STATISTICAL METHODS WITH APPLICATION TO FINANCE

a.y. 2025-2026

Linear Models

R. PAPPADÀ
rpappada@units.it

Department of Economics, Business, Mathematics and Statistics "B. de Finetti"
University of Trieste

April 1, 2026

Linear processes

Time series models with correlation can be constructed from white noise. Recall that

- we refer to a white noise series $\{x_t\}$ as a sequence of *uncorrelated* random variables with finite mean and variance and a well-defined distribution
- if $\{x_t\}$ is normally distributed with mean 0 and variance σ^2, the series is called a Gaussian white noise
- For a white noise series, all the theoretical ACFs are 0. In practice, if all sample ACFs are close to 0, then the series is a white noise series

Linear time series

Definition (Linear process)

A linear process $\{X_t\}$ is defined to be a linear combination of white noise variables a_t:

$$X_t = \mu + \psi_0 a_t + \psi_1 a_{t-1} + \cdots = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i} \qquad (1)$$

where μ is the mean of the process, $\psi_0 = 1$ and $\{a_t\} \sim \text{WN}(0, \sigma_a^2)$

The term “linear” is a reference to the summation of the coefficients ψ_i, each multiplied by a single white noise “error”.

The innovations

The “error” a_t (also called the *innovation* at time t) represents the effect of *new information* at time t

- if $\{a_t\} \sim \text{WN}(0, \sigma_a^2)$, then the effects of today’s new information is assumed to be independent of the effects of yesterday’s news
- a time series is linear if it can be written as a linear combination of all past innovations
- linear models can often provide accurate approximations in real applications

Stationarity

For the linear process in (1), *weak stationarity* requires that the weights on more remote past values are "small"

$$\mathsf{E}(X_t) = \mu, \qquad \mathsf{Var}(X_t) = \sigma_a^2 \sum_{i=0}^{\infty} \psi_i^2$$

and $\mathsf{Var}(X_t) < \infty$, so that $\{\psi_i^2\}$ must be a convergent sequence:

$$\psi_i^2 \to 0 \text{ as } i \to \infty$$

$\Rightarrow$ for a stationary series, the impact of the remote *innovation* a_{t-i} on X_t vanishes as i increases

Table of Contents

1. White noise and linear time series
2. Autoregressive models
   - AR(1) Model
   - AR(p) processes
   - Properties of AR Models
   - Nonstationary AR processes
3. Moving average models
   - MA(1) model
   - MA(q) processes
   - Properties of MA Models
4. Identifying AR and MA Models
   - Order determination
   - Information Criteria
   - Model Checking
5. Appendix

Definition

$\{X_t\}$ is an **autoregressive process of order 1**, AR(1), if

$$X_t = \phi_0 + \phi_1 X_{t-1} + a_t \quad (2)$$

where $\{a_t\} \sim \text{WN}(0, \sigma_a^2)$.

- *autoregression*: regression of the process on its own past values
- $\{X_t\}$ is correlated and the parameter ϕ_1 determines the amount of "feedback" of the past into the present
  - $\phi_1 = 0$ implies $X_t = \phi_0 + a_t$, so that X_t is weak WN(ϕ_0, σ_a^2)
- The random walk (with drift ϕ_0) is obtained as special case AR(1) for $\phi_1 = 1$.

Properties of AR(1)

Assuming that the series is weakly stationary, we have

$$\mathsf{E}(X_t) = \mu, \quad \mathsf{Var}(X_t) = \gamma_0$$

and $\mathsf{Cov}(X_t, X_{t+h}) = \gamma_h$, where μ and γ_0 are constants and γ_0 is a function of h, not t. Taking the expectation of Eq.(2), and using $\mathsf{E}(a_t) = 0$ we obtain the **mean**

$$\mathsf{E}(X_t) = \phi_0 + \phi_1 E(X_{t-1}) \tag{3}$$

Under the stationarity condition, $E(X_{t-1}) = E(X_t) = \mu$, and hence

$$\mathsf{E}(X_t) = \mu = \frac{\phi_0}{1-\phi_1}, \text{ for } \phi_1 \neq 1$$

The **intercept** is

$$\phi_0 = \mu(1-\phi_1)$$

and $E(X_t) = 0$ if and only if $\phi_0 = 0$.

Properties of AR(1)

Using $\phi_0 = \mu(1 - \phi_1)$, one can rewrite the AR(1) model as

$$X_t = \mu(1 - \phi_1) + \phi_1 X_{t-1} + a_t$$

that is

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + a_t \tag{4}$$

- The parameter μ is the mean of the process, hence $(X_t - \mu)$ has mean zero for all t
- $\phi_1(X_{t-1} - \mu)$: "memory" or "feedback" of the past into the present value of the process, ϕ_1 measuring the persistence of the dynamic dependence of an AR(1) process

Linearity

By repeated substitutions, Eq.(4) implies

$$X_t - \mu = a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \ldots$$

that is

$$X_t = \mu + \sum_{i=0}^{\infty} \phi_1^i \, a_{t-i}$$

$\Rightarrow X_t$ is a linear function of a_{t-i} for $i \geq 0$, that is, we can express an AR(1) model in the form of Eq.(1) with $\psi_i = \phi_1^i$.

Stationarity condition

The necessary and sufficient condition for the AR(1) model in Eq.(2) to be weakly stationary is $|\phi_1| < 1$.

Under stationarity

$$\text{Var}(X_{t-1}) = \text{Var}(X_t) = \gamma_0 \tag{5}$$

Then from Eq.(4), and using $\text{Cov}(X_{t-1}, a_t) = 0$

$$\begin{aligned}\text{Var}(X_t) &= \text{Var}[\phi_1(X_{t-1} - \mu) + a_t] \\ &= \phi_1^2 \text{Var}(X_{t-1} - \mu) + \text{Var}(a_t) \\ &= \phi_1^2 \text{Var}(X_{t-1}) + \sigma_a^2\end{aligned}$$

Using (5) we obtain

$$\gamma_0 = \phi_1^2 \gamma_0 + \sigma_a^2 \quad \Rightarrow \quad \boxed{\text{Var}(X_t) = \gamma_0 = \frac{\sigma_a^2}{1 - \phi_1^2}}$$

provided that $\phi_1^2 < 1 \Rightarrow -1 < \phi_1 < 1$

Autocorrelation Function

For a weakly stationary AR(1) model in Eq.(2) the ACF is

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1^k, \quad k \geq 0$$

where $|\phi_1| < 1$: the correlogram decays to zero more rapidly for small ϕ_1. Thus

$$\rho_0 = 1$$
$$\rho_1 = \phi_1$$
$$\rho_2 = \phi_1^2$$
$$\vdots$$

and the ACF satisfies the recursion

$$\rho_k = \phi_1 \rho_{k-1} \quad k = 1, 2, \ldots$$

- $\phi_1 > 0$: observations close together in time are positively correlated
- $\phi_1 < 0$: the ACF is exponentially decaying with alternating sign

ACF of stationary AR(1): Example

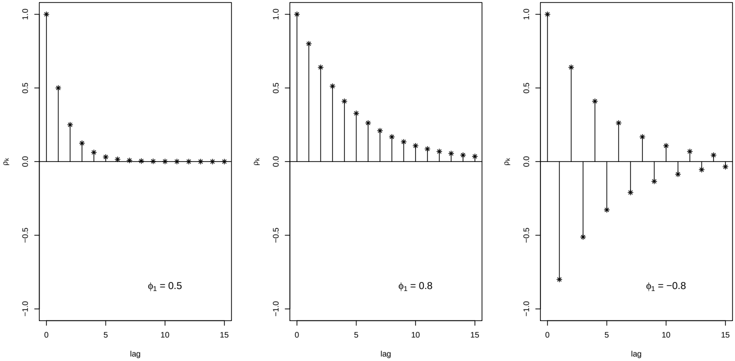


Figure 1: Theoretical ACF for three AR(1) models with mean zero: **(left)** $x_t = 0.5x_{t-1} + a_t$; **(middle)** $x_t = 0.8x_{t-1} + a_t$; **(right)** $x_t = -0.8x_{t-1} + a_t$.

AR(1): Simulated data

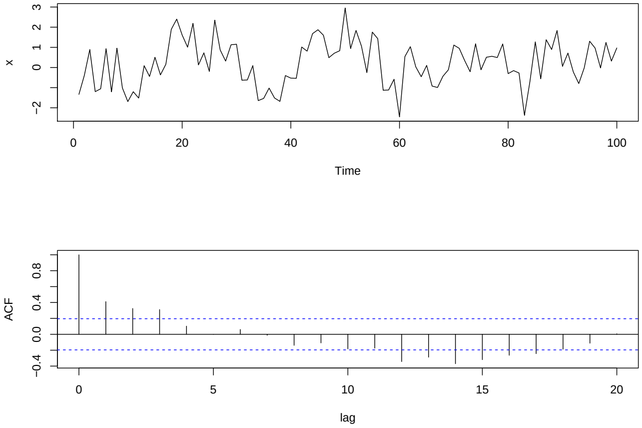


Figure 2: Time plot and ACF plot from one possible realisation of the AR(1) process $x_t = 0.5x_{t-1} + a_t$ (compare the ACF plot with Fig.1-left).

Table of Contents

1. White noise and linear time series
2. Autoregressive models
   - AR(1) Model
   - AR(p) processes
   - Properties of AR Models
   - Nonstationary AR processes
3. Moving average models
   - MA(1) model
   - MA(q) processes
   - Properties of MA Models
4. Identifying AR and MA Models
   - Order determination
   - Information Criteria
   - Model Checking
5. Appendix

The AR(p) model

The AR(1) model provides a limited range of behavior. We can use a model that regresses the current value of the process on several of the recent past values, say the last p values, not just the most recent.

The **autoregressive process of order p**, AR(p), is expressed as

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + a_t \tag{6}$$

where $\{a_t\}$ is white noise and the ϕ_i are the model parameters, with $\phi_p \neq 0$.

- similar to a multiple linear regression model with lagged values of the time series as the “x-variables”

The AR(p) model

The model can also be expressed as

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \cdots + \phi_p(X_{t-p} - \mu) + a_t \qquad (7)$$

where

- the mean is
$$\mathsf{E}(X_t) = \mu = \frac{\phi_0}{1 - (\phi_1 + \cdots + \phi_p)}$$
- $\phi_0 = \mu[1 - (\phi_1 + \cdots + \phi_p)]$ is called the “constant” or “intercept”
- $\mu = 0$ if and only if $\phi_0 = 0$

Example: $p = 2$

An **AR(2)** model assumes the form

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t \qquad (8)$$

or $X_t - \mu = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + a_t$ where $\mu = \dfrac{\phi_0}{1 - (\phi_1 + \phi_2)}$, provided that $\phi_1 + \phi_2 \neq 1$.

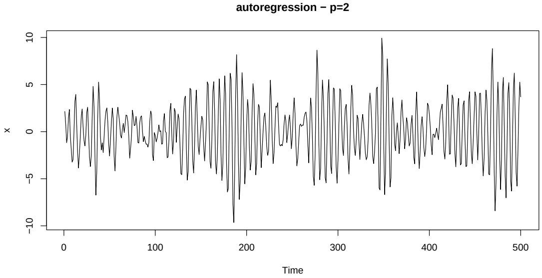


Figure 3: A simulated AR(2) process $x_t = x_{t-1} - 0.9x_{t-2} + a_t$ with parameters $\phi_1 = 1$ and $\phi_2 = -0.9$ and $\{a_t\}$ Gaussian white noise.

Table of Contents

1. White noise and linear time series
2. Autoregressive models
   - AR(1) Model
   - AR(p) processes
   - Properties of AR Models
   - Nonstationary AR processes
3. Moving average models
   - MA(1) model
   - MA(q) processes
   - Properties of MA Models
4. Identifying AR and MA Models
   - Order determination
   - Information Criteria
   - Model Checking
5. Appendix

Stationarity

Given the AR(p) model in Eq.(6), the equation

$$1 - \phi_1 z - \phi_2 z^2 - \dots \phi_p z^p = 0 \tag{9}$$

is called the **characteristic equation**. The roots of the characteristic equation must *all* exceed unity in *absolute value* for the process to be stationary:

- for the AR(1) Eq.(9) gives

  $$1 - \phi_1 z = 0$$

  whose solution $z_0 = 1/\phi_1$ must satisfy $|1/\phi_1| > 1$, that is $|\phi_1| < 1$
- for the AR(2) Eq.(9) gives

  $$1 - \phi_1 z - \phi_2 z^2 = 0$$

  Real-valued solutions of this equation are $z_{1,2} = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}$ and the condition is $|z_1| > 1$ and $|z_2| > 1$.

Stationarity: Examples

- The AR(1) model $x_t = \frac{1}{2}x_{t-1} + a_t$ is stationary because the root of

$$1 - \frac{1}{2}z = 0$$

  is $z = 2$, which is greater than 1.

- The AR(2) model $x_t = x_{t-1} - \frac{1}{4}x_{t-2} + a_t$ is stationary:

$$1 - z + \frac{1}{4}z^2 = 0 \to \frac{1}{4}(z^2 - 4z + 4) = \frac{1}{4}(z-2)^2 = 0$$

  the roots are therefore obtained as $z_1 = z_2 = 2$.

Unit-Root Nonstationarity

Exercise

The model

$$x_t = \frac{1}{2}x_{t-1} + \frac{1}{2}x_{t-2} + a_t$$

is non-stationary because one of the roots is unity.

Notice that the random walk

$$x_t = x_{t-1} + a_t$$

has characteristic equation $1 - z = 0$ with root $z = 1$ and it is non-stationary.

ACF of stationary AR(2): Example

A wide variety of ACFs are possible with two AR parameters

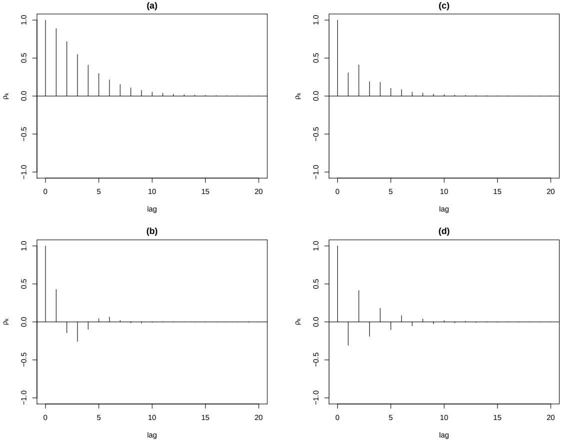


Figure 4: (a) $\phi_1 = 1.2$ and $\phi_2 = -0.35$, (b) $\phi_1 = 0.6$ and $\phi_2 = -0.4$, (c) $\phi_1 = 0.2$ and $\phi_2 = 0.35$, and (d) $\phi_1 = -0.2$ and $\phi_2 = 0.35$

Table of Contents

1. White noise and linear time series
2. Autoregressive models
   - AR(1) Model
   - AR(p) processes
   - Properties of AR Models
   - Nonstationary AR processes
3. Moving average models
   - MA(1) model
   - MA(q) processes
   - Properties of MA Models
4. Identifying AR and MA Models
   - Order determination
   - Information Criteria
   - Model Checking
5. Appendix

Nonstationarity

In some studies, interest rates, foreign exchange rates, or the price series of an asset are of interest. These series tend to be nonstationary.

- When $p = 1$, we know that the process is stationary if $|\phi_1| < 1$, which is equivalent to $|1/\phi_1| > 1$
- If $|\phi_1| \geq 1$, the **AR(1) process is nonstationary**: mean, variance, covariances and and correlations are not constant
- When $|\phi_1| > 1$, an AR(1) process has explosive behavior
- If $|\phi_1| = 1$ the model is said to be unit-root nonstationary because its AR polynomial has a unit root (Go to Appendix)

Nonstationarity and ACF

The ACF plot can help in identifying non-stationary time series:

- For a stationary time series, the sample ACF will drop to zero relatively quickly
- the sample ACF of non-stationary or “nearly non-stationary” processes will both decay (very) slowly towards zero

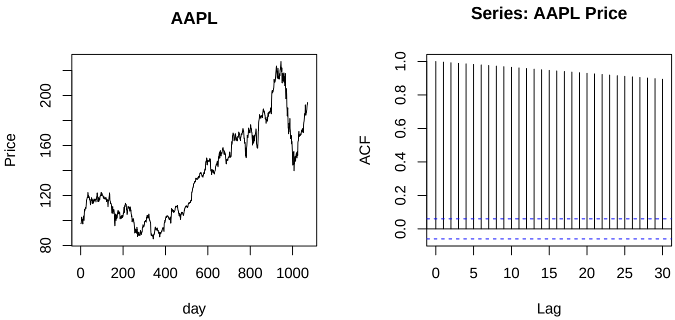


Figure 5: Time plot (left) and ACF (right) of daily closing prices of Apple stock from January 3, 2009 to April 5, 2013.

(Near) nonstationary behavior

A process can be "nearly nonstationary" if it is an ordinary AR process with roots near to the unit circle

Example: Consider some data generated from the AR(1) model with $\phi_1 = 0.95$

$$Y_t = 0.95 Y_{t-1} + a_t$$

The time plot and sample ACF will exhibit properties that are similar to those of data generated by a non-stationary *random walk*

$$Y_t = Y_{t-1} + a_t$$

The distinction between the two types of processes is sometimes difficult but relevant to forecasting issues.

Unit Root Testing

Assume we want to investigate whether the log price series from an asset follows a random walk

The first step is to look at the correlogram of the differenced series.

In addition, a formal test can be used to check whether a time series model that contains a unit root is suitable.

In general, **unit root tests** are used to verify the existence of a unit root in an AR(p) process. For instance, the null hypothesis of the *Augmented Dickey-Fuller* test is that there is a unit root; the alternative is that the process is stationary

From unit-root nonstationarity to stationarity

If the original data are non-stationary but the first differences are stationary, then a **unit root** is said to be present. The best-known example of unit-root nonstationary time series is the random walk model.

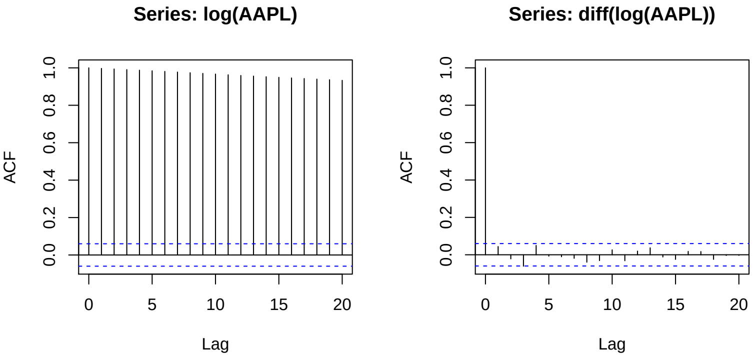


Figure 6: ACF of log prices (left) and of first differences of the logarithms (right) for the Apple stock price series.

Table of Contents

1. White noise and linear time series
2. Autoregressive models
   - AR(1) Model
   - AR(p) processes
   - Properties of AR Models
   - Nonstationary AR processes
3. Moving average models
   - MA(1) model
   - MA(q) processes
   - Properties of MA Models
4. Identifying AR and MA Models
   - Order determination
   - Information Criteria
   - Model Checking
5. Appendix

Introduction

- The idea behind AR processes is to feed past data back into the current value of the process, inducing some correlation at *all* lags;
- A useful alternative is a moving average (MA) model;
- MA models are useful in modeling asset returns in finance;
- In some cases there is a potential need for large values of p when fitting AR processes: a possibility is to add a moving average component to an AR process (**ARMA models**).

Definition

There are several ways to introduce MA models. One approach is to treat the model as a simple extension of white noise series.

The MA(1) (**moving average of order 1**) process is

$$X_t = a_t + \theta_1 a_{t-1} \tag{10}$$

where $\theta_1 \neq 0$ and the $\{a_t\}$ is white noise with zero mean and variance σ_a^2.

We may also write the MA(1) process with *intercept* μ in the form

$$X_t = \mu + a_t + \theta_1 a_{t-1} \tag{11}$$

Properties

Because MA processes consist of a finite sum of stationary white noise terms, they have a time-invariant mean and autocovariance.

For the MA(1) model in Eq.(11), the **mean** and **variance** functions are :

$$
\begin{aligned}
\mathsf{E}(X_t) &= \mathsf{E}(\mu + a_t + \theta_1 a_{t-1}) = \mu, \\
\mathsf{Var}(X_t) &= \mathsf{Var}(\mu + a_t + \theta_1 a_{t-1}) \\
&= \mathsf{Var}(a_t) + \mathsf{Var}(\theta_1 a_{t-1}) \\
&= \sigma_a^2 + \theta_1^2 \sigma_a^2 = \sigma_a^2(1 + \theta_1^2)
\end{aligned}
$$

where we use the fact that a_t and a_{t-1} are uncorrelated.

Autocovariance function

Consider the MA(1) model $X_t = a_t + \theta_1 a_{t-1}$ (assume for simplicity that $\mu = 0$). Then, the lag-k autocovariance is

$$\gamma(k) = \text{Cov}(X_t, X_{t+k}) = \text{Cov}(a_t + \theta_1 a_{t-1}, a_{t+k} + \theta_1 a_{t+k-1})$$

and using $\text{Cov}(a_t, a_{t+k}) = 0$ for all $k \neq 0$,

$$\gamma(k) = \begin{cases} \text{Cov}(X_t, X_t) = \sigma_a^2(1 + \theta_1^2) & k = 0 \\ \text{Cov}(a_t + \theta_1 a_{t-1}, a_{t+1} + \theta_1 a_t) = \sigma_a^2 \theta_1 & k = 1 \\ 0 & k > 1 \end{cases}$$

The moving average process is **stationary** for any value of the parameter θ_1.

MA(1): Simulated data

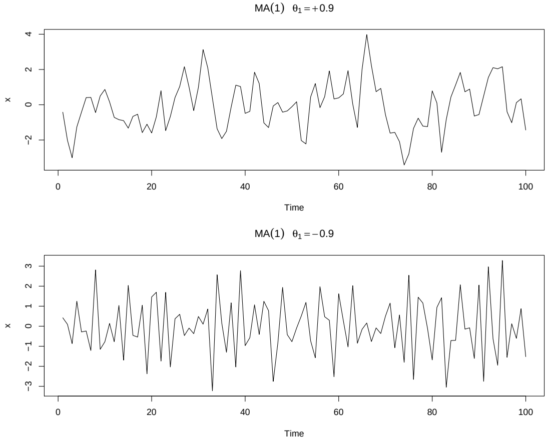


Figure 7: Simulated MA(1) models: $\theta_1 = 0.9$ (top) and $\theta_1 = -0.9$ (bottom).

Table of Contents

1. White noise and linear time series
2. Autoregressive models
   - AR(1) Model
   - AR(p) processes
   - Properties of AR Models
   - Nonstationary AR processes
3. Moving average models
   - MA(1) model
   - MA(q) processes
   - Properties of MA Models
4. Identifying AR and MA Models
   - Order determination
   - Information Criteria
   - Model Checking
5. Appendix

Definition

A **moving average process of order q**, MA(q), is a linear combination of the current white noise term and the q most recent past white noise terms and is defined by

$$X_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q} \quad (12)$$

where $a_t \sim \mathrm{WN}(0, \sigma_a^2)$ and $\theta_1, \theta_2, \ldots, \theta_q$ $(\theta_q \neq 0)$ are parameters.

Note that $\mathrm{E}(a_t) = \cdots = \mathrm{E}(a_{t-q}) = 0$. We may also write the MA(q) process with intercept μ in the form

$$X_t = \mu + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q}$$

The intercept μ is also the **mean** of X_t.

Table of Contents

1 White noise and linear time series

2 Autoregressive models
- AR(1) Model
- AR(p) processes
- Properties of AR Models
- Nonstationary AR processes

3 Moving average models
- MA(1) model
- MA(q) processes
- Properties of MA Models

4 Identifying AR and MA Models
- Order determination
- Information Criteria
- Model Checking

5 Appendix

Stationarity

The prior discussion applies to general MA(q) models, and we obtain two general properties:

- Unlike AR(p) models, the constant term of an MA model is the **mean** of the series

$$\mathrm{E}(X_t) = \mu$$

- the **variance** of an MA(q) model is

$$\mathrm{Var}(X_t) = \sigma_a^2(1 + \theta_1^2 + \cdots + \theta_q^2)$$

because each of the white noise terms has the same variance and the terms are mutually independent

Autocorrelation function

Autocorrelation Function. MA models are only linearly related to the first q lagged values.

▶ For the MA(1) model $X_t = a_t + \theta_1 a_{t-1}$ the ACF is

$$\rho_k = \begin{cases} \dfrac{\theta_1}{1+\theta_1^2} & k = 1 \\ 0 & k > 1 \end{cases}$$

Of course, $\rho_0 = 1$. Hence the MA(1) model has zero correlation at all lags except lag 1. In other words, the ACF of the MA(1) model *cuts off* at lag 1.

What are autocorrelations for the examples in Fig.7?

▶ For the MA(q) the autocorrelation is zero when $k > q$ because x_t and x_{t+k} then consist of sums of independent white noise terms and so have covariance zero: a MA(q) process is said to *cut off* after lag q.

Autocorrelation function: Example

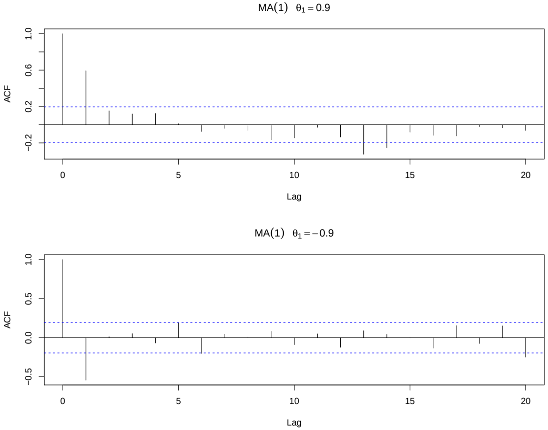


Figure 8: ACF for the series in Fig.7: x_t is correlated with x_{t-1}, but not with $x_{t-2}, x_{t-3}, \ldots$; when $\theta_1 = 0.9$ x_t and x_{t-1} are positively correlated, when $\theta_1 = -0.9$ x_t and x_{t-1} are negatively correlated.

Autocorrelation function: Example

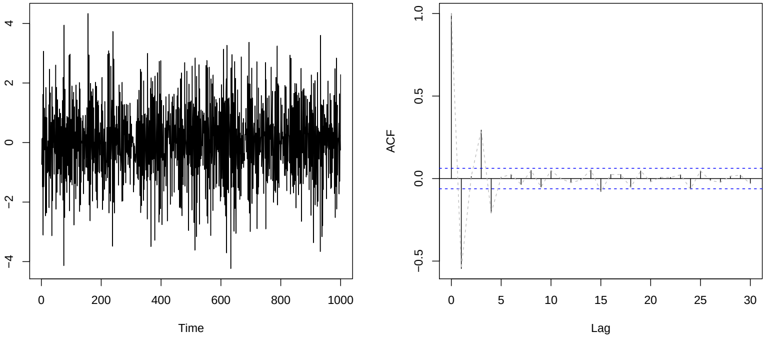


Figure 9: A realization of a MA(4) process and its ACF, with $\mu = 0$, $\theta_1 = -0.8$, $\theta_2 = 0.4$, $\theta_3 = 0.2$, $\theta_4 = -0.3$. Note that lag-four coefficient is significantly different from 0, while higher-order ACFs are 0.

Linearity

A MA(q) process

$$X_t = \mu + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q}$$

can be written as the linear process

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j \, a_{t-j}$$

with weights

- $\psi_0 = 1$
- $\psi_j = \theta_j, j = 1, \dots q$
- $\psi_j = 0, j > q$.

Invertibility

■ **Invertibility.** Consider a zero-mean MA(1) model $x_t = a_t + \theta_1 a_{t-1}$. We can rewrite it as

$$a_t = x_t - \theta_1 a_{t-1}$$

By repeated substitutions we obtain

$$a_t = x_t - \theta_1 x_{t-1} + \theta_1^2 x_{t-2} - \theta_1^3 x_{t-3} + \dots \tag{13}$$

Eq.(13) expresses the current shock at as a linear combination of the present and past values of x_t

- θ^j should go to 0 as j increases because the remote return x_{t-j} should have very little impact on the current shock, if any
- for an MA(1) model to be plausible, we require

$$|\theta_1| < 1$$

Such an MA(1) model is said to be invertible. If $|\theta_1| = 1$, then the MA(1) model is noninvertible.

Invertibility: Example

Consider the following first-order MA processes

$$A: \quad X_t = a_t + \theta a_{t-1}$$

$$B: \quad X_t = a_t + \frac{1}{\theta} a_{t-1}$$

These two different processes have exactly the same ACF (check it yourself!). Thus we cannot identify an MA process uniquely from a given ACF.

If $|\theta| < 1$, then model A is said to be *invertible* whereas model B is not $\Rightarrow$ The imposition of the invertibility condition ensures that there is a unique MA process for a given ACF.

Invertibility: MA(q) models

It can be shown that an MA(q) process is invertible if the roots of the equation

$$1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q = 0$$

all have modulus **greater than unity**. For example, in the case $q = 1$ we have that

$$1 + \theta_1 z = 0$$

has root $z = -1/\theta_1$. The process is then invertible if $\left| -\frac{1}{\theta_1} \right| > 1$, that is $|\theta_1| < 1$.

Table of Contents

1. White noise and linear time series
2. Autoregressive models
   - AR(1) Model
   - AR(p) processes
   - Properties of AR Models
   - Nonstationary AR processes
3. Moving average models
   - MA(1) model
   - MA(q) processes
   - Properties of MA Models
4. Identifying AR and MA Models
   - Order determination
   - Information Criteria
   - Model Checking
5. Appendix

ACF and Partial Autocorrelation

The ACF plot can be useful for determining the order of a moving average process:

- if $\rho_q \neq 0$ but $\rho_k = 0$ for $k > q$, then x_t follows an MA(q) model

For autoregressive processes, the ACF alone tells us little about the order of dependence. For instance, for the AR(1) model

$$\rho_k = \phi_1^k \quad (k \geq 0)$$

so that the autocorrelations are non-zero for all lags even though in the underlying model x_t only depends on the previous value x_{t-1}.

The **partial autocorrelation function (PACF)** of a stationary time series is a function of its ACF and can be useful for identifying the order of an AR process.

The Partial Autocorrelation Function

The k-th partial autocorrelation is the correlation that results after *removing the effect* of any correlations due to the terms at shorter lags.

Consider the AR(1) model

$$x_t = \phi_0 + \phi_1 x_{t-1} + a_t \tag{14}$$

For $k = 1$, the partial autocorrelation coefficient is simply equal to the autocorrelation coefficient

$$\phi_{1,1} = \rho_1 = \phi_1$$

Let $\hat{\phi}_{k,k}$ denote the estimate of $\phi_{k,k}$. $\hat{\phi}_{k,k}$ (lag-k sample PACF) can be calculated by fitting the regression model

$$Y_t = \phi_{0,k} + \phi_{1,k} Y_{t-1} + \cdots + \phi_{k,k} Y_{t-k} + a_{k,t}$$

lag-1 sample PACF of x_t is the estimate $\hat{\phi}_1$.

The Partial Autocorrelation Function (cont)

- lag-2 sample PACF of x_t shows the added contribution of x_{t-2} to x_t over the model (14)
- lag-3 sample PACF represents the added contribution of x_{t-3} to x_t over the AR(2) model $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + a_t$ and so on.

If Y_t is an AR(p) process, then

$$\phi_{k,k} = 0 \text{ for } k > p.$$

For instance, the partial ACF of an AR(1) process will be zero for all lags greater than 1.

It follows that the PACF plot can be useful when determining the order of the AR process

- a sign that a time series can be fit by an AR(p) model is that the sample PACF will be nonzero up to p and then will be nearly zero for larger lags

AR(1) PACF: Example

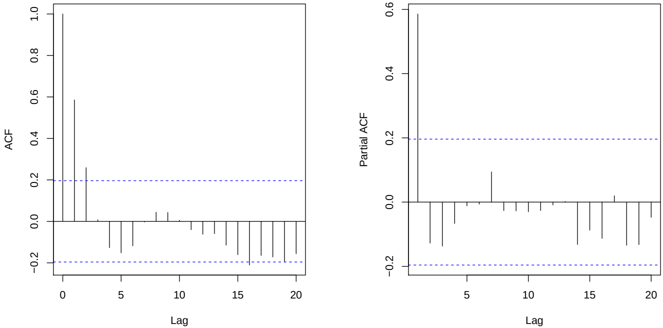


Figure 10: ACF and PACF for a simulated AR(1) process, $x_t = 0.7x_{t-1} + a_t$. Note that in the partial correlogram only the first lag is significant, which is usually the case when the underlying process is AR(1).

AR(2) PACF: Example

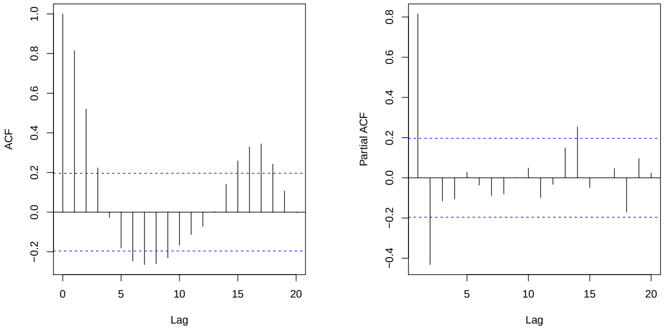


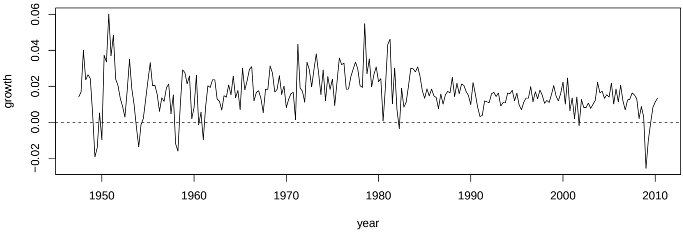
Figure 11: The ACF and PACF of an AR(2) model with $\phi_1 = 1.5$ and $\phi_2 = -0.75$.

Order determination: US Gross National Product

Consider the *quarterly growth rate* of US gross national product (**GNP**) from the second quarter of 1947 to the first quarter of 2010 ($T = 252$), obtained by taking first differences of the log series of GNP (in billions of dollars):

$$\text{growth rate} = \nabla y_t = y_t - y_{t-1}$$

where y_t is log-GNP.



Order determination: US Gross National Product

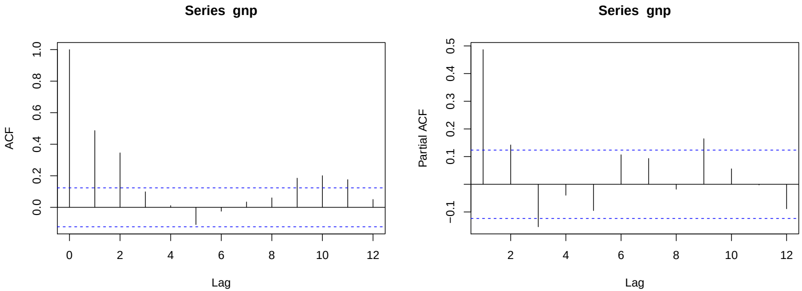


Figure 12: Sample ACF and PACF of the GNP growth rate series. The PACF plot suggests an AR(3) model for the data because the first three lags of sample PACF appear to be significant at the 5% level. There is a marginally significant PACF at lag 9, but this is likely due to random variation.

To sum up..

We have discussed the following properties:

- for MA models, ACF is useful in specifying the order because ACF cuts off at lag q for an MA(q) series;
- for AR models, PACF is useful in order determination because PACF cuts off at lag p for an AR(p) process;
- a MA series is always stationary, but for an AR series to be stationary, all the solutions of the characteristic equation must be greater than 1 in modulus.

Table of Contents

1. White noise and linear time series
2. Autoregressive models
   - AR(1) Model
   - AR(p) processes
   - Properties of AR Models
   - Nonstationary AR processes
3. Moving average models
   - MA(1) model
   - MA(q) processes
   - Properties of MA Models
4. Identifying AR and MA Models
   - Order determination
   - Information Criteria
   - Model Checking
5. Appendix

AIC

Models should be compared both by fit to the data and by model complexity.

Information criteria are based on the *likelihood*, i.e. the probability of obtaining the data given the model. They can be used to determine the order p of an AR process.

AIC (Akaike's Information Criterion) is defined as

$$\text{AIC} = -2 \times \log(\text{max.likelihood}) + 2 \times \text{number of parameters} \tag{15}$$

where `log(max.likelihood)` is the maximized value of the log-likelihood, that measures how well a model fits the data and the *penalty term* depends on the number of independent parameters

AIC and BIC

An alternative is to use the Bayesian Information Criterion (**BIC**):

$$\text{BIC} = -2 \times \text{log(max.likelihood)} + \log(n) \times \text{number of parameters} \quad (16)$$

The penalty for each parameter used is $\log(n)$, thus BIC penalizes the addition of extra parameters more severely than the AIC.

Example. For an AR(p) model with Gaussian error, one can use the approximation

$$\text{BIC}_p = n \log(\hat{\sigma}_a^2) + \log(n)(p+1); \quad \text{AIC}_p = n \log(\hat{\sigma}_a^2) + 2(p+1)$$

where $\hat{\sigma}_a^2$ is the estimated variance of a_t, and n is the number of observations

Selection rule

The AIC and BIC can be used to compare two fitted models: one selects the model that minimizes the AIC value. The same rule applies to BIC.

- To use AIC to select an AR model in practice, one computes $AIC(p)$ for $p = 0, \ldots, P$ where P is a positive integer, and selects the model that minimizes whichever criterion, AIC or BIC, is being used
- Since $\log(n) > 2$ provided, as is typical, that $n > 8$, BIC penalizes model complexity more than AIC does, and for this reason BIC tends to select simpler models than AIC.
- Several other possible criteria have been proposed, but in general we must take into account that there is not a ‘true’ model to select

Selection rule: US Gross National Product

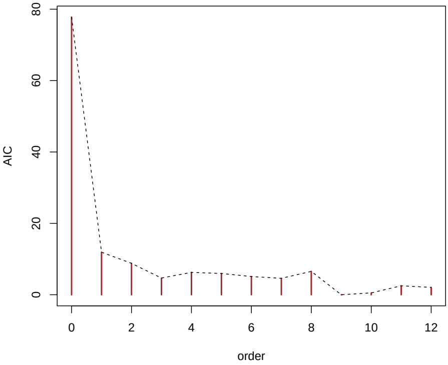


Figure 13: For the growth rate of US GNP, the AIC values with $P = 12$ are plotted against model order. The criterion identifies an AR(9) model for the series, but the plot shows that AIC would specify an AR(3) model if one focuses on lower order models (the AIC value has been adjusted so that the minimum is 0).

Table of Contents

1. White noise and linear time series
2. Autoregressive models
   - AR(1) Model
   - AR(p) processes
   - Properties of AR Models
   - Nonstationary AR processes
3. Moving average models
   - MA(1) model
   - MA(q) processes
   - Properties of MA Models
4. Identifying AR and MA Models
   - Order determination
   - Information Criteria
   - Model Checking
5. Appendix

Parameter Estimation

For the AR(1) model

$$X_t = \phi_1 X_{t-1} + a_t$$

ϕ_1 can be estimated by *maximum likelihood*. This technique finds the values of the parameters which maximise the probability of obtaining the data that we have observed. If $\hat{\phi}_1$ is the parameter estimate, the associated **residual** or estimated error term is

$$\begin{aligned}\hat{a}_t =& \text{observation - fitted value}\\ =& x_t - \hat{\phi}_1 x_{t-1}\end{aligned}$$

The residuals $\{\hat{a}_t\}$ form a time series that are used to check that the model really provide an adequate description of the data.

Residual Analysis

For the AR(p) model $X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + a_t$ denote the estimates of $\phi_0, \ldots, \phi_p$ by $\hat{\phi}_0, \ldots, \hat{\phi}_p$. The fitted model is

$$\hat{\phi}_0 + \hat{\phi}_1 x_{t-1} + \cdots + \hat{\phi}_p x_{t-p}$$

and the associated residual is

$$\hat{a}_t = x_t - \{\hat{\phi}_0 + \hat{\phi}_1 x_{t-1} + \cdots + \hat{\phi}_p x_{t-p}\} \quad t \geq p + 1$$

The variance of the a_t is also estimated and denoted by $\hat{\sigma}_a^2$.
If the model is adequate, then the residual series should behave as a white noise

- time plot of residuals and correlogram
- apply Ljung-Box test to residuals

GNP data: Fitted AR model

The R output from AR(3) fit (arima()) to the GNP growth rate series is:

```
Coefficients:
      ar1     ar2      ar3      intercept
     0.4386  0.2063  -0.1559     0.0163
s.e.  0.0620  0.0666   0.0626     0.0012
sigma^2 estimated as 9.549e-05: log likelihood = 808.56,
aic = -1607.12
```

Remark the 'intercept' is really the estimate of the mean μ!

The fitted AR(3) model is

$$x_t = 0.0083 + 0.4386 x_{t-1} + 0.2063 x_{t-2} - 0.1559 x_{t-3} + \hat{a}_t, \quad \hat{\sigma}_a^2 = 9.549 \times 10^{-5}$$

where $\hat{\phi}_0 = 0.0163(1 - (0.4386 + 0.2063 - 0.1559)) = 0.0083$, and standard errors of the estimates are 0.062, 0.067, 0.063, and 0.001, respectively.

GNP data: Model checking

In order to check if the parameters are statistically significant we look at the ratio of each parameter estimate and its standard error, for instance

$$t = \frac{\hat{\phi}_1}{s.e.(\hat{\phi}_1)}$$

The null hypothesis that the coefficient is zero $H_0 : \phi_1 = 0$ versus $H_1 : \phi_1 \neq 0$ is rejected at 5% level if

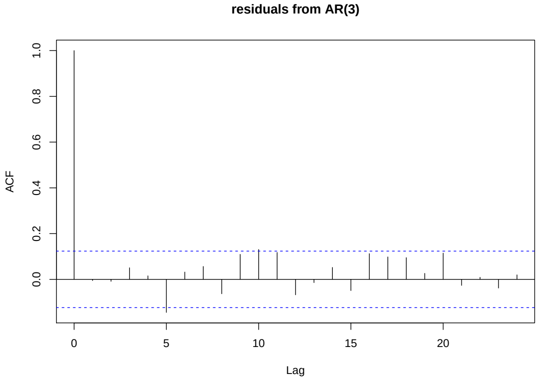
$$|t| > z_{1-\alpha/2}$$

where $z_{1-\alpha/2} \approx 2$. If we accept H_0, the AR coefficient is not significant at the 5% level, and one can refine the model by excluding that term. In the example of GNP data, for instance

$$\frac{0.4386}{0.0620} = 7.074$$

implying that the lag-1 AR coefficient is highly significant (all the parameters are found to be statistically significant).

GNP data: Model checking



If residual ACF shows additional serial correlations, then the model should be extended to take care of the those correlations.

Remark The **Ljung-Box Test** applied to residuals simultaneously tests that all autocorrelations in the residual series up to a given lag are zero.

Appendix: The random walk model

Let $\{y_t\}$ be a time series. The first-order differences can be written as

$$\nabla y_t = y'_t = y_t - y_{t-1}$$

where ∇ is the *difference operator*. When the series ∇y_t is a white noise $\{w_t\} \sim \mathrm{WN}(0, \sigma^2_{\mathrm{w}})$, then

$$y_t = y_{t-1} + w_t$$

Formally, the **random walk model** has the form

$$y_t = y_{t-1} + w_t \tag{17}$$

where $\{w_t\}$ is white noise with mean 0 and variance σ^2_w.

Appendix: The random walk model

The arbitrary initial condition is $(t = 0)$ $y_0 = 0$, so that $y_1 = w_1$. Then, we have

$$
\begin{aligned}
y_2 &= y_1 + w_2 = w_1 + w_2 \\
y_3 &= y_2 + w_3 = w_1 + w_2 + w_3 \\
\vdots &= \vdots \\
y_t &= w_1 + w_2 + \cdots + w_t = \sum_{i=1}^{t} w_i
\end{aligned}
$$

that is a cumulative sum of white noise variates.

If the steps w_i are normally distributed, then the process is called a normal random walk.

Second-order properties

Eq.(17) is a finite sum of white noise terms, each with zero mean and variance σ_w^2. Therefore,

- $\mathsf{E}(y_t) = \mathsf{E}(w_1 + \cdots + w_t) = 0$
- $\mathsf{Var}(y_t) = \mathsf{Var}(w_1 + \cdots + w_t) = t\sigma_w^2$

Given $k > 0$, the autocovariance function is given by

$$\gamma(t, t+k) = \mathsf{Cov}(y_t, y_{t+k}) = \mathsf{Cov}\left(\sum_{i=1}^{t} w_i, \sum_{j=1}^{t+k} w_j\right)$$
$$= \sum_{i=j} \mathsf{Cov}(w_i, w_j) = t\sigma_w^2$$

Simulated random walk series

Simulation from a time series model allows to inspect the main features of the model from plots, so that when historical data exhibit similar features, the model may be selected as a potential candidate.

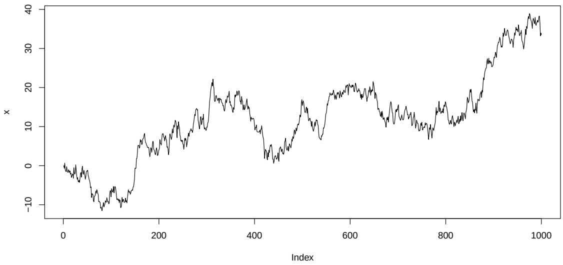


Figure 14: Time plot of a simulated random walk

For a random walk model we observe long periods of apparent trends up or down and sudden and inexplicable changes in direction.

Simulated random walk series

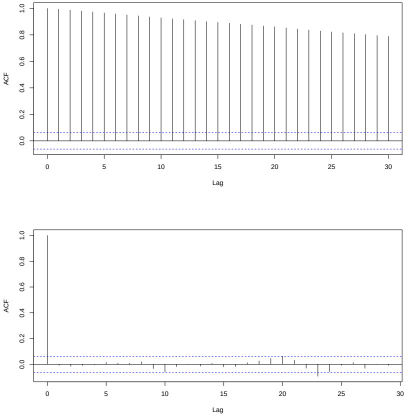


Figure 15: ACF plot for a simulated random walk (top) and first differences (bottom)

The model for log prices

The random walk model has been widely considered for the movement of the logarithm of prices:

$$p_t = p_{t-1} + w_t \tag{18}$$

where

- p_t is the log price of a particular stock at date t
- p_0 is a real number denoting the starting value of the process
- $w_t \sim \text{WN}(0, \sigma_a^2)$

In this case, the log price series is unit-root nonstationary.

The first-order differences from Eq. (18) give

$$\nabla p_t = p_t - p_{t-1} = w_t$$

Note that $p_t - p_{t-1} = \log\left(\frac{P_t}{P_{t-1}}\right)$ is the series **log returns**.

The drift

The log return series of a market index tends to have a small and positive mean. The random walk model can be adapted to allow for this by including a **drift parameter** δ:

$$p_t = \delta + p_{t-1} + w_t \tag{19}$$

In this case

$$p_t = \delta t + p_0 + w_1 + \cdots + w_t = \delta t + p_0 + \sum_{i=1}^{t} w_i$$

and

$$\mathsf{E}(p_t) = p_0 + \delta t,$$

$$\mathsf{Var}(p_t) = t\sigma_w^2$$

For a random walk with drift, the constant term δ becomes the time slope of the series.

The drift: example

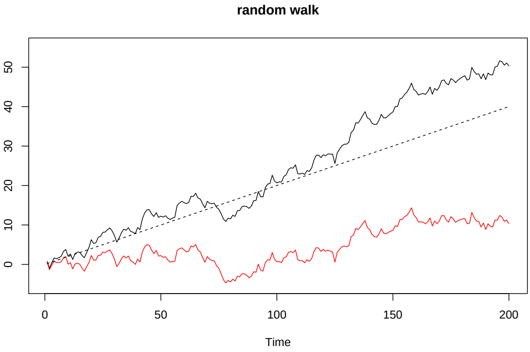


Figure 16: Normal Random Walk with $p_0 = 0$, $\sigma_w = 1$, with drift $\delta = 0.2$ (upper jagged line); without drift, $\delta = 0$ (red line). The dashed line is the straight line $0.2t$.

Go back to Nonstationarity