$$\mathbf{R}_{t0} = \mathbf{W}\mathbf{V}' \mathbf{R}_{t1} + \mathbf{e}_t,$$

which is a purely reduced-rank regression.

The Frisch-Waugh Theorem. For $y_t = b_1 x_{t1} + b_2 x_{t2} + \varepsilon_t$, the LSE $\widehat{b}_1$ for b_1 can be obtained in two steps:

1. Regress y_t on x_{t2}, leading to the residual $r_{t1} = y_t - \widehat{\beta}_1 x_{t2}$.
   Regress x_{t1} on x_{t2}, leading to the residual $r_{t2} = x_{t1} - \widehat{\beta}_2 x_{t2}$.
2. Obtain $\widehat{b}_1$ from $r_{t1} = b_1 r_{t2} + e_t$.

(ii) Profiling likelihood

In $\mathbf{R}_{t0} = \mathbf{W}\mathbf{V}'\mathbf{R}_{t1} + \mathbf{e}_t$, let $\mathbf{e}_t \sim_{iid} N(0, \mathbf{\Sigma}_e)$. The likelihood

$$L(\mathbf{V}, \mathbf{W}, \mathbf{\Sigma}_e) \propto |\mathbf{\Sigma}_e|^{-n/2} \exp\left\{ -\frac{n}{2}\text{tr}(\mathbf{\Sigma}_e^{-1}\mathbf{M})\right\}$$

$$\mathbf{M} \equiv \mathbf{M}(\mathbf{V}, \mathbf{W}) = \frac{1}{n}\sum_{t=1}^{n}(\mathbf{R}_{t0} - \mathbf{W}\mathbf{V}'\mathbf{R}_{t1})(\mathbf{R}_{t0} - \mathbf{W}\mathbf{V}'\mathbf{R}_{t1})'.$$

Thus

$$L(\mathbf{V}, \mathbf{W}) = \max_{\mathbf{\Sigma}_e \geq 0} L(\mathbf{V}, \mathbf{W}, \mathbf{\Sigma}_e) \propto |\mathbf{M}(\mathbf{V}, \mathbf{W})|^{-n/2},$$

$$\begin{aligned} L(\mathbf{V}) &\equiv \max_{\mathbf{W}} L(\mathbf{V}, \mathbf{W}) \equiv L(\mathbf{V}, \widetilde{\mathbf{W}}(\mathbf{V})) \\ &= |\mathbf{S}_{00} - \mathbf{S}_{01}\mathbf{V}(\mathbf{V}'\mathbf{S}_{11}\mathbf{V})^{-1}\mathbf{V}'\mathbf{S}_{10}|^{-n/2}, \end{aligned}$$

where $\widetilde{\mathbf{W}}(\mathbf{V}) = \mathbf{S}_{01}\mathbf{V}(\mathbf{V}'\mathbf{S}_{11}\mathbf{V})^{-1}$, $\mathbf{S}_{ij} = \frac{1}{n}\sum_{t=1}^{n}\mathbf{R}_{ti}\mathbf{R}_{tj}'$.

(iii) MLE for V and W

Since

$$|\mathbf{S}_{00} - \mathbf{S}_{01}\mathbf{V}(\mathbf{V}'\mathbf{S}_{11}\mathbf{V})^{-1}\mathbf{V}'\mathbf{S}_{10}| = |\mathbf{S}_{00}|\frac{|\mathbf{V}'(\mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01})\mathbf{V}|}{|\mathbf{V}'\mathbf{S}_{11}\mathbf{V}|},$$

maximize $L(\mathbf{V})$ is equivalent to

$$\min_{\mathbf{V}} \frac{|\mathbf{V}'(\mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01})\mathbf{V}|}{|\mathbf{V}'\mathbf{S}_{11}\mathbf{V}|} = \min_{\mathbf{C}} \frac{|\mathbf{C}'(\mathbf{I}_d - \mathbf{S}_{11}^{-1/2}\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\mathbf{S}_{11}^{-1/2})\mathbf{C}|}{|\mathbf{C}'\mathbf{C}|} = \prod_{j=1}^{r}(1 - \widehat{\lambda}_j),$$

where $\lambda_1 \geq \cdots \geq \lambda_r$ are the r largest eigenvalues of the matrix $\mathbf{S}_{11}^{-1/2}\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}\mathbf{S}_{11}^{-1/2}$.

Let $\widehat{\gamma}_1, \cdots, \widehat{\gamma}_r$ be the corresponding orthonormal eigenvectors. Then the MLEs are

$$\widehat{\mathbf{V}} = \mathbf{S}_{11}^{-1/2}(\widehat{\gamma}_1, \cdots, \widehat{\gamma}_r), \qquad \widehat{\mathbf{W}} = \mathbf{S}_{01}\widehat{\mathbf{V}}(\widehat{\mathbf{V}}'\mathbf{S}_{11}\widehat{\mathbf{V}})^{-1}.$$

(iv) Testing for the number of cointegration components r

Let $1 \geq \widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_d \geq 0$ be the eigenvalues of $\mathbf{S}_{11}^{-1}\mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{01}$. Johansen proposed two tests for the value of r based on those eigenvalues.

To test $H_0 : \mathrm{rank}(\mathbf{V}) \leq r \;\text{ vs }\; H_1 : \mathrm{rank}(\mathbf{V}) > r$, use trace statistic

$$\tau_1 = -n \sum_{j=r+1}^{d} \log(1 - \widehat{\lambda}_j).$$

H_0 is rejected for large values of τ_1.

Note. $\tau_1 \geq 0$, and $\tau_1 = 0$ under H_0, as then $\lambda_{r+1} = \cdots = \lambda_d = 0$.

Another approach: $H_0 : \mathrm{rank}(\mathbf{V}) = r \;\text{ vs }\; H_1 : \mathrm{rank}(\mathbf{V}) = r + 1.$

H_0 is rejected for large values of the test statistic

$$\tau_2 = -n \log(1 - \widehat{\lambda}_{r+1}).$$

Critical values for the two tests are tabulated in Hamilton (1994).

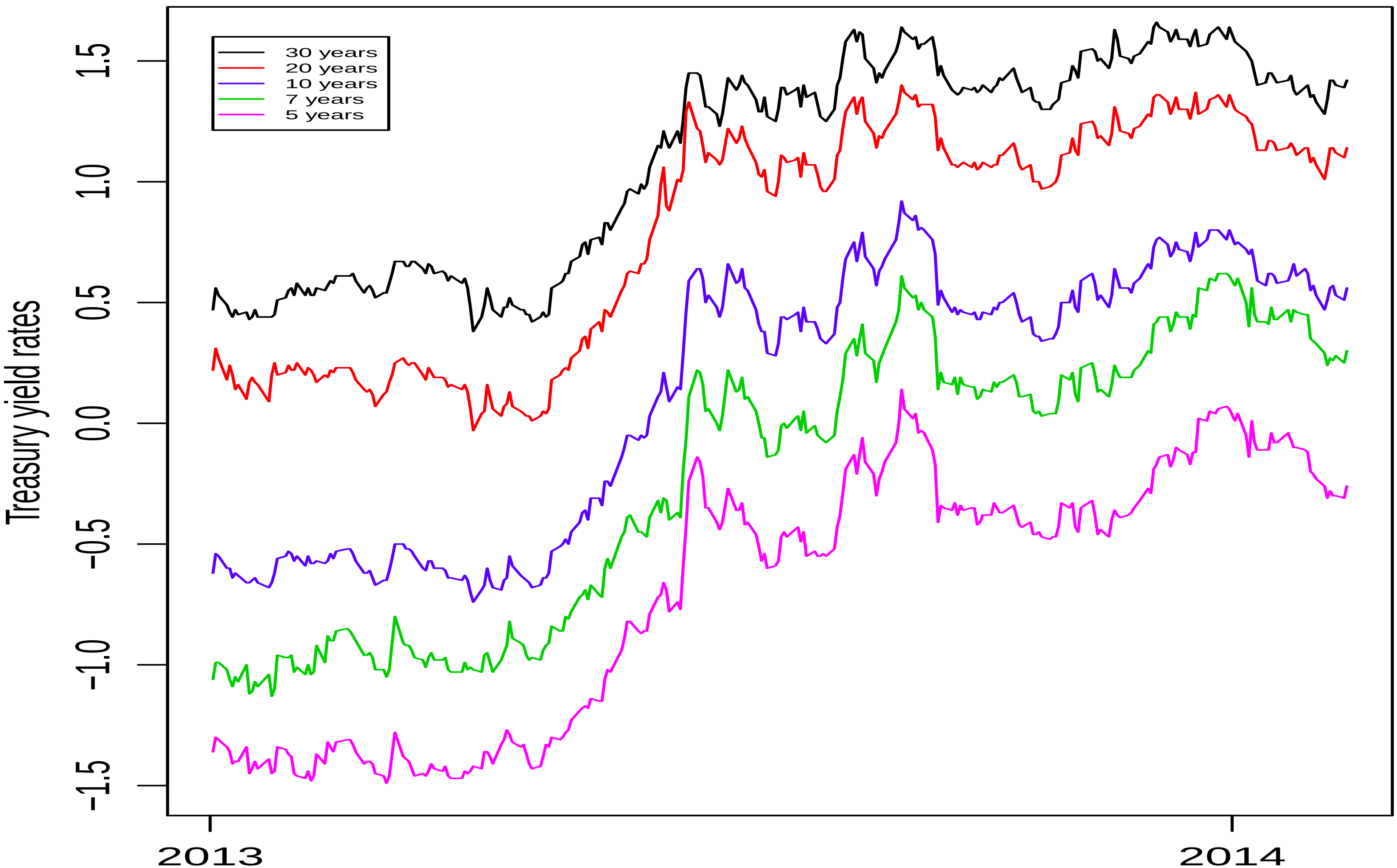
Implementation: R-package `urca`

Pfaff (2006). *Analysis of Integrated and Cointegrated Time Series with R*. Springer.

Illustration with a real data set: The daily U.S. Treasury real yield curve rates at fixed maturities of five, seven, ten, twenty and thirty years in the period of January 2, 2013 – February 11, 2014.

$n = 278,\ p = 5$

The data is stored in a 278×5 matrix `tbill` in R.


Treasury yield rates
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
2013
2014
30 years
20 years
10 years
7 years
5 years

To fit ECM, first check each component series is $I(1)$

Apply Augmented Dickey-Fuller (ADF) test using, for example, `ur.df` in the R-package `urca`

```
> t=ur.df(tbill[,1], type="none", lags=4,
        selectlags="AIC")
> summary(t)
```

This performs the ADF test to the first subseries

Option `type="none"`: choose no linear or constant trends, as differenced series shows no ssuch trends

Option `lags=4`: sets the upper bound for the AR order.

The tests for all five component series were not significant even at the 10% level; indicating no significant evidence against the hypothesis that there was at least one unit root in each of the five interest rates

Applying the same test to the differenced rates, the unit-root hypothesis is rejected at 1% significance level for all the five series.

Hence it is reasonable to assume that all the five yields are I(1) series.

To fit the data with the ECM and to apply the trace test,

```
> m1=ca.jo(tbill, type="trace", ecdet="none", K=2
             spec="transitory"); summary(m1)

Values of test statistic and critical values of t
          test 10pct  5pct  1pct
r <= 4 |  1.29  6.50  8.18 11.65
r <= 3 |  8.34 15.66 17.95 23.52
r <= 2 | 24.51 28.71 31.52 37.22
r <= 1 | 58.60 45.23 48.28 55.43
r = 0  | 95.33 66.49 70.60 78.87
```

H_0: rank$(\mathbf{V}) \geq r$ is rejected for $r = 1$ at the 1% significance level, and cannot be rejected for $r = 2$ even at the 10% level

Indicating two cointegration relations among the five yields series.

To apply the test with statistic τ_2:

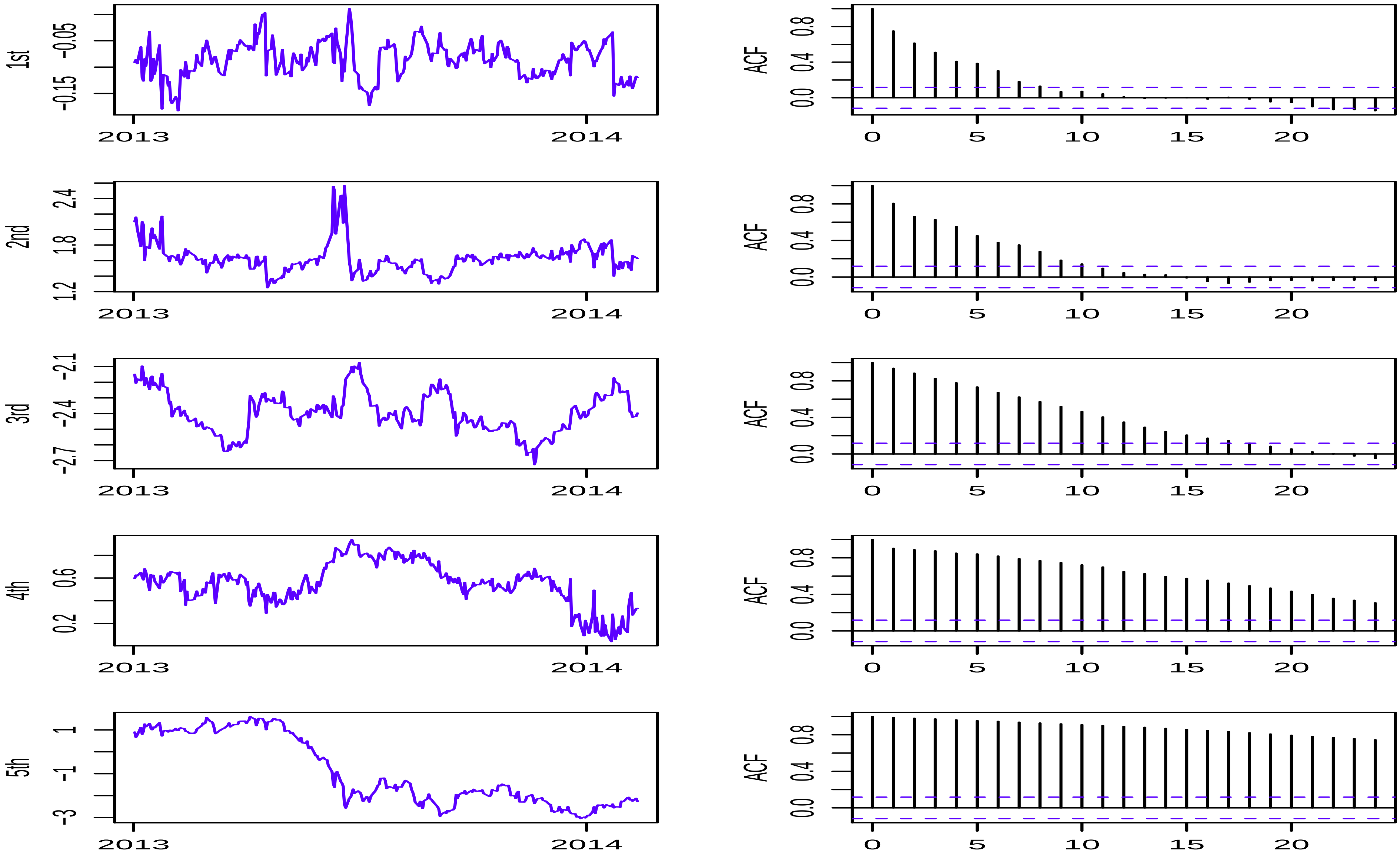
```
> m2=ca.jo(tbill, type="eigen", ecdet="none", K=2
      spec="transitory"); summary(m2)

Values of test statistic and critical values of t
          test 10pct  5pct  1pct
r <= 4 |  1.29  6.50  8.18 11.65
r <= 3 |  7.06 12.91 14.90 19.19
r <= 2 | 16.17 18.90 21.07 25.75
r <= 1 | 34.08 24.78 27.14 32.14
r = 0  | 36.73 30.84 33.32 38.78
```

H_0: $\text{rank}(\mathbf{V}) = r$ is rejected at the 1% significance level for $r = 1$, but cannot be rejected at the 10% level for $r = 2$.

Once again this test also indicates two cointegration directions.

To extract five candidate cointegrated variables $\widehat{\mathbf{V}}'\mathbf{X}_t$

```
> y = as.matrix(tbill)%*%as.matrix(m1@V)
```



The first two columns of `m1@V`, normalized respectively such that each is a unit vector (the sum of squared elements is one) , are

```
              ect1        ect2
5Year.l1    0.5808369   0.1115934
7Year.l1   -0.4399423   0.1382447
10Year.l1  -0.1716606  -0.7205056
20Year.l1  -0.3933011   0.6585049
30Year.l1   0.5337847  -0.1252131
```

The first cointegrated variable could be viewed tentatively as the sum of the contrast between the 5-year rate and the 7 year rate and the contrast between the 30 year rate and 20 year rate.

The second cointegrated variable is dominated by the contrast between the 20 year rate and the 10 year rate.

To refit the ECM with $r = 2$ fixed, run R-function `cajorls(m1, r=2)`, which returns the estimated coefficients in the fitted ECM:

	5Year.d	7Year.d	10Year.d	20Year.d	30Year
ect1	-0.0875	-0.0693	0.0952	0.1381	-0.072
ect2	0.1377	0.1261	-0.0056	-0.1166	0.073
constant	-0.0653	-0.0645	-0.0452	0.0251	-0.019
5Year.dl1	0.3726	0.5228	0.3455	0.2816	0.269
7Year.dl1	-0.5103	-0.7809	-0.1995	-0.1299	-0.175
10Year.dl1	0.4397	0.6145	-0.1263	-0.3357	-0.101
20Year.dl1	0.0153	-0.0309	0.1355	0.0828	0.129
30Year.dl1	-0.1897	-0.2339	-0.0130	0.2431	-0.025

In the above R-outputs, d indicates difference, `l1` indicates lag 1, and `dl1` indicates a differenced variable at lag 1.

For example, the fitted ECM model for the differenced 5 year rate is

$$\begin{aligned}\nabla X_{t1} &= -0.0875U_{t-1,1} + 0.1377U_{t-1,2} - 0.0653 \\ &\quad + 0.3726\nabla X_{t-1,1} - 0.5103\nabla X_{t-1,2} + 0.4397\nabla X_{t-1,3} \\ &\quad + 0.0153\nabla X_{t-1,4} - 0.1897\nabla X_{t-1,5},\end{aligned}$$

where $X_{t1}, \cdots, X_{t5}$ denote, respectively, the 5 year, 7 year, 10 year, 20 year and 30 year rates at time t, U_{t1} and U_{t2} are the two normalized cointegrated variables.

The standard errors of the estimated coefficients can be calculated using the function `abStdErr` available at

`http://orfe.princeton.edu/~jqfan/fan/FinEcon.html`

A new method by eigenanalysis

Setting: $\mathbf{y}_t$ is a $p \times 1$ time series, each component of $\mathbf{y}_t$ is weak $I(d)$ with $d \geq 0$.

Note. d may be different for different components.

Assumption: $\mathbf{y}_t = \mathbf{A}\mathbf{x}_t$, where $\mathbf{A}$ is $p \times p$ and unknown, and

$$\mathbf{x}_t = (\mathbf{x}_{t1}', \mathbf{x}_{t2}')',$$

where $\mathbf{x}_{t2}$ is $r \times 1$ and weakly stationary, all components of $\mathbf{x}_{t1}$ are $I(d)$ with $d \geq 1$, d may be different for different components.

Goal: to identify the cointegration rank r and the cointegrated process $\mathbf{x}_{t2}$.

Weak $I(d)$: $\mathbf{u}_t$ is called a weak $I(d)$ process if $\nabla^d \mathbf{u}_t$ is weakly stationary but $\nabla^{d-1} \mathbf{u}_t$ is not.

$\mathbf{A}$ and $\mathbf{x}_t$ in $\mathbf{y}_t = \mathbf{A}\mathbf{x}_t$ are not uniquely, as $(\mathbf{A}, \mathbf{x}_t)$ can be replaced by $(\mathbf{A}\mathbf{H}^{-1}, \mathbf{H}\mathbf{x}_t)$ for any invertible $\mathbf{H}$ of the form

$$\begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{0} & \mathbf{H}_{22} \end{pmatrix}$$

Let $\mathbf{A}$ be orthogonal, as any non-orthogonal $\mathbf{A}$ admits

$$\mathbf{A} = \mathbf{Q}\mathbf{U},$$

where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{U}$ is a upper-triangular matrix, and we may then replace $(\mathbf{A}, \mathbf{x}_t)$ by $(\mathbf{Q}, \mathbf{U}\mathbf{x}_t)$.

We always assume that $\mathbf{A}$ is orthogonal, then $\mathbf{x}_t = \mathbf{A}'\mathbf{y}_t$. Let

$$\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2).$$

Then $\mathbf{x}_{t2} = \mathbf{A}_2'\mathbf{y}_t$. $\mathcal{M}(\mathbf{A}_2)$ is called the cointegration space, which is uniquely defined.

Task: to determine r and to estimate $\mathcal{M}(\mathbf{A}_2)$

Basic idea: Let $\widehat{\mathbf{\Sigma}}_j = \frac{1}{n}\sum_{t=1}^{n-j}(\mathbf{y}_{t+j} - \bar{\mathbf{y}})(\mathbf{y}_t - \bar{\mathbf{y}})', \ \ \bar{\mathbf{y}} = \frac{1}{n}\sum_{t=1}^{n}\mathbf{y}_t.$

- For any $\mathbf{a} \in \mathcal{M}(\mathbf{A}_2), \ \mathbf{a}'\widehat{\mathbf{\Sigma}}_j\mathbf{a} \xrightarrow{P} \mathrm{Cov}(\mathbf{a}'\mathbf{y}_{t+j}, \mathbf{a}'\mathbf{y}_t)$ finite.
- For any $\mathbf{a} \notin \mathcal{M}(\mathbf{A}_2), \ \mathbf{a}'\mathbf{y}_t \sim I(d)$ for some $d \geq 1$, and

$$\mathbf{a}'\widehat{\mathbf{\Sigma}}_j\mathbf{a} = \begin{cases} O_e(n^{2d-1}) & E(\mathbf{a}'\mathbf{y}_t) = 0, \\ O_e(n^{2d}) & E(\mathbf{a}'\mathbf{y}_t) \neq 0, \end{cases}$$

where $U = O_e(V)$ indicates that $P(C_1 \leq |U/V| < C_2) \to 1$ for some constants $0 < C_1 < C_2 < \infty$.

Peãn & Poncela (2006).

The r directions in $\mathcal{M}(\mathbf{A}_2)$ makes $|\mathbf{a}'\widehat{\mathbf{\Sigma}}_j\mathbf{a}|$ as small as possible!

To combine the information over different lags without cancellation, let $\widehat{\mathbf{W}} = \sum_{j=0}^{j_0} \widehat{\mathbf{\Sigma}}_j \widehat{\mathbf{\Sigma}}_j'$.

Let $\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_p \geq 0$ be the eigenvalues of $\widehat{\mathbf{W}}$, let $\widehat{\gamma}_1, \cdots, \widehat{\gamma}_p$ be the corresponding eigenvectors. Define

$$\widehat{\mathbf{A}} = (\widehat{\mathbf{A}}_1, \widehat{\mathbf{A}}_2) = (\widehat{\gamma}_1, \cdots, \widehat{\gamma}_p), \qquad \widehat{\mathbf{x}}_{t2} = \widehat{\mathbf{A}}_2' \mathbf{y}_t.$$

$$\widehat{\mathbf{A}}'\widehat{\mathbf{W}}\widehat{\mathbf{A}} = \begin{pmatrix} \widehat{\mathbf{A}}_1'\widehat{\mathbf{W}}\widehat{\mathbf{A}}_1 & 0 \\ 0 & \widehat{\mathbf{A}}_2'\widehat{\mathbf{W}}\widehat{\mathbf{A}}_2 \end{pmatrix} = \sum_{j=0}^{j_0} (\widehat{\mathbf{A}}'\widehat{\mathbf{\Sigma}}_j\widehat{\mathbf{A}})(\widehat{\mathbf{A}}'\widehat{\mathbf{\Sigma}}_j'\widehat{\mathbf{A}})$$

$$= \sum_{j=0}^{j_0} \begin{pmatrix} \widehat{\mathbf{A}}_1'\widehat{\mathbf{\Sigma}}_j\widehat{\mathbf{A}}_1\widehat{\mathbf{A}}_1'\widehat{\mathbf{\Sigma}}_j'\widehat{\mathbf{A}}_1 + \widehat{\mathbf{A}}_1'\widehat{\mathbf{\Sigma}}_j\widehat{\mathbf{A}}_2\widehat{\mathbf{A}}_2'\widehat{\mathbf{\Sigma}}_j'\widehat{\mathbf{A}}_1 & 0 \\ 0 & \widehat{\mathbf{A}}_2'\widehat{\mathbf{\Sigma}}_j\widehat{\mathbf{A}}_2\widehat{\mathbf{A}}_2'\widehat{\mathbf{\Sigma}}_j'\widehat{\mathbf{A}}_2 + \widehat{\mathbf{A}}_2'\widehat{\mathbf{\Sigma}}_j\widehat{\mathbf{A}}_1\widehat{\mathbf{A}}_1'\widehat{\mathbf{\Sigma}}_j'\widehat{\mathbf{A}}_2 \end{pmatrix}.$$

Determining cointegration rank r

As $\widehat{\lambda}_i$ is at least $O_e(n^2)$ for all $1 \le i \le p-r$, and $\widehat{\lambda}_i = O_e(1)$ for all $p-r < i \le p$.

Hence as long as $1 \le r < p$, $\widehat{\lambda}_i/(n\widehat{\lambda}_p) \to \infty$ in probability for all $1 \le i \le p-r$, and $\widehat{\lambda}_i/(n\widehat{\lambda}_p) = o_p(1)$ for all $p-r < i \le p$.

Hence, we estimate r by

$$\widehat{r} = \max\{j : \widehat{\lambda}_{p+1-j}/(n\widehat{\lambda}_p) \le 1, \ 1 \le j \le p\}.$$

Alternatively, define $IC(l) = \sum_{j=1}^{l} \widehat{\lambda}_{p+1-j} + (p-l)\omega_n,$

where $\omega_n \to \infty$, $\omega_n/n^{4d-2} \to 0$ and d is the smallest integrated order. Let

$$\widetilde{r} = \arg\min_{1 \le l \le p} IC(l).$$

Note when $\omega_n = n\widehat{\lambda}_p$, $\quad \widetilde{r} = \widehat{r}.$

R-code

```
coin <- function(Y1,J0) {
# Y1: a n x p matrix
# J0:  No. of lags used in define \hat{W}

# Output
# X: nxp transformed p time series
# EigenV: p eigenvalues of \hat{W}
# kRatio: ratio estimate for cointegration rank
# kIC1: IC estimate for  cointegration rank with omega_n = n^{5/4} * \lambda
# kIC2: IC estimate for  cointegration rank with omega_n = n^{3/2} * \lambda

n=nrow(Y1); p=ncol(Y1)
Y=t(Y1)

# Part I -- Apply the transformation to recover X_t
S0=cov(t(Y)); V0=S0%*%S0
for(j in 1:J0) {
Y0=Y[,(j+1):n]; Y1=Y[,1:(n-j)]
S=cov(t(Y0), t(Y1)); V0=V0+S%*%t(S)
}
V0=V0/J0
```

```
t=eigen(V0, symmetric=T)
G=as.matrix(t$vectors)
X=t(G)%*%Y

# Part II estimate cointegration rank
ev=t$values
k1=p+1-min(which((ev[1:(p-1)]/(n*ev[p]))<=1, arr.ind=TRUE))

om1=(n^(5/4))*ev[p]; om2=(n^(3/2))*ev[p]
sev=vector(length=p); sev=ev[p];
for(i in 2:p) sev[i]=ev[p+1-i]+sev[i-1]
P=1:p
z=sev+(p-P)*om1
k2=which(z==min(z), arr.ind=TRUE)
z=sev+(p-P)*om2
k3=which(z==min(z), arr.ind=TRUE)

list(X=t(X), EigenV=ev, EigenVec=G, kRatio=k1, kIC1=k2, kIC2=k3)
}
```

Asymptotics I: p fixed

(i) $\mathcal{M}(\widehat{\mathbf{A}}_2)$ is a consistent estimator for $\mathcal{M}(\mathbf{A}_2)$ with known r

(ii) $\widehat{r}$ and $\widetilde{r}$ are consistent

Let

$$D(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2)) = \sqrt{1 - \frac{1}{r}\mathrm{tr}(\widehat{\mathbf{A}}_2\widehat{\mathbf{A}}_2'\mathbf{A}_2\mathbf{A}_2')}$$

Then $D(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2)) \in [0, 1]$. It is 0 iff $\mathcal{M}(\widehat{\mathbf{A}}_2) = \mathcal{M}(\mathbf{A}_2)$, and 1 iff $\mathcal{M}(\widehat{\mathbf{A}}_2) \perp \mathcal{M}(\mathbf{A}_2)$.

Put $\mathbf{x}_{t1} = \mathbf{A}_1'\mathbf{y}_t = (x_t^1, \cdots, x_t^{p-r})'$, and $x_t^j \sim I(d_j)$.

Let $z_t^j = \nabla^{d_j} x_t^j$, $\mathbf{z}_t = (z_t^1, \cdots, z_t^{p-r})'$, $\boldsymbol{\varepsilon}_t = (\mathbf{z}_t', \mathbf{x}_{t2}')'$.

$\boldsymbol{\varepsilon}_t$ is $p \times 1$ and weakly stationary.

For $\mathbf{t} = (t_1, \cdots, t_p)'$ with $0 < t_1 < \cdots < t_p \le 1$, let

$$\mathbf{S}_n(\mathbf{t}) \equiv (S_n^1(t_1), \cdots, S_n^p(t_p))' = \Big(\frac{1}{\sqrt{n}}\sum_{l=1}^{[nt_1]}(\varepsilon_l^1 - \mathrm{E}\varepsilon_1^1), \cdots, \frac{1}{\sqrt{n}}\sum_{l=1}^{[nt_p]}(\varepsilon_l^p - \mathrm{E}\varepsilon_1^p)\Big)'.$$

Condition 1.

(i) There exist a Gaussian process $\mathbf{W}(\mathbf{t}) = (W^1(t_1), \cdots, W^p(t_p))'$, with Var$(\mathbf{W}(\mathbf{1})) > 0$, for which $\mathbf{S}_n(\mathbf{t}) \overset{J_1}{\Longrightarrow} \mathbf{W}(\mathbf{t})$ on $D[0,1]$.

(ii) The sample covariance of $\mathbf{x}_{t2}$ satisfies

$$\sup_{0\le j\le j_0} \Big\|\frac{1}{n}\sum_{t=1}^{n-j}(\mathbf{x}_{t+j,2} - \bar{\mathbf{x}}^2)(\mathbf{x}_{t2} - \bar{\mathbf{x}}^2)' - \mathrm{Cov}(\mathbf{x}_{1+j,2}, \mathbf{x}_{1,2})\Big\|_2 \xrightarrow{p} 0,$$

where $\|\mathbf{H}\|_2 = \sup_{\|\mathbf{a}\|=1} \|\mathbf{H}\mathbf{a}\|$ is the L_2-norm of matrix $\mathbf{H}$.

Condition 1 holds if

- either $\{\boldsymbol{\varepsilon}_t\}$ α-mixing with $\sum_{m=1}^{\infty} \alpha_m^{1-1/\gamma} < \infty$, det$(\text{Var}(\boldsymbol{\varepsilon}_t)) \neq 0$, and $\mathrm{E}\|\boldsymbol{\varepsilon}_t\|^{2\gamma} < C$ for some constants $\gamma > 1$ and $C < \infty$,
- or $\boldsymbol{\varepsilon}_t = \sum_{j=0}^{\infty} \mathbf{C}_j \boldsymbol{\eta}_{t-j}$, where $\boldsymbol{\eta}_t$ are i.i.d. with a non-singular covariance matrix and $E\|\boldsymbol{\eta}_t\|^{4\gamma} < \infty$ for some constant $\gamma > 1$, and det$(\sum_{j=0}^{\infty} \mathbf{C}_j) \neq 0$, $\sum_{j=1}^{\infty} \sum_{i,m=1}^{p} |c_{j,im}| < \infty$, and $c_{j,im}$ denotes the (i, m)-th element of $\mathbf{C}_j$.

Theorem 1. Let r be known. Under Condition 1,

(i) $D(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2)) = O_p(n^{-2d+1})$ provided $|I_0| \geq 2$ or $|I_0| = 1$ and $\mathrm{E}z_t^{I_0} = 0$, and

(ii) $D(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2)) = O_p(n^{-2d})$ provided $|I_0| = 1$ and $\mathrm{E}z_t^{I_0} \neq 0$,

where $d = \min_{1 \leq i \leq p-r} d_i, \; I_0 = \{i : X_i \sim I(d), \, 1 \leq i \leq p - r\}$ and $|I_0|$ denotes the number of elements in I_0.

Theorem 2. Let Condition 1 hold.

(i) $\lim_{n\to\infty} P(\,\widehat{r} = r\,) = 1$ provided $1 \leq r < p$.

(ii) $\lim_{n\to\infty} P(\,\widetilde{r} = r\,) = 1$ provided $\lim_{n\to\infty}(\frac{j_0^2}{\omega_n} + \frac{\omega_n}{n^{4d-2}}) = 0$.

Asymptotics II: $p = o(n^c),\ c \in (0,\ 0.5)$

Condition 2.

(i) Suppose that the components of $\mathbf{z}_t$ are independent and $\mathbf{z}_t = 0$. For any $1 \le i \le p - r$, there exists an i.i.d normal sequence $\{\nu_t^i\}$, with Var$(\nu_t^i) = 1$, for which

$$\sup_{0\le t\le 1}\left|\sum_{s=1}^{[nt]}[z_s^i - \sigma_{ii}\nu_s^i]\right| = O_p(n^\tau),$$

where $\tau < 1/2$ is a constant, $b_1 \le \sigma_{ii}^2 \equiv \lim_{n\to\infty} \mathrm{Var}\left(\sum_{s=1}^n z_s^i\right)/n \le b_2$ for all $1 \le i \le p - r$, and $b_1,\ b_2$ are two positive constants.

(ii) The sample covariance of $\mathbf{x}_{t2}$ satisfies

$$\sup_{0\le j\le j_0}\left\|\frac{1}{n}\sum_{t=1}^{n-j}(\mathbf{x}_{t+j,2} - \bar{\mathbf{x}}^2)(\mathbf{x}_{t2} - \bar{\mathbf{x}}^2)' - \text{Cov}(\mathbf{x}_{1+j,2}, \mathbf{x}_{1,2})\right\|_2 \xrightarrow{p} 0.$$

(iii) For any $1 \le i \le p - r$ and $p - r + 1 \le j \le p$,

$$(S_n^i(t_1), S_n^j(t_2)) \overset{J_1}{\Longrightarrow} (W^i(t_1), W^j(t_2)),\ \text{ on } D[0,1]\times D[0,1],$$

$$\int_0^1 S_n^i(t)\, dS_n^j(t) \xrightarrow{d} \int_0^1 W^i(t)\, dW^j(t).$$

Theorem 3. Let k be known, $p = o(n^{1/2-\tau})$. Under Condition 2,

$$D(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{A}_2)) = O_p(p^{1/2} n^{-2d+1} (\lambda^*)^{-1}),$$

where λ^* is the smallest eigenvalue of a $(p-r) \times (p-r)$ positive definite matrix *depending on the underlying process intimately*.

Note. No explicit constraints on r and $p-r$ in Theorem 3. But when $p-r$ is fixed (i.e. $r/p \to 1$), λ^* is a finite and positive constant.

Theorem 4. Let Condition 2 hold and $p = o(n^{1/2-\tau})$. If

$$\lim_{n\to\infty} P\{\log n < \omega_n < (\lambda^* n^{2d-1})^2 / \log n\} = 1,$$

then $\lim_{n\to\infty} P(\,\widetilde{k} = k\,) = 1$.

Simulation I

In model $\mathbf{y}_t = \mathbf{A}\mathbf{x}_t = \mathbf{A}_1\mathbf{x}_{t1} + \mathbf{A}_2\mathbf{x}_{t2}$,

- all components of $\mathbf{x}_{t2}$ are AR(1) with coefficients generated from $U(-0.8, 0.8)$,
- all components of $\mathbf{x}_{t1}$ are ARIMA(1,2,1) with AR-coefficients $\sim U(0.3, 0.8)$ and MA-coefficients $\sim U(0, 0.95)$,
- all innovations are independent and $N(0, 1)$,
- all elements of $\mathbf{A}$ are generated from $U(-3, 3)$.

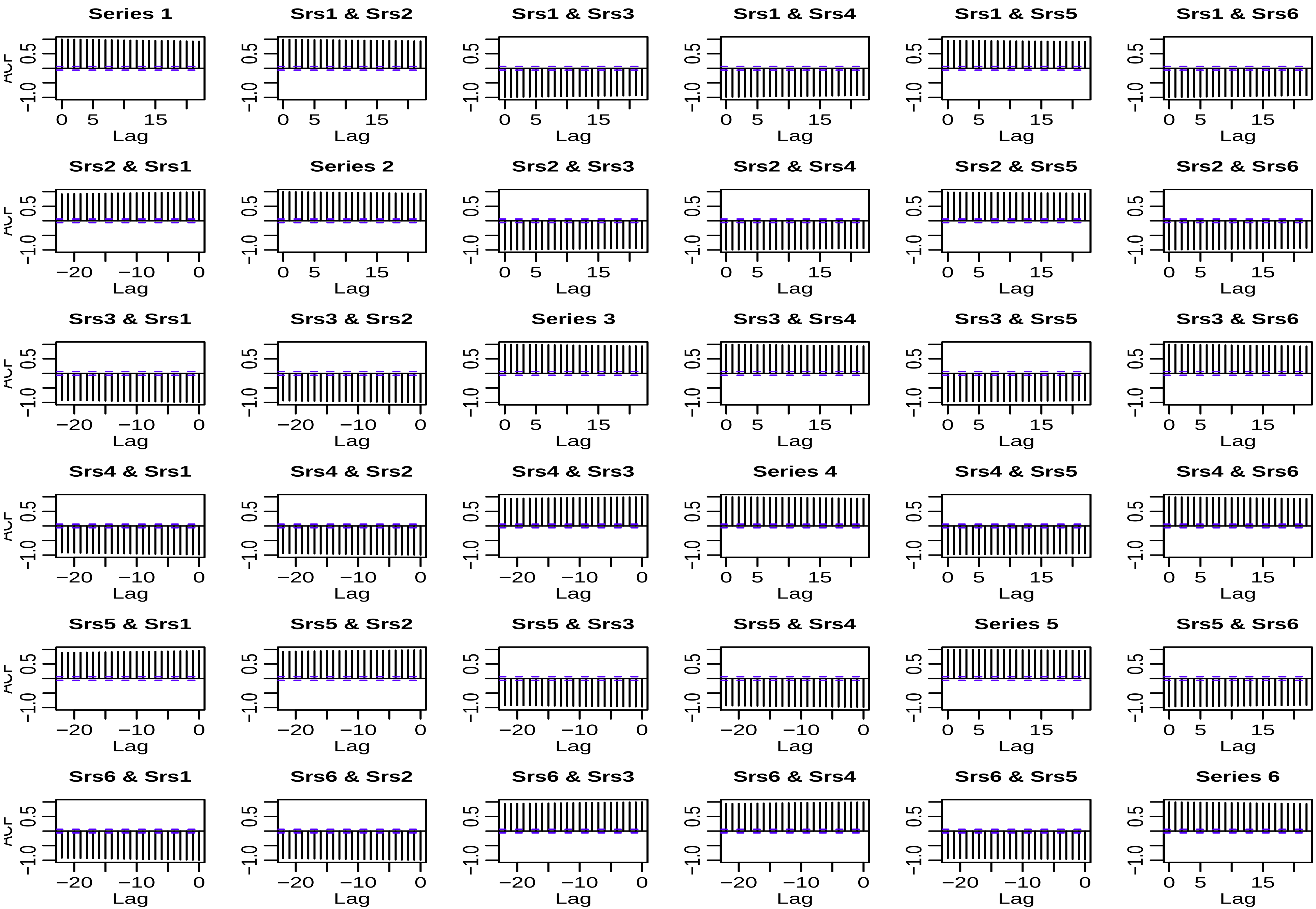
Note. $\mathbf{A}$ is not orthogonal.

True cointegration space: $\mathcal{M}(\mathbf{B}_2)$, where $(\mathbf{B}_1, \mathbf{B}_2) = (\mathbf{A}^{-1})'$.

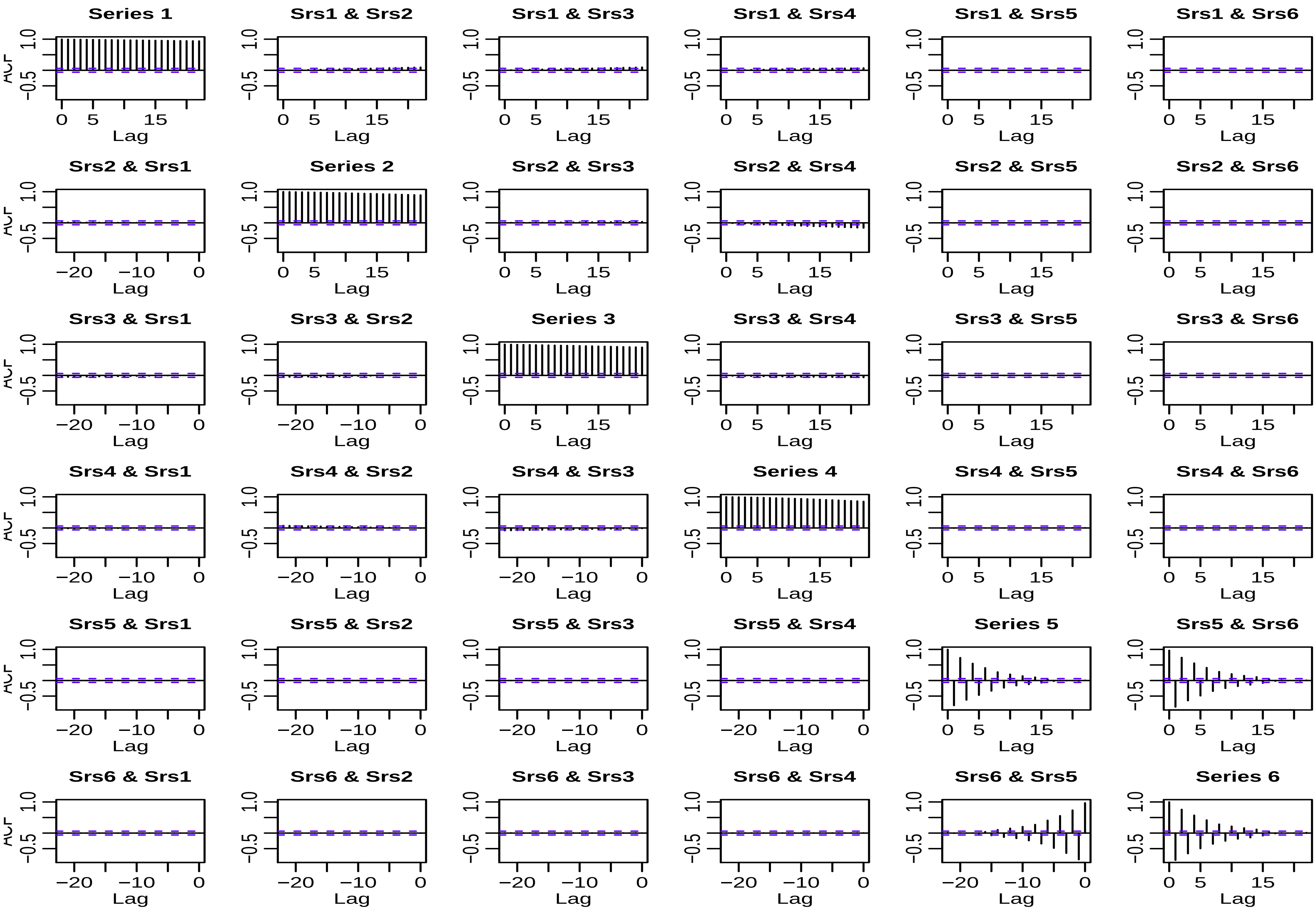
Estimation error:

$$D_1(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\mathbf{B}_2)) = \Big\{1 - \frac{1}{\max(r, \widehat{r})} \text{tr}\big(\widehat{\mathbf{A}}_2\widehat{\mathbf{A}}_2'\mathbf{B}_2(\mathbf{B}_2'\mathbf{B}_2)^{-1}\mathbf{B}_2'\big)\Big\}^{1/2}$$

Sample ACF/CCF of $\mathbf{y}_t$, $n = 1000,\ p = 6,\ r = 2$



Sample ACF/CCF of $\widehat{\mathbf{x}}_t$, $n = 1000,\ p = 6,\ r = 2$



Simulation replications: 1000 times

For IC-estimation for r, use

$$\omega_n^1 = n^{5/4}\widehat{\lambda}_p \qquad \text{or} \qquad \omega_n^2 = n^{3/2}\widehat{\lambda}_p$$

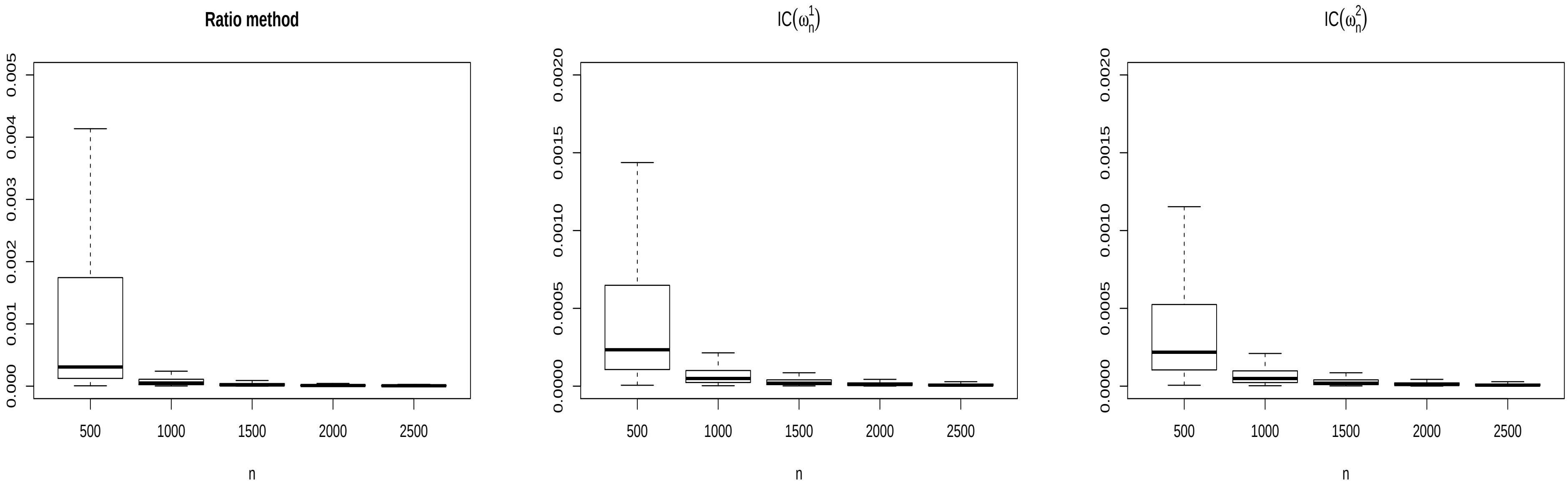
where $\widehat{\lambda}_p$ is the minimum eigenvalue of $\widehat{\mathbf{W}}$ (with $k_0 = 5$).

Note that when $\omega_n = n\widehat{\lambda}_p$, $\widehat{r} = \widetilde{r}$.

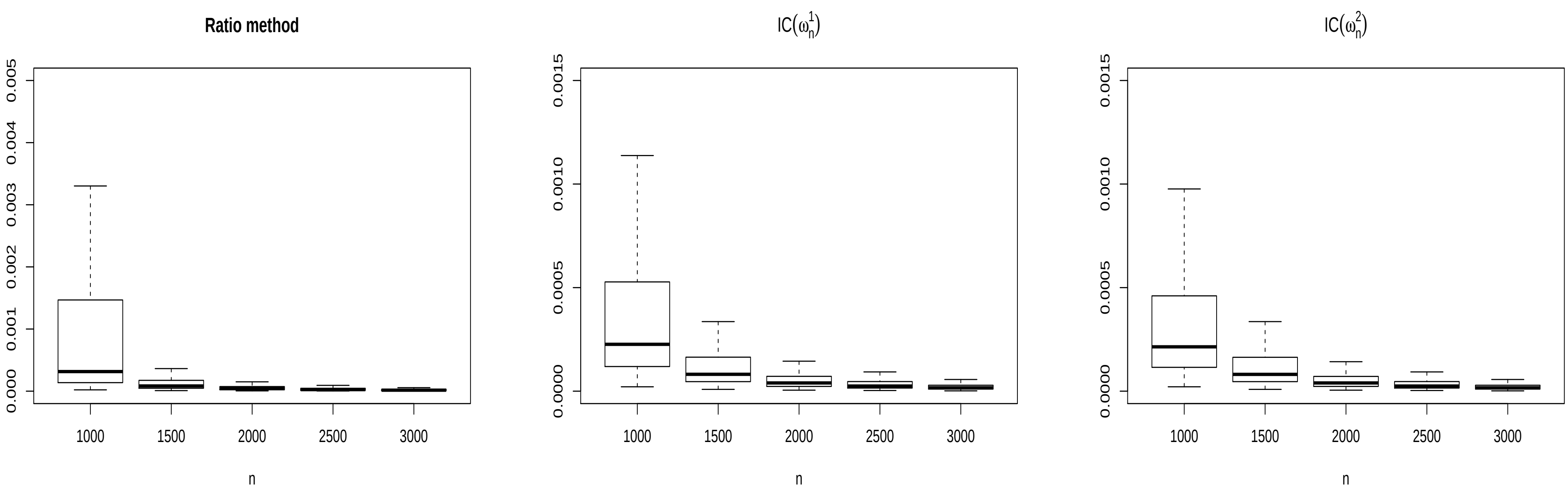
Relative frequencies for $\widehat{r} = r$ and $\widetilde{r} = r$

	n	300	500	1000	1500	2000	2500
p=6, r=4	Ratio	0.042	0.053	0.511	0.859	0.940	0.979
	IC(ω_n^1)	0.119	0.255	0.833	0.967	0.996	0.998
	IC(ω_n^2)	0.315	0.570	0.964	0.996	0.998	1.000
p=10, r=4	Ratio	0.019	0.154	0.837	0.967	0.988	0.993
	IC(ω_n^1)	0.155	0.501	0.956	0.994	0.998	0.998
	IC(ω_n^2)	0.412	0.796	0.994	0.999	0.998	0.996
p=20, r=6	Ratio	0.009	0.177	0.929	0.977	0.971	0.979
	IC(ω_n^1)	0.075	0.565	0.948	0.940	0.896	0.882
	IC(ω_n^2)	0.330	0.791	0.798	0.691	0.616	0.558
p=20, r=10	Ratio	0.000	0.005	0.479	0.873	0.946	0.951
	IC(ω_n^1)	0.000	0.050	0.857	0.974	0.991	0.993
	IC(ω_n^2)	0.003	0.410	0.972	0.996	0.995	0.999
p=20, r=14	Ratio	0.000	0.000	0.026	0.356	0.753	0.874
	IC(ω_n^1)	0.000	0.000	0.254	0.791	0.949	0.983
	IC(ω_n^2)	0.000	0.015	0.717	0.958	0.993	0.996

Boxplots of $D_1(\widehat{\mathcal{M}}(\mathbf{A}_2), \mathcal{M}(\mathbf{B}_2))$

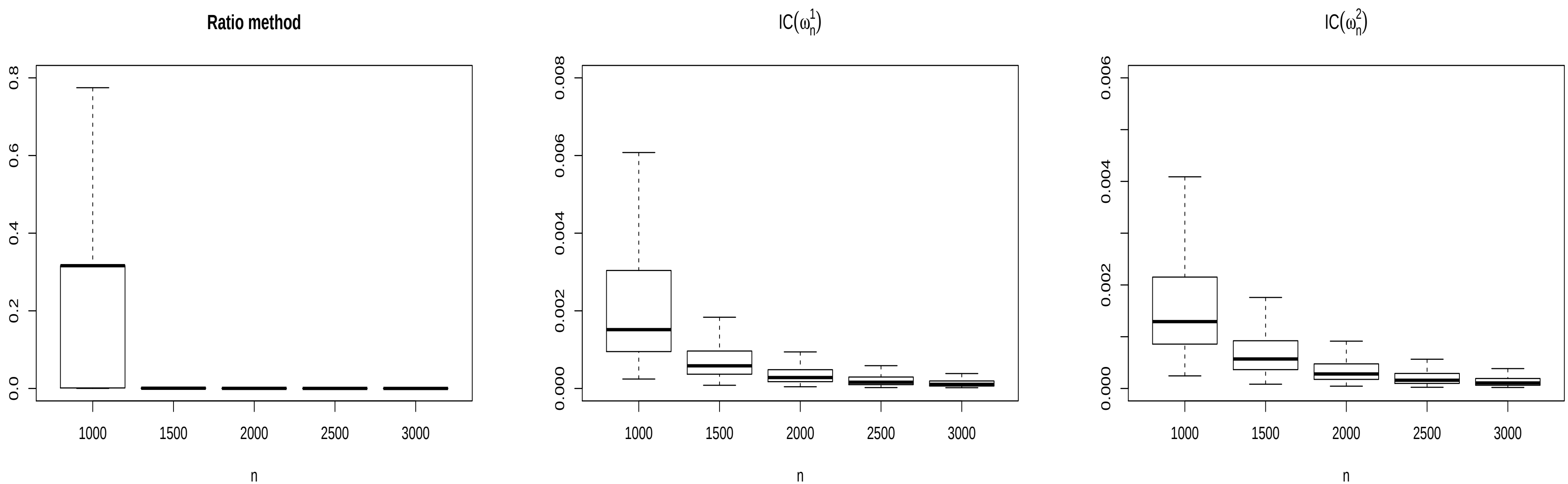


$$p = 6,\ r = 2,\ 500 \le n \le 2500$$



$$p = 10,\ r = 4,\ 1000 \le n \le 3000$$

Boxplots of $D_1(\widehat{\mathcal{M}}(\mathbf{A}_2), \mathcal{M}(\mathbf{B}_2))$



$$p = 20,\ r = 10,\ 1000 \le n \le 3000$$

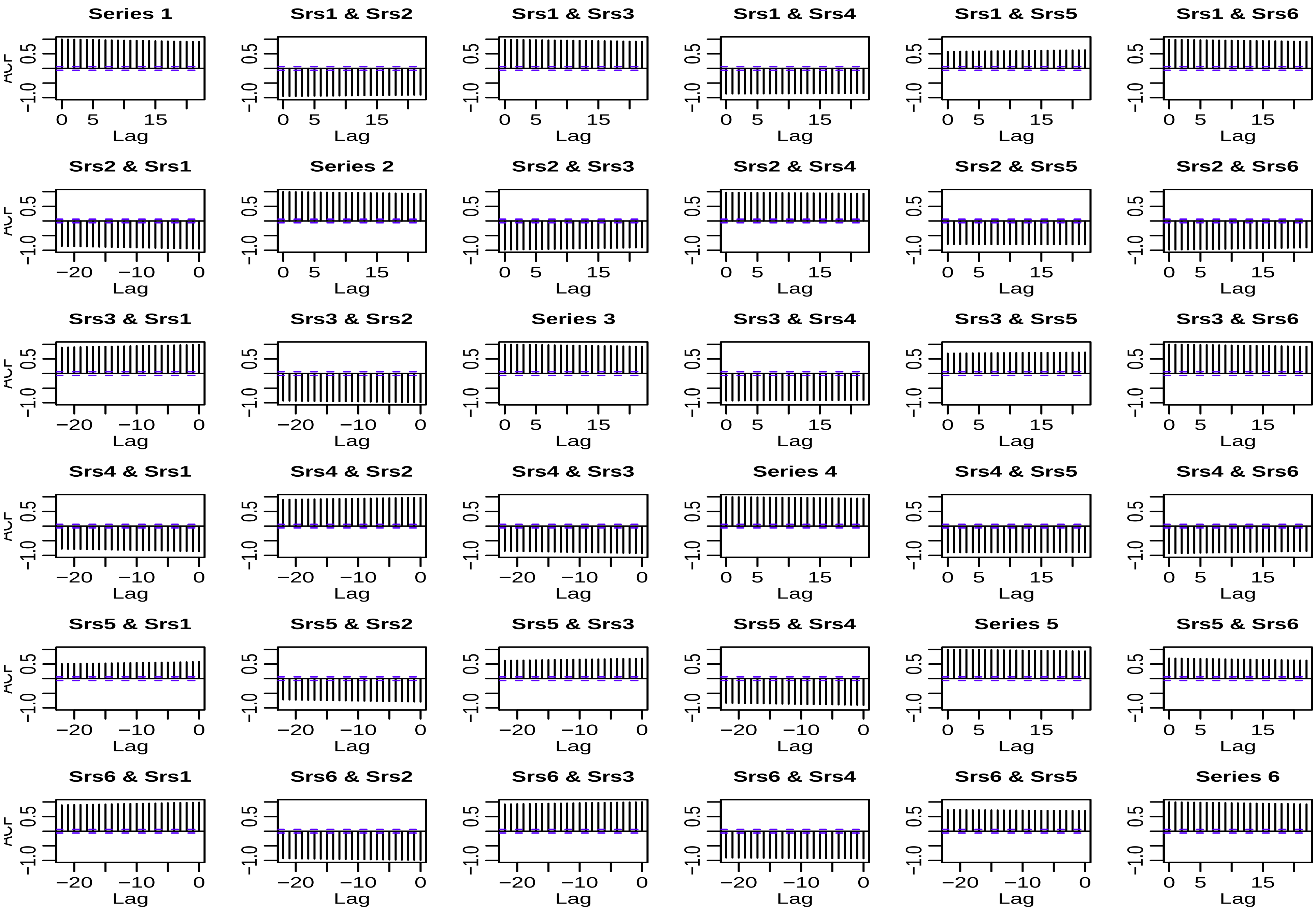
Simulation II

In model $\mathbf{y}_t = \mathbf{A}\mathbf{x}_t = \mathbf{A}_1\mathbf{x}_{t1} + \mathbf{A}_2\mathbf{x}_{t2}$,

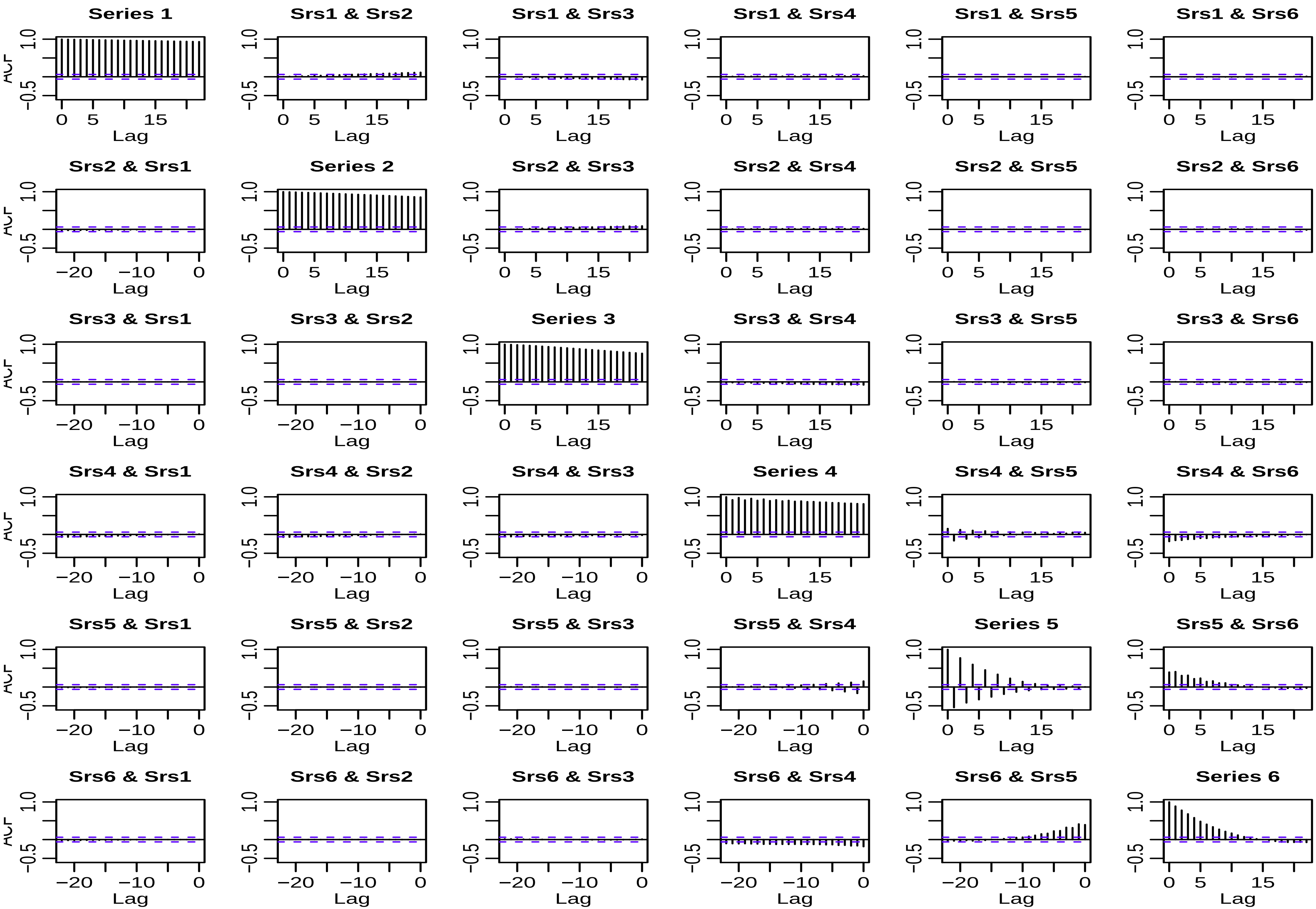
- all components of $\mathbf{x}_{t2}$ are AR(1) with coefficients $-0.8 + 1.6i/r$, $i = 1, \cdots, r$,
- s components of $\mathbf{x}_{t1}$ are ARIMA(1,1,1) with coefficients $(0.3 + 0.5i/s,\ 0.2 + 0.6i/s)$, $i = 1, \cdots, s$,
- $(p - r - s)$ components of of $\mathbf{x}_{t1}$ are ARIMA(0, 2,1) with coefficients generated from $U(-0.95,\ 0.95)$,
- all innovations are independent and $N(0, 1)$,
- all elements of $\mathbf{A}$ are generated from $U(-3, 3)$.

In IC-estimation for r, use $\omega_n^3 = n^{2/3}\widehat{\lambda}_p$ or $\omega_n^4 = n^{5/4}\widehat{\lambda}_p$.

Sample ACF/CCF of $\mathbf{y}_t$, $n = 1000,\ p = 6,\ r = 2,\ s = 2$



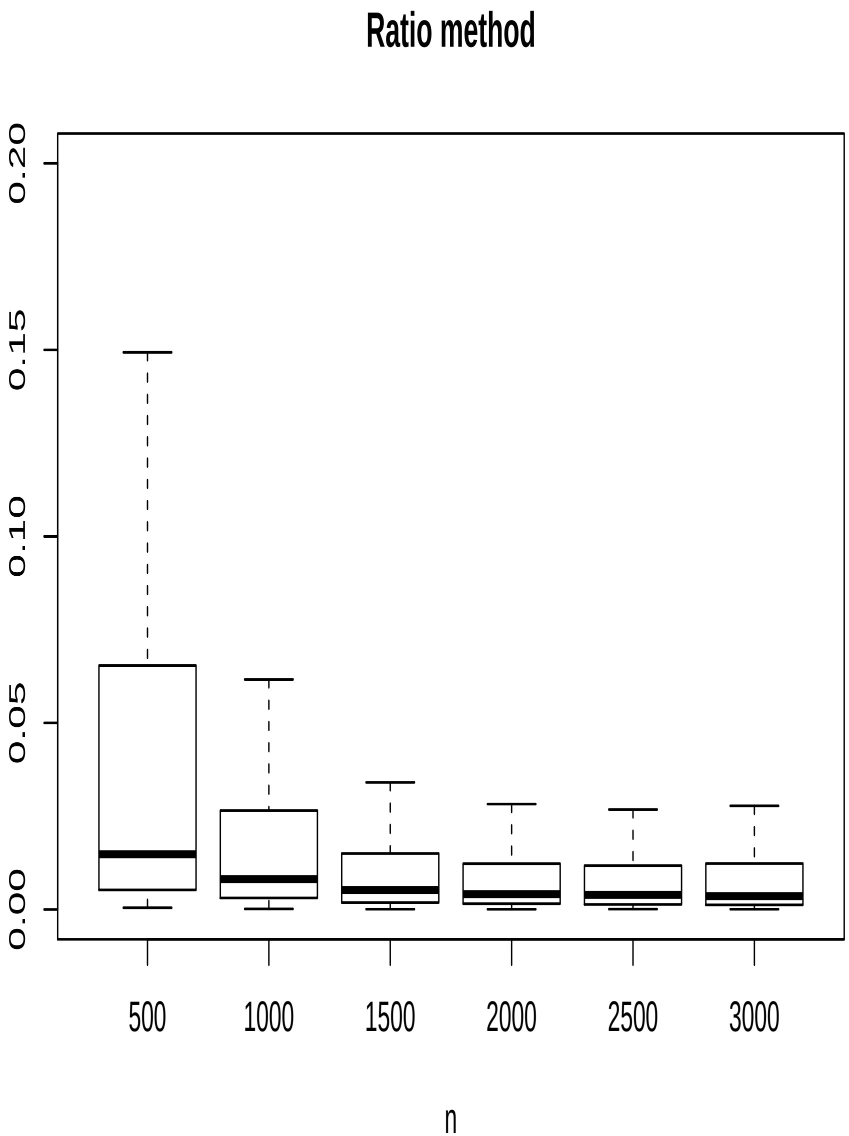
Sample ACF/CCF of $\widehat{\mathbf{x}}_t$, $n = 1000,\ p = 6,\ r = 2,\ s = 2$

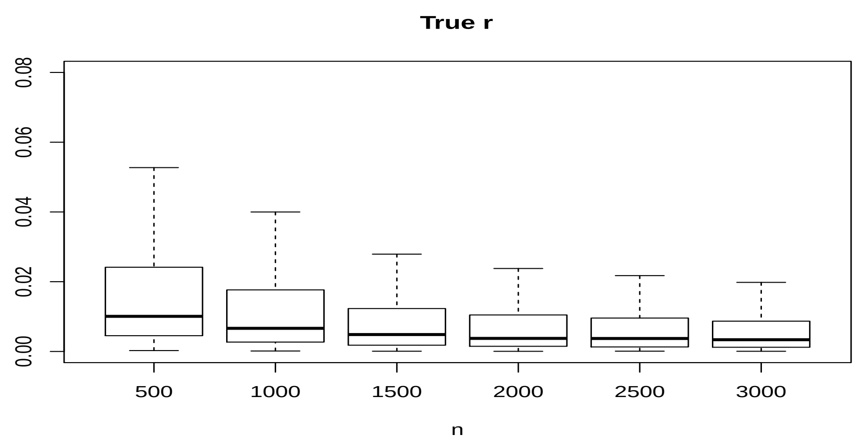


Relative frequencies for $\widehat{r} = r$ and $\widetilde{r} = r$

(p, k, s)	n	300	500	1000	1500	2000	2500
(6, 2, 2)	Ratio	0.711	0.778	0.873	0.918	0.909	0.873
	IC(ω_n^3)	0.476	0.522	0.623	0.749	0.846	0.893
	IC(ω_n^4)	0.788	0.841	0.877	0.857	0.783	0.687
(10, 6, 2)	Ratio	0.018	0.035	0.105	0.372	0.637	0.801
	IC(ω_n^3)	0.000	0.000	0.002	0.046	0.192	0.369
	IC(ω_n^4)	0.096	0.122	0.448	0.744	0.872	0.914
(20, 8, 2)	Ratio	0.000	0.004	0.209	0.630	0.652	0.525
	IC(ω_n^3)	0.000	0.000	0.003	0.136	0.456	0.609
	IC(ω_n^4)	0.037	0.046	0.590	0.523	0.367	0.226
(20, 10, 1)	Ratio	0.000	0.002	0.043	0.493	0.802	0.925
	IC(ω_n^3)	0.000	0.000	0.000	0.033	0.250	0.522
	IC(ω_n^4)	0.001	0.003	0.354	0.831	0.909	0.898
(20, 14, 2)	Ratio	0.000	0.000	0.000	0.060	0.295	0.560
	IC(ω_n^3)	0.000	.000	0.000	0.001	0.008	0.046
	IC(ω_n^4)	0.000	0.000	0.021	0.409	0.698	0.845

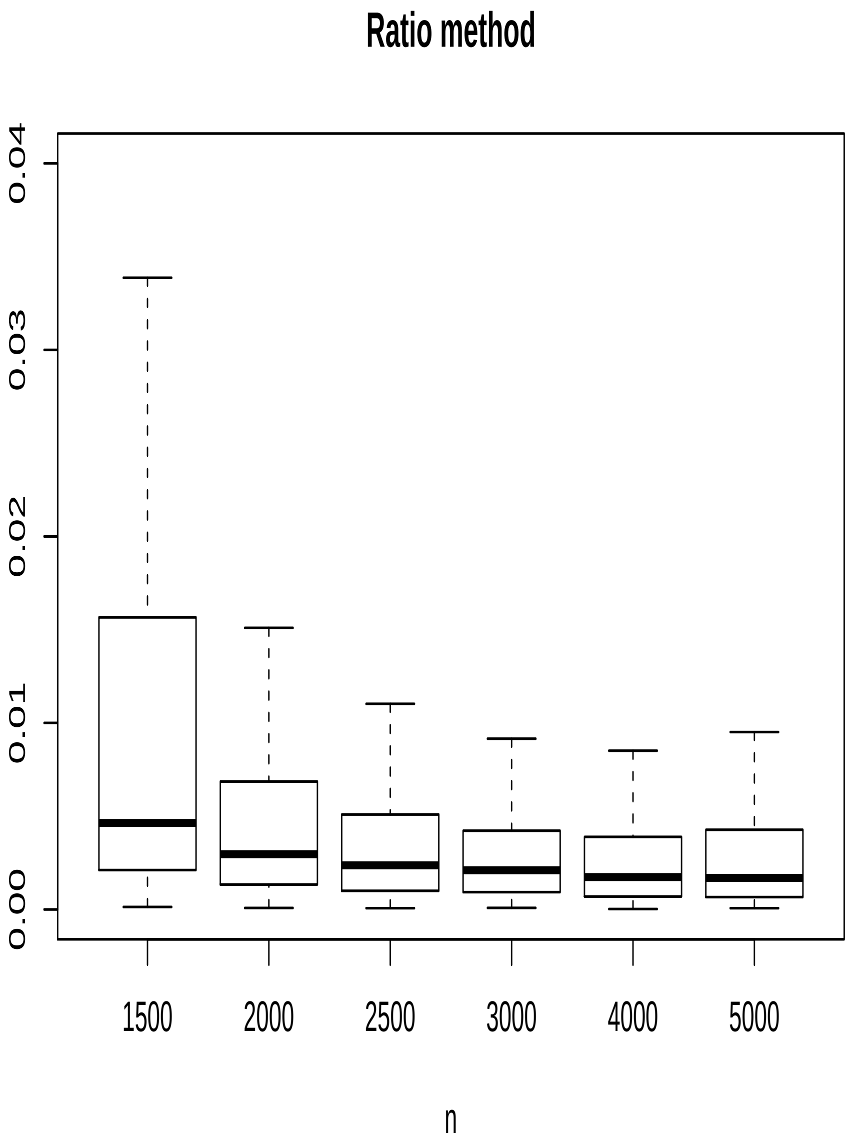
Boxplots of $D_1(\widehat{\mathcal{M}}(\mathbf{A}_2), \mathcal{M}(\mathbf{B}_2))$

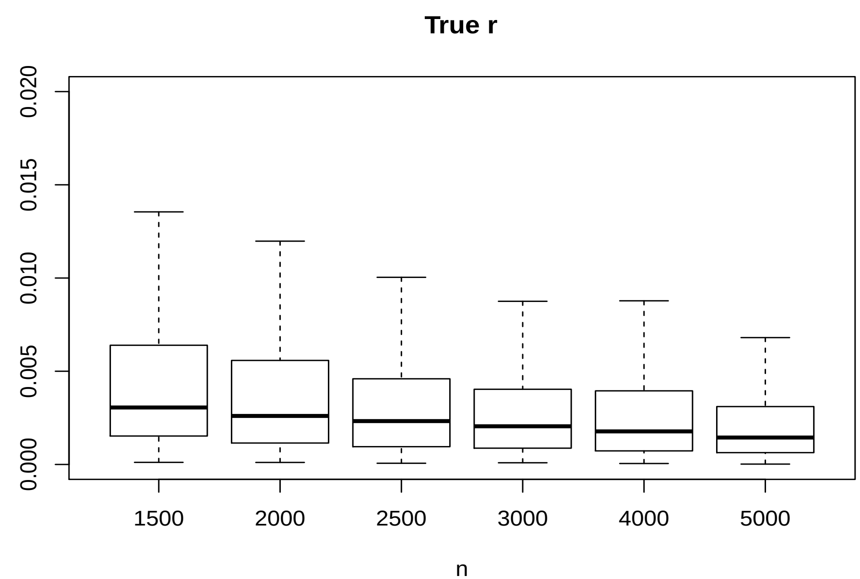




$(p, r, s) = (6, 2, 2),\ 500 \le n \le 2500$, with $\widehat{r}$ (left) and true r (right).

Boxplots of $D_1(\widehat{\mathcal{M}}(\mathbf{A}_2), \mathcal{M}(\mathbf{B}_2))$





$(p, r, s) = (10, 4, 1),\ 1500 \leq n \leq 5000$, with $\widehat{r}$ (left) and true r (right).

Simulation III

In model $\mathbf{y}_t = \mathbf{A}\mathbf{x}_t = \mathbf{A}_1\mathbf{x}_{t1} + \mathbf{A}_2\mathbf{x}_{t2}$,

- all components of $\mathbf{x}_{t2}$ are AR(1) with coefficients drawn from $U(-0.8, 0.8)$
- all components of $\mathbf{x}_{t1}$ are ARIMA(1,1,1) with AR coefficients drawn from $U(0.3, 0.8)$ and MA coefficients from $U(0, 0.95)$
- all innovations are independent and $N(0, 1)$,
- all elements of $\mathbf{A}$ are generated from $U(-3, 3)$.

In IC-estimation for r, use $\omega_n^1 = n^{5/4}\widehat{\lambda}_p$ or $\omega_n^2 = n^{3/2}\widehat{\lambda}_p$.

Relative frequencies for $\tilde{r} = r \equiv p/4$, average distance, 500 replications

p	n	500		1000		1500		2000		250
		Freq	Dist	Freq	Dist	Freq	Dist	Freq	Dist	Freq
8	Johansen	0.390	0.371	0.452	0.326	0.490	0.302	0.480	0.307	0.514
	ratio	0.748	0.174	0.848	0.105	0.884	0.081	0.886	0.079	0.890
	IC(ω_n^1)	0.654	0.217	0.780	0.136	0.802	0.123	0.818	0.112	0.852
	IC(ω_n^2)	0.448	0.338	0.572	0.136	0.628	0.222	0.656	0.206	0.690
12	Johansen	0.210	0.449	0.344	0.355	0.380	0.336	0.400	0.322	0.464
	ratio	0.658	0.236	0.794	0.138	0.770	0.151	0.844	0.102	0.840
	IC(ω_n^1)	0.556	0.261	0.708	0.168	0.748	0.145	0.796	0.114	0.824
	IC(ω_n^2)	0.366	0.358	0.444	0.299	0.518	0.258	0.536	0.247	0.610
20	Johansen	0.008	0.604	0.050	0.503	0.080	0.456	0.134	0.425	0.164
	ratio	0.404	0.390	0.544	0.299	0.620	0.243	0.704	0.184	0.730
	IC(ω_n^1)	0.390	0.342	0.554	0.245	0.670	0.183	0.686	0.154	0.768
	IC(ω_n^2)	0.232	0.417	0.346	0.331	0.400	0.294	0.456	0.256	0.472
28	Johansen	0	0.696	0	0.595	0.002	0.549	0.004	0.522	0.010
	ratio	0.234	0.489	0.386	0.372	0.462	0.332	0.558	0.280	0.582
	IC(ω_n^1)	0.252	0.407	0.454	0.274	0.546	0.228	0.610	0.195	0.700
	IC(ω_n^2)	0.176	0.442	0.270	0.350	0.334	0.304	0.358	0.284	0.436

A real data example

Consider 8 monthly US Industrial Production indices in January 1947 – December 1993 published by the US Federal Reserve:

the total index, manufacturing index, durable manufacturing, nondurable manufacturing, mining, utilities, products, materials.

The proposed method: $\widehat{\mathbf{x}}_t = \widehat{\mathbf{A}}'\mathbf{y}_t$ with $\widehat{r} = 3$,

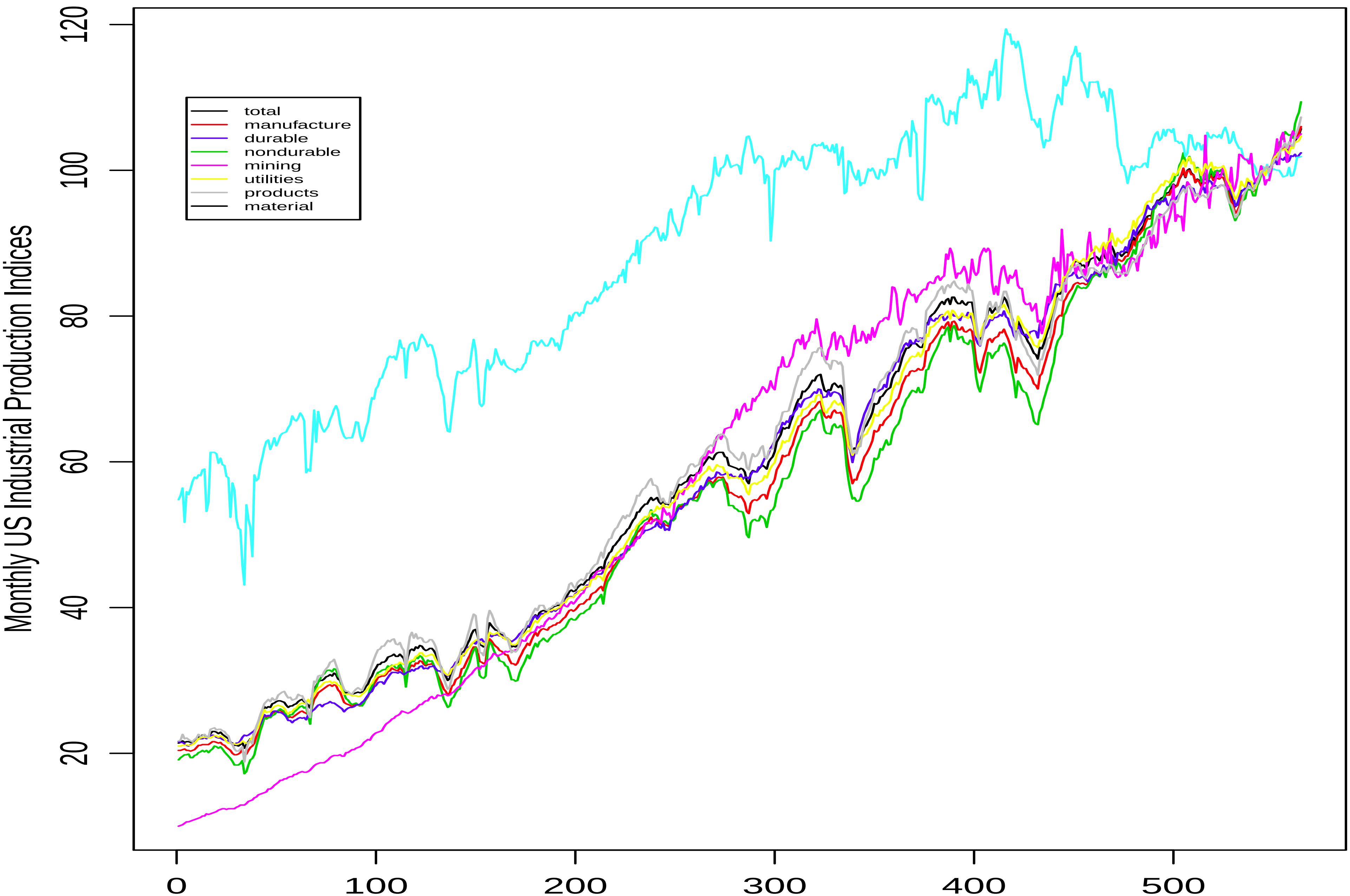
$\widetilde{r} = 3$ with $\omega_n = n^{5/4}\widehat{\lambda}_8$, and 4 with $\omega_n = n^{3/2}\widehat{\lambda}_8$.

Johansen's likelihood method: $\widehat{\mathbf{x}}_t = \widehat{\mathbf{B}}'\mathbf{y}_t$
with the estimated r equal 4

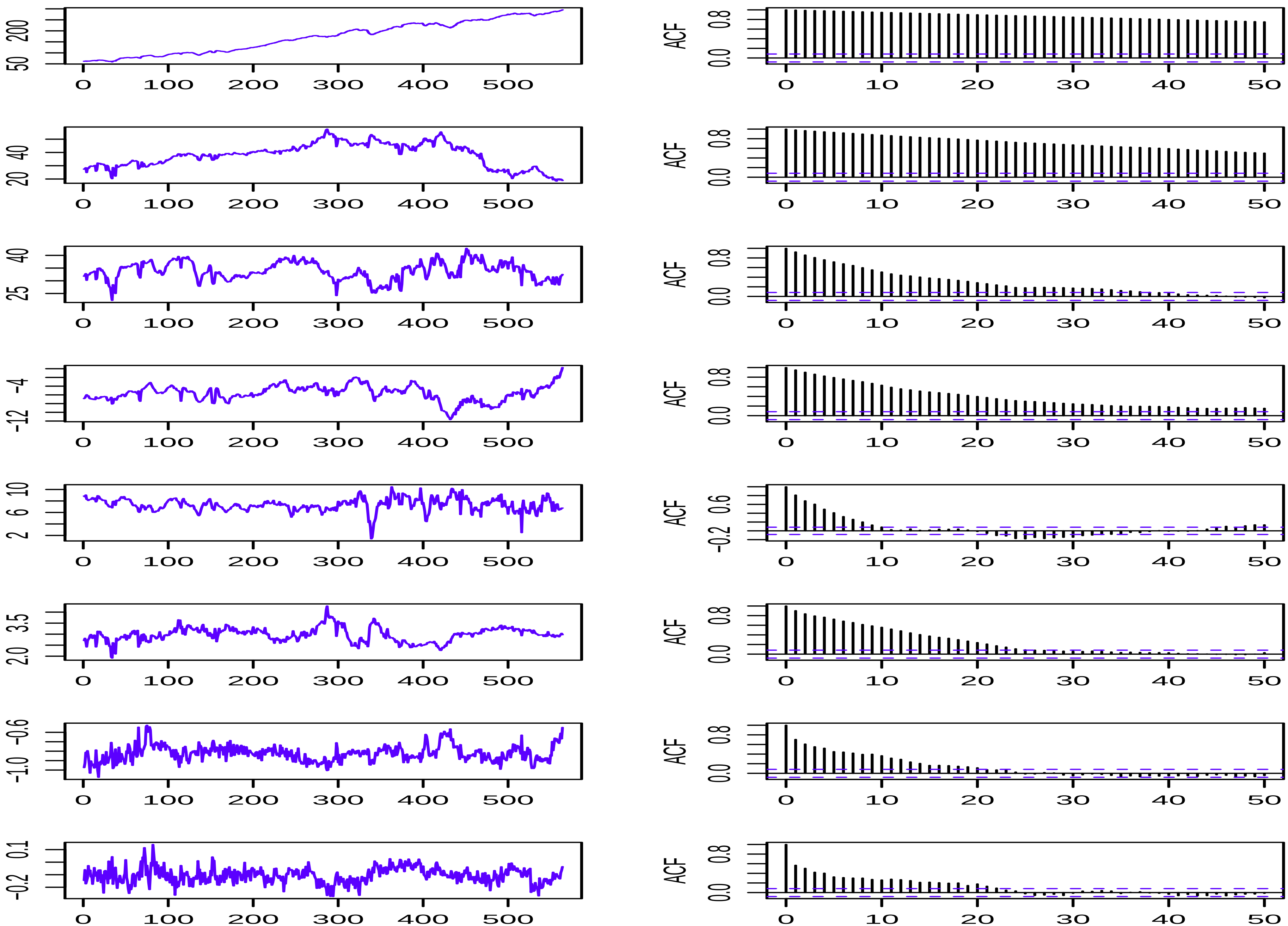
Let $\widehat{\mathbf{A}}_2$ and $\widehat{\mathbf{B}}_2$ consist of the last 4 columns of, respectively, $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$. Then

$$D_1(\mathcal{M}(\widehat{\mathbf{A}}_2), \mathcal{M}(\widehat{\mathbf{B}})) = \Big\{1 - \frac{1}{4}\text{tr}(\widehat{\mathbf{A}}_2\widehat{\mathbf{A}}_2'\widehat{\mathbf{B}}_2(\widehat{\mathbf{B}}_2'\widehat{\mathbf{B}}_2)^{-1}\widehat{\mathbf{B}}_2')\Big\}^{1/2} = \sqrt{1 - 0.9816} = 0.1357.$$

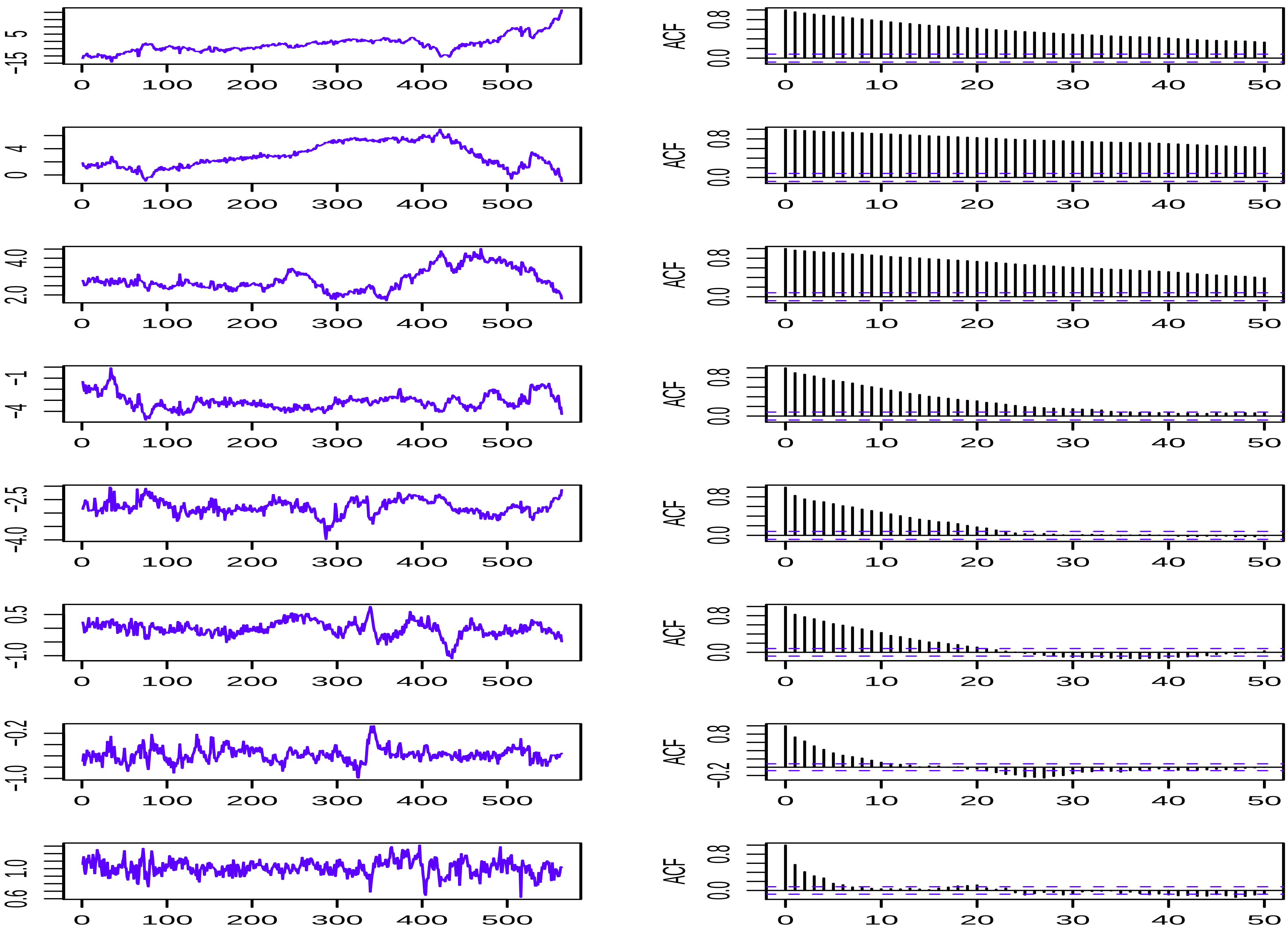
8 monthly U.S. Industrial Production indices in Jan 1947 - Dec 1993



Estimated $\widehat{\mathbf{x}}_t$ by the proposed method



Estimated $\widehat{\mathrm{x}}_t$ by Johansen's likelihood method



Extension to long memory cases

The method still applies when some component series have fractional integrated orders $d > 1/2$.

The asymptotic properties have been established with fixe p.