Discrete Time Series Analysis with ARMA Models

Veronica Sitsofe Ahiati (veronica@aims.ac.za)
African Institute for Mathematical Sciences (AIMS)

Supervised by Tina Marquardt
Munich University of Technology, Germany

June 6, 2007

Abstract

The goal of time series analysis is to develop suitable models and obtain accurate predictions for measurements over time of various kinds of phenomena, like yearly income, exchange rates or data network traffic. In this essay we review the classical techniques for time series analysis which are linear time series models; the autoregressive models (AR), the moving average models (MA) and the mixed autoregressive moving average models (ARMA). Time series analysis can be used to extract information hidden in data. By finding an appropriate model to represent the data, we can use it to predict future values based on the past observations. Furthermore, we consider estimation methods for AR and MA processes, respectively.

Taɖodzinu

Nusɔsrɔ tso ɣeyiɣiwo ƒe zɔzɔme ŋu (Time Series Analysis) ƒe taɖodzinue nye be woakpe ɖe mia ŋu be miadi mɔnu nyuitɔ anya nu si le dzɔdzɔ ge do ŋgɔ. Nu siwo woahia be mianya ƒe blibvo la me, miaƒe gagbagbawo ƒe tɔtrɔ kpakple mɔɖaŋunyawo ƒe zɔɖeɖe. Le nuŋlɔɖi sia me la, miadzro ɣeyiɣi vovovo siwo nye ɣeyiɣi tee la me. Esiawoe nye “autoregressive models” (AR), “Moving average models (MA), kple "autoregressive moving average models"(ARMA). Miate ŋu aza nusɔsrɔ sia atsɔ ahe nu siwo be ɖe miaƒe numekukuwo me ɖe go. Ne miete ŋu de dzesi mɔnu nyuitɔ la, miate ŋu aza nu siwo dzɔ va yi la, ake ɖe nu siwo ava dzɔ le etsɔ me la ŋu. Gawu la, agate ŋu ana miatrɔ asi le AR kple MA mɔnuwo ŋu wotɔxɛwotɔxɛe.

Contents

List of Figures

1. A Brief Overview of Time Series Analysis

1.1 Introduction

Time, in terms of years, months, days, or hours is a device that enables one to relate phenomena to a set of common, stable reference points. In making conscious decisions under uncertainty, we all make forecasts. Almost all managerial decisions are based on some form of forecast. In our quest to know the consequence of our actions, a mathematical tool called time series is developed to help guiding our decisions. Essentially, the concept of a time series is based on historical observation. It involves examining past values in order to try to predict those in the future. In analysing time series, successive observations are usually not independent, and therefore, the analysis must take into account the time order of the observation.

The classical method used to investigate features in a time series in the time domain is to compute the covariance and correlation function and in the frequency domain is by frequency decomposition of the time series which is achieved by a way of Fourier analysis.

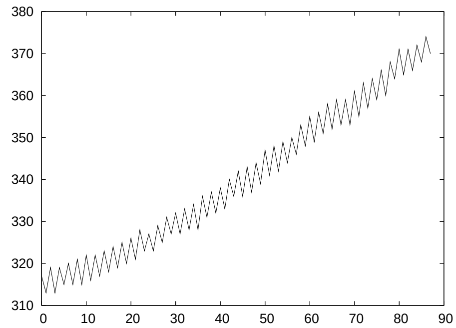
In this chapter, we will give some examples of time series and discuss some of the basic concepts of time series analysis including stationarity, the autocovariance and the autocorrelation functions. The general autoregressive moving average process (ARMA), its causality, invertibility conditions and properties will be discussed in Chapter 2. In Chapter 3 and 4, we will give a general overview of prediction and estimation of ARMA processes, respectively. In particular, we will consider a data example in Chapter 5.

1.2 Examples and Objectives of Time Series Analysis

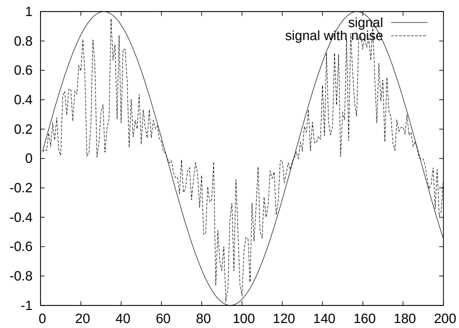
1.2.1 Examples of Time Series Analysis

Definition 1.1 *A* **time series** *is a collection of observations* $\{x_t\}_{t \in T}$ *made sequentially in time* t*. A* **discrete-time** *time series (which we will work with in this essay) is one in which the set* T *of times at which observations are made is discrete e.g.* $T = \{1, 2, \ldots, 12\}$ *and a* **continuous-time** *time series is obtained when observations are made continuously over some time interval e.g.* $T = [0, 1]$ *[BD02].*

Example 1.2 *Figure 1.1(a) is the plot of the concentration of carbon (iv) oxide (Maunal Loa Data set) in the atmosphere from 1959 to 2002. The time series changes with the fixed level and shows an overall upward trend and a seasonal pattern with a maximum concentration of carbon (iv) oxide in January and a minimum in July. The variation is due to the change in weather. During summer, plants shed leaves and in the process of decaying, release* CO_2 *into*



(a) The plot of the concentration of CO_2 in the atmosphere



(b) Example of a signal plus noise data plot

Figure 1.1: Examples of time series

the atmosphere and during winter, plants make use of CO_2 in the atmosphere to grow leaves and flowers. A time series with these characteristics is said to be non-stationary in both mean and variance.

Example 1.3 *Figure (1.1(b)) shows the graph of $\{X_t\} = N_t \sin(\omega t + \theta),\ t = 1, 2, ..., 200$, where $\{N_t\}_{t=1,2,...,200}$ is a sequence of independent normal random variables with zero mean and unit variance, and $\sin(\omega t + \theta)$ is the signal component. There are many approaches to determine the unknown signal components given the data X_t and one such approach is smoothing. Smoothing data removes random variation and shows trends and cyclic components. The series in this example is called signal plus noise.*

Definition 1.4 *A* **time series model** *for the observed data $\{x_t\}$ is a specification of the joint distributions of a sequence of random variables $\{X_t\}_{t \in T}$ of which $X_t = x_t$ is postulated to be a realization. We also refer to the stochastic process $\{X_t\}_{t \in T}$ as time series.*

1.2.2 Objectives of Time Series Analysis

A modest objective of any time series analysis is to provide a concise description of the past values of a series or a description of the underlying process that generates the time series. A plot of the data shows the important features of the series such as the trend, seasonality, and any discontinuities. Plotting the data of a time series may suggest a removal of seasonal components in order not to confuse them with long-term trends, known as *seasonal adjustment*. Other applications of time series models include the separation of noise from signals, forecasting future values of a time series using historical data and testing hypotheses. When time series observations are taken on two or more variables, it may be possible to use the variation in one time series to explain the variation of the other. This may lead to a deeper understanding of the mechanism generating the given observations [Cha04].

In this essay, we shall concentrate on autoregressive moving average (ARMA) processes as a special class of time series models most commonly used in practical applications.

1.3 Stationary Models and the Autocorrelation Function

Definition 1.5 *The* **mean function** *of a time series* $\{X_t\}$ *with* $E(X_t^2) < \infty$ *is*

$$\mu_X(t) = E(X_t), t \in \mathbb{Z},$$

and the **covariance function** *is*

$$\gamma_X(r, s) = \text{cov}(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))]$$

for all $r, s \in \mathbb{Z}$*.*

Definition 1.6 *A time series* $\{X_t\}_{t \in T}$ *is* **weakly stationary** *if both the mean*

$$\mu_X(t) = \mu_X,$$

and for each $h \in \mathbb{Z}$*, the covariance function*

$$\gamma_X(t + h, t) = \gamma_X(h)$$

are independent of time t*.*

Definition 1.7 *A time series* $\{X_t\}$ *is said to be* **strictly stationary** *if the joint distributions of* $(X_1, \ldots, X_n)$ *and* $(X_{1+h}, \ldots, X_{n+h})$ *are the same for all* $h \in \mathbb{Z}$ *and* $n > 0$*.*

Definitions 1.6 and 1.7 imply that if a time series $\{X_t\}$ is strictly stationary and satisifies the condition $E(X_t^2) < \infty$, then $\{X_t\}$ is also weakly stationary. Therefore we assume that $E(X_t^2) < \infty$. In this essay, *stationary* refers to *weakly stationary*.

For a single variable, the covariance function of a stationary time series $\{X_t\}_{t \in T}$ is defined as

$$\gamma_X(h) = \gamma_X(h, 0) = \gamma_X(t + h, t),$$

where $\gamma_X(\cdot)$ is the autocovariance function and $\gamma_X(h)$ its value at lag h.

Proposition 1.8 *Every Gaussian weakly stationary process is strictly stationary.*

Proof 1.9 *The joint distributions of any Gaussian stochastic process are uniquely determined by the second order properties, i.e., by the mean* μ *and the covariance matrix. Hence, since the process is weakly stationary, the second order properties do not depend on time* t *(see Definition 1.6). Therefore, it is strictly stationary.*

Definition 1.10 *The* **autocovariance function (acvf)** *and the* **autocorrelation function (acf)** *of a stationary time series* $\{X_t\}$ *at lag* h *are given respectively as:*

$$\gamma_X(h) = \text{cov}(X_{t+h}, X_t), \tag{1.1}$$

and

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \text{cor}(X_{t+h}, X_t). \tag{1.2}$$

The linearity property of covariances is that, if $E(X^2)$, $E(Y^2)$ and $E(Z^2) < \infty$, and a and b are real constants, then

$$\text{cov}(aX + bY, Z) = a\text{cov}(X, Z) + b\text{cov}(Y, Z) \tag{1.3}$$

Proposition 1.11 *If γ_X is the autocovariance function of a stationary process $\{X_t\}_{t\in\mathbb{Z}}$, then*

(i) $\gamma_X(0) \geq 0$,

(ii) $|\gamma_X(h)| \leq \gamma_X(0), \quad \forall \quad h \in \mathbb{Z}$,

(iii) $\gamma_X(h) = \gamma_X(-h)$.

Proof 1.12 *(i) From* $\text{var}(X_t) \geq 0$,

$$\gamma_X(0) = \text{cov}(X_t, X_t) = E(X_t - \mu_X)(X_t - \mu_X) = E(X_t^2) - \mu_X^2 = \text{var}(X_t) \geq 0.$$

(ii) From Cauchy-Schwarz inequality,

$$\begin{aligned} |\gamma_X(h)| &= |\text{cov}(X_{t+h}, X_t)| = |E(X_{t+h} - \mu_X)(X_t - \mu_X)| \\ &\leq [E(X_{t+h} - \mu_X)^2]^{\frac{1}{2}}[E(X_t - \mu_X)^2]^{\frac{1}{2}} \leq \gamma_X(0) \end{aligned}$$

(iii) $\gamma_X(h) = \text{cov}[X_t, X_{t+h}] = \text{cov}[X_{t-h}, X_t] = \gamma_X(-h)$
since $\{X_t\}$ *is stationary.*

Definition 1.13 *The* **sample mean** *of a time series $\{X_t\}$ with $x_1, \ldots, x_n$ as its observations is*

$$\overline{x} = \frac{1}{n}\sum_{t=1}^{n} x_t,$$

and the **sample autocovariance function** *and the* **sample autocorrelation function** *are respectively given as*

$$\widehat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \overline{x})(x_t - \overline{x}),$$

and

$$\widehat{\rho}(h) = \frac{\widehat{\gamma}(h)}{\widehat{\gamma}(0)},$$

where $-n < h < n$.

Remark 1.14 *The sample autocorrelation function is useful in determining the non-stationarity of data since it exhibits the same features as the plot of the data. An example is shown in Figure 1.2. It can be seen that the plot of the acf exhibits the same oscillations, i.e. seasonal effects, as the plot of the data.*

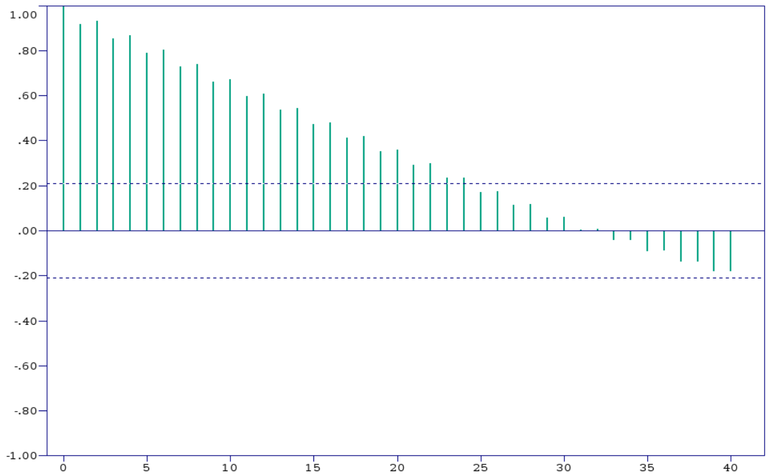


Figure 1.2: The sample acf of the concentration of CO_2 in the atmoshpere, see Remark 1.14

Definition 1.15 *A sequence of independent random variables $X_1, ..., X_n$, in which there is no trend or seasonal component and the random variables are identically distributed with mean zero is called* **iid noise**.

Example 1.16 *If the time series $\{X_t\}_{t\in\mathbb{Z}}$ is iid noise and $E(X_t^2) = \sigma^2 < \infty$, then the mean of $\{X_t\}$ is independent of time t, since $E(X_t) = 0$ for all t and*

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

is independent of time. Hence iid noise with finite second moments is stationary.

Remark 1.17 $\{X_t\} \sim IID(0, \sigma^2)$ *indicates that the random variables X_t, $t \in \mathbb{Z}$ are independent and identically distributed, each with zero mean and variance σ^2.*

Example 1.18 White noise *is a sequence $\{X_t\}$ of uncorrelated random variables such that each random variable has zero mean and a variance of σ^2. We write $\{X_t\} \sim WN(0, \sigma^2)$. Then $\{X_t\}$ is stationary with the same covariance function as the iid noise. A plot of white noise time series is shown in Figure 1.3.*

Example 1.19 *A sequence of iid random variables $\{X_t\}_{t\in\mathbb{N}}$ with*

$$P(X_t = 1) = p,$$

and

$$P(X_t = -1) = 1 - p$$

is called a **binary process**, *e.g. tossing a coin with $p = \frac{1}{2}$.*

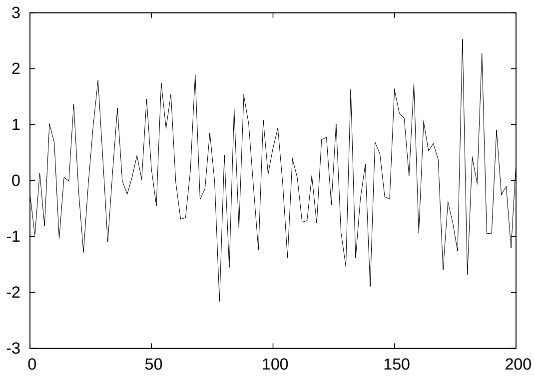


Figure 1.3: Simulated stationary white noise time series

A time series $\{X_t\}$ *is a* **random walk** *with* $\{Z_t\}$ *a white noise, if*

$$X_t = X_{t-1} + Z_t.$$

Starting the process at time $t = 0$ *and* $X_0 = 0$*,*

$$\begin{aligned} X_t &= Z_t + Z_{t-1} + \ldots + Z_1, \\ E(X_t) &= E(Z_t + Z_{t-1} + \ldots + Z_1) = 0. \\ \mathrm{var}(X_t) &= E(X_t^2) \\ &= E\left[(Z_t + Z_{t-1} + \ldots + Z_1)^2\right] \\ &= E(Z_t)^2 + E(Z_{t-1})^2 + \ldots + E(Z_1)^2 \text{ (independence)} \\ &= t\sigma^2. \end{aligned}$$

The time series $\{X_t\}$ *is non stationary since the variance increases with* t*.*

Example 1.20 *A time series* $\{X_t\}$ *is a* **first order moving average, MA(1)**, *if*

$$X_t = Z_t + \theta Z_{t-1}, \quad t \in \mathbb{Z}, \tag{1.4}$$

where $\{Z_t\} \sim WN(0, \sigma^2)$ *and* $|\theta| < 1$*.*
From Example 1.16,

$$\begin{aligned} E(X_t) &= E(Z_t + \theta Z_{t-1}) \\ &= E(Z_t) + \theta E(Z_{t-1}) = 0, \end{aligned}$$

$$\begin{aligned} E(X_t X_{t+1}) &= E\left[(Z_t + \theta Z_{t-1})(Z_{t+1} + \theta Z_t)\right] \\ &= E\left[Z_t Z_{t+1} + \theta Z_t^2 + \theta Z_{t-1} Z_{t+1} + \theta^2 Z_t Z_{t+1}\right] \\ &= E(Z_t Z_{t+1}) + \theta E(Z_t^2) + \theta E(Z_{t-1} Z_{t+1}) + \theta^2 E(Z_t Z_{t+1}) \\ &= \theta\sigma^2, \end{aligned}$$

and

$$\begin{aligned} E(X_t^2) &= E(Z_t + \theta Z_{t-1})^2 \\ &= E(Z_t^2) + 2\theta E(Z_t)E(Z_{t-1}) + \theta^2 E(Z_{t-1}^2) \\ &= \sigma^2(1+\theta^2) < \infty. \end{aligned}$$

Hence

$$\begin{aligned} \gamma_X(t+h,t) &= \operatorname{cov}(X_{t+h},X_t) = E(X_{t+h},X_t) \\ &= \begin{cases} \sigma^2(1+\theta^2), & \textit{if } h=0, \\ \sigma^2\theta, & \textit{if } h=\pm 1, \\ 0, & \textit{if } \mid h \mid > 1. \end{cases} \end{aligned}$$

Since $\gamma_X(t+h,t)$ *is independent of* t, $\{X_t\}$ *is stationary.*

The **autocorrelation function** *is given as:*

$$\begin{aligned} \rho_X(h) &= \operatorname{cor}(X_{t+h},X_t) = \frac{\gamma_X(h)}{\gamma_X(0)} \\ &= \begin{cases} 1, & \textit{if } h=0, \\ \dfrac{\theta}{1+\theta^2}, & \textit{if } h=\pm 1, \\ 0, & \textit{if } \mid h \mid > 1. \end{cases} \end{aligned}$$

Example 1.21 *A time series* $\{X_t\}$ *is a* **first-order autoregressive or AR(1)** *process if*

$$X_t = \phi X_{t-1} + Z_t,\ t \in \mathbb{Z}, \tag{1.5}$$

where $\{Z_t\} \sim WN(0,\sigma^2)$, $\mid \phi \mid < 1$, *and for each* $s < t$, Z_t *is uncorrelated with* X_s.

$$\begin{aligned} \gamma_X(0) &= E(X_t^2) \\ &= E\left[(\phi X_{t-1} + Z_t)(\phi X_{t-1} + Z_t)\right] \\ &= \phi^2 E(X_{t-1}^2) + E(Z_t^2) \\ &= \phi^2 \gamma_X(0) + \sigma^2. \end{aligned}$$

$Solving\ for\ \gamma_X(0),$

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

Since Z_t *and* X_{t-h} *are uncorrelated,* $\operatorname{cov}(Z_t, X_{t-h}) = 0, \quad \forall \quad h \neq 0.$
The autocovariance function at lag $h > 0$ *is*

$$\begin{aligned} \gamma_X(h) &= \operatorname{cov}(X_t, X_{t-h}) \\ &= \operatorname{cov}(\phi X_{t-1} + Z_t, X_{t-h}) \textit{ from Equation 1.3} \\ &= \phi \operatorname{cov}(X_{t-1}, X_{t-h}) + \operatorname{cov}(Z_t, X_{t-h}) = \phi \operatorname{cov}(X_{t-1}, X_{t-h}) \\ &= \phi \operatorname{cov}(\phi X_{t-2} + Z_{t-1}, X_{t-h}) = \phi^2 \operatorname{cov}(X_{t-2}, X_{t-h}) \end{aligned}$$

Continuing this procedure,

$$\begin{aligned} \gamma_X(h) &= \phi^h \gamma_X(0),\ h \in \mathbb{Z}_+ \\ \gamma_X(h) &= \phi^{|h|} \gamma_X(0),\ h \in \mathbb{Z}, \end{aligned}$$

due to symmetry of $\gamma_X(h)$ (see Proposition 1.11 (iii)). Therefore, the autocorrelation function is given as

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \frac{\phi^{|h|}\gamma_X(0)}{\gamma_X(0)} = \phi^{|h|},\ h \in \mathbb{Z}. \tag{1.6}$$

1.4 Removing Trend and Seasonal Components

Many time series that arise in practice are non-stationary but most available techniques are for analysis of stationary time series. In order to make use of these techniques, we need to modify our time series so that it is stationary. An example of a non-stationary time series is given in Figure 1.1(a). By a suitable transformation, we can transform a non-stationary time series into an approximately stationary time series. Time series data are influenced by a variety of factors. It is essential that such components are decomposed out of the raw data. In fact, any time series can be decomposed into a **trend component** m_t, a **seasonal component** s_t with known period, and a **random noise** Y_t as

$$X_t = m_t + s_t + Y_t,\ t \in \mathbb{Z}. \tag{1.7}$$

For nonseasonal models with trend, Equation (1.7) becomes

$$X_t = m_t + Y_t,\ t \in \mathbb{Z}. \tag{1.8}$$

For a nonseasonal time series, first order differencing is usually sufficient to attain apparent stationarity so that the new series $(y_1, \ldots, y_n)$ is formed from the original series $(x_1, \ldots, x_n)$ by $y_t = x_{t+1} - x_t = \nabla x_{t+1}$. Second order differencing is required using the operator ∇^2, where $\nabla^2 x_{t+2} = \nabla x_{t+2} - \nabla x_{t+1} = x_{t+2} - 2x_{t+1} + x_t$. The concept of **backshift operator B** helps to express differenced ARMA models, where $BX_t = X_{t-1}$. The number of times the original series is differenced to achieve stationarity is called the **order of homogeneity**. Trends in variance are removed by taking Logarithms of the time series data so that it changes from the trend in variance to that of the mean.

Remark 1.22 *For other techniques of removing trend and seasonal components and a detailed description, see [BD02], Chapter 1.5.*

Remark 1.23 *From now on, we assume that $\{X_t\}$ is a stationary time series, i.e., we assume that data has been transformed and there is no trend and seasonal component.*

2. ARMA Processes

In this chapter, we will introduce the general autoregressive moving average model (ARMA) and some properties, particularly, the spectral density and the autocorrelation function.

2.1 Definition of ARMA Processes

ARMA processes are a combination of autoregressive (AR) and moving average (MA) processes. The advantage of an ARMA process over AR and MA is that a stationary time series may often be described by an ARMA process involving fewer parameters than pure AR or MA process.

Definition 2.1 *A time series* $\{X_t\}$ *is said to be a* **moving average process of order** q **(MA(q))** *if it is a weighted linear sum of the last* q *random shocks so that*

$$X_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q},\ t \in \mathbb{Z}, \tag{2.1}$$

where $\{Z_t\} \sim WN(0, \sigma^2)$*. Using the backshift operator B, Equation (2.1) becomes*

$$X_t = \theta(B) Z_t, \tag{2.2}$$

where $\theta_1, \ldots, \theta_q$ *are constants and* $\theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q$ *is a polynomial in B of order* q*.*

A finite-order MA process is stationary for all parameter values, but an *invertibility condition* must be imposed on the parameter values to ensure that there is a unique MA model for a given autocorrelation function.

Definition 2.2 *A time series* $\{X_t\}$ *is an* **autoregressive process of order** p **(AR(p))** *if*

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + Z_t,\ t \in \mathbb{Z}, \tag{2.3}$$

where $\{Z_t\} \sim WN(0, \sigma^2)$*. Using the backshift operator* B*, Equation (2.3) becomes*

$$\phi(B) X_t = Z_t, \tag{2.4}$$

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ *is a polynomial in* B *of order* p*.*

Definition 2.3 *A time series* $\{X_t\}$ *is an* **ARMA(p,q) process** *if it is stationary and*

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \ldots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q},\ t \in \mathbb{Z}, \tag{2.5}$$

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

Using the backshift operator, Equation (2.5) becomes

$$\phi(B)X_t = \theta(B)Z_t,\ t \in \mathbb{Z}, \tag{2.6}$$

where $\phi(B)$, $\theta(B)$ are polynomials of order p, q respectively, such that

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p, \tag{2.7}$$

$$\theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q, \tag{2.8}$$

and the polynomials have no common factors.

Remark 2.4 *We refer to $\phi(B)$ as the* autoregressive polynomial *of order p and $\theta(B)$ as the* moving average polynomial *of order q.*

Theorem 2.5 *A stationary solution $\{X_t\}_{t\in\mathbb{Z}}$ of the ARMA Equation (2.6) exists and it is unique if and only if*

$$\phi(z) \neq 0 \quad \forall \quad |z| = 1. \tag{2.9}$$

The proof can be found in [BD87], Theorem 3.1.3.

2.1.1 Causality and Invertibility of ARMA Processes

Causality of a time series $\{X_t\}$ means that X_t is expressible in terms of Z_s, where $t > s$, and $Z_t \sim WN(0, \sigma^2)$. This is important because, we then only need to know the past values of Z_s for $s < t$ in order to determine the present value of X_t, i.e. we do not need to know the future values of the white noise sequence.

Definition 2.6 *An ARMA(p,q) process defined by Equation (2.6) is said to be* **causal** *if there exists a sequence of constants $\{\psi_i\}$ such that $\sum_{i=0}^{\infty} |\psi_i| < \infty$ and*

$$X_t = \sum_{i=0}^{\infty} \psi_i Z_{t-i},\ t \in \mathbb{Z}. \tag{2.10}$$

Theorem 2.7 *If $\{X_t\}$ is an ARMA(p,q) process for which the polynomials $\theta(z)$ and $\phi(z)$ have no common zeros, then $\{X_t\}$ is* **causal** *if and only if $\phi(z) \neq 0\ \forall\ z \in \mathbb{C}$ such that $|z|\ \leq 1$. The coefficients $\{\psi_i\}$ are determined by the relation*

$$\Psi(z) = \sum_{i=0}^{\infty} \psi_i z^i = \frac{\theta(z)}{\phi(z)}, \quad |z| \leq 1. \tag{2.11}$$

The proof can be found in [BD87], Theorem 3.1.1.

Equation (1.5) can be expressed as a moving average process of order ∞ (MA(∞)). By iterating, we have

$$\begin{aligned} X_t &= Z_t + \phi Z_{t-1} + \phi^2 X_{t-2} \\ &= Z_t + \phi Z_{t-1} + \ldots + \phi^k Z_{t-k} + \phi^{k+1} X_{t-k-1}. \end{aligned}$$

From Example 1.21, $|\phi| < 1$, $\{X_t\}$ is stationary and $\mathrm{var}(X_t^2) = E(X_t^2) = constant$. Therefore,

$$\| X_t - \sum_{i=0}^{k} \phi^i Z_{t-i} \|^2 = \phi^{2(k+1)} \| X_{t-k-1} \|^2 \longrightarrow 0 \ as \ k \to \infty. \tag{2.12}$$

From Equation (2.12),

$$X_t = \sum_{i=0}^{\infty} \phi^i Z_{t-i}, \quad t \in \mathbb{Z}. \tag{2.13}$$

Invertibility of a stationary time series $\{X_t\}$ means that Z_t is expressible in terms of X_s where $t > s$ and $Z_t \sim WN(0, \sigma^2)$.

Definition 2.8 *An ARMA(p,q) process defined by Equation 2.6 is said to be* **invertible** *if there exists a sequence of constants* $\{\pi_i\}$ *such that* $\sum_{i=0}^{\infty} |\pi_i| < \infty$ *and*

$$Z_t = \sum_{i=0}^{\infty} \pi_i X_{t-i}, \ t \in \mathbb{Z}. \tag{2.14}$$

Theorem 2.9 *If* $\{X_t\}$ *is an ARMA(p,q) process for which the polynomials* $\theta(z)$ *and* $\phi(z)$ *have no common zeros, then* $\{X_t\}$ *is* **invertible** *if and only if* $\theta(z) \neq 0 \ \forall \ z \in \mathbb{C}$ *such that* $|z| \ \leq 1$. *The coefficients* $\{\pi_i\}$ *are determined by the relation*

$$\Pi(z) = \sum_{i=0}^{\infty} \pi_i z^i = \frac{\phi(z)}{\theta(z)}, \quad |z| \leq 1. \tag{2.15}$$

See [BD87], Theorem 3.1.2 for the proof.

Equation (1.4) can be expressed as an autoregressive process of order ∞ (AR(∞)). By iterating, we have

$$\begin{aligned} Z_t &= X_t - \theta Z_{t-1} \\ &= X_t - \theta X_{t-1} + \theta^2 X_{t-2} + \ldots + \theta^k X_{t-k} + \theta^{k+1} Z_{t-k-1}. \end{aligned}$$

From Example 1.20, $|\theta| < 1$, $\{X_t\}$ is stationary and $\mathrm{var}(X_t^2) = E(X_t^2) = constant$. Therefore,

$$\| Z_t - \sum_{i=0}^{k} \theta^i X_{t-i} \|^2 = \theta^{2(k+1)} \| Z_{t-k-1} \|^2 \longrightarrow 0 \ as \ k \to \infty. \tag{2.16}$$

From Equation (2.16),

$$Z_t = \sum_{i=0}^{\infty} \theta^i X_{t-i}, \quad t \in \mathbb{Z}. \tag{2.17}$$

Example 2.10 *[Cha01] Suppose that* $\{Z_t\} \sim WN(0, \sigma^2)$ *and* $\{Z_t'\} \sim WN(0, \sigma^2)$ *and* $\theta \in (-1, 1)$. *From Example 1.20, the MA(1) processes given by*

$$X_t = Z_t + \theta Z_{t-1}, \quad t \in \mathbb{Z}, \tag{2.18}$$

and

$$X_t = Z_t' + \frac{1}{\theta} Z_{t-1}', \quad t \in \mathbb{Z}, \tag{2.19}$$

have the same autocorrelation function. Inverting the two processes by expressing Z_t *in terms of* X_t *gives*

$$Z_t = X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \dots \tag{2.20}$$
$$Z_t' = X_t - \theta^{-1} X_{t-1} + \theta^{-2} X_{t-2} - \dots \tag{2.21}$$

The series of coefficients of X_{t-k} *in Equation (2.20) converges since* $|\theta| < 1$ *and that of Equation (2.21) diverges. This implies the process (2.19) cannot be inverted.*

Example 2.11 *Let* $\{X_t\}$ *be an ARMA(2,1) process defined by*

$$X_t - X_{t-1} + \frac{1}{4} X_{t-2} = Z_t - \frac{1}{3} Z_{t-1}, \tag{2.22}$$

where $\{Z_t\}_{t\in\mathbb{Z}} \sim WN(0, \sigma^2)$. *Using the backshift operator* B, *Equation (2.22) can be written as*

$$(1 - \frac{1}{2}B)^2 X_t = (1 - \frac{1}{3}B) Z_t. \tag{2.23}$$

The AR polynomial $\phi(z) = (1 - \frac{1}{2}z)^2$ *has zeros, at* $z = 2$, *which lie outside the unit circle. Hence* $\{X_t\}_{t\in\mathbb{Z}}$ *is causal according to Theorem 2.7.*

The MA polynomial $\theta(z) = 1 - \frac{1}{3}z$ *has a zero at* $z = 3$, *also located outside the unit circle* $|z| \leq 1$. *Hence* $\{X_t\}_{t\in\mathbb{Z}}$ *is invertible from Theorem 2.9. In particular,* $\phi(z)$ *and* $\theta(z)$ *have no common zeros.*

Example 2.12 *Let* $\{X_t\}$ *be an ARMA(1,1) process defined by*

$$X_t - 0.5 X_{t-1} = Z_t + 0.4 Z_{t-1}, \tag{2.24}$$

where $\{Z_t\}_{t\in\mathbb{Z}} \sim WN(0, \sigma^2)$. *Using the backshift operator* B, *Equation (2.24) can be written as*

$$(1 - 0.5B) X_t = (1 + 0.4B) Z_t. \tag{2.25}$$

The AR polynomial $\phi(z) = 1 - 0.5z$ *has a zero at* $z = 2$ *which lies outside the unit circle. Hence* $\{X_t\}_{t\in\mathbb{Z}}$ *is causal according to Theorem 2.7.*

The MA polynomial $\theta(z) = 1 + 0.4z$ *has a zero at* $z = 2.5$ *which is located outside the unit circle* $|z| \leq 1$. *Hence* $\{X_t\}_{t\in\mathbb{Z}}$ *is invertible from Theorem 2.9. In particular,* $\phi(z)$ *and* $\theta(z)$ *have no common zeros.*

$$\begin{aligned} X_t &= Z_t + 0.4Z_{t-1} + 0.5X_{t-1} \\ &= Z_t + 0.4Z_{t-1} + 0.5\left(Z_{t-1} + 0.4Z_{t-2} + 0.5X_{t-2}\right) \\ &= Z_t + (0.4 + 0.5)Z_{t-1} + (0.5 \cdot 0.4)Z_{t-1} + 0.5^2X_{t-2} \\ &= Z_t + (0.4 + 0.5)Z_{t-1} + (0.5 \cdot 0.4)Z_{t-1} + 0.5^2\left(Z_{t-2} + 0.4Z_{t-3} + 0.5X_{t-3}\right) \\ &= Z_t + (0.4 + 0.5)Z_{t-1} + (0.5^2 + 0.5 \cdot 0.4)Z_{t-2} + 0.5^2 \cdot 0.4Z_{t-3} + 0.5^3X_{t-3}. \end{aligned}$$

Continuing this process, we get the causal representation of $\{X_t\}$ *as*

$$X_t = Z_t + \sum_{j=1}^{n}\left(0.5^{j-1} \cdot 0.4 + 0.5^j\right)Z_{t-j} + 0.5^nX_{t-n}$$

since 0.5^nX_{t-n} *tends to* 0 *as* n *tends to* ∞,

$$X_t = Z_t + 0.9\sum_{j=1}^{\infty}0.5^{j-1}Z_{t-j}, \quad t \in \mathbb{Z}$$

.

2.2 Properties of ARMA Processes

2.2.1 The Spectral Density

Spectral representation of a stationary process $\{X_t\}_{t\in\mathbb{Z}}$ decomposes $\{X_t\}_{t\in\mathbb{Z}}$ into a sum of sinusoidal components with uncorrelated random coefficients. The spectral point of view is advantageous in the analysis of multivariate stationary processes, and in the analysis of very large data sets, for which numerical calculations can be performed rapidly using the fast Fourier transform. The spectral density of a stationary stochastic process is defined as the Fourier transform of its autocovariance function.

Definition 2.13 *The spectral density of a discrete time series* $\{X_t\}$ *is the function* $f(\cdot)$ *given by*

$$f(\lambda) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}e^{-ih\lambda}\gamma(h), \quad -\pi \leq \lambda \leq \pi, \tag{2.26}$$

where $e^{i\lambda} = \cos(\lambda) + i\sin(\lambda)$ *and* $i = \sqrt{-1}$. *The sum in (2.26) converges since* $|e^{ih\lambda}|^2 = \cos^2(h\lambda) + \sin^2(h\lambda) = 1$ *converges and* $|\gamma(h)|$ *is bounded by Proposition 1.11 (ii). The period of* f *is the same as those of* $\sin$ *and* $\cos$.

The spectral density function has the following properties [BD02], Chapter 4.1.

Proposition 2.14 *Let f be the spectral density of a time series $\{X_t\}_{t\in\mathbb{Z}}$. Then*

(i) f *is even, i.e.,* $f(\lambda) = f(-\lambda)$,

(ii) $f(\lambda) \geq 0 \ \forall \ \lambda \in [-\pi, \pi]$,

(iii) $\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda) d\lambda = \int_{-\pi}^{\pi} \cos(h\lambda) f(\lambda) d\lambda, \quad h \in \mathbb{Z}.$

Theorem 2.15 *If $\{X_t\}$ is a causal ARMA(p,q) process satisfying Equation (2.6), then its spectral density is given by*

$$f_X(\lambda) = \frac{\sigma^2 |\theta(e^{-i\lambda})|^2}{2\pi |\phi(e^{-i\lambda})|^2}, \ -\pi \leqslant \lambda \leqslant \pi. \tag{2.27}$$

See [BD02], Chapter 4.4.

Example 2.16 *The MA(1) process given by*

$$X_t = Z_t + \theta Z_{t-1}, \quad t \in \mathbb{Z}, \tag{2.28}$$

has spectral density due to Equation (2.27) as

$$\begin{aligned} f_X(\lambda) &= \frac{\sigma^2}{2\pi}(1 + \theta e^{i\lambda})(1 + \theta e^{-i\lambda}) \\ &= \frac{\sigma^2}{2\pi}(1 + 2\theta \cos(\lambda) + \theta^2), \end{aligned}$$

since $2\cos(\lambda) = e^{-i\lambda} + e^{i\lambda}$.

2.2.2 The Autocorrelation Function

For stationary processes, the autocorrelation function (acf) $\rho_X(h)$ defined by Equation (1.2) measures the correlation at lag h between X_t and X_{t+h}.

Theorem 2.17 *Let $\{X_t\}_{t\in\mathbb{Z}}$ be an ARMA(p,q) process defined by Equation (2.6) with spectral density f_X. Then $\{X_t\}$ has autocovariance function γ_X given by*

$$\gamma_X(h) = \frac{\sigma^2}{2\pi} \int_{-\pi}^{\pi} e^{i\lambda h} \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2} d\lambda, \quad h \in \mathbb{Z}. \tag{2.29}$$

Proof 2.18 *The proof follows if we substitute Equation (2.27) in Proposition 2.14 (iii).*

Example 2.19 *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a causal MA(q) process given by Equation (2.1). The causality ensures that $\{X_t\}_{t\in\mathbb{Z}}$ can be written in the form*

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad \{Z_t\} \sim WN(0, \sigma^2). \tag{2.30}$$

$$\Psi(z) = \sum_{i=0}^{\infty} \psi_i z^i = \frac{\theta(z)}{\phi(z)} = 1 + \theta_1 z + \ldots + \theta_q z^q.$$

Hence

$$X_t = \sum_{j=0}^{q} \theta_j Z_{t-j}, \quad \{Z_t\} \sim WN(0, \sigma^2). \tag{2.31}$$

$$\begin{aligned} E(X_t X_{t+h}) &= E\left[\sum_{j=0}^{q} \theta_j Z_{t-j} \sum_{i=0}^{q} \theta_i Z_{t-i+h}\right] \\ &= \sum_{j=0}^{q}\sum_{i=0}^{q} \theta_j \theta_i E(Z_{t-j} Z_{t-i+h}) \\ &= \sigma^2 \sum_{j=0}^{q} \theta_j \theta_{j+|h|}, \end{aligned}$$

since $E(Z_{t-j}Z_{t-i+h}) = \sigma^2$ only when $i = j + |h|$. The autocovariance function of Equation (2.31) is therefore

$$\gamma(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q} \theta_j \theta_{j+|h|}, & \textit{if } |h| \le q, \\ 0, & \textit{otherwise}, \end{cases} \tag{2.32}$$

where $\theta_0 = 1$ and $\theta_j = 0$ for $j > q$.

Figures 2.1 and 2.4 show the plots of the acf of Example 2.16 for $\theta = 0.9$ and $\theta = -0.9$ respectively. The acf is zero after lag 1 in both plots and it is negative for $\theta = -0.9$ and positive for $\theta = 0.9$.

Figures 2.3 and 2.6 show the plots of the spectral density of Example 2.16 for $\theta = 0.9$ and $\theta = -0.9$ respectively. The spectral density of Figure 2.3 is large for low frequencies, and small for high frequencies (for $\theta > 0$), since the process has a large lag one positive autocorrelation as seen in Figure 2.1. Similarly, the spectral density of Figure 2.6 is negatively large for low frequencies since the process has a negative autocorrelation at lag one (since $\theta > 0$) as seen in Figure 2.4.

Figure 2.7 and 2.8 show the plots of the acf and spectral density of an ARMA(2,1) process with $\phi_1 = 0.5$, $\phi_2 = -0.4$ and $\theta = 0.9$. The spectral density is maximum at lag 1 and minimum at lag 3 since from Figure 2.7, the acf is maximum at lag 1 and minimum at lag 3.

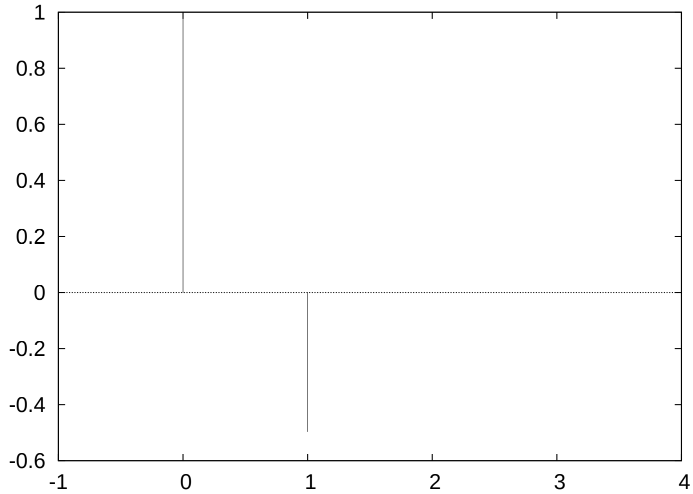


Figure 2.1: The acf for MA(1) with $\theta = -0.9$

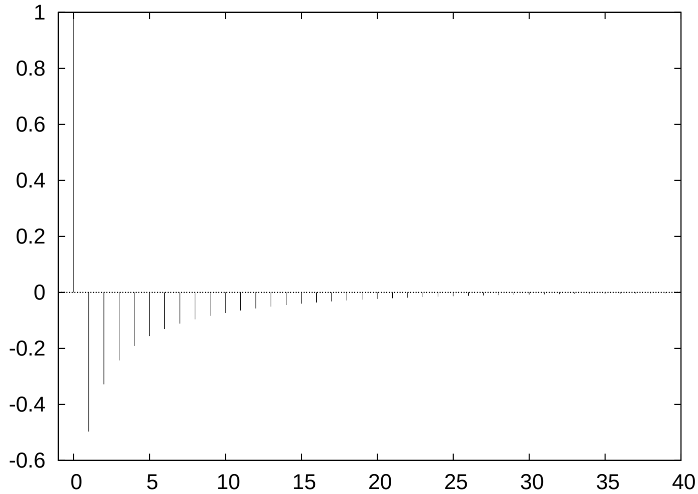


Figure 2.2: The pacf for MA(1) with $\theta = -0.9$

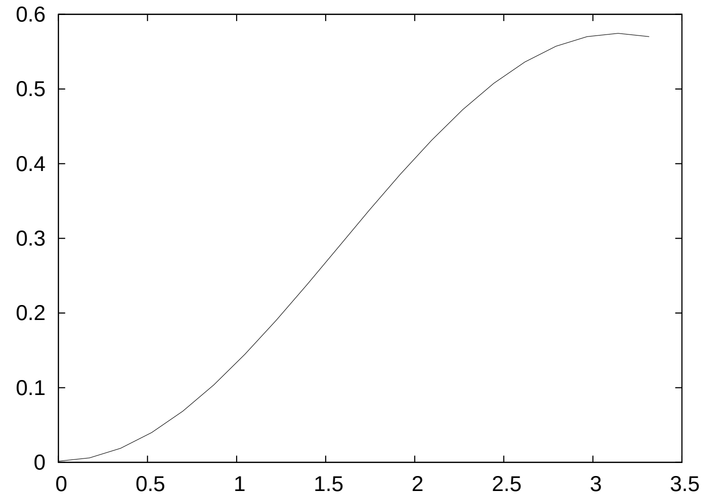


Figure 2.3: The spectral density of MA(1) with $\theta = -0.9$

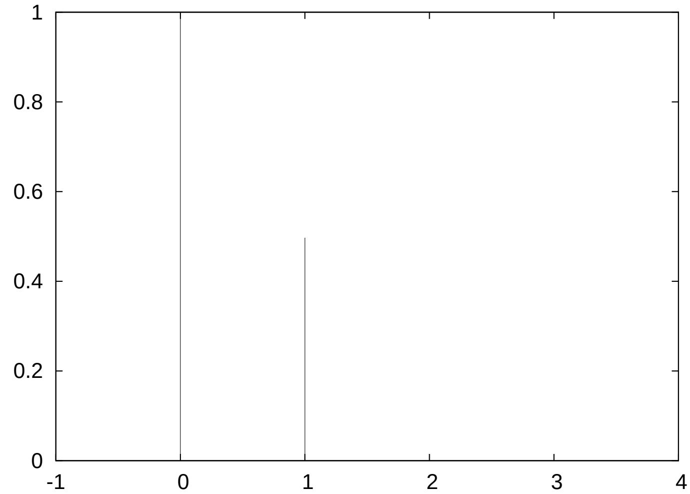


Figure 2.4: The acf for MA(1) with $\theta = 0.9$

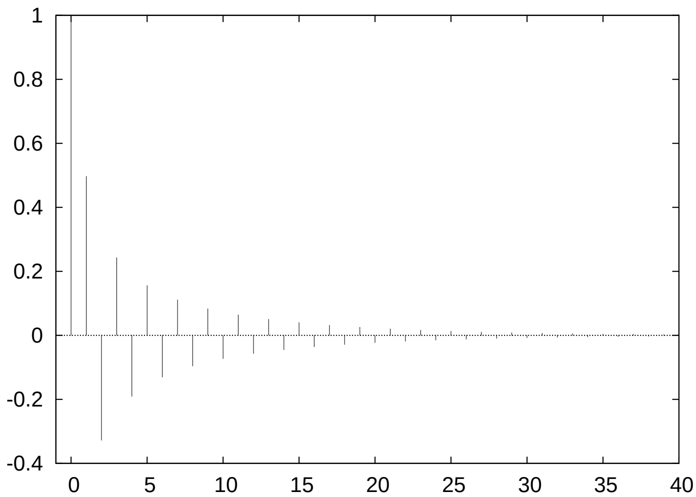


Figure 2.5: The pacf for MA(1) with $\theta = 0.9$

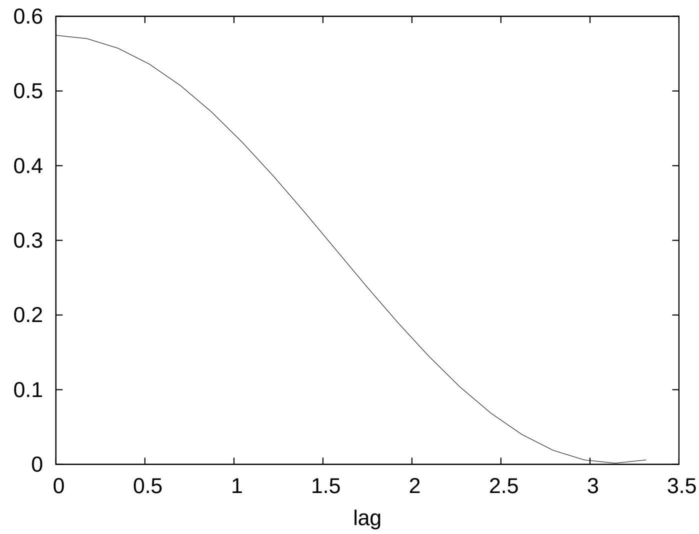


Figure 2.6: The spectral density of MA(1) with $\theta = 0.9$

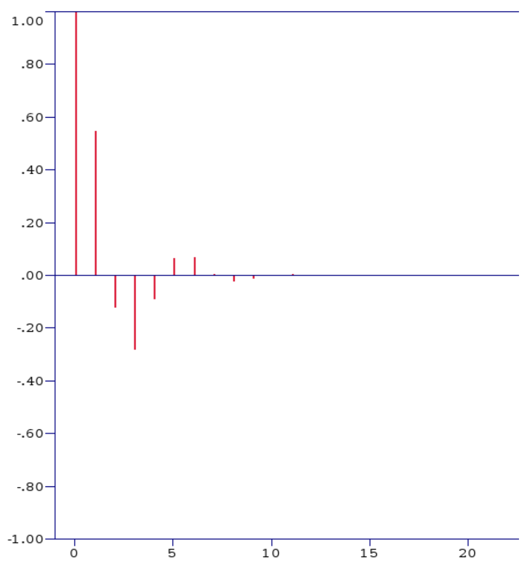


Figure 2.7: The acf of ARMA(2,1) with $\theta_1 = 0.9, \phi_1 = 0.5$ and $\phi_2 = -0.4$

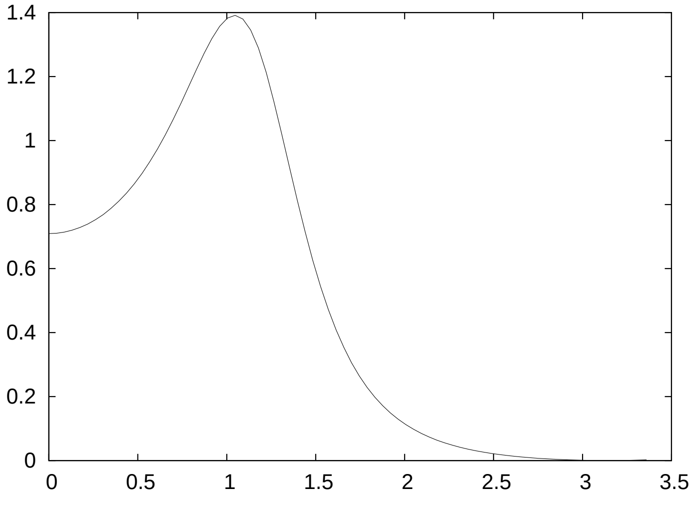


Figure 2.8: The spectral density of ARMA(2,1) with $\theta_1 = 0.9, \phi_1 = 0.5$ and $\phi_2 = -0.4$

3. Prediction

In this chapter, we investigate the problem of predicting the values $\{X_t\}_{t\geq n+1}$ of a stationary process in terms of $\{X_t\}_{t=1,\ldots,n}$.

Let $\{X_t\}$ be a stationary process with $E(X_t) = 0$, and autocovariance function γ. Suppose we have observations $x_1, x_2, \ldots, x_n$ and we want to find a linear combination of $x_1, x_2, \ldots, x_n$ that estimates x_{n+1}, i.e.

$$\widehat{X}_{n+1} = \sum_{i=0}^{n} \Phi_{n_i} X_i, \tag{3.1}$$

such that the mean squared error

$$E|X_{n+1} - \widehat{X}_{n+1}|^2 \tag{3.2}$$

is minimized.
Using the projection theorem Theorem 2.3.1 in [BD87], we can rewrite Equation (3.2) as

$$E\left[\left(X_{n+1} - \sum_{i=0}^{n} \Phi_{n_i} X_i\right) X_k\right] = 0, \quad \forall \quad k = 1, 2, \ldots, n. \tag{3.3}$$

From Equation (3.3), we have

$$E(X_{n+1}X_k) = \sum_{i=0}^{n} \Phi_{n_i} E(X_k X_i) \tag{3.4}$$

For $k = n$:

$$E(X_{n+1}X_n) = \gamma(1) = \sum_{i=0}^{n} \Phi_{n_i} E(X_n X_i) = \sum_{i=0}^{n} \Phi_{n_i} \gamma(n-i), \tag{3.5}$$

for $k = n - 1$:

$$E(X_{n+1}X_{n-1}) = \gamma(2) = \sum_{i=0}^{n} \Phi_{n_i} E(X_{n-1} X_i) = \sum_{i=0}^{n} \Phi_{n_i} \gamma(n-1-i), \tag{3.6}$$

continuing up to $k = 1$, we have

$$E(X_{n+1}X_1) = \gamma(n) = \sum_{i=0}^{n} \Phi_{n_i} E(X_1 X_i) = \sum_{i=0}^{n} \Phi_{n_i} \gamma(1-i). \tag{3.7}$$

Combining all the equations of the autocovariances, we have

$$\begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(n-1) \\ \gamma(n) \end{bmatrix} = \begin{bmatrix} \gamma(n-1) & \gamma(n-2) & \ldots & \gamma(1) & \gamma(0) \\ \gamma(n-2) & \gamma(n-3) & \ldots & \gamma(0) & \gamma(1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma(1) & \gamma(0) & \ldots & \gamma(n-3) & \gamma(n-2) \\ \gamma(0) & \gamma(1) & \ldots & \gamma(n-2) & \gamma(n-1) \end{bmatrix} \begin{bmatrix} \Phi_{n_1} \\ \Phi_{n_2} \\ \vdots \\ \Phi_{n_{n-1}} \\ \Phi_{n_n} \end{bmatrix} \tag{3.8}$$

$$\gamma_n = \Gamma_n \Phi_n \tag{3.9}$$

Due to the projection theorem, there exists a unique solution Φ_n if Γ_n is non-singular. This implies Γ_n^{-1} exists. Therefore

$$\Phi_n = \Gamma_n^{-1}\gamma_n. \tag{3.10}$$

We use the Durbin-Levinson, a recursive method, to calculate the prediction of X_{n+1}.

3.1 The Durbin-Levinson Algorithm

The Durbin-levinson algorithm is a recursive method for computing Φ_n and $v_n = E|X_n - \widehat{X}_n|^2$. Let $\widehat{X}_1 = 0$ and

$$\widehat{X}_{n+1} = \sum_{i=1}^{n} \Phi_{n_i} X_{n-i+1} \quad n = 1, 2, \ldots \tag{3.11}$$

$$= X_1\Phi_{n_n} + \ldots + X_n\Phi_{n_1}, \tag{3.12}$$

and the mean squared error of prediction be defined as

$$v_n = E(X_{n+1} - \widehat{X}_{n+1})^2, \tag{3.13}$$

where $v_0 = E(X_1)^2 = \gamma(0)$. $\Phi_n = (\Phi_{n_1}, \ldots, \Phi_{n_n})^T$ and v_n can be calculated recursively as follows:

Proposition 3.1 (The Durbin-Levinson Algorithm) *If $\{X_t\}$ is a stationary process with $E(X_t) = 0$ and autocovariance function γ such that $\gamma(0) > 0$ and $\gamma(h) \to 0$ as $h \to \infty$, then the coefficients Φ_{n_i} and the mean squared errors v_n given by Equations (3.12) and (3.13) satisfy*

$$\Phi_{nn} = \frac{1}{v_{n-1}}\left[\gamma(n) - \sum_{i=1}^{n-1} \phi_{n-1,i}\gamma(n-i)\right] \tag{3.14}$$

where

$$\begin{bmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{bmatrix} - \Phi_{nn} \begin{bmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix} \tag{3.15}$$

and

$$v_n = v_{n-1}[1 - \phi_{nn}^2] \tag{3.16}$$

with $\phi_{11} = \frac{\gamma(1)}{\gamma(0)}$ and $v_0 = \gamma(0)$.

Proof 3.2 *see [BD87], Chapter 5.2.*

4. Estimation of the Parameters

An appropriate ARMA(p,q) process to model an observed stationary time series is determined by the choice of p and q and the approximate calculation of the mean, the coefficients $\{\phi_j\}_{j=1,\dots,p}$, $\{\theta_i\}_{i=1,\dots,q}$ and the white noise variance σ^2. In this chapter, we will assume that the data has been adjusted by subtraction of the mean, and the problem thus transforms into fitting a zero-mean ARMA model to the adjusted data $\{X_t\}_{t\in\mathbb{Z}}$ for constant values of p and q.

4.1 The Yule-Walker Equations

Let $\{X_t\}_{t\in\mathbb{Z}}$ be the zero-mean causal autoregressive process defined in Equation (2.3). We will now find the estimators of the coefficient vector $\Phi = (\phi_1, \dots, \phi_p)^T$ and the white noise variance σ^2 based on the observations $x_1, \dots, x_n$. We assume that $\{X_t\}_{t\in\mathbb{Z}}$ can be expressed in the form of Equation (2.10), i.e. $\{X_t\}_{t\in\mathbb{Z}}$ is causal.

Multiply each side of Equation (2.3) for X_{t+1} by X_t, to get

$$X_t X_{t+1} = \sum_{j=1}^{p} \phi_j X_t X_{t-j+1} + X_t Z_{t+1}, \tag{4.1}$$

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

Taking expectations, we have

$$\begin{aligned} E(X_t X_{t+1}) &= \sum_{j=1}^{p} E\left(\phi_j X_t X_{t-j+1}\right) + E\left(X_t Z_{t+1}\right) && (4.2) \\ &= \sum_{j=1}^{p} \phi_j E\left(X_t X_{t-j+1}\right) + E\left(X_t Z_{t+1}\right), && (4.3) \end{aligned}$$

$E\left(X_t Z_{t+1}\right) = 0$ since the random noise of the future time $t+1$ is uncorrelated of X_t due to causality. Hence,

$$\gamma(1) = \sum_{j=1}^{p} \phi_j \gamma(j-1). \tag{4.4}$$

To get the autocovariance at lag 2, multiply each side of Equation (2.3) by X_{t-1}, to get

$$X_{t-1} X_{t+1} = \sum_{j=1}^{p} \phi_j X_{t-1} X_{t-j+1} + X_{t-1} Z_{t+1}, \tag{4.5}$$

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

Taking expectations, we have

$$E\left(X_{t-1}X_{t+1}\right) = \sum_{j=1}^{p} E\left(\phi_j X_{t-1}X_{t-j+1}\right) + E\left(X_{t-1}Z_{t+1}\right) \tag{4.6}$$

$$= \sum_{j=1}^{p} \phi_j E\left(X_{t-1}X_{t-j+1}\right) + E\left(X_{t-1}Z_{t+1}\right) \tag{4.7}$$

$$\gamma(2) = \sum_{j=1}^{p} \phi_j \gamma(j-2). \tag{4.8}$$

Continuing this process, we have the autocovariance at lag p as

$$\gamma(p) = \sum_{j=1}^{p} \phi_j \gamma(j-p). \tag{4.9}$$

Combining all the equations of the autocovariances

$$\begin{aligned}
\gamma(1) &= \phi_1\gamma(0) + \phi_2\gamma(1) + \ldots + \phi_p\gamma(p-1)\\
\gamma(2) &= \phi_1\gamma(1) + \phi_2\gamma(0) + \ldots + \phi_p\gamma(p-2)\\
\vdots &= \vdots \qquad \vdots\\
\gamma(p-1) &= \phi_1\gamma(p-2) + \phi_2\gamma(p-3) + \ldots + \phi_p\gamma(1)\\
\gamma(p) &= \phi_1\gamma(p-1) + \phi_2\gamma(p-2) + \ldots + \phi_p\gamma(0)
\end{aligned}$$

$$\begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(p-1) \\ \gamma(p) \end{bmatrix} = \begin{bmatrix} \gamma(0) & \gamma(1) & \ldots & \gamma(p-2) & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \ldots & \gamma(p-3) & \gamma(p-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma(p-2) & \gamma(p-3) & \ldots & \gamma(0) & \gamma(1) \\ \gamma(p-1) & \gamma(p-2) & \ldots & \gamma(1) & \gamma(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{p-1} \\ \phi_p \end{bmatrix} \tag{4.10}$$

or

$$\gamma_p = \Gamma_p \Phi, \tag{4.11}$$

where $\gamma(h) = \gamma(-h)$, Γ_p is the covariance matrix, $\gamma_p = (\gamma(1), \ldots, \gamma(p))^T$ and $\Phi = (\phi_1, \ldots, \phi_p)^T$. To get the estimators for the white noise variance σ^2, multiply Equation (2.3) by X_t and take expectations of both sides

$$\begin{aligned}
X_t^2 &= \phi_1 X_{t-1}X_t + \phi_2 X_{t-2}X_t + \ldots + \phi_p X_{t-p}X_t + Z_tX_t, \quad t \in \mathbb{Z},\\
E\left(X_t^2\right) &= \phi_1 E\left(X_{t-1}X_t\right) + \phi_2 E\left(X_{t-2}X_t\right) + \ldots + \phi_p E\left(X_{t-p}X_t\right) + E\left(Z_tX_t\right).
\end{aligned}$$

From the causality assumption,

$$\begin{aligned}
E(Z_tX_t) &= E(Z_t \sum_{j=0}^{\infty} \psi_j Z_{t-j})\\
&= \sum_{j=0}^{\infty} \psi_j E(Z_t Z_{t-j}) = \sigma^2.
\end{aligned}$$

Therefore,

$$\begin{aligned}\gamma(0) &= \phi_1\gamma(1) + \phi_2\gamma(2) + \ldots + \phi_p\gamma(p) + \sigma^2 \\ \sigma^2 &= \gamma(0) - \phi_1\gamma(1) - \phi_2\gamma(2) - \ldots - \phi_p\gamma(p)\end{aligned}$$

$$\sigma^2 = \gamma(0) - \Phi^T\gamma_p. \tag{4.12}$$

Equations (4.11) and (4.12) are the Yule-Walker equations which can be used to determine $\gamma(0), \ldots, \gamma(p)$ from σ^2 and Φ.

Replacing the covariances $\gamma(j)$, $j = 0, \ldots, p$ in Equations (4.11) and (4.12) by the corresponding sample covariances

$$\widehat{\gamma}(j) = \frac{1}{n}\sum_{k=1}^{n-h}(x_{k+h} - \overline{x})(x_k - \overline{x}), \quad 0 < h \leq n, \tag{4.13}$$

and

$$\widehat{\gamma}(h) = \widehat{\gamma}(-h), \quad -n < h \leq 0, \tag{4.14}$$

where $\overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$ is the sample mean of the sample $\{x_k\}_{k=0,\ldots,n}$. We obtain a set of equations for the Yule-Walker estimators $\widehat{\Phi}$ and $\widehat{\sigma}^2$, of Φ and σ^2, respectively

$$\widehat{\Gamma}_p\widehat{\Phi} = \gamma_p, \tag{4.15}$$

and

$$\widehat{\sigma}^2 = \widehat{\gamma}(0) - \widehat{\Phi}^T\widehat{\gamma}_p. \tag{4.16}$$

Dividing Equation (4.15) by $\widehat{\gamma}(0)$, we have

$$\widehat{R}_p\widehat{\Phi} = \widehat{\rho}_p, \tag{4.17}$$

where $\widehat{R}_p = \frac{\widehat{\Gamma}_p}{\widehat{\gamma}(0)}$ and $\widehat{R}_p = \widehat{R}_p^T$ since $\widehat{R}_p$ is a symmetric matrix. $\widehat{R}_p$ has a non-zero determinant if $\widehat{\gamma}(0) > 0$ ([BD87], Chapter 5.1). Hence from Equation (4.17),

$$\widehat{\Phi} = \widehat{R}_p^{-1}\widehat{\rho}_p. \tag{4.18}$$

Substituting Equation (4.18) into Equation (4.16), we have

$$\widehat{\sigma}^2 = \widehat{\gamma}(0) - (\widehat{R}_p^{-1}\widehat{\rho}_p)^T\widehat{\gamma}_p = \widehat{\gamma}(0)\left(1 - \widehat{\rho}_p^T\widehat{R}_p^{-1}\widehat{\rho}_p\right). \tag{4.19}$$

From Equation (4.17), we have $1 - \widehat{\phi}_1 z - \ldots - \widehat{\phi}_p z^p \neq 0$ for $|z| < 1$. Therefore, the fitted model

$$X_t - \widehat{\phi}_1 X_{t-1} - \ldots - \widehat{\phi}_p X_{t-p} = Z_t \quad \{Z_t\}, \sim WN(0, \widehat{\sigma}^2) \tag{4.20}$$

is causal. The autocovariances $\gamma_F(h), h = 0, \ldots, p$, of the fitted model therefore satisfy the $p+1$ linear equations

$$\gamma_F(h) - \widehat{\phi}_1\gamma_F(h-1) - \ldots - \widehat{\phi}_p\gamma_F(h-p) = \begin{cases} 0, & h = 1, \ldots, p, \\ \widehat{\sigma}^2, & h = 0. \end{cases} \tag{4.21}$$

From Equations (4.15) and (4.16) we have $\gamma_F(h) = \widehat{\gamma}(h), h = 0, 1, \ldots, p$. This implies that the autocovariances of the fitted model at lags $0, 1, \ldots, p$ coincide with the sample autocovariances.

Theorem 4.1 *If $\{X_t\}$ is the causal AR(p) process defined by Equation (2.3) with $\{Z_t\} \sim IID(0, \sigma^2)$, and if $\widehat{\Phi} = (\widehat{\phi}_1, \ldots, \widehat{\phi}_p)$. Then*

$$\sqrt{n}(\widehat{\Phi} - \Phi) \xrightarrow{d} N(0, \sigma^2 \Gamma_p^{-1}), \quad n \to \infty, \tag{4.22}$$

where $\Gamma_p = (\gamma(i-j))_{i,j=1,\ldots,p}$ is the (unknown) true covariance matrix, $N(0,\sigma)$ is the normal distribution with zero mean and variance σ and $\xrightarrow{d}$ denotes convergence in distribution. Furthermore,

$$\widehat{\sigma^2} \xrightarrow{p} \sigma^2, \quad n \to \infty$$

in probability.

Proof 4.2 *See [BD87], Chapter 8.10.*

Remark 4.3 *We can extend the idea leading to the Yule-Walker equations to general ARMA(p,q) process. However, the estimates for $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma^2$ are not consistent i.e. Theorem 4.1 does not hold anymore. One usually uses the Yule-Walker estimates as starting values in Maximum Likelihood Estimation.*

4.2 The Durbin-Levinson Algorithm

Levinson and Durbin derived an iterative way of solving the Yule-Walker equations (see Section 3.1). Instead of solving (4.11) and (4.12) directly, which involves inversion of $\widehat{R}_p$, the Levinson-Durbin algorithm fits AR models of successively increasing orders AR(1), AR(2), ..., AR(p) to the data. The fitted AR(m) process is then given by

$$X_t - \widehat{\phi}_{m1} X_{t-1} - \ldots - \widehat{\phi}_{mm} X_{t-m}, \quad \{Z_t\} \sim WN(0, \widehat{\sigma}_m), \tag{4.23}$$

where from Equations (4.18) and (4.19)

$$\widehat{\Phi}_m = (\widehat{\phi}_{m1}, \ldots, \phi_{mm}) = \widehat{R}_p^{-1} \widehat{\rho}_m, \tag{4.24}$$

and

$$\widehat{\sigma}_m = \widehat{\gamma}(0) \left(1 - \widehat{\rho}_m^T \widehat{R}_m^{-1} \widehat{\rho}_m\right). \tag{4.25}$$

Proposition 4.4 *If $\widehat{\gamma}(0) > 0$ then the fitted AR(m) model in Equation (4.24) for $m = 1, 2, \ldots, p$ is recursively calculated from the relations*

$$\widehat{\phi}_{mm} = \frac{1}{\widehat{\sigma}_{m-1}} \left[\widehat{\gamma}(m) - \sum_{j=1}^{m-1} \widehat{\phi}_{m-1,j} \widehat{\gamma}(m-j) \right], \tag{4.26}$$

$$\begin{bmatrix} \widehat{\phi}_{m1} \\ \vdots \\ \widehat{\phi}_{m,m-1} \end{bmatrix} = \widehat{\Phi}_{m-1} - \widehat{\phi}_{mm} \begin{bmatrix} \widehat{\phi}_{m-1,m-1} \\ \vdots \\ \widehat{\phi}_{m-1,1} \end{bmatrix} \tag{4.27}$$

and

$$\widehat{\sigma}_m = \widehat{\sigma}_{m-1}(1 - \widehat{\phi}^2_{mm}), \tag{4.28}$$

with $\widehat{\phi}_{11} = \widehat{\rho}(1)$ *and* $\widehat{\sigma}_1 = \widehat{\gamma}(0)[1 - \widehat{\rho}^2(1)]$.

Definition 4.5 *For* $n \geq 2$, $\alpha(n) = \Phi_{nn}$ *is called the partial autocorrelation function. For* $n = 1$, $\alpha(1) = \rho(1) = \text{cor}(X_t, X_{t+1})$. $\alpha(n)$ *measures the correlation between* X_t *and* X_{t+n} *taking into account the observations* $X_{t+1}, \ldots, X_{t+n-1}$ *lying in between.*

Proposition 4.6 *For an AR(p) model, the partial autocorrelation is zero after lag* p*, i.e.* $\alpha(h) = 0, \forall\ h > p$ *and for a MA(q) model,* $\alpha(h) \neq 0\ \forall\ h$.

Example 4.7 *Given that a time series has sample autocovariances* $\widehat{\gamma}(0) = 1382.2$, $\widehat{\gamma}(1) = 1114.4$, $\widehat{\gamma}(2) = 591.73$, *and* $\widehat{\gamma}(3) = 96.216$ *and sample autocorrelations* $\widehat{\rho}(0) = 1$, $\widehat{\rho}(1) = 0.8062$, $\widehat{\rho}(2) = 0.4281$, *and* $\widehat{\rho}(3) = -0.0.696$, *we use the Durbin-Levinson algorithm to find the parameters* ϕ_1, ϕ_2*, and* σ^2 *in the AR(2) model,*

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t, \quad \{Z_t\} \sim WN(0, \sigma^2), \tag{4.29}$$

where $Y_t = X_t - 46.93$ *is the mean-corrected series. We calculate* ϕ_1, ϕ_2*, and* σ^2 *from Proposition 4.4 as follows:*

$$\begin{aligned}
\widehat{\phi}_{11} &= \widehat{\rho}(1) = 0.8062 \\
\widehat{\sigma}_1 &= \widehat{\gamma}(0)(1 - \widehat{\rho}^2_1) = 1382.2(1 - 0.8062^2) = 483.83 \\
\widehat{\phi}_{22} &= \frac{1}{\widehat{\sigma}_1}(\widehat{\gamma}(2) - \widehat{\phi}_{11}\widehat{\gamma}(1)) = \frac{1}{483.82}(591.73 - 0.8062 \cdot 1114.4) = -0.6339 \\
\widehat{\phi}_{21} &= \widehat{\phi}_{11} - \widehat{\phi}_{22}\widehat{\phi}_{11} = 0.8062 + 0.6339 \cdot 0.8062 = 1.31725 \\
\widehat{\sigma}_2 &= \widehat{\sigma}_1(1 - \widehat{\phi}_{22}) = 483.82(1 - (-0.6339)^2) = 289.4.
\end{aligned}$$

Hence the fitted model is

$$Y_t = 1.31725 Y_{t-1} - 0.6339 Y_{t-2} + Z_t, \quad \{Z_t\} \sim WN(0, 289.4)$$

Therefore, the model for the original series $\{X_t\}$ *is*

$$\begin{aligned}
X_t &= 46.93 + 1.31725(X_{t-1} - 46.93) - 0.6339(X_{t-2} - 46.93) + Z_t \\
&= 14.86 + 1.31725 X_{t-1} - 0.6339 X_{t-2} + Z_t, \quad \{Z_t\} \sim WN(0, 289.4)
\end{aligned}$$

4.3 The Innovations Algorithm

We fit a moving average model

$$X_t = Z_t + \widehat{\theta}_{m1} Z_{t-1} + \ldots + \widehat{\theta}_{mm} Z_{t-m}, \quad \{Z_t\} \sim WN(0, \widehat{v}_m) \tag{4.30}$$

of orders $m = 1, 2, \ldots q$ by means of the Innovations algorithm just as we fit AR models of orders $1, 2, \ldots, p$ to the data $x_1, \ldots, x_n$ by the Durbin-Levinson algorithm.

Definition 4.8 *If $\widehat{\gamma}(0) > 0$ then the fitted MA(m) model in Equation (4.30) for $m = 1, 2, \ldots, q$, can be determined recursively from the relations*

$$\widehat{\theta}_{m,m-k} = \frac{1}{\widehat{v}_k}\left[\widehat{\gamma}(m-k) - \sum_{j=0}^{k-1}\widehat{\theta}_{m,m-j}\widehat{\theta}_{k,k-j}\widehat{v}_j\right], \quad k = 0, \ldots, q, \tag{4.31}$$

and

$$\widehat{v}_m = \widehat{\gamma}(0) - \sum_{j=0}^{k-1}\widehat{\theta}^2_{m,m-j}\widehat{v}_j \tag{4.32}$$

where $v_0 = \widehat{\gamma}(0)$.

Remark 4.9 *The estimators $\widehat{\theta}_{q1}, \ldots, \widehat{\theta}_{qq}$, obtained by the Innovations algorithm are usually not consistent in contrast to the Yule-Walker estimates. For MA(q) as well as for general ARMA(p,q) process, one therefore uses Maximum Likelihood Estimation (MLE) techniques which we will not introduce in this essay. We refer to [BD87], Chapter 8.7 for details on MLE.*

5. Data Example

Figure 5.1 shows the plot of the values of the SP500 index from January 3, 1995 to October 31, 1995 (Source: Hyndman, R. J. (n.d.) *Time Series Data Library*. Accessed on May 20, 2007). The series $\{X_t\}$ shows an overall upward trend, hence it is non-stationary. Figure 5.2 gives the plot of the autocorrelation and the partial autocorrelation of the series in Figure 5.1. The slow decay of the acf is due to the upward trend in the series. This suggests differencing at lag 1, i.e. we apply the operator $(1 - B)$ (see Chapter 1.3). The differenced series produces a new series shown in Figure 5.3. From the graph, we can see that the series is stationary. Figure 5.4 shows the plot of the autocorrelation function and the partial autocorrelation function of the differenced data which suggests fitting an autoregressive average model of order 2 (see Proposition 4.6).

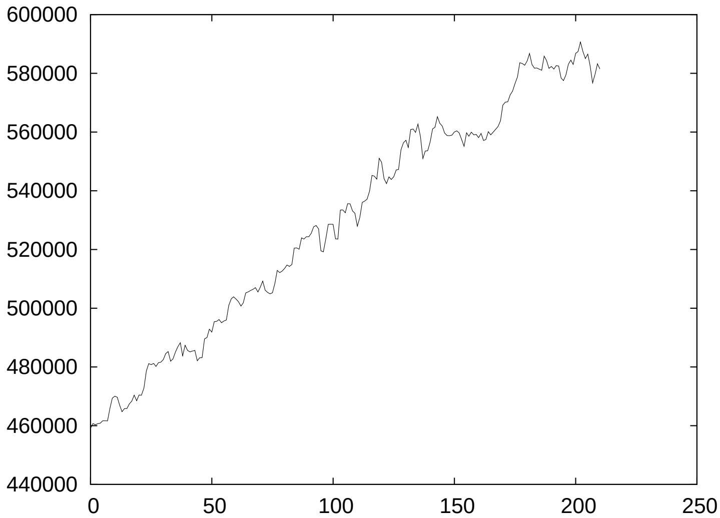


Figure 5.1: The plot of the original series $\{X_t\}$ of the SP500 index

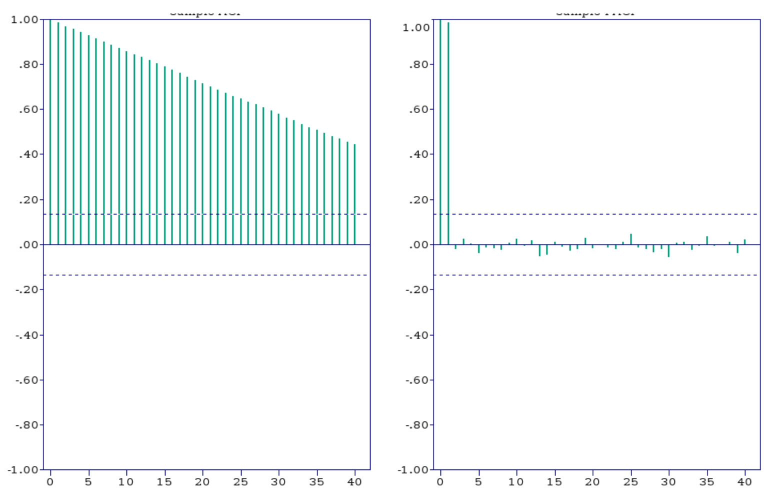


Figure 5.2: The acf and the pacf of Figure 5.1

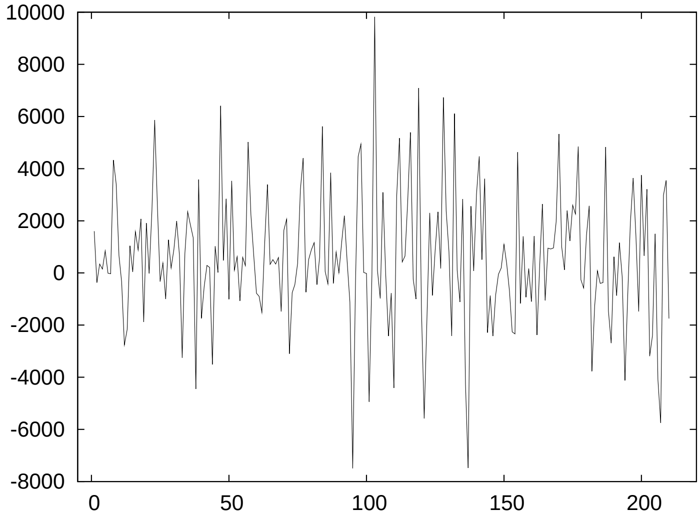


Figure 5.3: The differenced mean corrected series $\{Y_t\}$ in Figure 5.1

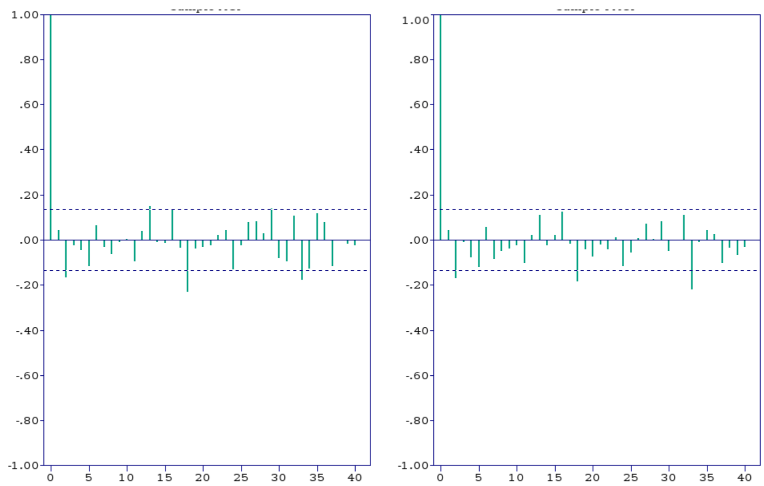


Figure 5.4: The acf and the pacf of Figure 5.3

Using the Yule-walker equations, we fit an AR(2) to the differenced series

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t, \quad t \in \mathbb{Z} \tag{5.1}$$

where

$$Y_t = (1 - B)X_t - 529500, \quad t \in \mathbb{Z}. \tag{5.2}$$

The sample autocovariance of the series in Figure 5.1 are $\widehat{\gamma}(0) = 6461338.3$, $\widehat{\gamma}(1) = 299806.1$, $\widehat{\gamma}(2) = -1080981.9$, and $\widehat{\gamma}(3) = -165410.3$, and the sample autocorrelation are $\widehat{\rho}(0) = 1$, $\widehat{\rho}(1) = 0.0464$, $\widehat{\rho}(2) = -0.1673$, and $\widehat{\rho}(3) = -0.0256$.

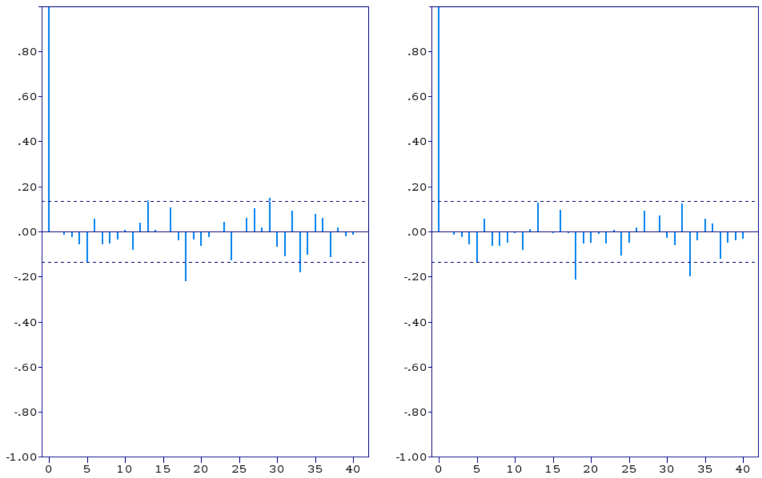


Figure 5.5: The plot of the acf and the pacf of the residuals

From Equation (4.18),

$$\begin{aligned}
\widehat{\Phi} = \widehat{R}_p^{-1}\widehat{\rho}_p &= \begin{bmatrix} 1 & \widehat{\rho}(1) \\ \widehat{\rho}(1) & 1 \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\rho}(1) \\ \widehat{\rho}(2) \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0.0464 \\ 0.0464 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.0464 \\ -0.1673 \end{bmatrix} \\
&= \begin{bmatrix} 1.00215 & -0.0465 \\ -0.0465 & 1.00215 \end{bmatrix} \begin{bmatrix} 0.0464 \\ -0.1673 \end{bmatrix} \\
&= \begin{bmatrix} 0.05428 \\ -0.1698 \end{bmatrix}
\end{aligned}$$

From Equation (4.19),

$$\begin{aligned}
\widehat{\sigma}^2 &= \widehat{\gamma}(0)\left(1 - \widehat{\rho}_p^T \widehat{R}_p^{-1}\widehat{\rho}_p\right) \\
&= 6461338.3\left[1 - (0.0464, -0.1673)(0.05428, -0.1698)^T\right] \\
&= 6261514
\end{aligned}$$

Hence the fitted model is

$$Y_t = 0.05428Y_{t-1} - 0.1698Y_{t-2} + Z_t, \quad \{Z_t\} \sim WN(0, 6261514). \tag{5.3}$$

Figure 5.5 shows the plot of the autocorrelation function of the residuals. From the graph, we observe that the acf and the pacf lie within the confidence interval of 95% which implies that $\{Z_t\}$ is white noise.

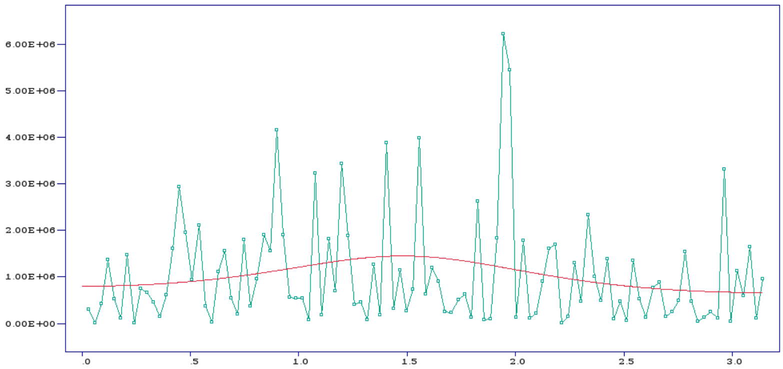


Figure 5.6: The spectral density of the fitted model

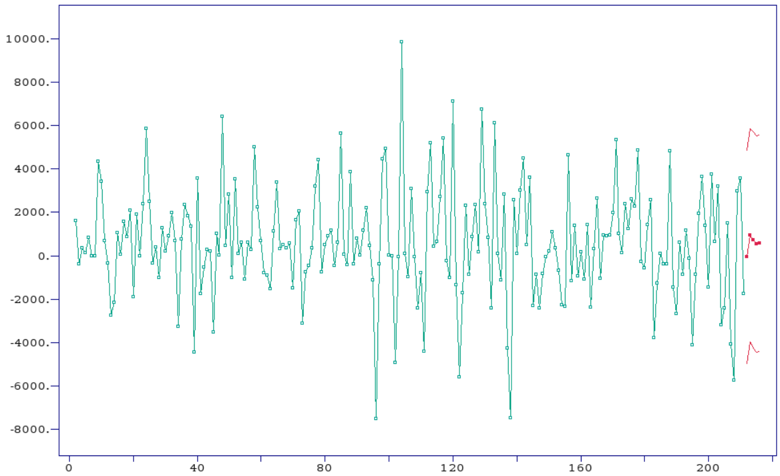


Figure 5.7: The plot of the forecasted values with 95% confidence interval

Figure 5.6 shows the plot of the spectral density of the fitted model. Our model is justified since the spectral density fits in the spectrum of the data.

Figure 5.7 shows the forecast of 5 successive values of the series in Figure 5.3 with a confidence interval of 95%.

6. Conclusion

This essay highlighted the main features of time series analysis. We first introduced the basic ideas of time series analysis and in particular the important concepts of stationary models and the autocovariance function in order to gain insight into the dependence between the observations of the series. Stationary processes play a very important role in the analysis of time series and due to the non-stationary nature of most data, we discussed briefly a method for transforming non-stationary series to stationary ones.

We reviewed the general autoregressive moving average (ARMA) processes, an important class of time series models defined in terms of linear difference equations with constant coefficients. ARMA processes play an important role in modelling time-series data. The linear structure of ARMA processes leads to a very simple theory of linear prediction which we discussed in Chapter 3. We used the observations taken at or before time n to forecast the subsequent behaviour of X_n.

We further discussed the problem of fitting a suitable AR model to an observed discrete time series using the Yule-Walker equations. The major diagnostic tool used is the sample autocorrelation function (acf) discussed in Chapter 1. We illustrated with a data example, the process of fitting a suitable model to the observed series and estimated its paramters. From our estimation, we fitted an autoregressive process of order 2 and with the fitted model, we predicted the next five observations of the series.

This essay leaves room for further study of other methods of forecasting and estimation of parameters. Particularly interesting are the Maximum Likelihood Estimators for ARMA processes.

Appendix A. Programs for Generating the Various Plots

A.1 Code for plotting data

```
from __future__ import division
from Numeric import *
from scipy import *
from scipy.io import *
import Gnuplot
import random
g = Gnuplot.Gnuplot(persist=1)

g('set yrange [-8000:10000]')
g('set xrange [-5:220]')
#g('set xzeroaxis lt 4 lw 2 ')
data=[]
## This command import the data from where it is saved.
co2 = asarray(array_import.read_array("/home/veronica/Desktop/co2.dat"),Int)
aa=[]
sp = asarray(array_import.read_array("/home/veronica/Desktop/sp.txt"),Int)
a=len(co2)
b=len(sp)
for j in sp:
    aa.append(j)


##This code gives the differenced data.
for i in arange(1,len(aa)):
    f=aa[i]-aa[i-1]
    data.append([i,f])


##plot1 = Gnuplot.PlotItems.Data(co2, with = 'lines')
##plot2 = Gnuplot.PlotItems.Data(sp, with = 'lines')
plot3=Gnuplot.PlotItems.Data(data, with = 'lines')
g.plot(plot3)
g.hardcopy(filename = 'name.eps',eps=True, fontsize=20)
```

A.2 Code for plotting the autocorrelation function of MA(q) process

```
from __future__ import division
from Numeric import *
from scipy import *
import Gnuplot
import random
g = Gnuplot.Gnuplot(debug=1)
g('set ylabel "X"' )
g('set xlabel "l"')
g('set xrange [-1:]')
g('set xzeroaxis lt 4 lw 2 ')

## g0 is the value of the autocorrelation function at lag 1.
## Sum(j,theta) gives the value of the autocorrelation function at lag h
def Sum(j,theta):
    h = 0
    if h == j:
        t = array(theta)
        g0 = sum(t**2)
        return g0
    else:
        h = 1
        g1 = 0
        for i in xrange(len(theta)):
            if (i+j) < len(theta):
                g1 = g1 + theta[i]*theta[i+j]
        return g1

data=[]
##input the MA Paramters as a list.
theta = [1,-0.9]
q = len(theta)-1
##This calculate the
for j in xrange(0,5):
    if j <=  q:
        rho = Sum(j,theta)
        if j == 0:
            g0 = rho
    else:
        rho = 0
    rho = rho/g0
    data.append([j,rho])
```

```
plot1 = Gnuplot.PlotItems.Data(data, with = 'impulses')
g.plot(plot1)
#g.hardcopy()
g.hardcopy(filename = 'name.eps',eps=True, fontsize=20)
```

A.3 Code for plotting the partial autocorrelation function of MA(1) process

```
from __future__ import division
from Numeric import *
from scipy import *
import Gnuplot
import random
g = Gnuplot.Gnuplot(debug=1)
##g('set terminal png')
g('set ylabel "X"' )
g('set xlabel "l"')
g('set xrange [-1:]')
g('set xzeroaxis lt 4 lw 2 ')
data=[]
theta=0.9
for k in arange(0,40):
    if k == 0:
        alpha = 1
    else:
        alpha =- (-theta)**k*(1-theta**2)/(1-theta**(2*(k+1)))
    data.append([k,alpha])


plot1 = Gnuplot.PlotItems.Data(data, with = 'impulses')
g.plot(plot1)
g.hardcopy(filename = 'name.eps',eps=True, fontsize=20)
```

A.4 Code for Plotting the Spectral Density Function of MA(1) process

```
from __future__ import division
from Numeric import *
```

```
from scipy import *
import Gnuplot
import random
g = Gnuplot.Gnuplot(persist=1)
g('set ylabel "X"' )
g('set xlabel "lag"')

data=[]
specify the value of theta
theta=
for l in arange(0, pi+pi/13, pi/18):
    f = 1/(2*pi)*(1+theta**2-2*theta*cos(l))
    data.append([l,f])

plot1 = Gnuplot.PlotItems.Data(data, with = 'lines')#, title = 'plot')
g.plot(plot1)
g.hardcopy(filename = 'name.eps',eps=True, fontsize=20)
```

Acknowledgements

To GOD be the GLORY.

I express my profound gratitude to my supervisor, Dr. Tina Marie Marquardt, for her help and providing many suggestions, guidance, comments and supervision at all stages of this essay. I express my indebtedness to Prof. F. K. Allotey, Prof. Francis Benya and staff of the Department of Mathematics, KNUST, for their support and advice.

Many thanks to AIMS staff and tutors for the knowledge I have learned from them. Without the help of my colleagues especially my best friends Mr. Wole Solana, Miss. Victoria Nwosu and Mr. Jonah Emmanuel Ohieku, my wonderful brothers Mr. Henry Amuasi, Mr. Samuel Nartey and Mr. Eric Okyere, my stay at AIMS would not have been a successful one. I am very grateful for their support.

I would not forget to appreciate my family for their help, love and support throughout my education.

GOD BLESS YOU ALL.

Bibliography

[BD87] P. J. Brockwell and R. A. Davis, *Time series: Theory and methods*, Springer-Verlag, 1987.

[BD02] P. J. Brockwell and R. A. Davis, *Introduction to time series and forecasting*, 2nd ed., Spinger-Verlag, New York, 2002.

[Cha01] C. Chatfield, *Time series forecasting*, sixth ed., Chapman and Hall/CRC, New York, 2001.

[Cha04] C. Chatfield, *The analysis of time series: An introduction*, Chapman and Hall/CRC, New York, 2004.