

# Chapter 2: Theoretical Background

Draft version

This chapter presents a brief introduction to the theory of time series analysis. It is meant to serve as a theoretical foundation of the concepts used in this thesis. Thus, the focus will be clearly on those specific concepts, and the chapter should to no means be considered a thorough overview of the complete field of study.

The chapter is divided into five sections. In the first section, a formal definition of time series is given. The second section discusses the main characteristics of time series. Section three describes the different components of time series data and presents methods to split a time series into these components. The fourth section introduces statistical models to forecast time series, in particular the *autoregressive integrated moving average* (ARIMA) model. Finally, the fifth section focusses on a specific niche within time series analysis that is used in this thesis, namely time series clustering.

## 2.1 Time series definition

According to Woodward, Gray, and Elliott (2017), a time series can be defined as follows:

**Definition 1** A *time series* is a special type of a stochastic process. A stochastic process  $\{Y(t); t \in T\}$  is a collection of random variables, where  $T$  is an index set for which all of the random variables are defined on the same sample space. When  $T$  represents time, the stochastic process is referred to as a time series. ■

Typically, observations are made at equally spaced time intervals, such as every hour, every day or every year. In such cases,  $T$  takes on a discrete set of values, and we refer to them as *discrete-time time series*. On the other hand, *continuous-time time series* arise when  $T$  takes on a continuous range of values (Brockwell and Davis 2002). In this thesis, only discrete-time time series are analyzed.

An observed realization of the stochastic process described in Definition 1, is referred to by Woodward, Gray, and Elliott (2017) as a *realization of a time series*. Other works, such as Brockwell and Davis (2002), Shumway and Stoffer (2011) and Hyndman and Athanasopoulos (2018), use the term *time series* for both the data and the stochastic process of which it is a realization. In this thesis, for the sake of simplicity, the latter approach is used, and no notational distinction will be made.

**Definition 2** A *time series* is a set of observed values  $\{y_t\}$  of a stochastic process  $\{Y(t); t \in T\}$ , where  $T$  represents time. ■

From the context it will be clear whether the term time series refers to the process (Definition 1) or its realization (Definition 2). When clarification is needed, it will be given locally.

## 2.2 Time series characteristics

### 2.2.1 Autocorrelation

Analyzing time series raises unique problems in statistical modeling and inference. In conventional statistics, methods rely on the assumptions of independent and identically distributed random variables. However, in time series analysis, observations made at nearby moments in time are likely to be related. That is, it is likely that there exist internal relationships within a time series. If these relationships are linear, they are called autocorrelation, as defined in Definition 3 (Shumway and Stoffer 2011).

**Definition 3** *Autocorrelation* measures the linear correlation between two points on the same time series observed at different times. ■

Given a sample  $\{y_1, y_2, \dots, y_n\}$  of a time series, the degree of dependence in the data can be assessed by computing the *autocorrelation function* (ACF), for each lag  $h$ , as defined in Equation 1 (Brockwell and Davis 2002).

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}, \quad -n < h < n$$

Where  $\gamma(h)$  is the autocovariance function of the sample at lag  $h$ , as defined in Equation 2.

$$\gamma(h) = n^{-1} \sum_{t=1}^{n-|h|} (y_{t+|h|} - \bar{y})(y_t - \bar{y}), \quad -n < h < n$$

Where  $n$  is the length of the sample,  $y_t$  is the data value at time  $t$ ,  $y_{t-|h|}$  is the data value at time  $t$  minus  $|h|$  time periods, and  $\bar{y}$  is the mean of the sample. Whenever a time series  $\{Y_t\}$  is stationary, a concept that is introduced in the next section, the ACF of the sample  $\{y_t\}$  can be used as an estimate for the ACF of  $\{Y_t\}$ .

Conventional statistical methods can not be appropriately applied to data that exhibit autocorrelation, since the independency assumption is violated. The primary objective of time series analysis, therefore, is to develop mathematical models that are appropriate to use for data with a temporal autocorrelation structure (Shumway and Stoffer 2011).

## 2.2.2 Stationarity

In time series analysis, a key role is played by time series whose statistical properties do not vary with time (Brockwell and Davis 2002). Such time series are referred to as *stationary time series*. The most restrictive form of stationarity is defined in Definition 4 (Woodward, Gray, and Elliott 2017).

**Definition 4** A time series  $\{Y(t); t \in T\}$  is *strictly stationary* if for any  $t_1, t_2, \dots, t_k \in T$  and any  $h \in T$ , the joint distribution of  $\{Y_{t_1}, Y_{t_2}, \dots, Y_{t_k}\}$  is identical to that of  $\{Y_{t_1+h}, Y_{t_2+h}, \dots, Y_{t_k+h}\}$ . ■

However, it is often too hard to mathematically establish the requirement of strict stationarity, since the involved distributions are not known. For most applications, a milder form of stationarity, which only imposes conditions on the first and second moment of a time series, is sufficient. This is officially known as *weak stationarity*, but, in time series analysis, usually just called *stationarity*. It is mathematically defined as in Definition 5 (Woodward, Gray, and Elliott 2017).

**Definition 5** A time series  $\{Y(t); t \in T\}$  is *stationary* if

- 1)  $E[Y_t] = \mu$  (constant for all  $t$ ).
- 2)  $Var[Y_t] = \sigma^2$  (constant for all  $t$ ).
- 3)  $\gamma(t_1, t_2)$  depends only on  $t_2 - t_1$ . ■

In words, this means that a time series is said to be stationary when the mean and variance are constant over time, and the autocovariance function only depends on the difference between two time points, and not on the time point itself. Note here that each time series that is strictly stationary (Definition 4), is, by definition, also stationary (Definition 5).

Stationarity is important, since, when studying real-world time series, it is common that only one realization of the series is available. That means that each random variable in the time series is represented by a single value. This makes it impossible to get an understanding of the underlying probability distributions of those random variables, unless it is assumed that their statistical properties are the same (i.e. the time series is stationary). In that case, the statistical properties of the whole sample can be used to estimate the statistical properties of each individual probability distribution. An understanding of the underlying probability distributions, in turn, is especially important when the goal is to forecast how the time series will behave

at future time points. When the statistical properties of the time series have been constant over time in the past, one can simply predict that they will remain constant in the future. Forecasting of time series is discussed in more detail in section 2.4.

In most practical applications, non-stationary time series are the rule rather than the exception. Luckily, by applying mathematical transformations, it is often possible to render a non-stationary time series as, approximately, stationary. This process is referred to as *stationarizing* a time series (Nau 2018). Stationarizing is used a lot in statistical forecasting methods, which are discussed in section 2.4.

### 2.2.3 Spectral density

Text

## 2.3 Time series components

### 2.3.1 Definitions

A time series can consist of various underlying patterns. Each of those patterns is considered a distinct component of the time series, with its own properties and behaviour. Splitting a time series into its components is known as *time series decomposition*. It enables a separate analysis of all the components, which helps to better understand the dynamics of a time series, but can also be useful in forecasting, as is shown later in this chapter.

Hyndman and Athanasopoulos (2018) define three main components of a time series: a trend-cycle component, a seasonal component and a remainder component. For simplicity, the trend-cycle component is usually called just the trend component, which is done in this thesis as well.

**Definition 4** The *trend component* is the combination of the trend and cyclical pattern of a time series. A trend exists when there is a long-term, not necessarily linear, increase or decrease in the data. A cyclical pattern occurs when the data exhibit rises and falls that are not of a fixed frequency. ■

**Definition 5** The *seasonal component* contains the seasonal pattern of a time series. A seasonal pattern occurs when a time series is affected by seasonal factors such as the time of the year or the day of the week. Seasonality is always of a fixed and known frequency. ■

**Definition 6** The *remainder component* is the remaining variation in a time series after the trend and seasonal components are removed. ■

There exist several different methods for the decomposition of a time series. Most of them are based on the classical decomposition method, which is discussed in the next section. A more sophisticated approach is known as STL, and is covered in section 2.2.2.3.

### 2.3.2 Classical decomposition

The oldest and simplest method for the decomposition of a time series is referred to as classical decomposition by Hyndman and Athanasopoulos (2018). They present a stepwise approach for the use of the method, which will be summarized in this section. Classical decomposition can be applied in two different forms. In the additive form, a time series is assumed to be the sum of its components, as shown in Equation 2.

$$y_t = T_t + S_t + R_t$$

In the multiplicative form, a time series is assumed to be the product of its components, as shown in Equation 3.

$$y_t = T_t \times S_t \times R_t$$

Where, for both Equation 2 and Equation 3,  $y_t$  is the data,  $T_t$  is the trend component,  $S_t$  is the seasonal component and  $R_t$  is the remainder component.

Additive decomposition is the appropriate form when the amplitude of the variation around the trend is relatively constant. On the other hand, when the amplitude of the variation around the trend changes with the level of the trend, multiplicative decomposition should be used.

In both the additive and multiplicative form of classical decomposition, the first step is to estimate the trend component. This is done by smoothing the data with a symmetric moving average filter of order  $m$ , where  $m$  is a non-negative integer. That is, the estimate of the trend component at time  $t$  is the average of all the data values within a window of  $m$  time periods centered at  $t$ , as shown in Equation 4. Usually,  $m$  is set to be equal to the seasonal period of the time series, which, in turn, is the number of observations per seasonal cycle. For example, when working with daily data that show a weekly seasonal pattern, the seasonal period will be 7.

$$\hat{T}_t = \frac{1}{m} \sum_{j=-k}^k y_{t+j}$$

Where  $k = (m - 1)/2$ . The detrended time series data are then calculated by removing the estimated trend component from the original data. In the case of additive decomposition by subtraction,  $y_t - \hat{T}_t$ , and in the case of multiplicative decomposition by division,  $y_t/\hat{T}_t$ .

The seasonal component is estimated by averaging the detrended data values per season, as shown in Equation 5. Using again the example of daily data with a weekly seasonal pattern, that would mean that the estimated seasonal component for a specific Monday is the average value of all Monday observations in the dataset, the estimated seasonal component for a specific Tuesday is the average value of all Tuesday observations in the dataset, and so on.

$$\hat{S}_t = \frac{1}{n_t} \sum_{i=1}^{n_t} (\omega_t)_i$$

Where  $\omega_t$  is a vector containing all the detrended values belonging to the same season as  $y_t$ , and  $n_t$  is the length of  $\omega_t$ . Usually, the estimated seasonal component values for each season are adjusted such that they add up to 0 in the case of additive decomposition and 1 in the case of multiplicative decomposition.

Finally, the remainder component is estimated by removing both the estimated trend component and the estimated seasonal component from the original time series. For additive decomposition, this is done by applying Equation 6.

$$\hat{R}_t = y_t - \hat{T}_t - \hat{S}_t$$

For multiplicative decomposition, Equation 7 is used.

$$\hat{R}_t = \frac{y_t}{\hat{T}_t \hat{S}_t}$$

Classical decomposition is generally praised for its simplicity, but has several disadvantages compared to some of the more modern decomposition methods (Hyndman and Athanasopoulos 2018). As a consequence of the use of a symmetric moving average filter, there are no trend component estimates available for the first few and last few observations of the time series. Therefore, also the remainder component estimate lacks these values. This is mainly problematic when forecasting, as is showed later in this chapter. Furthermore, the seasonal component stays constant over all the seasonal cycles, and cannot capture slight changes over

time. Especially when working with longer time series, this may be an inappropriate representation of the truth. Finally, classical decomposition is not robust to extreme values in the data.

### 2.3.3 STL decomposition

A widely used method that is based on classical decomposition, but deals with many of the limitations mentioned above, is known as STL. It stands for *A Seasonal-Trend decomposition procedure based on Loess*, and was developed by R. B. Cleveland et al. (1990). In this section, their methodology will be summarized. STL estimates all three components for every observation in a time series, and can also handle missing values in the data. Both the trend and seasonal component are robust and not distorted by extreme values. Furthermore, the seasonal component is not fixed, but can vary slightly over time.

As its name already implies, STL is based on loess, also known as locally-weighted regression. Loess was developed by Cleveland and Devlin (1988), and is a non-parametric regression technique, often used for smoothing, that fits weighted least squares regression curves to local subsets of a dataset. Joining them together forms the loess regression curve  $\hat{g}(x)$ . More specifically, for each value of  $x$ ,  $\hat{g}(x)$  is computed in the following way. First, a positive integer  $q$  is chosen, which defines the neighborhood width. That is, the  $q$  observations that are closest to  $x$  are selected as neighbors of  $x$ . Each of these observations is given a weight based on its distance to  $x$ , in a way that the closest observations get the highest weight. Let  $W$  be the tricube weight function as defined in Equation 8.

$$W(u) = \begin{cases} (1 - u^3)^3 & 0 \leq u < 1 \\ 0 & u \geq 1 \end{cases}$$

Then, the neighborhood weight for each observation  $x_i$  is calculated with Equation 9.

$$v_i = W\left(\frac{|x_i - x|}{\lambda_q(x)}\right)$$

Where  $\lambda_q(x)$  is the distance of the  $q$ th farthest observation from  $x$ . Then,  $\hat{g}(x)$  is calculated by fitting a polynomial regression of degree  $d$  to  $x$ , using weighted least squares with the neighborhood weights  $v_i$ . Usually,  $d$  is either 1 or 2, corresponding respectively to a locally-linear regression and a locally-quadratic regression. Since the loess regression curve is smooth, there is no need to compute  $\hat{g}(x)$  at all possible values of  $x$ . In general, the computation of  $\hat{g}(x)$  as described above is only performed at a finite set of locations, and interpolated elsewhere.

STL uses loess for several smoothing operations, that, when performed on a time series, lead to estimations of the trend, seasonal and remainder components of the data. The method is build up of two loops: an inner loop nested inside an outer loop. In the inner loop, the estimates of the seasonal and trend component are updated once, in a stepwise manner, which will be described below.

**Step 1.** The inner loop starts with computing the detrended time series data  $y_t - \hat{T}_t$  from the original time series data  $y_t$ . In the initial pass through the inner loop, there is no estimation of  $T_t$  yet, and  $\hat{T}_t$  is set equivalent to 0. That is, it is assumed there is no trend at all. This may be a rather poor estimate, but inside the loop, it will soon be updated to something more reasonable. In all successive passes through the loop, the estimated trend component that resulted from the previous loop is used.

**Step 2.** In the second step, the detrended time series is split up into subsets, with each subset containing all the data belonging to one specific season. That is, there will be  $n_p$  different subsets, where  $n_p$  is the number of observations per seasonal cycle. Each of those subsets is smoothed by loess, with  $q = n_s$  and  $d = 1$ .  $n_s$  is referred to as the seasonal smoothing parameter and its value must be chosen by the analyst. It basically determines how much the seasonal component is allowed to change over time. High values of  $n_s$  allow little variation, while low values can lead to overfitting. The smoothed values of all the subsets are then binded

back together into a temporary seasonal component  $C_t$ . Each end of  $C_t$  is extended  $n_p$  positions, such that  $C_t$  has  $2 \times n_p$  observations more than the original time series.

**Step 3.** In the third step, any trend that may have contaminated  $C_t$  is identified. This is done by applying a sequence of smoothers, called a low-pass filter, to  $C_t$ . It starts with a moving average of length  $n_p$ , followed by another moving average of length  $n_p$ , followed by a moving average of length 3, followed by a loess smoothing with  $q = n_l$  and  $d = 1$ . Just as earlier with  $n_s$ , the low-pass filter smoothing parameter  $n_l$  should be chosen by the analyst. The output of the third step is called  $L_t$ . Since moving averages are used, the first  $n_p$  observations and the last  $n_p$  observations of  $C_t$  will not have a smoothed value in  $L_t$ . However, this was already accounted for by extending  $C_t$  in step 2. That is,  $L_t$  is again of the same length as the original time series.

**Step 4.** In the fourth step, the seasonal component is estimated by detrending the temporary seasonal component. That is,  $\hat{S}_t = C_t - L_t$ .

**Step 5.** In the fifth step, the deseasonalized time series data  $y_t - \hat{S}_t$  are computed from the original time series data  $y_t$ .

**Step 6.** In the sixth and last step of the inner loop, the estimation of the trend component,  $\hat{T}_t$ , is calculated by loess smoothing the deseasonalized time series with  $q = n_t$  and  $d = 1$ . The trend smoothing parameter  $n_t$  should be chosen by the analyst.

The outer loop of STL starts with  $n_i$  iterations of the inner loop. The estimations of the trend and seasonal components that follow from the passes through the inner loop, are used to estimate the remainder component with Equation 10.

$$\hat{R}_t = y_t - \hat{T}_t - \hat{S}_t$$

For each observation in the time series, a robustness weight is calculated. This weight reflects how extreme the value of the remainder component of that observation is, in a way that a very extreme value will be given a very low, or even zero, weight. Let  $B$  be the bisquare weight function as defined in Equation 11.

$$B(u) = \begin{cases} (1 - u^2)^2 & 0 \leq u < 1 \\ 0 & u \geq 1 \end{cases}$$

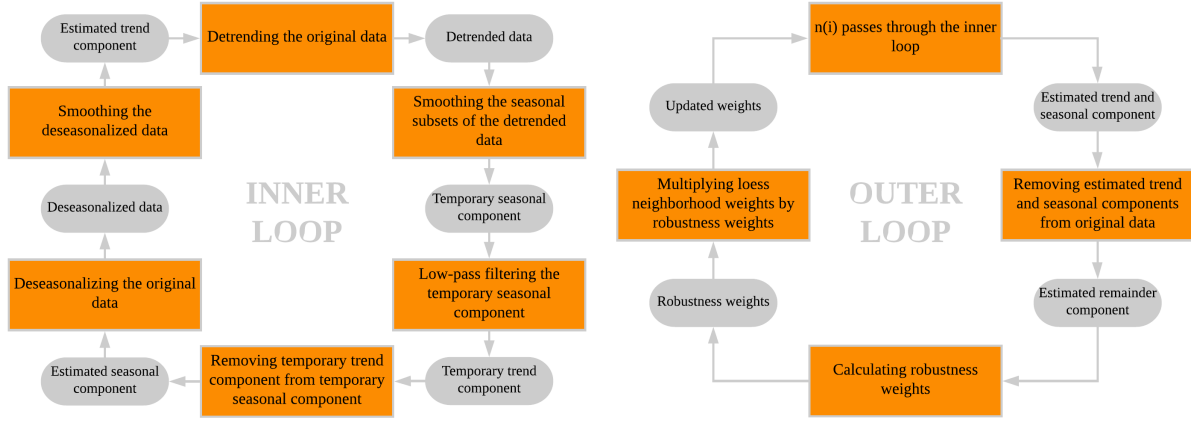
Then, the robustness weight at time point  $t$  is calculated with Equation 12.

$$\rho_t = B\left(\frac{|R_t|}{6\text{median}(|R_t|)}\right)$$

After the first pass of the outer loop, the next iteration starts again with  $n_i$  passes through the inner loop. However, in the loess smoothing in step 2 and step 6, each neighborhood weight  $v_i$  is now multiplied by its corresponding robustness weight  $\rho_t$ , such that extreme values have less influence on the estimates of the trend and seasonal components. Also, the estimated trend component that resulted from the last inner loop in the previous outer loop, is now used as first value of  $\hat{T}_t$ , rather than 0. In total, the outer loop will be carried out  $n_o$  times.

STL is designed for additive decomposition. However, a multiplicative version can be obtained by first log transforming the data, and finally back-transforming the components. This is based on the logarithm product rule, which states that  $\log(a) + \log(b)$  is equivalent to  $\log(a \times b)$  (Hyndman and Athanasopoulos 2018).

The complete methodology of STL as described above is summarized in Figure 1.



## 2.4 Time series forecasting

### 2.4.1 Forecasting models

Often, the main aim of time series analysis is to forecast future values of a time series. In some cases, this can be done by using external exploratory variables. One could for example try to forecast the profit of ice cream sales by using air temperature as an exploratory variable in a linear regression model. However, there are several reasons not to forecast time series in this way, as summed up by Hyndman and Athanasopoulos (2018). Firstly, the underlying system of the forecasted time series may not be sufficiently understood, and even if it is, the relations with exploratory variables may be too complex. Secondly, when forecasting future values of a time series, also the future values of the exploratory variables should be known, which means that each exploratory variable should be forecasted separately before the response variable can be forecasted. This may be too difficult to do accurately, and even when it is possible, it remains a very time consuming task. Especially when the only aim is to know what will happen, and not why it will happen, it is not worth the effort. Finally, modeling a time series with conventional statistical method like linear regression will likely result in model errors that exhibit autocorrelation, which implies that such models are not able to capture all the dynamics of the data. Thus, produced forecast are not as efficient, and, probably, not as accurate as can be.

Instead, in time series analysis, the internal dependence structure of a time series is used to forecast future values as a function of the current and past values (Shumway and Stoffer 2011). Obviously, this primarily requires a good understanding of that structure, which is obtained by describing the data with a time series model, as defined in Definition 7, adapted from Brockwell and Davis (2002).

**Definition 7** A *time series model* for an observed realization  $\{y_t\}$  of a time series  $\{Y_t\}$  is a specification of the joint distributions, or possibly only the means, variances and covariances, of the random variables that  $\{Y_t\}$  comprises. ■

One of the most famous and widely used group of time series models is known as the *autoregressive integrated moving average* (ARIMA) models, developed by Box and Jenkins (1970). In this thesis, ARIMA is used as well. The next section gives an extensive summary of its theory, based on Shumway and Stoffer (2011), Chapter 3, and Hyndman and Athanasopoulos (2018), Chapter 8.

## 2.4.2 ARIMA

### 2.4.2.1 Structure

An ARIMA model is a combination of an *autoregressive* (AR) and *moving average* (MA) model, preceded by a differencing operation on the original data. An autoregressive model of order  $p$ , commonly referred to as an  $AR(p)$  model, is based on the assumption that the current value of a time series is a linear combination of  $p$  previous values, as showed in Equation 13.

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t$$

Where  $y_t$  is the current value of the time series at time period  $t$ ,  $\epsilon_t$  is the random error (i.e. white noise) at time  $t$ ,  $\phi_1, \dots, \phi_p$  are model parameters and  $c$  is a constant.  $c = 0$  whenever  $\mu = 0$ , given that  $\mu$  is the mean of  $y_t$ . If  $\mu \neq 0$ ,  $c = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p)$ .

A moving average model of order  $q$ , commonly referred to as an  $MA(q)$  model, is based on the assumption that the current value of a time series is a linear combination of  $q$  previous forecast errors, as showed in Equation 14.

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

Where  $y_t$  is the current value of the time series at time period  $t$ ,  $\epsilon_t$  is the forecast error at time period  $t$ , which is assumed to be white noise,  $\theta_1, \dots, \theta_q$  are model parameters.

$AR(p)$  and  $MA(q)$  models can be combined into an autoregressive moving average model of order  $(p, q)$ , commonly referred to as  $ARMA(p, q)$ . That is, in such a model, the current value of a time series is a linear combination of both  $p$  previous values and  $q$  previous forecast errors, as showed in Equation 15.

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t$$

$ARMA(p, q)$  models require the forecasted time series to be stationary. When working with non-stationary time series, it is often possible to stationarize the series by differencing it one or more times. The first order difference of a time series is the series of changes from one time period to the next, as shown in Equation 16.

$$\nabla y_t = y_t - y_{t-1}$$

Where  $\nabla y_t$  is the first order difference of  $y_t$ . When the first order difference is still non-stationary, the second order difference  $\nabla^2 y_t$  can be computed by taking again the first order difference of  $\nabla y_t$ , and so on. The original non-stationary time series that needed to be differenced in order to get stationary, is called an *integrated* version of the stationary series. That is why a model that first stationarizes the data by applying a  $d$ -th order difference, before fitting an  $ARMA(p, q)$  model, is called an autoregressive integrated moving average model of order  $(p, d, q)$ , commonly referred to as  $ARIMA(p, d, q)$ . That is, in such a model, the current value of the  $d$ -th order difference of a time series is a linear combination of both  $p$  previous values and  $q$  previous forecast errors, as showed in Equation 17.

$$y_t = c + \phi_1 \nabla^d y_{t-1} + \dots + \phi_p \nabla^d y_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t$$

Where  $\nabla^d y_t$  is the  $d$ -th order difference of  $y_t$ . Note here that  $ARIMA(p, d, q)$  is a general form of all the other models discussed earlier in this section. For example, an  $AR(1)$  model can also be written as  $ARIMA(1,0,0)$ , an  $ARMA(2,1)$  as  $ARIMA(1,0,2)$ , and so on.

The process of finding an appropriate  $ARIMA(p, d, q)$  model that represents an time series is known as the Box-Jenkins modelling procedure and consists of three stages, named model selection, parameter estimation and model checking. All these stages are described separately in the next three subsections.



#### 2.4.2.2 Model selection

#### 2.4.2.3 Parameter estimation

#### 2.4.2.4 Model checking

#### 2.4.2.5 Forecasting

#### 2.4.2.6 Seasonal ARIMA models

### 2.4.3 Complex seasonality

### 2.4.4 Transformations

## 2.5 Time series clustering

Box, George E. P., and Gwilym M. Jenkins. 1970. *Time Series Analysis: Forecasting and Control*. Book. Holden-Day San Francisco.

Brockwell, Peter J., and Richard A. Davis. 2002. *An Introduction to Time Series and Forecasting*. Vol. 39. doi:10.1007/978-1-4757-2526-1.

Cleveland, Robert B., William S. Cleveland, Jean E. McRae, and Irma Terpenning. 1990. “STL: A Seasonal-Trend Decomposition Procedure Based on Loess.” *Journal of Official Statistics* 6 (1): 3–73.

Cleveland, William S., and Susan J. Devlin. 1988. “Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting.” *Journal of the American Statistical Association* 83 (403). Taylor & Francis: 596–610. doi:10.1080/01621459.1988.10478639.

Hyndman, Rob J., and George Athanasopoulos. 2018. *Forecasting: principles and practice*. OTexts. <https://otexts.org/fpp2/>.

Nau, Robert. 2018. “Stationarity and differencing.” <https://people.duke.edu/~rnau/411diff.htm>.

Shumway, Robert H., and David S. Stoffer. 2011. *Time Series Analysis and Its Applications With R Examples*. Vol. 9. doi:10.1007/978-1-4419-7865-3.

Woodward, Wayne A., Henry L. Gray, and Alan C. Elliott. 2017. *Applied Time Series Analysis with R*. CRC Press.