Part 1 — Review

Time series data occur in various fields, ranging in almost all areas

- Economic/financial: *share prices, export/input totals, average incomes, company profits or sales* in successive *hours/days/weeks/months/years*
- Physical science (meteorology, marine science, geophysics ...): *rainfall, air temperature, sunspot numbers* measured in successive *hours/days/weeks/months /years*
- Demography: *annual population data ...*
- Social science: *annual number of strikes, annual accidental deaths*
- Biology/Ecology: Canadian lynx data, ozone measurements
- .....

Features of time series data:

- The *time order* in the observations is important

- Successive observations are usually *not independent*!

Remark.

1. Due to the dependence in the data, it is possible to predict the future from the data available at present

2. X_t could be *almost independent of* X_{t+h} for large h. (Long-term prediction is difficult, if not impossible!)

Objectives of time series analysis:

1. enhence the understanding of the mechanism/dynamics generating the series

2. separate noise from signals

3. predict the future

4. control systems (not covered in this course)

Two approaches:

- analysis in time domain

- analysis in frequency domain — Fourier transforms

Plots and transformation

Time plot: plot X_t against t — the first step in analysis

(a) Discontinuities (eg. sudden changes of level — break the time series into homogeneous segments)

(b) Outlying observations (carefully checking for possible mistakes in recording process)

(c) Logarithmic transformation for positive data: $X_t \rightarrow \log X_t$ to stablise the variance when Var$(X_t) \nearrow$ as $E(X_t) \nearrow$. (Other transformation: Box-Cox transforms.)

After preliminary transformation, the data should be close to the form of *classical decomposition*:

$$X_t = m_t + s_t + Y_t,$$

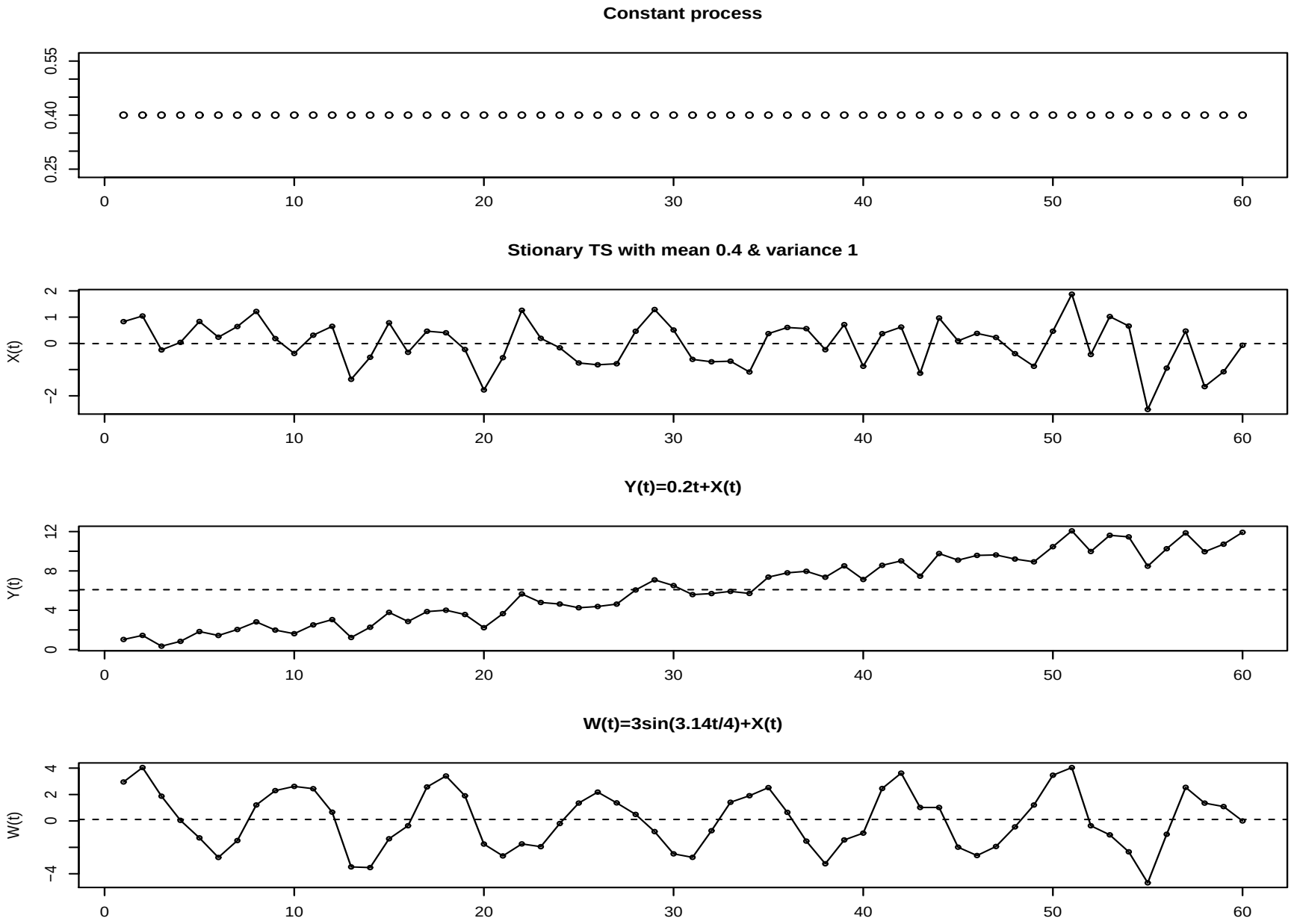
where

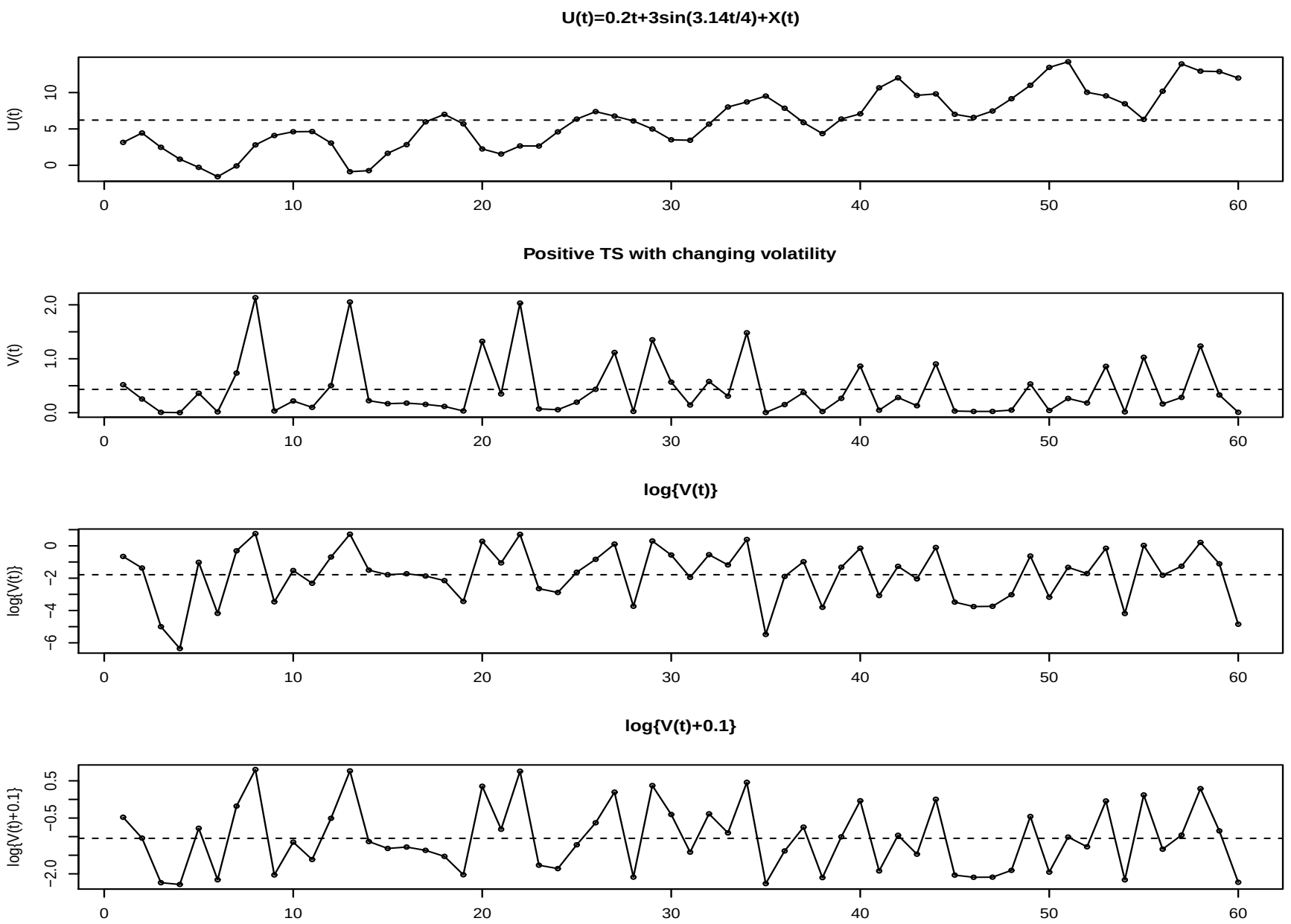
m_t — trend component (a slowly changing function)

s_t — seasonal component (a periodic function)

Y_t — stationary time series

Note. Both m_t and s_t are deterministic, Y_t is random.


Constant process
0.25
0.40
0.55
0
10
20
30
40
50
60
Stionary TS with mean 0.4 & variance 1
X(t)
−2
0
1
2
0
10
20
30
40
50
60
Y(t)=0.2t+X(t)
Y(t)
0
4
8
12
0
10
20
30
40
50
60
W(t)=3sin(3.14t/4)+X(t)
W(t)
−4
0
2
4
0
10
20
30
40
50
60


U(t)=0.2t+3sin(3.14t/4)+X(t)
U(t)
0
5
10
0
10
20
30
40
50
60
Positive TS with changing volatility
V(t)
0.0
1.0
2.0
0
10
20
30
40
50
60
log{V(t)}
log{V(t)}
−6
−4
−2
0
0
10
20
30
40
50
60
log{V(t)+0.1}
log{V(t)+0.1}
−2.0
−0.5
0.5
0
10
20
30
40
50
60

A general approach to time series analysis

(a) Plot the data, cheeking various features such as

- a trend
- a seasonal component
- any apparent sharp changes in hehaviour
- any outlying observations

(b) remove both trend and seasonal components, aiming for stationary residuals

(c) Choose a probability model to fit the residuals — *the main topic of this course*.

Stationarity

Definition. $\{X_t, t = 0, \pm 1, \pm 2, \cdots\}$ is (weakly) stationary if $E(X_t^2) < \infty$, and

(i) $E(X_t)$ is a constant independent of t, and
(ii) $\text{Cov}(X_t, X_{t+k})$ is independent of t for each k.

Definition. $\{X_t, t = 0, \pm 1, \pm 2, \cdots\}$ is strictly stationary if $(X_1, \cdots, X_n)$ and $(X_{1+k}, \cdots, X_{n+k})$ have the same joint distributions for all integers k and $n > 0$.

Remarks. (i) Stationarity only assumes that the first two moments of time series are time-invariant.

(ii) Strict stationarity implies stationarity provided $EX_t^2 < \infty$.

(iii) For Gaussian processes, the two definitions are equivalent.

Example. The MA(∞) process defined below is stationary.

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} \quad \text{for all } t,$$

where $\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2)$ and $\sum_{j=0}^{\infty} |a_j| < \infty$. Note

$$E|X_t| \le E|\varepsilon_1| \sum_{j=0}^{\infty} |a_j| < \infty.$$

Therefore, the above X_t is well-defined in the sense of convergence in probability.

Further, $EX_t = 0$, and

$$\text{Cov}(X_t, X_{t+k}) = \sum_{j,l=0}^{\infty} a_j a_l E(\varepsilon_{t-j}\, \varepsilon_{t+k-l}) = \sigma^2 \sum_{j=0}^{\infty} a_j a_{j+|k|},$$

which is independent of t.

Remark. (i) If $\{\varepsilon_t\}$ are i.i.d and $E|\varepsilon_t| < \infty$, the above MA($\infty$) process is strictly stationary.

(ii) Any MA(q) ($q < \infty$) process is stationary.

ARMA(p, q) model:

$$X_t = b_1 X_{t-1} + \cdots b_p X_{t-p} + \varepsilon_t + a_1 \varepsilon_{t-1} + \cdots + a_q \varepsilon_{t-q}, \qquad (1)$$

where $\{\varepsilon_t\} \sim$WN$(0, \sigma^2)$. It may be written as

$$b(B) X_t = a(B) \varepsilon_t,$$

where B is *backshift operator* defined as

$$B^k X_t = X_{t-k} \quad k = 0, \pm 1, \pm 2, \cdots,$$

and

$$b(z) = 1 - b_1 z - \cdots - b_p z^p, \quad a(z) = 1 + a_1 z + \cdots + a_q z^q.$$

Theorem. Suppose $b(\cdot)$ and $a(\cdot)$ do not have common factors. Then ARMA process $\{X_t, t = 0, \pm 1, \pm 2, \cdots\}$ defined above is stationary if $b(z) \neq 0$ for all complex numbers z such that $|z| \leq 1$.

Proof. We have a Taylor expansion

$$b(z)^{-1} = \sum_{j=0}^{\infty} c_j z^j \equiv c(z)$$

for all $|z| < \delta$, where $\sum_{j=0}^{\infty} |c_j| < \infty$. It is easy to see that $c(z)\, b(z) \equiv 1$. Therefore

$$X_t = c(B)b(B)X_t = c(B)a(B)\varepsilon_t = d(B)\varepsilon_t,$$

where $d(z) = c(z)a(z) = \sum_{j=0}^{\infty} d_j z^j$ with $\sum_{j=0}^{\infty} |d_j| < \infty$. This implies that $\{X_t\}$ ~MA(∞), therefore is stationary.

Remark. There exists a stationary solution of ARMA(p,q) equation (1), provided that $b(z) \neq 0$ for all $|z| = 1$ (i.e. no unit roots for $b(z) = 0$).

Causality: a time series $\{X_t\}$ is *causal*, or a causal function of $\{\varepsilon_t\}$ if for all t

$$X_t = \sum_{j=0}^{\infty} d_j \varepsilon_{t-j},$$

where $\sum_{j=0}^{\infty} |d_j| < \infty$.

Remark. (i) Causal ARMA processes are stationary but not vice versa.

(ii) Any stationary non-causal ARMA process can be represented as a causal ARMA process (with the same orders) in terms of a newly defined white noise.

(iii) In the above theorem, the condition that the process $\{X_t\}$ is double infinite in time is important. For example, the process defined by

$$X_t = 0.5 X_{t-1} + \varepsilon_t$$

for $t = 0, \pm 1, \pm 2, \cdots$ is stationary (also strictly stationary), where $\{\varepsilon_t\} \sim_{i.i.d} N(0, 1)$. However the process defined by the above equation for $t = 1, 2, \cdots$ only, and initiated at $X_0 \sim U(0, 1)$ is no longer stationary since $EX_t = 0.5^{t+1}$ for all $t \geq 0$. The process $\{X_t, t = 1, 2, \cdots\}$ will be stationary if and only if we start the process with $X_0 \sim N(0, 1/0.75)$.

(iv) When $b(z) \neq 0$ for all $|z| \leq 1$, $\{X_t, t = 0, \pm 1, \pm 2, \cdots\}$ is causal stationay, and $\{X_t, t = 1, 2, \cdots\}$ is called *stable* (regardless the initial values of $X_0, X_{-1}, X_{-2}, \cdots$).

(v) $X_t = 2X_{t-1} + \varepsilon_t$ defines a stationary but non-causal process $X_t = 0.5X_{t+1} - 0.5\varepsilon_{t+1}$, as

$$X_t = -0.5\varepsilon_{t+1} + 0.5(0.5X_{t+2} - 0.5\varepsilon_{t+2}) = -0.5\varepsilon_{t+1} - (0.5)^2\varepsilon_{t+2} + (0.5)^2 X_{t+2}$$
$$= -\sum_{j=1}^{k} (0.5)^j \varepsilon_{t+j} + (0.5)^k X_{t+k} = -\sum_{j=1}^{\infty} (0.5)^j \varepsilon_{t+j}.$$

Wold Decomposition Theorem. Any stationary process $\{X_t\}$ can be written as

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} + V_t,$$

where $\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2)$, $\sum_j a_j^2 < \infty$, $\{V_t\}$ is stationary, and also *deterministic* in the sense that V_t is entirely determined by its lagged values $V_{t-1}, V_{t-2}, \cdots$, and $\{\varepsilon_t\}$ and $\{V_t\}$ are two uncorrelated processes.

Remark. (i) $\{X_t\}$ is called *purely non-deterministic* if $V_t \equiv 0$.

(ii) If $\{X_t\}$ is a Gaussian process, both $\{\varepsilon_t\}$ and $\{V_t\}$ are also Gaussian.

(iii) When $V_t \equiv 0$, and $\{\varepsilon_t\}$ are independent with each other, X_t is called a linear processes.

Stationary Gaussian processes

Theorem. Let $\{X_t\}$ be stationary Gaussian time series.

(i) $\{X_t\} \sim$ MA(∞) with $\{\varepsilon_t\} \sim_{i.i.d.} N(0, \sigma^2)$ if it is a purely non-deterministic process

(ii) $\{X_t\} \sim$ MA(q) if it is a q-dependent process.

(iii) $\{X_t\} \sim$ AR(p) if given $\{X_{t-1}, \cdots, X_{t-p}\}$, X_t is independent of $\{X_{t-k}, k > p\}$.

Proof. (i) and (ii) follow from Wold decomposition theorem directly.

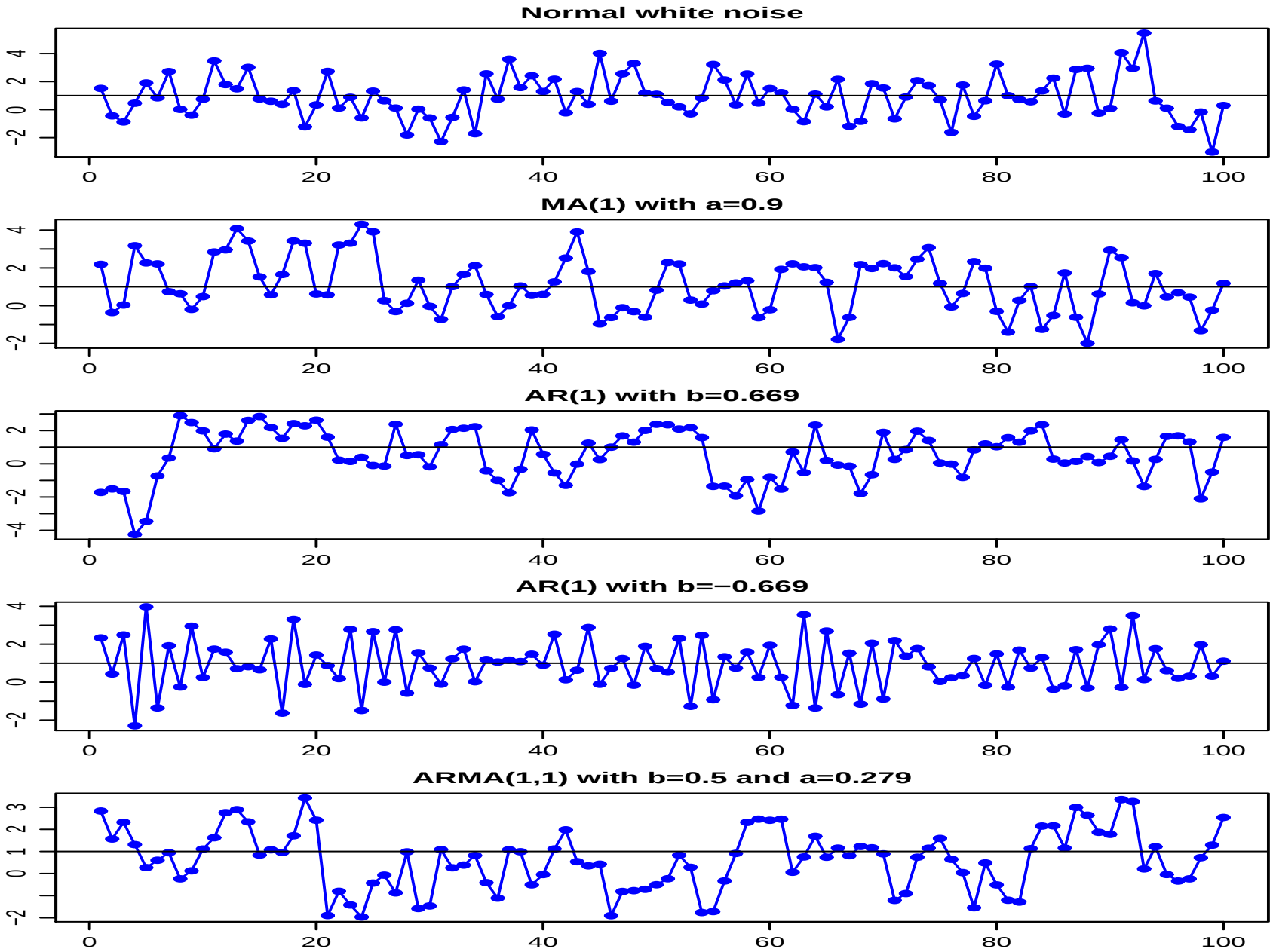
To prove (iii), let $\varepsilon_t = X_t - E(X_t | X_{t-1}, \cdots, X_{t-p})$.

Then ε_t is independent of $\{X_{t-k}, k \geq 1\}$ since Cov$(\varepsilon_t, X_{t-k}) = 0$ for $k \geq 1$. Therefore ε_t is also independent of $\{\varepsilon_{t-k}, k \geq 1\}$ since ε_{t-k} is a function of $\{X_{t-k}, X_{t-k-1}, \cdots\}$ only. Hence $\{\varepsilon_t\} \sim_{i.i.d.} N(0, \sigma^2)$. Due to the normality, $E(X_t | X_{t-1}, \cdots, X_{t-p})$ is a linear function of $X_{t-1}, \cdots, X_{t-p}$.

Generate 5 time series with the same marginal distribution $N(1, 1.81)$ from model

$$X_t - \mu = b(X_t - \mu) + \varepsilon_t + a\varepsilon_{t-1}, \quad \varepsilon_t \sim_{iid} N(0, 1),$$

The top panel is patternless as the points fluctuate randomly around the mean. In contrast the two successive points of MA(1) process in the second panel is positively correlated as $\rho(1) = 0.9/(1 + 0.9^2) = 0.497$, and there is the tendency that X_t and X_{t-1} stay on the same side of the mean. This tendency is much more pronounced for the AR(1) process with $b = 0.669$ in the 3rd panel and the ARMA(1, 1) process in the bottom panel. In both cases X_t is effectively an MA(∞) process with positive coefficients, and it tends to take long excursions before it returns to the mean — a clear indication of serial dependence in the data. The 4th panel shows a different pattern of AR(1): since $b = -0.669$, X_t oscillates around the mean in such a way that X_t and X_{t-1} tend to take opposite sites of the mean.


Normal white noise
MA(1) with a=0.9
AR(1) with b=0.669
AR(1) with b=−0.669
ARMA(1,1) with b=0.5 and a=0.279
0
20
40
60
80
100

Autocovariance and autocorrelation

For stationary time series $\{X_t\}$,

$$\gamma(k) = \mathrm{Cov}(X_{t+k}, X_t) = \mathrm{Cov}(X_{1+k}, X_1)$$

is called the *autocovariance function* (ACVF), and

$$\rho(k) = \gamma(k)/\gamma(0) = \mathrm{Corr}(X_{1+k}, X_1)$$

is called the *autocorrelation function* (ACF).

Theorem. (*Characterization of ACVF*).
A real-valued function $\gamma(\cdot)$ defined on the integers is the ACVF iff it is even and nonnegative definite in the sense that

$$\sum_{i,j=1}^{n} a_i a_j \gamma(i-j) \geq 0,$$

for any positive integer $n \geq 1$ and arbitary real numbers $a_1, \cdots, a_n$.

1. WN(0, σ^2): $\rho(k) = 0$ for all $k \neq 0$.
2. MA(∞):

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j},$$

where $\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2)$, $a_0 = 1$ and $\sum_{j=0}^{\infty} |a_j| < \infty$.

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} a_j a_{j+|k|}, \quad \rho(k) = \frac{\sum_{j=0}^{\infty} a_j a_{j+|k|}}{\sum_{j=0}^{\infty} a_j^2}.$$

3. MA(q):

$$\gamma(k) = \sigma^2 \sum_{j=0}^{q-|k|} a_j a_{j+|k|}, \quad \rho(k) = \frac{\sum_{j=0}^{q-|k|} a_j a_{j+|k|}}{\sum_{j=0}^{q} a_j^2} \quad \text{for } |k| \leq q,$$

and $\rho(k) = 0$ for $|k| > q$.

4. Causal ARMA(p,q): $b(B)X_t = a(B)\varepsilon_t$.

   (a) Write

$$X_t = \sum_{j=0}^{\infty} d_j \varepsilon_{t-j},$$

where

$$\begin{aligned} d_0 &= a_0\,(=1), \\ d_1 &= a_1 + d_0 b_1, \\ d_2 &= a_2 + d_0 b_2 + d_1 b_1, \\ d_3 &= a_3 + d_0 b_3 + d_1 b_2 + d_2 b_1, \\ &\cdots\cdots \end{aligned}$$

(We assume $a_j = 0$ for $j > q$ and $b_i = 0$ for $i > p$ in the above operation.) Use the results for MA(∞).

(b) Yule-Walker equation

$$\gamma(k) - b_1\gamma(k-1) - \cdots - b_p\gamma(k-p) = 0, \quad k \geq \max(p, q+1).$$

It has the general soluation

$$\gamma(k) = \alpha_1\xi_1^{-k} + \cdots + \alpha_p\xi_p^{-k},$$

where $\xi_1, \cdots, \xi_p$ are the p roots of equation $b(z) = 0$. (The causality implies $|\xi_j| > 1$.)

Theorem. (i) For causal ARMA processes, $\rho(k) \to 0$ at exponential rates as $|k| \to \infty$.

(ii) For MA(q) processes, $\rho(k) = 0$ for all $|k| > q$.

Estimation of ACVF and ACF

Given $\{X_1, \cdots, X_n\}$ from a stationary time series, we estimate the ACVF by its sample version

$$\widehat{\gamma}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n)$$

for $k = 0, 1, \cdots, n-1$, where $\bar{X}_n = \frac{1}{n} \sum_{t=1}^{n} X_t$, and estimate the ACF by

$$\widehat{\rho}(k) = \widehat{\gamma}(k)/\widehat{\gamma}(0).$$

Remark. (i) In practice, $n \geq 50$ and $k \leq n/4$ are usually required (Box and Jenkins, 1970, p.30).

(ii) The divisor n in the definition of $\widehat{\gamma}(k)$ may be replaced by $n-k$. But $\widehat{\rho}(k)$ and $\widehat{\gamma}(k)$ defined above are more preferable in practice because (i) $\{\widehat{\gamma}(k)\}$ is a nonnegative definite function if we define $\widehat{\gamma}(-k) = \widehat{\gamma}(k)$ for $k \geq 1$ and $\widehat{\gamma}(k) = 0$ for $|k| \geq n$ and $|\widehat{\rho}(k)| \leq 1$ (Those properties may be lost if we replace the divisor n by $n-k$), and (ii) the smaller variance of $\widehat{\gamma}(k)$ compensates its the larger bias when k becames large relative to n.

The sample ACF plays an active role in model selection and spectral analysis. For example, the ACF of an MA(q) process cuts off at q. But, its sample ACF will not have a clear cut at lag q due to random flunctuation. The proper statistical inference rests on the sampling distributions of the statistics involved. Let $\boldsymbol{\rho}(k) = (\rho(1), \cdots, \rho(k))^\tau$ and $\widehat{\boldsymbol{\rho}}(k)$ defined in the same way.

Theorem. Let $\{X_t\}$ be a stationary process defined as

$$X_t = \mu + \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j},$$

where $\{\varepsilon_t\} \sim \text{IID}(0, \sigma^2)$ and $\sum_{j=-\infty}^{\infty} |a_j| < \infty$.

(i) If $\sum_{j=-\infty}^{\infty} a_j \neq 0$, $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} N(0, \nu_1^2)$, where

$$\nu_1^2 = \sum_{j=-\infty}^{\infty} \gamma(j) = \sigma^2 (\sum_{j=-\infty}^{\infty} a_j)^2.$$

(ii) If $E\varepsilon_t^4 < \infty$, $\sqrt{n}\{\widehat{\gamma}(0) - \gamma(0)\} \xrightarrow{D} N(0, \nu_2^2)$, where

$$\nu_2^2 = 2\sigma^2 \sum_{j=-\infty}^{\infty} \rho^2(j) = 2\sigma^2\{1 + 2\sum_{j=1}^{\infty} \rho^2(j)\}.$$

(iii) If $E\varepsilon_t^4 < \infty$, $\sqrt{n}\{\widehat{\boldsymbol{\rho}}(k) - \boldsymbol{\rho}(k)\} \xrightarrow{D} N(0, \mathbf{W})$, where $\mathbf{W}$ is a $k \times k$ matrix with its (i, j)-th element given by Bartlett's formula

$$w_{ij} = \sum_{t=-\infty}^{\infty} \{\rho(t+i)\rho(t+j) + \rho(t-i)\rho(t+j)$$
$$+ 2\rho(i)\rho(j)\rho^2(t) - 2\rho(i)\rho(t)\rho(t+j) - 2\rho(j)\rho(t)\rho(t+i)\}.$$

(Proof. See §7.3 of Brockwell and Davis 1999.)

Remark. If $\{X_t\}$ is an MA(q) process (i.e. $a_j = 0$ for all $j < 0$ and $j > q$)

$$\sqrt{n}\,\widehat{\rho}(j) \xrightarrow{D} N\left(0, 1 + 2\textstyle\sum_{t=1}^{q} \rho^2(q)\right), \quad j > q.$$

Partial autocorrelation function (PACF)

Definition. Let $\{X_t\}$ be a stationary time series and $EX_t = 0$. The PACF is defined as $\pi(1) = \text{Corr}(X_1, X_2) = \rho(1)$ and

$$\pi(k) = \text{Corr}(R_{1|2,\cdots,k}, R_{k+1|2,\cdots,k})$$

for $k \geq 2$, where $R_{j|2,\cdots,k}$ is the residual of the linear regression of X_j on $(X_2, \cdots, X_k)$, namely

$$R_{j|2,\cdots,k} = X_j - (\alpha_2 X_2 + \cdots + \alpha_k X_k),$$

$$(\alpha_2, \cdots, \alpha_k) = \text{argmin}_{\beta_2,\cdots,\beta_k} E\{X_j - (\beta_2 X_2 + \cdots + \beta_k X_k)\}^2.$$

Remark. (i) For causal AR(p) models, $\pi(k) = 0$ for all $k > p$.

(ii) For $k \geq 1$, $\pi(k) = b_{kk}$, where

$$(b_{1k}, \cdots, b_{kk}) = \text{argmin}_{b_1,\cdots,b_k} E\{X_t - (b_1 X_{t-1} + \cdots + b_k X_{t-k})\}^2.$$

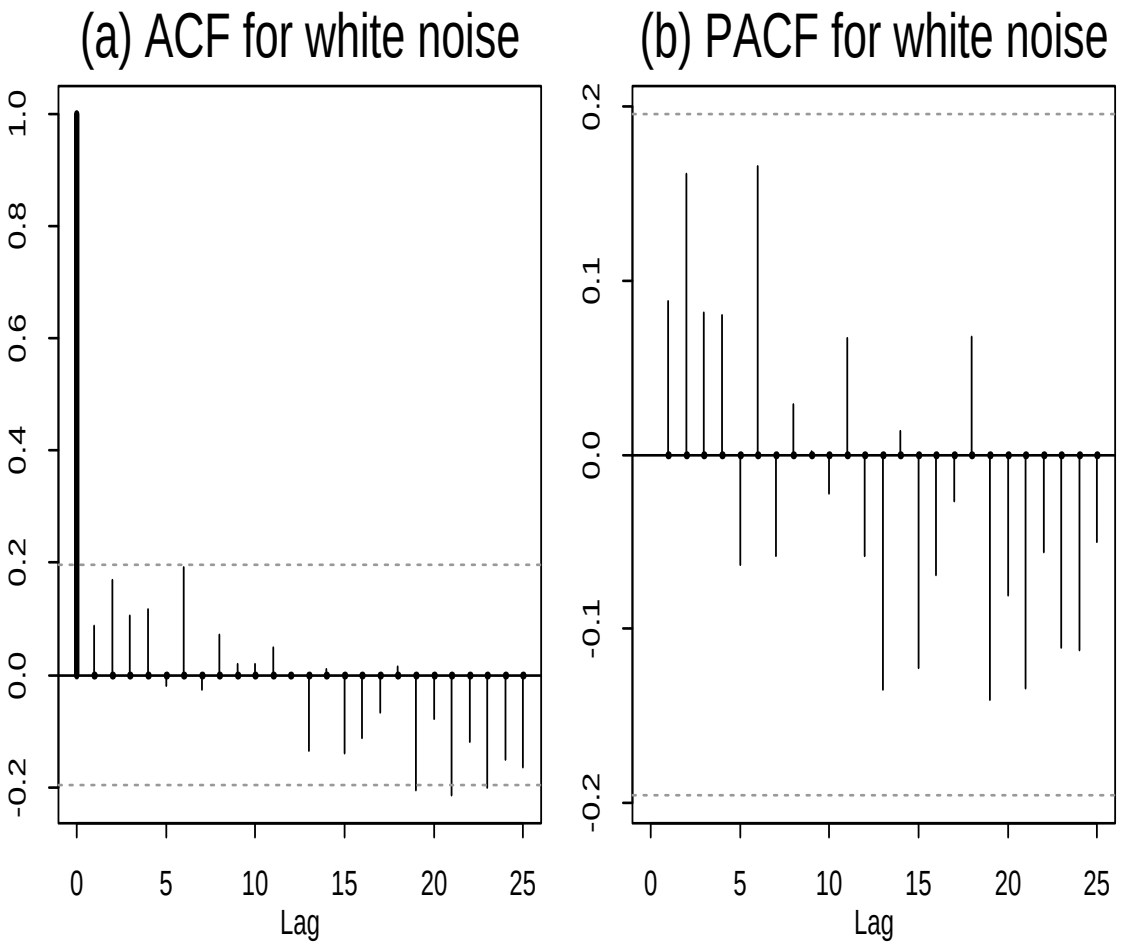
(Proof. See Corollary 5.2.1 of BD, 1991.)

(iii) In view of (ii) above, $\widehat{\pi}_i \equiv \widehat{b}_{kk} (= \widehat{b}_k)$ which is the estimator for the last coefficient in fitting the data with an AR(k) model.

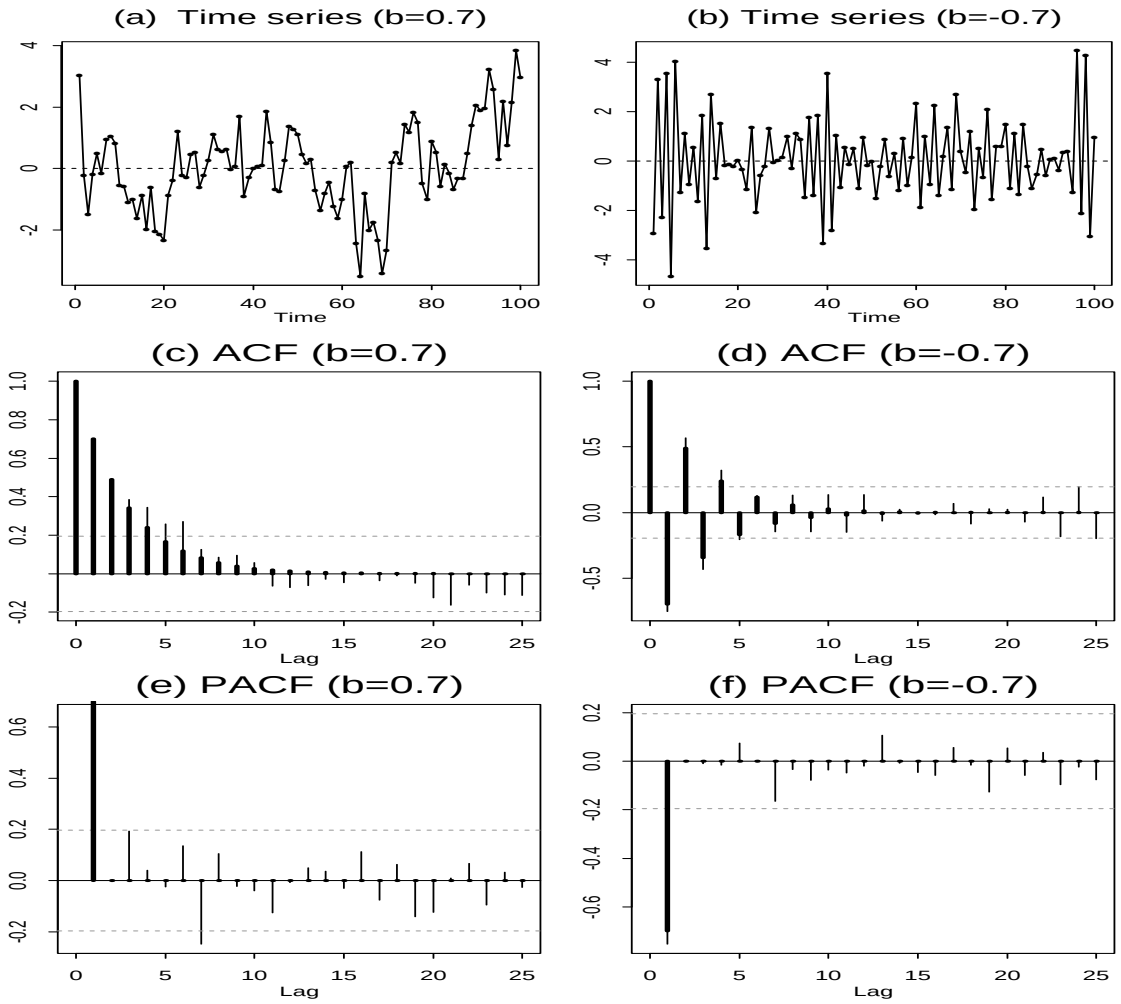
In following examples, we plot some estimated ACFs and PACFs (thin lines) based on samples of size $n = 100$ together with the true ACFs and PACFs (thick lines). We also super-impose the horizontal lines at $\pm 1.96/\sqrt{n}$. We always use standard normal while noise $\{\varepsilon_t\} \sim_{i.i.d} N(0, 1)$.

Note. For stationary ARMA processes, both $\widehat{\rho}(k)$ and $\widehat{\pi}(k)$ tend to overestimate $\rho(k)$ and $\pi(k)$ when k is large. (This could attribute to (i) the lack of information for large k's and (ii) the errors become 'overwhelming' when the true values are close to 0.)

1. Normal WN(0, 1)



2. AR(1) model $X_t = bX_{t-1} + \varepsilon_t$ with $b = 0.7$ and -0.7.

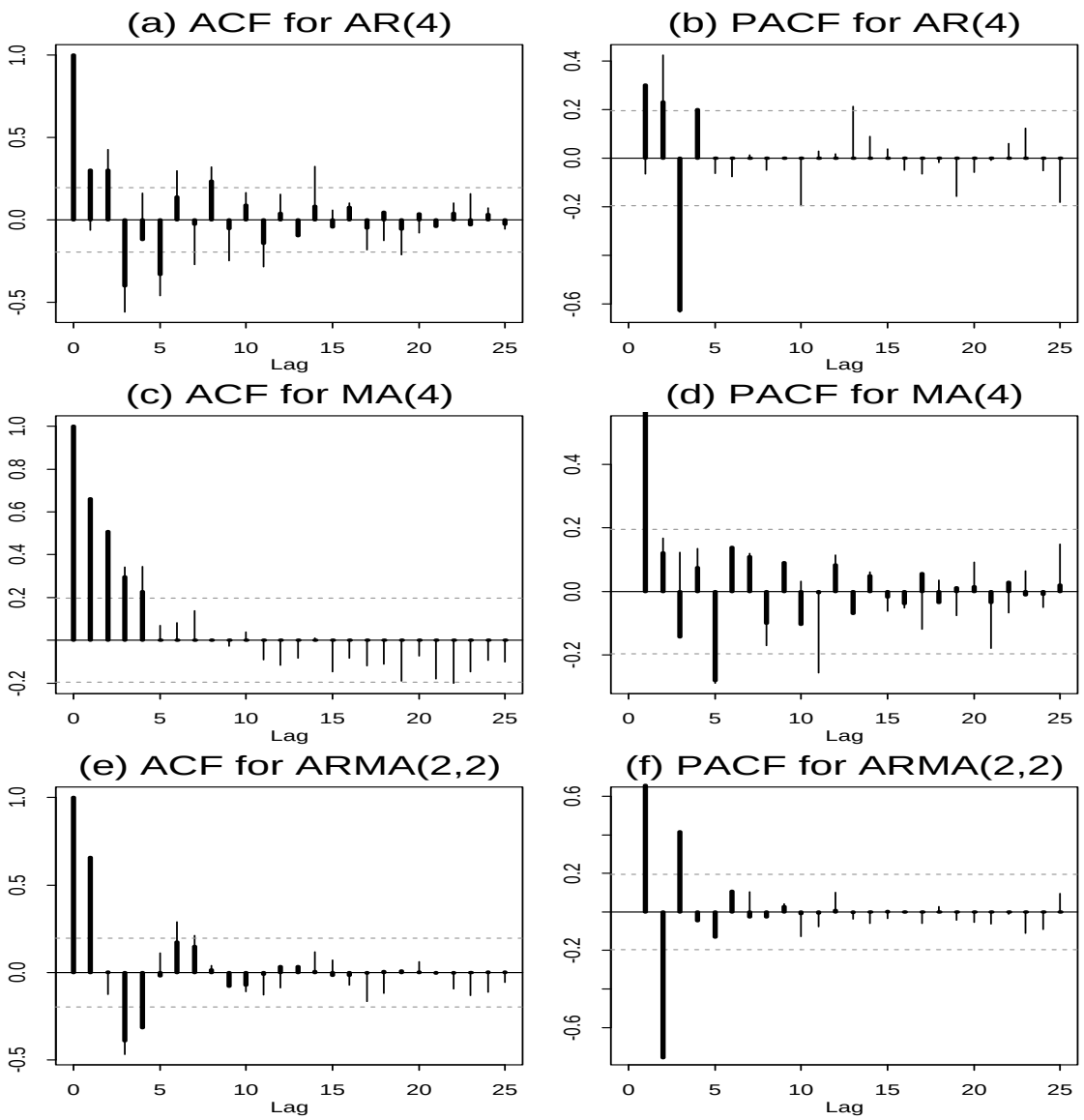


3. Three ARMA models:

AR(4): $X_t = 0.5X_{t-1} + 0.3X_{t-2} - 0.7X_{t-3} + 0.2X_{t-4} + \varepsilon_t$.

MA(4): $X_t = \varepsilon_t + 0.6\varepsilon_{t-1} + 0.6\varepsilon_{t-2} + 0.3\varepsilon_{t-3} + 0.7\varepsilon_{t-4}$.

ARMA(2,2): $X_t = 0.8X_{t-1} - 0.6X_{t-2} + \varepsilon_t + 0.7\varepsilon_{t-1} + 0.4\varepsilon_{t-2}$.


(a) ACF for AR(4)
(b) PACF for AR(4)
(c) ACF for MA(4)
(d) PACF for MA(4)
(e) ACF for ARMA(2,2)
(f) PACF for ARMA(2,2)
Lag

Model selection using ACF, PACF and EACF

1. For MA(q) processes, ACF cuts off at q

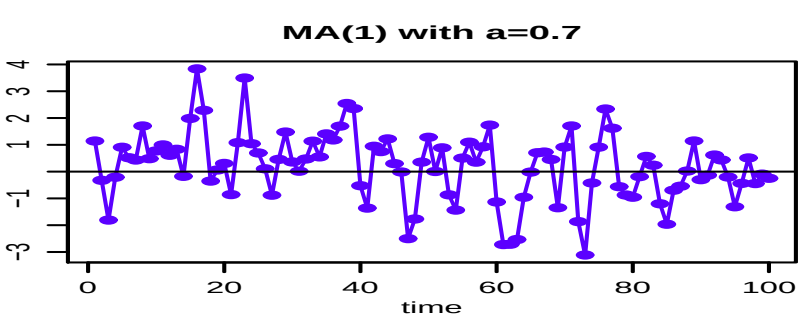
For an MA(q) process defined with innovations $\varepsilon_t \sim \text{IID}(0, \sigma^2)$ and $E(\varepsilon_t^4) < \infty$, it holds that as $T \to \infty$,

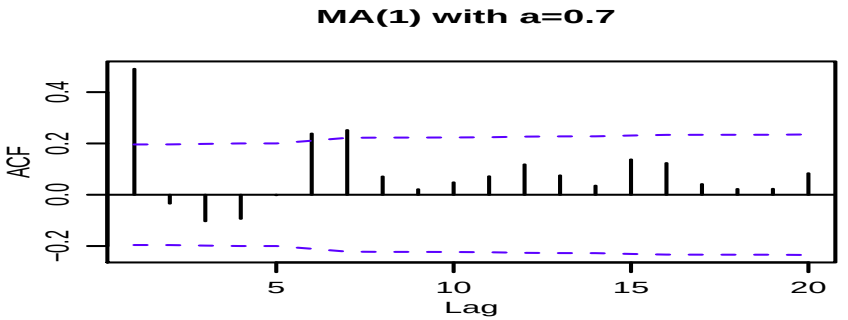
$$\sqrt{T}\,\widehat{\rho}(k) \xrightarrow{D} N\Big(0,\ 1 + 2\sum_{j=1}^{q} \rho(j)^2\Big), \quad \text{for all } k > q.$$

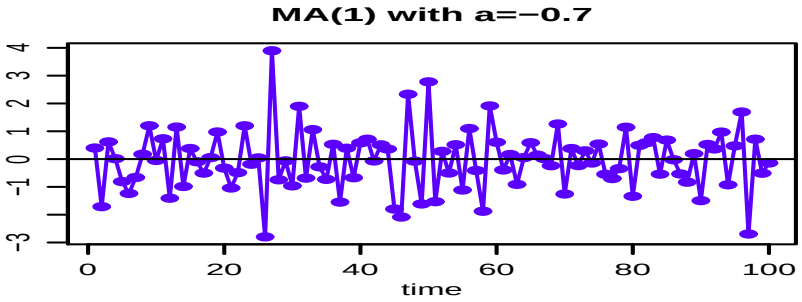
In sample ACF plots, we superimpose two confidence bounds

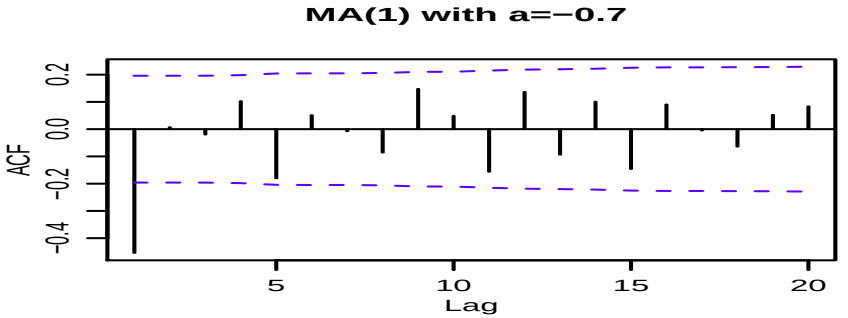
$$\pm 1.96\{1 + 2\sum_{j=1}^{k-1} \rho(j)^2\}^{1/2} \Big/ \sqrt{T} \tag{2}$$

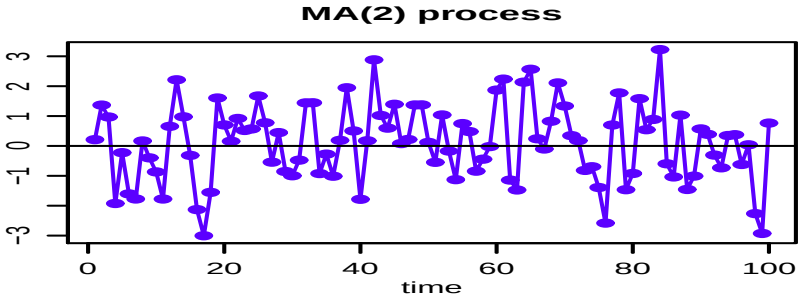
against k.

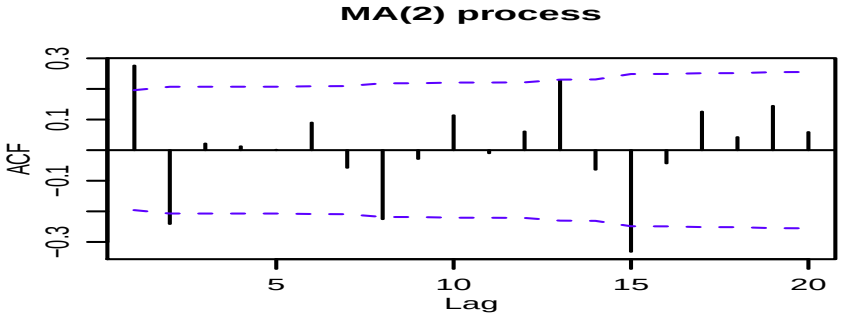

MA(1) with a=0.7
time

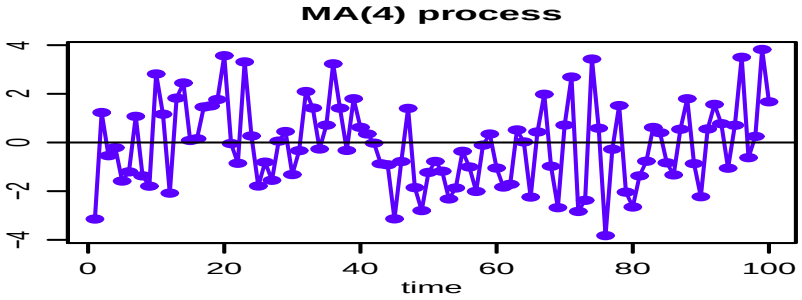

MA(1) with a=0.7
ACF
Lag

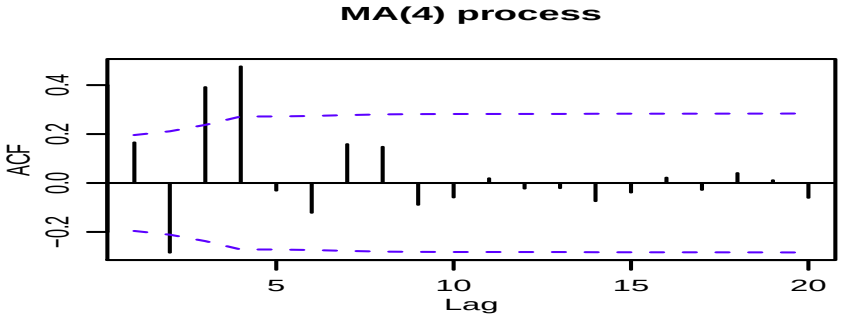

MA(1) with a=−0.7
time


MA(1) with a=−0.7
ACF
Lag


MA(2) process
time


MA(2) process
ACF
Lag


MA(4) process
time


MA(4) process
ACF
Lag

A note on R. To produce 1st ACF plot

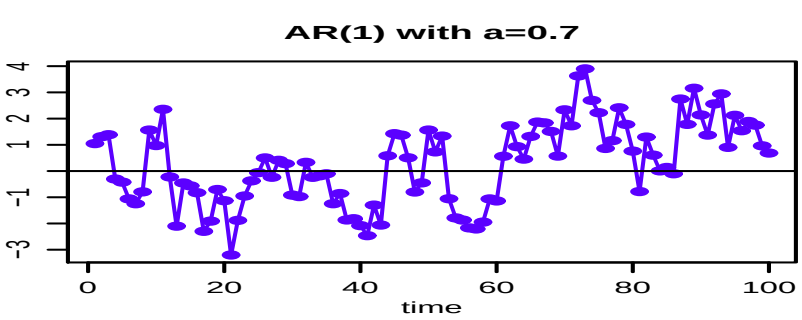
```
x = arima.sim(n=100, list(ma=0.7))
acf(x1, ci.type='ma', main="MA(1) with a=0.7")
```

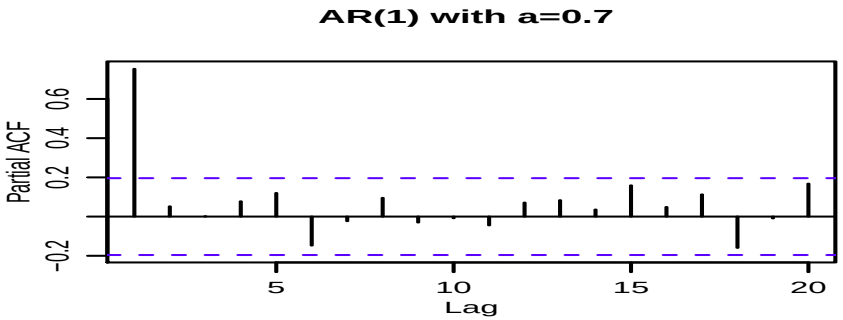
2. For AR(p), PACF cuts off at p.

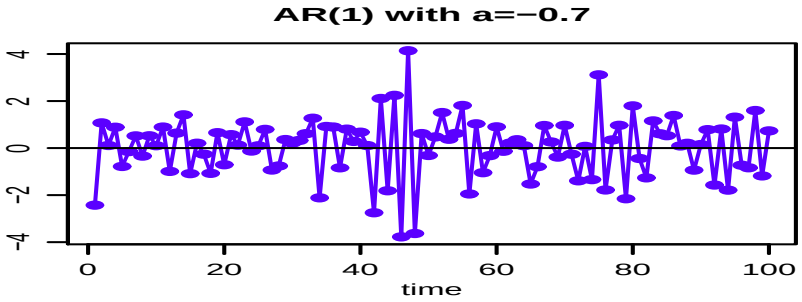
For an AR(p) process with innovations $\varepsilon_t \sim \text{IID}(0, \sigma^2)$ and $E(\varepsilon_t^4) < \infty$, it holds that as $T \to \infty$,

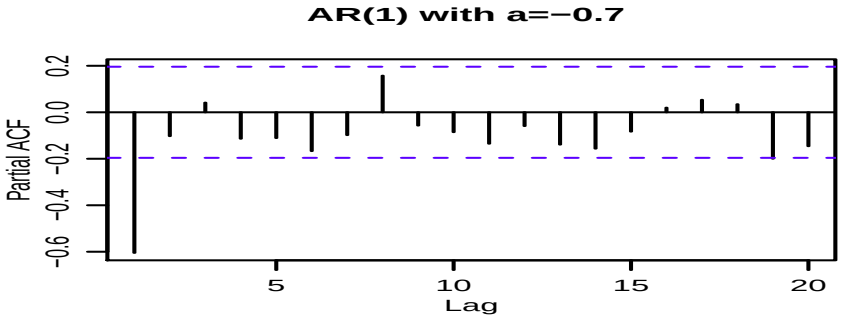
$$\sqrt{T}\,\widehat{\pi}(k) \xrightarrow{D} N\big(0,\ 1\big), \quad \text{for any } k > p.$$

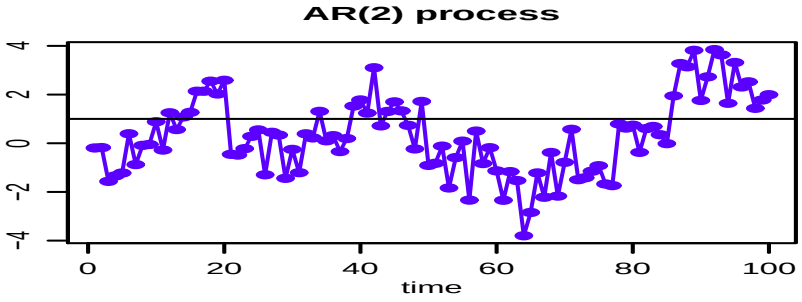
In sample PACF plots, we superimpose two confidence bounds $\pm 1.96/\sqrt{T}$.

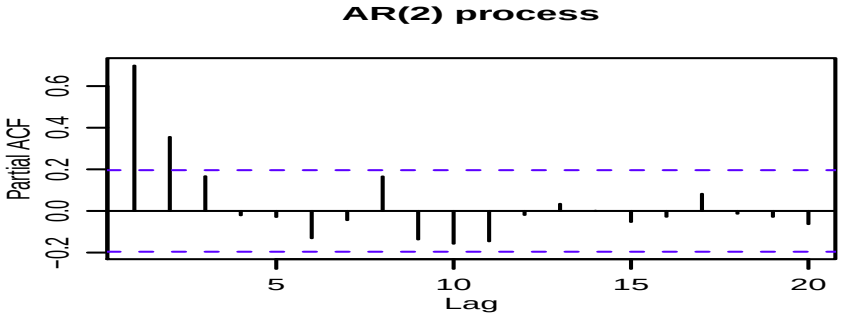

AR(1) with a=0.7
time

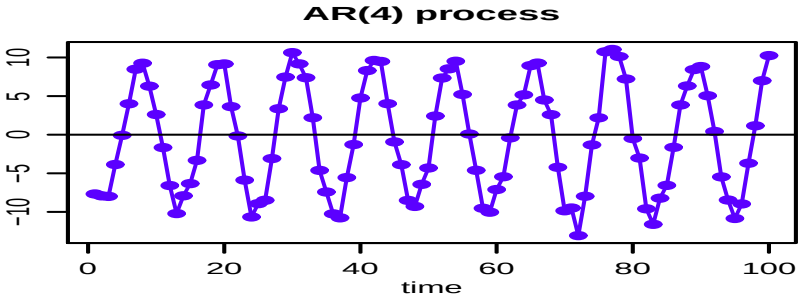

AR(1) with a=0.7
Partial ACF
Lag

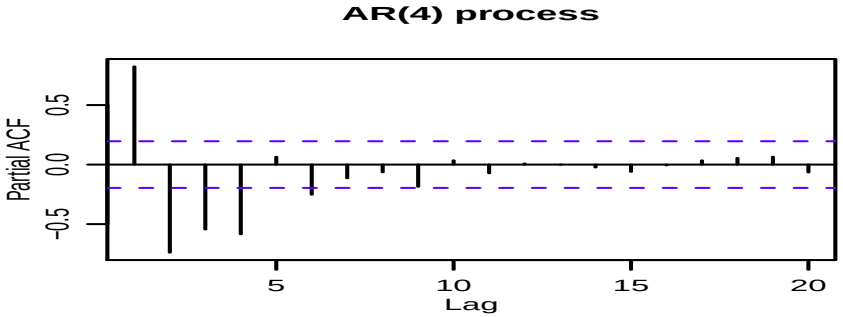

AR(1) with a=−0.7
time


AR(1) with a=−0.7
Partial ACF
Lag


AR(2) process
time


AR(2) process
Partial ACF
Lag


AR(4) process
time


AR(4) process
Partial ACF
Lag

3. Determine ARMA orders using EACF (Tsay and Tiao 1984, 1985)

Basic idea. (i) For a stationary ARMA(p, q) with mean 0, let

$$Z_t \equiv X_t - b_1 X_{t-1} - \cdots - b_p X_{t-p} \sim \text{MA}(q).$$

Hence the ACF of "residual" Z_t cuts off at q.

(ii) Since we do not know the AR order p, we try in the order of $p = 1, 2, ...$

(i) We consider stationary (causual) ARMA(1, q) model first:

$$X_t - bX_{t-1} = \varepsilon_t + a_1\varepsilon_{t-1} + \cdots + a_q\varepsilon_{t-q}, \quad \varepsilon_t \sim \text{WN}(0, \sigma_a^2).$$

The corresponding Yule-Walker equation is

$$\rho(k) - b\rho(k-1) = 0, \quad \text{for all } k > q.$$

Let $b^{(k)} = \rho(k+1)/\rho(k)$, and $Z_t^{(k)} = X_t - b^{(k)}X_{t-1}$, $t = 1, 2, \cdots$.

Then $b^{(q)} = b^{(q+1)} = \cdots = b$ and $b^{(k)} \neq b$ for $1 \leq k < q$. Hence

$$Z_t^{(q)} = Z_t^{(q+1)} = \cdots = \varepsilon_t + a_1\varepsilon_{t-1} + \cdots + a_q\varepsilon_{t-q} \sim \text{MA}(q),$$

but $Z_t^{(k)} \not\sim \text{MA}(q)$ for all $k < q$. Hence

the MA order is the smallest non-negative integer q
such that $Z_t^{(q+j)} \sim \text{MA}(q)$ for all $j \geq 0$.

Definition. The 1st EACF of Z_t is defined as for $k = 1, 2, \cdots$,

$$\rho(k; 1) = \text{ACF } of \ \{Z_t^{(k)}, \ t = 1, 2, \cdots\} \text{ at lag } k.$$

Remark. (i) For ARMA(1,1), $Z_t^{(k)} \sim$ MA(1) for $k \geq 1$, $\rho(1;1) \neq 0$, and $\rho(k;1) = 0$ for all $k > 1$.

(ii) For ARMA(1,2), $Z_t^{(k)} \sim$ MA(2) for $k \geq 2$, $\rho(2;1) \neq 0$, and $\rho(k;1) = 0$ for all $k > 2$.

(iii) For ARMA(1,q), $Z_t^{(k)} \sim$ MA(q) for $k \geq q$, $\rho(q;1) \neq 0$, and $\rho(k;1) = 0$ for all $k > q$.

The 1st EACF cuts off at q for ARMA$(1,q)$,
i.e. $\rho(k;1) = 0$, $\forall k > q$.

(ii) Consider stationary (causal) ARMA(p,q) model:

$$X_t - b_1 X_{t-1} - \cdots - b_p X_{t-p} = \varepsilon_t + a_1 \varepsilon_{t-1} + \cdots + a_q \varepsilon_{t-q},$$

where $\varepsilon_t \text{WN} i.i.d.(0, \sigma_a^2)$. The Yule-Walker equation is

$$\rho(k) - b_1 \rho(k-1) - \cdots - b_p \rho(k-p) = 0, \quad \text{for all } k > q.$$

Let $\mathbf{X}_{t-1} = (X_{t-1}, \cdots, X_{t-p})'$, $\mathbf{b} = (b_1, \cdots, b_p)'$, and

$$\boldsymbol{\rho}_k = \begin{pmatrix} \rho(k) \\ \rho(k+1) \\ \vdots \\ \rho(k+p-1) \end{pmatrix}, \quad \boldsymbol{\Sigma}_{k-1} = \begin{pmatrix} \rho(k-1) & \rho(k-2) & \cdots & \rho(k-p) \\ \rho(k) & \rho(k-1) & \cdots & \rho(k-p+1) \\ & \cdots & \cdots & \\ \rho(k+p-2) & \rho(k+p-3) & \cdots & \rho(k-1) \end{pmatrix},$$

By the Yule-Walker equation,

$$\boldsymbol{\rho}_k - \boldsymbol{\Sigma}_{k-1} \mathbf{b} = 0, \quad \text{namely} \quad \mathbf{b} = \boldsymbol{\Sigma}_{k-1}^{-1} \boldsymbol{\rho}_k \quad \forall\, k > q.$$

Let $\mathbf{b}^{(k)} = \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\rho}_{k+1}$ and $Z_t^{(k)} = X_t - \mathbf{X}_{t-1}' \mathbf{b}^{(k)}$. Then

$$\mathbf{b}^{(q)} = \mathbf{b}^{(q+1)} = \cdots = \mathbf{b}, \quad \text{and} \quad \mathbf{b}^{(k)} \neq \mathbf{b} \text{ for } 1 \leq k < q.$$

Hence

$$Z_t^{(q)} = Z_t^{(q+1)} = \cdots = \varepsilon_t + a_1\varepsilon_{t-1} + \cdots + a_q\varepsilon_{t-q} \sim \text{MA}(q),$$

but $Z_t^{(k)} \not\sim \text{MA}(q)$ for all $k < q$. Hence

the MA order is the smallest non-negative integer q
such that $Z_t^{(q+j)} \sim \text{MA}(q)$ for all $j \geq 0$.

Definition. The p-th EACF of Z_t is defined as for $k = 1, 2, \cdots,$

$$\rho(k; p) = \text{ACF of } \{Z_t^{(k)}, t = 1, 2, \cdots\} \text{ at lag } k.$$

For an ARMA(p, q) process, $\rho(k; p) = 0$ for all $k > q$.

One might expect that the above property could be extended to $\rho(k;\ell) = 0$ for all $k > q$ when $\ell > p$, since an ARMA(p, q) model may be viewed as an ARMA(ℓ, q) model with the last $\ell - p$ autoregressive coefficients equal to 0.

Unfortunately, or fortunately, this is not the case, due to the fact that the Yule-Walker equation used to estimate AR coefficients is under determined, and its sample version does not necessarily produce a consistent estimate. See Example 2.8 of Yao and Fan (2015) for a simple illustration with AR(1).

Instead the ℓ-th EACF cuts off at $q+(\ell-p)$ for ARMA(p, q) models, for $\ell > p$.

In practice we replace ACF $\rho(\cdot)$ by the sample ACF $\widehat{\rho}(\cdot)$, leading to sample EACF $\widehat{\rho}(k;\ell)$.

		MA				
		0	1	2	3	$\cdots$
	0	$\widehat{\rho}(1)$	$\widehat{\rho}(2)$	$\widehat{\rho}(3)$	$\widehat{\rho}(4)$	$\cdots$
	1	$\widehat{\rho}(1;1)$	$\widehat{\rho}(2;1)$	$\widehat{\rho}(3;1)$	$\widehat{\rho}(4;1)$	$\cdots$
AR	2	$\widehat{\rho}(1;2)$	$\widehat{\rho}(2;2)$	$\widehat{\rho}(3;2)$	$\widehat{\rho}(4;2)$	$\cdots$
	3	$\widehat{\rho}(1;3)$	$\widehat{\rho}(2;3)$	$\widehat{\rho}(3;3)$	$\widehat{\rho}(4;3)$	$\cdots$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

The *R*-function `eacf` in package TSA computes sample EACF automatically and prints out a coded table in the above form with $\widehat{\rho}(i,j)$ replaced by symbol "`x`" if $|\widehat{\rho}(i,j)| > 1.96/\sqrt{T-i-j}$, and "`o`" otherwise.

The ideal pattern of the sample EACF for ARMA(p, q)

	0	1	$\cdots$	q	$q+1$	$q+2$	$q+3$	$\cdots$
0	x	x	$\cdots$	x	x	x	x	$\cdots$
1	x	x	$\cdots$	x	x	x	x	$\cdots$
			$\cdots$		$\cdots$			
$p-1$	x	x	$\cdots$	x	x	x	x	$\cdots$
p	x	x	$\cdots$	o	o	o	o	$\cdots$
$p+1$	x	x	$\cdots$	x	o	o	o	$\cdots$
$p+2$	x	x	$\cdots$	x	x	o	o	$\cdots$
$p+3$	x	x	$\cdots$	x	x	x	o	$\cdots$
			$\cdots$		$\cdots$			

```
> install.packages("TSA")
> library(TSA)
> x <- arima.sim(n=100, list(ar=0.5, ma=0.279))+1
> eacf(x)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x o o o o o o o x x  x  o  o
1 x o o o o o o o o o o  o  o  o
2 o x o o o o o o o o o  o  o  o
3 o x x o o o o o o o o  o  o  o
4 x o o x x o o o o o o  o  o  o
5 o x o o o o o o o o o  o  o  o
6 o x o o o o o o o o o  o  o  o
7 x x o o o o o o o o o  o  o  o
```

Nonstationary Processes

1. Random Walks: $X_t = X_{t-1} + \varepsilon_t$ for $t \geq 1$, where $\varepsilon_t \sim \text{WN}(0, \sigma^2)$.

Assume $X_0 = c$, then

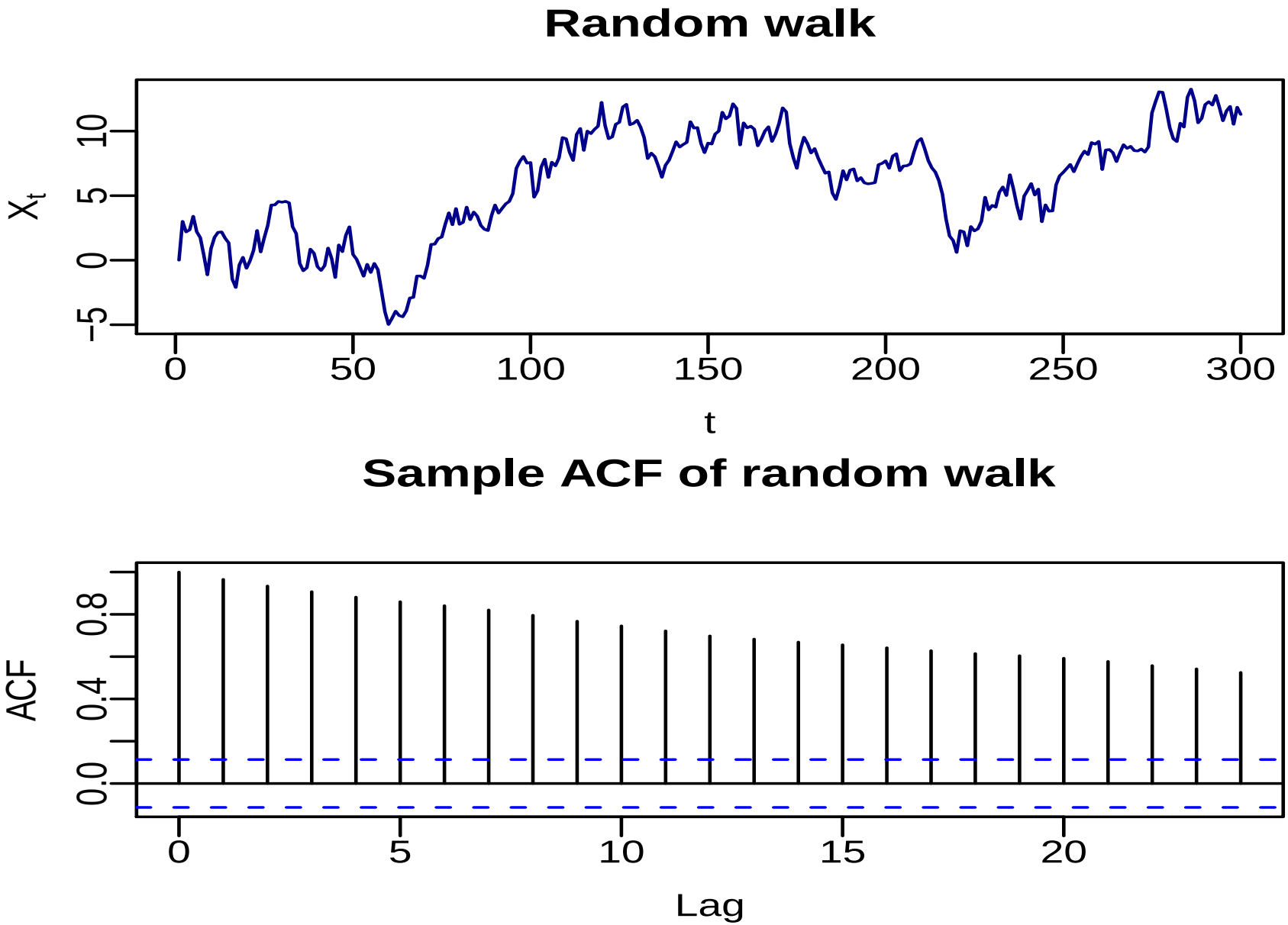
$$\text{Var}(X_t) = \text{Var}(X_{t-1} + \varepsilon_t) = \text{Var}(\varepsilon_1 + \cdots + \varepsilon_t) = t\sigma^2,$$

and for $k \geq 0$,

$$\text{Cov}(X_t, X_{t+k}) = \text{Var}(X_t) = t\sigma^2, \qquad \text{Corr}(X_t, X_{t+k}) = \{t/(t+k)\}^{1/2}$$

which is close to 1 for small k and large t.

The sample ACF a random walk is close to 1 at small time lags, and decays slowly when the lag increase.


Random walk
X_t
10
5
0
−5
0
50
100
150
200
250
300
t
Sample ACF of random walk
ACF
0.8
0.4
0.0
0
5
10
15
20
Lag

2. ARIMA(p,d,q): $X_t \sim$ ARIMA(p,d,q) if $\nabla^d X_t$ is stationary ARMA(p,q).

Consider ARIMA(0,1,1):

$$X_t - X_{t-1} = \varepsilon_t - \theta\varepsilon_{t-1}, \quad \varepsilon_t \sim \text{WN}(0,\sigma^2),$$

where $|\theta| < 1$.

$$\varepsilon_t = (1-\theta B)^{-1}(1-B)X_t = \{1-(1-\theta)(B+\theta B^2+\theta^2 B^3+\cdots)\}X_t.$$

Hence,

$$X_t = (1-\theta)(X_{t-1}+\theta X_{t-2}+\theta^2 X_{t-3}+\cdots)+\varepsilon_t.$$

This is in the form of an AR(∞) with coefficients $(1-\theta)\theta^k$ decaying exponentially.

Hence at time t, the "best" predictor for the future value X_{t+1} is

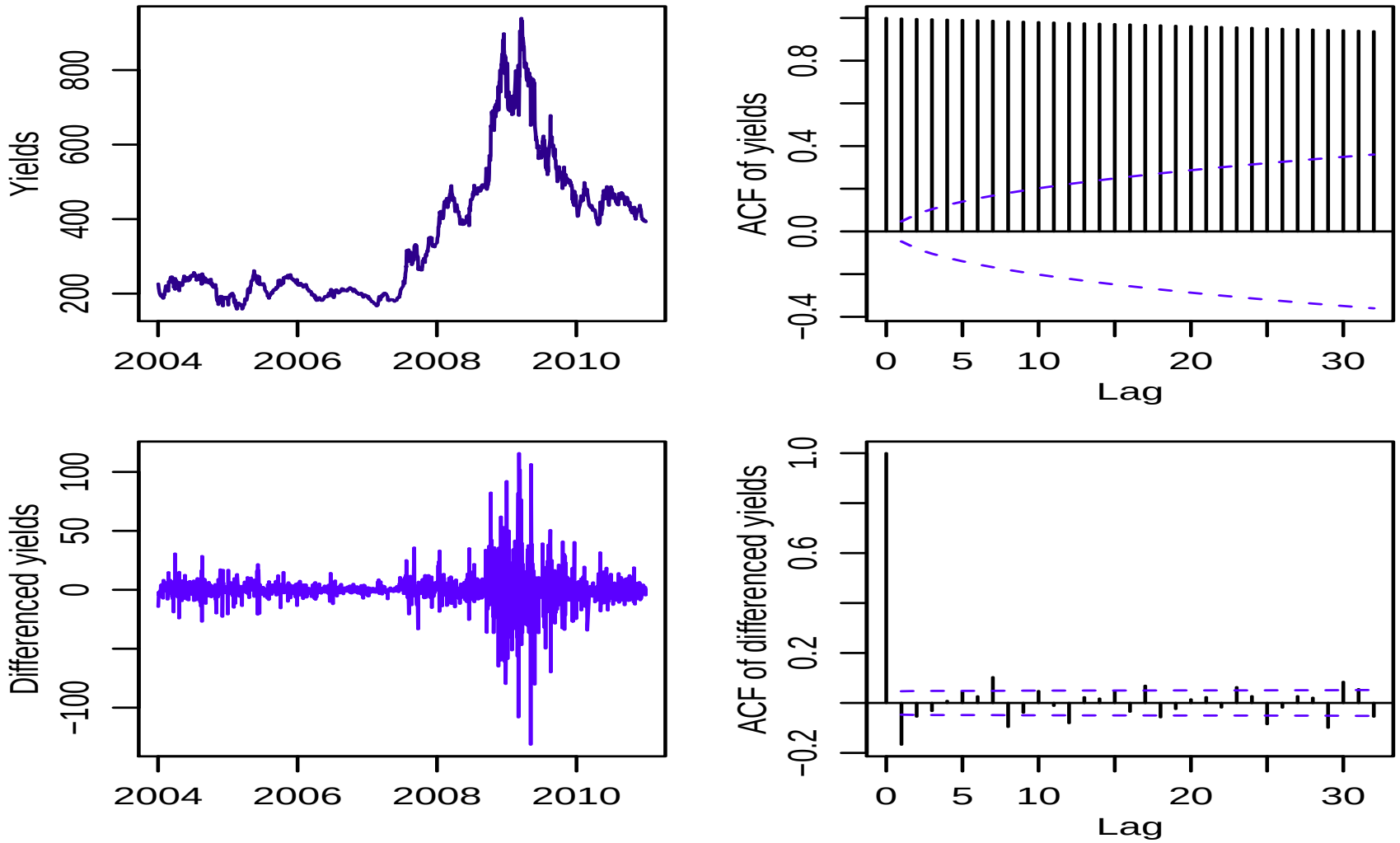
$$\widehat{X}_{t+1} = (1-\theta)(X_t + \theta X_{t-1} + \theta^2 X_{t-2} + \theta_3^3 X_{t-3} + \cdots), \qquad (3)$$

i.e. $\widehat{X}_{t+1}$ is the weighted average of the past values $X_t, X_{t-1}, \cdots$ with larger weights on more recent values.

Note $\sum_{j\geq 0}(1-\theta)\theta^j = 1$, i.e. all the weights sum up to 1.

(3) is the celebrated exponential smoothing.

The daily yields of a basket HY bonds in 2004 – 2010 and their differences: time series and sample ACF plots



Guidelines for model identification based on sample ACF, PACF and EACF.

1. If the sample ACF $\widehat{\rho}(k)$ are close to 1 for all small k and decay slowly, difference the data first.

2. If the sample ACF $\widehat{\rho}(k)$ is bounded by the bounds (2) for any $k > q$, fit an MA(q) model to the data.

3. If the sample PACF $\widehat{\pi}(k)$ is bounded by $\pm 1.96/\sqrt{T}$ for any $k > p$, fit an AR(p) model to the data.

4. If the sample PACF $\widehat{\rho}(q+j; p+i)$ is bounded by $\pm 1.96/\sqrt{T-i-j}$ for any $i \geq 0$ and $j > i$, fit an ARMA(p, q) model to the data.

Examples on model identification based on ACF, PACF and EACF.

```
> library("TSA") # To load all objects in package TSA to
                 # the current session
> help(package="TSA") # To access the info on objects in TSA
> ?eacf  # To bring up the help menu for eacf
> data(flow, package="TSA") # To load monthly river flow for
                       # Iowa River to the current session
> data(gold, package="TSA") # To load gold price series
```

We always let $\varepsilon_t \sim_{i.i.d.} N(0,1)$.

Example 1. MA(1): $X_t = \varepsilon_t - 0.7\varepsilon_{t-1}$.

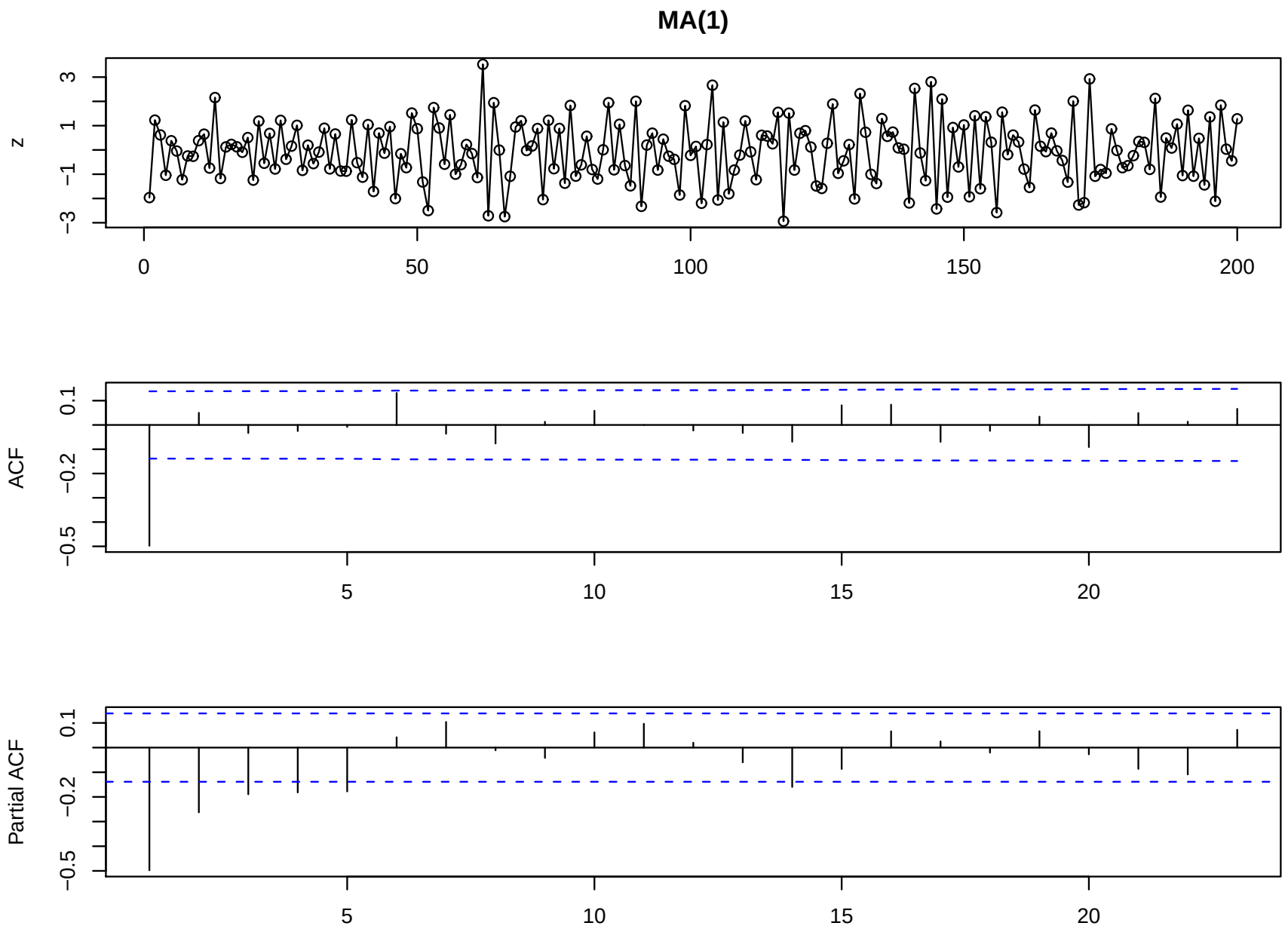
```
> x <- arima.sim(n=200, list(ma=-0.7))
> plot(x, type="o", main="MA(1)")
> acf(x, ci.type="ma")
> pacf(x)
> eacf(x)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x o o o o o o o o o o  o  o  o
1 x x o o o o o o o o o  o  o  o
2 x x x o o o o o o o o  o  o  o
3 x o o x o o o o o o o  o  o  o
4 x x o o x o o o o o o  o  o  o
5 x o o x o o o o o o o  o  o  o
6 o o x x o o o o o o o  o  o  o
7 x o x o o o o o o o o  o  o  o
```

ACF: MA(1)

PACF: AR(4)

EACF: MA(1)

Presentation of EACF:
$\widehat{\rho}_k(j)$ is marked as 'x' if $|\widehat{\rho}_k(j)| > 1.96/\sqrt{n-k-j}$, and 'o' otherwise.


MA(1)
z
3
1
−1
−3
0
50
100
150
200
ACF
0.1
−0.2
−0.5
5
10
15
20
Partial ACF
0.1
−0.2
−0.5
5
10
15
20

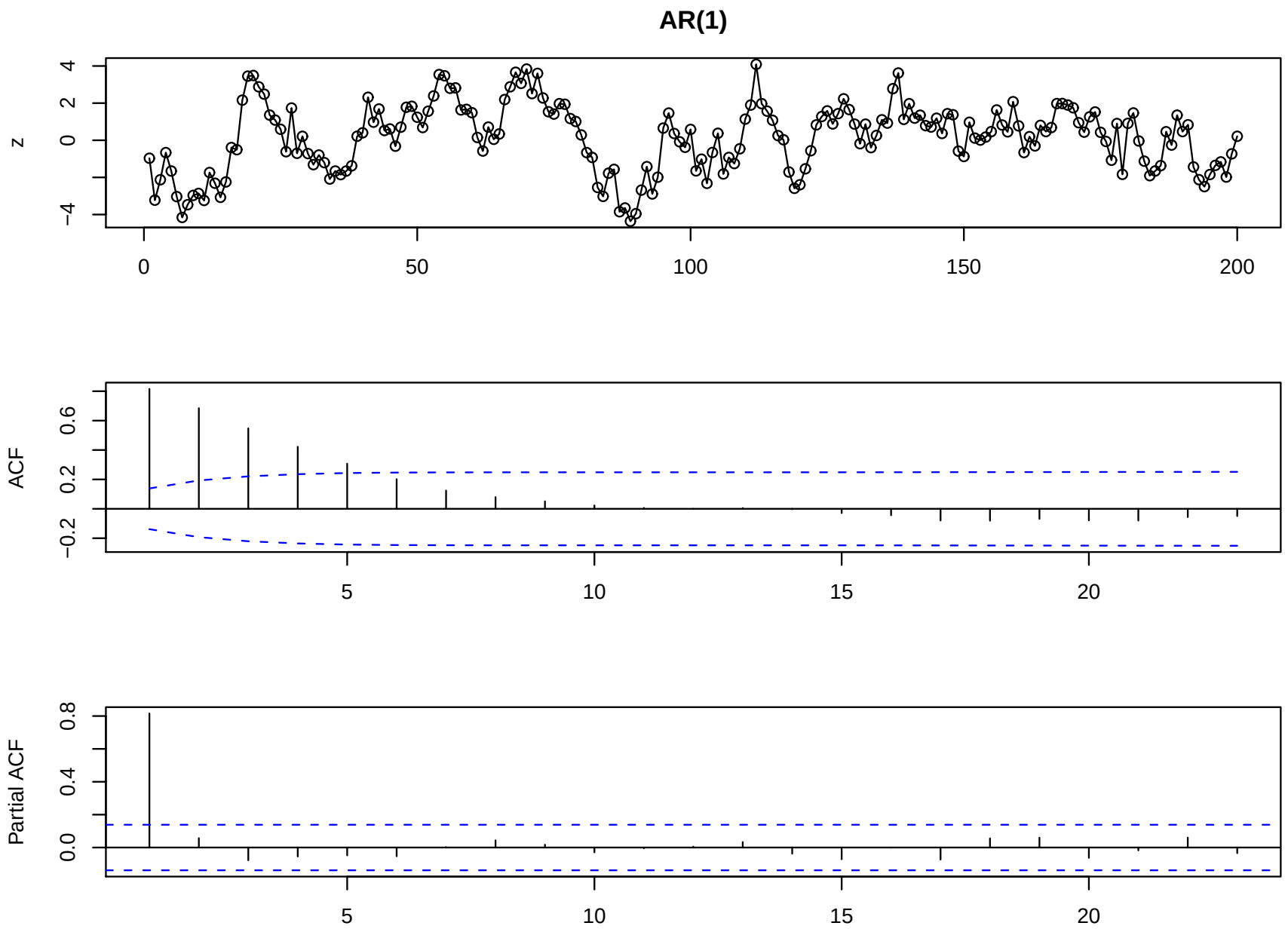
Example 2. AR(1): $X_t = 0.8X_{t-1} + \varepsilon_t$

```
> x<- arima.sim(n=200, list(ar=0.8))
> plot(x, type='o', main="AR(1)")
> acf(x, ci.type="ma")
> pacf(x)
> eacf(x)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x x x x x x o o o o  o  o  o
1 o o o o o o o o o o o  o  o  o
2 x o o o o o o o o o o  o  o  o
3 x o o o o o o o o o o  o  o  o
4 x x x o o o o o o o o  o  o  o
5 x x x x o o o o o o o  o  o  o
6 x x x o x o o o o o x  o  o  o
7 x x x o o o o o o o x  o  o  o
```

ACF: MA(5)

PACF: AR(1)

EACF: AR(1)


AR(1)
z
4
2
0
−4
0
50
100
150
200
ACF
0.6
0.2
−0.2
5
10
15
20
Partial ACF
0.8
0.4
0.0
5
10
15
20

Example 3. ARMA(1, 2)

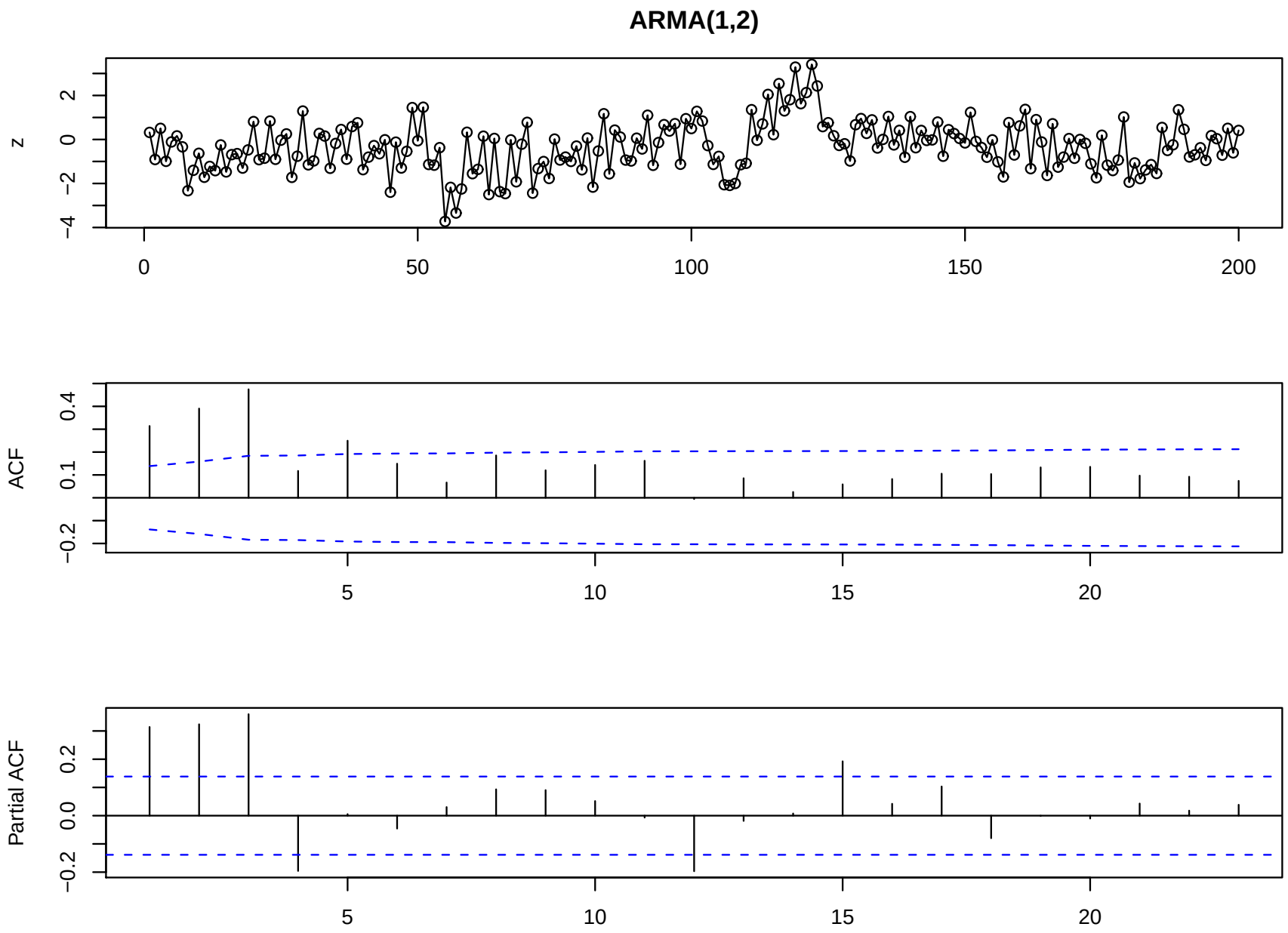
$X_t = 0.7X_{t-1} + \varepsilon_t - 0.6\varepsilon_{t-1} + 0.5\varepsilon_{t-2}$

```
> x <- arima.sim(n=200, list(ar=0.7, ma=c(-0.6, 0.5)))
> plot(x, type='o', main="ARMA(1,2)")
> acf(x, ci.type="ma")
> pacf(x)
> eacf(x)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x x x x o o o o o o  o  o  o
1 x x o o o o o o o o o  o  x  o
2 x x x o o o o o o o o  o  o  o
3 x x o o o o o o o o x  o  o  o
4 x o x x o o o o o o o  o  o  o
5 x o x o o o o o o x x  o  o  o
6 x o x o o o o o x o o  o  o  o
7 x o o x o o o o x o o  o  o  o
```

ACF: MA(5)

PACF: AR(4)?

EACF: MA(5), ARMA(1,2)?


ARMA(1,2)
z
−4
−2
0
2
0
50
100
150
200
ACF
−0.2
0.1
0.4
5
10
15
20
Partial ACF
−0.2
0.0
0.2
5
10
15
20

Example 4. ARIMA(0,1,1)

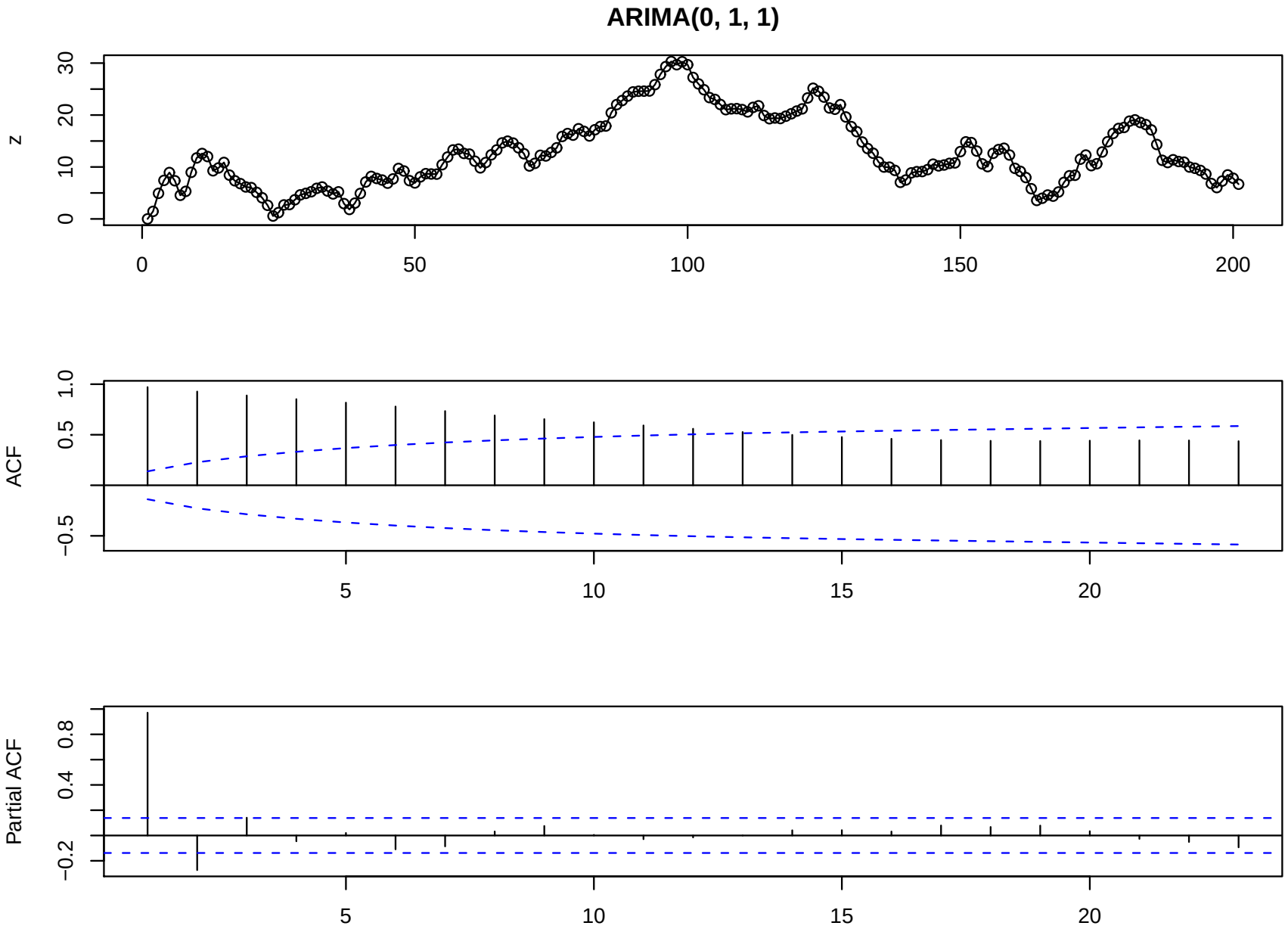
$X_t = X_{t-1} + \varepsilon_t - 0.8\varepsilon_{t-1}$

```
> x <- arima.sim(n=200, list(order=c(0,1,1), ma=0.8))
> plot(x, type='o', main="ARIMA(0, 1, 1)")
> acf(x, ci.type="ma")
> pacf(x)
> eacf(x)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x x x x x x x x x x  x  x  x
1 x o o o o o o o o o o  o  o  o
2 x x o o o o o o o o o  o  o  o
3 x x o o o o o o o o o  o  o  o
4 x x o o o o o o o o o  o  o  o
5 x o x o o o o o o o o  o  o  o
6 x x x x o o o o o o o  o  o  o
7 x x x x o o o o o o o  o  o  o
```

ACF close to 1: differencing

PACF: AR(2)???

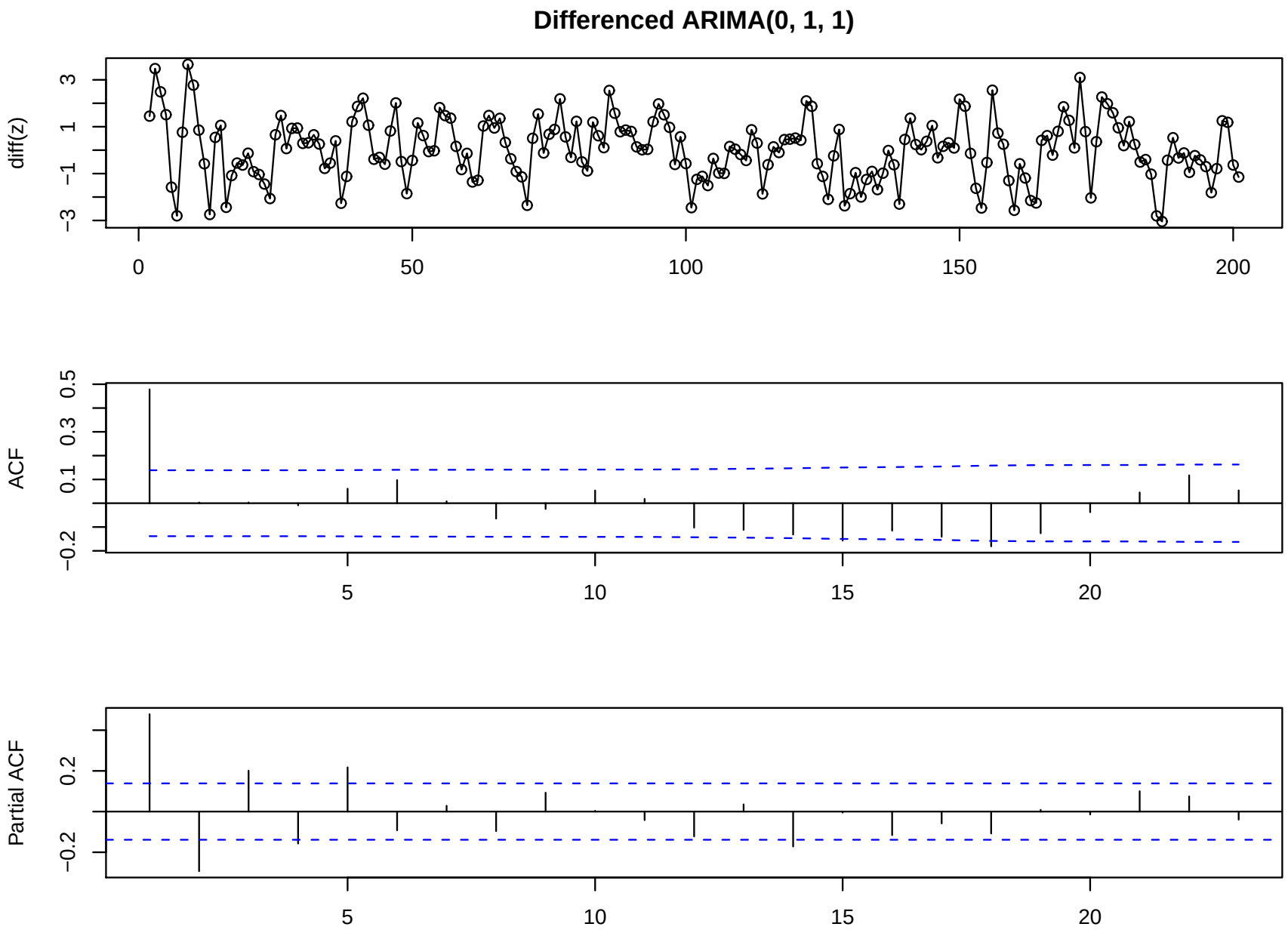
EACF: ARMA(1, 1) – interesting!


ARIMA(0, 1, 1)
z
0
10
20
30
0
50
100
150
200
ACF
1.0
0.5
−0.5
5
10
15
20
Partial ACF
0.8
0.4
−0.2
5
10
15
20

For the differenced data

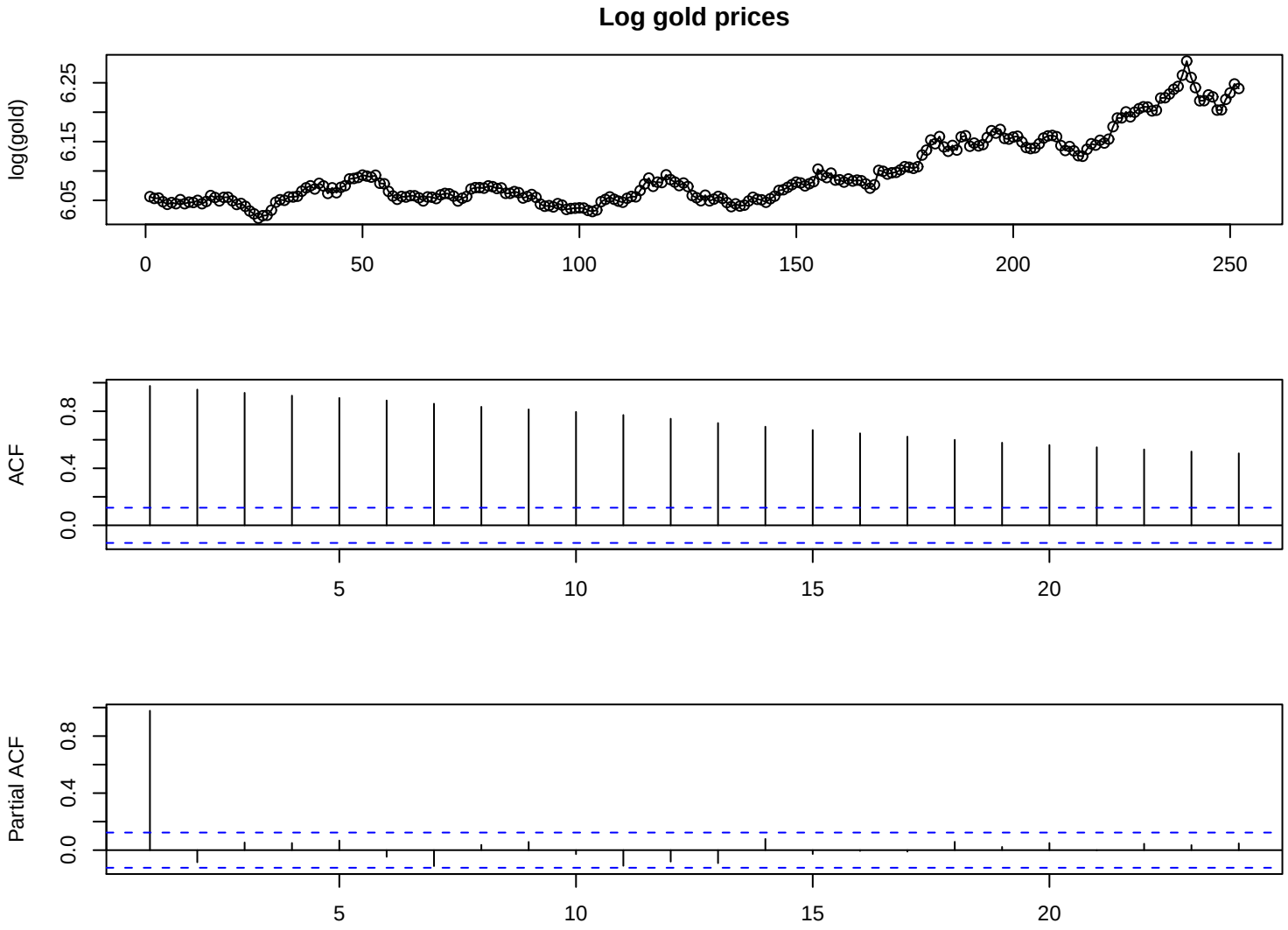
```
plot(diff(x), type="o", main="Differenced ARIMA(0, 1, 1)")
acf(diff(x), ci.type="ma")
pacf(diff(x))
eacf(diff(z))
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x o o o o o o o o o o  o  o  o
1 x x o o o o o o o o o  o  o  o
2 x o o o o o o o o o o  o  o  o
3 x o o o o o o o o o o  o  o  o
4 x x x x x o o o o o o  o  o  o
5 x o x o o o o o o o o  o  o  o
6 x o x o o o o o o o o  o  o  o
7 x x x o o o o o o o o  o  o  o
```

ACF: MA(1) PACF: AR(5) EACF: MA(1)


Differenced ARIMA(0, 1, 1)
diff(z)
−3
−1
1
3
0
50
100
150
200
ACF
−0.2
0.1
0.3
0.5
5
10
15
20
Partial ACF
−0.2
0.2
5
10
15
20

Example 5. `gold`: Gold prices

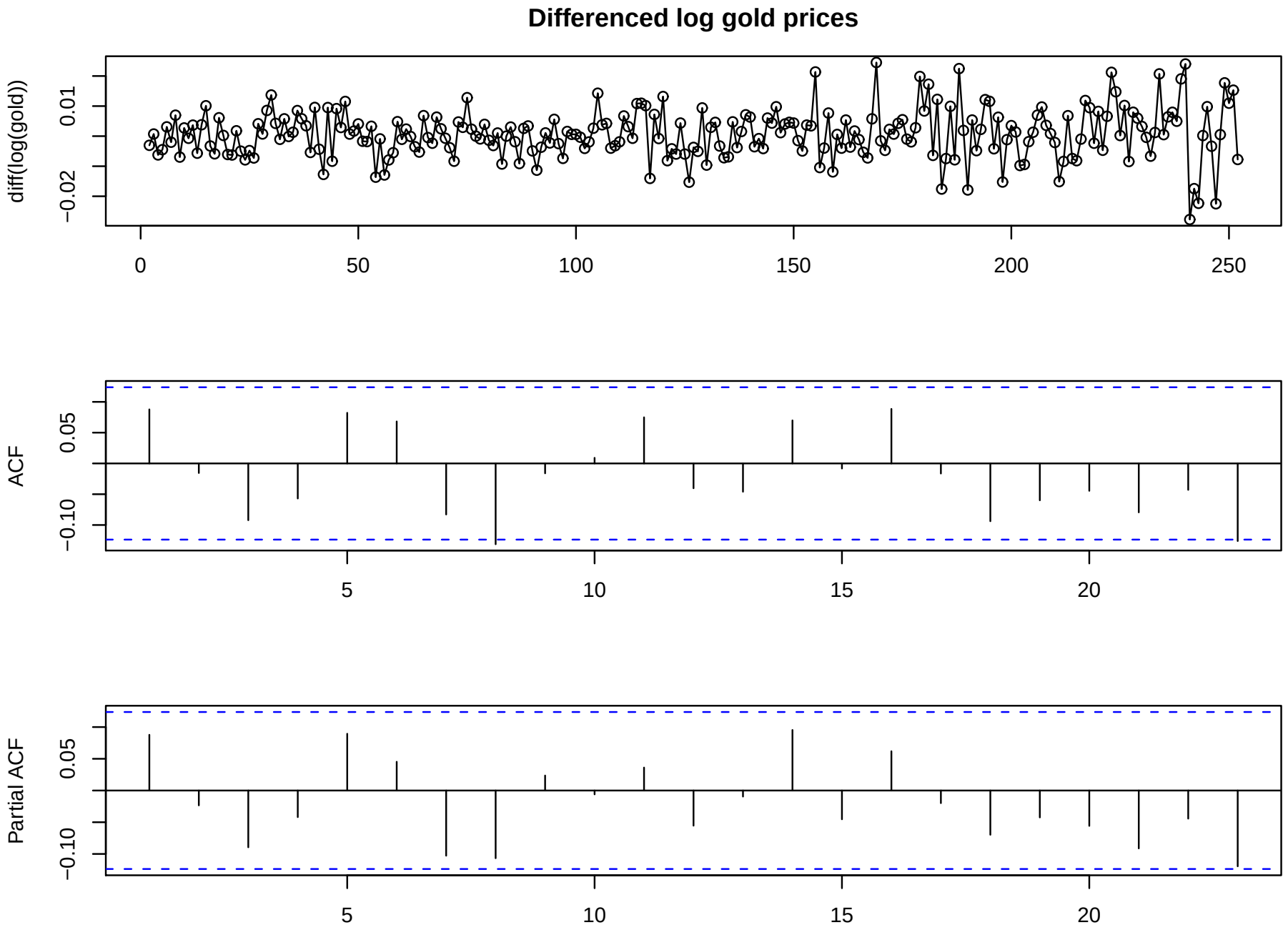
```
plot(log(gold), type='l', main="Log gold prices")
> acf(log(gold))
> pacf(log(gold))
> eacf(log(gold))
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x x x x x x x x x x  x  x  x
1 o o o o o o o x o o o  o  o  o
2 x o o o o o o o o o o  o  o  o
3 x x o o o o o o o o o  o  o  o
4 x x x o o o o o o o o  o  o  o
5 x x o x o o o o o o o  o  o  o
6 x x x o o o o o o o o  o  o  o
7 x x o o o o o o o o o  o  o  o
```



ACF are close to 1, so we look into the differenced data

```
> plot(diff(log(gold)), type='l', main="Differenced log gold prices")
> acf(diff(log(gold)))
> pacf(diff(log(gold)))
> eacf(diff(log(gold)))
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 o o o o o o o x o o o  o  o  o
1 x o o o o o o o o o o  o  o  o
2 x x o o o o o o o o o  o  o  o
3 x x x o o o o o o o o  o  o  o
4 x x o x o o o o o o o  o  o  o
5 x x o o o o o o o o o  o  o  o
6 x x x o o o o o o o o  o  o  o
7 x x x o o o x o o o o  o  o  o
```

Differenced log prices are white noise!


Differenced log gold prices
diff(log(gold))
−0.02
0.01
0
50
100
150
200
250
ACF
0.05
−0.10
5
10
15
20
Partial ACF
0.05
−0.10
5
10
15
20

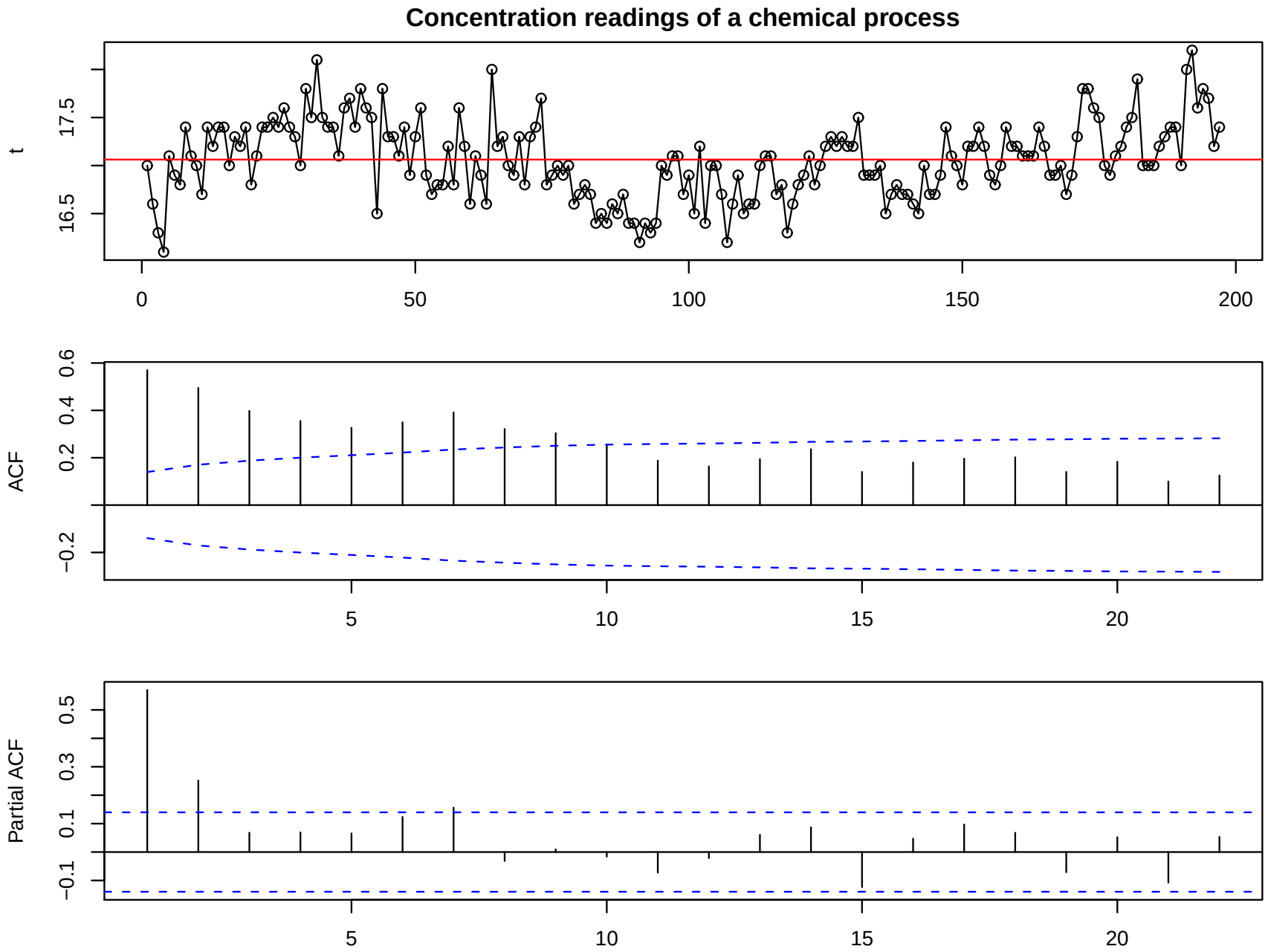
Example 6. Concentration readings of a chemical process – a classical example from Box and Jenkins' book.
Download the data set "cc.dat" from `stats.lse.ac.uk/q.yao/newTS/data/`, and place it at your `R` working folder/directory.

```
> t <- scan('cc.dat')
> plot(t, type='o', main='Concentration readings of a chemical process')
> abline(mean(t), 0, col='red')
> acf(t, ci.type="ma");  pacf(t);  eacf(t)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x x x x x x x x x x  x  x  x
1 x o o o o o x o o o o  o  o  x
2 x x o o o o x o o o o  o  o  o
3 x o o o o o x o o o o  o  o  o
4 x o o o o o o o o o o  o  o  o
5 x x x o x o o o o o o  o  o  o
6 x o x o x o o o o o o  o  o  o
```

ACF: MA(8) or MA(9)

PACF: AR(2) or AR(7)?

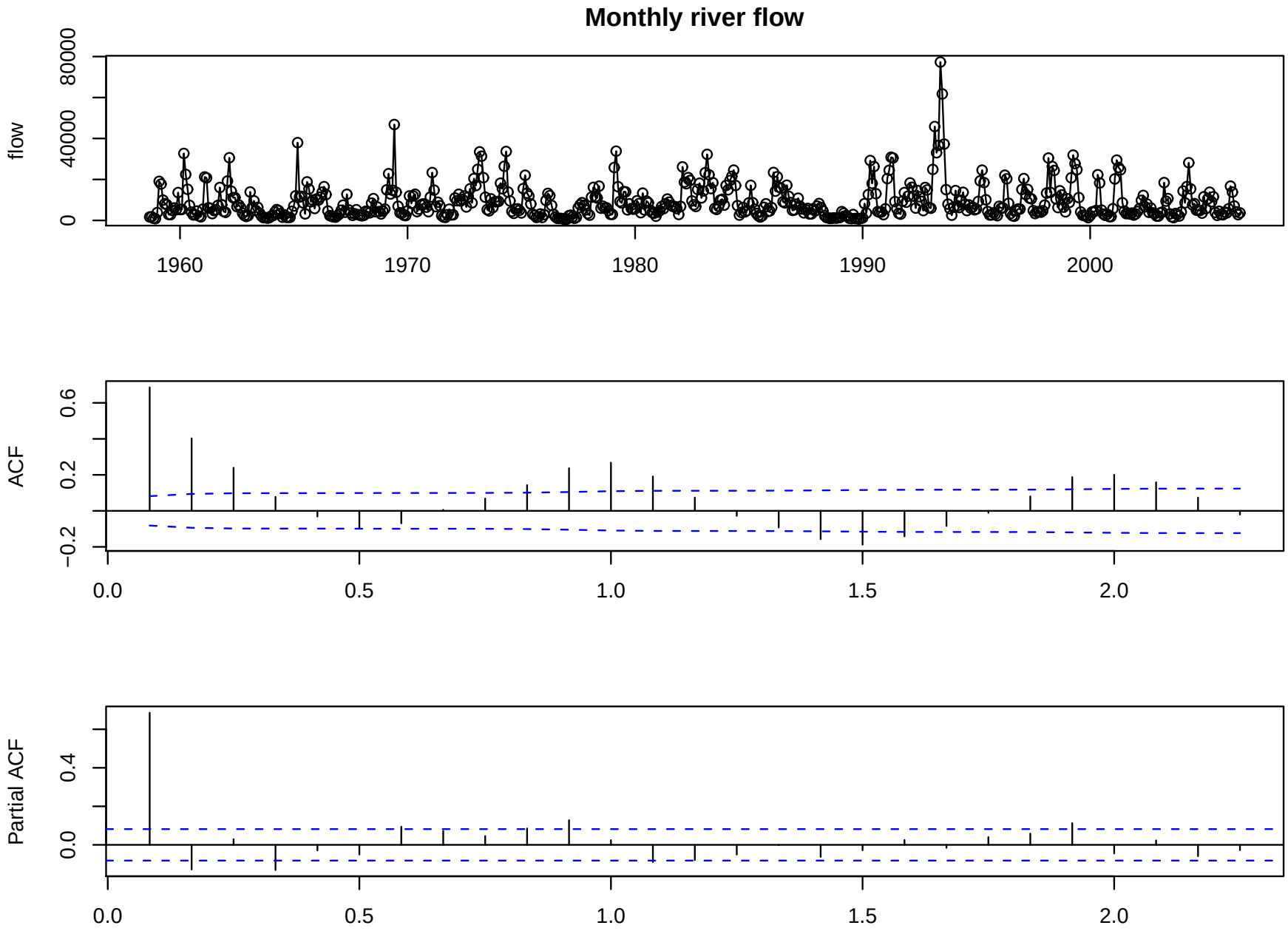
EACF: ARMA(2,2) or ARMA(1,1)


Concentration readings of a chemical process
t
16.5
17.5
0
50
100
150
200
ACF
−0.2
0.2
0.4
0.6
5
10
15
20
Partial ACF
−0.1
0.1
0.3
0.5
5
10
15
20

Example 7. `flow`: monthly river flow for Iowa River

```
> plot(flow, type='o', main="Monthly river flow")
> acf(flow, ci.type="ma");  pacf(flow);  eacf(flow)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x x o o x o o o x x  x  x  o
1 x o x x o x o o o o x  x  x  x
2 x o x o x o o o o o o  x  o  o
3 x x x o x o o o o o o  x  o  o
4 x x x o x o o o o o o  o  o  o
5 x x x x o x o o o o o  o  o  o
6 x x x x x o o o o o o  o  o  o
7 x o x o x x o o o o o  o  o  o
```

No simple models are identified. Actually ACF indicates strong periodicity. ARIMA framework does not appear to be adequate for this data.


Monthly river flow
flow
0
40000
80000
1960
1970
1980
1990
2000
ACF
−0.2
0.2
0.6
0.0
0.5
1.0
1.5
2.0
Partial ACF
0.0
0.4
0.0
0.5
1.0
1.5
2.0