Time Series (MATH5845)

Dr. Atefeh Zamani
Based on the notes by Prof. William T.M. Dunsmuir

T2 2025

Chapter 3

Simple Models For Time Series

Contents

In this chapter, we are going to

- present some simple models in time series:
  - Autoregressive of order 1,
  - Moving average of order 1,
  - mixed autoregressive-moving average of order (1,1).
- study the basic properties of these series
- do some inference about the mean and discuss the impact of autocorrelation on standard errors.

3.1 Some Simple Models

Lets have a quick review of the definitions of white noise and i.i.d. processes.

White Noise process: Let $\{X_t\}$ be a sequence of uncorrelated random variables with $\mu_X = 0$ and $Var(X_t) = \sigma^2$, abbreviated as $\{X_t\} \sim WN(0, \sigma^2)$. More precisely,

$$\gamma_X(h) = \begin{cases} \sigma^2 & h = 0 \\ 0 & h \neq 0 \end{cases}.$$

weakly stationary

This process is not strictly stationary in general but it is weakly stationary.

i.i.d. Noise Process: Let $\{X_t\}$ be a sequence of i.i.d. random variables with $\mu_X = 0$ and $Var(X_t) = \sigma^2$, abbreviated as $\{X_t\} \sim$i.i.d.$(0, \sigma^2)$. More precisely,

$$\gamma_X(h) = \begin{cases} \sigma^2 & h = 0 \\ 0 & h \neq 0 \end{cases},$$

This process is a strictly stationary time series.

3.1.1 Martingale and Martingale Difference

Note 3.1 *This section is based on Fabozzi et al. [2006].*

Definition 3.1 (Martingale) *A stochastic process $\{Y_t, t \in \mathbb{Z}\}$ is said to be a martingale if the following conditions hold:*

- $E|Y_t| < \infty$,
- * $E(Y_t|\overbrace{\sigma(Y_{t-1}, Y_{t-2}, \ldots, Y_1)}^{\text{information from past}}) = Y_{t-1}$, *immediate past.*

conditional expectation

$$E(X_t) = E(Y_t - Y_{t-1})$$
$$= E\,E(Y_t - Y_{t-1} | Y_{t-1}, \ldots)$$
$$= E\,\underbrace{E(Y_t | Y_{t-1}, \ldots)}_{\text{martingale } *} - E\,\underbrace{E(Y_{t-1} | Y_{t-1}, \ldots)}_{Y_{t-1}}$$
$$= E(Y_{t-1}) - E(Y_{t-1})$$
$$= 0 \checkmark \ (1)$$

Now, consider the difference sequence X_t defined as follows:

$$X_1 = Y_1, X_2 = Y_2 - Y_1, \ldots, X_t = Y_t - Y_{t-1}, \ldots.$$

Given the martingale property, we can see that X_t is a zero-mean (1), uncorrelated (2) process with the additional property that

$$(**) \quad \boxed{E(X_{t+1}|X_t, X_{t-1}, \ldots, X_1) = 0.}$$

correct

$x > s$

$$\mathrm{cov}(X_t, X_s) = E(X_t X_s) - E(X_t)E(X_s) = E(X_t X_s) = E(E(X_t X_s | \underbrace{X_{t-1}, X_{t-2}, \ldots}_{s<t}))$$
$$= E(X_s E(X_t | X_{t-1}, \ldots)) \overset{(**)}{=} E(X_s \cdot 0) = 0 \ (2)$$

3.1.1 Martingale and Martingale Difference

Note 3.1 *This section is based on Fabozzi et al. [2006].*

Definition 3.1 (Martingale) *A stochastic process* $\{Y_t,\ t \in \mathbb{Z}\}$ *is said to be a martingale if the following conditions hold:*

- $E|Y_t| < \infty$,
- $E(Y_t|\sigma(Y_{t-1}, Y_{t-2}, \ldots, Y_1)) = Y_{t-1}$, *immediate past.*

Now, consider the difference sequence X_t defined as follows:

$$X_1 = Y_1, X_2 = Y_2 - Y_1, \ldots, X_t = Y_t - Y_{t-1}, \ldots.$$

Given the martingale property, we can see that X_t is a zero-mean, uncorrelated process with the additional property that

$$E(X_{t+1}|X_t, X_{t-2}, \ldots, X_1) = 0.$$

3.1.1 Martingale and Martingale Difference

Note 3.1 *This section is based on Fabozzi et al. [2006].*

Definition 3.1 (Martingale) *A stochastic process* $\{Y_t,\ t \in \mathbb{Z}\}$ *is said to be a martingale if the following conditions hold:*

- $E|Y_t| < \infty$,
- $E(Y_t|\sigma(Y_{t-1}, Y_{t-2}, \ldots, Y_1)) = Y_{t-1}$, *immediate past.*

Now, consider the difference sequence X_t defined as follows:

$$X_1 = Y_1, X_2 = Y_2 - Y_1, \ldots, X_t = Y_t - Y_{t-1}, \ldots.$$

Given the martingale property, we can see that X_t is a zero-mean, uncorrelated process with the additional property that

$$E(X_{t+1}|X_t, X_{t-2}, \ldots, X_1) = 0.$$

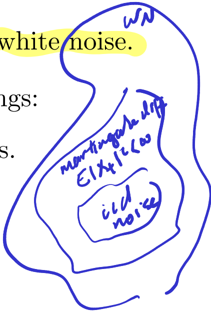
These properties can be used to define a martingale difference sequence.

Definition 3.2 (Martingale difference) *Martingale difference sequence is a zero mean, uncorrelated process such that its conditional mean at every step is zero.*

Note that the definition of martingale difference sequence is more restrictive than the definition of white noise insofar as not only its unconditional expectation is zero, as in the white noise case, but its conditional expectation is zero at every step.

- Any i.i.d. noise process with finite second moment is a martingale difference sequence.

- Any martingale difference sequence with finite second moment is white noise.



The converse, however, is not true. In fact, there exist both the followings:

- White noise processes that are not martingale difference sequences.

- Martingale difference sequences that are not i.i.d. sequences.

Note that

- There are white noise processes that have, for some time step, a conditional mean different from zero though the unconditional mean is always zero.
- Martingale difference sequences, though uncorrelated, are not independent.
- For **Gaussian processes with mean zero** the three concepts coincide.

Example 3.1 *Consider an i.i.d. process* $\{Z_t\} \sim i.i.d.(0, \sigma^2)$. *The process defined by the partial sums* $X_t = Z_1 + \ldots + Z_t = \sum_{j=1}^{t} Z_j$, $t = 1, 2, \ldots,$ *is a martingale because* $X_t = X_{t-1} + Z_t$ *and* $\sigma(X_{t-1}, \ldots, X_1) = \sigma(Z_{t-1}, \ldots, Z_1)$, *which results in that*

random walk with no drift

$$E(X_t|\sigma(X_{t-1}, \ldots, X_1)) = X_{t-1}$$

Note that in the case where $E(Z_t) = \mu \neq 0$, $t = 1, 2, \ldots,$ $\{X_t,\ t = 1, 2, \ldots\}$ *is no longer a martingale but the process* $\{Y_t = (X_t - \mu t),\ t = 1, 2, \ldots\}$ *with* $X_0 = 0$ *is a martingale.*

① $E|X_t| < \infty$

② $E(X_t \mid \sigma(X_{t-1}, \ldots)) = X_{t-1}$

$X_t = \sum_{i=1}^{t} Z_i$

$Z_i \sim iid(0, \sigma^2)$

$$E(X_t \mid \sigma(X_{t-1}, \ldots))$$

$$= E\left(\sum_{i=1}^{t} Z_i \mid \sigma(Z_{t-1}, \ldots)\right)$$

$$= E\left(\sum_{i=1}^{t-1} Z_i + Z_t \mid \sigma(Z_{t-1}, \ldots)\right)$$

$$= \sum_{i=1}^{t-1} Z_i + \underbrace{E(Z_t \mid \sigma(Z_{t-1}, \ldots))}_{E(Z_t) = 0}$$

?!

Z_t are iid

$t \neq s$ $E(Z_t \mid Z_s)$
$= E(Z_t)$
$= 0$

$$= X_{t-1}$$

3.1.2 The Moving Average of order 1 (MA(1))

Let $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$ and define, for $\theta \in \mathbb{R}$,

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

This process has mean function $\mu_X(t) \equiv 0$ and covariance function

$$\gamma_X(h) = \begin{cases} (1+\theta^2)\sigma^2 & h = 0 \\ \theta\sigma^2 & |h| = 1 \\ 0 & |h| > 1 \end{cases}$$

$$E(X_t) = E(Z_t + \theta Z_{t-1})$$
$$= \underbrace{E(Z_t)}_{0} + \theta \underbrace{E(Z_{t-1})}_{0} = 0 \quad ✓$$

$$\mathrm{cov}(X_t, X_{t+h}) = \mathrm{cov}(Z_t + \theta Z_{t-1}, Z_{t+h} + \theta Z_{t+h-1})$$

$$\frac{\mathrm{cov}(Z_t, Z_s)}{\gamma_Z(t-s)} = \begin{cases} \sigma^2 & t = s \\ 0 & t \neq s \end{cases}$$

$$= \mathrm{cov}(Z_t, Z_{t+h}) + \theta\, \mathrm{cov}(Z_t, Z_{t+h-1})$$

$$+ \theta\, \mathrm{cov}(Z_{t-1}, Z_{t+h}) + \theta^2 \mathrm{cov}(Z_{t-1}, Z_{t+h-1})$$

$$= \gamma_Z(h) + \theta\, \gamma_Z(h-1) + \theta \gamma_Z(h+1) + \theta^2 \gamma_Z(h)$$

$$= \begin{cases} \gamma_Z(0) + \theta^2 \gamma_Z(0) & h = 0 \\ \theta \gamma_Z(0) & h = 1 \\ \theta \gamma_Z(0) & h = -1 \\ 0 & |h| > 1 \end{cases}$$

$$= \begin{cases} (1+\theta^2)\sigma^2 & h = 0 \\ \theta \sigma^2 & h = \pm 1 \\ 0 & |h| > 1 \end{cases}$$

ACF

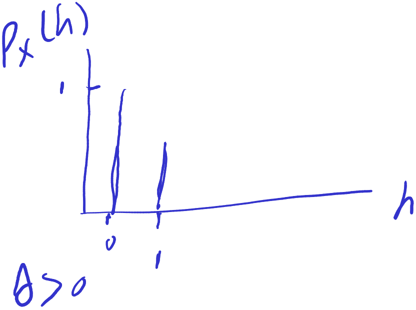
$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}$$

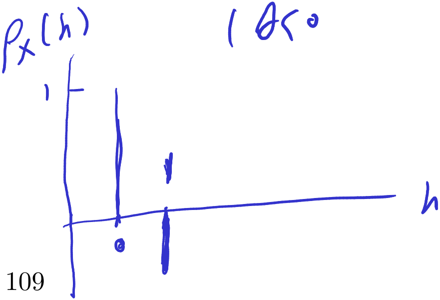
The autocorrelation function of X_t is

$$\rho_X(h) = \begin{cases} 1 & h = 0 \\ \frac{\theta}{(1+\theta^2)} & |h| = 1 \\ 0 & |h| > 1 \end{cases}$$ ✓

The process X_t is

- strictly stationary if $\{Z_t\} \sim$i.i.d.$(0, \sigma^2)$,
- weakly stationary if $\{Z_t\} \sim WN(0, \sigma^2)$.





Note 3.2 *To generate data from moving average time series, you can use the function* `filter` *from the* `stats` *package. The syntax of this function is*

```
filter(x, filter, method = c("convolution", "recursive"),sides = 2,
   circular = FALSE, init)
```

MA AR

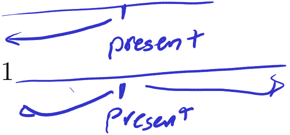
where

weights

- `x`*: a univariate or multivariate time series.*
- `filter`*: a vector of filter coefficients in reverse time order*
- `method`*: Either "convolution" (for MA) or "recursive" (for AR)*
- `sides`*: for convolution filters only. If sides = 1, coefficients are for past; if sides = 2, coefficients are centred around lag 0 (length of filter should be odd, if even, more of the filter is forward in time).*
- `circular`*: for convolution . If TRUE, wrap the filter around the ends of the series, otherwise assume external values are missing (NA).*

sides = 1

sides = 2



- `init`*: for recursive. Specifies the initial values of the time series. The default is a set of zeros.*

Example 3.2 (Simulated time series from MA(1))

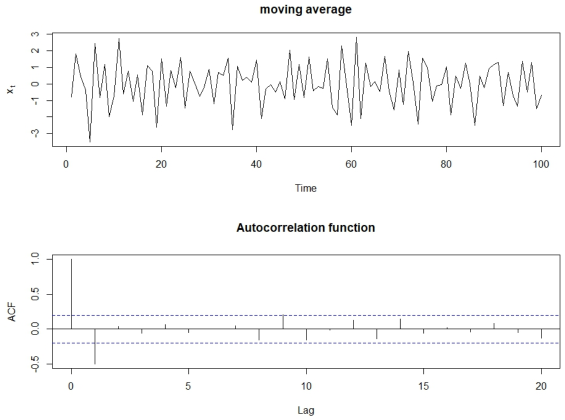


Figure 3.1: 100 simulated observations from $X_t = Z_t - 0.8Z_{t-1}$, with $Z_t \sim N(0, 1)$ along with the ACF

Example 3.3 *Consider the monthly employment in Trades Y_t. The combined seasonal and lag 1 differences appear to be stationary in the mean and variance.*

$Z_t = \nabla Y_t = Y_t - Y_{t-1} \longrightarrow$ drop linear trend

$X_t = \nabla_{12} Z_t = Z_t - Z_{t-12}$

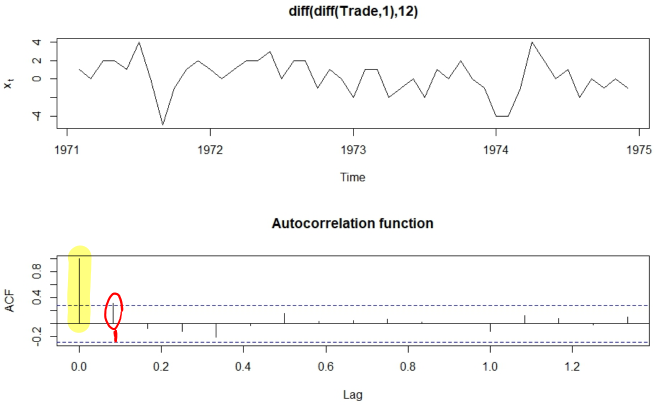


Figure 3.2: $\nabla_{12}\nabla Y_t$ along with the ACF.

Let

$$X_t = \nabla_{12}\nabla Y_t.$$

A good model for $\{X_t\}$ *is the MA(1) with* $\theta = 0.4173$ *and* $\sigma^2 = 3.078$. *Hence*

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + Z_t + 0.42Z_{t-1}$$

where $\{Z_t\} \sim i.i.d.N(0, 3.078)$.

3.1.3 General Linear Processes

We say $\{X_t\}$ is a general linear process if it can be represented as

weights

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad t = 0, \pm 1, \pm 2, \ldots . \tag{3.1}$$

MA(1)

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

$\psi_0 = 1$

$\psi_1 = \theta$

$\psi_j = 0$

where $\{Z_t\} \sim WN(0, \sigma^2)$ and $\{\psi_j\}$ is an absolutely summable sequence of constants:

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

$j = 0$ present

$j = -\infty, \ldots, -1$ future

$j = 1, 2, \ldots$ past

Note that $\{X_t\}$ is weakly stationary because $\mu_X(t) \equiv 0$ and

$$\gamma_X(h) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h}$$

both of which do not depend upon t. A special case of a linear process is the MA(∞) in which $\psi_j = 0$, $j < 0$.

$$E(X_t) = E\left(\sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}\right) = \sum_{j=-\infty}^{\infty} \psi_j \underbrace{E(Z_{t-j})}_{0} = 0 \quad \checkmark$$

$$cov(X_t, X_{t+h}) = cov\left(\sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \sum_{k=-\infty}^{\infty} \psi_k Z_{t+h-k}\right)$$

$$cov(Z_t, Z_s) = \begin{cases} 0 & t \neq s \\ \sigma^2 & t = s \end{cases}$$

$$= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \underbrace{cov(Z_{t-j}, Z_{t+h-k})}_{*}$$

$$cov(Z_{t-j}, Z_{t+h-k}) = \begin{cases} 0 & t-j \neq t+h-k \\ \sigma^2 & t-j = t+h-k \end{cases}$$

$$\cancel{t}-j = \cancel{t}+h-k \Rightarrow k = h+j \; *$$

$$= \sum_{j=-\infty}^{\infty} \psi_j \psi_{h+j} \underbrace{\gamma_Z(0)}_{\sigma^2}$$

$$= \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{h+j}$$ is a function of h

$\Rightarrow X_t$ is stationary

$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} = \sum_{j=-\infty}^{\infty} \psi_j B^j Z_t = \psi(B)$ general linear model

Using the backshift operator notation we can write the MA(∞) process as

$BX_t = X_{t-1}$

$B^j X_t = X_{t-j}$

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

where

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

Note 3.3 *The condition $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ ensures that the infinite sum in Equation* (3.1) *converges (with probability one), since $E|Z_t| \leq \sigma$ and*

$$E|X_t| \leq \sum_{j=-\infty}^{\infty} |\psi_j| E|Z_{t-j}| \leq \left(\sum_{j=-\infty}^{\infty} |\psi_j| \right) \sigma < \infty.$$

It also ensures that $\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$ and hence the series in (3.1) *converges in mean square, i.e., that X_t is the mean square limit of the partial sums $\sum_{j=-n}^{n} \psi_j Z_{t-j}$.*

Example 3.4 *Consider a white noise series* W_t, $t = 1, \ldots, 300$, *and let* X_t *be an average of* W_t *and its immediate neighbors in the past and future:*

$$X_t = \frac{1}{3}\left(\overset{\text{past}}{W_{t-1}} + \overset{\text{present}}{W_t} + \overset{\text{future}}{W_{t+1}}\right) \tag{3.2}$$

Figure 3.3 demonstrates a simulated white noise series along with X_t *and its auto-correlation function. It can be observed that* X_t *shows a smoother version of* W_t, *reflecting the fact that the slower oscillations are more apparent and some of the faster oscillations are taken out.*

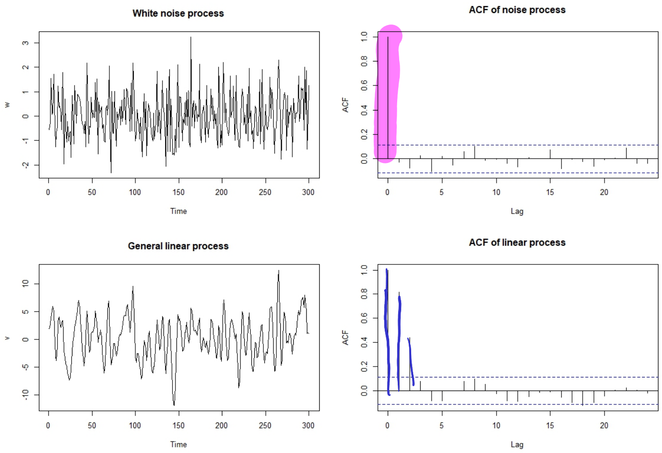


Figure 3.3: Simulated white noise and the resulting general linear processes along with their autocorrelation functions.

3.1.4 The Autoregression of order 1 (AR(1))

general linear model $X_t = \sum_{j=-\infty} \alpha_j Z_{t-j}$

Let $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$ and define, for $|\phi| < 1$,

$$X_t = \phi X_{t-1} + Z_t, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{3.3}$$

where $cov(X_s, Z_t) = 0$ (uncorrelated) for each $s < t$. The stationary solution to this system is corr between X_s and future noise

$X_t = \phi X_{t-1} + Z_t$
$= \phi(\phi X_{t-2} + Z_{t-1}) + Z_t$

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j} \tag{3.4}$$

Q 3.4

$\psi_0 = 1$

$\psi_j = \phi^j \quad j = 1, 2, \ldots$

$\psi_j = 0 \quad j = -\infty, \ldots, -1$

from which it follows that $\mu_X(t) \equiv 0$ and

$$\gamma_X(h) = \frac{\sigma^2 \phi^h}{1 - \phi^2}, \quad \rho_X(h) = \phi^h \tag{3.5}$$

which are geometrically decaying functions of separation lag h.

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \qquad \gamma_X(h) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h}$$

Consider X_t to be AR(1)

$$\sum_{j=0}^{\infty} r a^j = \frac{r}{1-a} \quad |a|<1$$

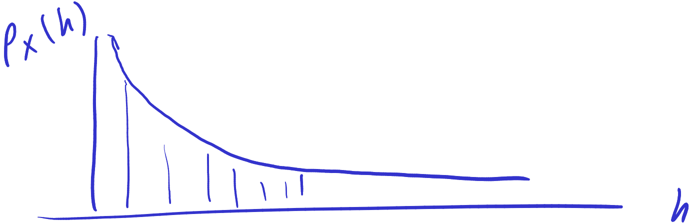
$$\gamma_X(h) = \sigma^2 \sum_{j=0}^{\infty} \phi^j \phi^{j+h}$$

$$= \sigma^2 \phi^h \sum_{j=0}^{\infty} \phi^{2j}$$

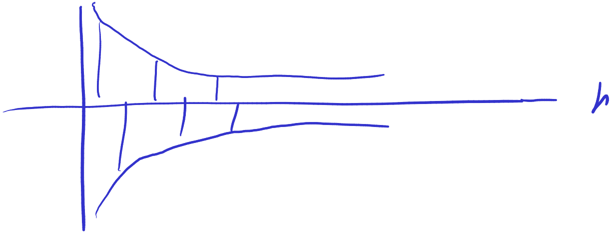
$$= \sigma^2 \phi^h \frac{1}{1-\phi^2} = \begin{cases} \dfrac{\sigma^2}{1-\phi^2} & h=0 \\[2ex] \dfrac{\sigma^2 \phi^h}{1-\phi^2} & |h|>0 \end{cases}$$

ACF

$$\Rightarrow \quad \rho_x(h) = \frac{\gamma_x(h)}{\gamma_x(0)} = \begin{cases} 1 & h = 0 \\ \phi^h & |h| > 0 \end{cases}$$



$0 < \phi < 1$



$-1 < \phi < 0$

Theorem 3.1 *Let* $\{Z_t\} \sim WN(0, \sigma^2)$ *and* $|\phi| < 1$. *Then the unique stationary solution to* $X_t = \phi X_{t-1} + Z_t$, $t = 0, \pm 1, \pm 2, \ldots,$ *is*

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

Proof. Consists of two steps. First show that this is a solution and it is stationary. Second show that it is unique. ■

- **Causal Solutions:** X_t can be expressed in terms of Z_s for $s \leq t$.
- **Non-causal Solutions:** X_t can be expressed in terms of Z_s for $s > t$.

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

- if $|\phi| < 1$, there is a unique causal stationary solution $X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$,
- if $|\phi| = 1$, there is no stationary solution,
- if $|\phi| > 1$, there is a unique non-causal stationary solution $X_t = -\sum_{j=1}^{\infty} \phi^{-j} Z_{t+j}$.

In order that time series models be useful for **forecasting** we require them to have causal solutions.

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

$\phi = 1$ $\longrightarrow$ $X_t = X_{t-1} + Z_t \longrightarrow$ not stationary

random walk

Example 3.5 (Simulated time series from AR(1)) *Figure 3.4 demonstrates a realization of* $\{X_1, \ldots, X_{100}\}$ *with* $\phi = 0.9$ *and* $Z_t \sim N(0, 1)$ *along with its autocorrelation function.*

$x_t = 0.9x_{t-1} + Z_t$

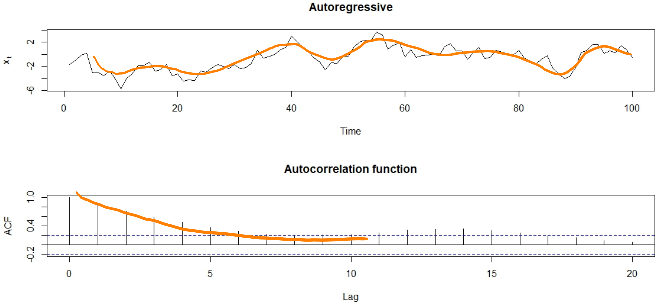


Figure 3.4: 100 simulated observations from $X_t = 0.9X_{t-1} + Z_t$ along with the ACF.

3.1.5 The Autoregressive Moving Average of order (1,1) (ARMA(1,1))

This time series, $\{X_t\}$, which is the mixed autoregressive and moving average times series, is denoted by ARMA(1,1), and obtained as the stationary solution to

order of MA

order of AR

$$X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{3.6}$$

where $Z_t \sim WN(0, \sigma^2)$ and $\phi + \theta \neq 0$.

$$B X_t = X_{t-1}$$

$$X_t - \Phi B X_t = Z_t + \theta B Z_t$$

$$\Rightarrow \underbrace{(1 - \Phi B)}_{\Phi(B)} X_t = \underbrace{(1 + \theta B)}_{\theta(B)} Z_t$$

Autoregressive operator

moving average operator

causal form

$$X_t = \underbrace{\Phi^{-1}(B)\,\theta(B) Z_t}_{1}$$

invertable

$$Z_t = \theta^{-1}(B)\,\Phi(B) X_t$$

AR operator $\phi(B) = (1-\phi B)$
MA operator $\theta(B) = (1+\theta B)$

The range of possible values of ϕ

Let $\phi(B) = 1 - \phi B$ and $\theta(B) = 1 + \theta B$.

- If $|\phi| < 1$, using the power series expansion, $\frac{1}{\phi(z)} = \frac{1}{1-\phi z} = \sum_{j=0}^{\infty} \phi^j z^j$, which has absolutely summable coefficients. Therefore,

1
$$\psi(B) = \phi(B)^{-1}\theta(B) = (1 + \phi B + \phi^2 B^2 + \cdots)(1 + \theta B) \equiv \sum_{j=0}^{\infty} \psi_j B^j.$$

By rewriting the terms, it can be shown that

$$\psi_j = \begin{cases} 1 & j = 0 \\ (\phi + \theta)\phi^{j-1} & j \geq 1 \end{cases},$$

and consequently, we conclude that the $MA(\infty)$ process

$$X_t = \psi(B)Z_t = Z_t + (\phi + \theta) \sum_{j=1}^{\infty} \phi^{j-1} Z_{t-j} \tag{3.7}$$

is the unique stationary solution of Equation (3.6).

$$\sum_{j=0}^{\infty} \phi^j B^j$$

① $\psi(B) = \phi^{-1}(B)\,\theta(B) = \frac{1}{\phi(B)}\,\theta(B)$

$$= (1+\phi B+\phi^2 B^2+\cdots)(1+\theta B)$$

$$= 1 + \underbrace{\phi B+\theta B}_{(\phi+\theta)B} + \underbrace{\phi^2 B^2 + \phi\theta B^2}_{\phi(\phi+\theta)B^2} + \cdots$$

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

 ARMA(1,1) $X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}$

$(1-\phi B) X_t = (1+\theta B) Z_t$

$X_t = \frac{1+\theta B}{1-\phi B} Z_t$

- If $|\phi| > 1$, using the geometric series $1/(1-x) = \sum_{j=0}^{\infty} x^j$ for $|x| < 1$, we can show that $1/(1-\phi z) = -\sum_{j=1}^{\infty} \phi^{-j} z^{-j}$ for $|\phi| > 1$ and $|z| \geq 1$. Therefore,

$$\psi(B) = \phi(B)^{-1}\theta(B) = \underbrace{(-\phi^{-1}B^{-1} - \phi^{-2}B^{-2} + \cdots)(1+\theta B)}_{-\phi^{-1}\theta \quad j=0} \equiv \sum_{j=0}^{\infty} \psi_j B^{-j}.$$

By rewriting the terms, it can be shown that

$$\psi_j = \begin{cases} -\theta\phi^{-1} & j = 0 \\ -(\phi+\theta)\phi^{-j-1} & j \geq 1 \end{cases},$$

and consequently,

$$X_t = -\theta\phi^{-1} Z_t - (\phi+\theta) \sum_{j=1}^{\infty} \phi^{-j-1} Z_{t+j}. \tag{3.8}$$

linear process

it is based on future and present

$\Rightarrow$ not causal

- If $\phi = \pm 1$, there is no stationary solution of Equation (3.6). Consequently, there is no such thing as an ARMA(1,1) process with $\phi = \pm 1$.

Note 3.4 *In summary,*

this course

1. *If $|\phi| < 1$, this is the unique stationary solution and it is causal since it depends only on Z_s for $s \leq t$.*

2. *If $|\phi| = 1$, there is no stationary solution.*

3. *$|\phi| > 1$, there is a unique stationary solution in terms of future values Z_s for $s > t$ and hence this solution is non-causal.*

$(1-\phi B)X_t = (1+\theta B)Z_t$

$\Rightarrow Z_t = \frac{1-\phi B}{1+\theta B} X_t$ ✓

The range of possible values of θ

Along with causality we have the concept of **invertibility**, which means that Z_t can be expressed in terms of X_s, $s \leq t$. The ARMA(1,1) process is invertible if $|\theta| < 1$.
Proof. Consider the power series expansion of $1/\theta(z)$, i.e., $\sum_{j=0}^{\infty}(-\theta)^j z^j$, which has absolutely summable coefficients. Therefore,

$$\pi(B) = \theta^{-1}(B)\phi(B) = (1 - \theta B + (-\theta)^2 B^2 + \cdots)(1 - \phi B) \equiv \sum_{j=0}^{\infty} \pi_j B^j.$$

$\Rightarrow Z_t = \pi(B) X_t$

$\frac{1}{1+\theta z} = \frac{1}{1-(-\theta z)} = \sum (-\theta z)^j$

By rewriting the terms, it can be shown that

$$\pi_j = \begin{cases} 1 & j = 0 \\ -(\phi+\theta)(-\theta)^{j-1} & j \geq 1 \end{cases},$$

and consequently, we can express the ARMA(1,1) time series as an AR(∞) as follows:

$$Z_t = \pi(B)X_t = X_t - (\phi+\theta)\sum_{j=1}^{\infty}(-\theta)^{j-1}X_{t-j}. \tag{3.9}$$

noise for time t

past observation

Thus the ARMA(1,1) process is invertible, since Z_t can be expressed in terms of the present and past values of the process $X_s,\ s \leq t$. ■

Similarly, it can be shown that the ARMA(1,1) process is **noninvertible** when $|\theta| > 1$, since then

$$Z_t = -\phi\theta^{-1} X_t + (\phi + \theta) \sum_{j=1}^{\infty} (-\theta)^{-j-1} X_{t+j}. \tag{3.10}$$

present noise

future time

$$Z_t = \pi(B) X_t$$

where $\pi(B) = \dfrac{1 - \phi B}{1 + \theta B}$

$$\frac{1}{1+\theta B} = \frac{1}{1-(-\theta B)} = \underbrace{\frac{1}{1-(-\theta B)}}\, - 1 + 1$$

$$= \frac{-\theta B}{1-(-\theta B)} + 1$$

$$= \frac{-\cancel{\theta B}}{\cancel{\theta B}} \frac{1}{\frac{1}{-\theta B} + 1} + 1$$

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

$$= -\sum_{j=1}^{\infty} \left(-\frac{1}{\theta B}\right)^{j}$$

Note 3.5 *We summarize these results as follows:*

This course

1. *If* $|\theta| < 1$, *then the ARMA(1,1) process is invertible, and* Z_t *is expressed in terms of* $X_s,\ s \leq t$.

2. *If* $|\theta| > 1$, *then the ARMA(1,1) process is noninvertible, and* Z_t *is expressed in terms of* $X_s,\ s \geq t$.

3. *If* $|\theta| = 1$, *the ARMA(1,1) process is somehow invertible (***NOT COVERED IN THIS COURSE***). Here, we use the term invertible if we can write* $Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$, *where* $\sum_{j=0}^{\infty} |\pi_j| < \infty$.

In general, a **linear** time series $\{X_t\}$ for which the following relation holds is said to be **invertible**:

$$Z_t = \sum_{j=0}^{\infty} \underbrace{\pi_j}_{\text{weight}} \underbrace{X_{t-j}}_{\text{past and present observations}}, \quad \text{where} \quad \sum_{j=0}^{\infty} |\pi_j| < \infty.$$

The ACVF for the stationary, causal ($|\phi| < 1$) and invertible ($|\theta| < 1$) ARMA(1,1) time series is

$$\gamma_X(h) = \begin{cases} (1 + \frac{(\theta+\phi)^2}{1-\phi^2})\sigma^2 & h = 0 \\ (\theta + \phi + \frac{(\theta+\phi)^2\phi}{1-\phi^2})\sigma^2 & |h| = 1 \\ \phi^{h-1}\gamma_X(1) & |h| > 1 \end{cases}$$

Application

Consider the segment of the daily Dow Jones Utilities Index shown in Figure 3.5 along with the lag 1 differenced series and their ACF's.

The ARMA(1,1) model for the differenced data when fitted using maximum likelihood in R is $\phi = 0.851$, $\theta = -0.526$ and $\sigma^2 = 0.143$. If Y_t represents the Dow Jones Utilities Index at time t then

$$\nabla Y_t = 0.851 \nabla Y_{t-1} + Z_t - 0.526 Z_{t-1}$$

where $\{Z_t\} \sim$i.i.d.N$(0, 0.143)$. Rewriting this equation we get

$$Y_t = 1.851 Y_{t-1} - 0.851 Y_{t-2} + Z_t - 0.526 Z_{t-1}.$$

Note that a mean term is not included in the model because it is not statistically significant, indicating that there is no significant upward or downward drift over the observation period. For comparison, the models AR(1) and MA(1) are also fitted to these data.

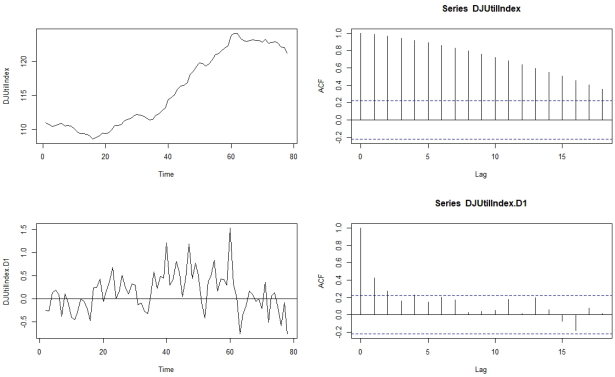


Figure 3.5: Dow Jones Utilities Index and its lag = 1 difference

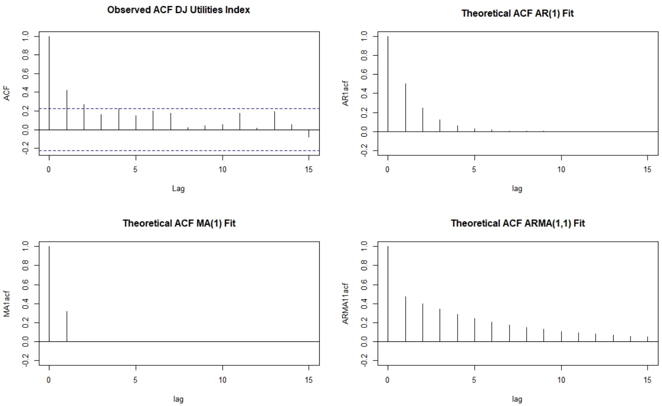


Figure 3.6: Sample and theoretical ACFs for lag 1 differenced DowJones Index

The three models were also fitted using the maximum likelihood method using `R`. The results are given in Table 3.1.

Model	$\hat{\phi}$	$\hat{\theta}$	$\hat{\sigma}^2$	AIC
ARMA(1,1)	0.8506 (0.1386)	-0.5257 (0.2550)	0.1434	75.38
AR(1)	0.4992 (0.1001)	-	0.1493	76.38
MA(1)	-	0.3600 (0.0858)	0.1639	83.42

Table 3.1: Estimated coefficients and their standard error, along with estimated σ^2 and AIC for ARMA(1,1), AR(1) and MA(1) models for Dow Jones Utilities Index.

Note 3.6

- *The ARMA model could be conceptualised as an AR(1) process observed with random noise. The observed ACF does show an initial drop and then an exponential decay from that point. This is what you can get from an ARMA(1,1) ACF.*

- *However, in this instance it is very hard to pick between the AR(1) and the*

ARMA(1,1) although the latter is marginally better or has a small advantage over the AR(1) model in terms of fit and estimated forecast error variance ($\hat{\sigma}^2$).

3.2 Sample Mean for Stationary Models

We know that for stationary time series,

$$var(\bar{X}_n) = \frac{1}{n} \sum_{h=-n+1}^{n-1} (1 - \frac{|h|}{n}) \gamma(h). \tag{3.11}$$

If $\sum |\gamma(h)| < \infty$, then

$$\lim_{n \to \infty} n \; var(\bar{X}_n) = \sum_{-\infty}^{\infty} \gamma(h).$$

In this section we derive expressions for this limit for the MA(1) and the AR(1).

3.2.1 Sample mean for MA(1)

For MA(1) process, we know that

$$\gamma_X(h) = \begin{cases} (1+\theta^2)\sigma^2 & h = 0 \\ \theta\sigma^2 & |h| = 1 \\ 0 & |h| > 1 \end{cases}$$

Here

$$\sum_{-\infty}^{\infty} \gamma(h) = \sigma^2(\theta + 1 + \theta^2 + \theta) = \sigma^2(1+\theta)^2,$$

so that

$$var(\bar{X}_n) \approx \frac{\gamma(0)}{n} \frac{(1+\theta)^2}{(1+\theta^2)} \equiv V(\theta).$$

- When $\theta < 0$ (negative lag one autocorrelation) the variance of the sample mean is smaller than for independent data ($\theta = 0$).

- When $\theta > 0$ (positive lag one autocorrelation) the variance is larger than the independent case.
- The intuition is that values that are negatively correlated tend to oscillate around the mean level and therefore is more precisely determined.

3.2.2 Sample mean for AR(1)

In AR(1) process, the autocovariance function is

$$\gamma_X(h) = \frac{\sigma^2 \phi^{|h|}}{1-\phi^2}, \qquad h = 0, \pm 1, \cdots,$$

and, consequently, using the geometric series, we have

$$\sum_{h=-\infty}^{\infty} \gamma(h) = \sigma^2 \sum_{h=-\infty}^{\infty} \frac{\phi^{|h|}}{1-\phi^2} = \frac{\sigma^2}{(1-\phi)^2} \equiv V(\phi).$$

- As $\phi \to 1$, the variance of the sample mean grows without bound.
- The intuition is that, as $\phi \to 1$, the process is tending to be like a random walk and each additional data point has very little additional information about the mean level.

3.2.3 Distribution of the sample mean

If $\{X_t\}$ is a Gaussian time series the sample mean is exactly normally distributed

$$\bar{X}_n \sim N\left(\mu, \frac{1}{n}\sum_{h=-n+1}^{n-1}\left(1-\frac{|h|}{n}\right)\gamma(h)\right)$$

and if we approximate the variance by $\sum_{-\infty}^{\infty}\gamma(h)$ we get

$$\bar{X}_n \approx N(\mu, \frac{1}{n}\sum_{-\infty}^{\infty}\gamma(h)).$$

- Even when the time series is **not Gaussian**, it is often the case that the central limit theorem can be proved and **this last approximate distribution can be used**.
- Approximate confidence intervals for the mean of a stationary time series can be calculated using the above results.

Example 3.6 (Approximate CI for μ in AR(1)) *In the AR(1), an approximate 95% confidence interval for the mean, μ, is*

$$\bar{x} \pm \frac{1.96}{\sqrt{n}} \frac{\sigma}{1-\phi}.$$

To calculate this we need to know the true values of σ and ϕ or we need good estimates (as derived in the next section).

Note that for the above interval:

1. *The width of this interval is*

$$w_T = \frac{2 \times 1.96}{\sqrt{n}} \frac{\sigma}{1-\phi}.$$

2. *If a confidence interval for the mean is calculated under the* **false assumption**

that the data are independent, then

$$\sum_{h=-\infty}^{\infty} \gamma(h) = \gamma(0) = \sigma/(1-\phi^2).$$

The width of the resulting interval is approximately (for large n)

$$w_F = \frac{2 \times 1.96}{\sqrt{n}} \frac{\sigma}{\sqrt{1-\phi^2}}$$

3. *The relative width of the* **false interval** *to the* **true interval** *is*

$$\frac{w_F}{w_T} = \frac{1-\phi}{\sqrt{1-\phi^2}} = \frac{\sqrt{1-\phi}}{\sqrt{1+\phi}}$$

4. *If correlation is strongly positive, say $\phi = 0.9$, then*

$$\frac{w_F}{w_T} = 0.229 \approx 0.23 \quad \equiv \quad w_F = 0.23 w_T.$$

Therefore, false interval is shorter than the true interval, leading to far greater precision being claimed for the sample mean estimate under the false independence assumption.

In conclusion,

1. *If* $\phi > 0$ *(positive correlation) the true interval is wider than the independent case* ($\phi = 0$)*. Positive correlation reduces the amount of information about the mean in the sample of size* n*.*

2. *If* $\phi < 0$ *(negative correlation) the true interval is narrower than in the independent case* ($\phi = 0$)*. Negative correlation provides more information about the mean than in the independent case since when the correlation is negative the successive values of the time series tend to oscillate around the mean.*

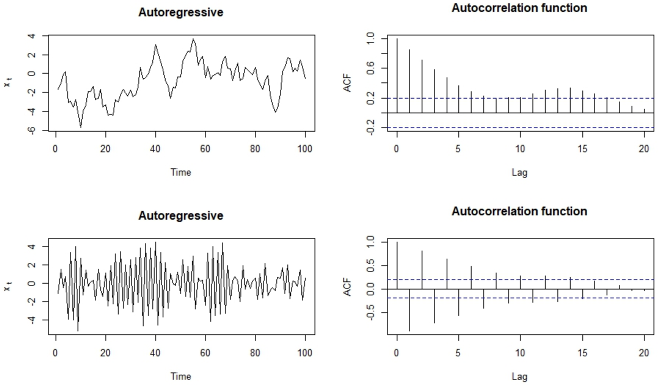


Figure 3.7: Sample path and ACF of simulated observations from AR(1) time series with $\phi > 0$ (up) and $\phi < 0$ (down)

3.2.4 Two Examples Illustrating Impact of Serial Dependence on Inference

Overshorts Data to Detect Leaky Petrol Tanks

The analysis summarized here is based on Brockwell and Davis [2002]. The data were collected as part of a large scale routine monitoring of gasoline stations in Colorado.

- X_t : the amount measured using a dipstick in the inground petrol storage tank at the end of day t
- A_t : the amount of petrol sold during the day minus any amount that was delivered during the day.

We expect that

$$X_t \approx X_{t-1} + \texttt{delivered}_t - \texttt{sold}_t = X_{t-1} - A_t.$$

Let the 'overshorts' be denoted by

$$Y_t = X_t - X_{t-1} + A_t$$

- Under the assumption of no measurement error and no leakage in the tank we should have $Y_t \equiv 0$.
- The series of overshorts Y_t is plotted for 57 consecutive days in Figure 3.8.
- In practice there is measurement variability. In this case the 'overshorts' Y_t will be samples from a distribution with mean μ.
- If the tank is not leaking into the subsoil then we should have $H_0 : \mu = 0$.
- To test this null hypothesis against the alternative that the tank is leaking $H_a : \mu < 0$, we could use the $t-$test

$$t = \frac{\bar{Y} - 0}{\mathrm{se}(\bar{Y})}$$

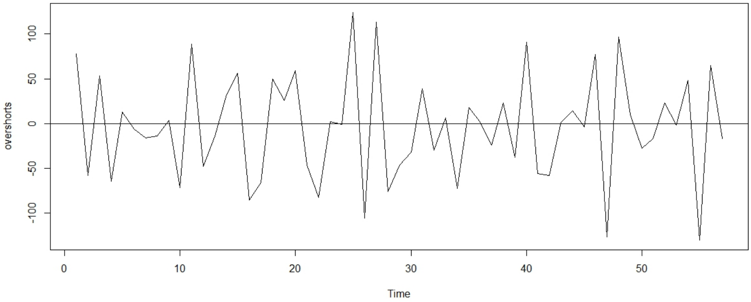


Figure 3.8: Overshorts series to detect leaky petrol tanks

- This is the same as fitting a (very) simple regression model

$$Y_t = \mu + U_t$$

in which the errors U_t are assumed to be independent $N(0, \sigma_U^2)$.

- The estimated mean is $\hat{\mu}_{LS} = \bar{Y} = -4.035088$ with reported standard error of 7.809927 and $t = -0.51666$ with associated P-value of 0.3037 which is greater than 0.05 and consequently not significant.

- The problem with this analysis is that it is based on a false assumption. Of the three assumptions about the U_t (independent, normally distributed and constant standard deviation σ) the independence assumption is the most critical. Independence is not true here!

Lets do the analysis using time series.

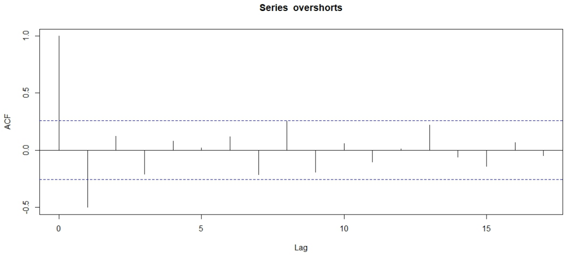


Figure 3.9: ACf of Overshorts time series

- A model for the observed autocorrelation is the simple moving average of degree 1:

$$Y_t = \mu + U_t \quad \text{where} \quad U_t = Z_t - \theta Z_{t-1}$$

in which the parameter θ controls the amount of correlation at one lag separation in time and the Z_t are independent $N(0, \sigma_Z^2)$.

- Fitting the mean μ and the MA(1) parameter θ jointly using MLE in `R`

$$\hat{\mu}_{MLE} = -4.779577 \; (s.e. = 1.0266), \; \hat{\theta} = -0.8473 \; (s.e. = 0.1205)$$

The standard error for $\hat{\mu}_{MLE}$ is 1.0266 and for testing H_0 we have

$$t = -4.779577/1.0266 \approx -4.655646$$

with associated p-value of < 0.001.

- We now have compelling evidence that the tank is leaking.

- Hence the naive analysis that ignores the possibility of serial dependence would lead to the conclusion that the tank does not leak.

Lake Huron Levels.

This example, also from Brockwell and Davis [2002], concerns the significance of time trends in the level of Lake Huron, one of the Great Lakes of North America. The time series consists of annual levels (in feet) reduced by 570 for the period 1875 to 1972 and is plotted in Figure 3.10.

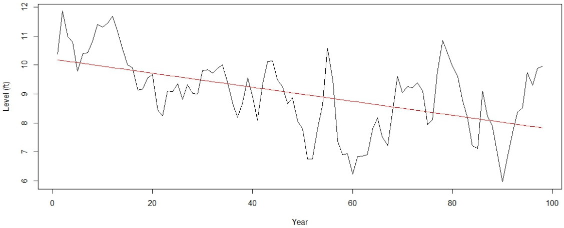


Figure 3.10: Level of Lake Huron measured in feet after subtraction of 570 showing the line fitted by least squares.

The series, denoted Y_t, is modelled using the simple linear trend, with time as the predictive variable:

$$Y_t = \alpha + \beta t + U_t,\ t = 1, \ldots, 98.$$

Ordinary least squares assumes that the errors U_t are **independent** $N(0, \sigma_U^2)$. Using ordinary least squares, we get

$$\hat{\alpha} = 10.202037\ (s.e. = 0.230111,\ p - value =< 2e - 16),$$

and

$$\hat{\beta} = -0.024201\ (s.e. = 0.004036,\ p - value =< 3.55e - 08),$$

which indicates a highly significant downward trend.

However, the least squares assumptions are not correct here.

Lets take a look at the residuals of the fitted model, presented in Figure 3.11.

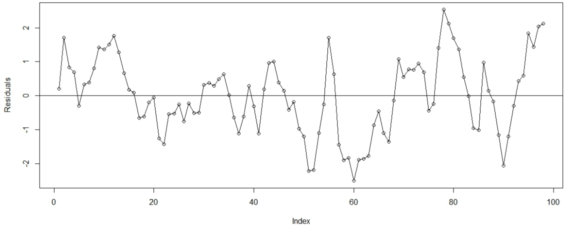


Figure 3.11: Residuals from fitting a line to Level of Lake Huron.

There are two interesting features of the graph of the residuals.

- The absence of any discernible trend.

- The smoothness of the graph.
  - There are long stretches of residuals that have the same sign.
  - This would be very unlikely to occur if the residuals were observations of i.i.d. noise with zero mean.
- Smoothness of the graph of a time series is generally indicative of the existence of some form of dependence among the observations.

Therefore, independence of U_t is wrong.

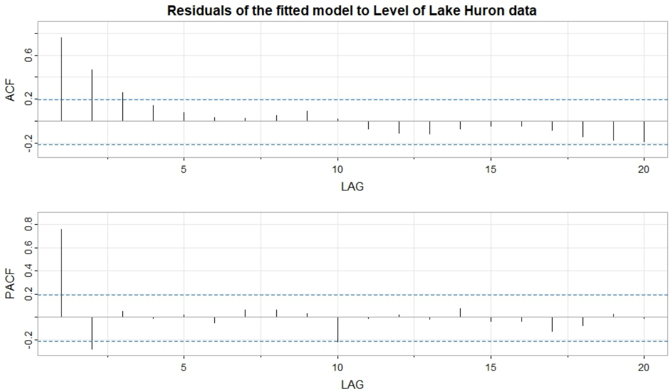


Figure 3.12: ACF and PACF of residuals of the fitted model to Level of Lake Huron data.

There is significant autocorrelation, Figure 3.12, between successive measurements which is well modelled by an autoregressive model of the form

$$U_t = \phi_1 U_{t-1} + \phi_2 U_{t-2} + Z_t$$

in which the Z_t are independent $N(0, \sigma_Z^2)$.

```
> LakeLevels.AR2<-arima(huron,order=c(2,0,0),xreg=Time)
> LakeLevels.AR2

Call:
arima(x = huron, order = c(2, 0, 0), xreg = Time)

Coefficients:
         ar1      ar2  intercept     Time
      1.0048  -0.2913    10.0915  -0.0216
s.e.  0.0976   0.1004     0.4636   0.0081
```

```
sigma^2 estimated as 0.4566:  log likelihood = -101.2,  aic = 212.4
```

- When autocorrelation is accounted for using an AR(2) model, $\hat{\beta} = -0.0216$ with $s.e. = 0.0081$, which reduced the test statistic to $t = -2.66$ (P-value of 0.008).
- This is strong evidence against the null hypothesis of no trend but is not as strong as is implied by least squares.
- Although adjusting for positive serial dependence was required, it did not change the substantive conclusion about the trend in lake levels in this instance.