Time Series Analysis

Time Series - An Introduction

- A time series is a sequence of observations ordered in time; observations are numbers (e.g. measurements).

- Time series analysis comprises methods that attempt to:
  - understand the underlying context of the data (where did they come from? what generated them?);
  - make forecasts (predictions).

Definitions/Setting

- A stochastic process is a collection of random variables $\{Y_t : t \in \mathbf{T}\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$.

- In time series modelling, a sequence of observations is considered as one realisation of an unknown stochastic process:

  1. can we infer properties of this process?

  2. can we predict its future behaviour?

- By time series we shall mean both the sequence of observations and the process of which it is a realization (language abuse).

- We will only consider discrete time series: observations $(y_1, \ldots, y_N)$ of a variable at different times ($y_i = y(t_i)$, say).

Setting (cont.)

- We will only deal with time series observed at regular time points (days, months etc.).

- We focus on pure univariate time series models: a single time series $(y_1, \ldots, y_N)$ is modelled in terms of its own values and their order in time. No external factors are considered.

- Modelling of time series which:

  - are measured at irregular time points, or

  - are made up of several observations at each time point (multivariate data), or

  - involve explanatory variables $\mathbf{x_t}$ measured at each time point,

  is based upon the ideas presented here.

Work plan

- We provide an overview of pure univariate time series models:
  - ARMA (‘Box-Jenkins’) models;
  - ARIMA models;
  - GARCH models.

- Models will be implemented in the public domain general purpose statistical language python.

References

1. Anderson, O. D. *Time series analysis and forecasting The Box-Jenkins approach*. Butterworths, London-Boston, Mass., 1976.

2. Box, George E. P. and Jenkins, Gwilym M. *Time series analysis: forecasting and control*. Holden-Day, San Francisco, Calif. 1976

3. Brockwell, Peter J. and Davis, Richard A. *Time series: theory and methods*, Second Edition. Springer Series in Statistics, Springer-Verlag. 1991.

4. Jonathan D. Cryer *Time series Analysis*. PWS-KENT Publishing Company, Boston. 1986.

Statistical versus Time series modelling

Problem: **Given a time series $(y_1, y_2, \ldots, y_N)$: (i) determine temporal structure and patterns; (ii) forecast non-observed values.**
Approach: **Construct a mathematical model for the data.**

- In statistical modelling it is typically assumed that the observations $(y_1, \ldots, y_N)$ are a sample from a sequence of independent random variables. Then
  - there is no covariance (or correlation) structure between the observations; in other words,
  - the joint probability distribution for the data is just the product of the univariate probability distributions for each observation;
  - we are mostly concerned with estimation of the mean behaviour μ_i and the variance σ_i^2 of the error about the mean, errors being unrelated to each other.

Statistical vs. Time series modelling (cont.)

- However, for a time series we cannot assume that the observations $(y_1, y_2, \ldots, y_N)$ are independent: the data will be serially correlated or auto-correlated, rather than independent.

- Since we want to understand/predict the data, we need to explain/use the correlation structure between observations.

- Hence, we need stochastic processes with a correlation structure over time in their random component.

- Thus we need to directly consider the joint multivariate distribution for the data, $p(y_1, \ldots, y_N)$, rather than just each marginal distribution $p(y_t)$.

Time series modelling

- If one could assume joint normality of $(y_1, \ldots, y_N)$ then the joint distribution, $p(y_1, \ldots, y_N)$, would be completely characterised by:
  - the means: $\mu = (\mu_1, \mu_2, \ldots, \mu_N)$;
  - the auto-covariance matrix $\mathbf{\Sigma}$, i.e. the $N \times N$ matrix with entries $\sigma_{ij} = \mathrm{cov}(y_i, y_j) = \mathbb{E}[(y_i - \mu_i)(y_j - \mu_j)]$.
- In practice joint normality is not an appropriate assumption for most time series (certainly not for most financial time series).
- Nevertheless, in many cases knowledge of μ and $\mathbf{\Sigma}$ will be sufficient to capture the major properties of the time series.

Time series modelling (cont.)

- Thus the focus in time series analysis reduces to understand the mean μ and the autocovariance $\mathbf{\Sigma}$ of the generating process (weakly stationary time series).

- In the applications both μ and $\mathbf{\Sigma}$ are unknown and so must be estimated from the data.

- There are N elements involved in the mean component μ and $N(N+1)/2$ distinct elements in $\mathbf{\Sigma}$: vastly too many distinct unknowns to estimate without some further restrictions.

- To reduce the number of unknowns, we have to introduce parametric structure so that the modelling becomes manageable.

Strict Stationarity

- The time series $\{Y_t : t \in \mathbb{Z}\}$ is strictly stationary if the joint distributions of $(Y_{t_1}, \ldots, Y_{t_k})$ and $(Y_{t_{1+\tau}}, \ldots, Y_{t_{k+\tau}})$ are the same for all positive integers k and all $t_1, \ldots, t_k, \tau \in \mathbb{Z}$.

- Equivalently, the time series $\{Y_t : t \in \mathbb{Z}\}$ is strictly stationary if the random vectors $(Y_1, \ldots, Y_k)$ and $(Y_{1+\tau}, Y_{2+\tau}, \ldots, Y_{k+\tau})$ have the same joint probability distribution for any time shift τ,

- Taking $k = 1$ yields that Y_t has the same distribution for all t.

- If $\mathbb{E}[|Y_t|^2] < \infty$, then $\mathbb{E}[Y_t]$ and Var(Y_t) are both constant.

- Taking $k = 2$, we find that Y_t and Y_{t+h} have the same joint distribution and hence cov(Y_t, Y_{t+h}) is the same for all h.

Weak Stationarity

- Let $\{Y_t : t \in \mathbb{Z}\}$ be a stochastic process with mean μ_t and variance $\sigma_t^2 < \infty$, for each t. Then, the autocovariance function is defined by:

$$\gamma(t,s) = \text{cov}\,(Y_t, Y_s) = \mathbb{E}[(Y_t - \mu_t)(Y_s - \mu_s)]\,.$$

- The stochastic process $\{Y_t : t \in \mathbb{Z}\}$ is weak stationary if for all $t \in \mathbb{Z}$ the following holds:

  - $\mathbb{E}\left[\,|Y_t|^2\right] < \infty, \qquad \mathbb{E}[Y_t] = m;$

  - $\gamma(r,s) = \gamma(r+t, s+t)$ for all $r, s \in \mathbb{Z}$.

- Notice that the autocovariance function of a weak stationary process is a function of only the time shift (or lag) $\tau \in \mathbb{Z}$:

$$\gamma_\tau = \gamma(\tau, 0) = \text{cov}\left(Y_{t+\tau}, Y_t\right), \qquad \text{for all} \quad t \in \mathbb{Z}.$$

In particular the variance is independent of time: $\text{Var}(Y_t) = \gamma_0$.

Autocorrelation

- Let $\{Y_t : t \in \mathbb{Z}\}$ be a stochastic process with mean μ_t and variance $\sigma_t^2 < \infty$, for each t. Then, the autocorrelation is defined by:

$$\rho(t,s) = \frac{\text{cov}\,(Y_t, Y_s)}{\sigma_t\,\sigma_s} = \frac{\gamma(t,s)}{\sigma_t\,\sigma_s}\,.$$

- If the function $\rho(t,s)$ is well-defined, its value must lie in the range $[-1, 1]$, with 1 indicating perfect correlation and -1 indicating perfect anti-correlation.

- The autocorrelation describes the correlation between the process at different points in time.

Autocorrelation Function (ACF)

- If $\{Y_t : t \in \mathbb{Z}\}$ is weak stationary then the autocorrelation depends only on the lag $\tau \in \mathbb{Z}$:

$$\rho_\tau = \frac{\text{cov}\left(Y_{t+\tau}, Y_t\right)}{\sigma_\tau\, \sigma_\tau} = \frac{\gamma_\tau}{\sigma^2}, \qquad \text{for all} \quad t \in \mathbb{Z},$$

  where $\sigma^2 = \gamma_0$ denotes the variance of the process.

- So weak stationarity (and therefore also strict stationarity) implies auto-correlations depend only on the lag τ and this relationship is referred to as the auto-correlation function (ACF) of the process.

Partial Autocorrelation Functions (PACF)

- For a weak stationary process $\{Y_t : t \in \mathbb{Z}\}$, the PACF α_k at lag k may be regarded as the correlation between Y_1 and Y_{1+k}, adjusted for the intervening observations $Y_1, Y_2, \ldots, Y_{k-1}$.

- For $k \geq 2$ the PACF is the correlation of the two residuals obtained after regressing Y_k and Y_1 on the intermediate observations $Y_2, Y_3, \ldots, Y_k$.

- The PACF at lag k is defined by $\alpha_k = \psi_{kk}$, $k \geq 1$, where ψ_{kk} is uniquely determined by:

$$\begin{pmatrix} \rho_0 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & \rho_0 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & & & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_0 \end{pmatrix} \begin{pmatrix} \psi_{k1} \\ \psi_{k2} \\ \vdots \\ \psi_{kk} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix}.$$

Stationary models

- Assuming weak stationarity, modelling a time series reduces to estimation of a constant mean $\mu = \mu_t$ and of a covariance matrix:

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{N-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{N-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{N-1} & \rho_{N-2} & \rho_{N-3} & \cdots & 1 \end{bmatrix}$$

- There are many fewer parameters in Σ $(N-1)$ than in an arbitrary, unrestricted covariance matrix.

- Still, for large N the estimation can be problematic without additional structure in Σ, to further reduce the number of parameters.

Auto-regressive Moving Average (ARMA) processes

- Weak stationary Auto-regressive Moving Average (ARMA) processes allow reduction to a manageable number of parameters.

- The simple structure of ARMA processes makes them very useful and flexible models for weak stationary time series $(y_1, \ldots, y_N)$.

- We assume that y_t has zero mean. Incorporation of non-zero mean is straightforward.

- Modelling of non-stationary data is based on variations of ARMA models.

First order auto-regressive processes: AR(1)

- The simplest example from the ARMA family is the first-order auto-regressive process denoted AR(1) i.e.

$$y_t = \varphi_1 y_{t-1} + \epsilon_t \,. \tag{1}$$

  Here ϵ_t constitute a white noise process i.e. zero mean 'random shocks' or 'innovations' assumed to be independent of each other and identically distributed with constant variance σ_ϵ^2.

- Equation (1) can be written symbolically in the more compact form

$$\varphi(B) y_t = \epsilon_t,$$

  where $\varphi(z) = 1 - \varphi_1 z$ and B is the backward shift or lag-operator defined by

$$B^m y_t = y_{t-m}.$$

AR(1) (cont.)

- The stationarity condition in an AR(1) process $y_t = \varphi_1 y_{t-1} + \epsilon_t$ amounts to $|\varphi_1| < 1$. Equivalently,

$$\varphi(z) = 1 - \varphi_1 z \neq 0, \qquad \text{for all} \quad z \in \mathbb{C} \quad \text{such that} \quad |z| \leq 1.$$

- By slight rearrangement and using the lag-operator, the AR(1) model $(1 - \varphi_1 B)y_t = \epsilon_t$ can be written as:

$$y_t = (1 - \varphi_1 B)^{-1}\epsilon_t = (1 + \varphi_1 B + \varphi_1^2 B^2 + \varphi_1^3 B^3 + \ldots)\epsilon_t.$$

  Notice that this series representation will converge as long as $|\varphi_1| < 1$.

AR(1) (cont.)

- For the AR(1) process it can be shown that:

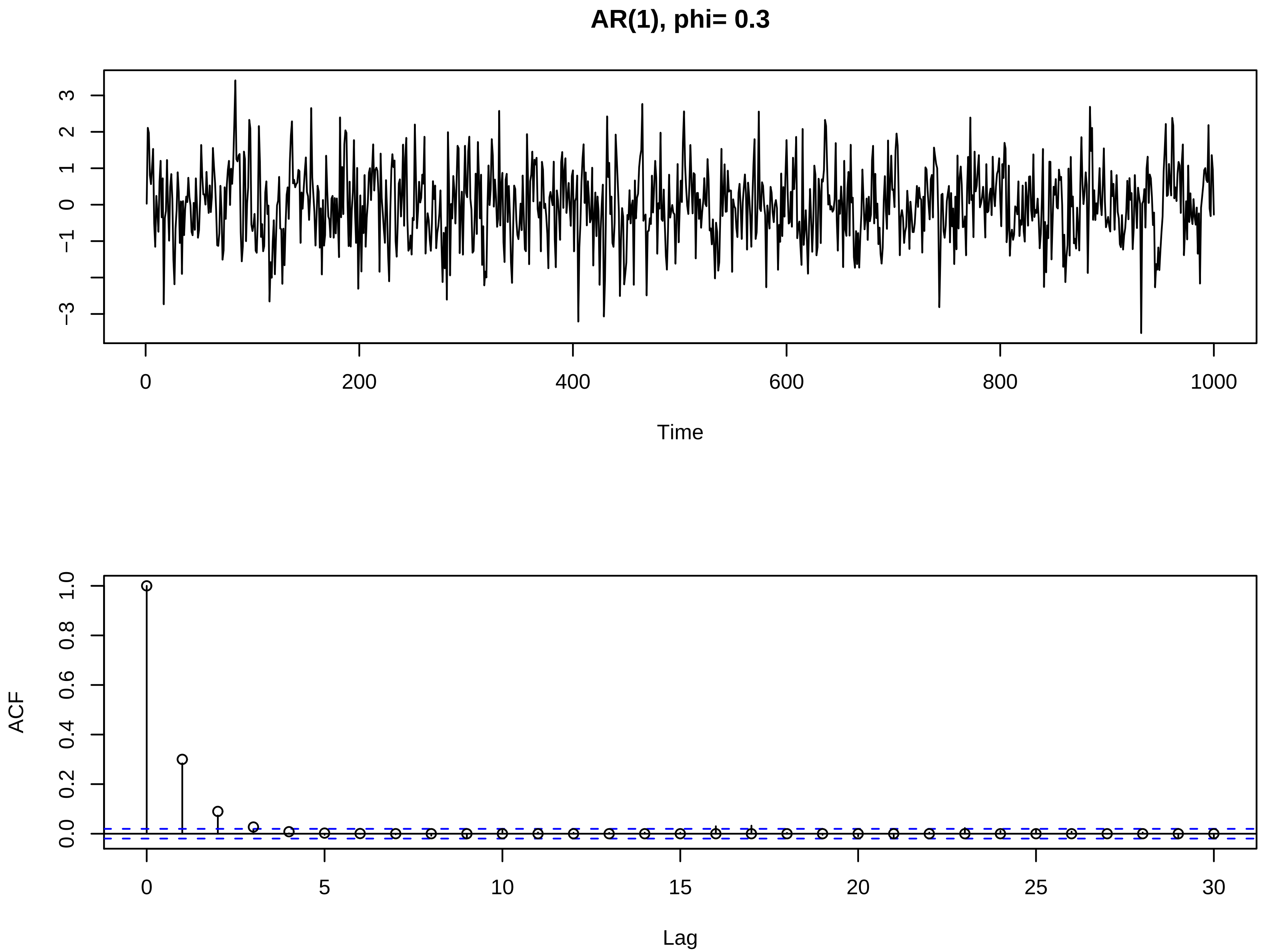
$$\text{Var}(y_t) = \gamma_0 = \sigma_\epsilon^2(1 + \varphi_1^2 + \varphi_1^4 + \ldots) = \frac{\sigma_\epsilon^2}{(1 - \varphi_1^2)},$$

$$\text{cov}(y_t, y_{t-k}) = \gamma_k = \gamma_{k-1}\varphi_1 \, , \quad k > 0,$$

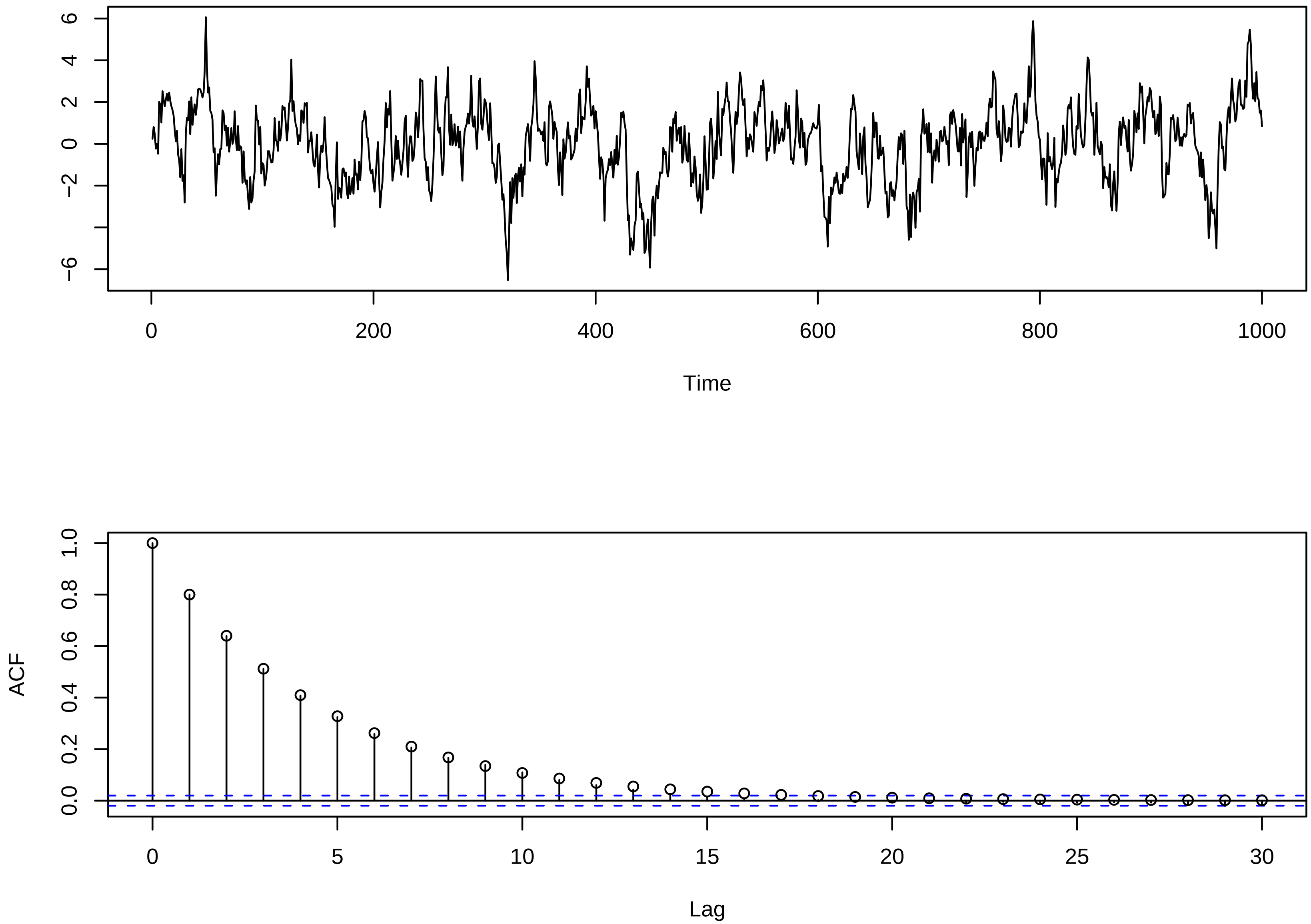
$$\rho_k = \frac{\gamma_k}{\gamma_0} = \varphi_1^k.$$

- Since $|\varphi_1| < 1$, the ACF ρ_k shows a pattern which is decreasing in absolute value. This implies that the linear dependence of two observations y_t and y_s becomes weaker with increasing time distance between t and s.

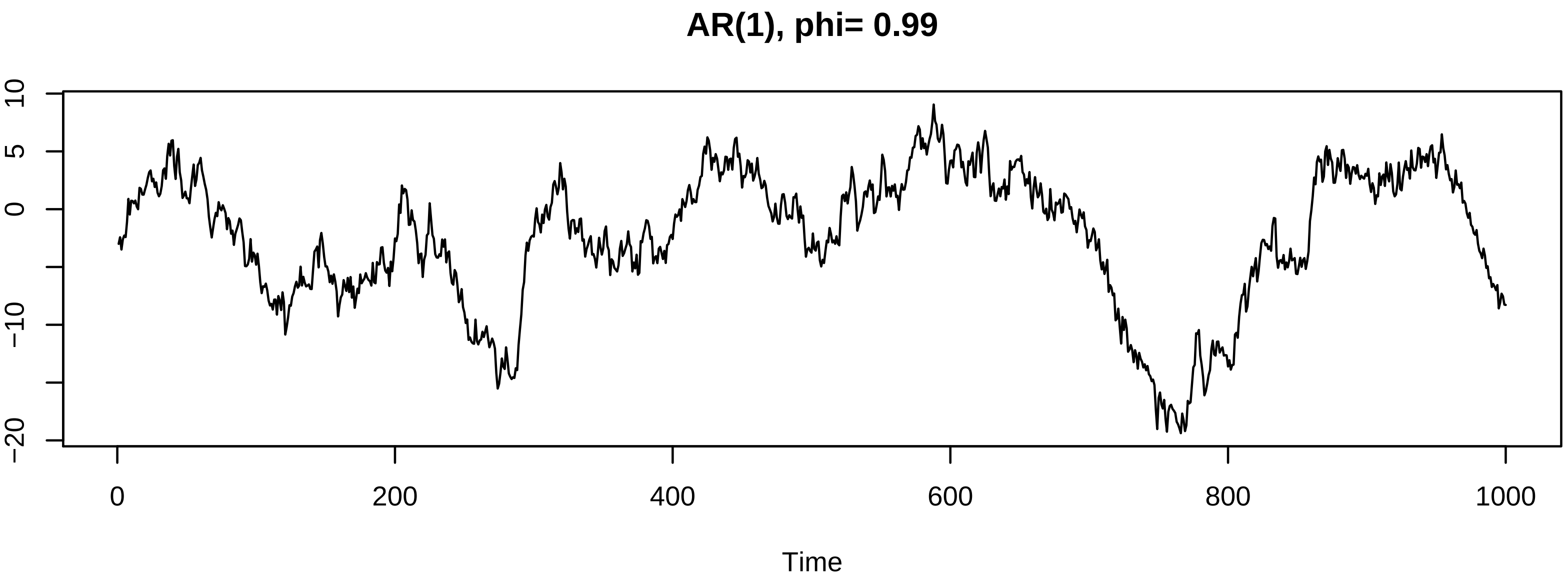
- If $0 < \varphi_1 < 1$, the ACF decays exponentially to zero, while if $-1 < \varphi_1 < 0$, the ACF decays in an oscillatory manner. Both decays are slow if φ is close to the non-stationary boundaries ± 1.


AR(1), phi= 0.3
3
2
1
0
−1
−3
0
200
400
600
800
1000
Time
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

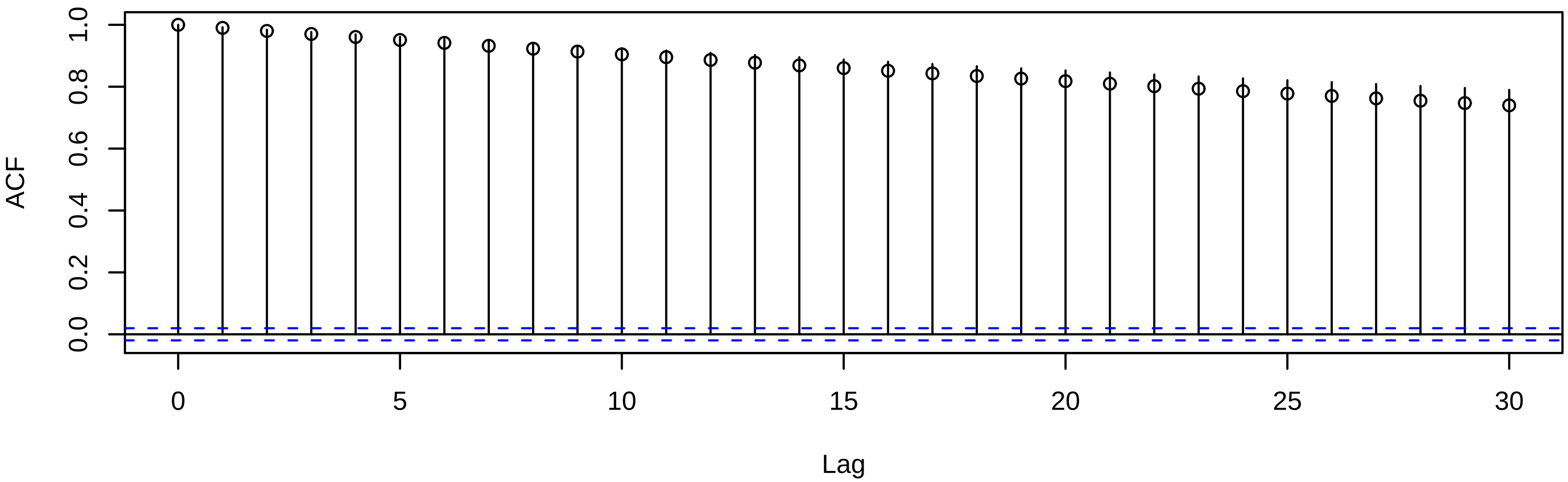
AR(1), phi= 0.8



6
4
2
0
−2
−6
0
200
400
600
800
1000
Time
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag



AR(1), phi= 0.99
10
5
0
−10
−20
0
200
400
600
800
1000
Time



ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
Lag

First order moving average processes: MA(1)

- A first-order moving-average process, MA(1), is defined by:

$$y_t = \epsilon_t - \theta_1 \epsilon_{t-1} = (1 - \theta_1 B)\epsilon_t.$$

- For the MA(1) process it can be shown that:

$$\mathrm{Var}(y_t) = \gamma_0 = (1 + \theta_1^2)\sigma_\epsilon^2$$

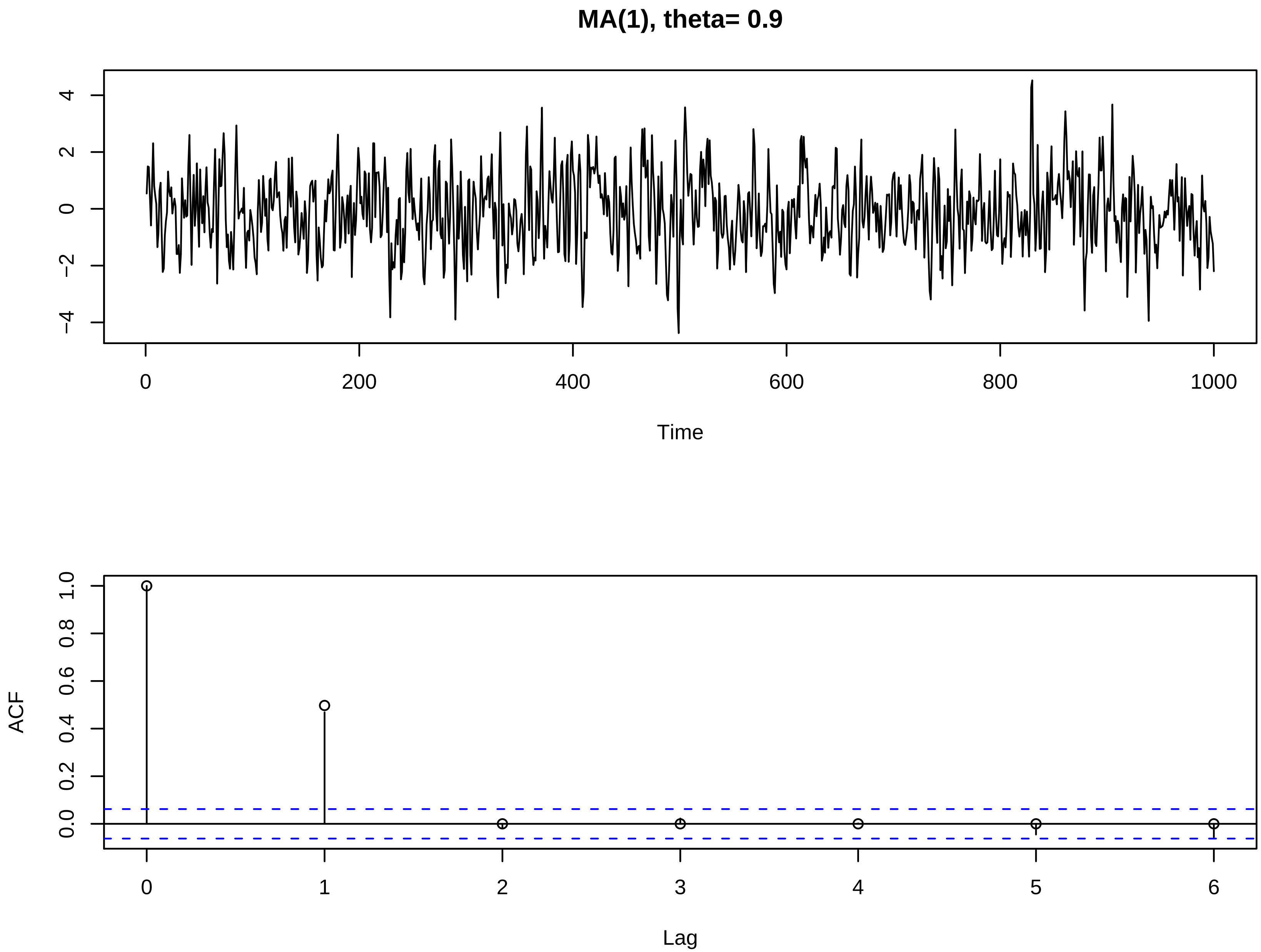
$$\mathrm{cov}(y_t, y_{t-1}) = \gamma_1 = -\theta_1 \sigma_\epsilon^2$$

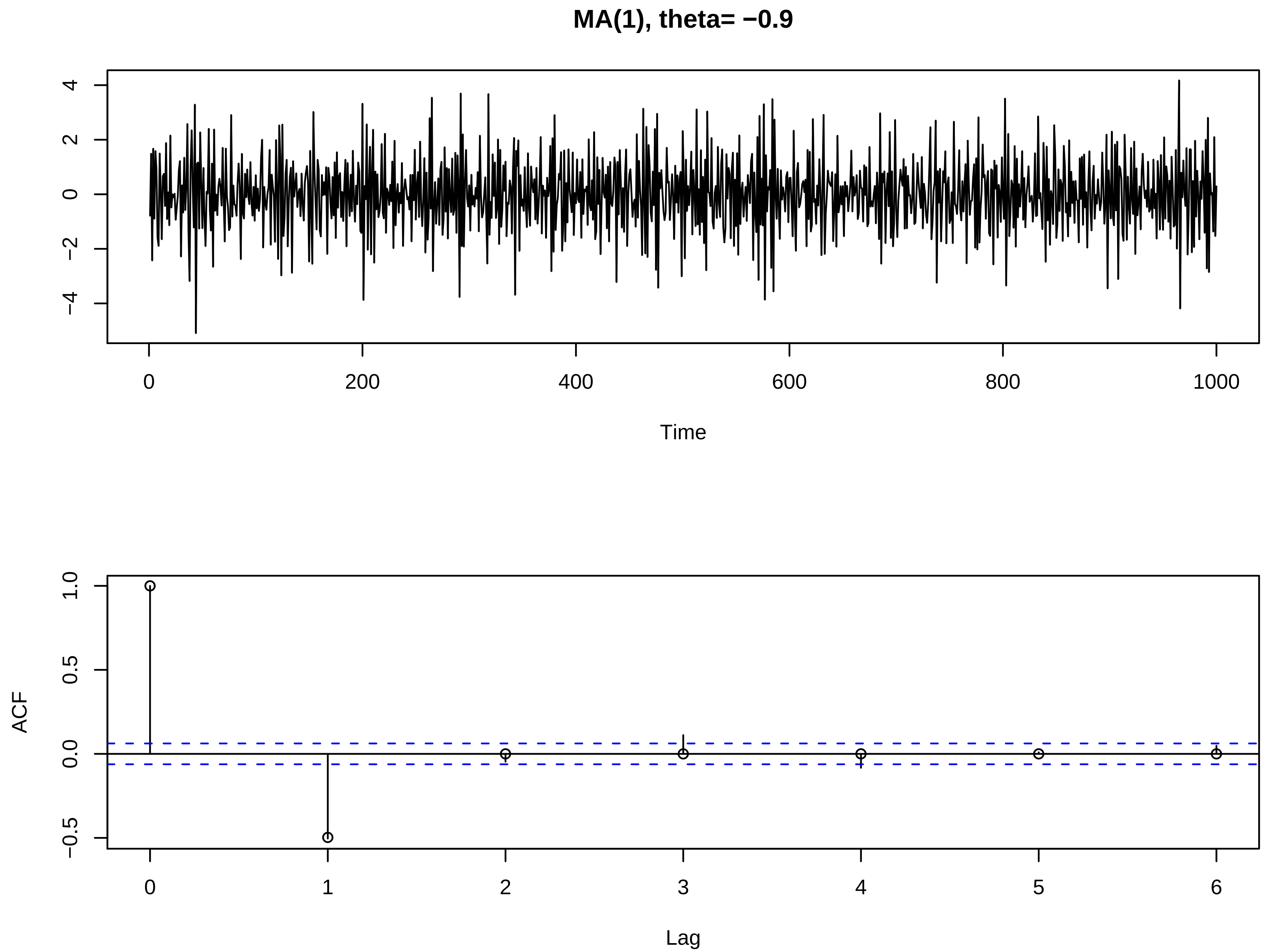
$$\mathrm{cov}(y_t, y_{t-k}) = 0, \qquad k > 1$$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{-\theta_1}{(1 + \theta_1^2)}$$

$$\rho_k = 0, \qquad k > 1\,.$$

- Note: the two observations y_t and y_s generated by a MA(1) process are uncorrelated if t and s are more than one observation apart.


MA(1), theta= 0.9
4
2
0
−2
−4
0
200
400
600
800
1000
Time
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
Lag


MA(1), theta= −0.9
4
2
0
−2
−4
0
200
400
600
800
1000
Time
ACF
1.0
0.5
0.0
−0.5
0
1
2
3
4
5
6
Lag

AR(p) and MA(q) processes

- Both the AR(1) and MA(1) processes impose strong restrictions on the pattern of the corresponding ACF.

- More general ACF patterns are allowed by autoregressive or moving average models of higher order.

- The AR(p) and MA(q) models are defined as follows:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \epsilon_t \quad \text{(AR(p) process)}$$

  and

$$y_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q} \quad \text{(MA(q) process)}$$

  The φ_i and θ_j, $i = 1, \ldots, p$; $j = 1, \ldots, q$ are parameters.

Autoregressive Moving Average Processes: ARMA(p,q)

- Combining the AR(p) and MA(q) processes we define an autoregressive moving average process ARMA(p,q):

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q} \,.$$

- Using the lag operator B, the ARMA(p,q) model may be written:

$$(1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p) y_t = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q) \epsilon_t$$

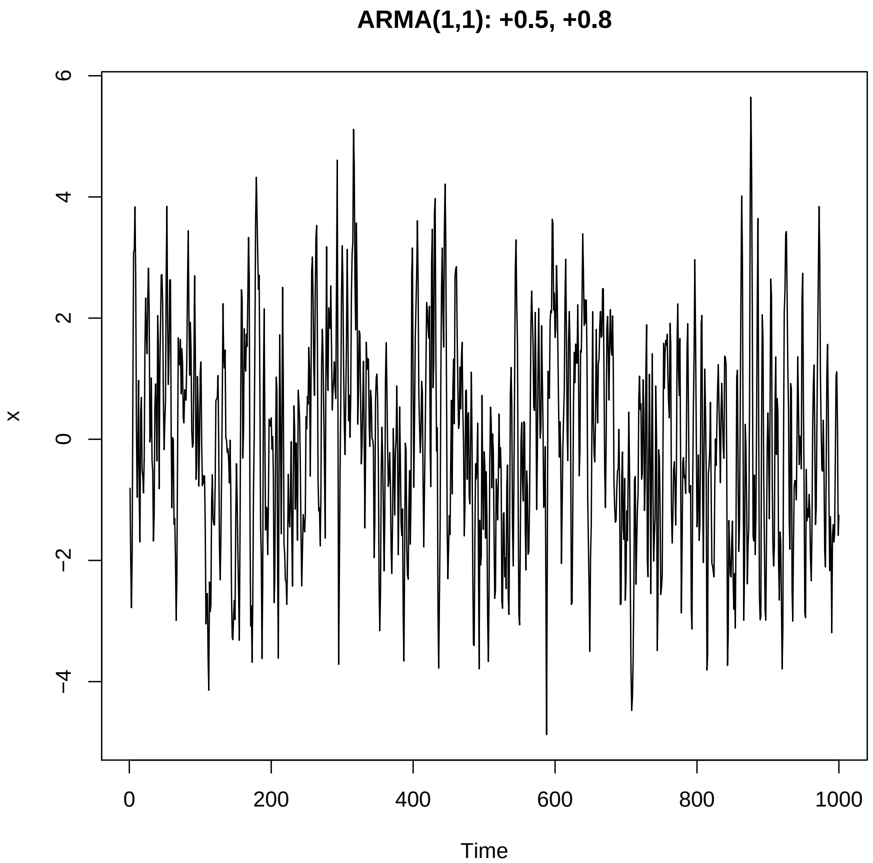
or more compactly as:

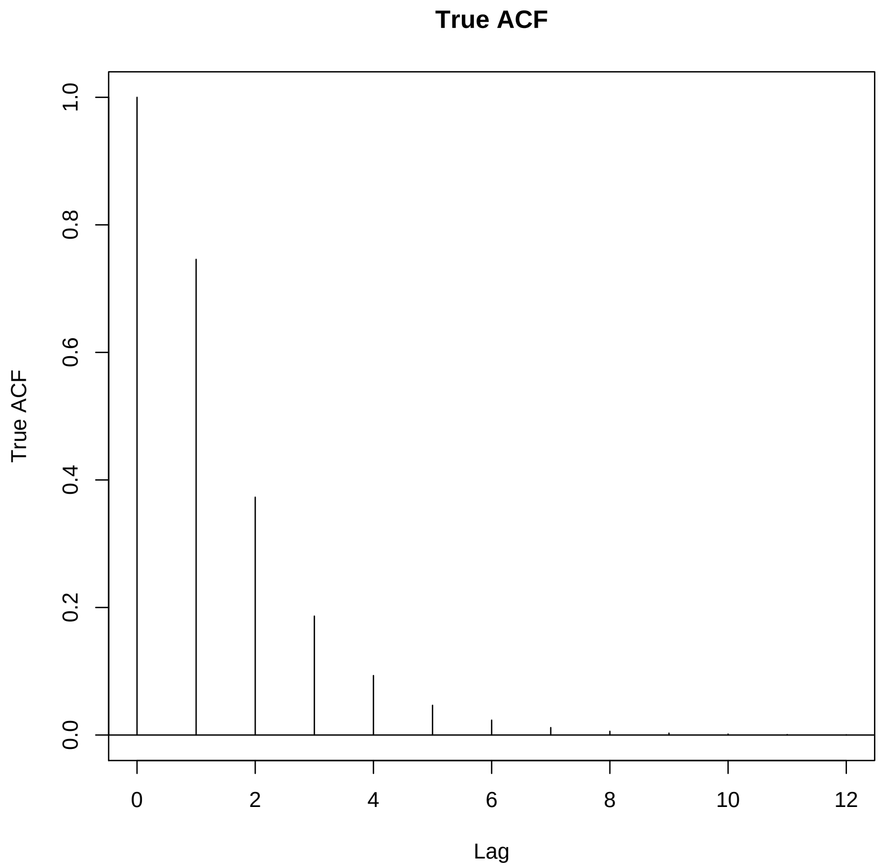
$$\varphi(B) y_t = \theta(B) \epsilon_t,$$

where

$$\varphi(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \ldots - \varphi_p z^p,$$

$$\theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \ldots - \theta_q z^q.$$


ARMA(1,1): +0.5, +0.8
x
Time
0
200
400
600
800
1000
−4
−2
0
2
4
6


True ACF
True ACF
Lag
0
2
4
6
8
10
12
0.0
0.2
0.4
0.6
0.8
1.0

Stationarity Conditions

- Assume that the polynomials $\theta(z)$ and $\varphi(z)$ have no common zeroes.
- An ARMA(p,q) model defined by $\varphi(B)y_t = \theta(B)\epsilon_t$ is stationary if

$$\varphi(z) = (1 - \varphi_1 z - \varphi_2 z^2 - \ldots - \varphi_p z^p) \neq 0, \quad \text{for } |z| \leq 1.$$

- For a stationary ARMA(p,q) process the polynomial $\varphi(B)$ can be 'inverted' and so y_t has a moving average representation of infinite order:

$$y_t = \sum_{j=0}^{\infty} \psi_j \, \epsilon_{t-j} \,, \tag{2}$$

where the coefficients ψ_j are determined by the relation

$$\frac{\theta(z)}{\varphi(z)} = \sum_{j=0}^{\infty} \psi_j \, z^j, \qquad |z| \leq 1.$$

We write equation (2) in compact form: $y_t = \varphi^{-1}(B)\theta(B)\epsilon_t$.

Invertibility Conditions

- The ARMA(p,q) model $\varphi(B)y_t = \theta(B)\epsilon_t$ is called invertible if there exists a sequence of constants $\{\pi_j\}$ such that $\sum_{j=0}^{\infty} \left|\pi_j\right| < \infty$ and

$$\epsilon_t = \sum_{j=0}^{\infty} \pi_j \, y_{t-j}. \tag{3}$$

- Assume that the polynomials $\theta(z)$ and $\varphi(z)$ have no common zeroes. Then the ARMA(p,q) process is invertible if and only if

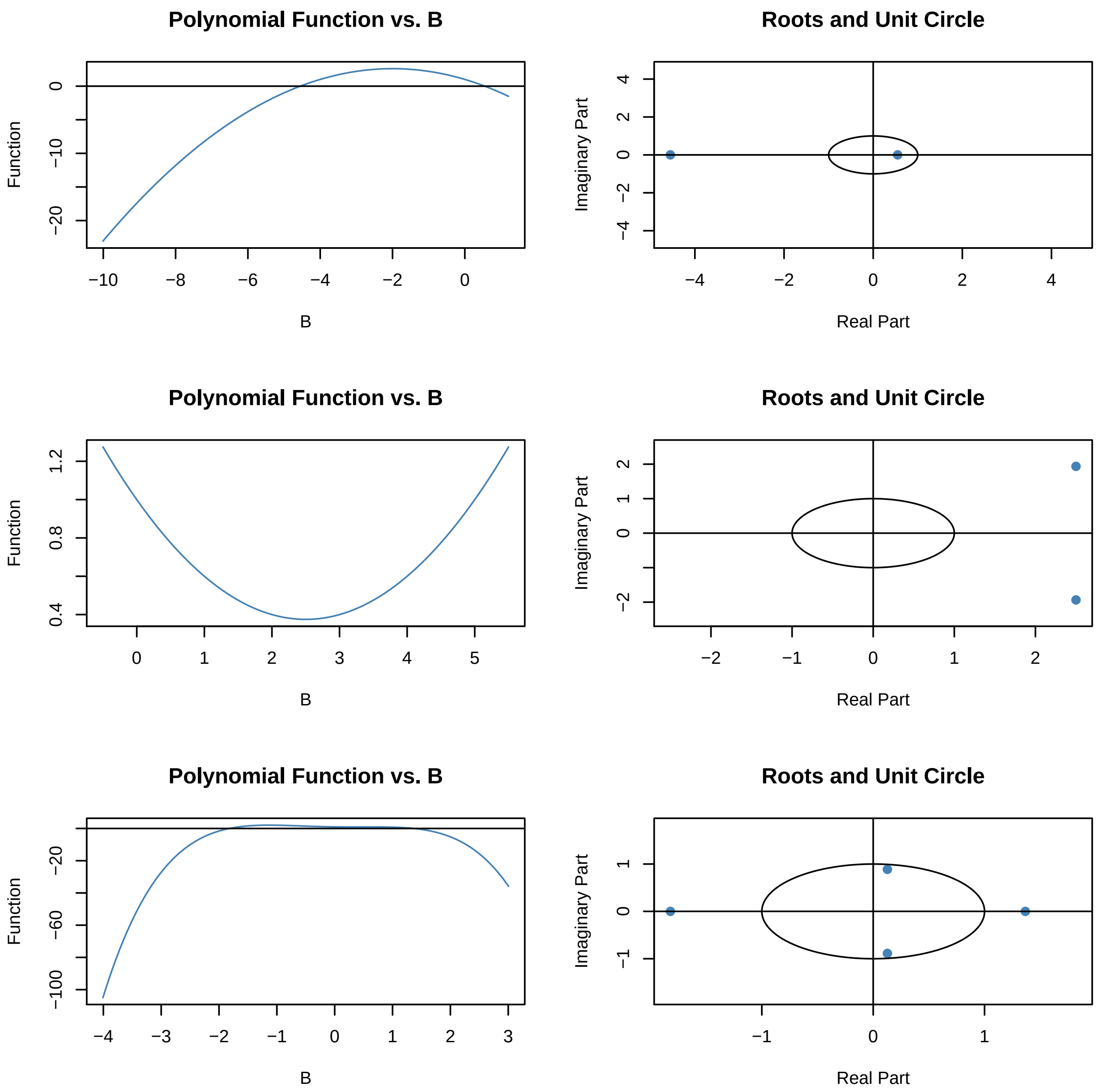
$$\theta(z) = (1 - \theta_1 z - \theta_2 z^2 - \ldots - \theta_q z^q) \neq 0, \quad \text{for } |z| \leq 1.$$

The coefficients π_j are determined by the relation

$$\frac{\varphi(z)}{\theta(z)} = \sum_{j=0}^{\infty} \pi_j \, z^j, \qquad |z| \leq 1.$$

We write equation (3) in the following compact form:

$$\theta^{-1}(B)\varphi(B)y_t = \epsilon_t \,.$$


Polynomial Function vs. B
Function
B
−10 −8 −6 −4 −2 0
0 −10 −20
Roots and Unit Circle
Imaginary Part
Real Part
−4 −2 0 2 4
Polynomial Function vs. B
Function
B
0 1 2 3 4 5
1.2 0.8 0.4
Roots and Unit Circle
Imaginary Part
Real Part
−2 −1 0 1 2
Polynomial Function vs. B
Function
B
−4 −3 −2 −1 0 1 2 3
−20 −60 −100
Roots and Unit Circle
Imaginary Part
Real Part
−1 0 1

Non-zero mean ARMA processes

- For ARMA models we have so far assumed a zero-mean stationary process.

- The generalisation of stationary non-zero constant mean ARMA(p,q) is straightforward:

  - Augmenting the stationary process with an additional parameter $\nu \neq 0$ one obtains: $\varphi(B)y_t = \nu + \theta(B)\epsilon_t$.

  - Inversion of $\varphi(B)$ then immediately yields the mean of y_t as: $\mu = \mathrm{E}(y_t) = \varphi^{-1}(B)\nu$.

  - Note that if $\varphi(B) = 1$ (which is the case for the pure MA(q) model) one has $\mu = \nu$.

Modelling using ARMA processes

Step 1. ARMA model identification;

Step 2. ARMA parameter estimation

Step 3. ARMA model selection ;

Step 4. ARMA model checking;

Step 5. forecasting from ARMA models.

ARMA model identification

- A plot of the data will give us some clue as to whether the series is not stationary.

- To analyse an observed stationary time series through an ARMA(p,q) model, the first step is to determine appropriate values for p and q.

- One of the basic tools in such model order identification are plots of the estimated ACF $\widehat{\rho}_k$ and PACF $\widehat{\alpha}_k$ against the lag k.

- The shape of these plots can help to discriminate between competing models.

ARMA model identification (cont.)

- The autocorrelations:
  - for a MA(q) process $\rho_k = 0$ for $k \geq q + 1$;
  - for an AR(p) process they decay exponentially.
  - for a mixed ARMA(p,q) we expect the correlations to tail off after lag $p - q$.

- These considerations assist in deciding whether $p > 0$ and, if not, to choose the value of q.

Estimators for ACF/PACF (see Ch. 7 in ref 3)

- Let $(y_1, y_2, \ldots, y_N)$ be a realization of a weak stationary time series.

- The sample autocovariance function is defined by

$$\widehat{\gamma}_k = \frac{1}{N} \sum_{t=1}^{N-k} (y_t - \bar{y})(y_{t+k} - \bar{y}) \qquad 0 \leq k \leq N\,,$$
$$\widehat{\gamma}_k = \widehat{\gamma}_{-k}\,, \qquad -N < h \leq 0,$$

where $\bar{y}$ is the sample mean

$$\bar{y} = \frac{1}{N} \sum_{j=1}^{N} y_j.$$

- The sample autocorrelation function is defined by

$$\widehat{\rho}_k = \frac{\widehat{\gamma}_k}{\widehat{\gamma}_0}\,, \qquad |k| < N.$$

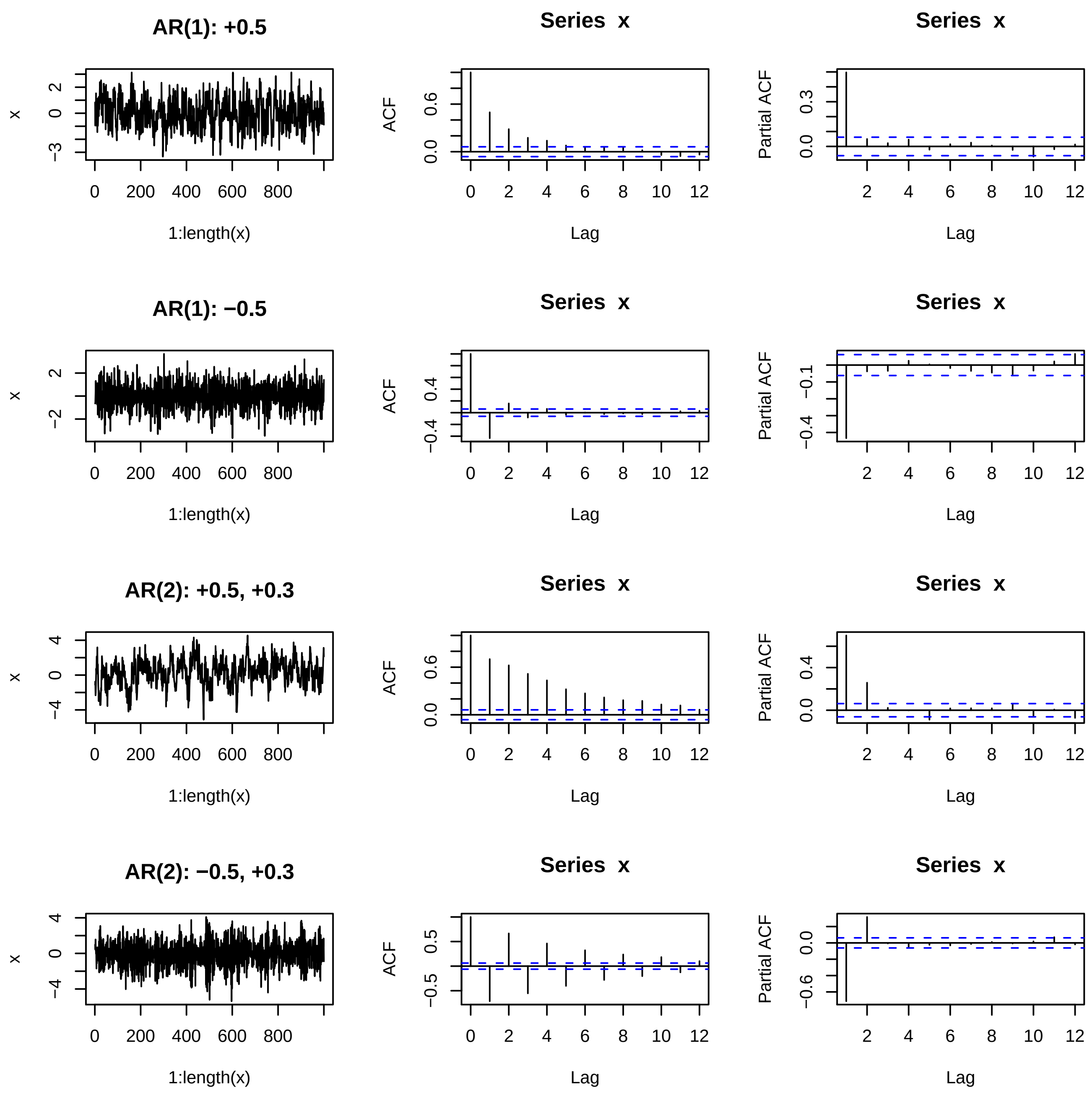
Estimators ACF/PACF (cont.)

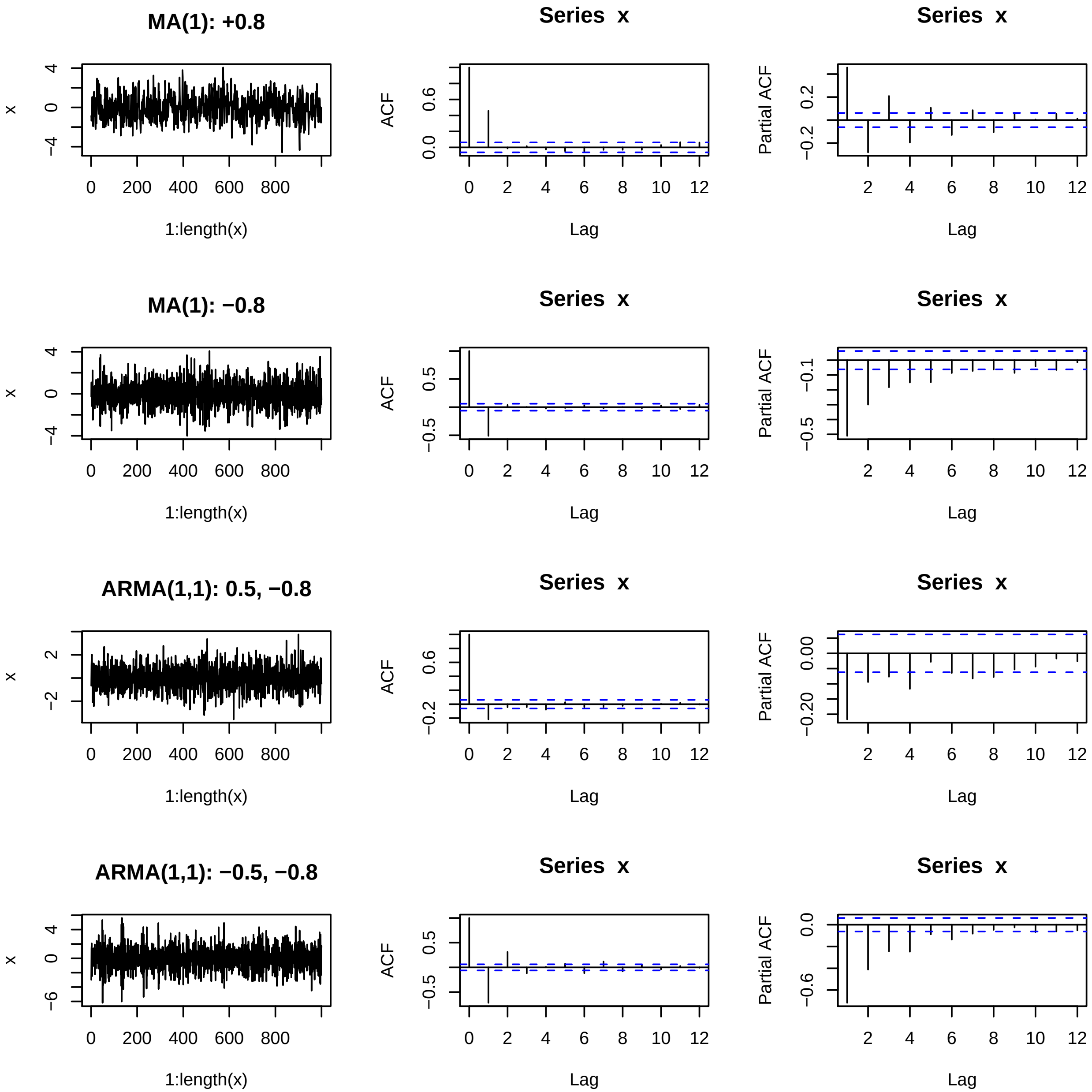
- The sample PACF at lag k can be computed as a function of the sample estimate of the ACF as:

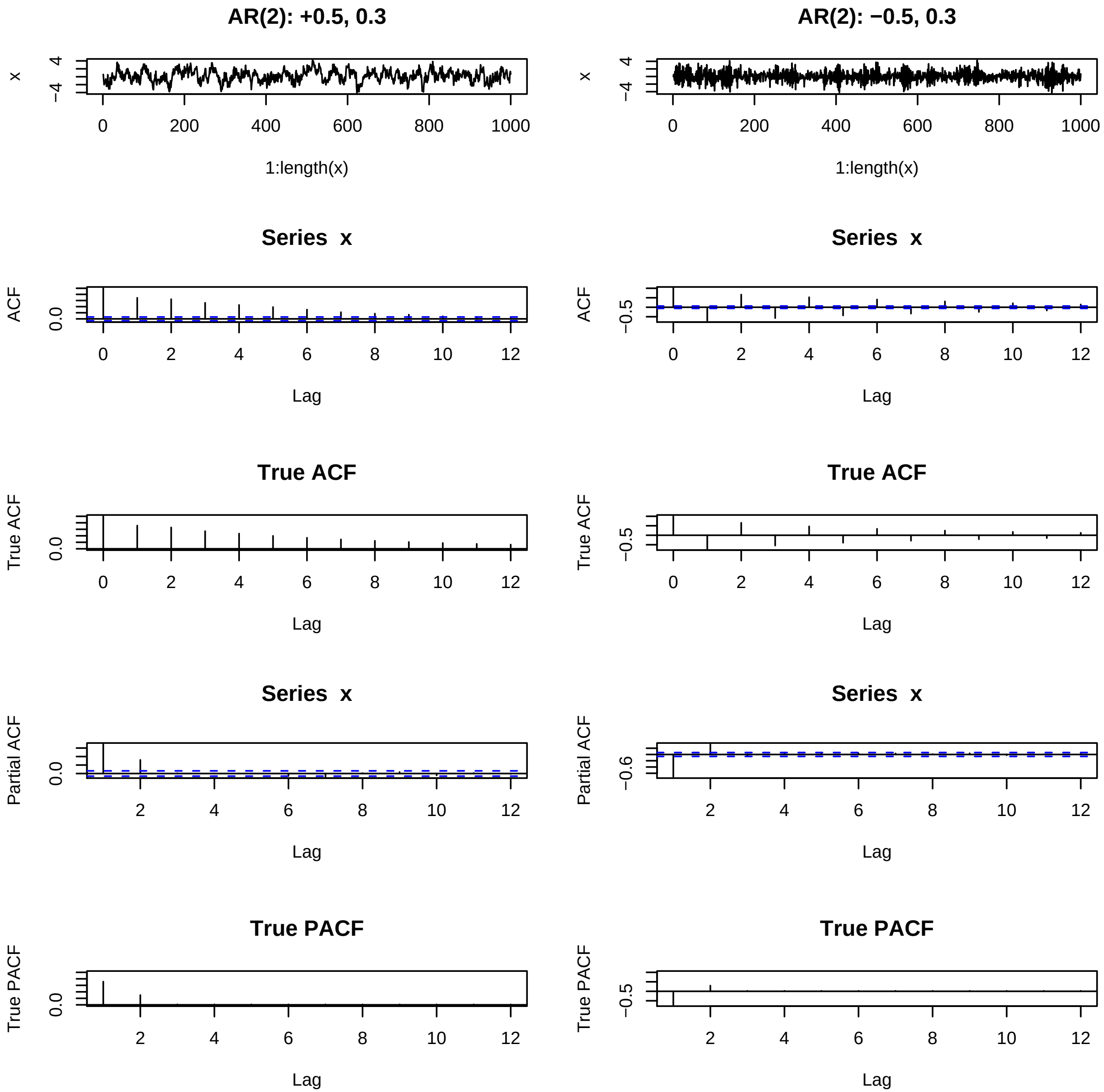
$$\widehat{\alpha}_k = \widehat{\psi}_{kk}\,, \qquad k \geq 1,$$

where $\widehat{\psi}_{kk}$ is uniquely determined by:

$$\begin{pmatrix} \widehat{\rho}_0 & \widehat{\rho}_1 & \widehat{\rho}_2 & \cdots & \widehat{\rho}_{k-1} \\ \widehat{\rho}_1 & \widehat{\rho}_0 & \widehat{\rho}_1 & \cdots & \widehat{\rho}_{k-2} \\ \vdots & & & & \vdots \\ \widehat{\rho}_{k-1} & \widehat{\rho}_{k-2} & \widehat{\rho}_{k-3} & \cdots & \widehat{\rho}_0 \end{pmatrix} \begin{pmatrix} \widehat{\psi}_{k1} \\ \widehat{\psi}_{k2} \\ \vdots \\ \widehat{\psi}_{kk} \end{pmatrix} = \begin{pmatrix} \widehat{\rho}_1 \\ \widehat{\rho}_2 \\ \vdots \\ \widehat{\rho}_k \end{pmatrix}.$$


AR(1): +0.5
AR(1): −0.5
AR(2): +0.5, +0.3
AR(2): −0.5, +0.3
Series x
x
ACF
Partial ACF
1:length(x)
Lag
0
200
400
600
800
2
4
6
8
10
12


MA(1): +0.8
Series x
Series x
MA(1): −0.8
Series x
Series x
ARMA(1,1): 0.5, −0.8
Series x
Series x
ARMA(1,1): −0.5, −0.8
Series x
Series x
x
1:length(x)
ACF
Partial ACF
Lag


AR(2): +0.5, 0.3
AR(2): −0.5, 0.3
1:length(x)
Series x
ACF
True ACF
Partial ACF
True PACF
Lag

ARMA Parameter estimation

- Fitting an ARMA(p,q) model requires estimation of:
  - the model parameters $(\varphi_1, \ldots, \varphi_p); (\theta_1, \ldots, \theta_q)$;
  - the mean μ (where this is non-zero) and
  - the variance, σ_ϵ^2, of the underlying white noise process ϵ_t.

- If we denote the full set of these parameters by a vector $\boldsymbol{\Theta}$ then we can proceed:
  - to write down a likelihood for the data $L(\boldsymbol{\Theta}; \mathbf{y}) = \mathbf{p}(\mathbf{y}; \boldsymbol{\Theta})$,
  - estimate the parameters by maximum likelihood and
  - derive standard errors and confidence intervals through the asymptotic likelihood theory results.

ARMA Parameter estimation (cont.)

- The usual way to proceed is to assume that $\epsilon_t \sim \mathrm{N}(0, \sigma_\epsilon^2)$.

- The resulting derivation of the likelihood function and the associated maximisation algorithm for the general ARMA(p,q) model is somewhat involved and we do not go into details here.

- The basic idea is to factorise the joint distribution $p(y_1, y_2, \ldots, y_N)$ as $p(y_1, y_2, \ldots, y_N) = p(y_1) \prod_{t=2}^{N} p(y_t | y_1, \ldots, y_{t-1})$.

- It may then be shown that $p(y_t | y_1, \ldots, y_{t-1})$ is normal with mean given by the predicted value $\widehat{y}_t$ of y_t and similarly that the marginal distribution $p(y_1)$ is normal with mean $\widehat{y}_1$.

- Then log likelihood can then be expressed in terms of the prediction errors $(y_t - \widehat{y}_t)$. This assists in developing algorithms to effect the maximisation.

ARMA Model Selection

- We want to find a model that fits the observed data as well as possible.

- Once fitted, models can then be compared by the use of a suitable penalised log-likelihood measure, for example Akaike's Information Criterion (AIC)

- There exists a variety of other selection criteria that have been suggested to choose an appropriate model.

- All these are similar differing only in the penalty adjustment involving the number of estimated parameters.

- As for the AIC, the criteria are generally arranged so that better fitting models correspond to lower values of the criteria.

ARMA Model checking

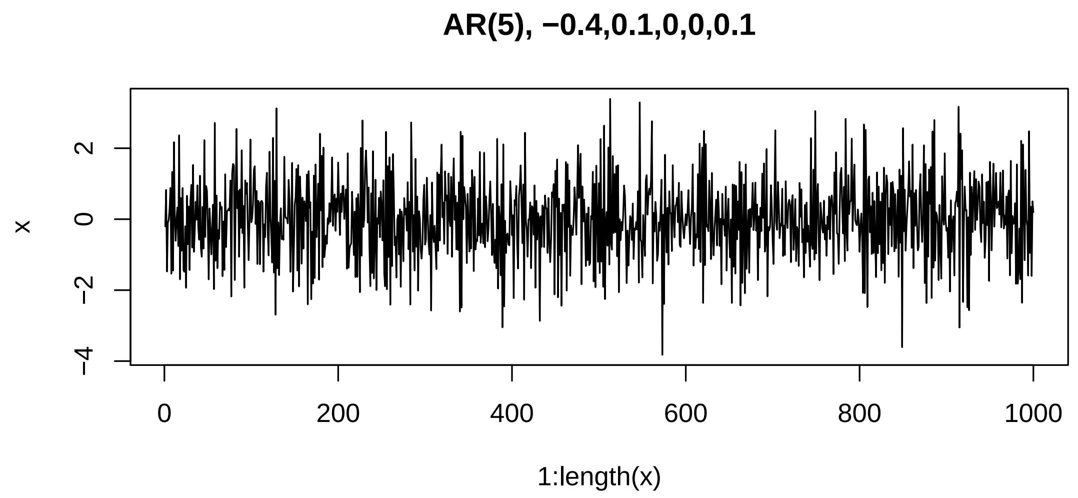
- The residuals for an ARMA model are estimated by subtraction of the adopted model predictions from the observed time series.

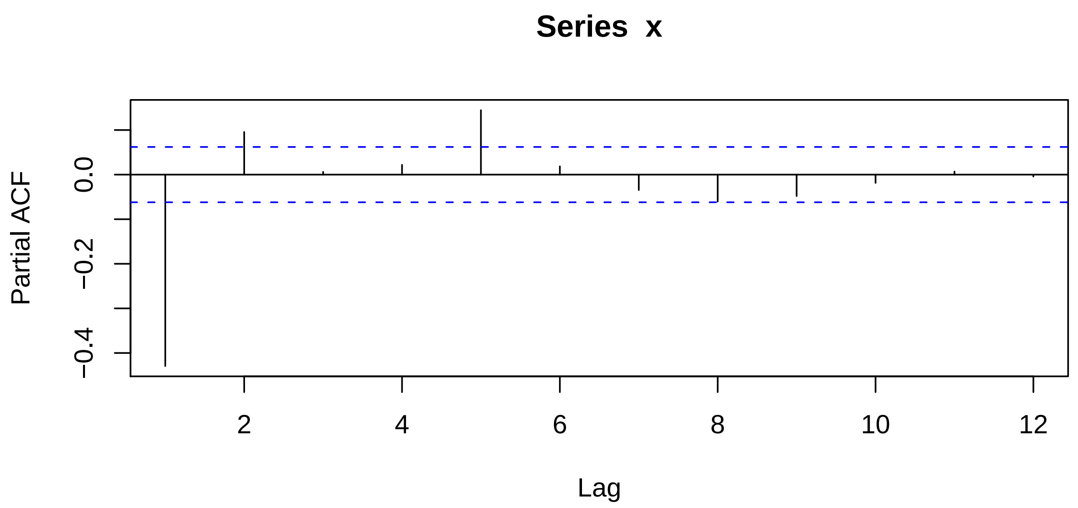
- If the model assumptions are valid then we would expect the (standard) residuals to be independent and normally distributed.

- In time series analysis it is important to check that there is no autocorrelation remaining in the residuals. Plots of residuals against the time ordering are therefore important.

- Various tests for serial correlation in the residuals are available.

Ex. 4


AR(5), −0.4,0.1,0,0,0.1
x
2
0
−2
−4
0
200
400
600
800
1000
1:length(x)


Series x
Partial ACF
0.0
−0.2
−0.4
2
4
6
8
10
12
Lag

Example 5

- The function `armaFit()` estimates the parameters of ARMA models (arguments are described on the help page).

- Consider the time series generated in Ex 4. from an AR(5) model with parameters:

$$\varphi_1 = -0.4, \quad \varphi_2 = 0.1, \quad \varphi_3 = \varphi_4 = 0, \quad \varphi_5 = 0.1.$$

- Examination of the PACF (see above) reveals significant correlation at lag 5, after which the correlation is negligible.

- This suggests to use an ARMA(p,q) model with $p = 5$, with q 1 or 2 (this is because the PACF of an MA(q) decreases exponentialy).

- We first apply the function `armaFit()` to estimate the parameters of an AR(5) model.

Example 5 (cont)

```
fit<-armaFit(x~ar(5),x,method="mle")
summary(fit)

Model:
 ARIMA(5,0,0) with method: CSS-ML

Coefficient(s):
      ar1        ar2        ar3        ar4        ar5  intercept
-0.419200   0.108544   0.006913  -0.004710   0.146163  -0.054552

Residuals:
     Min       1Q   Median       3Q      Max
-3.36283 -0.65182  0.02615  0.65574  3.19371


Moments:
Skewness Kurtosis
 -0.1242   0.1234
```

Example 5 (cont)

```
Coefficient(s):
           Estimate  Std. Error  t value Pr(>|t|)
ar1       -0.419200    0.031291  -13.397  < 2e-16 ***
ar2        0.108544    0.033978    3.195   0.0014 **
ar3        0.006913    0.034145    0.202   0.8396
ar4       -0.004710    0.034024   -0.138   0.8899
ar5        0.146163    0.031329    4.665 3.08e-06 ***
intercept -0.054552    0.027412   -1.990   0.0466 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

sigma^2 estimated as: 1.016
log likelihood:       -1427.07
AIC Criterion:        2868.15
```

Example 5 (cont)

- Note that `summary()` also provides the estimate of the variance σ^2 of the white noise process.

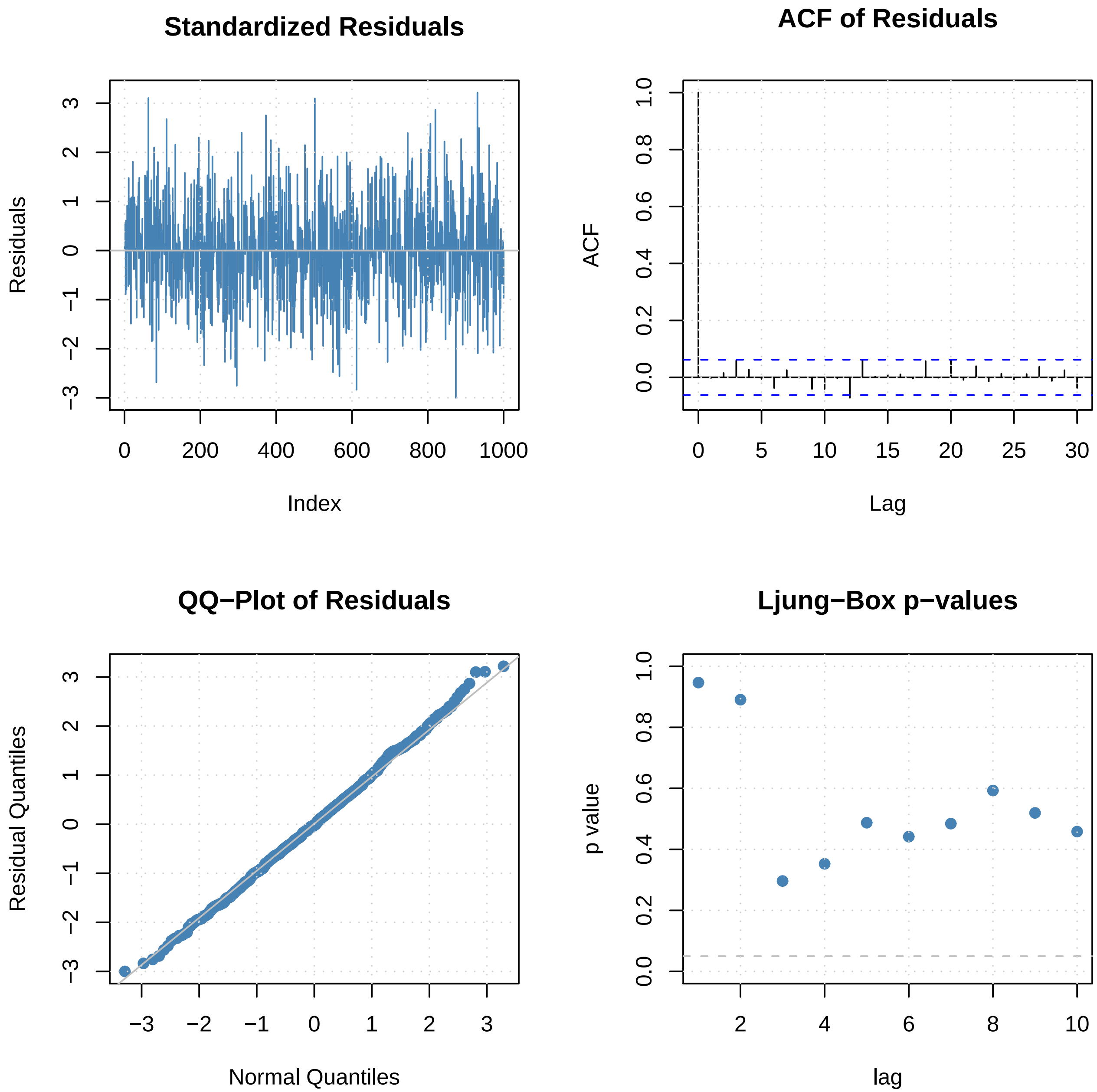
- The values of the AR coefficients of order 3 and 4 are small and the associated standard errors are large: as a consequence, these coefficients have large p-values (last column) and are not statistically significant according to a 5% t-test. It is therefore a good idea to fit an AR(5) process in which these coefficients (as well as the intercept) are fixed to zero. This can be specified with the parameter `fixed=c()`:

Example 5 (cont.)

```
fit<-armaFit(x~ar(5),x,fixed=c(NA,NA,0,0,NA,0),method="mle")
par(mfrow=c(2,2))
summary(fit)

Model:
 ARIMA(5,0,0) with method: CSS-ML

Coefficient(s):
      ar1          ar2          ar3          ar4          ar5    intercept
  -0.3564       0.1135       0.0000       0.0000       0.1231       0.0000

Residuals:
     Min       1Q   Median       3Q      Max
-3.13847 -0.66654 -0.01819  0.68648  3.36718
```

Example 5 (cont)

```
Moments:
Skewness Kurtosis
 0.07226 -0.02576

Coefficient(s):
           Estimate  Std. Error  t value Pr(>|t|)
ar1        -0.35642     0.03115  -11.441  < 2e-16 ***
ar2         0.11350     0.03120    3.637 0.000275 ***
ar3         0.00000     0.02861    0.000 1.000000
ar4         0.00000     0.03115    0.000 1.000000
ar5         0.12309     0.03120    3.945 7.98e-05 ***
intercept   0.00000     0.02861    0.000 1.000000
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

sigma^2 estimated as: 1.095
log likelihood:       -1464.51
AIC Criterion:        2937.02
```


Standardized Residuals
Residuals
Index
ACF of Residuals
ACF
Lag
QQ−Plot of Residuals
Residual Quantiles
Normal Quantiles
Ljung−Box p−values
p value
lag

See Ex. 5 (cont)

- The `summary()` method automatically plots the residuals, the autocorrelation function of the residuals, the standardized residuals, and the Ljung-Box statistic (test of independence).

- In order to investigate the model fit we could estimate the parameters for various ARMA(p,q) models with $p_{max} = 5$ and $q_{max} = 2$ for the same simulated time series and compare the relative fits through the AIC value (see the R script ex5.r).

Modeling with ARMA(p,q) models (summary)

- Model identification: Use the ACF and the PACF function to get indicators of p and q. The following can assist you to do that:

Procces	ACF	Partial ACF
AR(p)	Exp. decay or damped cos	zero after lag p
MA(q)	Cuts after lag q	Exp. decay or damped cos
ARM(p,q)	Exponential decay after $q-p$	Decay after $p-q$

- Parameter estimation: Estimate values for the model parameters $(\varphi_1, \ldots, \varphi_p); (\theta_1, \ldots, \theta_q)$, μ and σ_ϵ^2. (there are several ways one can do this e.g. the Yule-Walker method).

Modeling with ARMA(p,q) models (summary cont.)

- Model selection:
  - Fit ARMA(p,q) models by the maximum Likelihood estimates using the (Yule-Walke) estimates for the parameter as initial values of the maximisation algorithm.
  - Prevent over-fitting by imposing a cost for increasing the number of parameters in the fitted model. One way in which this can be done is by the information criterion of Akaike (AIC)
  - The model selected is the one that minimises the value of AIC.

- Model checking

  - The residuals of a fitted model are the scaled difference between an observed and a predicted value.

  - Goodness of fit is checked essentially by checking that the residuals are like white noise (i.e. mean zero i.i.d. process with constant variance).

  - There are several candidates for the residuals one is the computed in the course of determining the maximum likelihood estimates:

$$\hat{W}_t = \frac{Y_t - \hat{Y}_t(\hat{\varphi}, \hat{\theta}))}{r_{t-1}(\hat{\varphi}, \hat{\theta})^{1/2}},$$

    where $\hat{Y}_t(\hat{\varphi}, \hat{\theta})$ are the predicted values of Y_t, based on $Y_1, \ldots, Y_{N-1}$ for the fitted ARMA(p,q) model and $r_{t-1}(\hat{\varphi}, \hat{\theta})^{1/2}$ are the *sample mean squared errors*. Another is:

$$\hat{Z}_t = \hat{\theta}^{-1}(B)(\hat{\varphi})(B)Y_t.$$

Forecasting from ARMA models

- Given a series $(y_1, y_2, \ldots, y_N)$ up to time N, a prominent issue within time series analysis is:
  - to provide estimates of future values y_{N+h}, $h = 1, 2, \ldots$
  - conditionally on the available information, i.e. $y_N, y_{N-1}, y_{N-2}, \ldots$.

- Within the class of weak stationary ARMA(p,q) processes y_{N+h} is given by:

$$\begin{aligned} y_{N+h} = \nu + \varphi_1\, y_{N+h-1} + \ldots + \varphi_p\, y_{N+h-p} \\ + \epsilon_{N+h} - \theta_1\, \epsilon_{N+h-1} - \ldots - \theta_q\, \epsilon_{N+h-q} \qquad (*) \end{aligned}$$

Forecasting from ARMA models (cont.)

- An obvious forecast for y_{N+h} is
$$\widehat{y}_{N+h} = \mathbb{E}\left[y_{N+h} | y_N, y_{N-1}, y_{N-2}, \cdots\right]$$
i.e. its expected value given the observed series.

- The computation of this expectation follows a recursive scheme of substituting:
$$\widehat{y}_{N+j} = \begin{cases} y_{N+j} & , j \leq 0 \\ \widehat{y}_{N+j} & , j > 0 \end{cases}$$
into equation $(*)$ in place of y_{N+j} and taking $\epsilon_{N+j} = 0$ for $j > 0$.

Forecasting from ARMA models (cont.)

- For example for the ARMA(1,1) model with a non-zero mean, equation $(*)$ is: $y_{N+h} = \nu + \varphi_1 y_{N+h-1} + \epsilon_{N+h} - \theta_1 \epsilon_{N+h-1}$ so we obtain successively:

$$
\begin{aligned}
\hat{y}_{N+1} &= \nu + \varphi_1 y_N - \theta_1 \epsilon_N \\
\hat{y}_{N+2} &= \nu + \varphi_1 \hat{y}_{N+1} \\
&= \nu + \varphi_1(\nu + \varphi_1 y_N - \theta_1 \epsilon_N) \\
\hat{y}_{N+3} &= \nu + \varphi_1 \hat{y}_{N+2} \\
&= \nu + \varphi_1(\nu + \varphi_1(\nu + \varphi_1 y_N - \theta_1 \epsilon_N)) \\
\vdots \quad &\quad \vdots
\end{aligned}
$$

Iterating this scheme shows that with increased forecast horizon the forecast converges to the mean of the process μ.

Forecasting from ARMA models (cont.)

- Obtaining the sequence of forecast errors $\widehat{\epsilon}_{N+h} = y_{N+h} - \widehat{y}_{N+h}$ follows the same sort of scheme so that:

$$
\begin{aligned}
\widehat{\epsilon}_{N+1} &= y_{N+h} - \widehat{y}_{N+h} \\
&= \nu + \varphi_1 y_N + \epsilon_{N+1} - \theta_1 \epsilon_N - (\nu + \varphi_1 y_N - \theta_1 \epsilon_N) \\
&= \epsilon_{N+1}
\end{aligned}
$$

- Iterating along similar lines we obtain:

$$
\begin{aligned}
\widehat{\epsilon}_{N+2} &= \epsilon_{N+2} + (\varphi_1 - \theta_1)\epsilon_{N+1} \\
\widehat{\epsilon}_{N+3} &= \epsilon_{N+3} + (\varphi_1 - \theta_1)\epsilon_{N+3} + \varphi_1(\varphi_1 - \theta_1)\epsilon_{N+1}
\end{aligned}
$$

and so on.

Forecasting from ARMA models (cont.)

- The forecasts $\widehat{y}_{N+h}$ are unbiased and so the expected values of the forecast errors $\widehat{\epsilon}_{N+h}$ are zero.

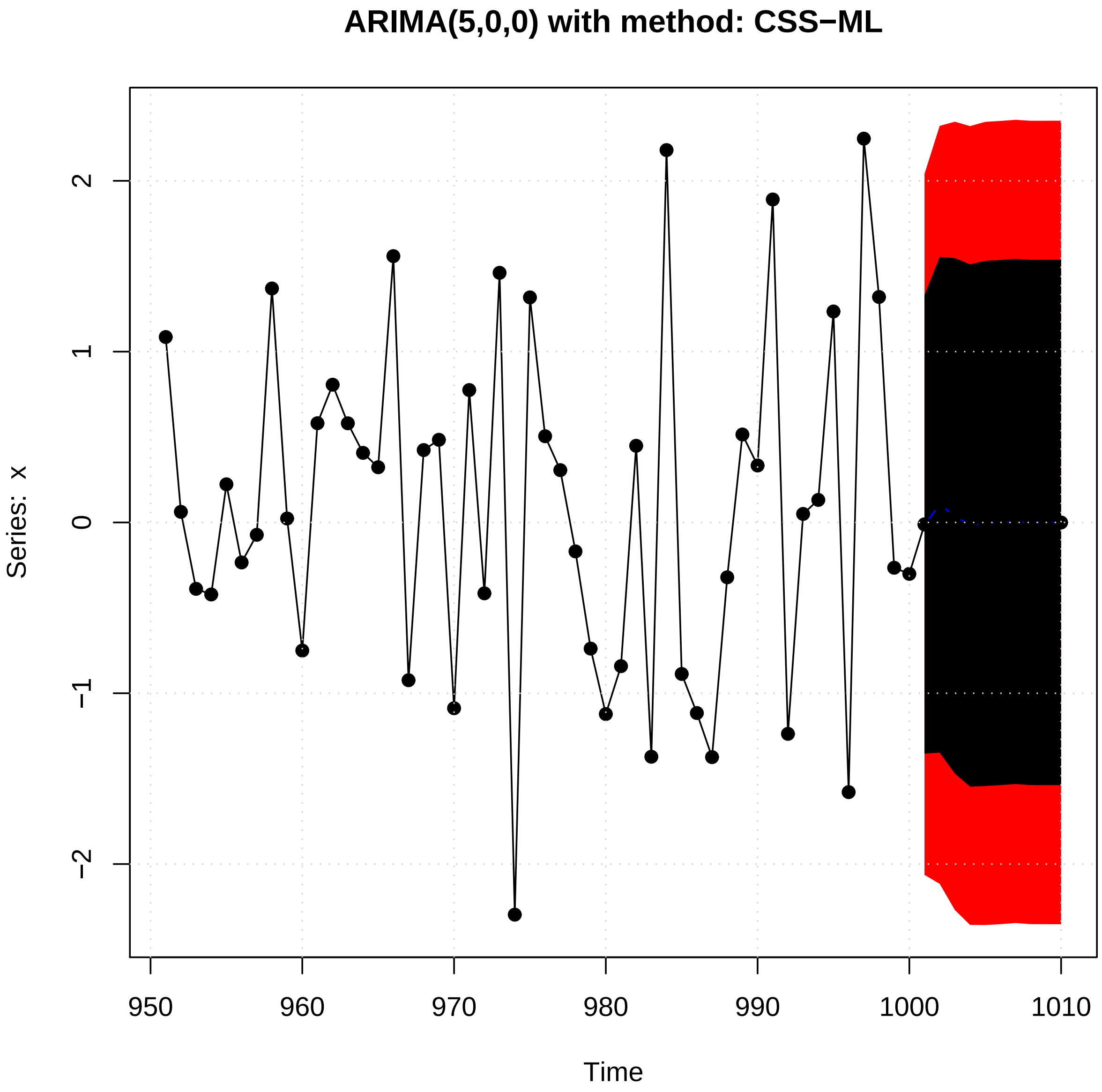
- The variance of the forecast error however increase with h.

- In the limit as h increases this variance converges to the unconditional variance of the process i.e. $\text{var}(y_t) = \sigma^2 = \gamma_0$.

- Clearly in practical forecasting from an ARMA(p,q) model the values of the parameters $(\varphi_1, \ldots, \varphi_p)$ and $(\theta_1, \ldots, \theta_q)$ will be unknown and these are replaced by their maximum likelihood estimates.

- Standard errors and confidence intervals for the forecasts may be derived from the general likelihood theory in the usual way.

See Ex. 6



ARIMA(5,0,0) with method: CSS−ML
Series: x
Time
950
960
970
980
990
1000
1010
−2
−1
0
1
2

Non–stationary processes

- Many time series encountered in practice may exhibit non-stationary behaviour. For example, there maybe non-stationarity in the mean component e.g. a time trend or seasonal effect in μ_t.

- We may think of this situation as the series consisting of a non-constant systematic (trend) component (usually some relatively simple function of time) and then a random component which is a zero-mean stationary series.

- Note that such a model is only reasonable if there are good reasons for believing that the trend is appropriate forever.

- There are several methods to eliminate trend and seasonal effects to generate stationary data.

ARIMA models

- Some types of time series the non-stationary behaviour of the mean μ_t is simple enough so that some differencing of the original series yields a new series which is stationary (so μ_t is constant).

- For example for financial time series (comprising log prices), first differencing (log returns) is often sufficient to produce a stationary time series with a constant mean.

- So the differenced series can be modelled directly by an ARMA process and no additional systematic component is required.

- This type of time series modelling where some degree of differencing is combined with an ARMA model is called Auto-regressive Integrated Moving Average (ARIMA) modelling.

ARIMA models (cont.)

- We have seen already that if the moduli of the roots of the characteristic equation of an ARMA(p,q) model lie inside the unit circle then the process will not be stationary.

- In general, if the modulus of a root is strictly inside the unit circle then this will lead to exponential or explosive behaviour in the series and no practical models result.

- If the modulus of the offending root lies on the circle then a more reasonable type of non-stationarity results. For example for the simple random walk

$$y_t = y_{t-1} + \epsilon_t.$$

  Note that the first difference of this series $y_t - y_{t-1}$ is a white noise process.

ARIMA models (cont.)

- This differencing idea can be generalised to the notion of using a model Y_t where the first difference of the process

$$X_t = (1 - B)Y_t = Y_{t-1} - Y_t$$

  is a stationary ARMA process, rather than white noise.

- More generally, if $d \geq 1$, Y_t is an ARIMA(p,d,q) process if $X_t = (1 - B)^d Y_t$ is an ARMA(p,q).

- An ARIMA(p,d,q) process Y_t satisfies:

$$\varphi^*(B)Y_t \equiv \varphi(B)(1 - B)^d Y_t = \theta(B)\epsilon_t,$$

  where $\varphi(z)$ and $\theta(z)$ are polynomials of degrees p and q, resp., and $\varphi(z) \neq 0$ for $|z| \leq 1$ and ϵ_t is a white noise process.

- An ARIMA model for a series y_t is one where a differencing operation on y_t leads to a series with stationary ARMA behaviour.

ARIMA models (cont.)

- A distinctive feature of the data which suggest the appropriateness of an ARIMA model is the slowly decaying positive sample ACF.

- Sample ACF with slowly decaying oscillatory behaviour are associated with models

$$\varphi^*(B)Y_t = \theta(B)\epsilon_t\,,$$

  in which φ^* has a zero near $e^{i\alpha}$ for some $\alpha \in (-\pi, \pi]$ other than $\alpha = 0$.

- In modeling using ARIMA processes the original series is simply differenced until stationarity is obtained and then the differenced series is modelled following the standard ARMA approach.

ARIMA models (cont.)

- Results may then be transformed back to the undifferenced original scale if required.

- Choice of an appropriate differencing parameter adds an extra dimension to model choice.

- For financial time series that have non-stationary behaviour, as mentioned earlier, first differencing (which leads to use of log returns), is usually sufficient to produce a time series with a stationary mean.

ARIMA models (summary)

- Plot the data to determine whether there is a trend. Of course this is only an indication, and what we see as a trend may be part of a very long–term circle.

- Use the sample ACF and PACF to determine whether it is possible to model the time series with an ARIMA model.

- Use differences to obtain an ARMA model.

- Model the differenced data using ARMA modelling.

See Ex. 7

ARCH and GARCH Modelling

- ARMA and ARIMA modelling is quite flexible and applicable. However, in some financial time series there are effects which cannot be adequately explained by these sorts of models.

- One particular feature is so called volatility clustering.

- This refers to a tendency for the variance of the random component to be large if the magnitude of recent ‘errors’ has been large and smaller if the magnitude of recent ‘errors’ has been small.

- This kind of behaviour requires non-stationarity in variance (i.e. heteroscedasticity) rather than in the mean

- This leads to alternative kinds of models to the ARIMA family which are referred to as ARCH and GARCH models.

ARCH and GARCH Modelling

- A dominant feature in many financial series is volatility clustering: The conditional variance of ϵ_t appears to be large if recent observations $\epsilon_{t-1}, \epsilon_{t-2}, ..$ are large in absolute value and small during periods where lagged innovations are also small in absolute value.

- This effect cannot be explained by ARIMA models which assume a constant variance.

- Autoregressive Conditionally Heteroscedastic(ARCH) models, (Engle 1982), were developed to model changes in volatility.

- These were extended to Generalised ARCH, or (GARCH) models (Bollerslev 1986).

ARCH Models

- Let x_t be the value of a stock at time t. The return, or relative gain, y_t, of the stock at time t is

$$y_t = \frac{x_t - x_{t-1}}{x_{t-1}}.$$

- Note, for financial series, return does not have a constant variance, with highly volatile periods tending to be clustered together – there is a strong dependence of sudden bursts of variability in a return on the time series' own past.

- Volatility models like ARCH, GARCH are used to study the returns y_t.

ARCH(1) Models

- The most simple ARCH model, the ARCH(1), models the return as

$$
\begin{aligned}
y_t &= \sigma_t \epsilon_t \\
\sigma_t^2 &= \omega + \alpha_1 y_{t-1}^2,
\end{aligned}
$$

  where $\epsilon_t \sim N(0,1)$.

- As with ARMA models, we impose constraints on the model parameters to obtain desirable properties: Sufficient conditions that guarantee $\sigma_t^2 > 0$ are $\omega > 0, \alpha_1 \geq 0$.

ARCH(1)(Properties)

- Conditionally on y_{t-1}, y_t is Gaussian: $y_t | y_{t-1} \sim N\left(0, \omega + \alpha_1 y_{t-1}^2\right)$.

- The returns $\{y_t\}$ have zero mean and they are uncorrelated.

- The squared returns $\left\{y_t^2\right\}$ satisfy:

$$y_t^2 = \omega + \alpha_1 y_{t-1}^2 + v_t,$$

  where the error process $v_t = \sigma_t^2 \left(\epsilon_t^2 - 1\right)$ is a white noise process.

- Hence

  – ARCH(1) models returns $\{y_t\}$ as a white noise process with non-constant conditional variance, and the conditional variance depends on the previous return.

  – the returns $\{y_t\}$ are uncorrelated, whereas their squares $\left\{y_t^2\right\}$ follow a non-Gaussian autoregressive process.

ARCH(1) Models (cont.)

- Moreover, the kurtosis of y_t is

$$\kappa = \frac{E\left[y_t^4\right]}{E\left[y_t^2\right]^2} = 3\frac{1 - \alpha_1^2}{\left(1 - 3\alpha_1^2\right)}$$

  which is always larger than 3, the kurtosis of the normal distribution.

- Thus, the marginal distribution of the returns, y_t, is leptokurtic, or has heavy tails.

- So outliers are more likely. This agrees with empirical evidence - outliers appear more often in asset returns than implied by an i.i.d sequence of normal random variates.

ARCH(1) Models (cont.)

- Estimation of the parameters ω and α_1 of the ARCH(1) model is accomplished using conditional MLE.

- The likelihood of the data $y_2, ..., y_n$ conditional on y_1, is given by

$$L(\omega, \alpha_1 | y_1) = \prod_{t=2}^{n} f_{\omega,\alpha_1}(y_t | y_{t-1}),$$

  where

$$f_{\omega,\alpha_1}(y_t | y_{t-1}) \sim N\left(0, \omega + \alpha_1 y_{t-1}^2\right),$$

  that is

$$f_{\omega,\alpha_1}(y_t | y_{t-1}) \propto \frac{1}{\left(\omega + \alpha_1 y_{t-1}^2\right)^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\left(\frac{y_t^2}{\omega + \alpha_1 y_{t-1}^2}\right)\right].$$

ARCH(1) Models (cont.)

- Hence, the objective function to be maximised is the conditional log-likelihood

$$l\left(\omega,\alpha_1|y_1\right)=\ln\left[L\left(\omega,\alpha_1|y_1\right)\right]\propto-\frac{1}{2}\sum_{t=2}^{n}\ln\left(\omega+\alpha_1y_{t-1}^2\right)-\frac{1}{2}\sum_{t=2}^{n}\left(\frac{y_t^2}{\omega+\alpha_1y_{t-1}^2}\right).$$

- Maximisation of this function is achieved using numerical methods (analytic expressions for the gradient vector and Hessian matrix of the log-likelihood functions can be obtained).

ARCH(m) Models (cont.)

- The general ARCH(m) model is defined by:

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \omega + \alpha_1 y_{t-1}^2 + \ldots + \alpha_m y_{t-m}^2, \end{aligned}$$

  where the parameter m determines the maximum order of lagged innovations which are supposed to have an impact on current volatility.

- Similar results to those from the ARCH(1) model hold:

$$y_t | y_{t-1}, \ldots, y_{t-m} \sim N\left(0, \omega + \alpha_1 y_{t-1}^2 + \ldots + \alpha_m y_{t-m}^2\right),$$

$$y_t^2 = \omega + \alpha_1 y_{t-1}^2 + \alpha_1 y_{t-1}^2 + \ldots + \alpha_m y_{t-m}^2 + v_t,$$

  where $v_t = \sigma_t^2 \left(\epsilon_t^2 - 1\right)$ is a shifted χ_1^2 random variable.

- y_t and v_t have a zero mean.

- Estimation of the parameters $\omega, \alpha_1, \ldots, \alpha_m$ is similar to that for ARCH(1)

Building ARCH models

- An ARIMA model is built for the observed time series to remove any serial correlation in the data.

- Examine the squared residuals to check for conditional heteroscedasticity.

- Use the PACF of squared residuals to determine the ARCH order.

As final remarks we should comments on some of the weaknesses:

- ARCH treats positive and negative returns in the same way (by past square returns).

- ARCH often over-predicts the volatility, because it responds slowly to large shocks.

GARCH(m,r) models

- Generalised ARCH models, GARCH (m,r) process (Boyerslev, 1982) are obtained by augmenting σ_t^2 with a component autoregressive in σ_t^2.

- For instance, a GARCH(1,1) model is

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \omega + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \end{aligned}$$

- Assuming $\alpha_1 + \beta_1 < 1$ and using similar manipulations as before, it can be shown that the GARCH(1,1) model admits a non-Gaussian ARMA(1,1) model for the squared process.

- Indeed:

$$\begin{aligned} y_t^2 &= \sigma_t^2 \epsilon_t^2 \\ \omega + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2 &= \sigma_t^2 \end{aligned}$$

GARCH(m,r) models (cont)

It can be seen that:

$$y_t^2 - \sigma_t^2 = \sigma_t^2 \left(\epsilon_t^2 - 1\right) = v_t$$

$$\Longrightarrow \; y_{t-1}^2 - \sigma_{t-1}^2 = \sigma_{t-1}^2 \left(\epsilon_{t-1}^2 - 1\right) = v_{t-1}$$

and then

$$y_t^2 - \omega - \alpha_1 y_{t-1}^2 - \beta_1 \sigma_{t-1}^2 = v_t$$

$$\Longrightarrow \; y_t^2 = \omega + \alpha_1 y_{t-1}^2 + \beta_1 y_{t-1}^2 + \beta_1 \left(\sigma_{t-1}^2 - y_{t-1}^2\right) + v_t$$

and so

$$y_t^2 = \omega + (\alpha_1 + \beta_1)\, y_{t-1}^2 - \beta_1 v_{t-1} + v_t.$$

GARCH(m,r) models (cont)

In general, the GARCH (m,r) model is

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \omega + \alpha_1 y_{t-1}^2 + \ldots + \alpha_m y_{t-m}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_r \sigma_{t-r}^2. \end{aligned}$$

Sufficient conditions for the conditional variance to be positive are obvious:

$$\omega > 0, \alpha_i, \beta_j \geq 0, i = 1, \ldots, m; j = 1, \ldots, r.$$

Using polynomials in the lag B, the specification of σ_t^2 may also be given by

$$\left(1 - \beta_1 B - \ldots - \beta_r B^r\right) \sigma_t^2 = \omega + \left(\alpha_1 B + \ldots + \alpha_m B^r\right) y_t$$

or

$$\left(1 - \beta\left(B\right)\right) \sigma_t^2 = \omega + \alpha\left(B\right) y_t.$$

GARCH(m,r) models (cont)

Assuming the zeros of the polynomial $(1-\beta(z))$ are larger than one in absolute value, the model can also be written as an ARCH process of infinite order:

$$\sigma_t^2 = (1-\beta(B))^{-1}\omega + (1-\beta(B))^{-1}\alpha(B)\, y_t.$$

Note that a GARCH(m,r) admits a non-Gaussian ARMA(m,r) model for the squared process:

$$y_t^2 = \omega + \sum_{i=1}^{\max(m,r)} (\alpha_i + \beta_i)\, y_{t-i}^2 + v_t - \sum_{i=1}^{r} \beta_i y_{t-i}^2.$$

Building and fitting GARCH models follows similarly to that discussed previously for ARCH models.

See Ex. 8