Chapter 4: Stationary TS Models

§4.1 Stationarity and Autocorrelation

- Consider a time series $\{X_t : t \in T\}$.
- Suppose that $(t_1, t_2, \ldots, t_n)$ is a vector of members of T.
- Then the vector $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ has the joint distribution function

$$F_{t_1,t_2,\ldots,t_n}(x_1, x_2, \ldots, x_n) = P(X_{t_1} \le x_1, X_{t_2} \le x_2, \ldots, X_{t_n} \le x_n).$$

- The collection $\{F_{t_1,t_2,\ldots,t_n}\}$ as $(t_1, t_2, \ldots, t_n)$ range over all the vectors of any finite length n with components from T, is called the collection of ***finite-dimensional distributions*** of $\{X_t : t \in T\}$

- This collection contains all information available about the series from the joint distributions of its constituent variables X_t.
- However, this way of describing a time series is often very complicated and impractical.
- A simpler, albeit generally incomplete, description of a time series is by the moments of the series, particularly, the first and second moments which are, in general, functions of time t:
  - ***Mean***

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

  - ***Variance***

$$\sigma^2_{X_t} = \mathrm{var}(X_t)$$

  - ***Autocovariance (ACVF)***

$$\gamma(t_1, t_2) = \mathrm{cov}(X_{t_1}, X_{t_2}), \quad \text{for all } t_1 \text{ and } t_2.$$

Note that the variance function is a special case of the autocovariance function when $t_1 = t_2$.

Definition 4.1

A time series $\{X_t\}$ is called ***strongly*** *(or* ***strictly) stationary*** *if for $n = 1, 2, \ldots$ and for all $\tau = 0, \pm 1, \pm 2, \ldots$ and for all times $t_1, t_2, \ldots, t_n$, the two families*

$$\{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\} \quad \text{and} \quad \{X_{t_1+\tau}, X_{t_2+\tau}, \ldots, X_{t_n+\tau}\}$$

have a common joint distribution.

□

- This condition states that finite dimensional distributions are invariant under time shifts.
- Note that if $\{X_t\}$ is strictly stationary, then
  - the distribution of X_t is the same for all t,
  - provided additionally that the mean and variance exist, we have

$$\mu_{X_t} = \mu, \qquad \sigma^2_{X_t} = \sigma^2,$$

  - the distribution of the vector (X_{t_1}, X_{t_2}) depends only on the time difference $t_2 - t_1$.
- So, the ACVF $\gamma(t_1, t_2)$ also depends only on $t_2 - t_1$, i.e.

$$\gamma(\tau) = \gamma(t + \tau, t) = \operatorname{cov}(X_{t+\tau}, X_t)$$

for all times τ and for all times t.

Definition 4.2

Similarly, we define so called ***autocorrelation function (ACF)*** *as*

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \operatorname{corr}(X_{t+\tau}, X_t) \quad \textit{for all} \;\; t, \tau. \tag{4.1}$$

Definition 4.3

A time series $\{X_t\}$ *with* $\mathrm{E}(X_t^2) < \infty$ *is called* ***weakly stationary*** *or just* ***stationary*** *if*

$$\mathrm{E}(X_{t_1}) = \mathrm{E}(X_{t_2})$$

and

$$\mathrm{cov}(X_{t_1}, X_{t_2}) = \mathrm{cov}(X_{t_1+\tau}, X_{t_2+\tau})$$

for all t_1, t_2 *and* τ.

□

If $\{X_t\}$ is a weakly stationary TS then obviously

- the expectation of X_t does not depend on t, i.e. $\mu_{X_t} = \mu$ for some μ and for all times t,
- the ACVF $\gamma(t+\tau, t) = \gamma(\tau, 0) = \gamma(\tau)$ may be viewed as a function of a single variable τ.

- Note that
$$\gamma(0) = \mathrm{var}(X_t),$$
that is, the variance is also constant for all t.

Remark 4.1

- If the first two moments exist and are finite, then strict stationarity implies weak stationarity.
- Note that the multivariate normal distribution is completely specified by its first and second moments, i.e. μ_{X_t} and $\gamma(t_1, t_2)$. It therefore follows that weak stationarity implies strict stationarity for Gaussian time series.
- However, μ and $\gamma(\tau)$ may not adequately describe a stationary processes which is very “non-Gaussian”.

Remark 4.2

The following example shows that for non-Gaussian processes *the weak stationarity does not imply strict stationarity*.

Example 4.1

Let $Z_t \underset{iid}{\sim} \mathcal{N}(0,1)$. Define

$$X_t = \begin{cases} Z_t, & \text{if } t \text{ is even,} \\ \frac{1}{\sqrt{2}}(Z_t^2 - 1) & \text{if } t \text{ is odd.} \end{cases}$$

Then

$$\mathrm{E}(X_t) = \begin{cases} \mathrm{E}\, Z_t = 0, & \text{if } t \text{ is even,} \\ \mathrm{E}\left[\frac{1}{\sqrt{2}}(Z_t^2 - 1)\right] = \frac{1}{\sqrt{2}}\,\mathrm{E}[Z_t^2 - 1] = 0 & \text{if } t \text{ is odd.} \end{cases}$$

Also,

$$\mathrm{var}(X_t) = \begin{cases} \mathrm{var}(Z_t) = 1, & \text{if } t \text{ is even,} \\ \mathrm{var}(\frac{1}{\sqrt{2}}(Z_t^2 - 1)) = \frac{1}{2}\,\mathrm{var}(Z_t^2) = 1 & \text{if } t \text{ is odd,} \end{cases}$$

Example 4.1cont-d

and, since X_t and $X_{t+\tau}$ are independent for any $\tau \neq 0$, we obtain that for any $\tau \neq 0$,

$$\mathrm{cov}(X_t, X_{t+\tau}) = 0.$$

- Hence $\{X_t\}$ is a weakly stationary TS.
- Are the X_t identically distributed?
- Note that $P(X_1 < -1/\sqrt{2}) = 0 < P(X_2 < -1/\sqrt{2})$, hence the distributions of X_1 and X_2 are different and the series X_t is not strictly stationary.

Example 4.2

I.i.d. noise

Suppose that $\{X_t\}$ is a sequence of r.v.s which are independent and identically distributed (i.i.d.). Then the joint c.d.f. can be written as

$$\begin{aligned} P(X_{t_1} \le x_1, \ldots, X_{t_n} \le x_n) \\ &= P(X_{t_1} \le x_1) \ldots P(X_{t_n} \le x_n) \\ &= F(x_1) \cdots F(x_n). \end{aligned}$$

Since the joint distribution does not depend on the choice of the indices $\{t_1, \ldots, t_n\}$, it follows that $\{X_t\}$ is strictly stationary.

Example 4.2cont-d:

If X_t has finite second moment $\mathrm{E}(X_t^2) < \infty$, then $\{X_t\}$ is also weakly stationary. Let $\mathrm{var}(X_t) = \sigma^2$, then

$$\gamma(\tau) = \begin{cases} \sigma^2, & \text{if } \tau = 0, \\ 0, & \text{if } \tau \neq 0. \end{cases}$$

Also, the conditional distribution of $X_{n+\tau}$ given values of $(X_1, \ldots, X_n)$ is

$$P(X_{n+\tau} \leq x | X_1 = x_1, \ldots, X_n = x_n) = P(X_{n+\tau} \leq x).$$

This confirms that knowledge of the past has no value for predicting future in this case.

Definition 4.4

A sequence $\{X_t\}$ *of uncorrelated r.v.s, each with zero mean and variance* σ^2 *is called* ***white noise***. *It is denoted by*

$$\{X_t\} \sim WN(0, \sigma^2).$$

Example 4.3

- White noise meets the requirements of the definition of weak stationarity.
- Note that the TS in Example 4.1 is white noise.
- Note that if a Gaussian TS $\{X_t\}$ is white noise, then $\{X_t\}$ is a sequence of Gaussian iid r.vs, which we denote by

$$\{X_t\} \underset{iid}{\sim} \mathcal{N}(0, \sigma^2).$$

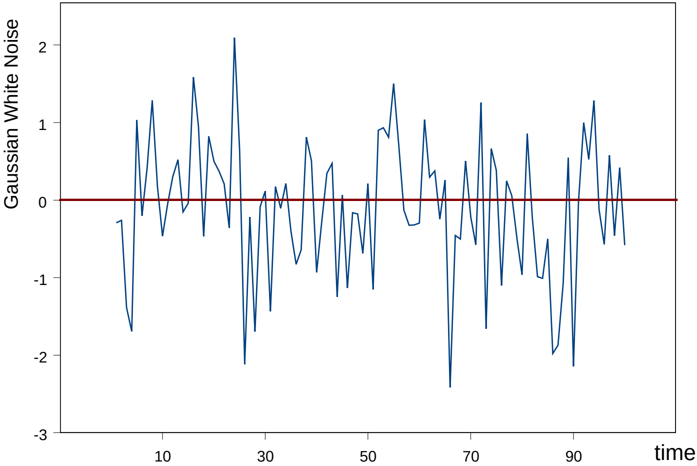


Figure 4.1: Simulated Gaussian White Noise Time Series

Remark 4.3

- Note that every iid series with mean 0 and variance σ^2 is $WN(0, \sigma^2)$, but not conversely. That is, in general
- *zero correlation does not imply independence.*
- Gaussian white noise is however an iid process.

Example 4.4

MA(1) process

Let

$$X_t = Z_t + \theta Z_{t-1}, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{4.2}$$

where

$$\{Z_t\} \sim WN(0, \sigma^2),$$

and $\theta \neq 0$ is a constant. Then $\{X_t\}$ is called **first order moving average**, which we denote by **MA(1)**.

Is the MA(1) a weakly stationary series?

Example 4.4cont-d:

From equation 4.2 we obtain

$$\mathrm{E}(X_t) = \mathrm{E}(Z_t + \theta Z_{t-1}) = \mathrm{E}(Z_t) + \theta\,\mathrm{E}(Z_{t-1}) = 0.$$

Now, we need to check if the autocovariance function does not depend on time, i.e., it depends only on lag τ.

$$\begin{aligned}
\mathrm{cov}(X_t, X_{t+\tau}) = & \quad \mathrm{cov}(Z_t + \theta Z_{t-1}, Z_{t+\tau} + \theta Z_{t-1+\tau}) \\
= & \quad \mathrm{E}[(Z_t + \theta Z_{t-1})(Z_{t+\tau} + \theta Z_{t-1+\tau})] \\
& \quad -\underbrace{\mathrm{E}(Z_t + \theta Z_{t-1})}_{=0}\underbrace{\mathrm{E}(Z_{t+\tau} + \theta Z_{t-1+\tau})}_{=0} \\
= & \quad \mathrm{E}(Z_t Z_{t+\tau}) + \theta\,\mathrm{E}(Z_t Z_{t-1+\tau}) + \theta\,\mathrm{E}(Z_{t-1} Z_{t+\tau}) \\
& \quad +\theta^2\,\mathrm{E}(Z_{t-1} Z_{t-1+\tau}).
\end{aligned}$$

Example 4.4cont-d:

Now, considering all possible values of the lag τ we obtain

$$\mathrm{cov}(X_t, X_{t+\tau}) = \begin{cases} (1+\theta^2)\sigma^2, & \text{if } \tau = 0, \\ \theta\sigma^2, & \text{if } \tau = \pm 1, \\ 0, & \text{if } |\tau| > 1. \end{cases} \tag{4.3}$$

Hence, the (auto)covariance does not depend on t and we can write the autocovariance as $\gamma_X(\tau)$, i.e. function of lag τ only:

$$\gamma_X(\tau) = \mathrm{cov}(X_t, X_{t+\tau}) \quad \text{for any } \tau.$$

The conclusion is that *MA(1) is a weakly stationary process*.

Example 4.4cont-d:

Also, from (4.3) we obtain the form of the autocorrelation function

$$\rho_X(\tau) = \begin{cases} 1, & \text{if } \tau = 0, \\ \frac{\theta}{1+\theta^2} & \text{if } \tau = \pm 1, \\ 0, & \text{if } |\tau| > 1. \end{cases} \tag{4.4}$$

Figure 4.2 shows an MA(1) process which is simulated by using a Gaussian white noise and $\theta = 0.5$.

□

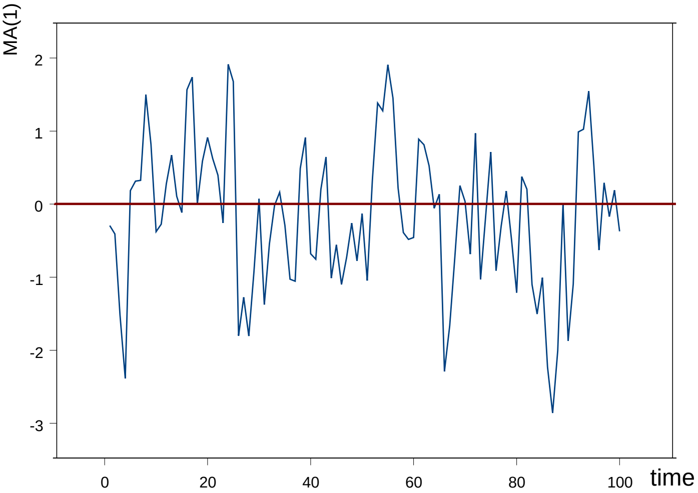


Figure 4.2: Simulated MA(1) Time Series

§4.1.1 Sample Autocovariance and Autocorrelation

- The ACVF and ACF are helpful tools for assessing the degree, or time range, of dependence and recognising if a TS follows a well-known model.
- However, in practice we generally are not given the ACVF or ACF, but
- are given a sample from, or realisaion of, a time series.
- When we try to fit a model to the observed realisation, we often use *sample autocovariance and autocorrelation functions* which are defined in terms of the observed data.

Definition 4.5

Let $x_1, \ldots, x_n$ be observations of a TS. The ***sample autocovariance function*** *is defined as*

$$\widehat{\gamma}(\tau) = \frac{1}{n} \sum_{t=1}^{n-|\tau|} (x_t - \bar{x})(x_{t+|\tau|} - \bar{x}), \quad -n < \tau < n \tag{4.5}$$

where

$$\bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t.$$

The ***sample autocorrelation function*** *is defined as*

$$\widehat{\rho}(\tau) = \frac{\widehat{\gamma}(\tau)}{\widehat{\gamma}(0)}, \quad -n < \tau < n. \tag{4.6}$$

Remark 4.4

For lag $\tau \geq 0$ the sample autocovariance function is approximately equal to the sample covariance of the $n - \tau$ pairs $(x_1, x_{1+\tau}), \dots, (x_{n-\tau}, x_n)$. Note that, in (4.5), we divide the sum by n, not by $n - \tau$ and also we use the overall mean $\bar{x}$ for both x_t and $x_{t+\tau}$.

Remark 4.5

The sample autocovariance function $\widehat{\gamma}(\tau)$ and the sample autocorrelation function $\widehat{\rho}(\tau)$ are the most commonly used estimators of the theoretical autocovariance function $\gamma(\tau)$ and autocorrelation function $\rho(\tau)$ respectively.

- A graph of the sample autocorrelation (autocovariance) function is called a **correlogram** (**covariogram**).
- Figures 4.3 and 4.4, respectively, show the correlogram of the Gaussian white noise time series given in Figure 4.1 and the correlogram of the MA(1) TS with $\theta = 0.5$ calculated from the white noise.
- As expected, there is no significant correlation for lag $\tau \geq 1$ for the white noise, but there is one for the MA(1) for lag $\tau = 1$.

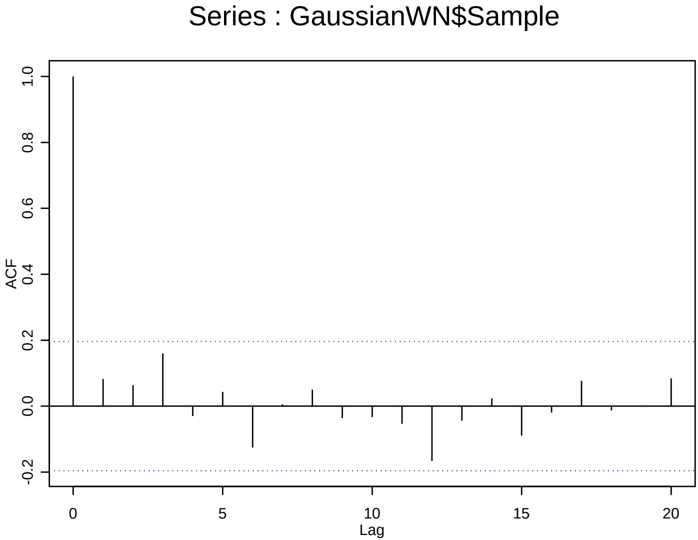


Figure 4.3: Correlogram of the Simulated Gaussian White Noise Time Series

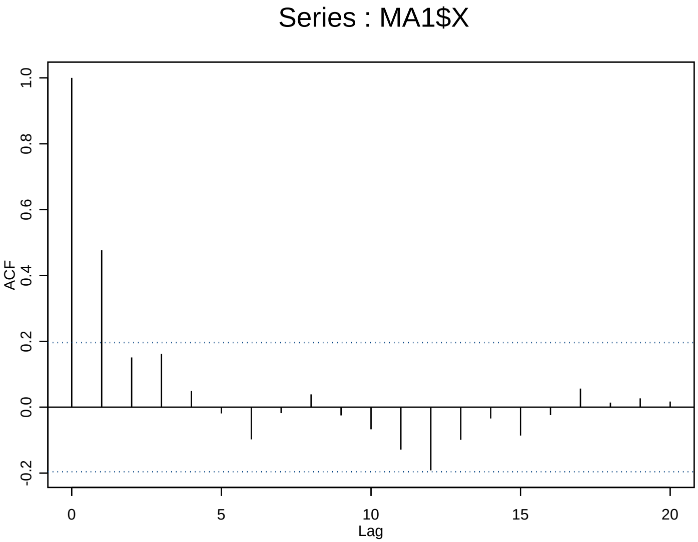


Figure 4.4: Correlogram of the Simulated MA(1) Time Series

§4.2 Properties of ACVF and ACF

First we examine some basic properties of the Autocovariance function (ACVF).

Proposition 4.1

The ACVF of a stationary TS is a function $\gamma(\cdot)$ such that

1. $\gamma(0) \geq 0$,
2. $|\gamma(\tau)| \leq \gamma(0)$ *for all* τ,
3. $\gamma(\cdot)$ *is even, i.e.,*

$$\gamma(\tau) = \gamma(-\tau), \quad \textit{for all } \tau.$$

Definition 4.6

We say that a real-valued function κ *defined on the integers is nonnegative definite if*

$$\sum_{i,j=1}^{n} a_i \kappa(i-j) a_j \geq 0 \tag{4.7}$$

for all positive integers n *and real-valued vectors* $\boldsymbol{a} = (a_1, \ldots, a_n)^{\mathrm{T}}$.

Theorem 4.1

A real-valued function defined on the integers is the autocovariance function of a stationary TS if and only if it is even and nonnegative definite.

Proof

We will only show that the ACVF of a stationary TS $\{X_t\}$ is nonnegative definite and omit the rest.

Let $\boldsymbol{a} = (a_1, \ldots, a_n)^{\mathrm{T}}$ be a real n-dimensional vector. Define a matrix

$$\boldsymbol{V} = \begin{pmatrix} \gamma(0) & \gamma(1-2) & \ldots & \gamma(1-n) \\ \gamma(2-1) & \gamma(0) & \ldots & \gamma(2-n) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \ldots & \gamma(0) \end{pmatrix}.$$

It is easy to see that $\boldsymbol{V}$ is the covariance matrix of the vector $\boldsymbol{X} = (X_1, \ldots, X_n)^{\mathrm{T}}$. Then,

$$0 \le \mathrm{var}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}) = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{a} = \sum_{i,j=1}^{n} a_i \gamma(i-j) a_j.$$

Hence $\gamma(\tau)$ is a nonnegative definite function.

$\square$

§4.3 Moving Average Process MA(q)

Definition 4.7

$\{X_t\}$ is a ***moving-average process of order q*** *if*

$$X_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q},$$

where

$$Z_t \sim WN(0, \sigma^2)$$

and $\theta_1, \ldots, \theta_q$ *are constants.*

Remark 4.6

X_t is a linear combination of $q+1$ white noise variables and we say that it is **q-correlated**, that is X_t and $X_{t+\tau}$ are uncorrelated for all lags $\tau > q$.

Remark 4.7

If $\{Z_t\}$ is an i.i.d process then X_t is a strictly stationary TS since for any n and for any $t_1, t_2, \ldots, t_n$

$$(\boldsymbol{Z}_{t_1}^{\mathrm{T}}, \boldsymbol{Z}_{t_2}^{\mathrm{T}}, \ldots, \boldsymbol{Z}_{t_n}^{\mathrm{T}}) \stackrel{D}{=} (\boldsymbol{Z}_{t_1+\tau}^{\mathrm{T}}, \boldsymbol{Z}_{t_2+\tau}^{\mathrm{T}}, \ldots, \boldsymbol{Z}_{t_n+\tau}^{\mathrm{T}})$$

for all τ, where $\boldsymbol{Z}_t^{\mathrm{T}} = (Z_{t-q}, Z_{t+1-q}, \ldots, Z_{t-1}, Z_t)$.
Then *also* $\{X_t\}$ is called **q-dependent**, that is X_t and $X_{t+\tau}$ are independent for all lags $\tau > q$.

Remark 4.8 (Some obvious observations:)

- IID noise is a 0-dependent TS.
- White noise is a 0-correlated TS.
- MA(1) is 1-correlated TS, and it is also 1-dependent if the WN $\{Z_t\}$ is an iid noise.
- If $\{Z_t\} \underset{iid}{\sim} \mathcal{N}(0, \sigma^2)$, then X_ts are also normally distributed and hence we have a strictly stationary Gaussian TS.

Remark 4.9

The MA(q) process can also be written in the following equivalent form

$$X_t = \theta(B)Z_t, \tag{4.8}$$

where the **moving average operator**

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q \tag{4.9}$$

defines a linear combination of the first q powers of the backward shift operator $B^k Z_t = Z_{t-k}$.

Example 4.5 (MA(2) process)

This process is written as

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} = (1 + \theta_1 B + \theta_2 B^2) Z_t. \tag{4.10}$$

What are the properties of MA(2)? As it is a combination of a zero mean white noise, it also has zero mean, i.e.,

$$\mathrm{E}\, X_t = \mathrm{E}(Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}) = 0.$$

It is easy to calculate the covariance of X_t and $X_{t+\tau}$. We get

$$\gamma(\tau) = \operatorname{cov}(X_t, X_{t+\tau}) = \begin{cases} (1 + \theta_1^2 + \theta_2^2)\sigma^2 & \text{for } \tau = 0, \\ (\theta_1 + \theta_1\theta_2)\sigma^2 & \text{for } \tau = \pm 1, \\ \theta_2 \sigma^2 & \text{for } \tau = \pm 2, \\ 0 & \text{for } |\tau| > 2, \end{cases}$$

which shows that the autocovariances depend on lag, but not on time.

Example 4.5cont-d:

Dividing $\gamma(\tau)$ by $\gamma(0)$ we obtain the autocorrelation function,

$$\rho(\tau) = \begin{cases} 1 & \text{for } \tau = 0, \\ \frac{\theta_1+\theta_1\theta_2}{1+\theta_1^2+\theta_2^2} & \text{for } \tau = \pm 1, \\ \frac{\theta_2}{1+\theta_1^2+\theta_2^2} & \text{for } \tau = \pm 2 \\ 0 & \text{for } |\tau| > 2. \end{cases}$$

$\square$

In summary: MA(2) process is a weakly stationary, 2-correlated TS.

The following graph shows MA(2) processes obtained from the simulated Gaussian white noise shown in Figure 4.1 for various values of the parameters (θ_1, θ_2).
The blue series (with smaller variance) is

$$x_t = z_t + 0.5z_{t-1} + 0.5z_{t-2},$$

while the purple series (with larger variance) is

$$x_t = z_t + 5z_{t-1} + 5z_{t-2},$$

where z_t are realizations of an i.i.d. Gaussian noise.

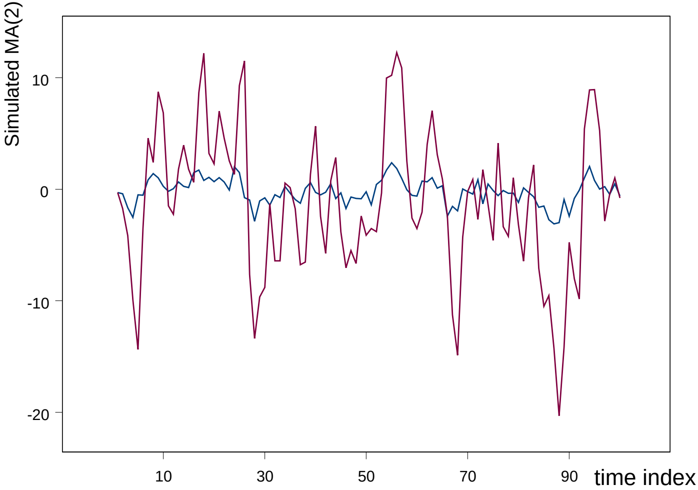
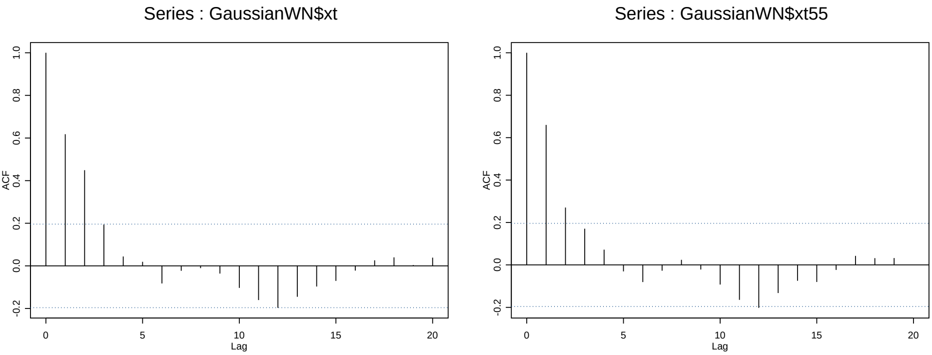


Figure 4.5: Two simulated MA(2) processes, both from the white noise shown in Figure 4.1, but for different sets of parameters.



(a) (b)

Figure 4.6: (a) Sample ACF for $x_t = z_t + 0.5z_{t-1} + 0.5z_{t-2}$ and (b) for $x_t = z_t + 5z_{t-1} + 5z_{t-2}$.

As we can see, very different processes can be obtained by varying the parameters. This is an important property of MA(q) processes, which gives a very large family of models. This property is reinforced by the following Proposition.

Proposition 4.2

If $\{X_t\}$ is a stationary q-correlated time series with mean zero, then it can be represented as an MA(q) process.

$\square$

The following theorem gives the form of ACF for a general MA(q).

Theorem 4.2

An MA(q) process as in Definition 4.7 is a weakly stationary TS with the ACVF

$$\gamma(\tau) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|\tau|} \theta_j \theta_{j+|\tau|}, & \text{if } |\tau| \le q, \\ 0, & \text{if } |\tau| > q, \end{cases} \tag{4.11}$$

where θ_0 is defined to be 1.

□

It follows from the above theorem that the ACF is given by

$$\rho(\tau) = \begin{cases} \sum_{j=0}^{q-|\tau|} \theta_j \theta_{j+|\tau|} / \sum_{j=0}^{q} \theta_j^2 & \text{if } |\tau| \le q, \\ 0, & \text{if } |\tau| > q, \end{cases} \tag{4.12}$$

Remark 4.10

The ACVF (ACF) of a MA(q) has a distinct “cut-off” at lag $\tau = q$.

Remark 4.11

Note that an arbitrary constant, say μ, can be added to the right hand side of Definition 4.7 to give a TS with mean μ. This does not affect the ACF and has been omitted for simplicity.

§4.3.1 Non-uniqueness of MA Models

Consider an example of MA(1)

$$X_t = Z_t + \theta Z_{t-1}$$

whose ACF is

$$\rho(\tau) = \begin{cases} 1, & \text{if } \tau = 0, \\ \frac{\theta}{1+\theta^2} & \text{if } \tau = \pm 1, \\ 0, & \text{if } |\tau| > 1. \end{cases}$$

- Note that for $q = 1$, the maximum value of $|\rho(1)|$ is 0.5. This can be verified directly from the above formula for the ACF.
- Treating $\rho(1)$ as a function of θ we can calculate its extrema. Let

$$f(\theta) = \frac{\theta}{1 + \theta^2}.$$

- Then

$$f'(\theta) = \frac{1 - \theta^2}{(1 + \theta^2)^2}.$$

- The derivative is equal to 0 at $\theta = \pm 1$ and the function f attains its maximum at $\theta = 1$ and minimum at $\theta = -1$. We have

$$f(1) = \frac{1}{2}, \quad f(-1) = -\frac{1}{2}.$$

- This fact can be helpful in recognizing MA(1) processes. In fact, MA(1) with $|\theta| = 1$ may be uniquely identified from the autocorrelation function.

However, it is easy to see that θ and $\frac{1}{\theta}$ give the same ACF! Take, for example 5, and $\frac{1}{5}$. In both cases

$$\rho(\tau) = \begin{cases} 1 & \text{if } \tau = 0, \\ \frac{5}{26} & \text{if } \tau = \pm 1, \\ 0 & \text{if } |\tau| > 1. \end{cases}$$

Also, the pair $\sigma^2 = 1, \theta = 5$ gives the same ACVF as the pair $\sigma^2 = 25, \theta = \frac{1}{5}$, namely

$$\gamma(\tau) = \begin{cases} (1+\theta^2)\sigma^2 = 26, & \text{if } \tau = 0, \\ \theta\sigma^2 = 5, & \text{if } \tau = \pm 1, \\ 0, & \text{if } |\tau| > 1. \end{cases}$$

Hence, the MA(1) processes

$$X_t = Z_t + \frac{1}{5} Z_{t-1}, \quad Z_t \underset{iid}{\sim} \mathcal{N}(0, 25)$$

and

$$X_t = Y_t + 5 Y_{t-1}, \quad Y_t \underset{iid}{\sim} \mathcal{N}(0, 1)$$

are the same.

- Since we can only observe variables X_t and not the noise variables, *we cannot distinguish between these two models.*
- Is there any reason to prefer one of the two processes over the other?
- If yes, which one?

Next, we develop the notion of *invertibility* which will help us answer these questions.

§4.3.2 Invertibility of MA Processes

The MA(1) process can be expressed in terms of lagged values of X_t by substituting repeatedly for lagged values of Z_t.

Thus, the substitution $Z_t = X_t - \theta Z_{t-1}$ yields

$$
\begin{aligned}
Z_t = X_t - \theta Z_{t-1} &= X_t - \theta(X_{t-1} - \theta Z_{t-2}) \\
&= X_t - \theta X_{t-1} + \theta^2 Z_{t-2} \\
&= X_t - \theta X_{t-1} + \theta^2 (X_{t-2} - \theta Z_{t-3}) \\
&= X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \theta^3 Z_{t-3} \\
&= \ldots \\
&= X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \theta^3 X_{t-3} + \theta^4 X_{t-4} + \ldots + (-\theta)^n Z_{t-n}.
\end{aligned}
$$

This can be rewritten as

$$(-\theta)^n Z_{t-n} = Z_t - \sum_{j=0}^{n-1} (-\theta)^j X_{t-j}.$$

However, if $|\theta| < 1$, then

$$\mathrm{E}\left(Z_t - \sum_{j=0}^{n-1} (-\theta)^j X_{t-j}\right)^2 = \mathrm{E}\left(\theta^{2n} Z_{t-n}^2\right) = \theta^{2n}\sigma^2 \underset{n\to\infty}{\longrightarrow} 0$$

and we say that the sum is convergent *in the mean square sense*.

Hence, we obtain another representation of this model

$$Z_t = \sum_{j=0}^{\infty} (-\theta)^j X_{t-j}.$$

- This turns out to be a representation of another class of models, called *infinite autoregressive (AR) models*.
- Thus, we inverted MA(1) to an infinite AR.
- This was possible due to the assumption that $|\theta| < 1$.
- MA(1) processes satisfying this assumption are called **invertible**.
- For various reasons (related to estimation and prediction) it is desirable that a TS be invertible.
- Thus, in the previous example (§4.3.1) we would choose the instance with $\sigma^2 = 25, \theta = \frac{1}{5}$.

§4.4 Linear Processes

Definition 4.8

Let $Z_t \sim WN(0, \sigma^2)$, $-\infty < t < \infty$. A TS $\{X_t\}$ is called a ***linear process*** *if it has the representation*

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \tag{4.13}$$

for all t, where $\{\psi_j\}$ is a sequence of constants such that $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

$\square$

Remark 4.12

The condition $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ ensures that X_t is well-defined, i.e. that the infinite sum converges with probability one, or *almost surely*, and also in the mean square sense, that is

$$\mathrm{E}\,(X_t - \sum_{j=-n}^{n} \psi_j Z_{t-j})^2 \to 0 \quad \text{as } n \to \infty.$$

It can be shown that $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ ensures convergence of the infinite sum in (4.13) also in the mean. Consequently, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ allows us to establish that $\mathrm{E}(X_t) = \sum_{j=-\infty}^{\infty} \mathrm{E}(\psi_j Z_{t-j}) = 0$.

Remark 4.13

MA(∞) is a linear process with $\psi_j = 0$ for $j < 0$, that is MA(∞) has the representation

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}.$$

Note that the formula (4.13) can be written using the backward shift operator B. We have

$$Z_{t-j} = B^j Z_t.$$

Hence

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} = \sum_{j=-\infty}^{\infty} \psi_j B^j Z_t.$$

Denoting

$$\psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j, \tag{4.14}$$

we can write the linear process in a neat way

$$X_t = \psi(B) Z_t.$$

The operator $\psi(B)$ is a linear filter, which when applied to a stationary process produces a stationary process. This fact is proved in the following proposition.

Proposition 4.3

Let $\{Y_t\}$ be a stationary TS with mean zero and autocovariance function γ_Y. If $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, then the process

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} = \psi(B) Y_t \qquad (4.15)$$

is stationary with mean zero and autocovariance function

$$\gamma_X(\tau) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(\tau - k + j). \qquad (4.16)$$

Proof

- The assumption $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ assures convergence of the series.
- When combined with the Dominated Convergence Theorem and Cauchy-Schwarz-Bunyakovsky inequality (not examinable), this assumption also allows us to change the order of expectation and infinite summations in the expressions given below:

First, since $\mathrm{E}\, Y_t = 0$, we have

$$\mathrm{E}\, X_t = \mathrm{E}\left(\sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} \right) = \sum_{j=-\infty}^{\infty} \psi_j \, \mathrm{E}(Y_{t-j}) = 0$$

and

Proof cont-d:

$$
\begin{aligned}
\mathrm{E}(X_t X_{t+\tau}) &= \mathrm{E}\left[\left(\sum_{j=-\infty}^{\infty} \psi_j Y_{t-j}\right)\left(\sum_{k=-\infty}^{\infty} \psi_k Y_{t+\tau-k}\right)\right] \\
&= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \, \mathrm{E}(Y_{t-j} Y_{t+\tau-k}) \\
&= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(\tau - k + j).
\end{aligned}
$$

Hence, $\{X_t\}$ is a stationary TS with the autocovariance function given by formula (4.16).

$\square$

Corrolary 4.1

If $\{Y_t\}$ is a white noise process, then $\{X_t\}$ given by (4.15) *is a stationary linear process with zero mean and the ACVF*

$$\gamma_X(\tau) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+\tau}. \tag{4.17}$$

□

§4.5 Autoregressive Processes AR(p)

The idea behind the autoregressive models is to explain the present value of the series, X_t, by a function of p past values, $X_{t-1}, X_{t-2}, \ldots, X_{t-p}$.

Definition 4.9

An ***autoregressive process of order p*** *is written as*

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + Z_t, \qquad (4.18)$$

where $\{Z_t\}$ is white noise, i.e., $\{Z_t\} \sim WN(0, \sigma^2)$ and Z_t is uncorrelated with X_s for any $s < t$.

Remark 4.14

We assume for simplicity of notation that the mean of X_t is zero. If the mean is $\mathrm{E}\, X_t = \mu \neq 0$, then we replace X_t by $X_t - \mu$ or, equivalently, we can write

$$X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + Z_t,$$

where

$$\alpha = \mu(1 - \phi_1 - \ldots - \phi_p).$$

Other ways of writing AR(p) model use:

Vector notation: Denote

$$\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_p)^{\mathrm{T}},$$
$$\boldsymbol{X_{t-1}} = (X_{t-1}, X_{t-2}, \ldots, X_{t-p})^{\mathrm{T}}.$$

Then the formula (4.18) can be written as

$$X_t = \boldsymbol{\phi}^{\mathrm{T}} \boldsymbol{X}_{t-1} + Z_t.$$

Backshift operator: Namely, writing the model (4.18) in the form

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \ldots - \phi_p X_{t-p} = Z_t,$$

and applying $BX_t = X_{t-1}$ we get

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p) X_t = Z_t$$

or, using the concise notation we write

$$\phi(B) X_t = Z_t, \tag{4.19}$$

where $\phi(B)$ denotes the **autoregressive operator**

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p.$$

Then the AR(p) can be viewed as a solution to the equation (4.19), i.e.,

$$X_t = \frac{1}{\phi(B)} Z_t. \tag{4.20}$$

§4.5.1 AR(1)

- According to the Definition 4.9 an autoregressive process of order 1 has to satisfy

$$X_t = \phi X_{t-1} + Z_t, \tag{4.21}$$

where $Z_t \sim WN(0, \sigma^2)$ and ϕ is a non-zero constant.
- There are (infinitely) many processes satisfying (4.21) (since we can freely choose, say, X_0.)
- Are any of these AR(1) processes stationary?
- If a stationary AR(1) processes exists, E(X_t) would have to be zero.
- Also, we are interested in stationary processes with finite variance $v = \mathrm{E}(X_t^2) < \infty$.

Corollary 4.1 says that an infinite combination of white noise variables is a stationary process. Here, due to the recursive form of the TS we can write AR(1) in such a form. Namely

$$\begin{aligned} X_t &= \phi X_{t-1} + Z_t \\ &= \phi(\phi X_{t-2} + Z_{t-1}) + Z_t \\ &= \phi^2 X_{t-2} + \phi Z_{t-1} + Z_t \\ &\vdots \\ &= \phi^k X_{t-k} + \sum_{j=0}^{k-1} \phi^j Z_{t-j}. \end{aligned}$$

This can be rewritten as

$$\phi^k X_{t-k} = X_t - \sum_{j=0}^{k-1} \phi^j Z_{t-j}.$$

What would we obtain if we have continued the backwards operation, i.e., what happens when $k \to \infty$?

Taking the expectation we obtain

$$\lim_{k\to\infty} \mathrm{E}\left(X_t - \sum_{j=0}^{k-1} \phi^j Z_{t-j}\right)^2 = \lim_{k\to\infty} \phi^{2k} v = 0$$

provided that $|\phi| < 1$. Hence, a stationary AR(1) does exist as a limit

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j} \tag{4.22}$$

in the mean square sense. This is a linear process (4.13) with

$$\psi_j = \begin{cases} \phi^j & \text{for } j \geq 0, \\ 0 & \text{for } j < 0. \end{cases} \tag{4.23}$$

- This technique of iterating backwards works well for AR(1) but not for higher orders.
- A more general way to convert the series into a linear process form is the following method of *matching coefficients*.

The AR(1) model is

$$\phi(B)X_t = Z_t,$$

where $\phi(B) = 1 - \phi B$ and $|\phi| < 1$. We want to write the model as a linear process

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} = \psi(B)Z_t,$$

where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$.

It means we want to find the coefficients ψ_j. Substituting Z_t from the AR model into the linear process model we obtain

$$X_t = \psi(B)Z_t = \psi(B)\phi(B)X_t. \qquad (4.24)$$

In full, the coefficients can be written as

$$\begin{aligned}
1 &= (1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \ldots)(1 - \phi B) \\
&= 1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \ldots \\
&\quad -\phi B - \psi_1 \phi B^2 - \psi_2 \phi B^3 - \psi_3 \phi B^4 - \ldots \\
&= 1 + (\psi_1 - \phi)B + (\psi_2 - \psi_1\phi)B^2 + (\psi_3 - \psi_2\phi)B^3 + \ldots
\end{aligned}$$

Now, equating coefficients on the LHS and RHS of this equation we see that all the coefficients of B^j, $j \geq 1$, must be zero, i.e.,

$$\begin{aligned}
\psi_1 &= \phi \\
\psi_2 &= \psi_1 \phi = \phi^2 \\
\psi_3 &= \psi_2 \phi = \phi^3 \\
&\vdots \\
\psi_j &= \psi_{j-1} \phi = \phi^j.
\end{aligned}$$

So, we obtained the linear process form of the AR(1)

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}.$$

Remark 4.15

Note, that from the equation (4.24) it follows that $\psi(B)$ is an inverse of $\phi(B)$ written formally as

$$\psi(B) = \frac{1}{\phi(B)}. \tag{4.25}$$

For an AR(1) we have

$$\psi(B) = \frac{1}{1 - \phi B} = 1 + \phi B + \phi^2 B^2 + \phi^3 B^3 + \dots \tag{4.26}$$

So, since $|\phi| < 1$, we have $\sum_{j=0}^{\infty} |\phi^j| < \infty$ and hence AR(1) is stationary with

$$\mathrm{E}\, X_t = \sum_{j=0}^{\infty} \phi^j \, \mathrm{E}(Z_{t-j}) = 0 \tag{4.27}$$

and autocovariance function given by (4.17), that is

$$\gamma(\tau) = \sigma^2 \sum_{j=0}^{\infty} \phi^j \phi^{j+\tau} = \sigma^2 \phi^\tau \sum_{j=0}^{\infty} \phi^{2j}.$$

The infinite sum in this expression is the sum of a geometric progression as $|\phi| < 1$, i.e.,

$$\sum_{j=0}^{\infty} \phi^{2j} = \frac{1}{1 - \phi^2}.$$

Corollary 4.1 also gives us the ACVF of the newly defined stationary AR(1) process.

$$\gamma(\tau) = \frac{\sigma^2 \phi^{|\tau|}}{1 - \phi^2}. \tag{4.28}$$

Then the variance of AR(1) is

$$\gamma(0) = \frac{\sigma^2}{1 - \phi^2}.$$

Hence, the autocorrelation function of AR(1) is

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \phi^{|\tau|}. \tag{4.29}$$

However, when $|\phi| = 1$, there is no stationary solution of (4.21) Figures 4.7, 4.9 and 4.8, 4.10 show simulated AR(1) processes for four different values of the coefficient ϕ (equal to -0.9, 0.9, -0.5 and 0.5) and the respective sample ACF functions.

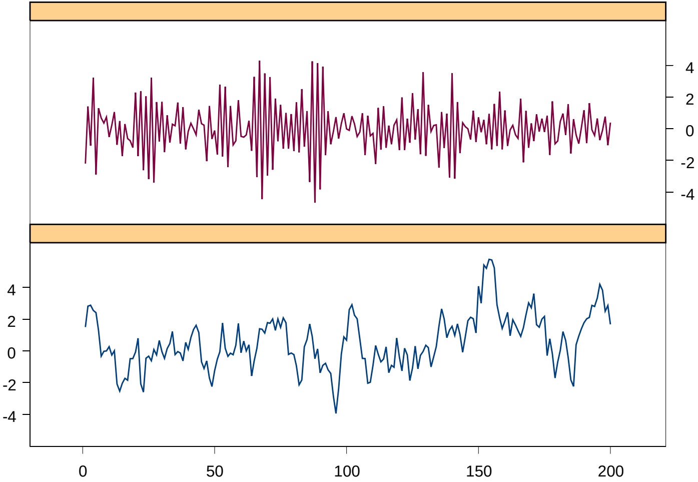
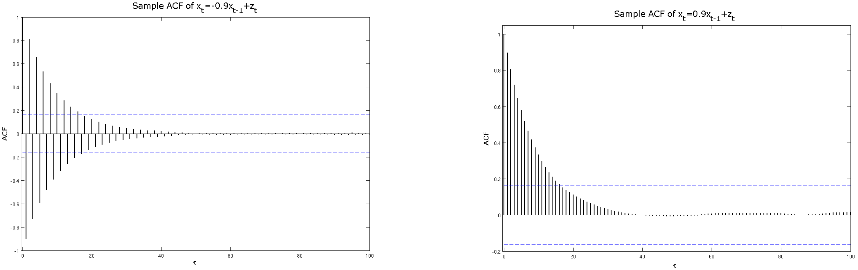


Figure 4.7: Simulated AR(1) processes for $\phi = -0.9$ (top) and for $\phi = 0.9$ (bottom).



(a) (b)

Figure 4.8: Sample ACF for AR(1): (a) $x_t = -0.9x_{t-1} + z_t$ and (b) $x_t = 0.9x_{t-1} + z_t$.

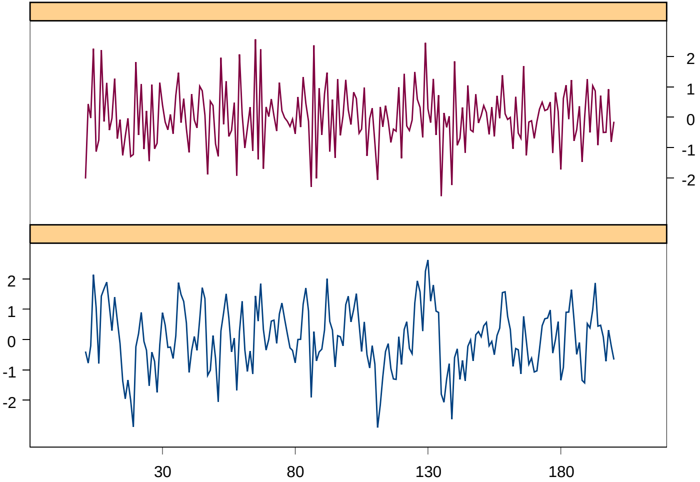


Figure 4.9: Simulated AR(1) processes for $\phi = -0.5$ (top) and for $\phi = 0.5$ (bottom).

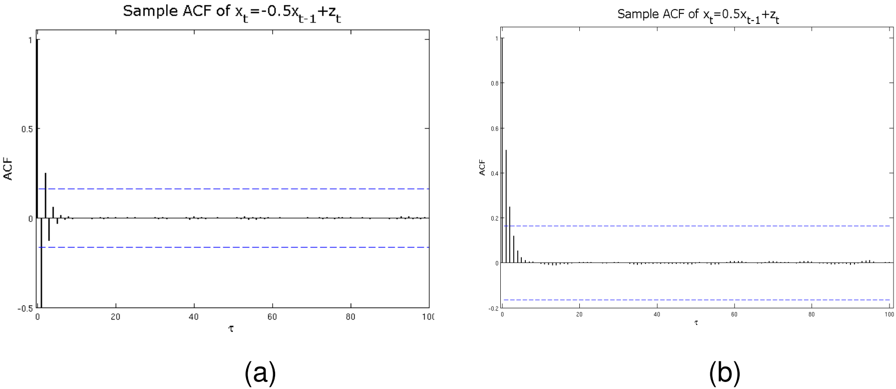


Figure 4.10: Sample ACF for AR(1): (a) $x_t = -0.5x_{t-1} + z_t$ and (b) $x_t = 0.5x_{t-1} + z_t$.

Looking at these graphs we can see that

- for the positive values of ϕ we obtain smoother TS than for the negative ones.
- Also, the ACFs are very different. We see that if ϕ is negative the neighboring observations are negatively correlated, but those two time points apart are positively correlated.
- In fact, if ϕ is negative the neighboring TS values have typically opposite signs. This is more evident if ϕ is close to -1.

§4.5.2 Random Walk

This is a TS which at each point of time moves randomly away from its current position. The model can then be written as

$$X_t = X_{t-1} + Z_t, \tag{4.30}$$

where Z_t is a white noise variable with zero mean and constant variance σ^2. The model has the same form as AR(1) process, but since $\phi = 1$, it is not stationary. Such process is called Random Walk.

Repeatedly substituting for past values gives

$$
\begin{aligned}
X_t &= X_{t-1} + Z_t \\
&= X_{t-2} + Z_{t-1} + Z_t \\
&= X_{t-3} + Z_{t-2} + Z_{t-1} + Z_t \\
&= \dots \\
&= X_0 + \sum_{j=0}^{t-1} Z_{t-j}.
\end{aligned}
$$

If the initial value, X_0, is constant, then the mean value of X_t is equal to X_0, that is

$$
\mathrm{E}\, X_t = \mathrm{E}\left[X_0 + \sum_{j=0}^{t-1} Z_{t-j}\right] = X_0.
$$

So, the mean is constant, but as we see next, the variance and covariance do depend on time, not just on lag.

$$
\begin{aligned}
\operatorname{var}(X_t) &= \operatorname{var}\left(X_0 + \sum_{j=0}^{t-1} Z_{t-j}\right) \\
&= \operatorname{var}\left(\sum_{j=0}^{t-1} Z_{t-j}\right) \\
&= \sum_{j=0}^{t-1} \operatorname{var}(Z_{t-j}) = t\sigma^2
\end{aligned}
$$

since Z_t are uncorrelated. Also,

$$
\begin{aligned}
\operatorname{cov}(X_{t_1}, X_{t_2}) &= \operatorname{cov}\left(\sum_{j=0}^{t_1-1} Z_{t_1-j}, \sum_{k=0}^{t_2-1} Z_{t_2-k}\right) \\
&= \mathrm{E}\left[\left(\sum_{j=0}^{t_1-1} Z_{t_1-j}\right)\left(\sum_{k=0}^{t_2-1} Z_{t_2-k}\right)\right] \\
&= \sigma^2 \min\{t_1, t_2\}.
\end{aligned}
$$

- Two simulated series of this form are shown in Figure 4.11.
- As we can see, the random walk meanders away from its starting value in no particular direction. It does not exhibit any clear trend, but at the same time is not stationary.
- However, the first difference of random walk is stationary as it is just white noise, namely

$$\nabla X_t = X_t - X_{t-1} = Z_t.$$

- The differenced random walk and its sample ACF are shown in Figure 4.12.

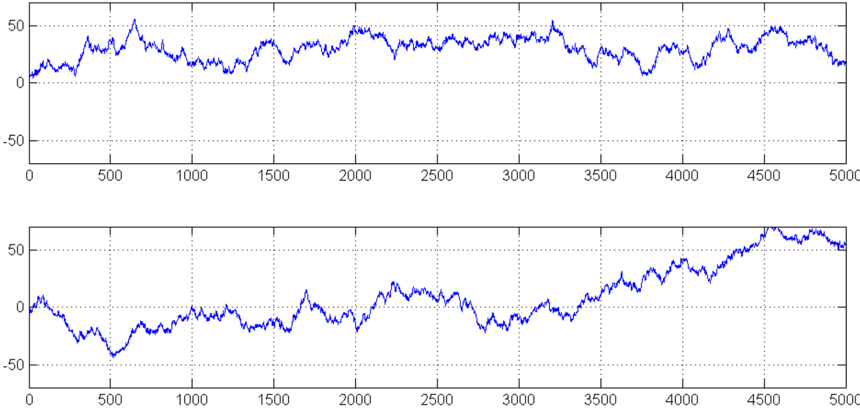


Figure 4.11: Two simulated Random Walks with $x_0 = 0$, $x_t = x_{t-1} + z_t$, $Z_t \sim N(0, 1)$.

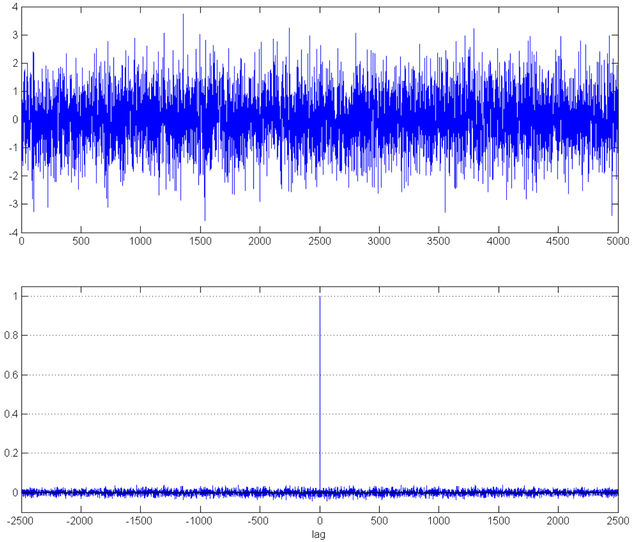


Figure 4.12: Differenced Random Walk (from bottom of Figure 4.11) $\nabla x_t = z_t$ (top) and its sample ACF (bottom).

§4.5.3 Explosive AR(1) Model and Causality

- As we have seen in the previous section, random walk, which is AR(1) with $\phi = 1$, is not a stationary process. So, the question is if a stationary AR(1) process with $|\phi| > 1$ exists? Also, what are the properties of AR(1) models for $\phi > 1$?
- Clearly, the sum $\sum_{j=0}^{k-1} \phi^j Z_{t-j}$ will not converge in mean square sense as $k \to \infty$ and we will not get a linear process representation of the AR(1).

However, if $|\phi| > 1$ then $\frac{1}{|\phi|} < 1$ and we can express a past value of the TS in terms of a future value rewriting

$$X_{t+1} = \phi X_t + Z_{t+1}$$

as

$$X_t = \phi^{-1} X_{t+1} - \phi^{-1} Z_{t+1}.$$

Then, substituting for X_{t+j} several times we obtain

$$\begin{aligned} X_t &= \phi^{-1} X_{t+1} - \phi^{-1} Z_{t+1} \\ &= \phi^{-1}(\phi^{-1} X_{t+2} - \phi^{-1} Z_{t+2}) - \phi^{-1} Z_{t+1} \\ &= \phi^{-2} X_{t+2} - \phi^{-2} Z_{t+2} - \phi^{-1} Z_{t+1} \\ &= \ldots \\ &= \phi^{-k} X_{t+k} - \sum_{j=1}^{k} \phi^{-j} Z_{t+j} \end{aligned}$$

- Since $|\phi^{-1}| < 1$ and since we seek a stationary process, we obtain

$$X_t = -\sum_{j=1}^{\infty} \phi^{-j} Z_{t+j},$$

which is a future dependent stationary TS. This however, does not have much practical meaning because it requires knowledge of future values to define the present value.

- When the current value of a process does not involve any observations from the future, e.g. AR(1) with $|\phi| < 1$, we say that such a process is **causal**.
- Figure 4.13 shows a simulated causal series $x_t = 1.02x_{t-1} + z_t$. As we can see the values of the time series quickly become large in magnitude, even for ϕ just slightly above 1. Such process is called **explosive**.

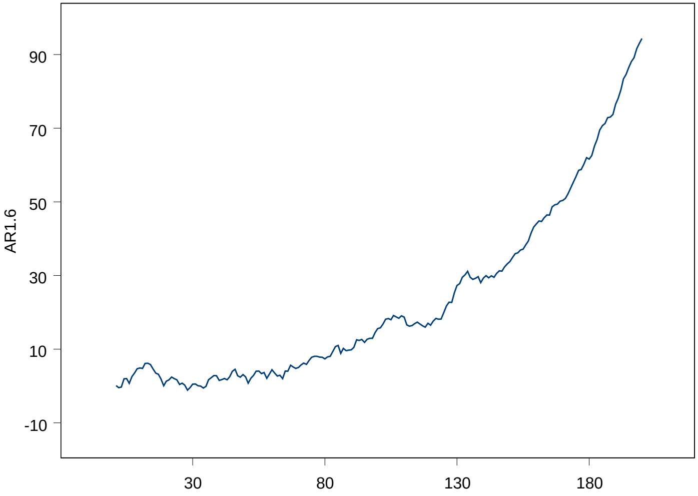


Figure 4.13: A simulated causal explosive AR(1): $x_t = 1.02x_{t-1} + z_t$.

§4.6 Autoregressive Moving Average Model ARMA(1,1)

This section is an introduction to a wide class of models ARMA(p,q) which we will consider in more detail later in this course.

The special case, ARMA(1,1), is defined by linear difference equations with constant coefficients as follows.

Definition 4.10

Let ϕ and θ be non-zero constants. A TS $\{X_t\}$ is an ***ARMA(1,1) process*** *if it satisifies*

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}, \quad \text{for every } t,\ -\infty < t < \infty, \tag{4.31}$$

where $\{Z_t\} \sim WN(0, \sigma^2)$.

Such a model may be viewed as a generalization of the two previously introduced models AR(1) and MA(1). Compare

$$\begin{aligned} \text{AR(1)} \quad & X_t = \phi X_{t-1} + Z_t \\ \text{MA(1)} \quad & X_t = Z_t + \theta Z_{t-1} \\ \text{ARMA(1,1)} \quad & X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1} \end{aligned}$$

Hence, when $\phi = 0$ then ARMA(1,1) $\equiv$ MA(1) and we denote such a process as ARMA(0,1). Similarly, when $\theta = 0$ then ARMA(1,1) $\equiv$ AR(1) and we denote such process as ARMA(1,0).

Here, as in the MA and AR models, we can use the backshift operator to write the ARMA model more concisely as

$$\phi(B)X_t = \theta(B)Z_t, \tag{4.32}$$

where $\phi(B)$ and $\theta(B)$ are the linear filters

$$\phi(B) = 1 - \phi B, \quad \theta(B) = 1 + \theta B.$$

§4.6.1 Causality and invertibility of ARMA(1,1)

For which values of the parameters ϕ and θ does a stationary causal ARMA(1,1) process exist?

What about invertibility of ARMA(1,1)?

A solution to 4.31, or to 4.32, can be formally written as

$$X_t = \frac{1}{\phi(B)}\theta(B)Z_t.$$

However, for $|\phi| < 1$ we have (see Remark 4.15)

$$
\begin{aligned}
\frac{1}{\phi(B)}\theta(B) &= (1 + \phi B + \phi^2 B^2 + \phi^3 B^3 + \dots)(1 + \theta B) \\
&= 1 + \phi B + \phi^2 B^2 + \phi^3 B^3 + \dots + \theta B + \phi\theta B^2 + \\
&\quad \phi^2\theta B^3 + \phi^3\theta B^4 + \dots \\
&= 1 + (\phi + \theta)B + (\phi^2 + \phi\theta)B^2 + (\phi^3 + \phi^2\theta)B^3 + \dots \\
&= 1 + (\phi + \theta)B + (\phi + \theta)\phi B^2 + (\phi + \theta)\phi^2 B^3 + \dots \\
&= \sum_{j=0}^{\infty} \psi_j B^j,
\end{aligned}
$$

where $\psi_0 = 1$ and $\psi_j = (\phi + \theta)\phi^{j-1}$ for $j = 1, 2, \dots$.

Thus, we can write the solution to 4.32 in the form of an MA(∞) model, i.e.,

$$X_t = Z_t + (\phi + \theta) \sum_{j=1}^{\infty} \phi^{j-1} Z_{t-j}. \tag{4.33}$$

By Corollary 4.1, this is a stationary process, and it is evidently causal.

Now, suppose that $|\phi| > 1$. Then, by similar arguments as in the AR(1) model, it can be shown that

$$X_t = -\theta\phi^{-1} Z_t - (\phi + \theta) \sum_{j=1}^{\infty} \phi^{-j-1} Z_{t+j}.$$

Here too, we obtained a noncausal process which depends on future noise values, hence of no practical value.

- If $|\phi| = 1$ then there is no stationary solution to 4.32 (neither causal nor non-causal).
- While causality means that the process $\{X_t\}$ is expressible in terms of past and present values of $\{Z_t\}$, the dual property of invertibility means that the process $\{Z_t\}$ is expressible in the past and present values of $\{X_t\}$.
- Is ARMA(1,1) invertible?

ARMA(1,1) model is

$$\phi(B)X_t = \theta(B)Z_t.$$

Writing the solution for Z_t we have

$$Z_t = \frac{1}{\theta(B)}\phi(B)X_t = \frac{1}{1+\theta B}(1-\phi B)X_t. \tag{4.34}$$

It can be made rigorous that the inverse of the operator $(1 + \theta B)$ exists if and only if $|\theta| < 1$, in which case

$$\frac{1}{1 + \theta B} = \sum_{j=0}^{\infty} (-\theta)^j B^j,$$

When this infinite expansion is applied to 4.34, it gives

$$\begin{aligned} Z_t &= \sum_{j=0}^{\infty} (-\theta)^j B^j (1 - \phi B) X_t \\ &= X_t - (\phi + \theta) \sum_{j=1}^{\infty} (-\theta)^{j-1} X_{t-j}. \end{aligned}$$

Thus, ARMA(1,1) is invertible if and only if $|\theta| < 1$.

When combined, these two properties, causality and invertibility, determine the admissible region for the values of parameters ϕ and θ, which is the square

$$-1 < \phi < 1$$

$$-1 < \theta < 1.$$

§4.6.2 ACVF and ACF of ARMA(1,1)

The fact that we can express an ARMA(1,1) TS as a linear process of the MA form

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j},$$

where Z_t is a white noise, is very helpful in deriving the ACVF and ACF of the process. By Corollary 4.1 we have

$$\gamma(\tau) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|\tau|}.$$

For a stationary ARMA(1,1) the coefficients ψ_j are given in §(5) as follows:

$$\psi_0 = 1$$

$$\psi_j = (\phi + \theta)\phi^{j-1} \quad \text{for } j = 1, 2, \ldots$$

Knowing ψ_j, we can now derive expressions for $\gamma(0)$ and $\gamma(1)$.

$$
\begin{aligned}
\gamma(0) &= \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 \\
&= \sigma^2 \left[1 + (\phi + \theta)^2 \sum_{j=1}^{\infty} \phi^{2(j-1)} \right] \\
&= \sigma^2 \left[1 + (\phi + \theta)^2 \sum_{j=0}^{\infty} \phi^{2j} \right] \\
&= \sigma^2 \left[1 + \frac{(\phi + \theta)^2}{1 - \phi^2} \right]
\end{aligned}
$$

and

$$
\begin{aligned}
\gamma(1) &= \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+1} \\
&= \sigma^2 \Big[(\phi+\theta) + (\phi+\theta)(\phi+\theta)\phi + (\phi+\theta)\phi(\phi+\theta)\phi^2 + \\
&\quad (\phi+\theta)\phi^2(\phi+\theta)\phi^3 + \ldots \Big] \\
&= \sigma^2 \Big[(\phi+\theta) + (\phi+\theta)^2\phi(1+\phi^2+\phi^4+\ldots) \Big] \\
&= \sigma^2 \left[(\phi+\theta) + (\phi+\theta)^2\phi \sum_{j=0}^{\infty} \phi^{2j} \right] \\
&= \sigma^2 \left[(\phi+\theta) + \frac{(\phi+\theta)^2\phi}{1-\phi^2} \right]
\end{aligned}
$$

Similar derivations for $|\tau| \geq 2$ give

$$\gamma(\tau) = \phi^{|\tau|-1}\gamma(1). \tag{4.35}$$

Hence, we can calculate the autocorrelation function $\rho(\tau)$. For $|\tau| = 1$ we obtain

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{(\phi + \theta)(1 + \phi\theta)}{1 + 2\phi\theta + \theta^2} \tag{4.36}$$

and for $|\tau| \geq 2$ we have

$$\rho(\tau) = \phi^{|\tau|-1}\rho(1). \tag{4.37}$$

From these formulae we can see that when $\phi = -\theta$ the ACF $\rho(\tau) = 0$ for $\tau = 1, 2, \ldots$ and the process is just a white noise.

Graph 4.14 shows the admissible region for the parameters ϕ and θ for a stationary ARMA(1,1) series, and indicates the regions when we have special cases of ARMA(1,1), which are white noise, AR(1) and MA(1).

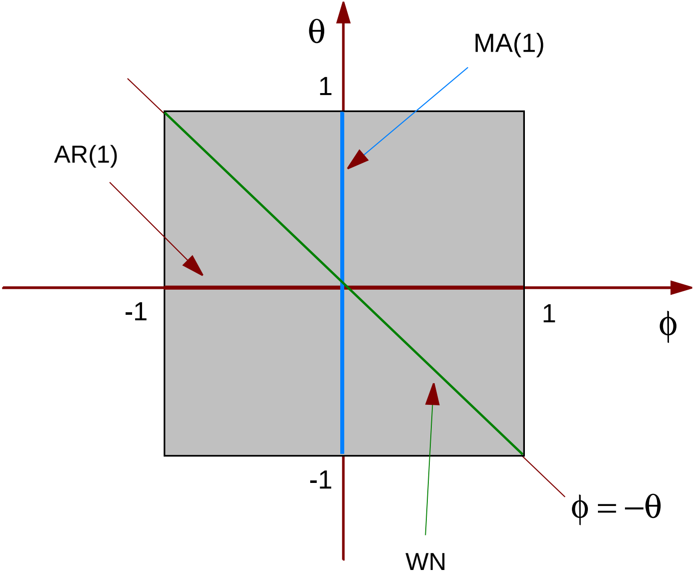


Figure 4.14: Admissible parameter region for ARMA(1,1)

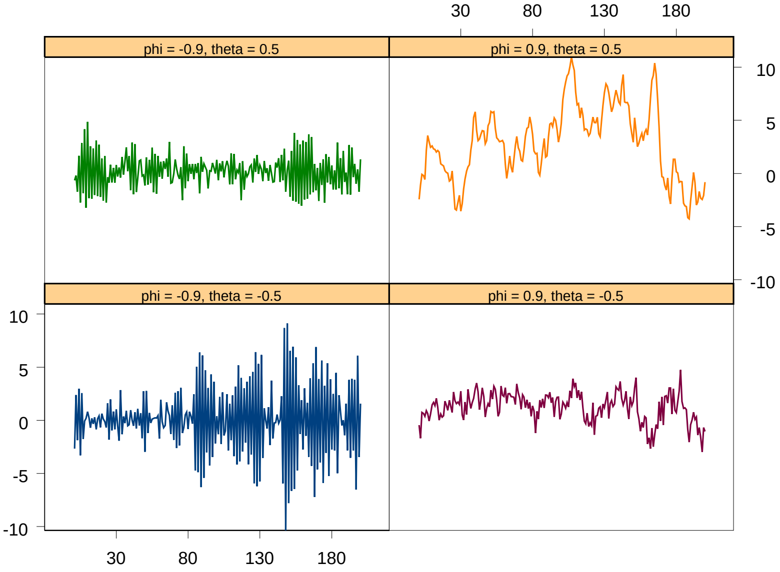


Figure 4.15: ARMA(1,1) for various values of the parameters ϕ and θ.

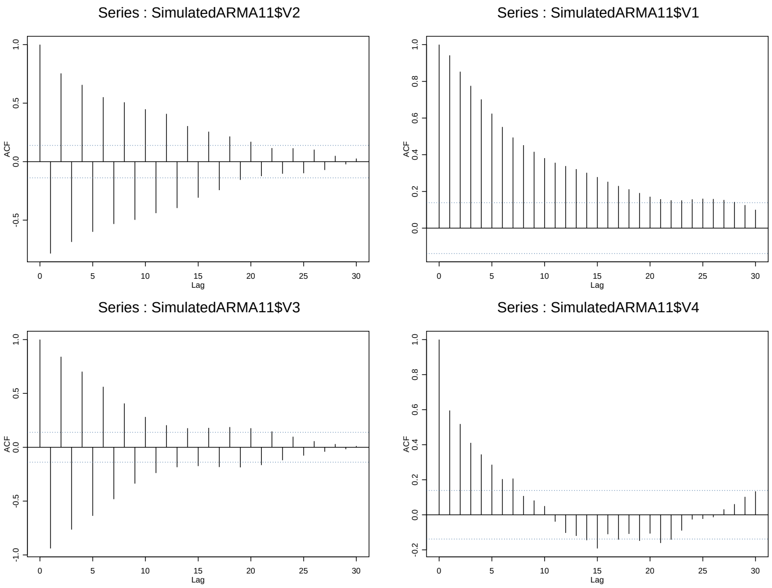


Figure 4.16: ACF of the ARMA(1,1) processes with the parameter values as in Figure 4.15, respectively.

Box, G.E.P., & Jenkins, G.M. 1976. Time series analysis: Forecasting and control. revised edition. Holden Day.

Brockwell, P.J., & Davis, R.A. 2002. An introduction to time series and forecasting. second edition. Springer-Verlag.

Chatfield, C. 2004. The analysis of time series: An introduction. sixth edition. Chapman and Hall.

J.Hansen, & S.Lebedeff. 1987. Global trends of measured surface air temperature. Journal of geophysical research, **92**, 13.345–13.372.