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 the previous chapter, we introduce white noise and i.i.d. noise processes. In this chapter, we are going to present some simple models in time series including the autoregressive of order 1, the moving average of order 1 and the mixed autoregressive-moving average of order (1,1). The basic properties of these series are developed including their sample path behaviour and their autocorrelation functions. Applications to inference about the mean level and trend of a time series are presented and the impact of autocorrelation on standard errors highlighted.

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 mean zero and variance σ^2. We write this as $\{X_t\} \sim WN(0, \sigma^2)$. The mean and autocovariance functions are $\mu_X(t) \equiv 0$ and

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

respectively. Note that 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 mean zero and variance σ^2. We write this as $\{X_t\} \sim$i.i.d.$(0, \sigma^2)$. This process is a strictly stationary time series and the mean and autocovariance functions are the same as the mean and autocovariance functions for the white noise processes.

3.1.1 Martingale and Martingale Difference

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

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

- $E|Y_t| < \infty, \quad t \in \mathbb{Z}$,
- $E(Y_t|\sigma(Y_{t-1}, Y_{t-2}, \ldots, Y_1)) = Y_{t-1}, \quad t \in \mathbb{Z}$.

That is, the conditional expectation of Y_t given its past $\sigma(Y_{t-1}, Y_{t-2}, \ldots, Y_1)$ is equal to the immediate past[1].

Note 3.2 *Martingales do not necessarily have finite variance.*

Suppose that a time series Y_t is a martingale. Consider the difference sequence X_t defined as follows:

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

[1] In definition of Martingale, $F_{t-1} = \sigma(Y_{t-1}, \cdots, Y_1)$ is the sigma field generated from the past of the process. Generally, a sigma field (σ-field) generated by a set of random variables is the smallest collection of events that includes all the possible outcomes of those random variables and is closed under the operations of countable union, countable intersection, and complementation. For this course, we are not going into detailed definition of sigma fields and just consider it as the whole information we can get from this set of random variables.

Given the martingale property, it is possible to demonstrate that X_t is a zero-mean, uncorrelated process with the additional property that

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

These properties can be used to define a martingale difference sequence. A **martingale difference sequence** is a zero mean, uncorrelated process such that its conditional mean at every step is zero. From this definition, we see 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. In fact, we can establish the following implication chain:

- 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.

Both of the above are quite important. The first implies 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. The second implies that martingale difference sequences, though uncorrelated, are not independent and can therefore exhibit correlation in higher moments. For **Gaussian processes with mean zero** the three concepts coincide: white noise, i.i.d. noise, and martingale difference sequences imply each other.

The above considerations are important if we are to distinguish between noise and innovation processes (More information is provided in Section 8.9.3 of Spanos [2019] but it is out of the scope of this course).

Note 3.3 *Uncorrelatedness of martingale differences is due to the fact that, for $m < n$, $E(X_n X_m) = 0$, since*

$$\begin{aligned} E(X_n X_m) &= E(E(X_m X_n | X_m, \cdots, X_1)) \\ &= E(X_m E(X_n | X_m, \cdots, X_1)) \\ &= E(X_m \times 0) = 0. \end{aligned}$$

Example 3.1 *(Random walk with i.i.d. noise is martingale) 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*

$$\begin{aligned} E(X_t|\sigma(X_{t-1}, \ldots, X_1)) &= E(X_{t-1} + Z_t|\sigma(X_{t-1}, \ldots, X_1)) \\ &= E(X_{t-1}|\sigma(X_{t-1}, \ldots, X_1)) + E(Z_t|\sigma(X_{t-1}, \ldots, X_1)) \\ &= X_{t-1} + E(Z_t|\sigma(Z_{t-1}, \ldots, Z_1)) \\ &= X_{t-1} + E(Z_t) \\ &= X_{t-1} \end{aligned}$$

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 (Why?).

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

Let $\{Z_t\} \sim$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

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

Besides, the autocorrelation function of X_t is

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

Note that for the autocovariance and mean functions to be as above we only require Z_t to be uncorrelated in which case X_t would be a weakly stationary time series. When $\{Z_t\} \sim$i.i.d.$(0, \sigma^2)$ the process X_t is strictly stationary.

Note 3.4 *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)
```

where

- `x`*: a univariate or multivariate time series.*
- `filter`*: a vector of filter coefficients in reverse time order (as for AR or MA coefficients).*
- `method`*: Either "convolution" or "recursive" (and can be abbreviated). If "convolution" a moving average is used: if "recursive" an autoregression is used.*
- `sides`*: for convolution filters only. If sides = 1 the filter coefficients are for past values only; if sides = 2 they are centred around lag 0. In this case the length of the filter should be odd, but if it is even, more of the filter is forward in time than backward.*
- `circular`*: for convolution filters only. If TRUE, wrap the filter around the ends of the series, otherwise assume external values are missing (NA).*
- `init`*: for recursive filters only. Specifies the initial values of the time series just prior to the start value, in reverse time order. The default is a set of zeros.*

Example 3.2 (Simulated time series from MA(1)) *Figure 3.1 demonstrates a realization of $\{X_1, \ldots, X_{100}\}$ with $\theta = -0.8$ and $Z_t \sim N(0, 1)$ along with its autocorrelation function. (What does the autocorrelation function look like if $\theta = 0.8$?)*

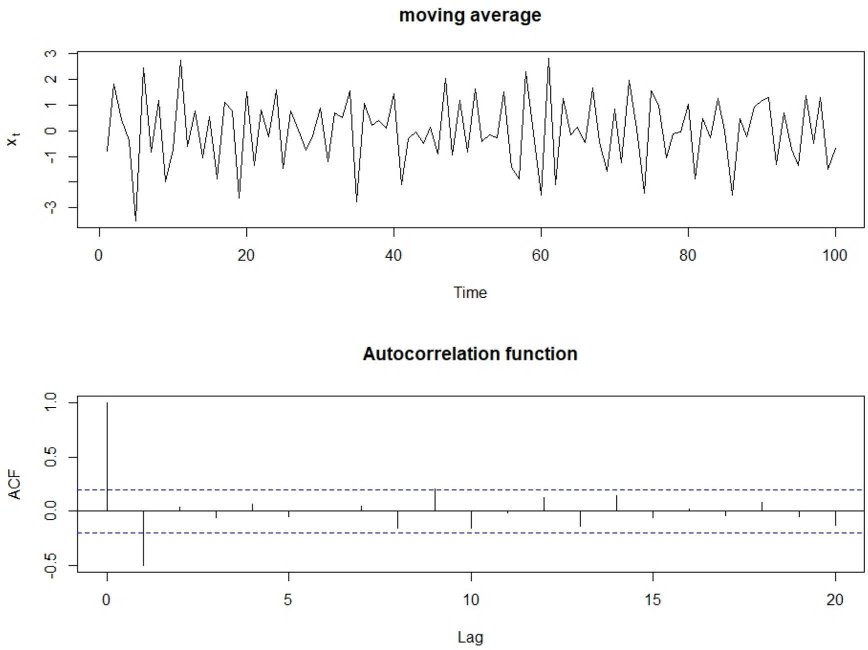

Figure 3.1: 100 simulated observations from $X_t = Z_t - 0.8Z_{t-1}$ along with the ACF.

Example 3.3 *To illustrate the MA(1) process, consider the monthly employment in Trades introduced in Chapter 1. The combined seasonal and lag 1 differences appear to be stationary in the mean and variance (Figure 3.2). Let $\{Y_t\}$ be the original series and 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

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

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

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

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

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

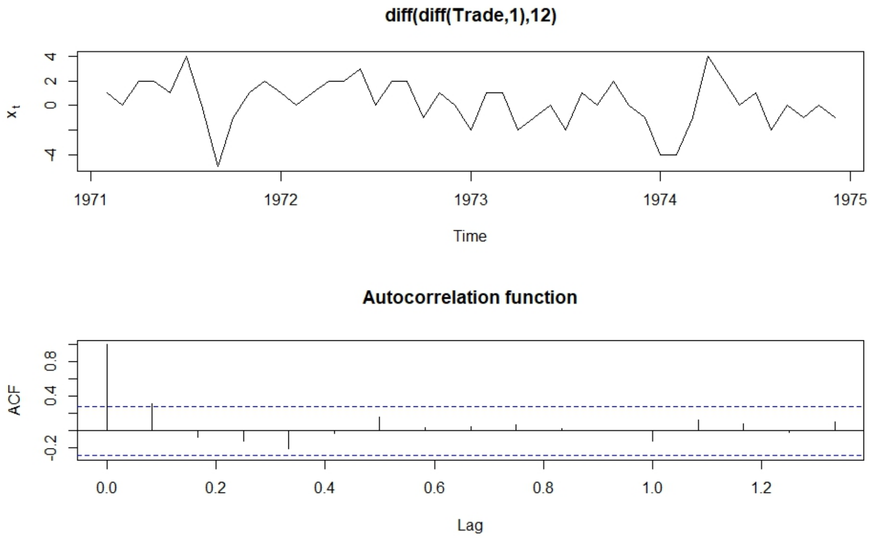

Figure 3.2: The monthly employment in Trades data after combined seasonal and lag 1 differences along with the ACF.

both of which do not depend upon t.

A special case of a linear process is the MA(∞) in which $\psi_j = 0$, $j < 0$. Using the backshift operator notation we can write the MA(∞) process as

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

where

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

Note 3.5 *The condition* $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ *ensures that the infinite sum in Equation* (3.1) *converges (with probability one), since* $E|Z_t| \le \sigma$ *and*

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

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* $\sum_{j=-n}^{n} \psi_j Z_{t-j}$.

MORE IN CHAPTER 3.

Example 3.4 *Consider a white noise series* W_t, $t = 1, \dots, 300$, *and, for each* t, *let* X_t *be an average of* W_t *and its immediate neighbors in the past and future. That is, let*

$$X_t = \frac{1}{3}(W_{t-1} + W_t + W_{t+1}) \tag{3.2}$$

Figure 3.3 demonstrates a simulated white noise series along with X_t *and its autocorrelation 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.*

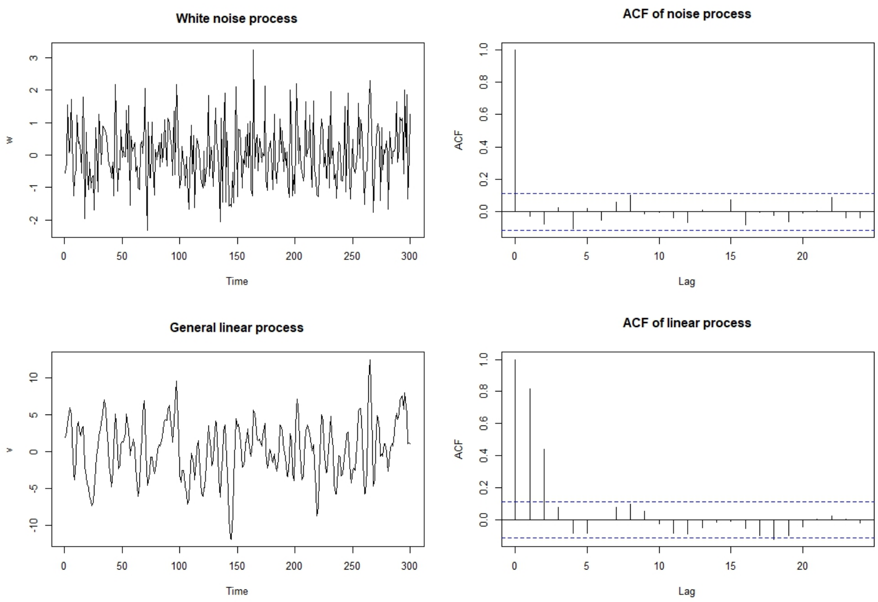

Figure 3.3: Simulated white noise and the resulting general linear processes along with their autocorrelation functions.

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

Let $\{Z_t\}$ ∼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 Z_t is uncorrelated with X_s for each $s < t$. The stationary solution to this system of difference equations is

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

from which it follows that $\mu_X(t) \equiv 0$ and the autocovariance and autocorrelation functions are

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

which are geometrically decaying functions of separation lag h.

Theorem 3.1 *Let $\{Z_t\}$ ∼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, Brockwell and Davis [2002], Page 46. ■

It is not possible to define a stationary solution to $X_t = \phi X_{t-1} + Z_t$ when $|\phi| = 1$. However it is possible to define a stationary solution when $|\phi| > 1$. To develop this we need some ideas from power series in the backshift operator. The unique stationary solution to the AR(1) when $|\phi| > 1$ is

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

and since this depends on future values of the Z_t processm, we refer to this as the *non-causal stationary* solution. Details in next Chapter.

- **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 $s > t$.

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

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. (What does the autocorrelation function look like if* $\phi = -0.9$*?)*

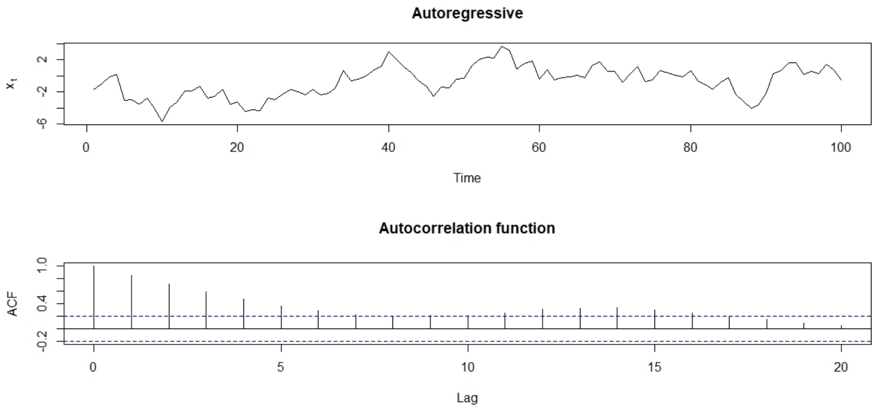

Figure 3.4: 100 simulated observations from $X_t = 0.9X_{t-1} + Z_t$ along with the ACF.

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

The 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

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

where $Z_t \sim\ WN(0, \sigma^2)$ and $\phi + \theta \neq 0$. Equation (3.6) can also be written based on the operator B, as

$$(1 - \phi B)X_t = (1 + \theta B)Z_t. \tag{3.7}$$

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, we have

$$\frac{1}{\phi(z)} = \frac{1}{1 - \phi z} = \sum_{j=0}^{\infty} \phi^j z^j,$$

  which has absolutely summable coefficients. Therefore,

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

  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

$$\begin{aligned} X_t &= \psi(B)Z_t \\ &= Z_t + (\phi + \theta)\sum_{j=1}^{\infty} \phi^{j-1} Z_{t-j} \end{aligned} \tag{3.8}$$

  is the unique stationary solution of Equation (3.6).

- 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,

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

  By rewriting the terms, it can be shown that

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

  and consequently,

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

- 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 3.6 *In summary,*

1. *If* $|\phi| = 1$, *there is no stationary solution for Equation* (3.6).
2. *A stationary solution of Equation* (3.6) *exists if and only if* $|\phi| \neq 1$.
3. *If* $|\phi| < 1$, *then the unique stationary solution is given by* (3.8). *In this case we say that* $\{X_t\}$ *is causal or a causal function of* $\{Z_t\}$, *since* X_t *can be expressed in terms of the current and past values* $Z_s,\ s \leq t$.
4. *If* $|\phi| > 1$, *then the unique stationary solution is given by* (3.9). *The solution is noncausal, since* X_t *is then a function of* $Z_s,\ s \geq t$.

The range of possible values of θ

As mentioned before, causality means that X_t is expressible in terms of $Z_s,\ s \leq t$. The dual concept of **invertibility** means that Z_t can be expressed in terms of X_s, $s \leq t$. The ARMA(1,1) process defined by Equation (3.6) is invertible if $|\theta| < 1$. To demonstrate this, 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,

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

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:

$$\begin{aligned} Z_t &= \pi(B)X_t \\ &= X_t - (\phi+\theta)\sum_{j=1}^{\infty}(-\theta)^{j-1}X_{t-j}. \end{aligned} \tag{3.10}$$

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 $X_s,\ s \leq t$. An argument like the one used to show noncausality when $|\phi| > 1$ shows that the ARMA(1,1) process is **noninvertible** when $|\theta| > 1$, since then

$$Z_t = -\phi\theta^{-1}X_t + (\phi+\theta)\sum_{j=1}^{\infty}(-\theta)^{-j-1}X_{t+j}. \tag{3.11}$$

Note 3.7 *We summarize these results as follows:*

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 invertible in the more general sense that Z_t is a mean square limit of finite linear combinations of $X_s,\ s \leq t$, although it cannot be expressed explicitly as an infinite linear combination of $X_s,\ s \leq t$. However, in this course, we use the term invertible if we can write $Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$, where $\sum_{j=0}^{\infty} |\pi_j| < \infty$. If you are interested, you can refer to Brockwell and Davis [1991], Section 4.4, for more information.*

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

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad \text{where } \sum_{j=0}^{\infty} |\pi_j| < \infty.$$

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

To illustrate the ARMA(1,1) 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.

Note 3.8 *The Dow Jones Utilities Index data is available in the file* `DowJonesUtil.txt`*. Besides, it is part of* `itsmr` *package in* `R` *and can be loaded directly.*

For comparison, the models AR(1) and MA(1) are also fitted to these data. The estimated autocorrelation function and the theoretical autocorrelation functions for the three fitted models are given in Figure 3.6. 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.

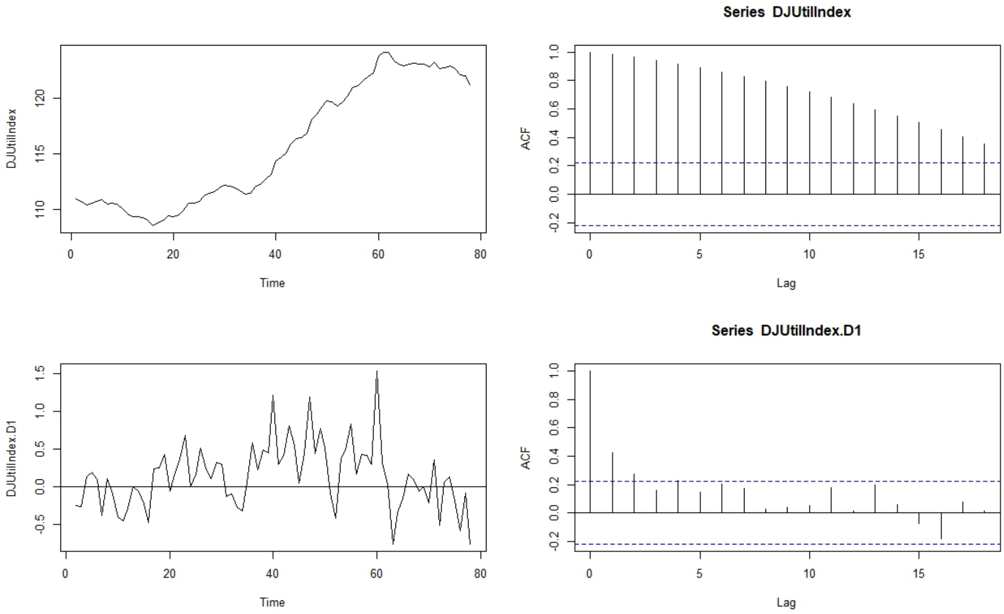

Figure 3.5: Dow Jones Utilities Index and its lag = 1 difference

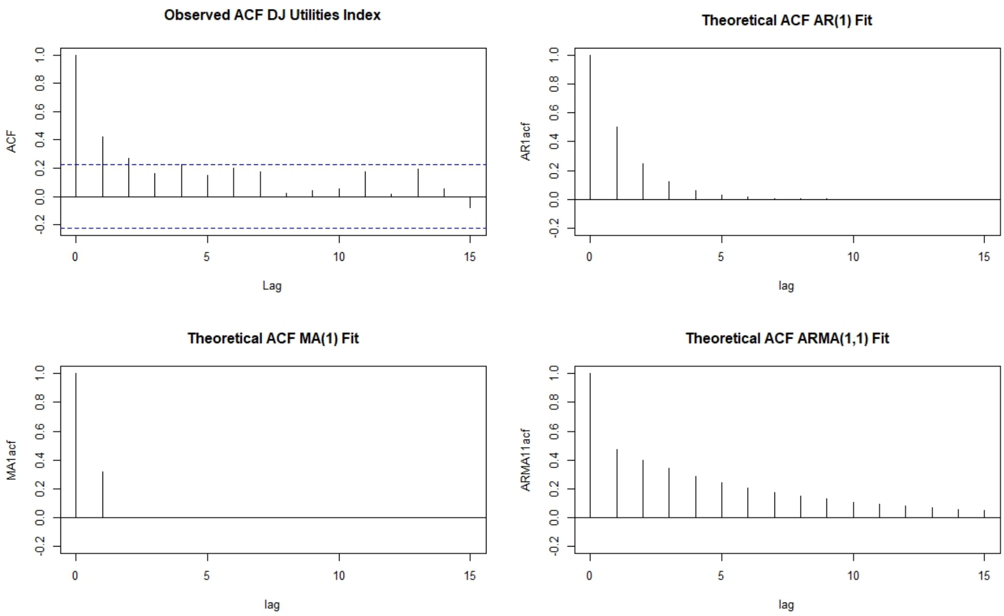

Figure 3.6: Sample and theoretical ACFs for lag 1 differenced Dow Jones Utilities Index

Note 3.9

- *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 has a slight edge in fit and estimated forecast error variance ($\hat{\sigma}^2$).*

3.2 Further Notes on the Sample Mean for Stationary Models

Previously, we derived the general expression (2.10) for the variance $var(\bar{X}_n)$ of the sample mean of a stationary time series of length n and gave the limit, as $n \to \infty$, as $\sum_{-\infty}^{\infty} \gamma(h)$. 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

$$var(\bar{X}_n) \approx \frac{\gamma(0)}{n} \frac{(1+\theta)^2}{(1+\theta^2)} \equiv V(\theta), \tag{3.12}$$

and we note that when $\theta < 0$ (negative lag one autocorrelation) the variance of the sample mean is smaller than for independent data ($\theta = 0$) and 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

$$\begin{aligned} \sum_{h=-\infty}^{\infty} \gamma(h) &= \sigma^2 \sum_{h=-\infty}^{\infty} \frac{\phi^{|h|}}{1-\phi^2} \\ &= \frac{\sigma^2}{(1-\phi)^2} \equiv V(\phi). \end{aligned} \tag{3.13}$$

Note that 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

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

where the variance is given in (2.10) 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 confidence intervals for μ in AR(1)) *In the AR(1), we have an approximate 95% confidence interval for the mean, μ, given by*

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

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*

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

If a confidence interval for the mean is calculated under the false assumption that the data are independent the usual estimate of standard deviation of the observed process X_t would be used resulting in an estimate of $(\gamma_X(0))^{1/2} = \sigma/(1-\phi^2)^{1/2}$. The width of the resulting interval is approximately (for large n)

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

and the relative width of the false interval to the true interval is

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

which, if correlation is strongly positive, say $\phi = 0.9$, the false interval is 0.23 times the true interval leading to far greater precision being claimed for the sample mean estimate under the false independence assumption.

2. *If $\phi > 0$ (positive correlation) the 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.*

3. *If $\phi < 0$ (negative correlation) the 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.*

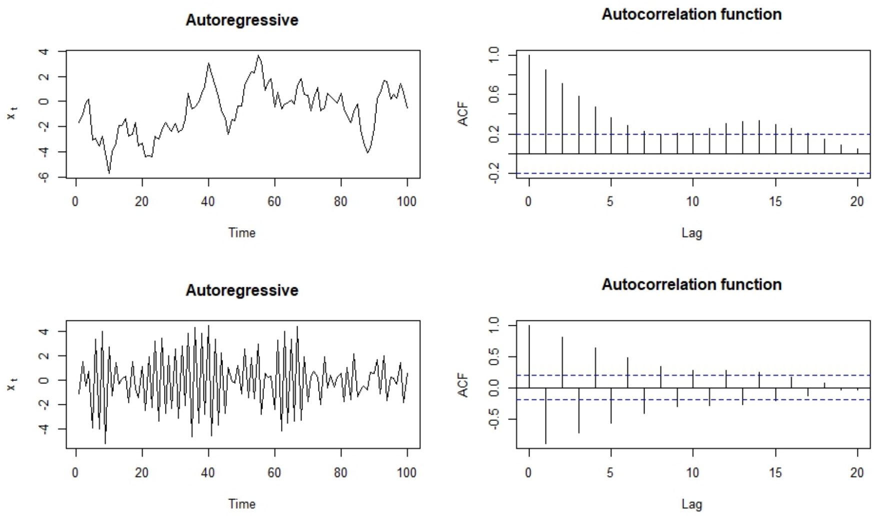

Figure 3.7: Sample path and ACF of simulated observations from AR(1) time series with $\phi > 0$ (up) and $\phi < 0$ (down)

3.3 Impact of Serial Dependence on Inference

The following two examples illustrate the 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.

Let X_t be the amount measured using a dipstick in the inground petrol storage tank at the end of day t and let A_t be 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

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

so that 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 and the data is available in the file `overshorts.txt`.

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

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

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

$$Y_t = \mu + U_t$$

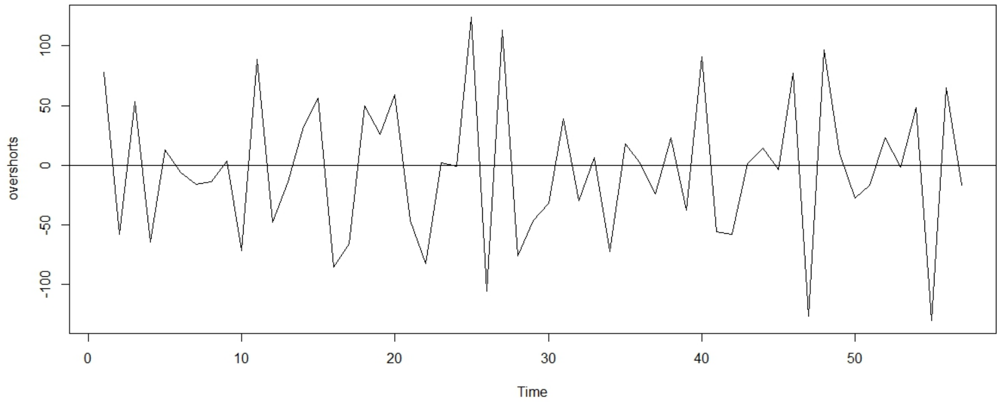

Figure 3.8: Overshorts series to detect leaky petrol tanks

in which the errors U_t are assumed to be independent $N(0, \sigma_U^2)$ (normally distributed with zero mean and standard deviation σ_U). 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. A model for the observed autocorrelation, Figure 3.9, 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 $N(0, \sigma_Z^2)$.

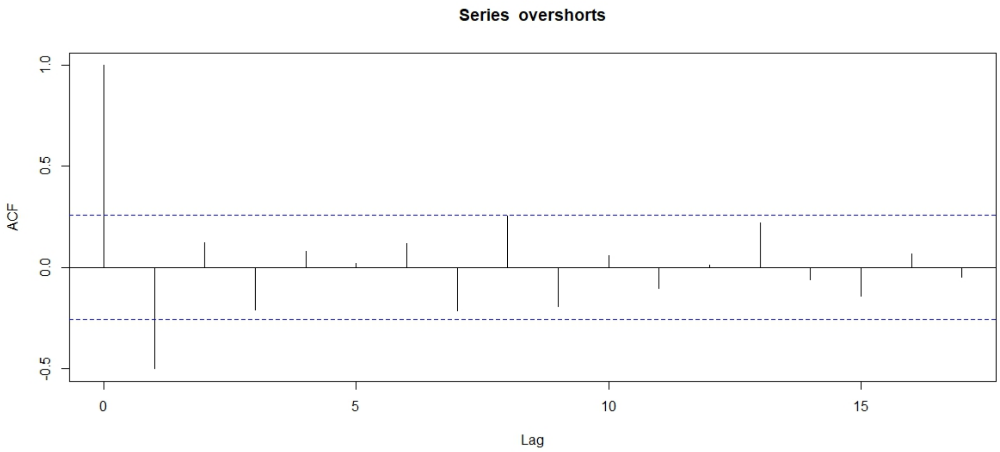

Figure 3.9: ACf of Overshorts time series

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.

Note 3.10 *The analysis of this time series is not complete. We get back to it later in this course.*

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. This data is part of `itsmr` Package and can be loaded directly in R.

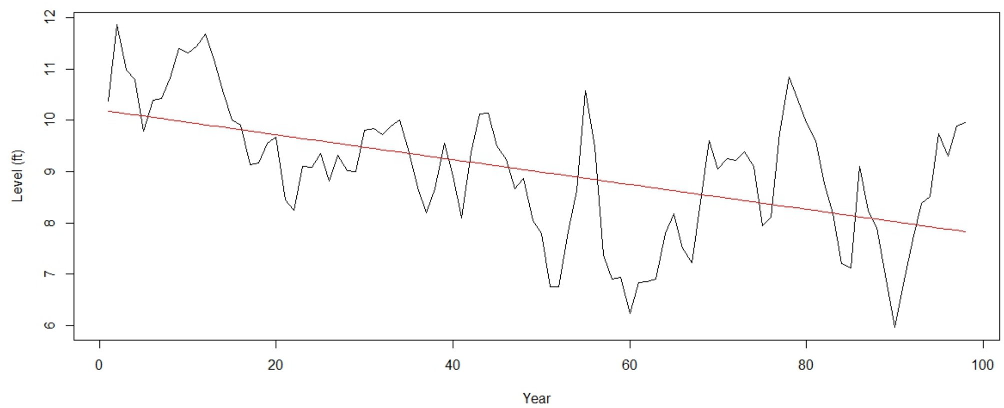

Figure 3.10: Level of Lake Huron measured in feet after subtraction of 570 showing the line fitted by least squares.

The series, denoted Y_t, is modelled using the simple linear trend

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

Ordinary least squares assumes that the errors U_t are **independent** $N(0, \sigma_U^2)$ (normally distributed with zero mean and standard deviation σ_U). 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. To make it clear, lets take a look at the residuals of the fitted model, presented in Figure 3.11. There are two interesting features of the graph of the residuals.

- The absence of any discernible trend.
- The smoothness of the graph.

In particular, 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.

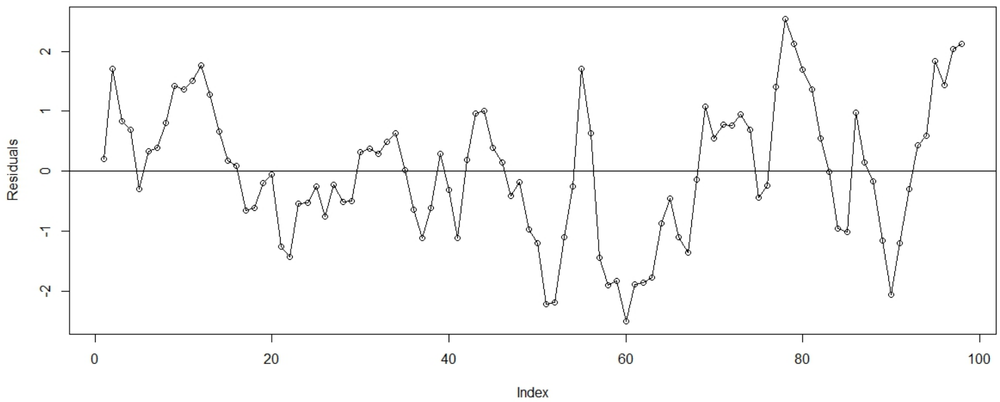

Figure 3.11: Residuals from fitting a line to Level of Lake Huron.

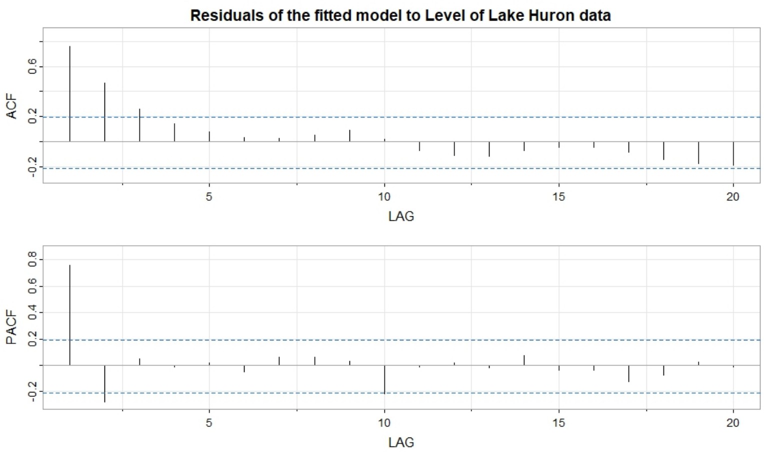

Figure 3.12: ACF and PACF of residuals of the fitted model to Level of Lake Huron data.

Therefore, independence of U_t is wrong. 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 $N(0, \sigma_Z^2)$.

The fitted model, using the `arima` function in `R`, give estimates and standard errors as follows:

```
> 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 the point estimate is $\hat{\beta} = -0.0216$ with double the standard error of 0.0081 and reduced $t = -2.66$ which has a P-value of 0.008. This is strong evidence against the null hypthesis of no trend but is not as strong as is implied by least squares. In conclusion, while adjusting for positive serial dependence was required, it did not change the substantive conclusion about the trend in lake levels in this instance. However in other examples it can and therefore correction for any serial dependence should always be done.

3.4 Exercises

Exercise 3.1 *Show that a sequence of martingale differences is an uncorrelated sequence.*

Exercise 3.2 *Consider an i.i.d. process* $\{Z_t\} \sim i.i.d.(1, \sigma^2)$ *and define* $M_t = \prod_{j=1}^{t} Z_j$, $t = 1, 2, \ldots$. *Show that* M_t *is martingale. If* $E(Z_t) = \mu \neq 1$, *redefine* M_t *such that it remains a martingale.*

Exercise 3.3 *Derive the autocovariance and autocorrelation functions for the MA(1). Show that the maximum absolute value of the lag 1 correlation is* $|\rho_X(1)| = 0.5$ *and that this occurs when* $|\theta| = 1$. *What happens to the autocorrelations when* $|\theta| > 1$*?*

Exercise 3.4 *Derive Equations* (3.4) *and* (3.5) *for the AR(1).*

Exercise 3.5 *For the MA(1) case sketch the asymptotic variance* $V(\theta)$, *defined n* (3.12), *against* θ *and comment.*

Exercise 3.6 *For the AR(1) case sketch the asymptotic variance* $V(\phi)$, *defined n* (3.13), *against* ϕ *and comment.*

3.5 Tutorial: Week 2

Exercise 3.7 (Dow Jones Utilities Index) *Using the data in the file* `DowJonesUtil.txt`

- *replicate the analysis and conclusions in Section 3.1.5.*
- *Check the assumptions that the residuals from the ARMA(1,1) model are normal and uncorrelated. Start by plotting four graphs in one frame: Original series, its ACF, lag 1 differenced series, its ACF. Comment.*
- *Fit the AR(1), the MA(1) and the ARMA(1,1). Compare your results with those reported in Section 3.1.5.*
- *Plot the observed ACF and the model ACF (theoretical using the model estimates) for the three models considered. Discuss.*

Exercise 3.8 (Overshorts analysis) *Consider the overshorts data in* `overshorts.txt`.

- *Perform a standard t-test (use the* `R` *command* `t.test(..)`*) and draw your conclusions about the hypothesis that the mean level of overshorts is zero versus the alternative that the tank is not leaking (one-sided alternative since we really do not expect the tank to be gaining, unless the pump meter is dodgy!).*
- *Look at the sample ACF for the overshorts series, discuss if the MA(1) model seems appropriate, fit this model using the* `arima`*. Command and reconsider your previous conclusion concerning the hypothesis that the tank is not leaking.*
- *Check the properties of the residuals.*
- *Plot the model ACF for the fitted model and compare with the observed ACF.*

Exercise 3.9 (Lake Huron analysis) *The data are in* `LakeHuronLevels.txt` *and an* `R`*-script which performs most of the analysis required is in* `LakeHuronAnalysis.Rmd`.

- *Use this file to understand how to assess serial dependence in simple regression settings and to include modelling serial dependence in the regression modelling.*
- *Experiment with alternative specifications of the autocorrelation in the residuals and record your conclusions.*