Financial Time Series

Claremont Graduate University
Qidi Peng

Fall 2016

Contents

Chapter 1

Basic Concepts

1.1 Introduction to Financial Time Series

Motivation: Time series is an important tool to model and forecast financial instruments, as well as the features in business and economic environment. Main goal: Introduce financial time series; review the basic probability and statistics concepts.

Definition 1.1.1 (Time series) *A time series is a sequence of random variables*

$$X_1, X_2, \ldots.$$

We denote it by $(X_t)_{t\geq 1}$, where t is called time index.

Note that there is no restriction on the set of time indices. For example $(X_t)_{t\in\mathbb{Z}}$, $(X_t)_{t\in 2\mathbb{N}}$ are also well-defined times series (here $\mathbb{Z}$ denotes the set of integers and $2\mathbb{N}$ denotes the set of positive even numbers). Examples of where one can use time series:

- Daily log returns of Apple stock, from the year 2004 to 2013 (10 years). One can model these values as part of the time series

$$X_1, X_2, \ldots, X_{3650}, \ldots.$$

- Daily Daw Jones Industrial Average Index.
- Quarterly earnings of Facebook company, from 2012 to 2016 (5 years):

$$X_1, X_2, \ldots .X_{20}, \ldots.$$

- Exchange rate between currencies Euro VS US Dollar (daily).

Through all these examples, one can consider using time series to model the values and forecast their future behavior. Also through all these examples, we notice some common features of times series $(X_t)_{t\geq 1}$:

1. X_t is a random variable for each t. This is because financial model has much more uncertainty than mechanics and physics models.

2. X_t and $X_{t'}$ are basically not independent when $t \neq t'$. Because it is the fact in the world of finance that historical instrument values have more or less some impact on future's values. In financial time series, we usually measure this relationship by correlation.

3. In a time series, the time index t goes to infinity. It turns out that to describe the structure of a time series is quite different from describing the structure of a random vector. For the latter structure, it is sufficient to provide joint probability distribution.

Then how to describe the structure of a time series $(X_t)_t$ (here t belongs to some set of indices, like $\mathbb{Z}$, $\mathbb{N}$.)?

Idea 1 Autoregressive models.

It makes sense to believe that, the feature's value at each day t - X_t, can be explained by its own values at previous days:

$$X_t = c + a_1 X_{t-1} + a_2 X_{t-2} + \ldots + a_p X_{t-p} + \underbrace{w_t}_{error\ term\ of\ the\ regression}, \quad (1.1.1)$$

where

- The number of predictors $p \geq 0$.
- The constant c is called trend, the real numbers $a_1, \ldots, a_p$ are coefficients of the regression.
- $(w_t)_t$ is a white noise. Recall that a a white noise $(w_t)_t$ is a time series satisfying

  $$E(w_t) = 0, \ Var(w_t) = \sigma^2, \text{ and } Cov(w_t, w_{t'}) = 0 \text{ for any } t \neq t'.$$

  The model (1.1.1) is called autoregressive model of order p. We denote it by

  $$(X_t)_t \sim AR(p).$$

Idea 2 Moving average models.

It is also natural to believe that the feature's value at each day t - X_t, can be explained by other features at previous days. Since these latter features are unknown, we use a white noise $(w_t)_t$ to model their total impact:

$$X_t = \mu + b_1 w_{t-1} + b_2 w_{t-2} + \ldots + b_q w_{t-q} + \underbrace{w_t}_{\textit{error term of the regression}}, \quad (1.1.2)$$

where

- The constant μ, like c in (1.1.1), is the trend; $b_1, \ldots, b_q$, like $a_1, \ldots, a_p$ in (1.1.1), are the coefficients of the linear regression.
- $(w_t)_t$ denotes a white noise with variance $Var(w_t) = \sigma^2$.

The model (1.1.2) is called moving average of order q and we denote by

$$(X_t)_t \sim MA(q).$$

Idea 3 Autoregressive moving average models.

This idea involves extending $AR(p)$ and $MA(q)$ by combining them. Assume that X_t can be explained by its own historical values and some other unknown factors' historical values, then X_t follows an autoregressive moving average model. More precisely,

$$X_t = \underbrace{c}_{\textit{trend}} + \underbrace{\sum_{i=1}^{p} a_i X_{t-i}}_{\textit{AR term}} + \underbrace{\sum_{j=1}^{q} b_j w_{t-j}}_{\textit{MA term}} + \underbrace{w_t}_{\textit{error term of regression}}. \quad (1.1.3)$$

We denote this model by

$$(X_t)_t \sim ARMA(p, q).$$

Observations: $ARMA(p, 0) \sim AR(p)$, $ARMA(0, q) \sim MA(q)$.
We will study $ARMA(p, q)$ in great details in the following chapters.

1.2 Basic Concepts in Probability

1.2.1 Probability Space and Random Variables

Definition 1.2.1 *A probability space consists of the triple* $(\Omega, \mathcal{F}, P)$.

- Ω *is called sample space. It is the collection of all elementary events of an experiment. Each subset of* Ω *is called an event.*

- $\mathcal{F}$ *is called the set of events. It is the collection of all subsets of* Ω.

- P *is called the probability measure. It is in fact a function* $\mathcal{F} \to [0,1]$.

Examples of probability space $(\Omega, \mathcal{F}, P)$:

1. $\Omega = \{0,1\}$; $\mathcal{F} = \{\{0\},\{1\},\emptyset,\Omega\}$;
$$P(\{0\}) = P(\{1\}) = \frac{1}{2}.$$
2. $\Omega = \{\text{“}rainy'', \text{“}sunny''\}$; $\mathcal{F} = \{\{\text{“}rainy''\},\{\text{“}sunny''\},\emptyset,\Omega\}$;
$$P(\{\text{“}rainy''\}) = 0.1,\ \ P(\{\text{“}sunny''\}) = 0.9.$$
3. (This example is out of program) $\Omega = [0,1]$; $\mathcal{F} = \{all\ subsets\ of\ [0,1]\}$; we can define P by: for any $x \in \mathbb{R}$,
$$P((-\infty, x]) = \mu((-\infty, x] \cap [0,1]),$$
where μ is the Lebesgue measure. $\mu(I)$ denotes the length of the interval I (note that $\mu(\emptyset) = 0$).

It is usually inconvenient to describe an event in $\mathcal{F}$, we thus introduce random variables to help to denote events.

Definition 1.2.2 (Random variables) *A random variable* X *is a function* $\Omega \to \mathbb{R}$.

Examples of random variables: note that every random variable X captures a probability space.

1. For the probability space $(\Omega, \mathcal{F}, P)$ such that $\Omega = \{\text{“}rainy'', \text{“}sunny''\}$; $\mathcal{F} = \{\{\text{“}rainy''\},\{\text{“}sunny''\},\emptyset,\Omega\}$;
$$P(\{\text{“}rainy''\}) = 0.1,\ \ P(\{\text{“}sunny''\}) = 0.9,$$
we can define a random variable X by
$$X(\text{“}rainy'') = 0 \text{ and } X(\text{“}sunny'') = 1.$$
Then $P(\{\text{“}rainy''\})$ and $P(\{\text{“}sunny''\})$ have their shortcut notations:
$$P(X = 0) = 0.1,\ \ P(X = 1) = 0.9.$$
2. (This example is out of program) For the probability space $(\Omega, \mathcal{F}, P)$ such that $\Omega = [0,1]$; $\mathcal{F} = \{all\ subsets\ of\ [0,1]\}$; we can define a random variable X such that
$$X(u) = u, \text{ for } u \in [0,1].$$
Then $P([a,b]) = P(a \le X \le b)$ and $P((-\infty, x]) = P(X \le x)$.

From now on, we always use a random variable to denote outcome of an experiment. For instance $X =$ underlying asset price, or $X =$ exchange rate.
Remark: The most important random variable in financial time series is normal random variable (or equivalently Gaussian random variable).

1.2.2 Moments of Random Variable

In financial times series, we mainly study the first order moment (expected value) and the second order moments (variance, covariance) of random variable.

Definition 1.2.3 (Expected value or equivalently, expectation) *The expected value of a random variable X denotes its mean value and is defined by*

$$E(X) = \begin{cases} \sum\limits_{all\ possible\ k} kP(X=k) & \text{if } X \text{ is discrete;} \\ \int_{-\infty}^{+\infty} xf(x)dx & \text{if } X \text{ is continuous having density } f. \end{cases}$$

Definition 1.2.4 (Variance, covariance and correlation) *The variance of a random variable X measures the dispersion of the realizations of X to the center $E(X)$:*

$$Var(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2.$$

The covariance of two random variables X, Y measure the linearity between X and Y:

$$Cov(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y).$$

The covariance may take any real values, then it is not convenient to compare the linearity of two pairs of random variables through covariance function. To overcome this we rely more on correlation, which plays the same role as covariance but takes values in $[-1, 1]$:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}.$$

We also denote by $\sigma(X) = \sqrt{Var(X)}$ the standard deviation of X.

Remark that the moments of a random variable do not necessarily exist. However from now on we assume that the first order and second order moments of the time series considered in this course exist!
Here are more properties on the correlation needed to be known:

- When $\rho(X, Y) = 1$, then there exist $a > 0$ and $b \in \mathbb{R}$ such that $Y = aX + b$. We say that X and Y are perfectly positively correlated.
- When $\rho(X, Y) = -1$, then there exist $a < 0$ and $b \in \mathbb{R}$ such that $Y = aX + b$. We say that X and Y are perfectly negatively correlated.
- When $\rho(X, Y) = 0$, then we say that X and Y are uncorrelated or linearly independent. Recall that $\rho(X, Y) = 0$ is equivalent to $E(XY) = E(X)E(Y)$.

- Independence implies linearly independence, but generally speaking, linearly independence does not implies independence. Recall that X, Y are independent if and only if $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for any events A, B. In this course, we are only interested in linear dependency.

In terms of calculating the moments of X, the following formula are useful:

- The expected value $E(\cdot)$ is linear:

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i),$$

  for any real numbers $a_1, \ldots, a_n$ and any random variables $X_1, \ldots, X_n$. In particular,

$$E(aX + b) = aE(X) + b.$$

- The variance $Var(\cdot)$ is not linear:

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i) + 2 \sum_{1 \le i < j \le n} a_i a_j Cov(X_i, X_j),$$

  for any real numbers $a_1, \ldots, a_n$ and any random variables $X_1, \ldots, X_n$. In particular,

$$Var(aX + b) = a^2 Var(X).$$

- The covariance function $Cov(\cdot, \cdot)$ is bilinear:

$$\begin{aligned} &Cov(a_1X_1 + a_2X_2, b_1Y_1 + b_2Y_2) \\ &= a_1b_1Cov(X_1, Y_1) + a_1b_2Cov(X_1, Y_2) \\ &\quad + a_2b_1Cov(X_2, Y_1) + a_2b_2Cov(X_2, Y_2), \end{aligned}$$

  for any real numbers a_1, a_2, b_1, b_2 and any random variables X_1, X_2, Y_1, Y_2. To be more general, we have

$$Cov\left(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j\right) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j Cov(X_i, Y_j).$$

1.3 Stationarity of Financial Time Series

There is a special class of time series: second order stationary time series. Most of the study on time series are based on second order stationary time series.

Definition 1.3.1 (Stationarity) *Consider a time series $(X_t)_t$.*

- *Stationarity in the mean. If for all t,*

$$E(X_t) = c(constant),$$

then $(X_t)_t$ is said to be stationary in the mean.

- *If $(X_t)_t$ is stationary in the mean and moreover, for any fixed k,*

$$Cov(X_t, X_{t+k}) \text{ remains invariant for all } t,$$

equivalently,

$$\rho(X_t, X_{t+k}) = \rho_k, \text{ which does not depend on } t,$$

then $(X_t)_t$ is called a second order stationary time series and the function ρ_k is called autocorrelation function (in terms of k).

Another way of defining a second order stationary time series $(X_t)_t$ is that for any fixed k, the random vector

$$(X_t, X_{t+1}, \dots, X_{t+k})$$

have the same means and covariance matrix for all t. A direct consequence of this second order stationary property is:

$$E(X_t) = constant \text{ and } Var(X_t) = constant.$$

1.4 Estimation of the Moments

Unfortunately, the first and second moments of a time series $(X_t)_t$ are not straightforwardly observed in the real data world. One only observes a finite number of values of $X_1, X_2, \dots, X_n$ (for example, the feature values at day 1 to day n). These values are called a sample. In statistics, we estimate the moments starting from an observed sample.
Given a second order stationary time series $(X_t)_t$. Assume a sample $(X_1, X_2, \dots, X_n)$ is available.

- The sample mean is an estimate of the expected value $E(X_t)$:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}.$$

  This estimate is unbiased, i.e. $E(\overline{X}) = E(X_t)$ for all t.

- The sample variance is an estimate of the variance $Var(X)$:

$$S^2(X) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}.$$

This estimate is unbiased, i.e. $E(S^2(X)) = Var(X_t)$ for all t.

- The sample autocovariance of X_t, X_{t+k} is an estimate of the covariance $Cov(X_t, X_{t+k})$:

$$c_k = \frac{1}{n}\sum_{t=1}^{n-k}(X_t - \overline{X})(X_{t+k} - \overline{X}).$$

Note that c_k is not unbiased estimate of $Cov(X_t, X_{t+k})$ since $E(c_k) \neq Cov(X_t, X_{t+k})$, however c_k converges to $Cov(X_t, X_{t+k})$ as n goes to infinity:

$$c_k \xrightarrow[n\to\infty]{in\ probability} Cov(X_t, X_{t+k}).$$

In this case we say that the sample autocovariance c_k is a consistent estimate of $Cov(X_t, X_{t+k})$.

Also observe that c_0 estimates $Var(X_t)$ then we define the sample autocorrelation r_k by

$$r_k = \frac{c_k}{c_0},$$

which is a consistent estimate of the autocorrelation ρ_k:

$$r_k \xrightarrow[n\to\infty]{in\ probability} \rho_k.$$

Chapter 2

Basic Stochastic Models

Motivation: We introduce decomposition of a time series and some basic adequate models for financial time series. They are often used as a standard against which the performance of more complicated models can be assessed.
Main goal: Know the concepts of backward shift operator and correlogram, know the correlation structures, fitting approaches, simulations and/or predictions of white noise, random walks and $AR(p)$.

2.1 Decomposition of a Time Series

A general time series $(X_t)_t$ can be usually decomposed to 3 parts: trend, seasonal effect and error term. There are 2 simple decomposition models: additive model and multiplicative model.

Definition 2.1.1 (Additive decomposition model) *A simple additive decomposition model is given by*

$$X_t = m_t + s_t + z_t, \tag{2.1.1}$$

where, at time t,

- *X_t is the observed feature value (a random variable).*
- *m_t is the trend (a real number).*
- *s_t is a seasonal effect (another time series). Note that the seasonal effect is not necessarily periodic! It has a sense much weaker than a periodic term.*
- *z_t is the error item (often denoted by a white noise).*

For example,

1. In $AR(p)$,

$$X_t = \underbrace{c}_{stands\ for\ m_t} + \underbrace{(a_1 X_{t-1} + \ldots + a_p X_{t-p})}_{stands\ for\ s_t} + \underbrace{w_t}_{stands\ for\ z_t}.$$

2. In $MA(q)$,

$$X_t = \underbrace{\mu}_{stands\ for\ m_t} + \underbrace{(b_1 w_{t-1} + \ldots + b_q w_{t-q})}_{stands\ for\ s_t} + \underbrace{w_t}_{stands\ for\ z_t}.$$

Observe that most financial series of prices are not additive decomposition model, because their seasonal effects tend to increase as the trend increases. It is then more appropriate to think about multiplicative decomposition models. For example the short term interest rate and European asset values follow a multiplicative decomposition model.

Definition 2.1.2 (Multiplicative decomposition model) $(X_t)_t$ *is called a multiplicative decomposition model if it can be written as*

$$X_t = m_t s_t + z_t,$$

where m_t, s_t, z_t *are respectively trend, seasonal effect and error term.*

Remark: A multiplicative model can be simply transformed to an additive model, through the logarithm transformation: if the series $(X_t)_t$ is positive and if

$$X_t = m_t s_t + z_t,$$

then

$$\log(X_t) = m'_t + s'_t + z'_t,$$

for some other trend m'_t, s'_t, z'_t. This is the reason why we often take log returns in quantitative finance.

Exercise: A stock prices is given as

$$X_t = e^{\mu_t + \sigma Z_t} \text{ at day } t,$$

where μ_t is a deterministic function and $(Z_t)_t$ is a sequence of normal random variables. Determine the trend and seasonal effect of X_t.

2.2 White Noise

Motivation: In a general regression model

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p + \epsilon,$$

the error term ϵ is called *residual*. In a time series regression model, the residual ϵ_t becomes time dependent, and we often use a white noise to fit $(\epsilon_t)_t$, i.e., $(\epsilon_t)_t$ is replaced by $(w_t)_t$ in time series.
Where to apply white noise in finance: fitting residuals of regressions, moving average models, fitting time independent financial series (lottery, some short term security prices,etc.).

Definition 2.2.1 (White noise) *A white noise $(w_t)_t$ is a time series such that*

- $E(w_t) = 0$ *for all* t.
- $Cov(w_t, w_{t'}) = \begin{cases} \sigma^2 & \textit{if } t = t'; \\ 0 & \textit{if } t \neq t'. \end{cases}$

It is worth noting that:

- Definition 2.2.1 contains no message on the probability distribution of w_t, so a white noise can have any probability distribution (with existing variance)! In particular, $(w_t)_t$ is called a *Gaussian white noise* if w_t is normally distributed:

$$w_t \sim \mathcal{N}(0, \sigma^2), \text{ for each } t.$$

- By definition, any two members w_t, $w_{t'}$ of a white noise are generally uncorrelated, not independent! But they are independent if the white noise is a Gaussian white noise, because "uncorrelated components of a **Gaussian vector** are in fact independent random variables"[1].

Second-order properties of white noise $(w_t)_t$:

- The expected value $\mu_w = 0$.
- The autocovariance function

$$\gamma_k := Cov(w_t, w_{t+k}) = \begin{cases} \sigma^2 & \text{if } k = 0; \\ 0 & \text{if } k \neq 0. \end{cases}$$

- The autocorrelation function

$$\rho_k := \rho(w_t, w_{t+k}) = \begin{cases} 1 & \text{if } k = 0; \\ 0 & \text{if } k \neq 0. \end{cases}$$

[1]This result can be found in almost all probability textbooks for advanced courses. I suggest to refer to "Probability and Measure", by Patrick Billingsley.

2.3 Random Walk

Motivation: Random walk is a simple time series, defined as the sum of infinite number of members of a white noise at time t. An investment philosophy believes that security prices are completely unpredictable, especially in the short term. Random walk theory states that both fundamental analysis and technical analysis are wastes of time, as securities behave randomly. Thus, the theory holds that it is impossible to outperform the market by choosing the "correct" securities; it is only possible to outperform the market by taking on additional risk. Critics of random walk theory contend that empirical evidence shows that security prices do indeed follow particular trends that can be predicted with a fair degree of accuracy. The theory originated in 1973 with the book - "A Random Walk Down Wall Street". A random walk model is different from $ARMA(p, q)$, but these 2 models are both generalized by *autoregressive integrated moving average models* $ARIMA(p, d, q)$, which will be introduced later.
Where to fit random walks: short term exchange rate, short term security prices.

2.3.1 Definition and Properties of Random Walk

Definition 2.3.1 (Random walk) *Let* $(X_t)_t$ *be a time series.* $(X_t)_t$ *is a random walk if*

$$X_t = X_{t-1} + w_t, \tag{2.3.1}$$

where $(w_t)_t$ *is a white noise (not necessarily Gaussian white noise).*

Remarks:

1. From equation (2.3.1) we can equivalently write

$$X_t = w_t + w_{t-1} + \ldots = \sum_{k=0}^{\infty} w_{t-k}. \tag{2.3.2}$$

   This equation can be explained as: the value of X_t depends on the cumulative impacts of all historical unknown factors.

2. In practice, the above series (2.3.2) will not be infinite but will start at some time $t = 1$:

$$X_t = w_t + \ldots + w_1 = \sum_{k=1}^{t} w_k. \tag{2.3.3}$$

   This is in fact a particular moving average model (with trend $\mu = 0$ and all coefficients equal to 1) of order $t - 1$: $MA(t - 1)$. In other words, a random walk can be *approximated* by a moving average model.

3. The second-order moments of the random walk defined in (2.3.2) **DON'T EXIST**. For instance the variance of X_t in (2.3.2) is given as

$$Var\left(\sum_{k=0}^{\infty} w_{t-k}\right) = \lim_{N\to\infty}\sum_{k=0}^{N-1} Var(w_{t-k}) = \lim_{N\to\infty}\sum_{k=0}^{N-1}\sigma^2 = \lim_{N\to\infty} N\sigma^2 = \infty.$$

Therefore we study the second-order properties of $(X_t)_t$ defined in (2.3.3) instead.

Second-order properties of the random walk $(X_t)_{t\geq 1}$ defined in (2.3.3):

- The expected value

$$\mu_X = E(X_t) = \sum_{k=0}^{t-1}\underbrace{E(w_{t-k})}_{=0} = 0.$$

- The variance of X_t depends on t!

$$Var(X_t) = Var\left(\sum_{k=1}^{t} X_k\right) = \sum_{k=1}^{t} Var(X_k) = \sum_{k=1}^{t}\sigma^2 = t\sigma^2.$$

- The random walk $(X_t)_{t\geq 1}$ is NOT second-order stationary, because its autocovariance function is time-dependent: for $k \geq -(t-1)$,

$$\begin{aligned}
\gamma_k(t) &:= Cov(X_t, X_{t+k}) = Cov\left(\sum_{i=1}^{t} w_i, \sum_{j=1}^{t+k} w_j\right) \\
&= \sum_{i=1}^{t}\sum_{j=1}^{t+k} Cov(w_i, w_j) = \sum_{i=j, 1\leq i\leq t, 1\leq j\leq t+k} \sigma^2 \\
&= \begin{cases} t\sigma^2 & \text{if } k \geq 0; \\ (t+k)\sigma^2 & \text{if } k < 0. \end{cases}
\end{aligned}$$

- The autocorrelation function is accordingly time-dependent: for $k \geq -(t-1)$,

$$\begin{aligned}
\rho_k(t) &:= \frac{Cov(X_t, X_{t+k})}{\sqrt{Var(X_t)Var(X_{t+k})}} = \frac{Cov(X_t, X_{t+k})}{\sqrt{t(t+k)}\sigma^2} \\
&= \begin{cases} \sqrt{\frac{t}{t+k}} & \text{if } k \geq 0; \\ \sqrt{\frac{t+k}{t}} & \text{if } k < 0. \end{cases}
\end{aligned}$$

2.3.2 Representation of Random Walks via Backward Shift Operators

In time series analysis, there is a simple and elegant way to represent a regressive time series. It is through the so called "backward shift operators" B. Using B, most of the classical time series $(X_t)_t$ can be represented under the form

$$P(B)X_t = w_t,$$

where P is some polynomial.

Definition 2.3.2 (Backward shift operator) *Let* $(X_t)_t$ *be a time series. The* backward shift operator B *is defined by*

$$BX_t = X_{t-1} \text{ for all } t.$$

Properties of B:

- B is linear:
$$B(aX_t + bY_t) = aBX_t + bBY_t = aX_{t-1} + bY_{t-1},$$
for all real numbers a, b and for all time series $(X_t)_t$ and $(Y_t)_t$.

- The n-th composition of B is called *backward shift operator of lag* n. For any $n \geq 1$,
$$B^n X_t = \underbrace{B \dots B}_{n \text{ times}} X_t = X_{t-n}.$$
As a convention,
$$B^0 X_t = X_t.$$
One can then denote $B^0 = 1$.

- For any $m, n \geq 0$,
$$(aB^m + bB^n)X_t = aB^m X_t + bB^n X_t = aX_{t-m} + bX_{t-n}.$$

Representation of random walk $(X_t)_t$:

- For $(X_t)_t$ defined in (2.3.2), we have

$$\begin{aligned} X_t &= w_t + w_{t-1} + \dots = w_t + Bw_t + B^2 w_t + \dots \\ &= (1 + B + B^2 + \dots)w_t = \left(\frac{1}{1-B}\right) w_t. \end{aligned}$$

- For $(X_t)_t$ defined in (2.3.3), we have

$$\begin{aligned} X_t &= w_t + w_{t-1} + \dots + w_1 = w_t + Bw_t + B^2 w_t + \dots + B^{t-1} w_t \\ &= (1 + B + B^2 + \dots + B^{t-1})w_t = \left(\frac{1-B^t}{1-B}\right) w_t. \end{aligned}$$

2.4 Autoregressive Models

2.4.1 $AR(p)$ without Drift

In this subsection we only study $AR(p)$ without drift (trend) in detail. The stationarity and second-order properties are derived.

Definition 2.4.1 (Autoregressive models without drift) $(X_t)_t$ *is called an autoregressive model without drift (or trend) of order p, if*

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \ldots + a_p X_{t-p} + w_t. \tag{2.4.1}$$

Backward shift operator representation of $AR(p)$: From (2.4.1) we see

$$X_t - a_1 X_{t-1} - a_2 X_{t-2} - \ldots - a_p X_{t-p} = w_t.$$

Equivalently,

$$(1 - a_1 B - a_2 B^2 - \ldots - a_p B^p) X_t = w_t. \tag{2.4.2}$$

We call

$$\theta_p(B) = 1 - a_1 B - a_2 B^2 - \ldots - a_p B^p$$

the polynomial (of order p) of backward shift operator representing $AR(p)$.
Remarks:

- The random walk is the special case of $AR(1)$ with $a_1 = 1$.
- For each fixed time t, (2.4.1) is nothing else than a linear regression of p regressors (factors, predictors,...). Since each variable X_t is regressed on its past values, it is then called "autoregressive".
- As all other linear regression model, the parameters $a_1, \ldots, a_p, \sigma^2$ can be estimated by using least squares method.

2.4.2 Second Order Property of $AR(1)$

An $AR(1)$ can be written as

$$(1 - aB) X_t = w_t.$$

Equivalently,

$$X_t = (1 - aB)^{-1} w_t = \sum_{k=0}^{\infty} (aB)^k w_t = \sum_{k=0}^{\infty} a^k B^k w_t = \sum_{k=0}^{\infty} a^k w_{t-k}.$$

- Expectation of $AR(1)$: by the linearity of expectation (and the dominating convergence theorem, to allow $E\sum_{k=0}^{\infty}(\cdot) = \sum_{k=0}^{\infty} E(\cdot)$. No problem if you don't know this.),

$$E(X_t) = E\left(\sum_{k=0}^{\infty} a^k w_{t-k}\right) = \sum_{k=0}^{\infty} E(a^k w_{t-k}) = 0.$$

- Variance of $AR(1)$: by the linear independency of white noises (and again the dominating convergence theorem to allow $Var\sum_{k=0}^{\infty}(\cdot) = \sum_{k=0}^{\infty} Var(\cdot)$.)

$$Var(X_t) = Var\left(\sum_{k=0}^{\infty} a^k w_{t-k}\right) = \sum_{k=0}^{\infty} a^{2k} Var(\epsilon_{t-k}) = \sigma^2 \sum_{k=0}^{\infty} a^{2k}.$$

Note that the geometric series $\sum_{k=0}^{\infty} a^{2k}$ is convergent to $\frac{1}{1-a^2}$ **IF AND ONLY IF** $|a| < 1$! Hence,

$$Var(X_t) = \begin{cases} \frac{\sigma^2}{1-a^2} & \text{if } |a| < 1; \\ \infty & \text{if else.} \end{cases}$$

- Autocovariance and autocorrelation of $AR(1)$: knowing that the covariance function exists only when $|a| < 1$, we then assume it and go on. Let $|a| < 1$, then

$$\begin{aligned} \gamma_k &:= Cov(X_t, X_{t+k}) = Cov\left(\sum_{i=0}^{\infty} a^i w_{t-i}, \sum_{j=0}^{\infty} a^j w_{t+k-j}\right) \\ &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} a^{i+j} Cov(\epsilon_{t-i}, \epsilon_{t+k-j}) \\ &= \sum_{j-i=k, i,j\geq 0}^{\infty} a^{i+j}\sigma^2, \end{aligned}$$

because

$$Cov(\epsilon_{t-i}, \epsilon_{t+k-j}) = \sigma^2 \text{ if and only if } t-i = t+k-j \Longleftrightarrow j-i = k.$$

Case 1 If $k \geq 0$, then $j = i+k \geq 0$, we have

$$\gamma_k = \sigma^2 \sum_{j=i+k, i\geq 0}^{\infty} a^{i+j} = \sigma^2 \sum_{i=0}^{\infty} a^{2i+k} = \frac{\sigma^2 a^k}{1-a^2}.$$

Case 2 If $k < 0$, then $i = j - k \geq 0$, we have

$$\gamma_k = \sigma^2 \sum_{i=j-k, j\geq 0}^{\infty} a^{i+j} = \sigma^2 \sum_{j=0}^{\infty} a^{2j-k} = \frac{\sigma^2 a^{-k}}{1-a^2}.$$

We conclude that

$$\gamma_k = \frac{\sigma^2 a^{|k|}}{1-a^2}.$$

Accordingly, the autocorrelation

$$\rho_k = \frac{\gamma_k}{\sqrt{Var(X_t)Var(X_{t+k})}} = \frac{\frac{\sigma^2 a^{|k|}}{1-a^2}}{\frac{\sigma^2}{1-a^2}} = a^{|k|}.$$

Observations:

- When $|a| < 1$, $AR(1)$ is stationary, because γ_k does not depend on t.
- The fact that $|a| < 1$ is equivalent to that the equation

$$1 - aB = 0$$

has root $|B| > 1$. Then we can say, $AR(1)$ is stationary when the polynomial $1 - aB$ has root strictly greater than unity in absolute value (out of the unit disc).

- More generally, the following result holds:

Theorem 2.4.1 *An $AR(p)$*

$$(1 - a_1 B - a_2 B^2 - \ldots - a_p B^p)X_t = w_t$$

is stationary if and only if ALL the roots of the polynomial

$$\theta_p(B) = 1 - a_1 B - a_2 B^2 - \ldots - a_p B^p$$

are strictly greater than unity, i.e., $|B| > 1$ for all B satisfying

$$1 - a_1 B - a_2 B^2 - \ldots - a_p B^p = 0.$$

The polynomial $\theta_p(B)$ is called ***characteristic function of*** *$AR(p)$.*

It may happen that some root is complex valued, saying $z = a + bi$, then $|z| = \sqrt{a^2 + b^2}$.

A simple proof of the theorem: Let $r_1, \dots, r_p$ (can be complex valued) be the p roots of the polynomial $\theta_p(B)$, i.e.,

$$1 - a_1 B - a_2 B^2 - \dots - a_p B^p = (1 - r_1^{-1}B)(1 - r_2^{-1}B)\dots(1 - r_p^{-1}B).$$

Then we have

$$X_t = \frac{1}{(1 - r_1^{-1}B)(1 - r_2^{-1}B)\dots(1 - r_p^{-1}B)} w_t.$$

If $|r_p| \le 1$, then

$$\frac{1}{1 - r_p^{-1}B} w_t = \sum_{k=0}^{\infty} r_p^{-k} B^k w_t = \sum_{k=0}^{\infty} r_p^{-k} w_{t-k}$$

has variance

$$\sigma^2 \sum_{k=0}^{\infty} (|r_p|^{-2})^k = \infty, \text{ because } |r_p|^{-2} \ge 1.$$

As a result, the corresponding $AR(p)$ is not stationary. No root $|r| \le 1$ is allowed, for the $AR(p)$ to be stationary.
Examples:

1. A white noise $X_t = w_t$ is stationary because the equation

$$1 = 0$$

has no roots $|B| \le 1$.

2. A random walk $X_t = X_{t-1} + w_t$ is non-stationary because the characteristic function

$$1 - B = 0$$

has root $B = 1$, which is in the unit disc.

3. The $AR(1)$ model $X_t = \frac{1}{2}X_{t-1} + w_t$ is stationary because the characteristic function $1 - \frac{1}{2}B = 0$ has single root $B = 2$, which is greater than 1 in absolute value.

4. The $AR(2)$ model $X_t = X_{t-1} - \frac{1}{4}X_{t-2} + w_t$ is stationary, because the characteristic function

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

has root $B = 2$, which is greater than 1 in absolute value.

5. The $AR(2)$ model $X_t = \frac{1}{2}X_{t-1} + \frac{1}{2}X_{t-2} + w_t$ is non-stationary because the roots of
$$1 - \frac{1}{2}B - \frac{1}{2}B^2 = -\frac{1}{2}(B-1)(B+2) = 0$$
are $B = -2$ and $B = 1$, where $B = 1$ is not exceeding unity!

6. The $AR(2)$ model $X_t = -\frac{1}{4}X_{t-2} + w_t$ is stationary because
$$1 + \frac{1}{4}B^2 = 0$$
has 2 roots $B = 2i$, $B = -2i$, both satisfying $|B| = 2 > 1$.

Note that the R function *polyroot* finds roots of polynomials so that it can be used to determine whether an $AR(p)$ is stationary.

2.4.3 Second Order Property of $AR(p)$

In general, we have for an $AR(p)$ model $\theta_p(B)X_t = c + w_t$ with c being the trend (a real number),
$$E(X_t) = \frac{c}{\theta_p(1)} \textit{ and } Var(X_t) = \theta_p(B)^{-1}Var(w_t),$$
where we denote by $aB^kVar(w_t) = a^2Var(w_{t-k}) = a^2$.
Example: mean and variance of $AR(2)$. Consider an $AR(2)$ model
$$(1 - a_1B - a_2B^2)X_t = c + w_t.$$
We have

- the expectation of X_t
$$E(X_t) = \frac{c}{1 - a_1 - a_2}.$$
- the variance of X_t
$$Var(X_t) = \frac{1}{1 - a_1B - a_2B^2}Var(w_t) = \frac{\sigma^2}{1 - a_1^2\left(\frac{1+a_2}{1-a_2}\right) - a_2^2}.$$
For $AR(2)$ to be stationary we necessarily have
$$a_1 + a_2 \neq 1 \text{ and } a_1^2\left(\frac{1+a_2}{1-a_2}\right) + a_2^2 < 1.$$
These conditions are equivalent to the 2 roots of
$$1 - a_1B - a_2B^2 = 0,$$
given by
$$B_{1\cdot 2} = \frac{-a_1 \pm \sqrt{a_1^2 + 4a_2}}{2a_2}$$
satisfy $|B_1| > 1$ and $|B_2| > 1$.

Chapter 3

Regression

Motivation: Trends and seasonal variations of times series can be deterministic or stochastic. In this chapter we provide a popular approach to fit deterministic and seasonal variations - regression.
Main goal: Know some theoretical approaches to fit deterministic trends and seasonal variations; know how to fit and simulate deterministic trends and seasonal variations in *R*.

3.1 Linear Models

Recall that an additive decomposition of a time series involves representing a time series (X_t) by

$$X_t = m_t + s_t + z_t,$$

where

- m_t is the trend;
- s_t is the seasonal effect;
- z_t is the error term.

It is often believed that, m_t is a deterministic function, s_t is a stochastic process and z_t is a white noise. However, real world's models are much more complicated than the above one. In other words, the 3 patterns (m_t, s_t, z_t) of a decomposition could all be stochastic processes, and there might be no clear boundary among them. In this chapter, we mainly study how to fit the trend m_t, using linear regression models. We denote all the rest part (i.e. $s_t + z_t$) by z_t, which is not necessarily a white noise any more.

Definition 3.1.1 (Linear model) *A model for a time series $(X_t)_t$ is called* ***linear model*** *if it can be expressed as*

$$X_t = a_0 + a_1 u_{1,t} + a_2 u_{2,t} + \ldots + a_m u_{m,t} + z_t, \tag{3.1.1}$$

where

- $u_{i,t}$ *denotes the* i*th predictor of the linear regression. Through this chapter we assume* $u_{i,t}$ *are deterministic values (not random variables).*
- a_i *is the coefficient (or weight) of the* i*th predictor.*
- z_t *denotes the error at time* t*. It is also called residual of the regression.*

Examples:

1. The straight line model $X_t = a_0 + a_1 t + w_t$ is a linear model with $u_{1,t} = t$ and $a_2 = \ldots = a_m = 0$, $z_t = w_t$.
2. The polynomial function model $X_t = a_0 + a_1 t + a_2 t^2 + \ldots + a_m t^m + w_t$ is a linear model with $u_{i,t} = t^i$ for $i = 1, \ldots, m$ and $z_t = w_t$.

Remarks:

1. Many non-linear models can be transformed to a linear model. For example, the exponentials (often used to fit interest rates)

$$X_t = e^{a_0 + a_1 t + w_t}$$

   is equivalent to the linear model below, after a logarithmic transformation:

$$\log(X_t) = a_0 + a_1 t + w_t.$$

2. Generally speaking, a linear model is not stationary because its trend is often time-dependent. However, its differenced series is "more stationary" than the original process. For example the linear model

$$X_t = a_0 + a_1 t + w_t$$

   is not stationary, because its expectation

$$E(X_t) = a_0 + a_1 t$$

   depends on time t. However the differenced series of (X_t) is given by

$$Y_t = X_t - X_{t-1} = a_1 + w_t - w_{t-1}$$

   which is stationary, because it is a moving average model $MA(1)$.

3. To explain the deep fact that "The differenced series is more stationary than the original series", we take another example: consider the linear model driven by a polynomial function trend

$$X_t = a_0 + a_1 t + a_2 t^2 + w_t.$$

- **The first order differenced series** of (X_t) is given by

  $$\nabla X_t := X_t - X_{t-1} = a_1 + a_2(t^2 - (t-1)^2) + w_t - w_{t-1} = (a_1 - a_2) + 2a_2 t + w_t - w_{t-1}.$$

  ∇X_t is not stationary because its expectation is time-dependent.
- **The second order differenced series** of (X_t) is given by

  $$\begin{aligned}\nabla^2 X_t &:= \nabla X_t - \nabla X_{t-1} = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) \\ &= (2a_2 t + w_t - w_{t-1}) - (2a_2(t-1) + w_{t-1} - w_{t-2}) \\ &= 2a_2 + w_t - 2w_{t-1} + w_{t-2} \\ &= 2a_2 + \nabla^2 w_t.\end{aligned}$$

  Observe that $\nabla^2 X_t \sim MA(2)$ so it is stationary.

More generally, we claim that the following is true:

Proposition 3.1.2 *Let the linear model (X_t) expressed as*

$$X_t = a_0 + a_1 t + a_2 t^2 + \ldots + a_m t^m + w_t,$$

then the mth order differenced series of (X_t) is given as

$$\nabla^m X_t = a_m m! + \nabla^m w_t = a_m m! + \sum_{i=0}^{m} \binom{m}{i} (-1)^i w_{t-i} \sim MA(m)$$

and it is stationary.

Proof. Please check the concept "finite difference approach" and the following web for the finite difference of polynomials:
http://www.trans4mind.com/personal_development/mathematics/series/polynomialEquationDifferences.htm

3.2 Fitted Models: Autocorrelation and Estimation of Sample Statistics

3.2.1 Relationship Between Time Series and its Differenced Series

If the first order differenced series (∇X_t) of a time series (X_t) is stationary, then (X_t) can be represented as sum of stationary time series as

$$X_t = (X_t - X_{t-1}) + (X_{t-1} - X_{t-2}) + \ldots = \nabla X_t + \nabla X_{t-1} + \ldots = \sum_{i=0}^{\infty} \nabla X_{t-i}.$$

If the second order differenced series $(\nabla^2 X_t)$ of (X_t) is stationary, then again, (X_t) can be represented as sum of stationary time series as

$$X_t = \sum_{i=0}^{\infty} \left(\sum_{j=0}^{\infty} \nabla^2 X_{t-i-j} \right).$$

In general we have the following statement:

Proposition 3.2.1 *If the mth order differenced series $(\nabla^m X_t)$ of (X_t) is stationary, then (X_t) can be represented as sum of stationary time series as*

$$X_t = \sum_{i_1=0}^{\infty} \dots \sum_{i_m=0}^{\infty} \nabla^m X_{t-i_1-i_2-\dots-i_m}.$$

The above result reveals a deep fact of the nature: any movement (can be huge) is caused by the cumulative effects of plenty of tiny amounts of random factors, which are zero mean (positive and negative influences are of equal chance) and same variance (stationary). The Roman Lucretius's scientific poem "On the Nature of Things" (check Wikipedia for "Brownian motion") has a remarkable description of random walk of dust particles in verses 113 - 140 from Book II. He uses this as a proof of the existence of atoms:

"Observe what happens when sunbeams are admitted into a building and shed light on its shadowy places. You will see a multitude of tiny particles mingling in a multitude of ways... their dancing is an actual indication of underlying movements of matter that are hidden from our sight... It originates with the atoms which move of themselves (i.e., spontaneously). Then those small compound bodies that are least removed from the impetus of the atoms are set in motion by the impact of their invisible blows and in turn cannon against slightly larger bodies. So the movement mounts up from the atoms and gradually emerges to the level of our senses, so that those bodies are in motion that we see in sunbeams, moved by blows that remain invisible."

The motivation of representing a time series by sum of stationary series is for simulation and forecasting, since it is much easier to simulate and forecast a stationary time series than an arbitrary time series.

Example 3.2.2 *Let (X_t) be a linear model driven by the polynomial function:*

$$X_t = a_0 + a_1 t + a_2 t^2 + \dots + a_m t^m + w_t, \ \textit{for } t \geq 1,$$

then (X_t) can be represented as

$$\begin{aligned} X_t &= \sum_{i_1,\ldots,i_m\geq 0, i_1+\ldots+i_m\leq t-1}^{\infty} \nabla^m X_{t-i_1-i_2-\ldots-i_m} \\ &= a_m m! \sum_{k=0}^{t-1} \binom{k}{i_1, i_2, \ldots, i_m} + \sum_{i_1,\ldots,i_m\geq 0, i_1+\ldots+i_m\leq t-1}^{\infty} \nabla^m w_{t-i_1-i_2-\ldots-i_m}, \end{aligned}$$

where

$$\binom{k}{i_1, i_2, \ldots, i_m} := \frac{k!}{i_1! i_2! \ldots i_m!}$$

denotes the multinomial coefficient.

3.3 Properties of Sample Statistics of a Stationary Series

According to the last section, we know that it suffices to estimate the first and second order moments of a stationary series, in order to fit the general linear model (X_t). The main goal of this subsection is thus to derive the following result:

Proposition 3.3.1 *Let (Y_t) be a stationary time series with $E(Y_t) = \mu$ and $Var(Y_t) = \sigma^2$. Observe a trajectory $Y_1, Y_2, \ldots, Y_n$ is observed, then*

(a) *The sample mean $\overline{Y} = \frac{Y_1+\ldots+Y_n}{n}$ is an unbiased estimate of μ. Namely,*

$$E(\overline{Y}) = \mu.$$

(b) *However the sample mean $\overline{Y}$ is not necessarily efficient ("an estimate is efficient" means its variance converges to 0, as $n \to \infty$), because*

$$Var(\overline{Y}) = \frac{\sigma^2}{n}\left(1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\rho_k\right),$$

where ρ_k denotes the autocorrelation function of (Y_t).

Proof. The proof of (a) is straightforward. We now focus on proving (b). Note that $Y_1, Y_2, \ldots, Y_n$ are NOT independent. We write

$$\begin{aligned}
Var(\overline{Y}) &= Var\left(\frac{Y_1 + Y_2 + \ldots + Y_n}{n}\right) \\
&= \frac{1}{n^2} Cov\left(\sum_{i=1}^{n} Y_i, \sum_{j=1}^{n} Y_j\right) \\
&= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} Cov\left(Y_i, Y_j\right) \\
&= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{i-j} \\
&= \frac{1}{n^2} \left(\sum_{i=1}^{n} \gamma_0 + 2 \sum_{1 \le i,j \le n, i>j} \gamma_{i-j}\right) \\
&= \frac{1}{n^2} \left(n\gamma_0 + 2 \sum_{k=1}^{n-1} \sum_{i=k+1}^{n} \gamma_k\right) \\
&= \frac{1}{n^2} \left(n\gamma_0 + 2 \sum_{k=1}^{n-1} (n-k)\gamma_k\right).
\end{aligned}$$

Plug $\gamma_0 = \sigma^2$ and $\gamma_k = \sigma^2 \rho_k$ into the above equation, we obtain

$$Var(\overline{Y}) = \frac{\sigma^2}{n} \left(1 + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \rho_k\right).$$

To show that $\overline{Y}$ is not necessarily efficient, we provide one counter-example: assume $\rho_k = 1$ for all k, then

$$\begin{aligned}
Var(\overline{Y}) &= \frac{\sigma^2}{n} \left(1 + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \rho_k\right) \\
&= \frac{\sigma^2}{n} \left(1 + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right)\right) \\
&= \frac{\sigma^2}{n} \left(1 + 2 \left(n - 1 - \frac{\sum_{k=1}^{n-1} k}{n}\right)\right) \\
&= \sigma^2 \left(\frac{1}{n} + 2 \left(1 - \frac{1}{n} - \frac{n-1}{2n}\right)\right) \xrightarrow[n \to \infty]{} \sigma^2 \neq 0.
\end{aligned}$$

However, when $(Y_t)_t$ is a stationary $AR(1)$, $\overline{Y}$ is efficient, because in this case $\rho_k \sim |a|^k$ with $|a| < 1$.

3.4 Linear Regression - Least Squares Estimates

In this subsection we introduce ordinary least squares (OLS) estimation approach. We take one example to show how this approach works.
Consider a simple linear regression

$$X_t = a_0 + a_1 u_t + z_t,$$

where

- a_0, a_1 are unknown parameters.
- u_t is the predictor explaining the values of X_t, it can be either deterministic or random.
- z_t is the error term with zero mean and $Var(z_t) = \sigma^2$. σ^2 is often supposed to be unknown.

Problem: observing $X = (X_1, X_2, \ldots, X_n)$ and $U = (u_1, u_2, \ldots, u_n)$, we would estimate a_0, a_1 and σ^2.
OLS: The approach of OLS involves in determining a_0, a_1 such that the residual sum of squares (RSS)

$$RSS = \sum_{t=1}^{n} \left(X_t - a_0 - a_1 u_t\right)^2$$

is minimized.
Solution: Let

$$f(a_0, a_1) = \sum_{t=1}^{n} \left(X_t - a_0 - a_1 u_t\right)^2.$$

f is differentiable over $\mathbb{R}^2$. Therefore one can solve

$$\begin{cases} \frac{\partial f}{\partial a_0} = 0; \\ \frac{\partial f}{\partial a_1} = 0. \end{cases}$$

The above equations could be written as

$$\begin{cases} -2\sum_{t=1}^{n} \left(X_t - a_0 - a_1 u_t\right) = 0; \\ -2\sum_{t=1}^{n} \left(X_t - a_0 - a_1 u_t\right) u_t = 0. \end{cases}$$

Further the solutions of (a_0, a_1) can be obtained as

$$\begin{cases} a_0 = \overline{X} - a_1 \overline{U}; \\ a_1 = \frac{\frac{\sum_{t=1}^{n} X_t u_t}{n} - \overline{XU}}{\frac{\sum_{t=1}^{n} u_t^2}{n} - \overline{U}^2} = \frac{\widehat{Cov}(U,X)}{\widehat{Var}(U)}, \end{cases}$$

where $\widehat{Cov}$ and $\widehat{Var}$ denote the empirical covariance and empirical variance respectively.
Note that

$$\begin{cases} \frac{\partial^2 f}{\partial a_0^2} = 2 > 0; \\ \frac{\partial^2 f}{\partial a_1^2} = 2\sum_{t=1}^n u_t^2 > 0. \end{cases}$$

Therefore the solutions a_0, a_1 minimize RSS. We then denote the estimates by $(\hat{a}_0, \hat{a}_1)$. In order to estimate σ^2, one just needs to apply the moment method:

$$\hat{\sigma}^2 = \frac{\widehat{RSS}}{n-2} = \frac{\sum_{i=1}^n (X_t - \hat{a}_0 - \hat{a}_1 u_t)^2}{n-2}.$$

Here $\widehat{RSS}$ is the estimate of RSS and $n-2$ makes the estimate $\hat{\sigma}^2$ unbiased.

3.5 Linear Seasonal Indicator Variables Model

Recall the additive decomposition model:

$$X_t = m_t + s_t + z_t.$$

In this section we fit the seasonal effect using the so-called indicator variables model, where s_t equals some constant in each season. Let s be the number of seasons in a time period. For example, if a time period is 1 year and the time t is monthly, then $s = 12$ (months); if the time is weekly, then $s = 52$. The linear seasonal indicator variables model is then defined by

$$X_t = m_t + \beta_i + z_t, \text{ if } t \text{ belongs to the } i\text{th season},$$

where $\beta_1, \ldots, \beta_s$ are s unknown constants.
As one example, let us consider a monthly data set X_t, then the seasonal indicator variables model can be given by

$$\begin{cases} X_t = at + \beta_1 + z_t, & \text{if } t = 1, 13, 25, \ldots; \\ X_t = at + \beta_2 + z_t, & \text{if } t = 2, 14, 26, \ldots; \\ \cdots & \cdots \\ X_t = at + \beta_{12} + z_t, & \text{if } t = 12, 24, 36, \ldots; \end{cases}$$

If the n values $X_1, \ldots, X_n$ are observed, then

$$s_t = \beta_i, \text{ for } t = i, i+12, \ldots, i+12k,$$

with k being the greatest integer such that $i + 12k \leq n$.
Estimation of $a, \beta_1, \ldots, \beta_s, \sigma^2$: In the above linear model, one can use OLS to estimate all unknown parameters $a, \beta_1, \ldots, \beta_s, \sigma^2$. But for estimating each

β_i, only the corresponding data set on $t = i, i+s, \ldots, i+ks$ are used. For example, let us denote by

$$I_i = \{i, i+s, \ldots, i+ks\},$$

with k being the greatest integer such that $i + ks \leq n$. And we denote by

$$X_{I_i} = \{X_t\}_{t \in I_i}.$$

Then for each i, using OLS to estimate the parameters a, β_i, σ^2 from

$$X_t = at + \beta_i + z_t,$$

we obtain

$$\begin{cases} \hat{\beta}_i = \overline{X}_{I_i} - \hat{a}_i \overline{I}_i; \\ \hat{a}_i = \frac{\widehat{Cov}(I_i, X_{I_i})}{\widehat{Var}(I_i)}; \\ \hat{\sigma_i^2} = \frac{\sum_{t \in I_i}(X_t - \hat{a}_i t - \hat{\beta}_i)}{\# I_i}; \end{cases}$$

where $\hat{a}_i$ and $\hat{\sigma_i^2}$ denote the ith estimate of a and σ^2 respectively, and $\# I_i$ denotes the number of elements in I_i. $\hat{\beta}_1, \ldots, \hat{\beta}_s$ are hence obtained. Finally we provide a low variance estimate of a and σ^2 by taking

$$\hat{a} = \frac{\sum_{i=1}^{s} \hat{a}_i}{s}; \; \hat{\sigma^2} = \frac{\sum_{i=1}^{s} \hat{\sigma_i^2}}{s}.$$

3.6 Harmonic Seasonal Models

We introduce another powerful seasonal models: harmonic seasonal model. Contrary to the previous one, this model is time-continuous. The basic idea of fitting a harmonic seasonal model is to use a harmonic function to fit the seasonal effect s_t. Recall that one simple harmonic model is given by

$$s_t = A \sin(2\pi f t + \phi),$$

where

- A is amplitude. It controls the lower bound and upper bound of the function: $|s_t| \leq A$.
- f is frequency. By definition $f = \frac{1(cycle)}{period}$ determines the number of cycles at each period.
- ϕ is the phase shift. It explains the initial position (or value) of the time series.

The fitting problem involves estimating the unknown parameters A, ϕ. By using the fact that

$$\sin(x+y) = \cos(s)\sin(y) + \sin(x)\cos(y),$$

one can write

$$s_t = a_s \sin(2\pi f t) + a_c \cos(2\pi f t),$$

where

$$a_s = A\cos(\phi),\ a_c = A\sin(\phi).$$

The model becomes a linear regression, hence OLS can be applied.
Now we introduce the general model: for a time series $\{X_t\}_t$ with s seasons, there are up to $[s/2]$ ($[s/2]$ denotes the floor number of $s/2$: the greatest integer less than or equal to $s/2$) possible cycles which could explain the values of s_t:

$$\begin{aligned} X_t &= m_t + \sum_{i=1}^{[s/2]} A\sin\left(2\pi t \frac{i}{s} + \phi\right) + z_t \\ &= m_t + \sum_{i=1}^{[s/2]} \left(s_i \sin\left(2\pi t\frac{i}{s}\right) + c_i \cos\left(2\pi t \frac{i}{s}\right)\right) + z_t, \end{aligned}$$

where s_i, c_i, $i = 1, \ldots, [s/2]$ and σ^2 are unknown parameters.
Why does the number of cycles only go up to $[s/2]$? To answer this question let's simply assume s is even, then for $s/2 < i \leq s$,

$$A\sin\left(2\pi t\frac{i}{s} + \phi\right) = A\sin\left(2\pi t\frac{i - s/2}{s} + \phi + \pi t\right) = (-1)^t A \sin\left(2\pi t \frac{j}{s} + \phi\right),$$

with $j = i - s/2 \leq s/2$. The above item turns out to be the jth cycle with $j \leq s/2$, which has been considered in the model.
Estimation of the parameters s_i, c_i, $i = 1, \ldots, [s/2]$: Note that if m_t is a linear model, the the harmonic seasonal model becomes a linear regression model, one can thus use OLS to solve it.

3.7 Inverse Transformation and Bias Correction

We arise our attention to the following issue:
Consider a time series $\{X_t\}$, for which the logarithmic transformation $Y_t = \log(X_t)$ can be fitted. The fitted values of Y_t is denoted by $\hat{Y}_t$. Then the fitted values of X_t are naturally

$$\hat{X}_t = e^{\hat{Y}_t}.$$

Assume that $\hat{Y}_t$ is unbiased, i.e., $E(\hat{Y}_t) = Y_t$. However $\hat{X}_t$ becomes biased, since by using the Jensen's inequality we have

$$E(\hat{X}_t) = E(e^{\hat{Y}_t}) \geq e^{E(\hat{Y}_t)} = e^{Y_t} = X_t.$$

We say $\hat{X}_t$ overestimates Y_t. Accordingly, the parameters in X_t are all estimated with bias. Therefore, after a non-linear transformation, the estimates of the parameters of X_t should be "corrected", by using some correction terms. For example,

- when the residual series z_t of the fitted log-regression model is Gaussian white noise, a factor $e^{\sigma^2/2}$ can be used:

$$\hat{X}_t' = e^{\frac{\sigma^2}{2}} \hat{X}_t.$$

- In general, however, the distribution of the residuals is often unavailable or of a fat-tail, in which case a correction factor can be determined empirically using the mean of the anti-log of the residual series. In this approach, the adjusted item can be given as the average of the exponential of residuals:

$$\hat{X}_t' = \frac{\sum_{t=1}^{n} e^{\hat{Z}_t}}{n} \hat{X}_t,$$

  where $\hat{Z}_t = X_t - \hat{a}_0 - \hat{a}_1 u_{1,t} - \hat{a}_m u_{m,t}$ is the tth residual of the linear model.

Chapter 4

Stationary Models

Motivation: A time series' residuals $\{z_t\}_t$ are often correlated in time t. In this chapter we introduce more alternative stationary models to describe $\{z_t\}$.
Main goal: Know the concept and properties of moving average and moment properties of $ARMA(p, q)$.

4.1 Moving Average Model

4.1.1 $MA(q)$ process: definition and properties

Definition 4.1.1 ($MA(q)$) *A moving average model $MA(q)$ of order q (without trend) is a time series $\{X_t\}_t$ satisfying*

$$X_t = b_1 w_{t-1} + b_2 w_{t-2} + \ldots + b_q w_{t-q} + w_t = \sum_{i=1}^{q} b_i w_{t-i} + w_t,$$

where $\{w_t\}_t$ is a white noise with variance σ^2 and $b_1, \ldots, b_q$ are real numbers.

Here are some observations on $MA(q)$:

1. X_t is a linear regression of $w_{t-1}, \ldots, w_{t-q}$. In finance, it can be explained as: today's price depends on the values of of the past q days of some unknown factor.

2. Remember that for each X_t, w_t denotes its error term.

3. Using the backward shift operator B, one can write

$$X_t = (1 + b_1 B + b_2 B^2 + \ldots b_q B^q) w_t.$$

   Please compare this expression to that of an $AR(p)$.

The following property is essential for $MA(q)$.

Theorem 4.1.2 *A $MA(p)$ is second-order stationary.*

Proof. Let $X(t) \sim MA(q)$. We only show the theorem is true when the trend is zero. Indeed, when the trend is a nonzero constant, all the results still hold. Now It is sufficient to derive its expectation and covariance to show both are relevant to time t.

- The expectation of X_t is given as

$$E(X_t) = E(b_1 w_{t-1} + b_2 w_{t-2} + \ldots + b_p w_{t-q} + w_t) = 0,$$

which is independent of t.

- Let $b_0 = 1$, we then can write

$$X_t = \sum_{i=0}^{q} b_i w_{t-i}.$$

The covariance $Cov(X_t, X_{t+k})$ for $k \in \mathbb{Z}$ can thus be given as below.

$$\begin{aligned} Cov(X_t, X_{t+k}) &= Cov\left(\sum_{i=0}^{q} b_i w_{t-i}, \sum_{j=0}^{q} b_j w_{t+k-j}\right) \\ &= \sum_{i=0}^{q}\sum_{j=0}^{q} b_i b_j Cov(w_{t-i}, w_{t+k-j}). \end{aligned}$$

Two cases follow, with respective to the value of k.

Case 1 If $|k| \le q$. Then there exist $i, j \in \{0, 1, \ldots, q\}$ such that $k = j - i$.

Case 1.1 If $0 \le k \le p$, then we let $j = k + i$ and $i = 0, \ldots, q - k$ to obtain

$$\begin{aligned} Cov(X_t, X_{t+k}) &= \sum_{i=0}^{q}\sum_{j=0}^{q} b_i b_j Cov(w_{t-i}, w_{t+k-j}) \\ &= \sum_{i=0}^{q-k}\sum_{j=k+i} b_i b_j \times \sigma^2 + \sum_{i=0}^{q-k}\sum_{j\neq k+i} b_i b_j \times 0 \\ &= \sum_{i=0}^{q-k} b_i b_{k+i} \sigma^2. \end{aligned}$$

Case 1.2 If $-p \le k < 0$, we then let $j = k + i$ with $i = -k, \ldots, q$ to obtain

$$\begin{aligned}
Cov(X_t, X_{t+k}) &= \sum_{i=0}^{q} \sum_{j=0}^{q} b_i b_j Cov(w_{t-i}, w_{t+k-j}) \\
&= \sum_{i=-k}^{q} \sum_{j=k+i} b_i b_j \times \sigma^2 + \sum_{i=0}^{q-k} \sum_{j \ne k+i} b_i b_j \times 0 \\
&= \sum_{i=-k}^{q} b_i b_{k+i} \sigma^2 \\
&= \sum_{j=0}^{q+k} b_{j-k} b_j \sigma^2 \text{ (we have set } j = i + k).
\end{aligned}$$

We conclude from Cases 1 and 2 that

$$Cov(X_t, X_{t+k}) = \sum_{i=0}^{q-|k|} b_i b_{|k|+i} \sigma^2, \text{ for } |k| \le q.$$

Case 2 If $|k| > q$, then $j = k + i$ never holds, we then have

$$Cov(X_t, X_{t+k}) = 0.$$

This can be also explained as: X_t is some function of $(w_t, w_{t-1}, \ldots, w_{t-q})$; X_{t+k} is some function of $(w_{t+k}, w_{t+k-1}, \ldots, w_{t+k-q})$. The two sets of white noises are disjoint when $|k| > q$, so they are independent, the covariance is hence zero.

We finally conclude that the autocovariance function

$$\gamma(k) = Cov(X_t, X_{t+k}) = \begin{cases} \sum_{i=0}^{q-|k|} b_i b_{|k|+i} \sigma^2, & \text{for } |k| \le q \\ 0, & \text{for } |k| > q. \end{cases}$$

It does not depend on t, so $MA(q)$ is second-order stationary. Another observation is that

$$Var(X_t) = \gamma(0) = \sum_{i=0}^{q} b_i^2 \sigma^2 = \left(1 + \sum_{i=1}^{q} b_i^2\right) \sigma^2.$$

4.1.2 Invertibility of $MA(q)$

We know that a $MA(q)$ (without trend) can be expressed as

$$X_t = \theta_q(B) w_t,$$

where $\theta_q(B)$ is a polynomial of degree q of B:

$$\theta_q(B) = 1 + b_1 B + b_2 B^2 + \ldots + b_q B^q.$$

It is also known that any $MA(q)$ is second-order stationary. Now we discuss of the following problem: **can X_t be expressed as an autoregressive model?** This problem refers to the **invertibility** of $MA(q)$.

Definition 4.1.3 (Invertible time series) *If $X_t \sim MA(q)$ can be expressed as a* ***stationary*** *$AR(\infty)$ below:*

$$w_t = \frac{1}{\theta_q(B)} X_t = (1 + a_1 B + a_2 B^2 + \ldots) X_t,$$

then $\{X_t\}_t$ is invertible.

Remarks:

1. Invertibility is an essential approach to check whether an $AR(\infty)$ is stationary.

2. Mathematically, invertibility is equivalent to saying "the inverse polynomial $\frac{1}{\theta_q(B)}$ has its Taylor expansion at 0, and this Taylor series is convergent when taking $B = 1$".

3. In finance, the behavior of an instrument $X_t \sim MA(q)$ is invertible shows that X_t depends on all its past values, but the impact of X_{t-k} on X_t decreases quite fast as the lag k increases. Stationarity means implies that forecasting of X_t using its past values can be made. If a time series is not invertible, then the forecasting of X_t won't be accurate at all from using its past values.

Example 4.1.4 *Consider a $MA(1)$:*

$$X_t = (1 + bB) w_t.$$

It is equivalent to

$$w_t = X_t - b w_{t-1} = X_t - b X_{t-1} + b^2 w_{t-2} = \ldots = X_t - b X_{t-1} + b^2 X_{t-2} - \ldots + (-1)^k b^k w_{t-k}.$$

The error term

$$(-1)^k b^k w_{t-k}$$

converges to 0 if and only if $|b| < 1$. Equivalently, the root of $1 - bB = 0$ satisfies $|B| > 1$. Therefore the root $|B| > 1$ is a sufficient and necessary condition for $\{X_t\}_t$ to be invertible.

More generally, we have the following result:

Theorem 4.1.5 *A $MA(q)$ is invertible if and only if its characteristic function $\theta_p(B)$ has all its roots satisfying $|B| > 1$.*

The proof is very similar to that of stationarity of $AR(p)$, so we omit it.

4.2 Mixed Models: $ARMA(p,q)$

Definition 4.2.1 **(**$ARMA(p,q)$**)** *Denote by $ARMA(p,q)$ an autoregressive moving average model of orders (p,q). We say a time series $X_t \sim ARMA(p,q)$ (without trends) if X_t has the following form:*

$$\begin{aligned} X_t &= a_1 X_{t-1} + \ldots a_p X_{t-p} + w_t + b_1 w_{t-1} + \ldots + b_q w_{t-q} \\ &= \underbrace{\sum_{i=1}^{p} a_i X_{t-i} + \sum_{j=1}^{q} b_j w_{t-j}}_{\textit{seasonal effect}} + \underbrace{w_t}_{\textit{error term}} . \end{aligned}$$

Remark that by using the backward shift operator B, an $ARMA(p,q)$ can be denoted by

$$\theta_p(B) X_t = \phi_q(B) w_t,$$

where the two characteristic functions

$$\theta_p(B) = 1 - a_1 B - a_2 B^2 - \ldots - a_p B^p$$

and

$$\phi_q(B) = 1 + b_1 B + b_2 B^2 + \ldots + b_q B^q.$$

Below are the important observations on $ARMA(p,q)$.

1. An $ARMA(p,q)$ is second-order stationary if and only if all roots of $\theta_p(B) = 0$ satisfy $|B| > 1$.

2. An $ARMA(p,q)$ is invertible if and only if all roots of $\phi_q(B) = 0$ satisfy $|B| > 1$.

3. When θ and ϕ share a common factor, then a **stationary model** $ARMA(p,q)$ can be simplified. For example, the model

$$(1 - \frac{1}{2}B)(1 - \frac{1}{3}B) X_t = (1 - \frac{1}{2}B) w_t$$

   can be written as

$$(1 - \frac{1}{3}B) X_t = w_t.$$

   It turns out that the $ARMA(2,1)$ is simplified to $AR(1)$.

Next we take another 3 examples to show how it works (why is it necessary to assume stationarity? It will be funny).

Example 1: Consider an $ARMA(1,1)$:

$$(1 - \frac{1}{2}B) X_t = (1 - \frac{1}{2}B) w_t.$$

Show that $X_t = w_t$.
Proof. We write the model as

$$X_t - \frac{1}{2}X_{t-1} = w_t - \frac{1}{2}w_{t-1}.$$

Equivalently,

$$X_t - w_t = \frac{1}{2}\left(X_{t-1} - w_{t-1}\right), \text{ for all } t.$$

As a consequence we can say

$$X_t - w_t = \left(\frac{1}{2}\right)^k \left(X_{t-k} - w_{t-k}\right) \xrightarrow[k\to\infty]{} 0.$$

Hence $X_t = w_t$. The $ARMA(1, 1)$ is reduced to $AR(0)$ or $MA(1)$.
Example 2: Consider the $ARMA(1, 1)$ given as

$$(1 - B)X_t = (1 - B)w_t.$$

We will show that it is not necessary that $X_t = w_t$. In fact we have

$$X_t - X_{t-1} = w_t - w_{t-1}.$$

Equivalently,

$$X_t - w_t = X_{t-1} - w_{t-1}, \text{ for all } t.$$

Hence $X_t - w_t$ is a constant (not depending on t)! We denote this constant by c (not necessarily zero), therefore

$$X_t = c + w_t.$$

Example 3: Consider

$$(1 - 2B)X_t = (1 - 2B)w_t.$$

Then one has

$$X_t - w_t = 2(X_{t-1} - w_{t-1}) = \ldots = 2^k(X_{t-k} - w_{t-k}) \neq 0, \text{ as } k \to \infty.$$

Therefore $X_t \neq w_t$, the above $ARMA(1, 1)$ can't be simplified to $ARMA(0, 0)$.

4.2.1 Second-order Properties of $ARMA(p, q)$

It is basically quite difficult to derive the second-order moments of a general $ARMA(p, q)$. Many algorithms exist, but no general closed formula is known (indeed it does not exist). However in some simple cases, the backward shift operator B could be helpful for obtaining the covariance of $ARMA(p, q)$.

In this subsection we consider a stationary $ARMA(1, 1)$, given as

$$X_t = aX_{t-1} + w_t + bw_{t-1},$$

with $-1 < a < 1$. Our goal is to derive its covariance function $Cov(X_t, X_{t+k})$. The above model can be written as

$$X_t = (1 - aB)^{-1}(1 + bB)w_t.$$

Hence,

$$\begin{aligned} X_t &= \left(\sum_{i=0}^{\infty} a^i B^i\right)(1 + bB)w_t \\ &= \left(1 + \sum_{i=0}^{\infty} a^{i+1}B^{i+1} + \sum_{i=0}^{\infty} a^i b B^{i+1}\right) w_t \\ &= w_t + (a + b)\sum_{i=1}^{\infty} a^{i-1} w_{t-i}. \end{aligned}$$

For simplifying notation purpose we denote by $\alpha_0 := \frac{1}{(a+b)}$ (if $a + b = 0$, the one has directly $X_t = w_t$) and $\alpha_i = a^{i-1}$ for $i \geq 0$. Therefore

$$X_t = (a + b)\sum_{i=0}^{\infty} \alpha_i w_{t-i}.$$

$$\begin{aligned} Cov(X_t, X_{t+k}) &= Cov\left((a + b)\sum_{i=0}^{\infty} \alpha_i w_{t-i}, (a + b)\sum_{j=0}^{\infty} \alpha_j w_{t+k-j}\right) \\ &= (a + b)^2 \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \alpha_i \alpha_j Cov\left(w_{t-i}, w_{t+k-j}\right). \end{aligned}$$

For $k \geq 0$, we let $j = i + k$ for $i = 0, 1, \ldots$. Then

$$Cov(X_t, X_{t+k}) = (a + b)^2 \sum_{i=0}^{\infty} \alpha_i \alpha_{i+k}\sigma^2 = (a + b)\alpha_k\sigma^2 + (a + b)^2\sigma^2 a^k(1 - a^2)^{-1}.$$

When $k < 0$, one can use similar computation to obtain

$$Cov(X_t, X_{t+k}) = (a + b)a^{-k}\sigma^2 + (a + b)^2\sigma^2 a^{-k}(1 - a^2)^{-1}.$$

Finally, for all $k \in \mathbb{Z}$,

$$Cov(X_t, X_{t+k}) = (a + b)\alpha_{|k|}\sigma^2 + (a + b)^2\sigma^2 a^{|k|}(1 - a^2)^{-1}.$$

Chapter 5

Non-stationary Models

Motivation: Non-stationary trends and seasonal effects can both result in non-stationarity of the time series. In this chapter we introduce autoregressive integrated moving average models ($ARIMA(p, d, q)$) to model a non-stationary time series.
Main goal: Know fitting the non-stationary models $ARIMA$, $ARCH$ and $GARCH$.

5.1 Non-seasonal ARIMA Models

Definition 5.1.1 (Integrated series) *A series $\{X_t\}_t$ is* integrated of order *d, denoted as $I(d)$, if the dth difference of $\{X_t\}$ is a white noise, i.e. $\nabla^d X_t = w_t$.*

For example,

- the white noise $\{w_t\}$ is integrated of order 0. We then denote $\{w_t\}_t = I(0)$;
- the random walk $X_t = X_{t-1} + w_t$ is integrated of order 1, since $\nabla X_t = X_t - X_{t-1} = w_t$. We denote $\{X_t\}_t = I(1)$;
- the $AR(2)$ $X_t = -2X_{t-1} + X_{t-2} + w_t$ is integrated of order 2, since $\nabla^2 X_t = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) = w_t$. We denote $\{X_t\}_t = I(2)$.

Observe that $\nabla^d = (1 - B)^d$, where B denotes the backward shift operator, then one can denote

$$(1 - B)^d X_t = w_t,$$

if $\{X_t\}_t$ is integrated of order d. Also remark that the series integrated of order d is not unique. Indeed, if $\{X_t\}_t$ is $I(d)$, so is $X_t + Y_t$ for all Y_t being $I(r)$, with $r < d$. The *autoregressive integrated moving average model* ($ARIMA$), is defined based on high-order differenced series.

Definition 5.1.2 **(**$ARIMA(p, d, q)$**)** *A time series* $\{X_t\}_t$ *follows an* $ARIMA(p, d, q)$ *model if the dth differences of* $\{X_t\}_t$ *are an* $ARMA(p, q)$ *process. Mathematically, we say*

$$\{X_t\}_t \sim ARIMA(p, d, q)$$

if and only if

$$(1 - B)^d X_t \sim ARMA(p, q).$$

Remarks:

(1) $ARIMA(p, d, q)$ is quite a general model, it extends $ARMA(p, q)$. In fact $ARMA(p, q) = ARIMA(p, 0, q)$.

(2) Accordingly the integrated autoregressive model is denoted by $ARI(p, d) = ARIMA(p, d, 0)$ and the integrated moving average model is denoted by $IMA(d, q) = ARIMA(0, d, q)$.

(3) By definition we can from now on use characteristic polynomials to denote an $ARIMA(p, q, d)$:

$$\theta_p(B)(1 - B)^d X_t = \psi_q(B) w_t.$$

(4) By definition an $ARIMA(p, d, q)$ is also $ARIMA(p + d - r, r, q)$ for $r = 0, 1, 2, \ldots, d - 1$. This is because one can write

$$\theta_p(B)(1 - B)^d X_t = \psi_q(B) w_t$$

by

$$\underbrace{\left(\theta_p(B)(1 - B)^{d-r}\right)}_{new\ \theta_{p+d-r}(B)}(1 - B)^r = \psi_q(B) w_t.$$

In particular, if $r = 0$, we get

$$ARIMA(p, d, q) = ARMA(p + d, q).$$

We see that $ARIMA(p, d, q)$ plays the role of reducing the degree of AR characteristic polynomial.

Examples:

(1) The time series $X_t = X_{t-1} + w_t + b w_{t-1}$ can be written as

$$(1 - B)X_t = (1 + bB)w_t,$$

then $\{X_t\}_t$ follows $ARIMA(0, 1, 1)$ or $IMA(1, 1)$.

(2) The time series $X_t = (1+a)X_{t-1} - aX_{t-2} + w_t$ can be represented as

$$(1-aB)(1-B)X_t = w_t.$$

Then it is an $ARIMA(1,1,0)$ or equivalently $ARI(1,1)$. However, if $a = 1$, one can express $\{X_t\}_t$ as

$$(1-B)^2 X_t = w_t,$$

equivalently, $\{X_t\}_t \sim ARIMA(0,2,0) = I(2)$.

Applications of $ARIMA(p,d,q)$
In finance and marketing, any non-seasonal non-stationary time series sample data can be fitted using an $ARIMA(p,d,q)$, such as cosmetics production series, number of sold item (non-seasonal) series, short-term interest or currency exchange rate series. Many non-stationary $ARMA(p,q)$ can be fitted by an $ARIMA(p,d,q)$, this gives a chance to forecast non-stationary series.

5.2 Seasonal ARIMA Models

The $ARIMA$ process can be extended to include seasonal terms, giving a non-stationary seasonal $ARIMA$ ($SARIMA$) process. Seasonal $ARIMA$ models are powerful tools in the analysis of time series as they are capable of modeling a very wide range of series. Much of the methodology was pioneered by Box and Jenkins in the 1970's.

Definition 5.2.1 **(**$ARIMA(p,d,q)(P,D,Q)_s$**)** *A seasonal ARIMA, denoted by* $ARIMA(p,d,q)(P,D,Q)_s$*, can be expressed as*

$$\Theta_P(B^s)\theta_p(B)(1-B^s)^D(1-B)^d X_t = \Psi_Q(B^s)\psi_q(B)w_t,$$

where

- Θ_P*,* θ_p*,* Ψ_Q *and* ψ_q *are polynomials of degrees* P, p, Q, q*.*
- s *is the number of seasons, such as* 12 *(months),* 52 *(weeks) or* 365 *(days).*

Remarks:

(1) Roughly speaking, $\{X_t\}_t \sim ARIMA(p,d,q)(P,D,Q)_s$ says the dth differenced series (of $\{X_t\}_t$)'s Dth differenced series has seasonal period s. So $ARIMA(p,d,q)(P,D,Q)_s$ covers quite a huge range of time series such as $ARIMA(p,d,q)$ and indicator variable seasonal linear models. Observe that we have two other equivalent ways to explain the seasonal

$ARIMA$:
Way 1: we can write

$$\Theta_P(B^s)(1-B^s)^D Y_t = \Psi_Q(B^s)\psi_q(B)w_t,$$

with Y_t satisfying

$$Y_t = \theta_p(B)(1-B)^d X_t \sim ARIMA(Ps, D, Qs+q).$$

Way 2: we can write

$$\Theta_P(B^s)(1-B^s)^D Y_t = \Psi_Q(B^s)w_t,$$

with Y_t satisfying

$$Y_t = \frac{1}{\psi_q(B)}\theta_p(B)(1-B)^d X_t \sim ARIMA(P, D, Q)_s.$$

(2) The $ARIMA(p,d,q)(P,D,Q)_s$ is generally not stationary. It can be stationary if $D = d = 0$ and the roots of the left hand side polynomial has all its roots exceeding unity in absolute values.

Examples:

(1) $X_t = aX_{t-12} + w_t$ is called an AR model with a seasonal period of 12 units. It is can be used to describe monthly series, with a period of 1 year. Using characteristic polynomials it can be written as

$$(1-aB^{12})X_t = w_t.$$

Compared to the definition of $ARIMA(p,d,q)(P,D,Q)_s$, it can be expressed as

$$\Theta_1(B^{12})X_t = w_t,$$

hence $\{X_t\}_t$ follows $ARIMA(0,0,0)(1,0,0)_{12}$.

(2) Consider $X_t = X_{t-1} + aX_{t-12} - aX_{t-13} + w_t$. It can be written as

$$(1-aB^{12})(1-B)X_t = w_t.$$

Or equivalently,

$$\Theta_1(B^{12})(1-B)X_t = w_t,$$

then it can be expressed as $ARIMA(0,1,0)(1,0,0)_{12}$.

5.3 Some Other Non-stationary Financial Series: ARCH and GARCH

A wide range of financial indices and credit ratings contain volatilities (variance parameter). Sometime it is reasonable to assume that these volatilities are time-dependent. For example, a random walk (starting from some time t_0)'s variance at each time t is growing with respect to t, then we say it has time-dependent volatility. If a series exhibits periods of increased variance, like electricity production, the series is then called conditional heteroskedastic. In discrete time series analysis, if the volatility changes via time in a regular way, we can consider an autoregressive conditional heteroskedastic model to fit the error term.

Definition 5.3.1 (Volatility) *If a financial time series has the decomposition*

$$X_t = \underbrace{m_t}_{trend} + \underbrace{s_t}_{seasonal\ effect} + \underbrace{\sigma_t w_t}_{error\ term},$$

then the coefficient function σ_t of the white noise w_t in the error term is called volatility. *It measures the conditional variance of the series X_t.*

For example, consider the following time-dependent $AR(1)$ process:

$$X_t = X_{t-1} + \sigma_t w_t, \text{ where } \sigma_t \text{ is a deterministic function.}$$

We have

$$Var(X_t|X_{t-1}) = Var(\sigma_t w_t) = \sigma_t^2 Var(w_t) = \sigma_t^2 \sigma^2,$$

which is time-dependent. We see the value of the volatility σ_t heavily explains the conditional variance $Var(X_t|X_{t-1})$. If σ_t is growing as t increases, we then say $\{X_t\}$ is conditional heteroskedastic.

Definition 5.3.2 ($ARCH(p)$ model) *An $ARCH(p)$ process is given by*

$$\epsilon_t = w_t \sqrt{a_0 + \sum_{i=1}^{p} a_i \epsilon_{t-i}^2},$$

where

- $a_0 > 0$, $a_1, \ldots, a_p \geq 0$ *are model parameters. The positivity guarantees that the square root returns a real number.*

- $\{w_t\}_t$ *is by default a white noise with zero mean and **variance** $\sigma^2 = 1$.*

Remark: If $w_t \sim \mathcal{N}(0,1)$, then we can say

$$\epsilon_t = w_t \sqrt{a_0 + \sum_{i=1}^{p} a_i \epsilon_{t-i}^2}$$

is equivalent to

$$\epsilon_t^2 = w_t^2 \left(a_0 + \sum_{i=1}^{p} a_i \epsilon_{t-i}^2 \right).$$

And roughly, since $E(w_t^2) = 1$, we have

$$\epsilon_t^2 \approx a_0 + \sum_{i=1}^{p} a_i \epsilon_{t-i}^2 + w_t',$$

with w_t' being another white noise. This shows the time series $Y_t = \epsilon_t^2 \sim AR(p)$. **In conclusion, the $ARCH(p)$ is approximately a square root process of $AR(p)$.**

Example: For $\{\epsilon_t\}_t \sim ARCH(1)$, we can show that the variance sequence

$$Var(\epsilon_t) = a_0 + a_1 Var(\epsilon_{t-1})$$

which is eventually non-decreasing.
Further, a more general model, called generalized ARCH $((GARCH(p,q)))$ is defined by

$$\epsilon_t = w_t \sqrt{h_t},$$

where

$$h_t = a_0 + \sum_{i=1}^{p} a_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} b_j h_{t-j}.$$

Note that $GARCH(p,0) = ARCH(p)$. $GARCH(p,q)$ is widely used in financial applications, such as $S\&P$ Index (calculated from the stock prices of 500 large corporations), including the Dow Jones Industrial Average Index. In fact it is so far the most general (and efficient) model among discrete time series which is used to fit the error terms exhibiting the conditional heretoskedastic feature.

Chapter 6

Long-Memory Processes

Motivation: Some financial time series exhibit correlations at high lag, this phenomenon is called long-memory, and the time series is called long-memory process. The modeling and forecasting techniques for a long-memory are quite different from those of a stationary-based time series.
Main goal: Know fitting a long-memory process using fractional differencing series.

6.1 Long-memory property

Let $\{X_t\}_t$ be a stationary time series, one can observe that its autocorrelation ρ_k falls down to zero very quickly, such as

$$|\rho_k| \sim a^k, \text{ with } |a| < 1,$$

which runs faster to zero than any polynomial[1]. However there exists a class of stationary times series, of which the autocorrelation ρ_k goes to zero SLOWLY. These processes are usually not stationary. Long-memory process is one of such time series.

Definition 6.1.1 *A stationary time series $\{X_t\}_t$ with long-memory has an autocorrelation function ρ_k that satisfies*

$$|\rho_k| \sim k^{-\lambda},$$

where $\lambda \in (0, 1)$ is some constant. The closer λ is to 0, the more pronounced is the long-memory.

The long-memory basically tells that, the current asset price depends heavily on the historical prices from a long time ago (so that the correlation of high

[1]Mathematically, $\rho_k \sim a^k$ is defined to be $\frac{\rho_k}{a^k} \xrightarrow[k\to\infty]{} c$ for some $c \neq 0$.

lag is not negligible). Another way to define the long-memory property is: the stationary process $\{X_t\}$ is of long-memory if the autocorrelation function is summable:

$$\sum_{k=1}^{+\infty} |\rho_k| = +\infty.$$

Examples:

1. If $\rho_k = k^{-2}$ for $k = 1, 2, \ldots$, then

$$\sum_{k=1}^{+\infty} |\rho_k| = \sum_{k=1}^{+\infty} k^{-2} = \frac{\pi^2}{6} < +\infty.$$

2. If $\rho_k = k^{-1/2}$ for $k = 1, 2, \ldots$, then for any $N \geq 1$,

$$\sum_{k=1}^{N} |\rho_k| = \sum_{k=1}^{N} \int_k^{k+1} k^{-1/2}\, \mathrm{d}x \geq \sum_{k=1}^{N} \int_k^{k+1} x^{-1/2}\, \mathrm{d}x = \int_1^{N+1} x^{-1/2}\, \mathrm{d}x.$$

   Taking limits to both sides leads to

$$\sum_{k=1}^{+\infty} |\rho_k| \geq \int_1^{+\infty} x^{-1/2}\, \mathrm{d}x = +\infty,$$

   which shows $\{X_t\}_t$ is a short-memory process (ρ_k goes to zero fast.).

6.2 Fractional Differencing

As one typical example of long-memory process, we introduce fractionally differenced ARIMA.

Definition 6.2.1 *A fractionally differenced ARIMA $\{X_t\}_t$, denoted by $\{X_t\}_t \sim FARIMA(p, d, q)$, is defined by*

$$\phi_p(B)(1-B)^d X_t = \psi_q(B) w_t, \tag{6.2.1}$$

where ϕ_p is a polynomial of degree p; ψ_q is a polynomial of degree q; and d is a fraction satisfying $-\frac{1}{2} < d < \frac{1}{2}$. When $0 < d < \frac{1}{2}$, $\{X_t\}_t$ is a long-memory process.

Example: when $\{X_t\}_t \sim FARIMA(0, d, 0)$, with $d \in (-\frac{1}{2}, \frac{1}{2})$, we can show that $\{X_t\}_t$ is stationary and

$$\rho_k \sim k^{2d-1}.$$

Then $\{X_t\}_t$ is a long-memory process if $\lambda = 1 - 2d \in (0, 1)$, equivalently, $d \in (0, \frac{1}{2})$.

6.3 Fitting FARIMA

There are 2 ways to fit $FARIMA$.

Method 1 (6.2.1) is equivalent to

$$(1-B)^d X_t = (\phi_p(B))^{-1}\psi_q(B)w_t.$$

Now we denote by $Y_t = (1-B)^d X_t$. Then observe that, by a Taylor expansion around B,

$$(1-B)^d = 1 - dB + \frac{d(d-1)}{2!}B^2 - \frac{d(d-1)(d-2)}{B^3} + \ldots$$

In order to generate X_t, we only consider the expansion up to lag 40 (the error is then negligible):

$$(1-B)^d \approx 1 - dB + \frac{d(d-1)}{2!}B^2 - \frac{d(d-1)(d-2)}{B^3} + \ldots + \frac{d(d-1)\ldots(d-39)}{40!}B^{40}. \tag{6.3.1}$$

Now we introduce the algorithm to generate $\{X_t\}_t$:
Algorithm 1:

Step 1 Generate $\{w_t\}_t$, at the meanwhile we obtain the values of $Y_t = (\phi_p(B))^{-1}\psi_q(B)w_t$.

Step 2 Solve $\{X_t\}_t$ from (6.3.1).

Method 2 We can also write (6.2.1) as

$$X_t = (\phi_p(B))^{-1}\psi_q(B)(1-B)^{-d}w_t.$$

Then $\{X_t\}_t$ can be straightforwardly generated.
Algorithm 2:

Step 1 Generate $\{w_t\}_t$.

Step 2 Compute $\{X_t\}_t$ by

$$X_t = (\phi_p(B))^{-1}\psi_q(B)(1-B)^{-d}w_t.$$

Note that in the second algorithm we have to approximate $(1-B)^{-d}$ by

$$(1-B)^{-d} \approx 1 + dB + \frac{-d(-d-1)}{2!}B^2 - \frac{-d(-d-1)(-d-2)}{B^3} + \ldots + \frac{-d(-d-1)\ldots(-d-39)}{40!}B^{40}.$$

Applications of long-memory processes:

- Nile minima flows: Hurst (1951) found that $X_t =$ **the annual minimum water level (mm)** against **time** t (from year 622 to 1284) can be described as
$$X_t = m_t + s_t + z_t,$$
where the stochastic part s_t exhibits long-memory property. More precisely, $\{s_t\}_t$ is stationary with its autocorrelation
$$\rho_k \sim k^{\lambda}, \ \lambda \in (0, 1).$$

- Bank loan rate: it is observed that $X_t =$ **monthly percentage US Federal Reserve Bank prime loan rate** against **time** t (monthly, from January 1949 to November 2007) exhibits long-memory property.