Evaluating the Performance of Time Series Models for Forecasting Electricity Consumption in Finland

Johannes Viherä
Master’s Thesis in Information Systems
Supervisor: Markku Heikkilä
Faculty of Social Sciences, Business and Economics
Åbo Akademi University
Turku 2023

Abstract

Subject: Information Systems	
Writer: Johannes Viherä	
Title: Evaluating the Performance of Time Series Models for Forecasting Electricity Consumption in Finland	
Supervisor: Markku Heikkilä	
Abstract: The electricity market in Finland is in a surrounding unseen in decades. This raises the question about the performance of forecasting tools in an unpredictable environment. This study aims to investigate the performance of autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average (SARIMA) models in forecasting electricity consumption in Finland. A literature review was conducted to provide a comprehensive understanding of time series analysis and the application of the selected models. The results revealed that SARIMA models generally provided higher accuracy in forecasting electricity consumption, with accuracy ranging from 90.7% to 95.5%. In contrast, ARIMA models had an accuracy of between 88.7% to 94.3%. The SARIMA models were more successful in predicting fluctuations in consumption, while ARIMA models were better suited for datasets with reduced noise.	
Keywords: Electricity consumption, Time series analysis, Weighted moving average, Autoregressive models, ARMA, ARIMA, SARIMA, Box-Jenkins approach	
Date: xx.4.2023	**Number of pages**: 79

Contents

1 Introduction

This chapter provides a succinct overview of the topic, explains its relevance, and outlines the structure and objectives of this thesis. Additionally, it presents the research questions and the methods employed to address them.

1.1 Introduction

In February 2022, Russia initiated an attack on Ukraine, serving as the final catalyst for the escalation of energy prices in Europe. In the ensuing weeks, Western nations began isolating Russia from international trade. Given that Russia accounted for 40% of Europe's imported gas demand in 2021 (IEA, 2022), inflation soared to unprecedented levels.

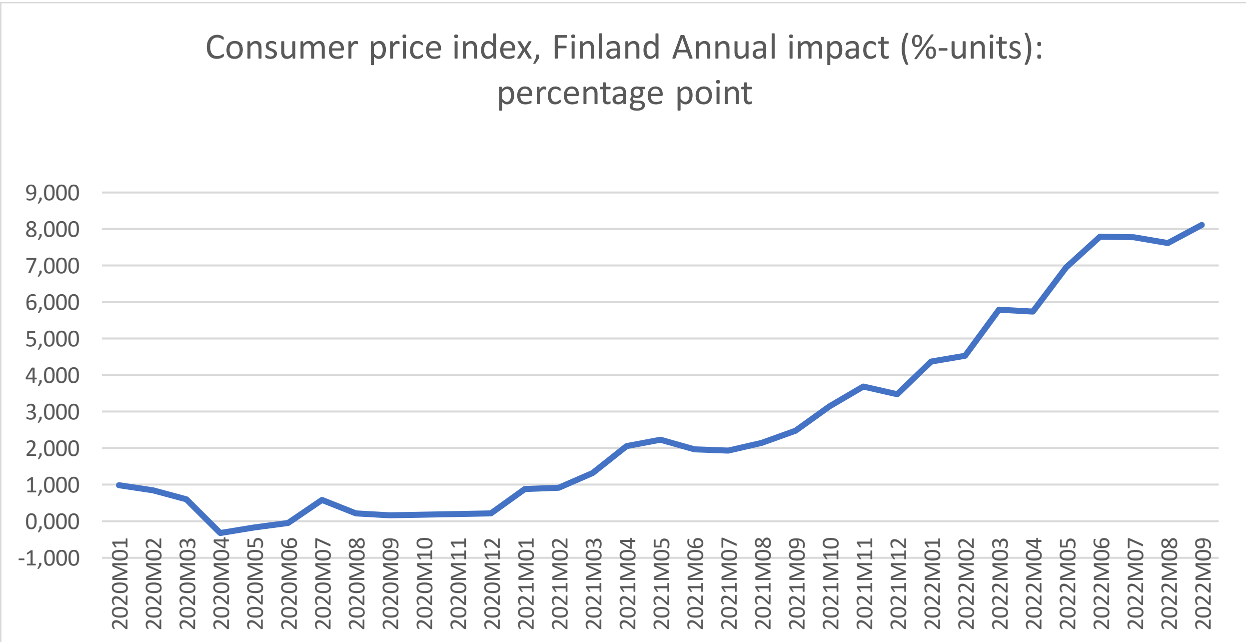

Figure 1: Consumer price index Finland (Tilastokeskus 2022)

Apart from petrol prices, the cost of electricity has also received widespread media attention. Consumers have faced challenges due to abrupt increases in electricity bills, while electricity providers have suffered losses due to contracts signed prior to the unexpected shift in production. Consequently, electricity providers have raised their prices. For instance, Helen, a Helsinki-based energy company, implemented an average price increase of 58% for their permanent contracts in August 2022. (Lassila, 2022)

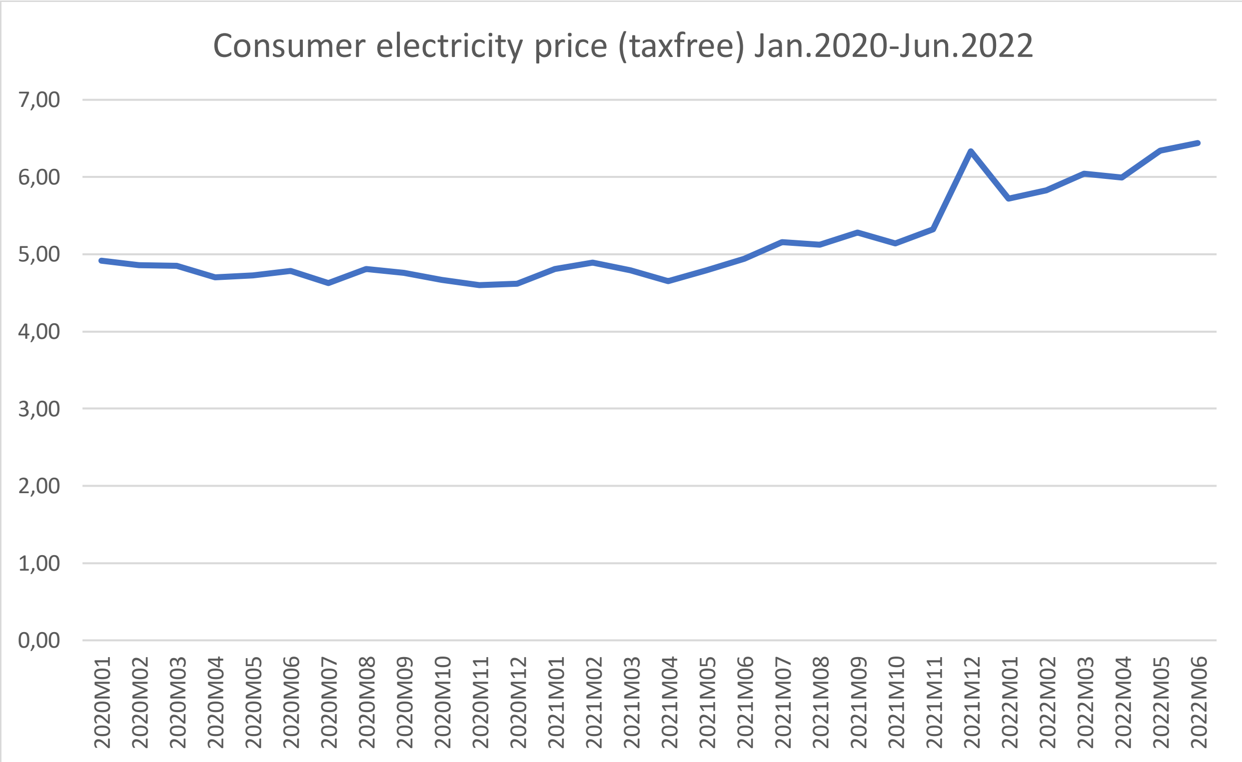

Figure 2: Consumer Electricity Price (Tilastokeskus, 2022)

The current state of the European electricity market is distinctive, resulting from a combination of already surging prices and a sudden energy shortage. Electricity consumption in Finland exhibits a seasonal pattern, primarily due to the high energy demand for heating during winter. An intriguing aspect of this situation is whether the soaring prices will compel individuals and companies in Finland to reduce their energy consumption or whether electricity usage will remain at similar levels as in previous years.

1.2 Objective

This study aims to provide insights into the performance of models in a unique situation that is not necessarily predictable by trends. While these models have been adapted for forecasting electricity prices, consumption, and production, the recent market situation and its impact on model performance has yet to be evaluated. To achieve this research objective, the thesis will address the following research questions:

1. *Which of the Time series models give the better accuracy forecasting the electricity consumption?*
2. *Which of the models suit the dataset better?*

1.3 Method

The thesis will conduct a literature review on the theory underpinning time series analysis and the methods related to forecasting using time series models. The literature review will commence with a broader focus on time series analysis theory before narrowing down to specific models selected and their practical application in the empirical study conducted after the literature review.

Both research questions will be answered in the empirical study, where autoregressive integrated moving average (ARIMA) models and seasonal autoregressive integrated moving average (SARIMA) models will be compared in terms of their forecasting performance. Forecasting models will be developed, and data will be preprocessed using Jupyter Notebook, which employs Python as the programming language. The models will utilize Finnish electricity consumption data. The study will compare the two models based on the accuracy of their forecasts for future electricity consumption.

1.4 Structure of thesis

The forthcoming thesis will be organized as follows: A literature review will first be conducted, focusing on the theory of time series analysis and the forecasting methods employed in the empirical study. The methods will be elucidated, the data used in the study will be described, and the results will be discussed. Upon presenting the study's findings, potential avenues for future research will be explored.

2 Theory

This chapter aims to establish a theoretical framework for time series analysis, addressing the process of developing models and generating forecasts. The chapter commences with an introduction to the fundamental concepts of time series and their characteristics, followed by a discussion of forecasting methods and their practical applications.

2.1 Time Series

A time series is defined as observations arranged chronologically (Kirchgässner, Wolters & Hassler, 2012). Cryer (1986) posits that the objective of time series analysis is to comprehend or model the mechanism responsible for the occurrences within the time series or to analyze quantitative data for the purpose of generating predictions or forecasting future values of the series.

Persons (1919) identifies four components of time series:

1) Trend – Long-term development.
2) Cycle – A cyclical component with periods exceeding one year.
3) Seasonal Cycle - A cyclical component with periods within a year.
4) The residual – Seemingly random values that do not belong to any of the cycles.

2.1.1 Trend

In time series analysis, measurements are frequently taken at regular intervals, such as yearly, monthly, daily, or hourly. Although values in a time series can generally be random, the possibility of a shift to relatively higher or lower values over time exists. This change in the time series is referred to as a trend, which typically results from alterations in factors surrounding the values. (Anderson, Sweeney, and Williams, 1999)

Anderson et al. (1999) delineate various trend patterns:

1) Linear trend
2) Nonlinear trend
3) Linear declining trend
4) No trend

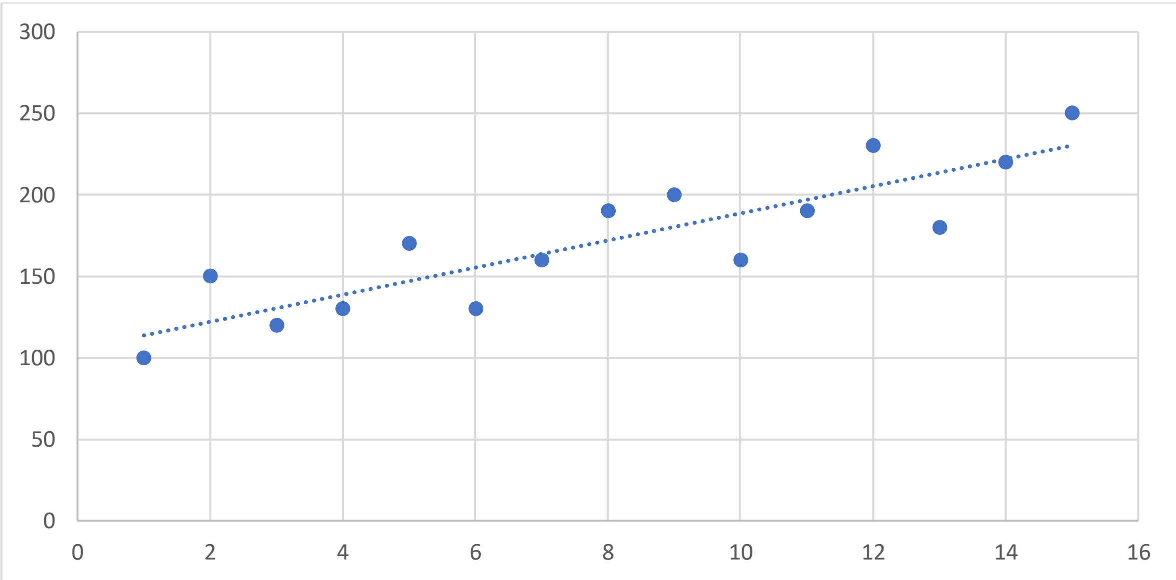

Figure 3: Linear Trend

In figure 3, the fluctuation in values on the y-axis is irregular, but as we progress further on the x-axis, the values are higher. This gradual growth in the values is an upward trend in the time series. This trend is visualized with a trendline in the figure.

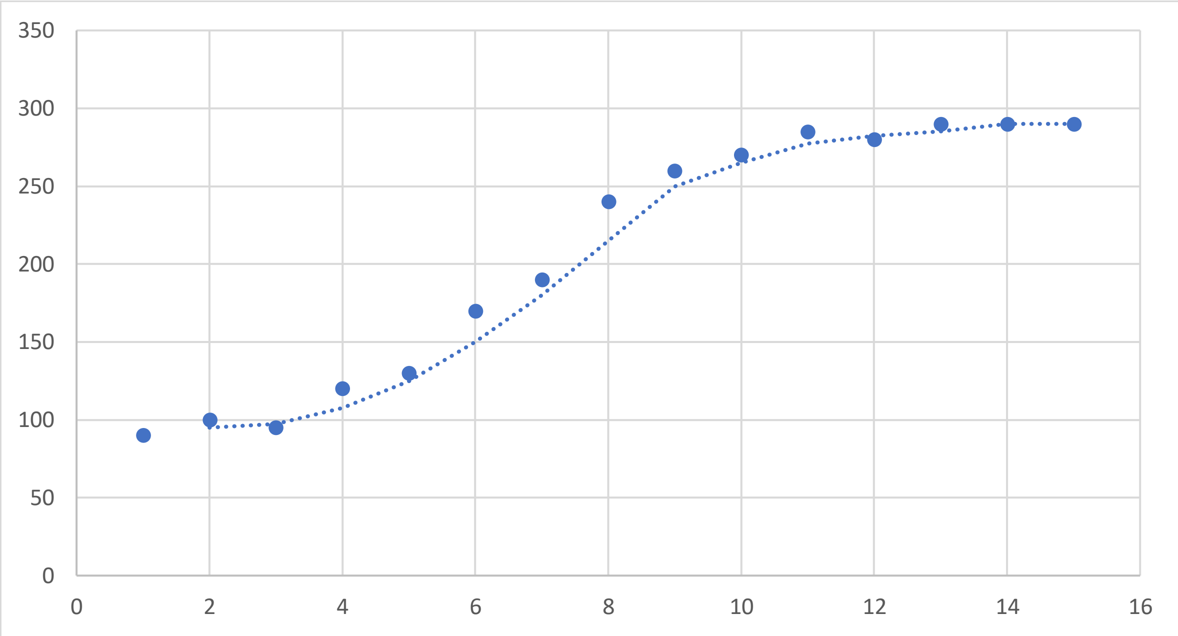

Figure 4: Nonlinear Trend

Figure 4 demonstrates a Nonlinear trend. The trend consists of steady growth, which is followed by a phase of rapid increase in values and a leveling at the end of the trend.

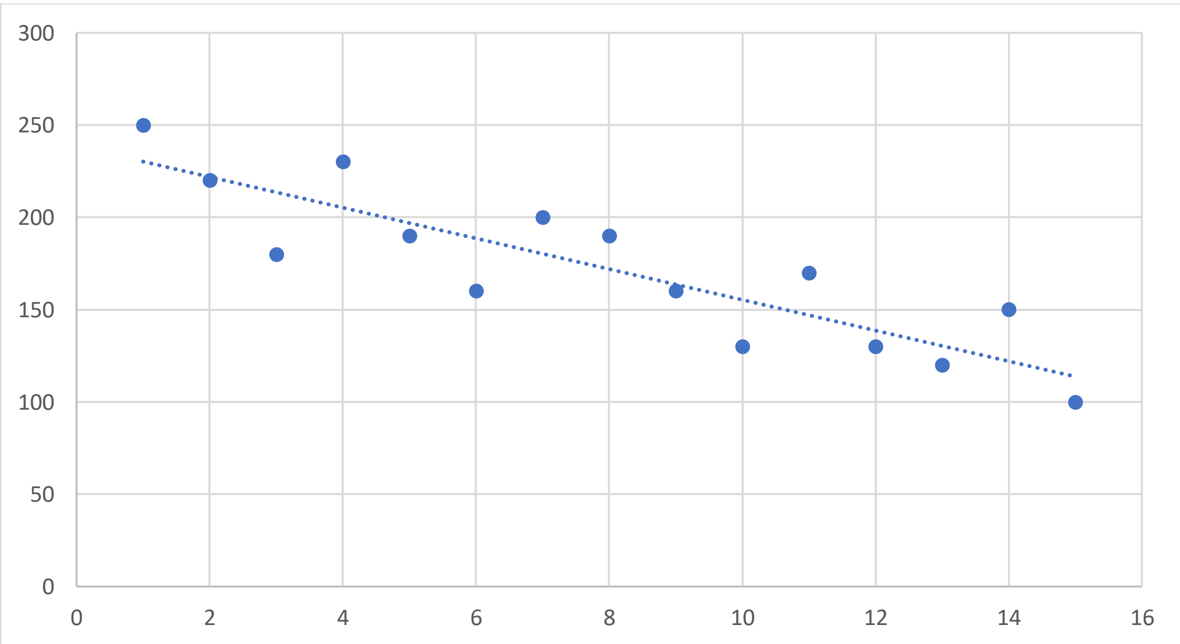

Figure 5: Linear Declining Trend

Anderson et al. (1999) point out that trends can also be declining.

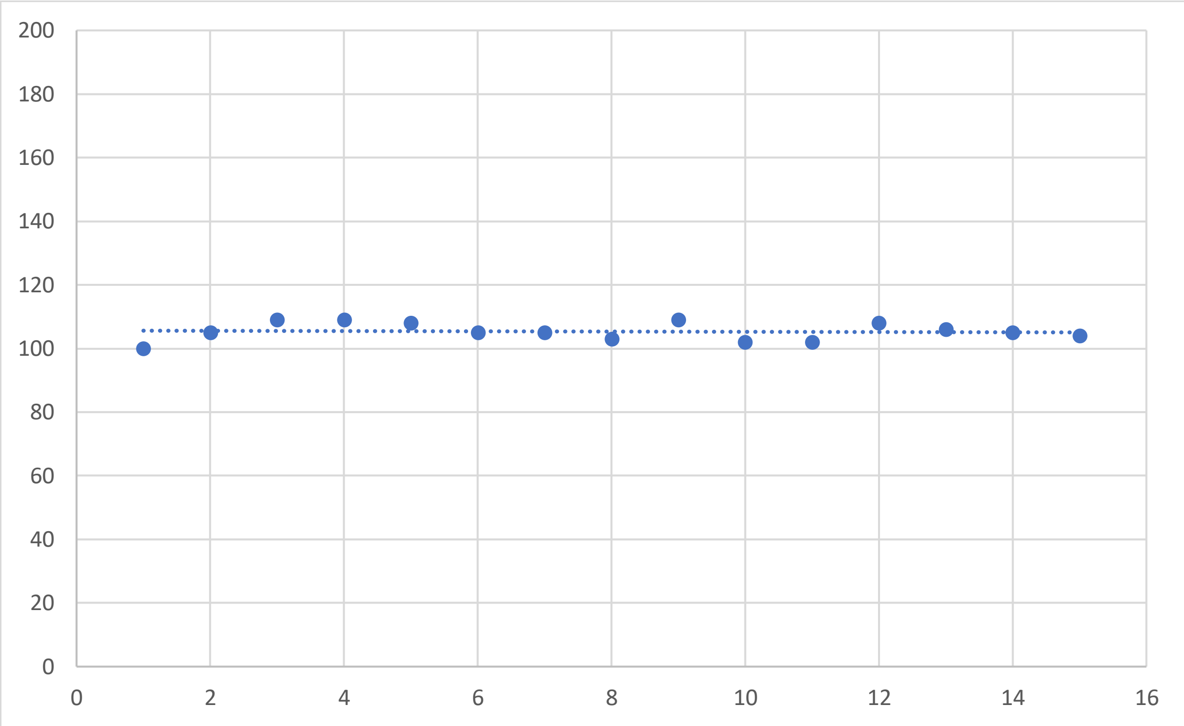

Figure 6: No Trend

It is also possible that the data has no trend. The horizontal trendline visualizes the missing trend in figure 6.

2.1.2 Cycle

Although a trend may be discernible over an extended period, most data points do not precisely align with the trend line. The data generally fluctuates and appears on both sides of the trendline. If the data exhibits a recurring sequence exceeding a calendar year, it can be considered a cyclical component (Anderson et al., 1999)

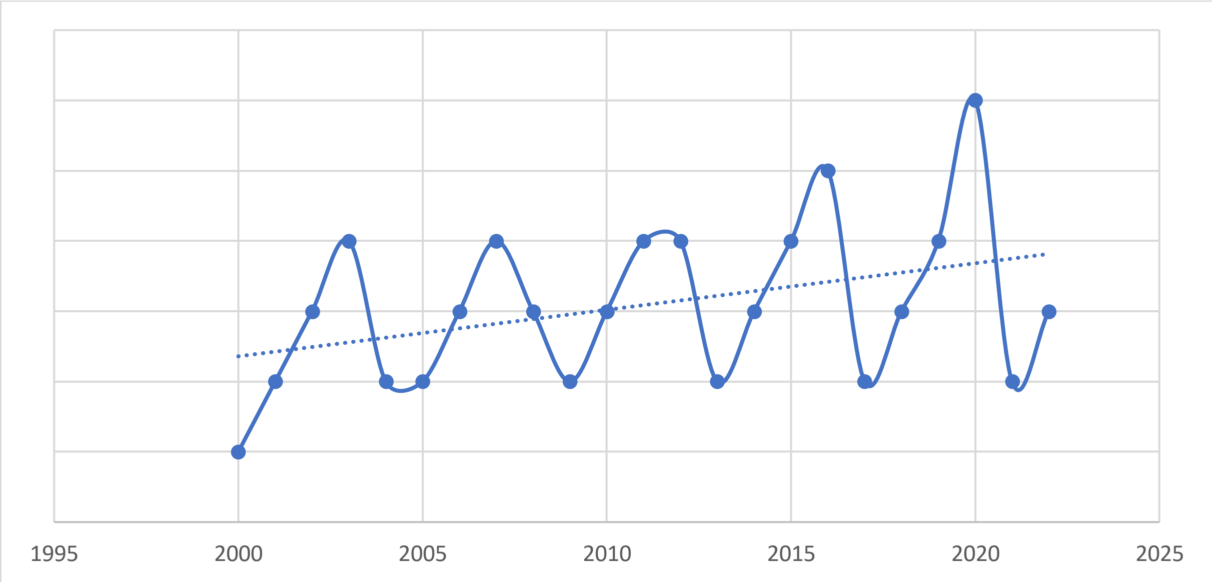

Figure 7:Trend and Cyclical Components of a Time Series

Numerous time series exhibit cyclical components. Anderson et al. (1999) cite economic data as an example of a time series that often contains multi-year-lasting fluctuations in variables. For instance, if the inflation rate oscillates between moderate and high, several time series depicting prices or price variation have an increasing trend line with values regularly fluctuating above and below the trend line. (Anderson et al., 1999)

2.1.3 Seasonality

In contrast to the cyclical component of time frames, which is identifiable as regular variation over a multi-year time frame, time series may demonstrate a consistent pattern over one-year periods. As an example, Anderson et al. (1999) name the sales volume of seasonally used products, such as swimming pool equipment which reaches its peak sales during spring and summer. The component displaying seasonal fluctuation is named the seasonal component. Anderson et al. (1999) point out that seasonality can occur on a monthly, weekly, or hourly basis.

2.1.4 The residual

The residual, or irregular component, accounts for values unaffected by trend, cycle, or seasonality. Often referred to as noise, irregular values result from short-term, unforeseen, and nonrecurring factors influencing the time series. Owing to the random characteristics of these values, they are unforeseeable and fall outside the scope of predictions made for the time series. (Anderson et al., 1999)

2.2 Time Series Analysis

Time series analysis entails the process of forecasting or predicting future values within a time series. Forecasting methods can be classified as quantitative or qualitative. Quantitative methods can be employed if:

1. Past information of the variable being predicted is available
2. The information available can be quantified
3. One can assume that the pattern recognizable will continue in the future

If all the necessary prerequisites are satisfied, it is possible to generate forecasts or predictions of future values utilizing either a time series approach or a causal approach (Anderson, Sweeney, & Williams, 1999). In cases where historical data are solely comprised of past values, the analytical process is referred to as a time series method. This method aims to discern patterns within the data and exploit them to produce future value predictions. Time series methods exclusively base their predictions on historical values and the forecasts derived from them (Anderson et al., 1999).

Causal forecasting techniques presuppose the existence of a cause-and-effect relationship between one or more variables. As an illustration, Anderson et al. (1999) cite the connection between advertising expenditures and sales volume, with the former typically exhibiting a positive correlation with the latter, meaning that increased advertising leads to higher sales. This relationship can be characterized by a regression, which can subsequently be employed to forecast future sales volumes.

Qualitative forecasting methods depend on expert opinion. The judgment and expertise of individuals well-versed in a specific field can be harnessed to generate forecasts. The advantage of qualitative forecasting techniques lies in their applicability to variables that are not quantifiable or for which historical data are either unavailable or ill-suited to the task. (Anderson et al., 1999)

2.3 Time series forecasting models

Forecasting is essential in addressing various practical issues, such as budgeting for staff and resources, making investment decisions, and determining appropriate attire for the following day. Hyndman and Athanasopoulos (2018) state that some phenomena are easier to predict than others, and the predictability of a specific event or quantity depends on several factors, including:

1. The level of understanding of the factors contributing to the event or quantity.

2. The amount of data available.
3. The extent to which forecasts influence the object being forecasted.

Hyndman et al. (2018) provide the example of electricity demand forecasting. Demand is primarily driven by weather and other factors with smaller contributions to values. Ample data is available for factors such as wind speed and temperature, as well as historical electricity demand. Moreover, the technology and expertise required to develop models linking demand to key driver variables yield generally accurate forecasts.

Forecasting conditions can vary significantly in terms of time horizons, determining factors, data patterns, and other aspects. Methods can range from simple, like the naïve method using the most recent observation as a forecast, to complex, such as neural networks. The choice of method depends on the available data and the predictability of the forecast subject. (Hyndman et al., 2018)

2.3.1 Forecast accuracy

Evaluating forecast accuracy is crucial when selecting a model. Hyndman et al. (2018) define a forecast error as the difference between an observed value and a forecasted value. In this context, "error" does not imply a mistake but rather an unpredictable component of an observation. Anderson et al. (1999) describe the forecasting error using Equation 2.1:

$$Error = Observed\ Value - Forecast$$

Equation 2.1

According to Hyndman et al. (2018), forecast errors should be measured on the same scale as the data to avoid scale-dependent accuracy methods that cannot be used for comparison across series with different units of measurement.

Hyndman et al. (2018) present the mean absolute error (MAE) as a commonly used error measure, given by Equation 2.2:

$$MAE = mean(|e_t|)$$

Equation 2.2

The MAE is popular for comparing forecast methods for a single time series or multiple series with similar units due to its simplicity and ease of computation.

Anderson et al. (1999) introduce the mean squared error (MSE) as another approach for measuring forecast accuracy, defined by Equation 2.3:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_{i-x_i})^2$$

Equation 2.3

Where n is the number of observations, y_i is the observed value at time i and x_i is the predicted value at time i.

Percentage errors, which are independent of the units in which the data is measured, are used to compare performance between data sets with different units of measurement. The percentage error p_t is given by $p_t = 100e_t/y_t$, where y_t is the measure at time t. The most commonly used measure is mean absolute percentage error:

$$MAPE = mean(|p_t|)$$

Equation 2.4

The measures that are based on percentage errors get infinite or undefined values if the measure is 0, or alternatively extreme values if some of the measures are close to zero. Further disadvantages with percentage-based error include that they often assume that the unit of measurement has a meaningful zero. For example, in cases of measures on the interval scale, such as temperature, the use of percentage error is not recommended. (Hyndman et al., 2018)

Relative error measures seek to remove the scale of data, which can be problematic in some cases. Relative errors have a statistical distribution with undefined mean and infinite variance, and they can only be computed when multiple forecasts exist for the same time series. (Hyndman & Koehler, 2006)

Hyndman and Koehler (2006) introduce a measurement of error that circumvents the problems associated with relative error measurements. By scaling the error using the values derived from the Mean Absolute Error (MAE) in the forecast method, the definition of scaled error is established as:

$$q_t = \frac{e_t}{\frac{1}{n-1}\sum_{i=2}^{n}|Y_i - Y_{i-1}|}$$

Equation 2.5

A value less than one for q_t means that the forecast is at a more precise level compared to the average values obtained from forecast method in MAE. If the value is greater than one, the forecast can be regarded as worse than the average values obtained from the forecast method in MAE.

The mean absolute scaled error is calculated by:

$$MASE = mean(|q_t|)$$

Equation 2.6

When MASE is less than 1, the method produces smaller errors than the forecast with values obtained for the measure. (Hyndman, 2006)

Hyndman and Koehler (2006) propose employing scaled errors as a standard technique for comparing forecast accuracy across different scales. This approach has a meaningful scale, is not susceptible to previously mentioned issues, and is widely applicable. Hyndman and Koehler (2006) note that the only situation leading to undefined or infinite values for scaled errors would occur if all observation points were equal. Regression models often use R^2 as a measure of model goodness, representing the proportion of variation in values explained by the regression model. (Colton & Bower, 2002)

2.3.2 Smoothing methods

Anderson et al. (1999) introduce three forecasting methods: moving averages, weighted moving averages, and exponential smoothing. These methods share the common feature of eliminating fluctuations in data caused by irregular components in time series, and thus are referred to as smoothing methods. These methods are optimal for stable or stationary time series, which exhibit no significant trend or cyclical/seasonal fluctuations in measures. Smoothing methods are straightforward to use and generally provide high accuracy for short-range forecasts, such as predictions for the subsequent period.

2.3.2.1 Moving Averages

The moving averages method employs the average value of the most recent data values in a time series to predict the next period:

$$Moving\ average = \frac{\sum(most\ recent\ n\ data\ values)}{n}$$

Equation 2.7

The term "moving" arises from each new observation replacing the oldest one in the average calculation, causing the average to move forward as new observations are added to the time series (Anderson et al., 1999). For the moving averages method, the number of data values (n) included in the moving average must be chosen. Moving averages with different numbers of values vary in their forecasting accuracy. Anderson et al. (1999) suggest approaching the problem of determining the number of values in the moving average through trial and error to identify the length of the moving average that minimizes errors.

2.3.2.2 Weighted Moving Averages

The simple moving average includes an assumption about equal weights for the values in the calculation, whereas weighted moving average involves determining varying weights for each of the calculation values. After determining the weights for the calculation, an average for n values is calculated and used as a forecast. Generally speaking, the weight for the most recent value is the highest, and the older the value is, the less the weight the value receives. The sum of weights should be equal to 1. (Anderson et al., 1999)

According to Anderson et al. (1999), if we believe that the most recent value can be regarded as a better predictor of the future compared to the oldest one in the calculation, it should be given the largest weight. Moving forward in the calculation, the weight should gradually decrease while going backwards in values selected for the calculation. Anderson et al. (1999) point out that if the time series is highly variable, the optimal solution would be to select approximately equal weights for the values.

2.3.2.3 Exponential Smoothing

Exponential smoothing has minimal data requirements and is therefore recommended as a method when a forecast is required for a large number of items. The method uses a weighted average of past values in the time series as in the smoothing methods previously presented. The difference between exponential smoothing and weighted moving averages is that instead of multiple weights, it utilizes only one weight for the most recent observation. The weights for other data values are determined automatically by the formula and they become smaller as the observations move to the past. The formula for the basic exponential smoothing model is presented below:

$$F_{t+1} = \alpha Y_t + (1 - \alpha) F_t$$

Equation 2.8

where:

F_{t+1} = forecast of the time series for period t + 1

Y_t = actual value of time series in period t

F_t = forecast of the time series for period t

α = smoothing costant (0 ≤ α ≤ 1)

The equation demonstrates how the forecast for period t + 1 is determined by the weighted average of the actual value and forecast for period. The weight given for the actual value in period t is α and the weight for the forecast in period t is 1- α. Anderson et al. (1999)

2.3.2.4 Smoothing methods in practice

Hansun and Kristanda (2017) examined the forecasting efficacy of simple moving averages (SMA), weighted moving averages (WMA), and exponential moving averages (EMA) in foreign exchange transactions. Their investigation encompassed three currency pairs: EUR/USD, AUD/USD, and GBP/USD, with mean squared error (MSE), mean absolute percentage error (MAPE), and mean absolute scaled error (MASE) employed to compare the performance of the models.

Their findings indicated that EMA was the most suitable method for Forex data, as it exhibited the lowest MSE, MAPE, and MASE values across all currency pairs. WMA ranked second in accuracy across all error measure types, while SMA demonstrated the poorest performance. Nonetheless, Hansun and Kristanda (2017) asserted that all models performed adequately in the test.

Reghunath and Raghavan (2005) showcased the application of simple moving averages in their research on water resources in the Nethravathi River region of southern India. They utilized simple moving averages to filter the data on monthly water levels in the area's observation wells to reveal concealed long-term trends in water table fluctuations. A 12th-order moving average was employed in their study.

Lauren and Harlili (2014) integrated simple moving averages with machine learning models to forecast trend patterns in stock closing prices. Their investigation utilized moving averages with durations of 5 and 10 days for short-term averages, and 20 and 50 days for long-term averages. Various combinations of these durations were tested in the experiment, with the 5 and 20-day combination yielding the most favorable outcomes and the highest likelihood of generating profit in stock trading.

Moving averages are extensively employed in the technical analysis of stocks. Detry and Gregoire (2001) define technical analysis as a method whereby future stock prices are predicted by identifying predefined patterns in historical price data. By employing moving averages of varying lengths and comparing short-term moving average prices to long-term moving average prices, the analysis generates buy or sell signals: if the short-term moving average exceeds the long-term moving average, the subsequent day is deemed a buy day, and vice versa.

Detry and Gregoire (2001) replicated a prior study that demonstrated the capacity of variable-length moving averages to predict stock price changes the following day. The experimental results corroborated the earlier study's conclusions, revealing that variable-length moving averages were capable of accurately predicting prices in 13 out of 15 instances.

2.3.3 Autoregression

Autoregressive models are forecast methods where the forecasted variable is predicted by using a combination of linear variables from the past values. The term autoregression comes from the model being a regression of the value compared to itself. (Hyndman et al., 2018)

Autoregressive model of order p can be written as:

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

Equation 2.9

where ε_t is white noise. This model is referred to as an AR(p) model. (Hyndman et al., 2018)

Hyndman et al. (2018) demonstrate autoregressive models and their flexibility: the figures below show AR (1) and AR (2) models. Changes in parameters $\phi_1 \ldots \phi_p$ result in two different looking patterns in time series:

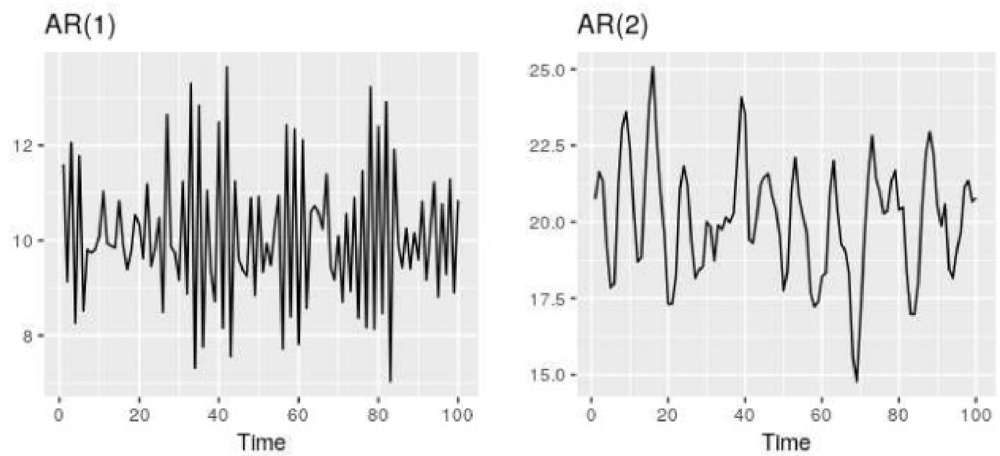

Figure 8: Autoregressive models AR (1) and AR (2) -Hyndman & Athanasopoulos (2018), Forecasting: Principles and Practice, Chapter 8.3

Whereas variance error ε_t affects only on the scale on the series.

AR (1) is written by $y_t = 18 - 0{,}8y_{t-1} + \varepsilon_t$

and AR (2) is written by $y_t = 8 + 1{,}3y_{t-1} + 0{,}7y_{t-2} + \varepsilon_t$

where ε_t is normally distributed white noise with a variance of 1 and mean 0.

For an AR (1) model, when:

$\phi_1 = 0,\ y_t$ is equal to white noise.

$\phi_1 = 1$ and c = 0, y_t is equivalent to a random walk.

$\phi_1 = 1$ and c $\neq 0$, y_t is equivalent to a random walk with drift.

$\phi_1 < 0$, y_t tends to fluctuate between positive and negative values.
(Hyndman et al. 2018)

Autoregressive models are usually restricted to stationary data, which will be discussed in the next chapter. In these cases, there are constraints of the parameter values:

For AR (1): -1 < ϕ_1 <1.

For AR (2): -1 < ϕ_2 <1, ϕ_1+ϕ_2 < 1

(Hyndman et al. 2018)

2.3.4 Stationary Time Series

Hyndman et al. (2018) characterize stationary time series as those that exhibit the overall nature of the time series without reliance on the specific portion being observed. In essence, the time series should maintain a consistent scale, irrespective of the point at which it is examined.

Consequently, time series exhibiting seasonal fluctuations or trends cannot be considered stationary, as these patterns differentially influence values at distinct time points. Hyndman et al. (2018) also note that a white noise series is stationary, as it is not dependent on time and should exhibit randomness at any given moment. It is important to recognize that time series with cyclic patterns can be deemed stationary if the aforementioned criteria (absence of seasonality or trend patterns) are satisfied. This is attributable to cycles lacking a fixed length, precluding the interpretation of future peak or trough points based solely on time series observation. In summary, stationary time series lack predictable patterns and appear as horizontal lines with constant variance. (Hyndman et al. 2018)

2.3.4.1 Differencing

Differencing is the act of transforming a time series into a stationary series by using differences between consecutive values instead of the values themselves. Differencing may help in stabilizing the mean and variance of the time series by eliminating the changes in the level of observations. This reduces or eliminates the trend and seasonality from the time series. (Hyndman et al. 2018)

Hyndman et al. (2018) present the random walk model as a method of differencing:

$$y_t' = y_t - y_{t-1}$$

Equation 2.10

What is notable is that differenced series have T-1 values. The cause of this is the inability to calculate y_t' for the first observation.

The differenced series can be regarded as white noise; hence it can be written as:

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

Equation 2.11

The equation can be rearranged as the random walk model:

$$y_t = y_{t-1} + \varepsilon_t$$

Equation 2.12

Random walk models are widely used for non-stationary series in the field of finance and economics. The typical random walks include both lengthy periods of apparent ups or downs and unpredictable changes in direction. Predictions created applying these models are equal to the last observation as the future movements are unpredictable. (Hyndman et al. 2018)

Hyndman et al. (2018) present seasonal differencing as an alternative way to create stationary time series. The method calculates differences between observations and their counterparts from the same part of the seasonal cycle:

$$y'_t = y_t - y_{t-m}$$

Equation 2.13

Where *m* is the number of seasons, as known as lag-m differences. The name comes from the method, where the observation after a lag of m periods is subtracted from the value.

If the data appears to be white noise, then the use of formula:

$$y_t = y_{t-m} + \varepsilon_t$$

Equation 2.14

is suggested for the original data. As in equation 2.12, the forecasts for this model are equal to the last observation.

A way to determine if there is a need for differencing is to take a unit root test, which are statistical tests for hypothesizes. These tests determine whether differencing is required. There are multiple tests available, some of which might give different answers compared to others, due to different assumptions in the tests. These tests are usually included in packages for statistical programs. (Hyndman et al 2018)

2.3.4.2 Backshift notation

The backwards operator (B), or lag (L) in some references is expressed by:

$$By_t = y_{t-1}$$

Equation 2.15

Which means that B operating on y_t shifts the data back one period. A shift back of two periods is expressed by:

$$B(By_t) = B^2 y_t = y_{t-2}$$

Equation 2.16

Operating with data from different periods require different backshifts, for example for monthly data a shift back of twelve periods could be used to receive the same month previous year. (Hyndman et al. 2018)

Operator B can be used to describe the process of differencing. A first difference can be expressed by:

$$y'_t = y_t - y_{t-1} = y_t - \mathrm{B}y_t = (1 - B)y_t$$

Equation 2.17

and in general, the difference in *d*th order can be expressed by:

$$(1 - B)^d y_t$$

Equation 2.18

The backshift operator can be used to combine differences, as it can be treated using algebraic rules. In particular, the terms can be multiplied together if B is involved. (Hyndman et al. 2018)

2.3.5 Moving average models

As an alternative of using the latest values of the forecasted variable in a regression, models with moving averages use forecast errors in a model resembling regressions. Moving average model of order q:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

Equation 2.19

where ε_t is white noise. (Hyndman et al. 2018)

Moving average models should not be mixed with the moving averages in smoothing, discussed earlier in the thesis. According to Hyndman et al. (2018), moving average models are used for forecasting of future values, as moving average smoothing is used to estimate trends.

Hyndman et al. (2018) demonstrate the moving average models in the figure below including MA (1) and MA(2):

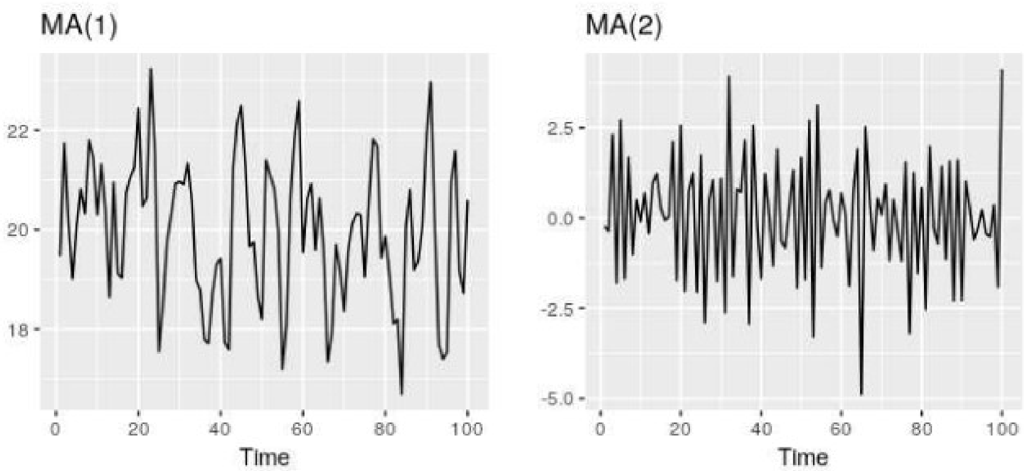

Figure 9: Moving average models MA (1) and MA (2) -Hyndman & Athanasopoulos (2018), Forecasting: Principles and Practice, Chapter 8.4

MA (1) is written by $y_t = 20 + \varepsilon_t + 0{,}8\varepsilon_{t-1}$

and MA (2) is written by $y_t = \varepsilon_t - \varepsilon_{t-1} + 0{,}8\varepsilon_{t-2}$

where ε_t is normally distributed white noise with a variance of 1 and a mean of 0.

changing the parameters $\theta_1, \ . \ . \ . \ , \theta_q$ will result in different time series patterns. Equally to autoregressive models discussed earlier, variance error ε_t affects only on the scale of the series. (Hyndman et al. 2018)

2.3.6 ARIMA

Siami-Namini and Namin (2018) present stochastic models utilized in time series forecasting. Auto-Regressive Moving Average (ARMA) is the most well-known method for a single time series time series data, which is a method combining autoregressive and moving average models. Auto-Regressive Integrated Moving Average (ARIMA) is a type of ARMA, which includes integrated differencing.

Siami-Namini et al. (2018), describe the key elements of the model as following:

- Autoregression (AR). A regression model utilizing observations and lagged observations (p).
- Integrated (I). The part of the model with the purpose of making a time series stationary (d).
- Moving Average (MA). An approach taking the dependency between variables and residual errors into account (q).

Hence, if a simple form of an AR model of order p can be written as:

$$x_t = c + \sum_{i=1}^{p} \emptyset_i x_{t-i} + \varepsilon_t$$

Equation 2.20

Where x_t is the stationary variable, c is constant, the terms of $\emptyset_i$ are autocorrelation coefficients at lags 1…. p and the residuals, ε_t, is white noise with a mean of 0 and variance of σ_ε^2. An MA model of order q can be written in the form:

$$x_t = \mu + \sum_{i=0}^{q} \theta_i \varepsilon_{t-i}$$

Equation 2.21

Where μ is the expectation of x_t, θ_i are the weights applied to the values and ε_t, is white noise with a mean of 0 and variance of σ_ε^2. Adding these models together will create an ARMA model of order (p, q):

$$x_t = c + \sum_{i=1}^{p} \emptyset_i x_{t-i} + \varepsilon_t + \sum_{i=0}^{q} \theta_i \varepsilon_{t-i}$$

Equation 2.22

Where $\emptyset_i \neq 0$, $\theta_i \neq 0$, and $\sigma_\varepsilon^2 > 0$. The parameters p and q stand for AR and MA. (Siami-Namini et al. 2018)

ARIMAs are capable of using non-stationary time series data, as the "integrated" (I) component uses differencing to convert the time series to a stationary one. The general form of an ARIMA model is ARIMA (p, d, q). (Siami-Namini et al. 2018)

As in ARMA models, in ARIMA models p equals the order of the autoregressive part and q the order of the moving average part. The "integration" added to ARIMAs is linked to the term d, which stands for the degree of differencing involved. A full ARIMA model can be written as:

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Equation 2.23

Where y'_t is the differenced time series. The model includes both lagged values and errors of y_t. It should be taken into consideration that the stationary and invertibility conditions that applied to moving average models and autoregressive models also apply to ARIMAs. (Hyndman et al. 2018)

Hyndman et al. (2018) further analyze the formula by listing special cases:

ARIMA (0, 0, 0) = White noise

ARIMA (0, 1, 0) with no constant = Random walk

ARIMA (0, 1, 0) with a constant = Random walk with drift

ARIMA (p, 0, 0) = Autoregression

ARIMA (0, 0, q) = Moving average

According to Hyndman et al. (2018) combining the components of the models in order to create more advanced ones make the use of backshift notations easier. The equation 2.23 can be written in backshift notation as:

$$\left(1 - \phi_1 B - \cdots + \phi_p B^p\right)\ (1 - B)^d{}_{y_t} = c + (1 + \theta_1 B + \cdots + \theta_q B^q)$$

Equation 2.24

The constant c effects on the long-term forecasts obtained from ARIMA models. Hyndman et al. (2018) list the effects on forecasts as:

- If c = 0 and d = 0, the long-term forecasts will go to zero.
- If c = 0 and d = 1, the long-term forecasts will go to a non-zero constant.
- If c = 0 and d = 2, the long-term forecasts will follow a straight line.
- If c ≠ 0 and d = 0, the long-term forecasts will go to the mean of the data.
- If c ≠ 0 and d = 1, the long-term forecasts will follow a straight line.
- If c ≠ 0 and d = 2, the long-term forecasts will follow a quadratic trend. (Hyndman et al., 2018, p.313)

Hyndman et al. (2018) posit that the value of d is associated with the size of the prediction interval. When d is equal to zero, the standard deviation of long-term forecasts eventually converges with the standard deviation of the historical data. The value of p plays a significant role in the context of data exhibiting cyclic trends.

Chen & Lei (2016) assert that ARIMA models are characterized by their simplicity and adaptability, leading to their increased popularity. The versatility of ARIMA has facilitated its implementation in numerous forecasting studies.

Nochai and Nochai (2006) investigated the application of an ARIMA model for predicting oil palm prices. They employed an ARIMA (2,1,0) model for farm prices, ARIMA (1,0,1) for wholesale prices, and ARIMA (3,0,0) for general oil palm prices. The study was prompted by the oil crisis in Thailand during the early 2000s, when the government encouraged the use of oil palm and other biodiesel alternatives. Model evaluation was conducted using MAPE, which ranged between 13,23 and 5,27, indicating high forecast accuracy.

Chen, Yuan, and Shu (2008) compared an ARIMA model with simple exponential smoothing and two-parameter exponential smoothing to forecast property crime rates in a Chinese town. The ARIMA model demonstrated superior fitting and accuracy compared to the other methods, with a MAPE of 948.

2.3.7 SARIMA

Seasonal ARIMA models (SARIMA) are ARIMAs including additional seasonal terms. According to Hyndman et al. (2018) seasonal ARIMA consists of:

$$ARIMA\ (p, d, q)\ (P, D, Q)m$$

Equation 2.25

Where *m* = number of observations per year. In the equation, (p, d, q) is the non-seasonal part of the model, and $(\mathrm{P}, \mathrm{D}, \mathrm{Q})\mathrm{m}$ is the seasonal part.

The seasonal part of the model consists of terms similar to the non-seasonal terms but involve a backshift in the seasonal period. Hydman et al. (2018) demonstrate the effect by writing $SARIMA(1,1,1)(1,1,1)_4$ for quarterly data (m=4) as follows:

$$(1-\phi_1 B)(1-\Phi_1 B^4)(1-B)\,(1-B)^4{}_{y_t} = (1+\theta_1 B)(1+\Theta_1 B^4)\varepsilon_t$$

Equation 2.26

Which means that the seasonal terms are multiplied by the non-seasonal terms. (Hyndman et.al 2018)

Chang and Liao (2010) used SARIMA models in the forecasting of tourism departures to Taiwan. The data consisted of departures from 3 destinations: Hong Kong, Japan, and USA. After unit root testing, the data was differenced and empirical examination for models was conducted. The models selected were:

- $\mathrm{SARIMA}(0,1,1)(1,0,1)_{12}$ for Hong Kong
- $\mathrm{SARIMA}(0,1,1)(1,0,0)_{12}$ for Japan
- $\mathrm{SARIMA}(1,1,0)(0,0,1)_{12}$ for the USA
- $\mathrm{SARIMA}(1,1,1)(1,0,0)_{12}$ for total tourism in Taiwan.

The MAPE for these models was low, ranging from 5,63% to 8,9%, which indicates that the models were highly accurate.

Eni (2015) used SARIMAS to forecast rainfall in Nigeria. Due to the rainfall causing flooding in Nigeria, it is important to receive early warnings of rain in order to optimally use and manage water resources. SARIMA (1, 1, 1) (0, 1, 1). According to Eni (2015), the results of forecasting rainfall for one year were promising, and the SARIMA model proved to be capable in forecasting rainfall.

2.4 Box-Jenkins Approach

Identifying optimal models for time series analysis is a crucial step in conducting analyses or obtaining further information about the time series. A widely recognized strategy for identifying an appropriate model is the Box-Jenkins approach (Cryer, 1986).

Box and Jenkins (1976) propose a three-step strategy for finding an optimal model for a dataset:

1. Model Identification
2. Model Estimation
3. Model Diagnostics

Each step may be employed multiple times to identify the optimal model for the time series (Cryer, 1986). Bleikh and Young (2016) introduce a preliminary step before the main steps: determining the stationarity of the time series. According to Bleikh and Young (2016), Box and Jenkins recommend differentiating non-stationary time series to attain stationarity. Additionally, Bleikh and Young (2016) cite forecasting as a fourth step in the process.

2.4.1 Model Identification

Box & Jenkins (1976) state that during the identification phase of model building, no precise formulation for the problem is available. Consequently, statistically ineffective methods must be utilized. At this stage, Box and Jenkins (1976) suggest employing graphical methods while bearing in mind that these methods only provide preliminary results for subsequent stages of the process.

As ARIMA models constitute a diverse group, selecting parameters such as the degree of differencing (d), the degree of the autoregressive process (p), and the degree of the moving average process (q) can be challenging (Bleikh

& Young, 2016). Model identification not only aims to narrow down the types of models used but also provides estimates of the model parameters. This is particularly useful during the estimation stage to provide initial values for the processes (Box & Jenkins, 1976). Methods used for model identification include information criteria, partial autocorrelation functions, and full autocorrelation functions (Magnus & Fosu, 2011). In practice, as there is no knowledge of the real autocorrelation and partial correlation functions, sample autocorrelation and sample partial autocorrelation formulas should be estimated based on the data. (Bleikh & Young, 2016)

2.4.2 Model Estimation

Box & Jenkins (1976) assert that the identification process frequently overlaps with the estimation process. The identification process results in a tentative formulation of the model, and the subsequent step is to obtain efficient estimates of the parameters.

For simple autoregressive models, some straightforward techniques in model estimation can be employed due to the linear relationships between autocorrelations and autoregressive model parameters. As the process is complex and should be executed by computers, most statistical software include features enabling the use of the Box-Jenkins method. Non-linear least squares and maximum likelihood estimations are among the primary approaches for this procedure. (Bleikh & Young, 2016)

2.4.3 Model Diagnostics

Upon estimating the parameters, standard errors and confidence intervals must be derived. It is advised to incorporate additional parameters into the model and examine their significance, a process referred to as over-fitting. A prevalent method for this involves analyzing the residuals and their behavior within the model. (Bleikh & Young, 2016)

Box and Jenkins (1976) assert that the diagnostic check serves as a means of verifying the model's appropriateness for the specific task. Should signs of inefficiency be present in the model, the process offers insights on how the model might be altered to achieve the desired level of effectiveness. Once the correct model has been identified, a more intricate version is constructed, which brings the initial model identification into question due to the inclusion of supplementary parameters in the more detailed model. If the more complex model fails to demonstrate improved efficiency or if the additions do not yield superior results, it can be inferred that the initially identified model is optimal.

3 Empirical study

This chapter will provide a short introduction to the data used in the empirical study and analyze it. The chapter will continue with a description of the models used and present the results of the study.

3.1 Introduction to the data

The electricity system of Finland consists of power plants, the transmission grid, regional and distribution networks. The power system in Finland is part of the inter-Nordic system that spreads to Sweden, Norway, and Eastern Denmark. Part of the Finish electricity network are links to both Estonia and Russia. (Fingrid, 2017)

Finland is a part of the Nordic wholesale electricity market, where 70% of the trade takes place in the power exchange located in Oslo. The market covers the Nordic countries and the Baltic states but will be expanding to cover the whole of Europe, as the European union is aiming to create a European market. The internal market is part of the EU’s energy union and objectives concerning the security of energy supply. (Ministry of economic affairs and employment 2022)

Supply and total consumption of electricity, 2021*

Supply	GWh	%
Nuclear power	22 649	26,1
Hydro power	15 550	17,9
Wind power	8 133	9,4
Solar power	305	0,4
Net imports	17 768	20,5
Other heating power	22 370	25,8
Total	**86 775**	**100**
Total consumption	**GWh**	**%**
Industry and construction	37 706	43,5
Households and agriculture	24 535	28,3
Services and public consumption	21 906	25,2
Transmission and distribution losses	2 629	3,0
Total	**86 775**	**100**

* preliminary data

Table 1: Supply and total consumption of electricity, 2021 (Tilastokeskus 2022)

According to Tilastokeskus (2022) the Finish electricity consumption consisted in 2021 of 20,5% imported energy and the most important source of domestic energy was nuclear power with a share of 26,1%. Most of the energy is consumed by industry and construction (43,5%), and the households together with agriculture consume 28,3% of the energy.

3.2 Description of the dataset

The dataset used in the study is consists of hourly electricity consumption measured in megawatt hours in Finland and is sourced from Fingrid. The dataset includes hourly electricity consumption data from 2017 to the end of 2022, spanning a total of five years. The data is provided in csv format and contains a total of 44169 data points.

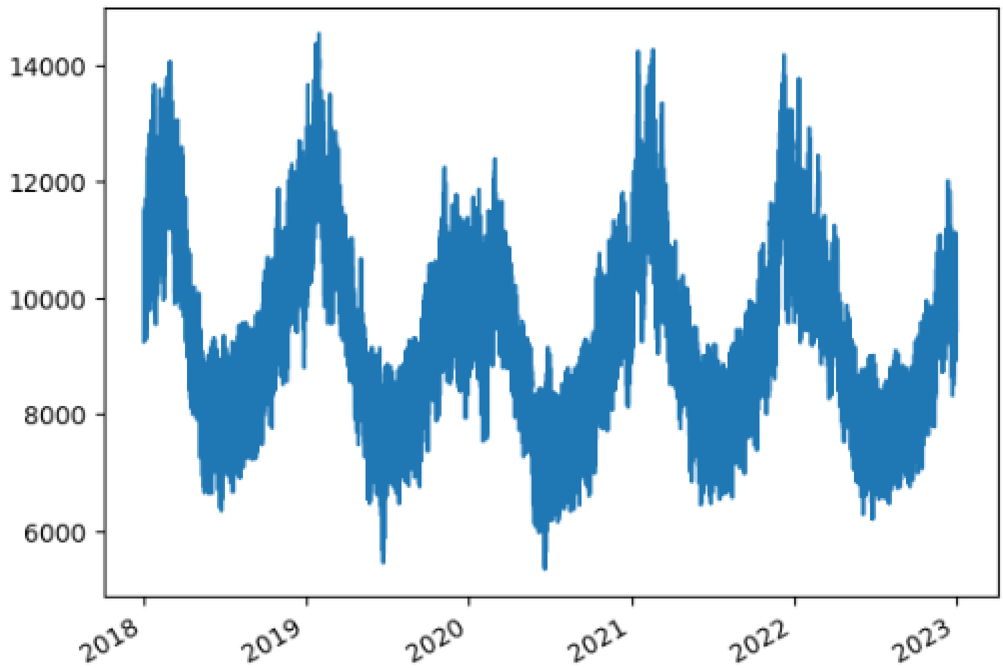

Figure 10: Consumption of electricity in Finland

The values in the dataset have a mean of 9381,09 and a standard deviation of 1561,54. The minimum value in the dataset is 5341, and the maximum value is 14542. The distribution of values in the dataset is skewed to the right, with a 25th percentile value of 822, a 50th percentile value of 9138, and a 75th percentile value of 10471.

Visually inspected, the data is clearly seasonal, and it fluctuates between the cold winter season and warmer summer season. One could also notice that seasonality is not only yearly, but the consumption also varies between shorter periods. No clear trend pattern is recognizable from the data, which could indicate that the models will not need excessive amounts of differencing.

The data was preprocessed by aggregating the values into consumption by week, fortnight and month. The aggregation was done using the function resample in pandas, a library used for data analysis in python. The aggregation was done for the purpose of saving the amount of computing power required for the models to work, and for the sake of comparing the accuracy of models on different periods.

3.2.1 Consumption, one-week periods

As the data is resampled for weekly consumption of electricity, amount of datapoints drop from 44169 to 261. The mean for the data is 1587560,1724, the standard deviation is 252778,07708. The values fluctuate between 1161397 and 2768911.

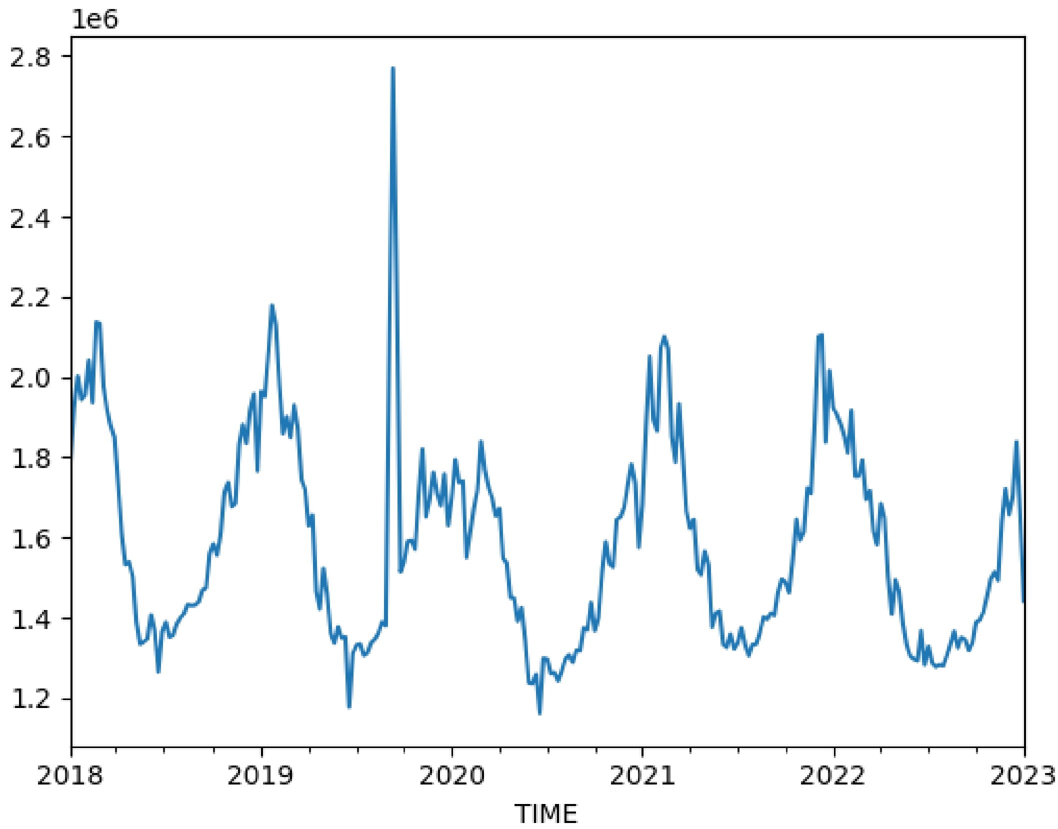

Figure 11 : Electricity Consumption in Finland, Weekly Data

Visually inspecting the data, it has no particular trend, which would mean that the time series is almost stationary. By using a rolling mean to visualize data, one can expect to catch a trend pattern by smoothing out the noise in the data.

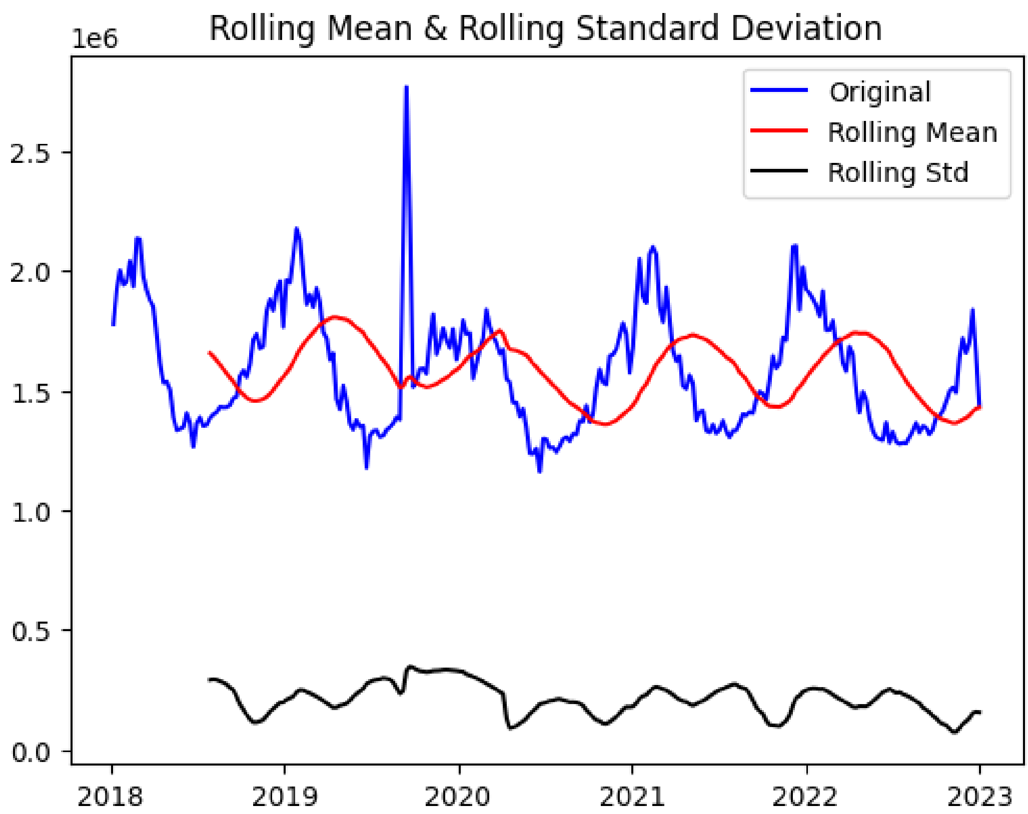

Figure 12: Electricity Consumption in Finland, Weekly Data, Rolling Mean and Rolling Standard Deviation

Figure 12 shows a weak, downward trend pattern. A pattern would indicate that the time series is non-stationary and will need differencing in order to fit the models.

Another way of inspecting the data is to use the decompose function in python. This function separates the trend, seasonality and residual from the time series.

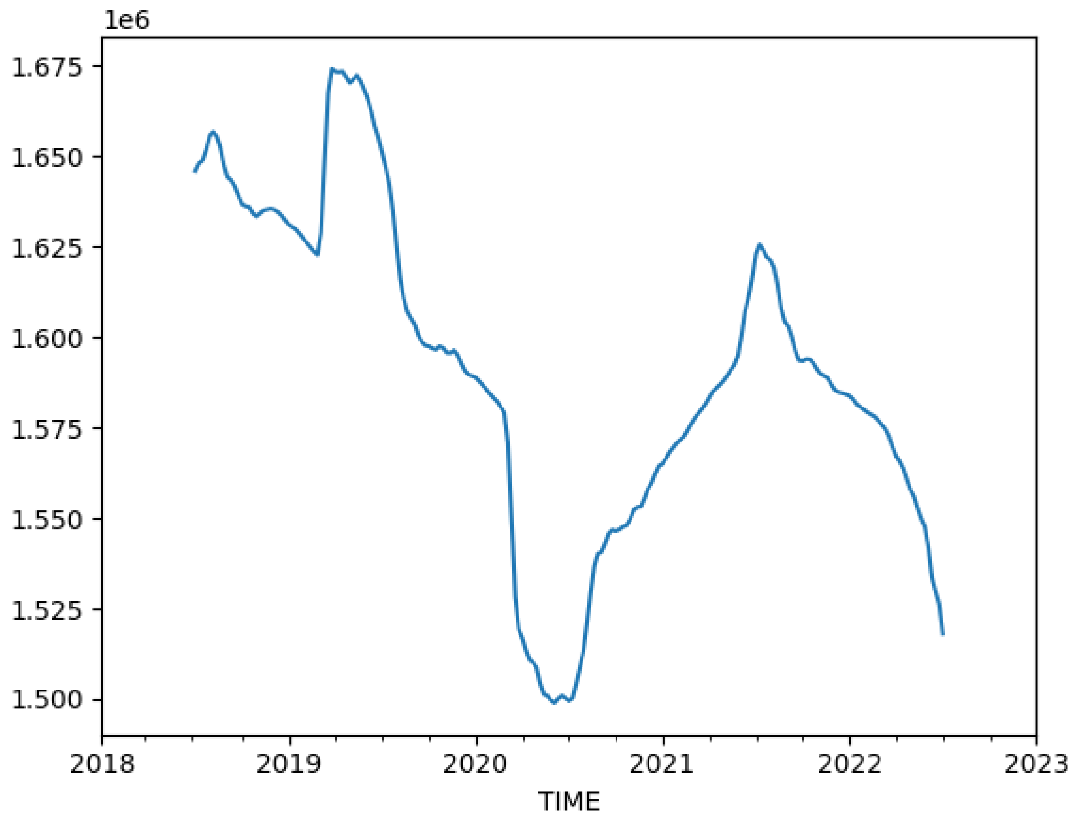

Figure 13: Electricity Consumption in Finland, Weekly Data Decomposed Trend Pattern

Decomposing the time series confirms the suspicion of a downward trend pattern in the data. In order to receive confirmation of the need for differencing, a unit root test was conducted. The selected test was Augmented Dickey Fuller (ADF), which is included in python libraries used for the study. The ADF test results in a p-value of 0,02 and critical values of 1%: -3,456 for 1%, -2,873for 5% and -2.573 for 10%. The p-value is below the significance value of 0,05 and thus, the null hypothesis is rejected, and the time series is stationary.

An autocorrelation plot on the non-differenced data shows that the values correlate with each other, the nearer others they are measured in the period.

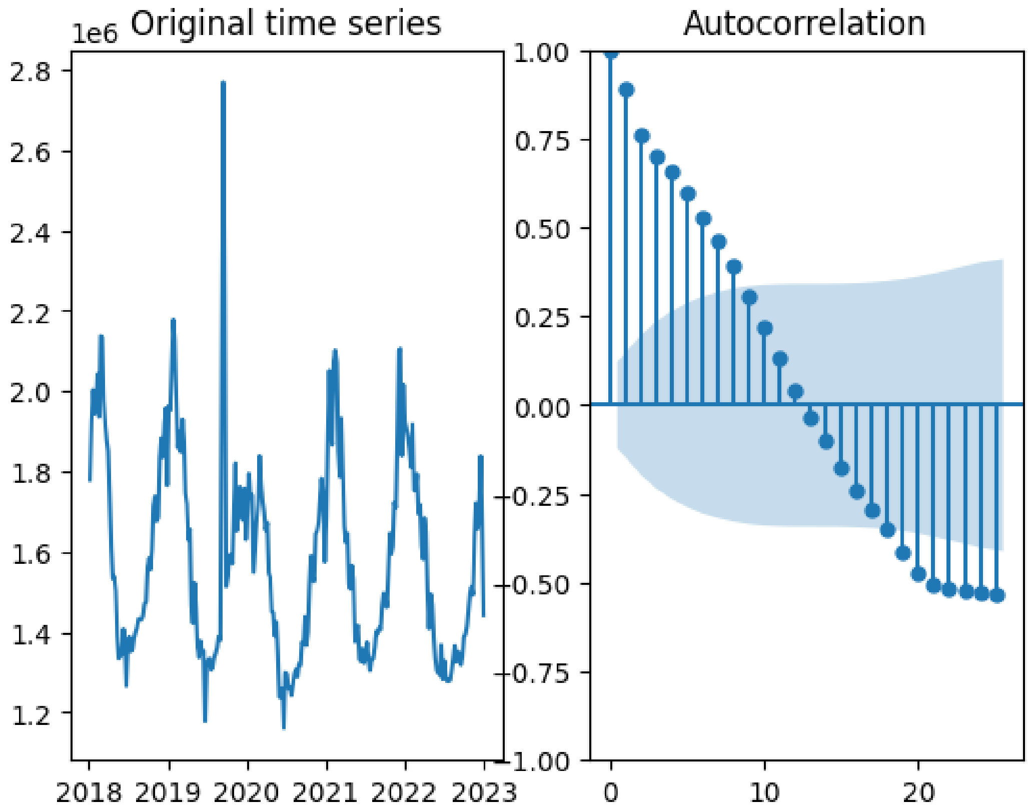

Figure 14: Electricity Consumption in Finland, Weekly Data, Autocorrelation

Basically, looking at figure 14, one could conclude that the weather has an impact on the values. When the lag increases, the correlation decreases. This is logical as the temperature and other weather conditions have an effect on the consumption, and they tend to change slowly. This type of autocorrelation is not ideal for the Arima model; thus, some differencing is needed.

By differencing the time series once, the autocorrelations in the data is looking very different compared to figure 14.

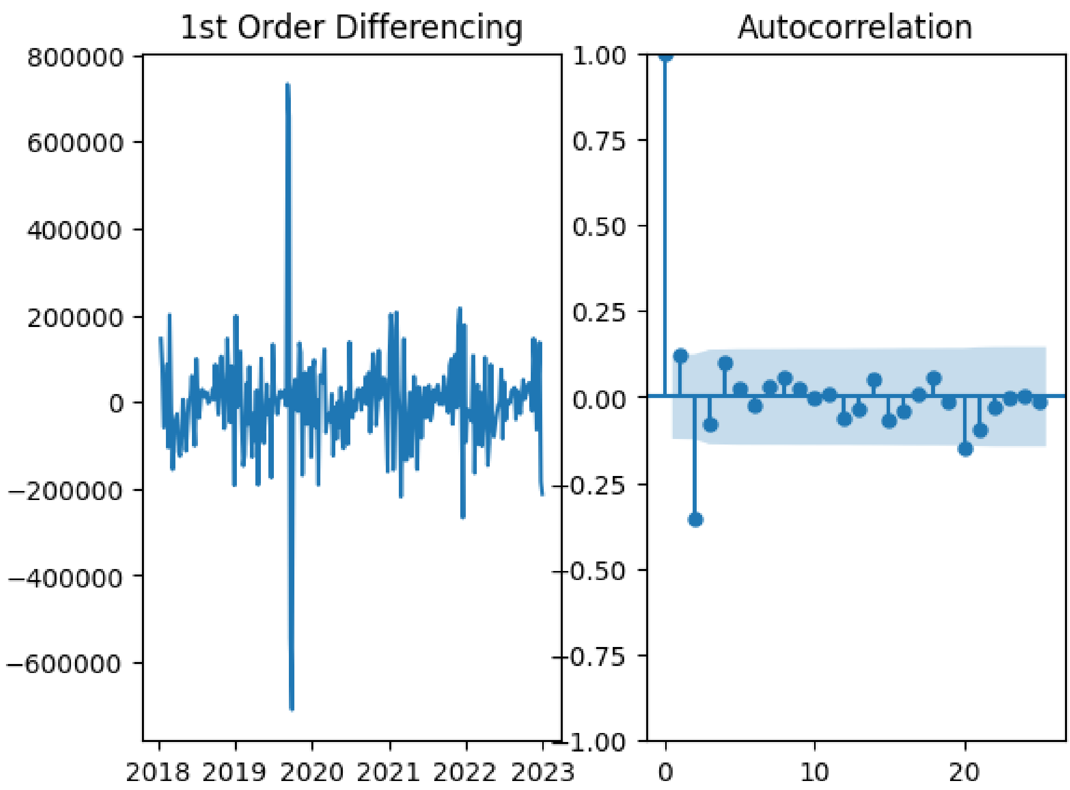

Figure 15: Electricity Consumption in Finland, Weekly Data, Autocorrelation of 1st Order Differenced Data

In figure 15, the correlations drop significantly due to the unit of measure swhiching from absolute values to change in values. The previous value does not determine the next value like in the original data.

The p and q in an Arima model can be estimated with visual inspections of the autocorrelation plot and partial autocorrelation plot. The partial autocorrelation plot measures the correlation of values after removing the effect of other lags. Partial autocorrelations can be used to estimate the last coefficient in an autoregressive model. The visual determination of p and q is not possible if both of the values are positive. (Hyndman et al. 2018)

The data may follow an ARIMA (p, d,0) if the autocorrelation for differenced data is exponentially decaying or sinusoidal and if the partial autocorrelation has a significant spike at lag p, but no significant lags after it. One can expect an ARIMA (0, d, q) if the partial autocorrelation is exponentially decaying or sinusoidal and in the autocorrelation plot there is a significant spike at lag q, but significant lags after it. (Hyndman et al. 2018)

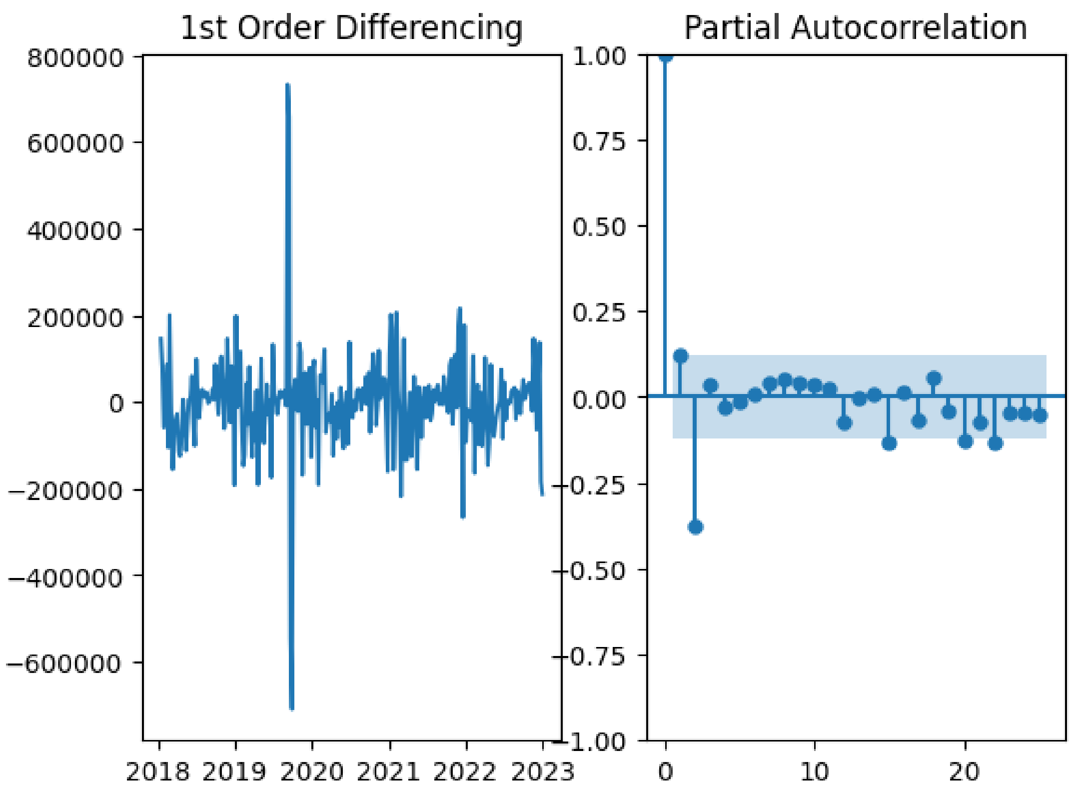

Figure 16: Electricity Consumption in Finland, Weekly Data, Partial Autocorrelation of 1st Order Differenced Data

Inspecting the partial autocorrelation plot together with the plot in figure 15, one can see the values exponentially decaying, when in both of the plots, the correlation drops from 1 to below significance level. The assumptions for ARIMA (0, d, q) and ARIMA (p,d,0) were that either the autocorrelation plot or the partial autocorrelation plot display a significant lag and none thereafter. Inspecting the autocorrelation plot, one can see that the 2nd lag is beyond the level of significance, which would indicate that q is 2. On the other hand, inspecting the partial autocorrelation plot in figure 15, the 2nd lag is significant as well, which would indicate that p is 2. As both of the figures indicate a value > 0, they rule each other out and p or q cannot be determined visually.

As the study is comparing seasonal ARIMAs to ARIMAs, the seasonality of the data is a subject of interest. By using the function decompose, the seasonality pattern can be separated from the time series.

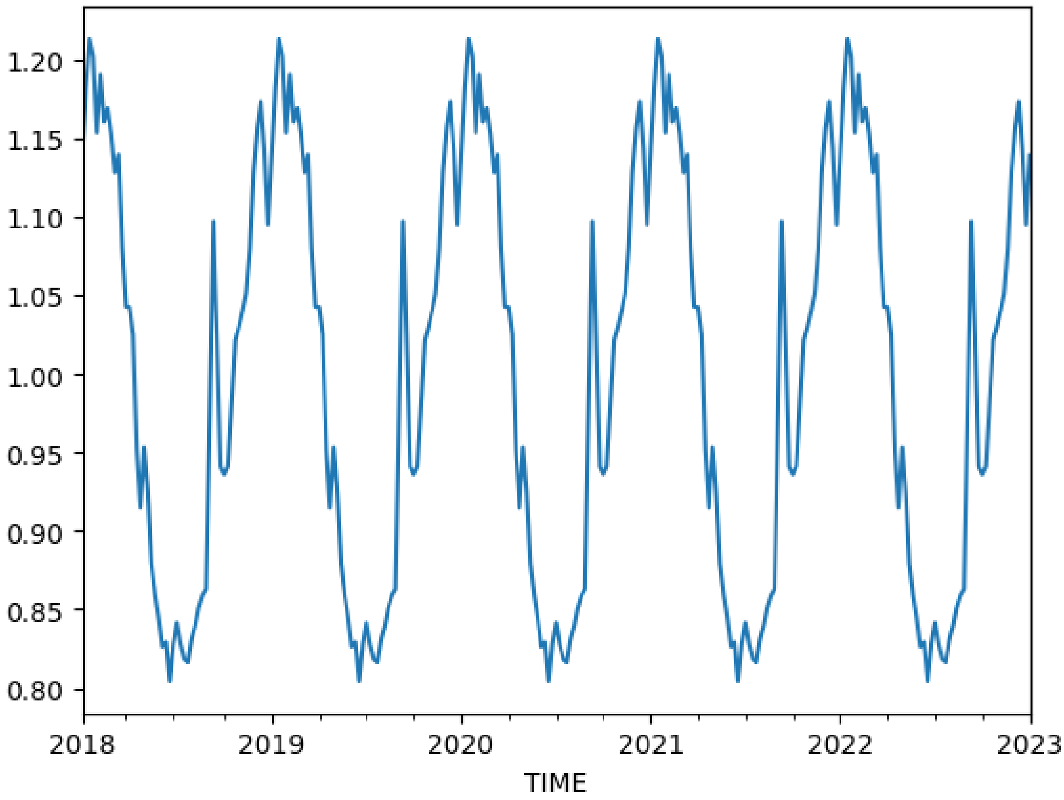

Figure 17: Electricity Consumption in Finland, Weekly Data Decomposed Seasonality Pattern

As expected, the data has a seasonality pattern that can be explained with the heating consuming electricity during the winter season. The pattern also shows some yearly spikes in consumption that separate from the main pattern. There is a spike downwards approximately in the middle of Q2 and a spike upwards at the end of Q3.

The residual of the time series is the irregularities in the data. By decomposing the time series, it can be visualized as well.

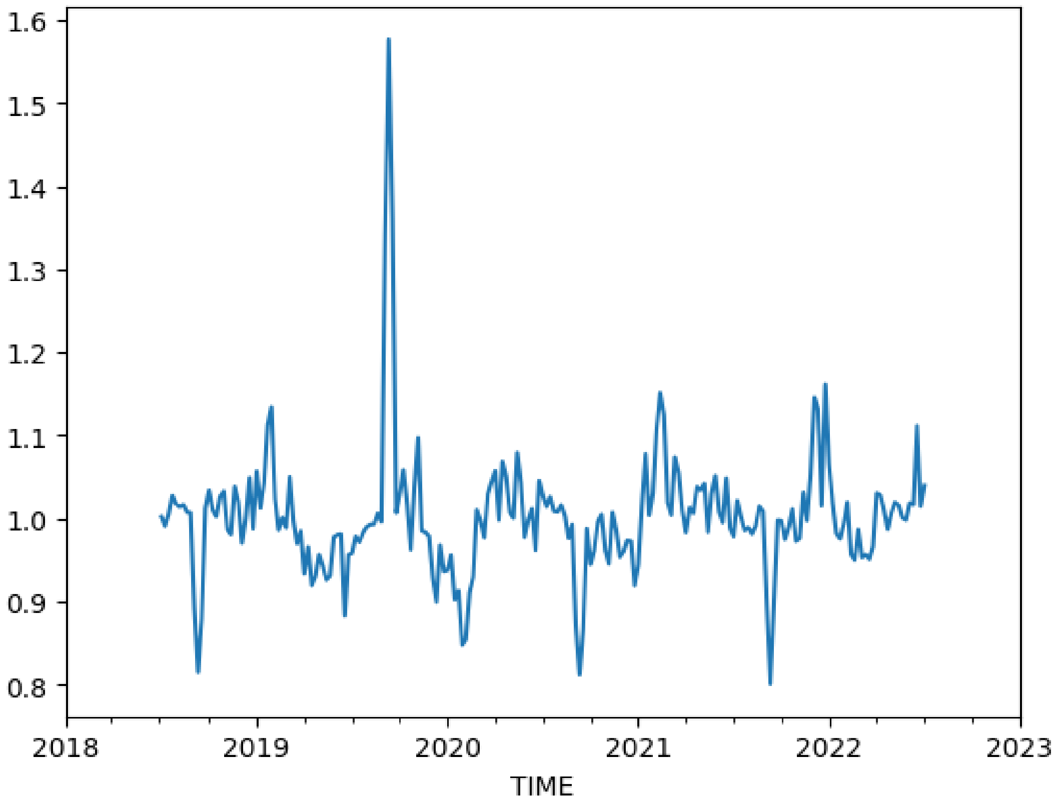

Figure 18: Electricity Consumption in Finland, Weekly Data Decomposed Residual Pattern

In figure 18, it can be interpreted that the distinguished spike in consumption in the third quarter of 2019 is irregular as can be expected. Other notable irregularities can be observed during the third quarters of the year. These can be caused for example by warmer periods during the fall.

3.2.2 Consumption, two-week periods

The time series aggregated to fortnight periods decreases the number of observations to 131. The mean of values for the dataset is 3163001,56, standard deviation 506371,23 and the values fluctuate between 1778723 and 4880804.

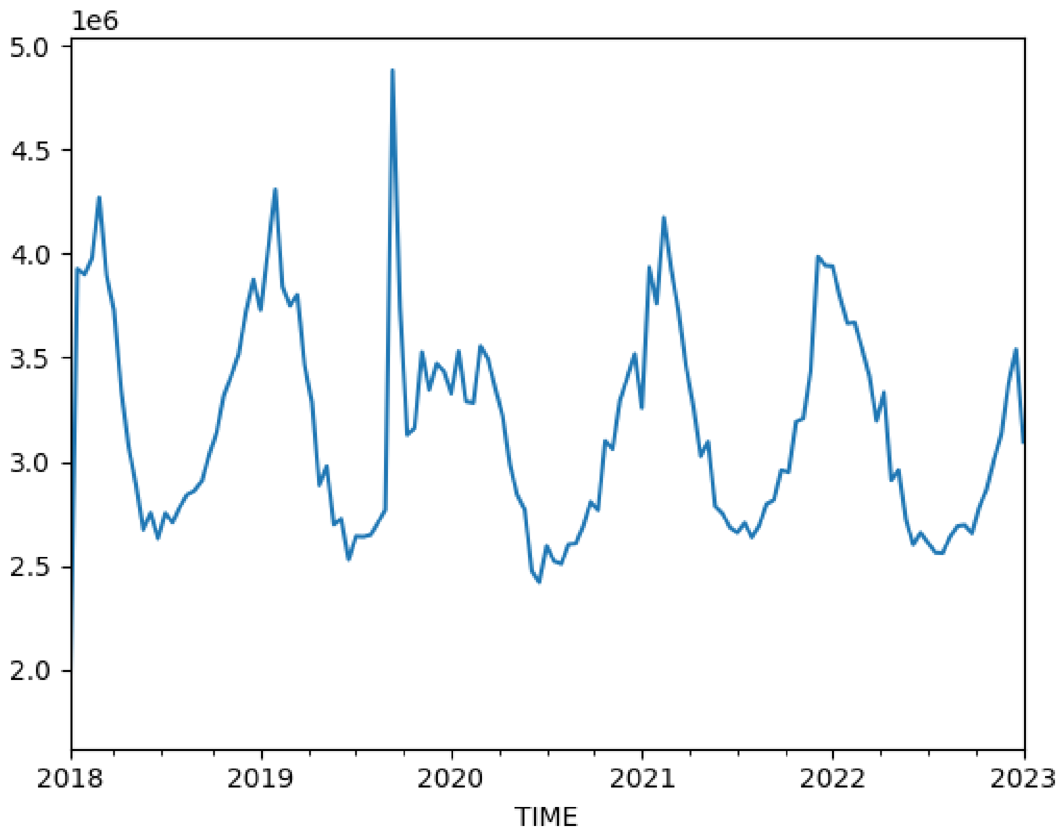

Figure 19: Electricity Consumption in Finland, Data for Two-Week Periods

Compared to the weekly data, the variation between the values are smoothed out a bit.

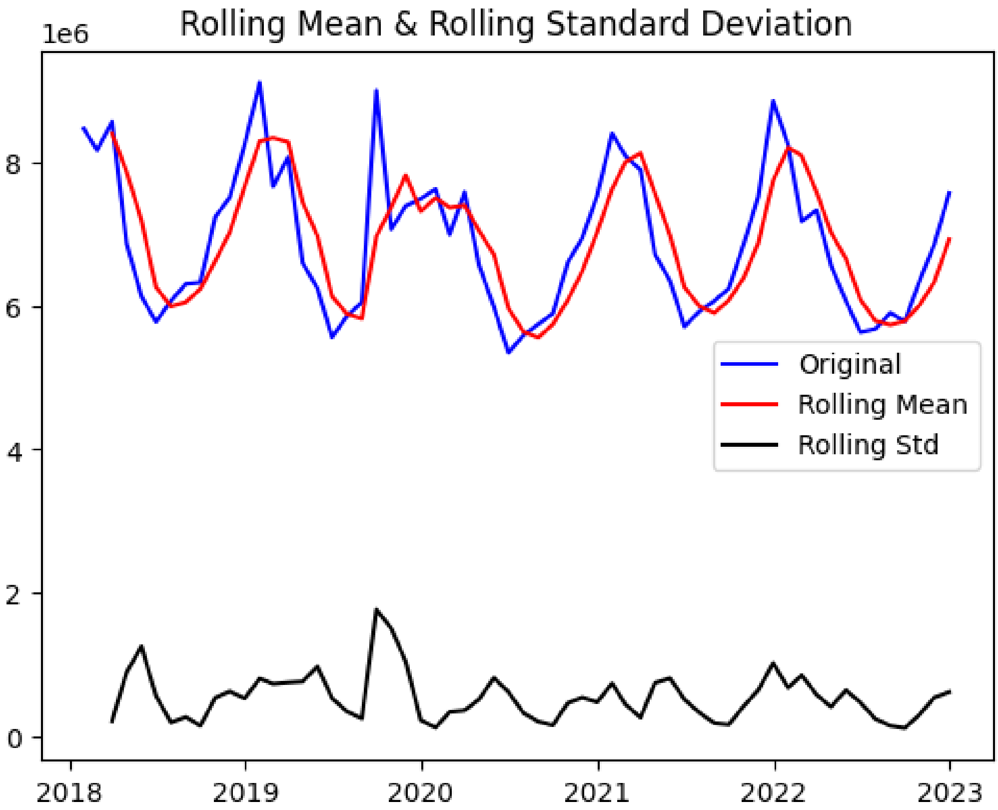

Figure 20: Electricity Consumption in Finland, Data for Two-Week Periods, Rolling Mean and Standard Deviation

Compared to the rolling mean of weekly data in figure 12, one can see that the rolling mean in figure 20 is closer to the actual values. Unlike in figure 12, the rolling standard deviation in figure 20 has two higher spikes during the decrease in observations during 2018 and the increase in consumption in 2019.

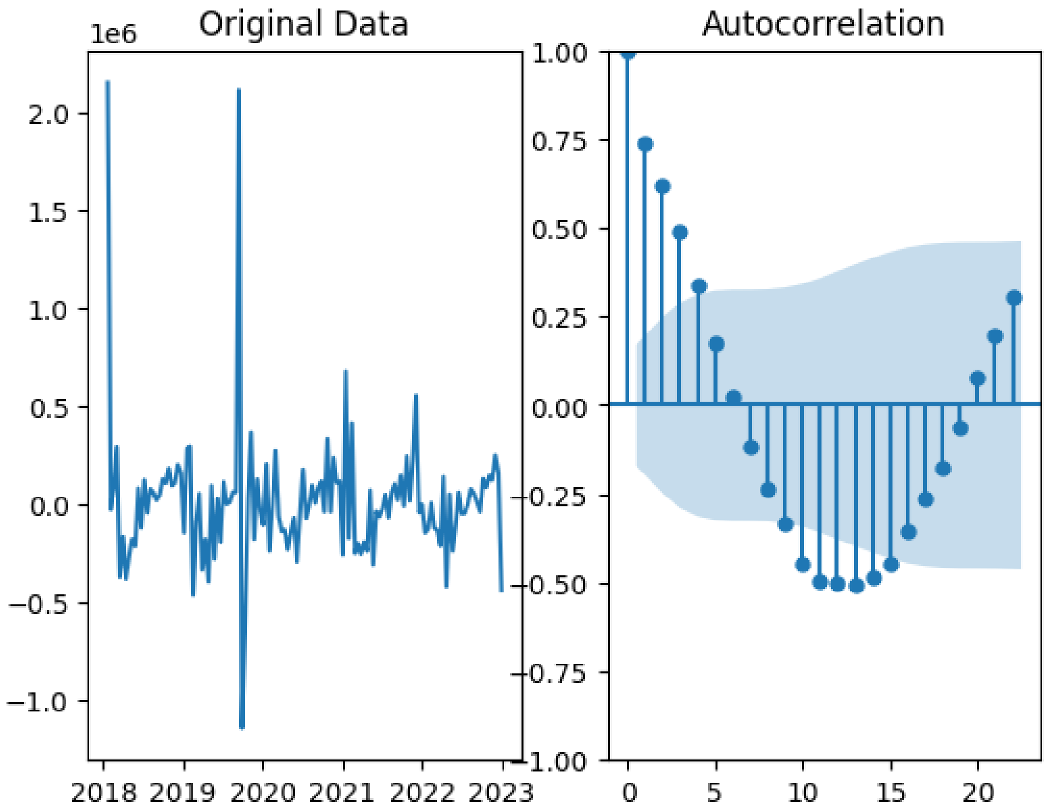

Figure 21: Electricity Consumption in Finland, Data for Two-Week Periods, Autocorrelation

The pattern in the autocorrelation plot is similar in nature compared to the pattern in figure 14. It must me noted that measuring the data on two-week periods means that the observations of Autocorrelation in span the period of approximately one year, which explains the increase in autocorrelation towards the end of the observations in figure 21.

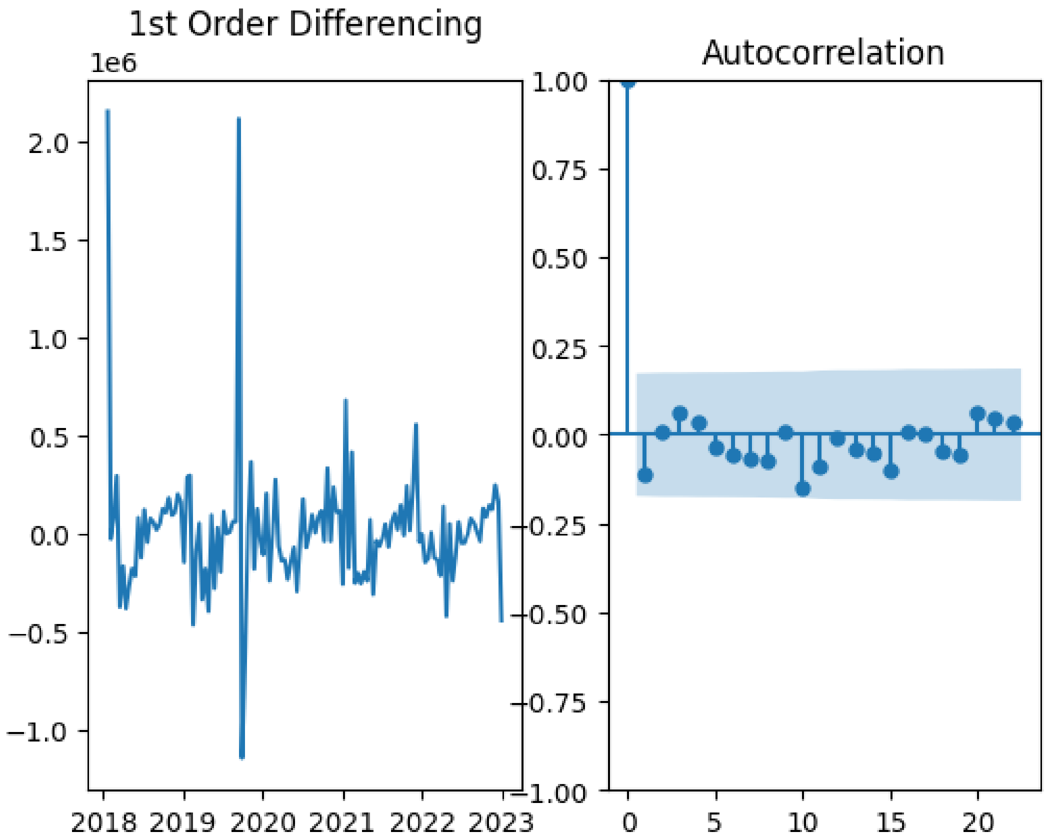

Figure 22: Electricity Consumption in Finland, Data Differenced in the 1st Order for Two-Week Periods, Autocorrelation

The differencing in first order reveals that most of the autocorrelations are negative, which could indicate that the time series is too different. The autocorrelation in figure 22 plot does not display values with significant correlation. Comparing the figure 22 plot to the plot in figure 15, the amount of negative correlation has increased.

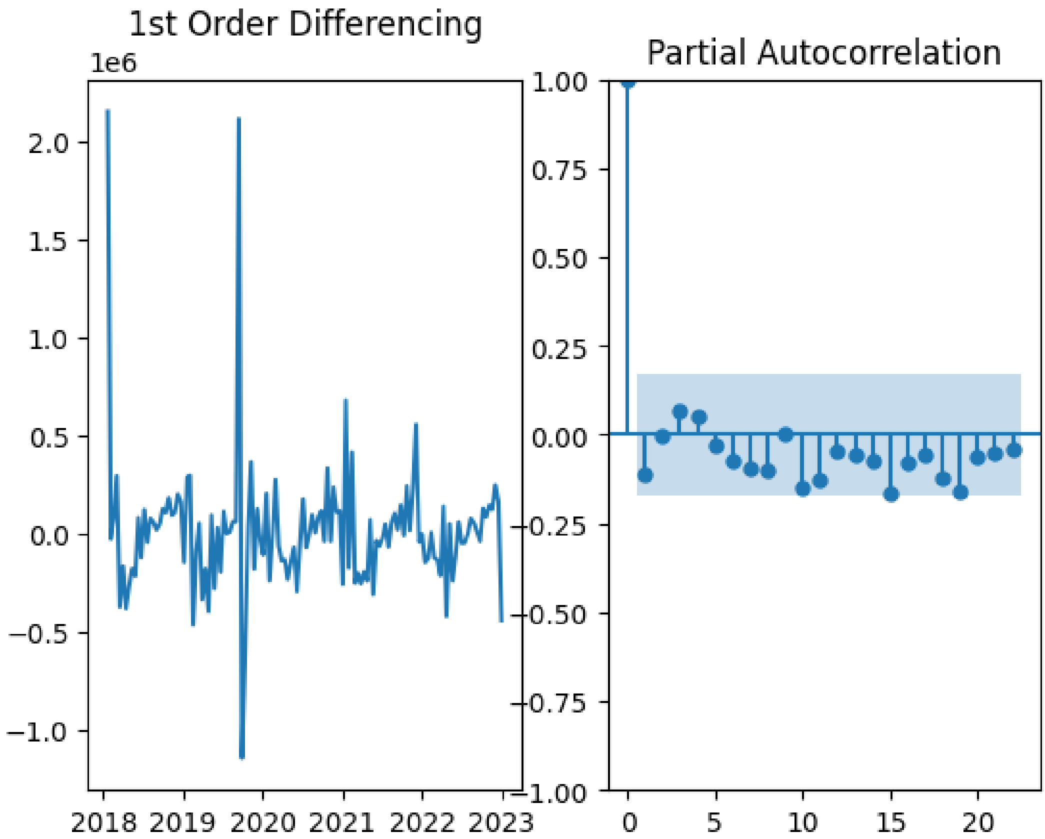

Figure 23: Electricity Consumption in Finland, Data Differenced in the 1st Order for Two-Week Periods, Partial Autocorrelation

The partial autocorrelation in figure 23 has no significant values, and as in figure 22, the amount of negative correlation has increased.

The ADT test for the data gives a result of 0,0000072, which is clearly below the level of significance and therefore the null hypothesis can be rejected. The decomposed time series illustrates a small downwards trend, as in weekly data, but according to the ADT test it is not significant enough for the time frame not being stationary.

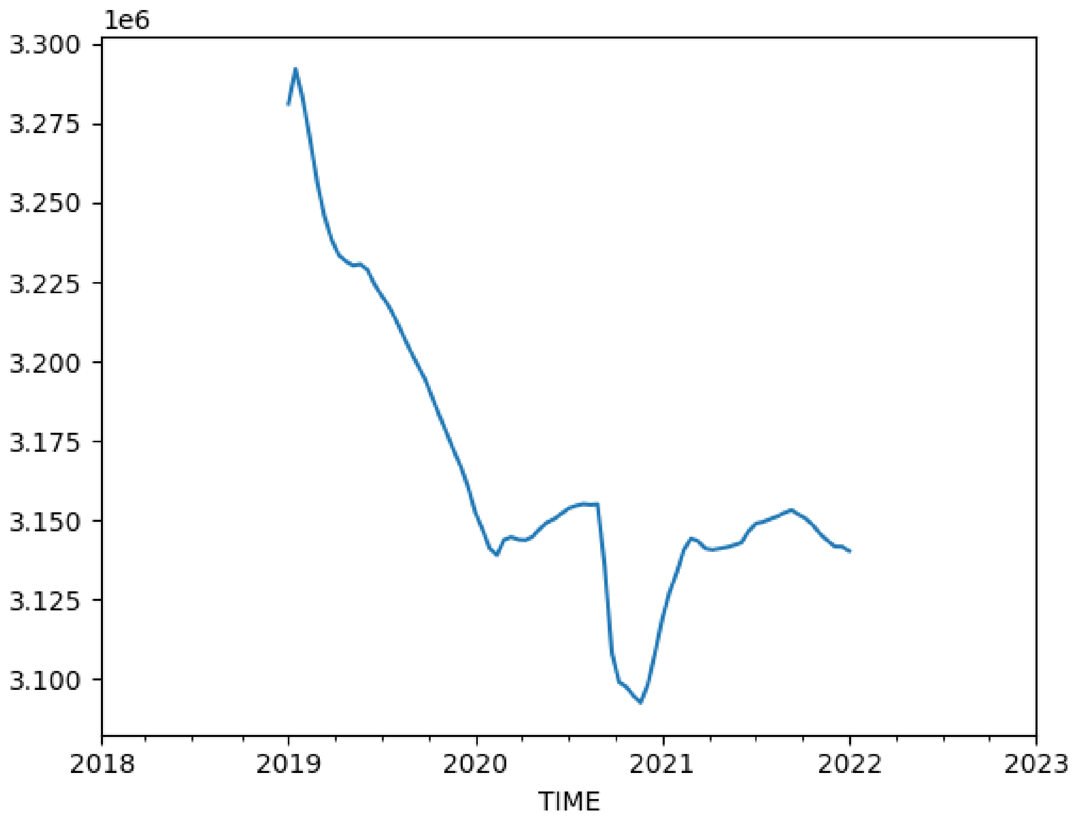

Figure 24: Electricity Consumption in Finland, Data for two-week periods, Decomposed Trend Pattern

One reason for the decreasing trend in figure 24 could be that decomposing function does not see the peak in consumption during 2019 as residual. As the consumption for the beginning of 2019 was higher than usual, it is regarded as a starting point for a trend that decreases during the year.

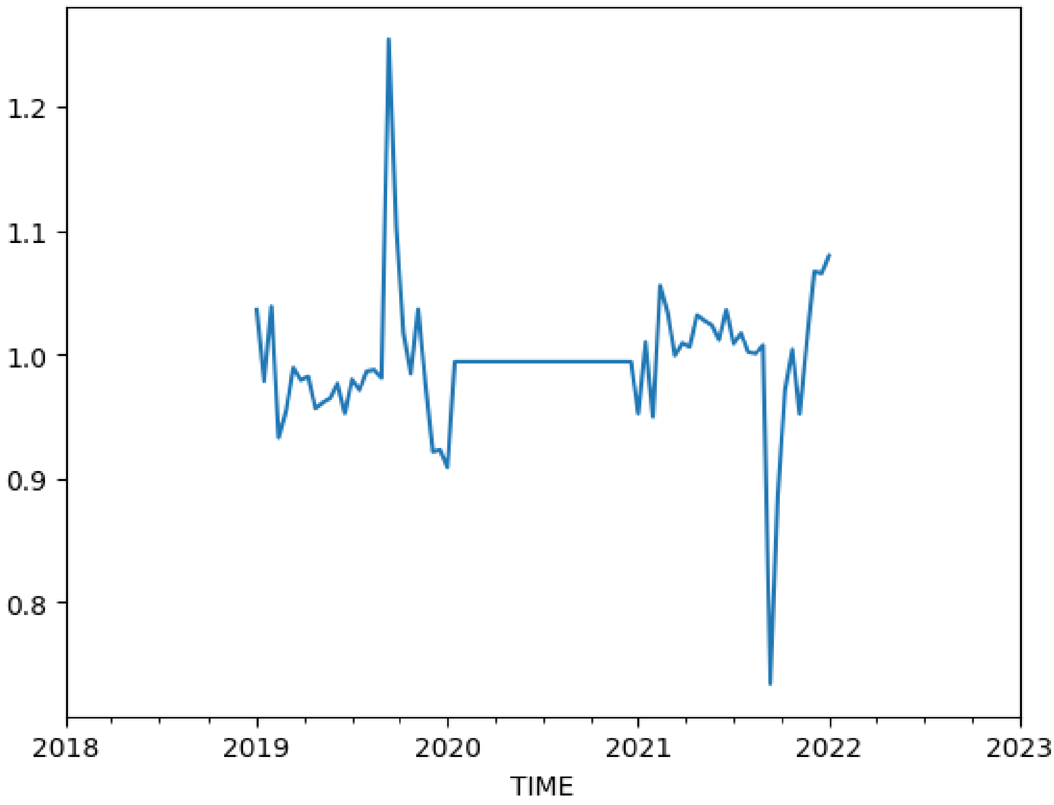

Figure 25: Electricity Consumption in Finland, Data for two-week periods, Decomposed Residual Pattern

The residual plot does not confirm this, as the peak of consumption in 2019 is acknowledged in figure 25. An interesting observation in the residual plot could be made that the end of 2021 has a spike down. Overall, the amount of residual decreased vastly compared to figure 18.

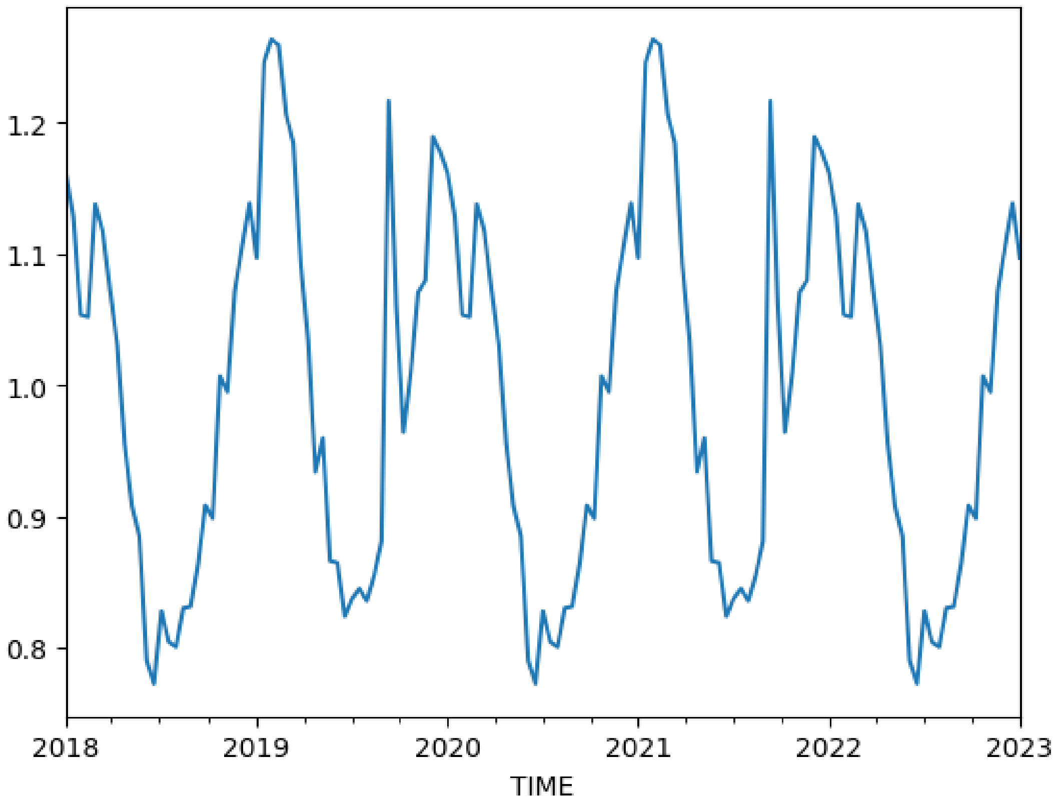

Figure 26: Electricity Consumption in Finland, Data for two-week periods, Decomposed Seasonality Pattern

In comparison to the decomposed seasonality of the time series for weekly data in figure 17, the pattern is could be regarded as less regular. One noticeable observation is that the spikes during fall of 2019 and 2021 are higher compared to the spikes illustrated for same time periods in figure 17.

3.2.3 Monthly consumption

Resampled data for monthly consumption of electricity includes 60 datapoints. The mean for the data is 6905886,75, the standard deviation is 1001736.82. The values fluctuate between 5348632 and 9120657.

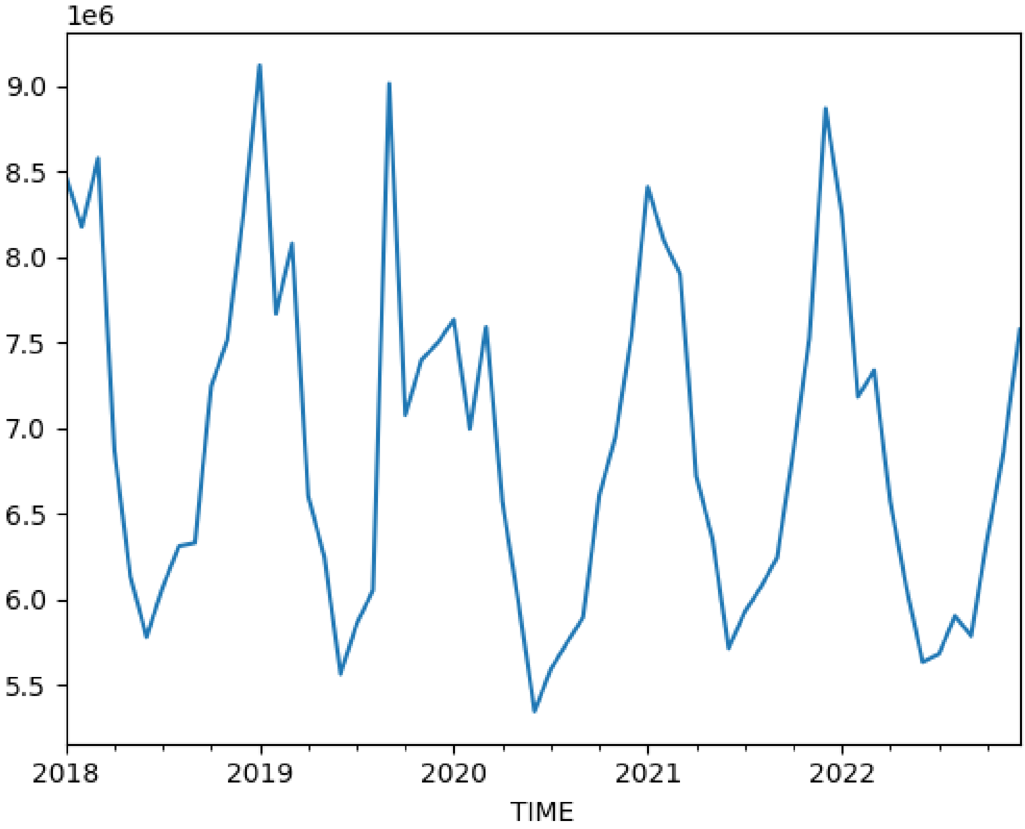

Figure 27: Electricity Consumption in Finland, Data for monthly periods

As expected, the variance in observations is smoothed further.

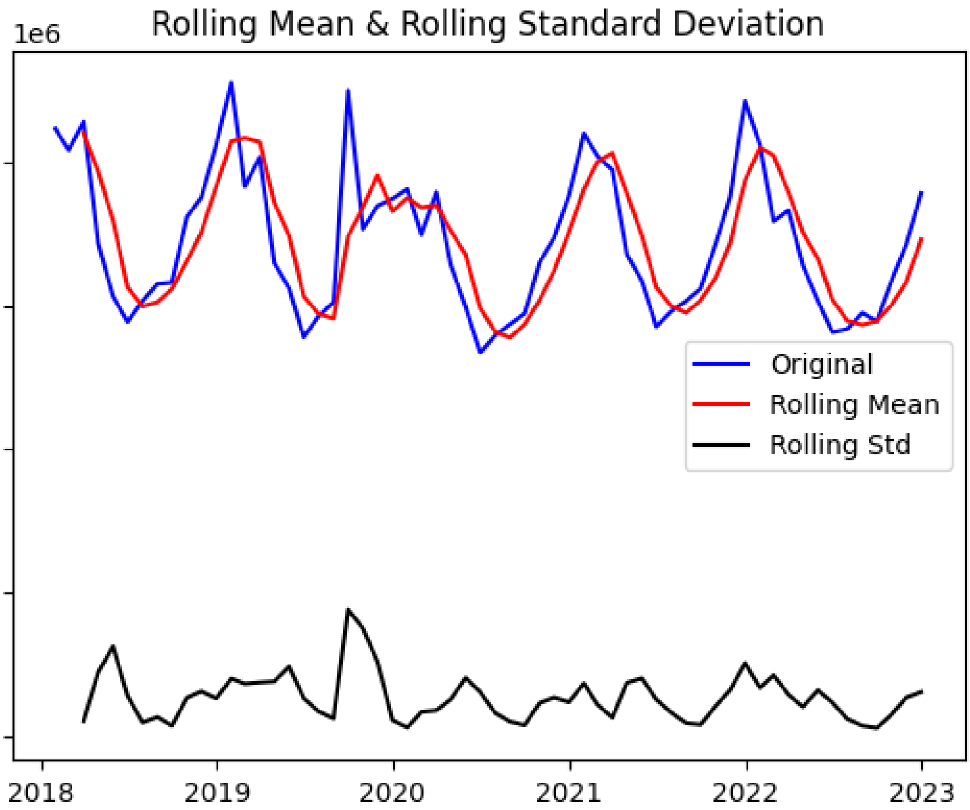

Figure 28: Electricity Consumption in Finland, Data for Monthly Periods, Rolling Mean and Std

The rolling mean of measures is close to the actual values and the same peaks in rolling standard deviations that are illustrated in figure 20 are included.

The ADF test results in a p-value of 0,00000037 with rejects the null hypothe sis and the time series can be regarded as stationary.

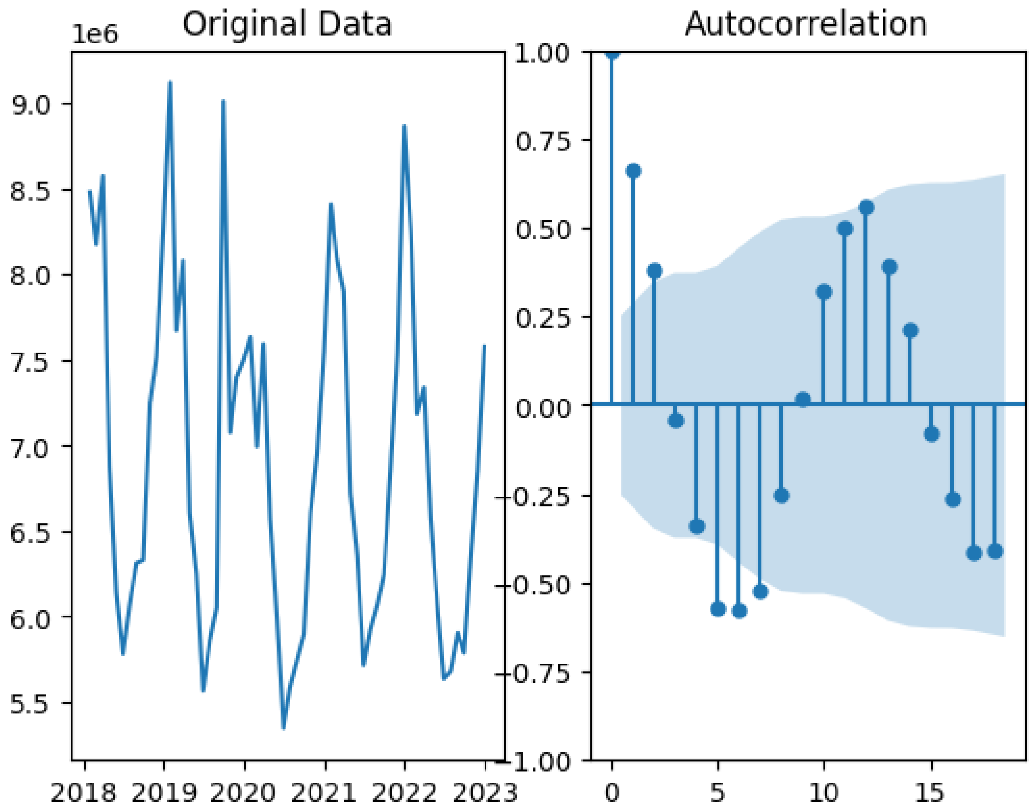

Figure 29: Electricity Consumption in Finland, Data for Monthly Periods, Autocorrelation

The observation period for the autocorrelation plot spans over 1,5 years, therefore the correlations decrease and increase during the period.

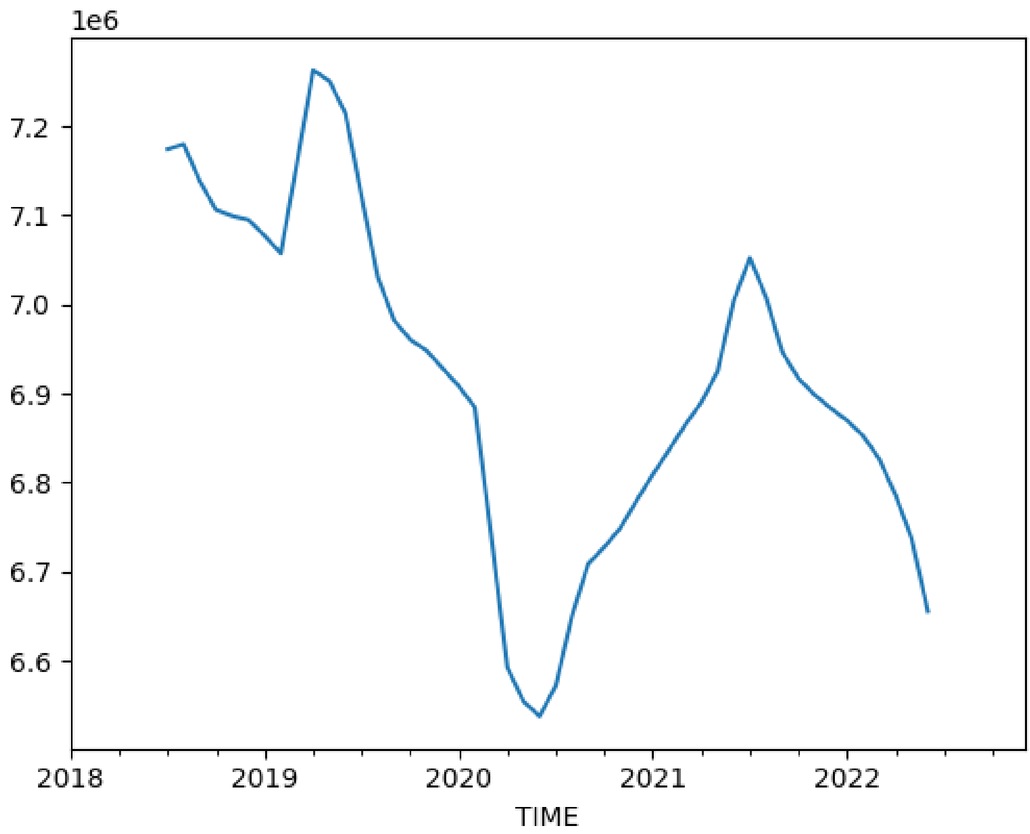

Figure 30: Electricity Consumption in Finland, Data for Monthly Periods, Decomposed Trend Pattern

The decomposed trend pattern yet again displays a decreasing trend. What is notable in Figure 30 is that the trend begins from the late 2018, compared to the start of 2019 in figure 24.

The seasonal pattern for monthly data is looking clearer compared to seasonal patterns in figure 26 and 17.

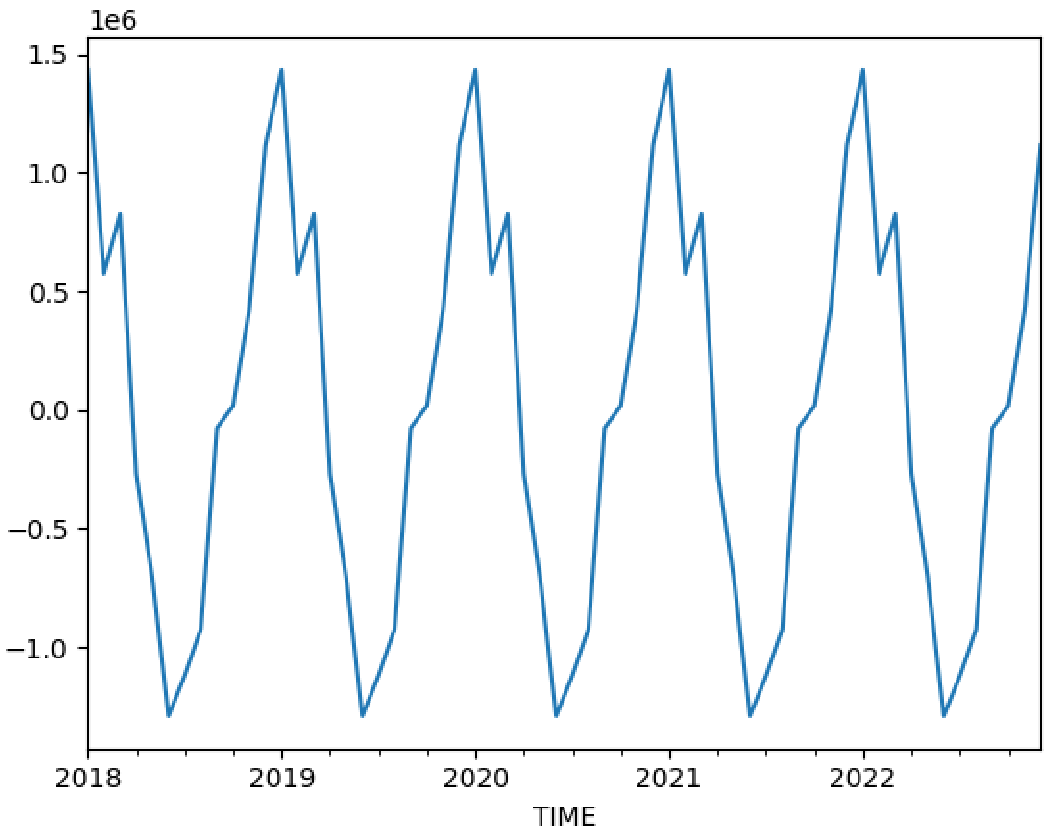

Figure 31: Electricity Consumption in Finland, Data for Monthly Periods, Decomposed Seasonal Pattern

The seasonal pattern is regular, and in figure 31 no other spikes can be detected, as in figure 26 and 17. The downward spike during the start of the year could be due to Christmas holidays, as the utilization rate of industrial machinery drops.

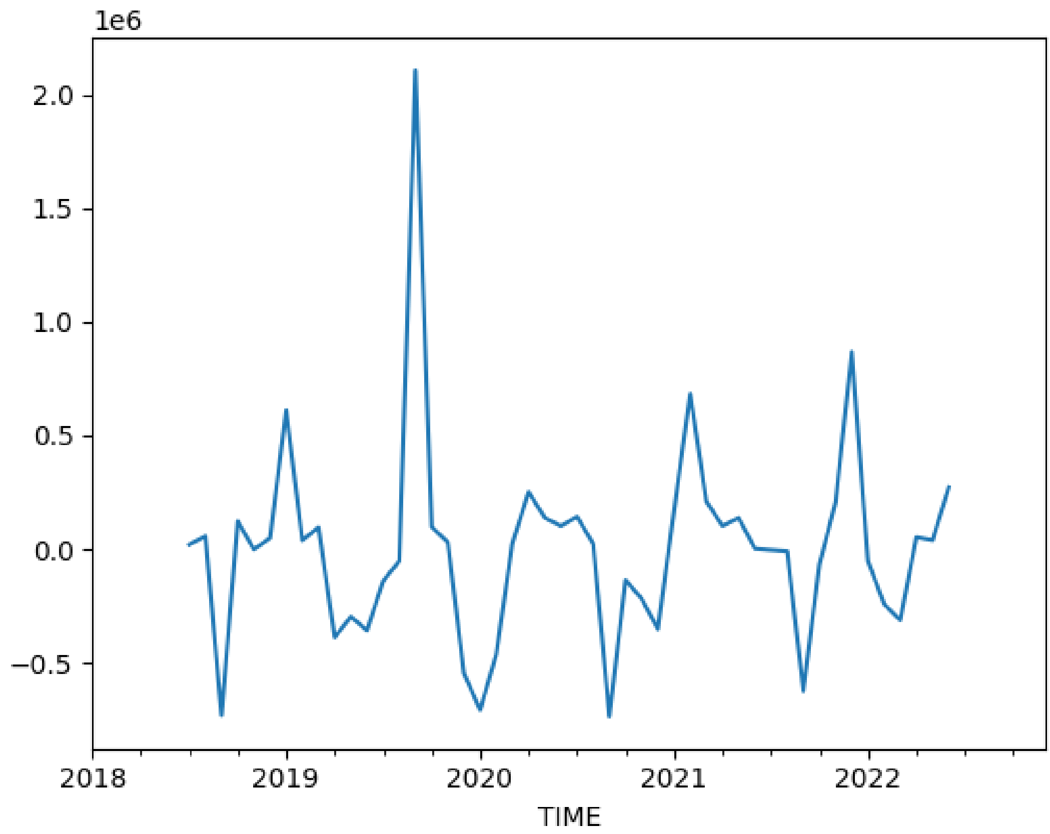

Figure 32: Electricity Consumption in Finland, Data for Monthly Periods, Decomposed Residual Pattern

Surprisingly, the residual plot has more fluctuation compared to figure 25. Notably, the average residual value is close or under zero.

3.3 Models

The models used in empirical study are ARIMA and Seasonal ARIMA by pmdarima. The models are constructed by using the auto_arima function. The start values for p and q are 0 and the max values are 5 for p and q. The function loops trough the possible combinations for p, d and q and selects the best fitting model for the final forecast based on their AIC value.

For seasonal ARIMA, the model additionally seeks the best fitting values for P, D and Q. The seasonal period for the models is 52 observations for the weekly data, 26 observations for the fortnight data and 12 observations for the monthly data.

The data is split into test and train data with the purpose of conducting a forecast of 6 months.

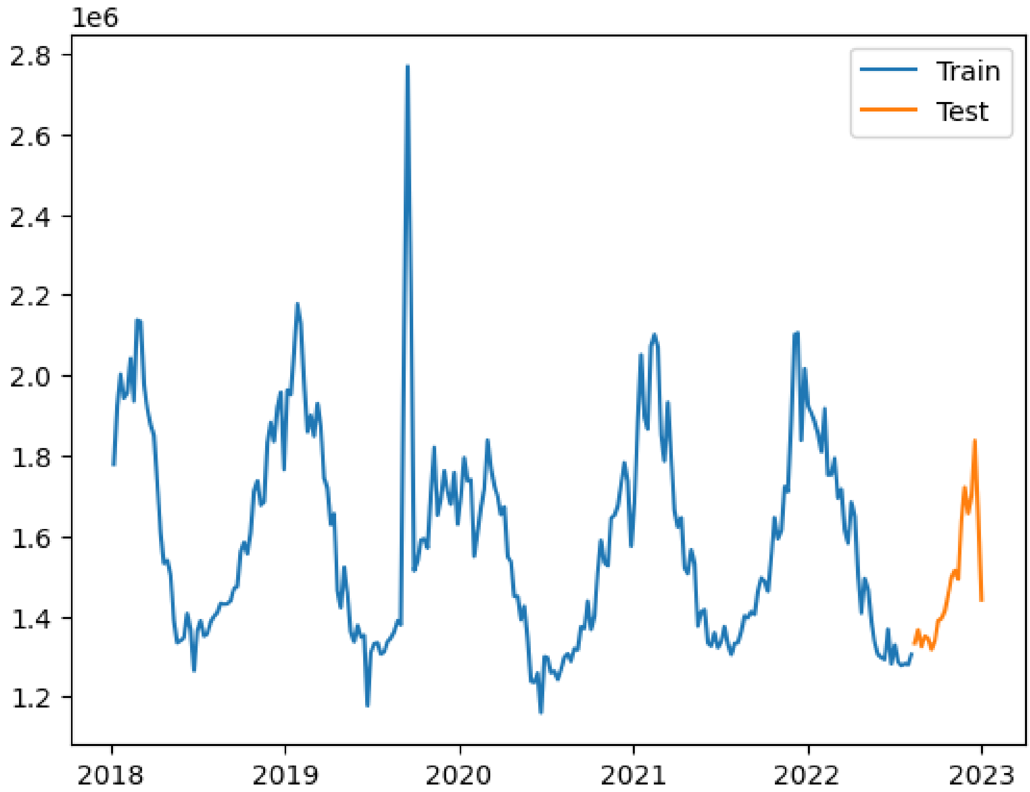

Figure 33: Test-Train Split, Weekly Data

For the weekly data, the train data consists of 240 observations and test data of 21 observations. The size of train data for fortnightly consumption is 120 observations and the size of test data is 11 observations. The monthly consumption data is split to train data of 54 observations and test data of 6 observations.

The accuracy of the forecasting models is measured by mean absolute percentage error. As other forecasting error metrics are dependent on the scale of observations used and the study is conducted to 3 datasets with different time periods, mean absolute percentage error enables the possibility to compare the performance of all models.

3.4 Results

Auto_arima selected the following models for the datasets:

Weekly data	ARIMA (2, 1, 0)	SARIMA (2, 1, 0)(1, 0, 0) 52
Two-week period data	ARIMA (1, 0, 0)	SARIMA (1, 0, 0)(0, 0, 2)26
Monthly data	ARIMA (2, 0, 5)	SARIMA (1, 0, 0)(1, 0, 0)12

Table 2: Models Chosen for the Study

For weekly data, the optimal ARIMA model had a p of 2, d of 1 and q of 0. The seasonal terms were 1 for P and 0 for D and Q. The terms selected for fortnight data were 1 for p, and 2 for Q. The terms d, q, P and D were 0. For the monthly data, the terms for optimal models had some variation as the ARIMA had the terms 2 for p, 0 for d and 5 for q whereas the seasonal ARIMA had 1 for p and the terms d and q were 0. The seasonal terms for monthly data were 1 for P and 0 for D and Q.

3.4.1 Weekly data results

Mean absolute percentage error on the weekly time period for ARIMA (2, 1, 0) was 11,3%, which translates to the accuracy of the forecast being 88,7%. For the SARIMA (2, 1, 0)(1, 0, 0) 52, the mean absolute percentage error was 9,3 %, which translates to the accuracy of 90,7%.

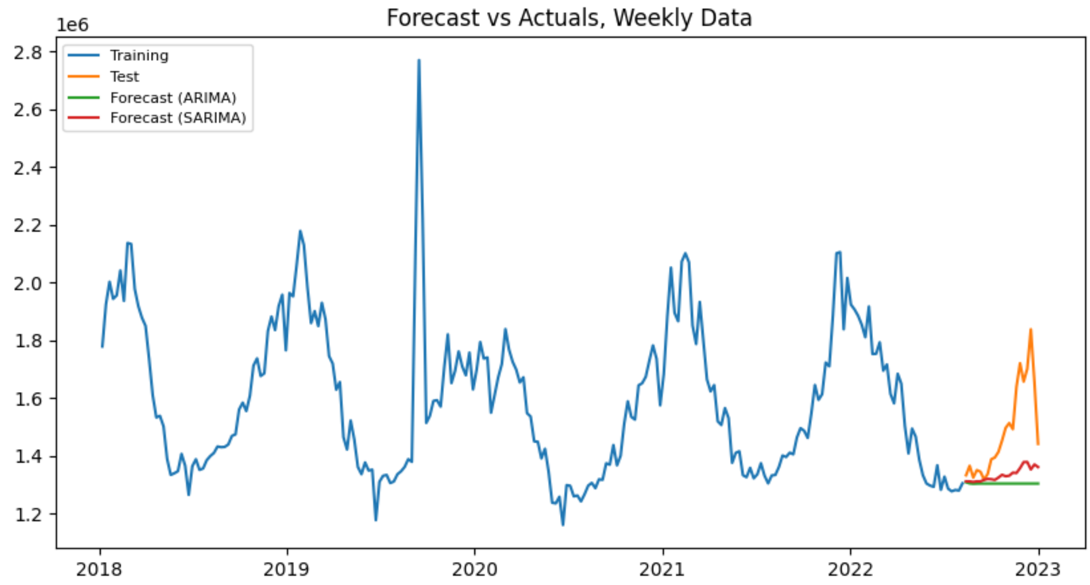

Figure 34: Forecast Values Visualized, Weekly Data

Figure 34 visualizes the performance of both models and compares them to the actual values, labeled test. The forecasting period starts from the summer of 2022, which is typically at the bottom of yearly consumption. The visualization illustrates the inability of the ARIMA to foresee the increase of consumption, and the predicted values have no fluctuation. The SARIMA model is able to predict some increase in consumption, presumably based on the seasonality.

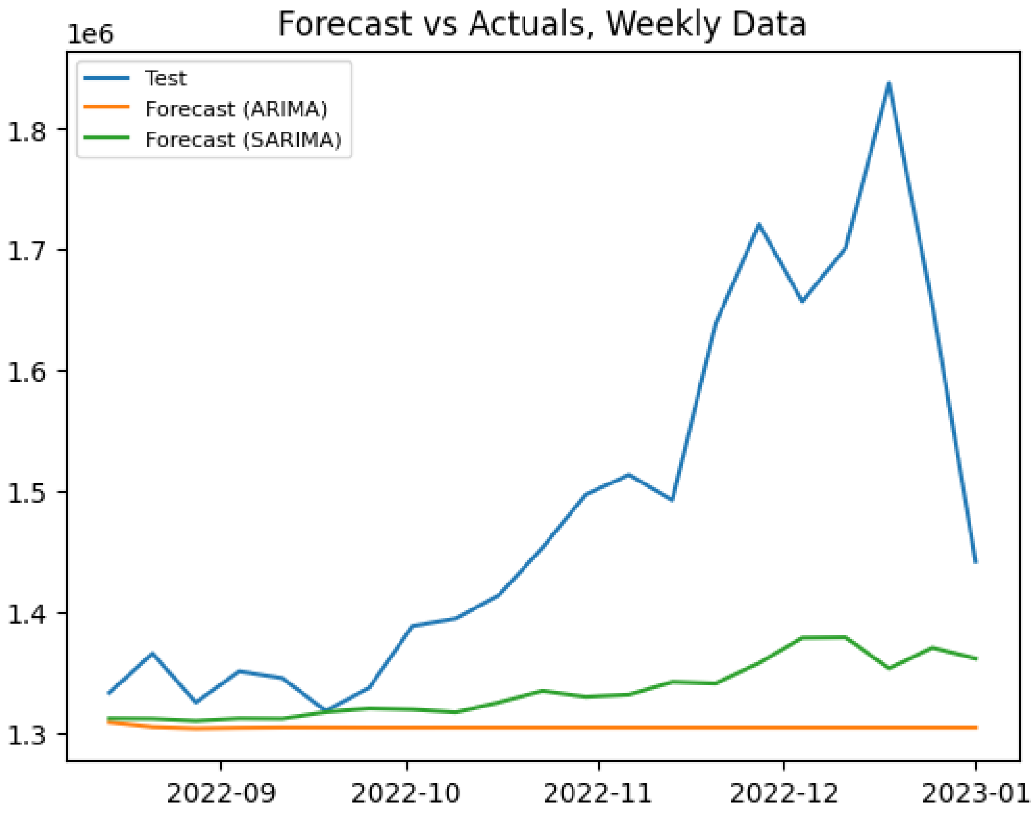

Figure 35: Forecast Values Visualized, Weekly Data Scale 2

The visual inspection on a closer scale reveals that the fluctuations in the values of ARIMA are miniscule, and the difference to forecast values from the SARIMA model is evident.

3.4.2 Two-Week Period Results

Mean absolute percentage error on the two-week time period for ARIMA (1, 0, 0) was 6,3%, which translates to the accuracy of the forecast being 93,7%. For the SARIMA (1, 0, 0)(0, 0, 2)26, the mean absolute percentage error was 4,54 %, which means the accuracy of 95,46%.

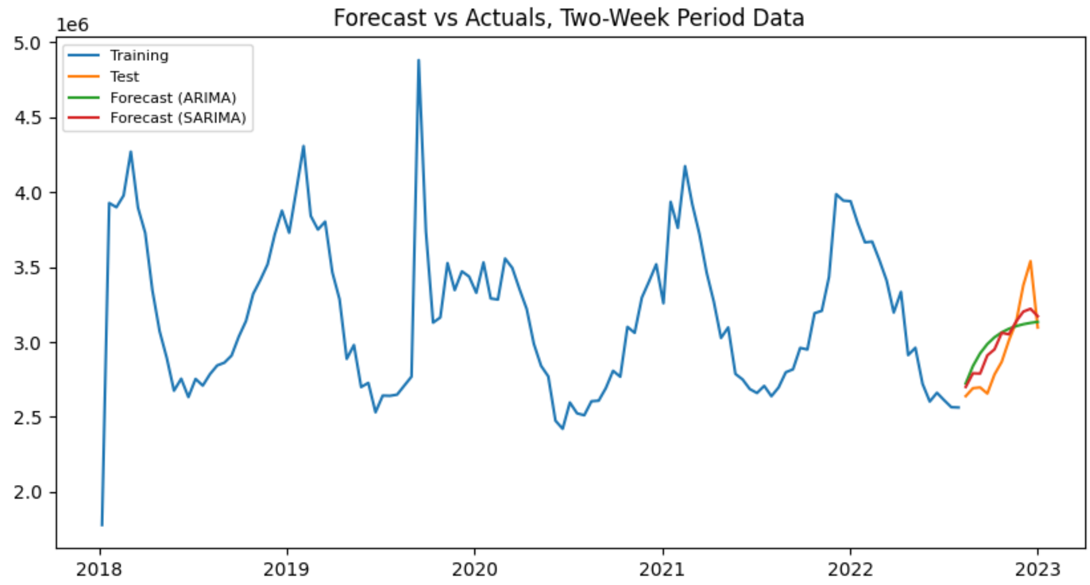

Figure 36: Forecast Values Visualized, Data for Two-Week Periods

Figure 36 visualizes the predictions for both of the values and compares them to the actual values. Both of the models were able to predict the yearly increase in consumption and both of the models can be regarded as accurate. Based on the visual inspection of figure 36, the Sarima model was able to predict fluctuation in the consumption, whereas the ARIMA was only able to predict the trend pattern for the rest of the year.

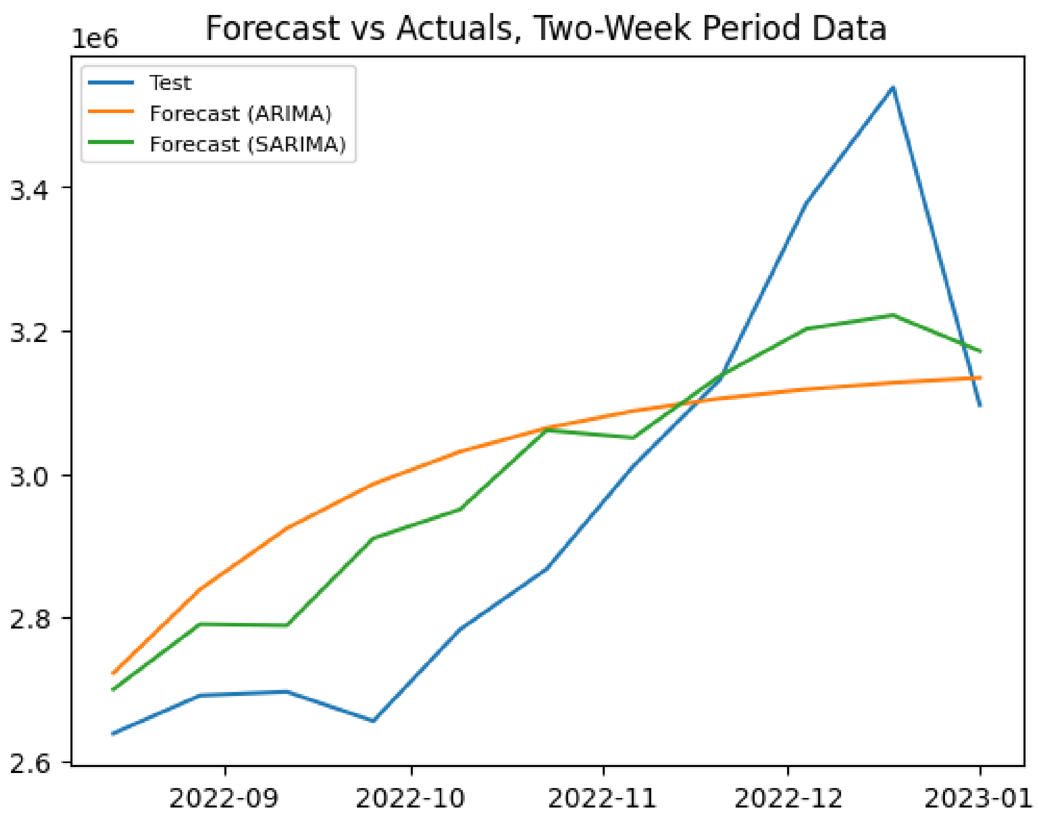

Figure 37: Forecast Values Visualized, Data for Two-Week Periods Scale 2

Closer inspection of the values confirms the expectation that the SARIMA model was able to keep the forecasted values closer to the actual values, whereas the ARIMA predicted a continuous trend value.

3.4.3 Monthly Data Results

Mean absolute percentage error on the two-week time period for ARIMA (2, 0, 5) was 5,73%, which translates to the accuracy of the forecast being 94,27%. For the SARIMA (1, 0, 0)(1, 0, 0)12, the mean absolute percentage error was 7,96 %, which means the accuracy of 92,04%. What is notable for this timeframe is that the ARIMA model performed better than the SARIMA.

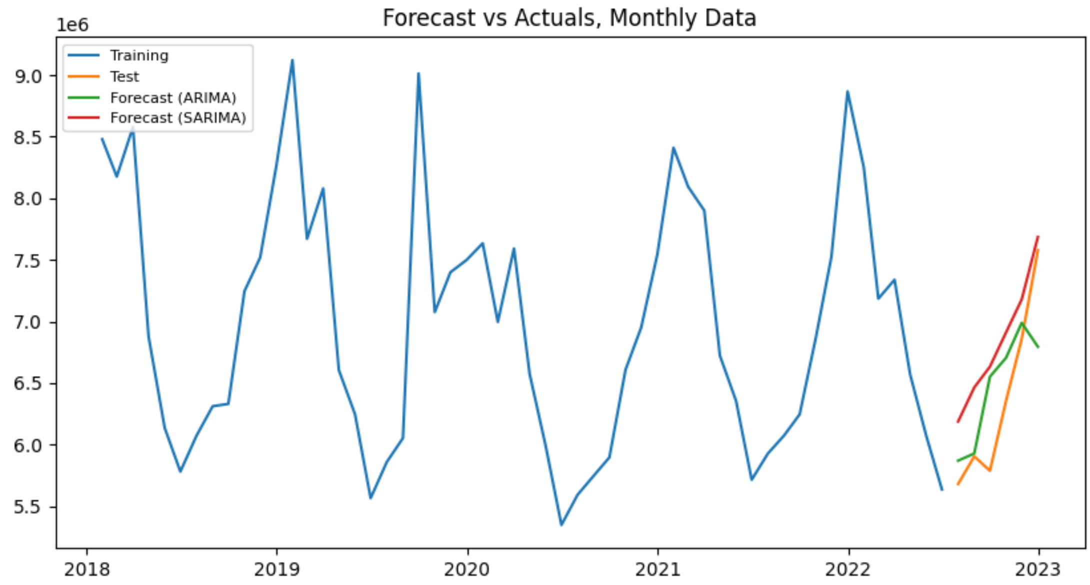

Figure 38: Forecast Values Visualized, Monthly Data

Inspecting figure 38, it is clear that the lower accuracy of the SARIMA model is due to its forecasting values being too high. The forecasted values for the ARIMA model are higher compared to actual values, but the line is notably closer than the line for SARIMAs predicted values.

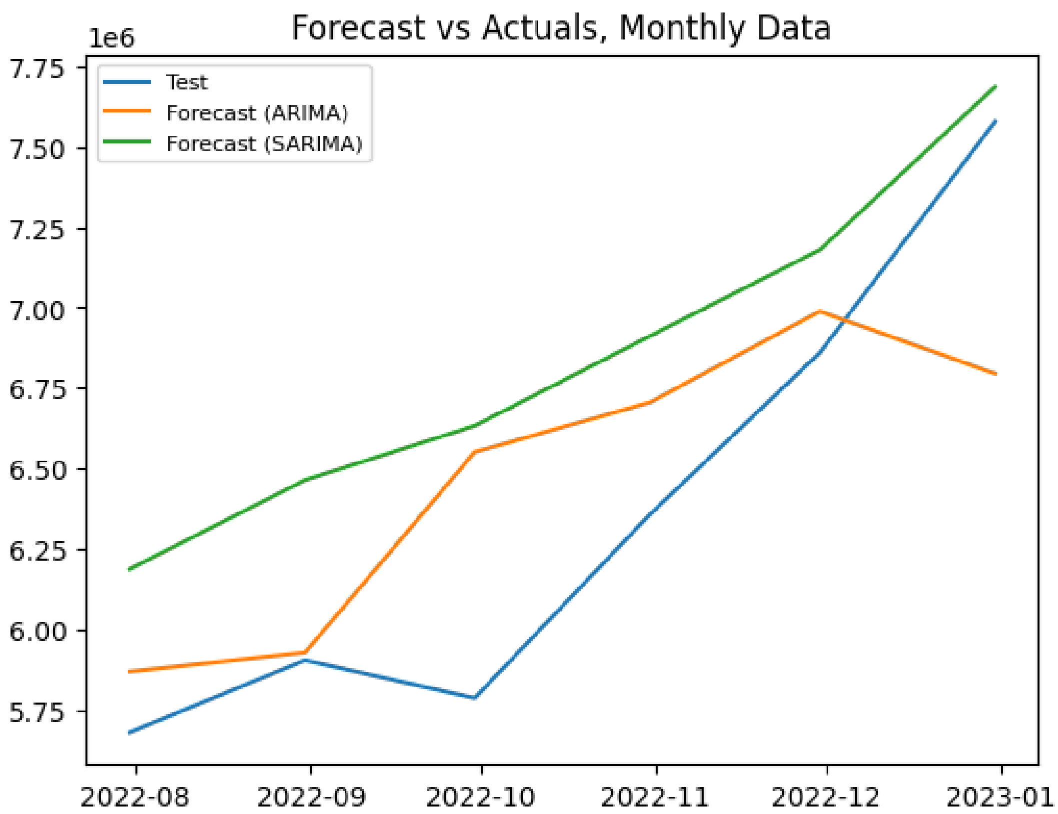

Figure 39: Forecast Values Visualized, Monthly Data, scale 2

Closer inspection of the values confirms that the SARIMA did predict too high values for the electricity consumption, thus resulting in a significantly lower accuracy. What is notable in the ARIMA line is that the last value it predicted broke the trend of increasing consumption.

3.4.4 Summary of results

The study was conducted using optimal ARIMA and SARIMA models to predict electricity consumption data at weekly, two-week, and monthly intervals. For weekly data, the SARIMA (2, 1, 0)(1, 0, 0) 52 model was more accurate than the ARIMA(2, 1, 0) model. For two-week data, the SARIMA (1, 0, 0)(0, 0, 2)26 model was more accurate than the ARIMA(1, 0, 0) model. For monthly data, the ARIMA (2, 0, 5) model was more accurate than the SARIMA(1, 0, 0)(1, 0, 0)12 model.

Period	ARIMA MAPE	SARIMA MAPE
Weekly data:	11,3%	9,3%
Two-week period data	6,3%	4,54%
Monthly data:	5,73%	7,96%

Table 3: Study Results Summary

Overall, visualizations of the forecasted values compared to actual values showed that the SARIMA models were better able to predict fluctuations in consumption compared to the ARIMA models. However, the ARIMA model for monthly data was more accurate than the SARIMA model, due to the SARIMA model predicting higher values.

4 Conclusions and Discussion

This chapter discusses and concludes the results of the empirical study by answering the research questions. Following the discussion, recommendations for future research is made.

4.1 Research Questions

The aim of the study was to compare ARIMAs to seasonal ARIMAs and by doing that, answer the research questions. The thesis has succeeded to some degree in answering these questions, and the results of the empirical study were enough to give some expected answers, but also some surprising results.

1. *Which of the Time series models give the better accuracy forecasting the electricity consumption?*

Based on the results presented in the study, the SARIMA models were more accurate than the ARIMA models in predicting electricity consumption. The mean absolute percentage error (MAPE) for the SARIMA models was lower, indicating that they had a higher level of accuracy than the ARIMA models. The accuracy of the SARIMA models ranged from 90.7% to 95.5%, while the ARIMA models had an accuracy of between 88.7% to 94.3%. Thus, it can be concluded that the SARIMA models gave better accuracy in forecasting electricity consumption.

2. *Which of the models suit the dataset better?*

Overall, the SARIMA models were found to be more accurate than the ARIMA models in forecasting electricity consumption, while both the ARIMA and SARIMA models selected by the auto_arima algorithm were found to suit the datasets well. It is important to note that the performance of the models may vary depending on the specific dataset being analyzed, and it may be necessary to try different models to determine which one is most suitable for a particular dataset.

Perhaps the most interesting find of the study was that the for the monthly data, the SARIMA model could not outperform the ARIMA. Whether it was due to the ARIMA model performing better than expected or the SARIMA model performing worse than expected is up to debate. Based on the decomposed data, one could perhaps expect that due to the monthly data having such a clear seasonality pattern, the SARIMA would have performed better.

Another explanation for the difference in performance could be drawn yet again from the decomposed time series on monthly level. The amount of residual is significantly reduced when the observation values are aggregated for a longer time period. As the models are based on autoregressive moving average models, which can be disturbed by the residual values, the importance of following the existing trend might have overcome the importance of following a seasonal pattern. In other words, as the noise in the data is reduced, following the short-term trend leads to better results.

4.2 Future Research

Some areas of possible future research can be suggested based on the results of this study. As the unexpected result of SARIMAs losing in accuracy for on the monthly time period raises some questions about the optimal performance of these models.

The performance of SARIMAs could be the subject of interest to focus on. As the model performs well on other timeframes and data, it would be interesting to find an answer about the optimal amount of training data for these models. The effect of splitting data into train and test data could be examined to find the appropriate level of relative amount of data needed for the model to function properly. An alternative view on the subject could be the absolute number of observations needed for training the models in order to establish an accurate forecast.

Another standpoint regarding SARIMAs would be the seasonality factor. Should the model be given the cycle length of the larger fluctuation in the seasonality, or would it perform as well or better compared to ARIMAs if it is given the frequency of smaller cycles in the data? This is based on the finding that on monthly scale, the non-seasonal terms with minimized AIC-values were different for SARIMA and ARIMA, which was unexpected as well.

5 Summary in Swedish – Svensk sammanfattning

5.1 Utvärdering av prestanda hos tidsseriemodeller för prognostisering av elförbrukning i Finland

5.2 Introduktion

I februari 2022 genomförde Ryssland en attack på Ukraina, vilket fungerade som en slutlig katalysator för de stigande energipriserna i Europa. Efterföljande veckor började västvärlden isolera Ryssland från internationell handel. Detta resulterade i hög inflation och priserna på el påverkades som ett resultat. Situationen på marknaden är unik, på grund av att inom fem år, skiftade priset av el från en låg nivå på grund av Covid-19 till en högre nivå.

Syftet med studien är att svara på följande forskningsfrågor:

Vilken tidsseriemodell ger bättre precision för att prognostisera elförbrukning?

samt

Vilken av använda modell passar bäst för elförbruksdata?

En litteraturöversikt av teorin bakom tidsserieanalys och prognosmetoder genomförs innan en jämförelse utförs mellan två modeller - autoregressivt integrerat glidande medelvärde och säsongsmässigt autoregressivt integrerat glidande medelvärde - som använder elförbrukningsdata från Finland. Jupyter notebooks och Python används för att utveckla och bearbeta modellerna. Studien jämför de två modellerna utifrån prognosernas noggrannhet för elförbrukning i framtiden. Slutligen presenteras resultaten.

5.3 Litteraturöversikt

I tidsserieanalys tas mätvärden ofta vid regelbundna intervall, till exempel årligen, månatligen, dagligen eller timvis (Cryer, 1986). Även om värdena i en tidsserie generellt kan vara slumpmässiga, finns det möjlighet att de skiftar till relativt högre eller lägre värden över tid. Denna förändring i tidsserien kallas trend. Cykliska komponenter är återkommande sekvenser som sträcker sig över ett kalenderår. Säsongskomponenter visar regelbundna mönster över ettårsperioder. Den oregelbundna komponenten är de värden som inte påverkas av trender, cykler eller säsongsvariationer. (Anderson, Sweeney, and Williams, 1999)

Tidsserieanalys är en metod för att förutse framtida värden. Om vissa förutsättningar är uppfyllda kan kvantitativa metoder användas. En metod är glidande medelvärde, där det genomsnittliga värdet för de senaste värdena används som en prognos för nästa period. Ett annat sätt är autoregressiva modeller, där framtida värden förutspås genom att använda en kombination av tidigare värden. För att skapa en tidsserieanalys krävs att den är stationär, det vill säga att den inte är beroende av tid och inte har några förutsägbara mönster. Icke-stationära tidsserier kan omvandlas till stationära genom differentiering. (Anderson et.al, 1999)

En populär strategi för att hitta en lämplig modell för tidsserieanalys är Box-Jenkins-metoden. Den består av tre huvudsteg: identifiering, skattning och kontroll av modellen. Dessa steg kan användas flera gånger för att hitta den optimala modellen. (Bleikh & Young, 2016)

5.4 Empirisk studie

Finlands elsystem består av kraftverk, överföringsnätet, regionala nät och distributionsnät. Elsystemet är en del av det inter-nordiska systemet som sträcker sig till Sverige, Norge och östra Danmark. En del av det finska elnätet har förbindelser till både Estland och Ryssland. Finland är en del av den nordiska elmarknaden där 70 % av handeln sker på kraftbörsen i Oslo. Marknaden täcker de nordiska länderna och de baltiska staterna, men kommer att utvidgas för att täcka hela Europa, då EU strävar efter att skapa en europeisk marknad.

Enligt Statistikcentralen (2022) bestod Finlands elförbrukning 2021 av 20,5 % importerad energi och den viktigaste inhemska energikällan var kärnkraft med en andel av 26,1 %. Största delen av energin konsumeras av industrin och byggsektorn (43,5 %), medan hushållen och jordbruket tillsammans konsumerar 28,3 % av energin.

Studiens data består av timvisa elförbrukningsdata i Finland från Fingrid mellan 2018 och slutet av 2022. Data är visuellt säsongsbetonade och varierar mellan kalla vinter- och varma sommarperioder.

För att jämföra noggrannheten hos modellerna på olika tidsperioder aggregerades datamaterialet till veckovis, tvåveckors och månatlig förbrukning. ARIMA- och säsongsbetonade ARIMA-modeller användes för att göra prognoser för en period av 6 månader. Valet av modellernas parametrar utfördes med funktionen Auto_arima, som värderar olika kombinationer av parametrar för modellerna och väljer de optimala för studien.

Auto_arima valde följande modeller för datamängderna:

- Veckodata: ARIMA (2, 1, 0), SARIMA (2, 1, 0)(1, 0, 0) 52
- Tvåveckorsdata: ARIMA (1, 0, 0), SARIMA (1, 0, 0)(0, 0, 2) 26
- Månadsdata: ARIMA (2, 0, 5), SARIMA (1, 0, 0)(1, 0, 0) 12

Modellernas noggrannhet mäts med procentuella fel, kalkylerat med skillnaden mellan prognostiserade värden och faktiska värden i data.

5.5 Resultat

För veckodata och tvåveckorsdata var prognosen med SARIMA mer exakt än ARIMA. För månadsdata var ARIMA mer exakt än SARIMA. Sammantaget visade visualiseringar av prognosvärden jämfört med faktiska värden att SARIMA-modellerna var bättre på att förutsäga fluktuationer i konsumtion jämfört med ARIMA-modellerna. Dock var ARIMA-modellen för månadsdata mer exakt än SARIMA-modellen, eftersom SARIMA-modellen förutspådde högre värden.

Studiens syfte var att jämföra ARIMA-modeller med säsongsbetonade ARIMA-modeller (SARIMA) och därigenom besvara forskningsfrågorna. Avhandlingen har till viss del lyckats besvara dessa frågor, och resultaten från den empiriska studien gav både förväntade och överraskande resultat.

Forskningsfrågan om vilken tidsseriemodell som prognostiserar elförbrukning med bäst precision kan besvaras med att SARIMA-modellerna var mer exakta än ARIMA-modellerna för att förutsäga elförbrukningen. MAPE var lägre för SARIMA-modellerna, vilket indikerar att de hade en högre noggrannhet än ARIMA-modellerna. Noggrannheten för SARIMA-modellerna varierade mellan 90,7 % och 95,5 %, medan ARIMA-modellerna hade en noggrannhet mellan 88,7 % och 94,3 %. Därför kan det konstateras att SARIMA-modellerna gav bättre noggrannhet vid prognos av elförbrukningen.

För at svara den andra forskningsfrågan kan det sägas att på generellt sett var SARIMA-modellerna mer exakta än ARIMA-modellerna för att förutsäga elförbrukningen. Både ARIMA- och SARIMA-modellerna som valdes av auto_arima-algoritmen ansågs passa data bättre. I framtida forskning kan det vara intressant att undersöka SARIMA-modellernas prestanda i förhållande till träningsdata och säsongsmönster för att optimera deras användning.

6 References

Anderson, D. R., Sweeney, D. J., & Williams, T. A. (1999). *Statistics for business and economics* (7. ed.). South-Western.

Bleikh, H. Y., & Young, W. L. (2016). *Time series analysis and adjustment: Measuring, modelling and forecasting for business and economics*. Routledge.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control* (Rev. ed.). Prentice-Hall.

Chang, Y. W., & Liao, M. Y. (2010). A seasonal ARIMA model of tourism forecasting: The case of Taiwan. *Asia Pacific journal of Tourism research*, *15*(2), 215-221.

Chen, P., Yuan, H., & Shu, X. (2008, October). Forecasting crime using the arima model. In *2008 fifth international conference on fuzzy systems and knowledge discovery* (Vol. 5, pp. 627-630). IEEE.

Chen, S.X. Lei, L. Tu, Y. 2016. Functional Coefficient moving average model with applications to forecasting chinese cpi, Statistica Sinica, volume 26.

Colton, J. A., & Bower, K. M. (2002). Some misconceptions about R2. *International Society of Six Sigma Professionals, EXTRAOrdinary Sense*, *3*(2), 20-22.

Cryer, J. D. (1986). *Time series analysis* (Vol. 286). Boston: Duxbury Press.

Detry, P. J., & Gregoire, P. (2001). Other evidences of the predictive power of technical analysis: the moving averages rules on European indexes. *Available at SSRN 269802*.

Eni, D. (2015). Seasonal ARIMA modeling and forecasting of rainfall in Warri Town, Nigeria. *Journal of Geoscience and Environment Protection*, *3*(06), 91.

Fingrid. (2017, May 18). Electricity System of Finland. Retrieved October 31, 2022, from https://www.fingrid.fi/en/grid/power-transmission/electricity-system-of-finland/

Hansun, S., & Kristanda, M. B. (2017, November). Performance analysis of conventional moving average methods in forex forecasting. In *2017 International Conference on Smart Cities, Automation & Intelligent Computing Systems (ICON-SONICS)* (pp. 11-17). IEEE.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Hyndman, Rob J., and Anne B. Koehler. "Another look at measures of forecast accuracy." *International journal of forecasting* 22.4 (2006): 679-688.

IEA (2022), *Energy Fact Sheet: Why does Russian oil and gas matter?*, IEA, Paris https://www.iea.org/articles/energy-fact-sheet-why-does-russian-oil-and-gas-matter

Kirchgässner, G., Wolters, J., & Hassler, U. (2012). *Introduction to modern time series analysis*. Springer Science & Business Media.

Lassila, A. (2022, August 22). Energia: Helsingin kaupungin energiayhtiö Heleniltä Vanhoihin Sähkösopimuksiin jättikorotus. Helsingin Sanomat. Retrieved October 25, 2022, from https://www.hs.fi/talous/art-2000009025452.html

Lauren, S., & Harlili, S. D. (2014, August). Stock trend prediction using simple moving average supported by news classification. In *2014 International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA)* (pp. 135-139). IEEE.

Ministry of economic affairs and employment. (2022). Electricity market - ministry of economic affairs and employment. Retrieved October 31, 2022, from https://tem.fi/en/electricity-market

Nochai, R., & Nochai, T. (2006, June). ARIMA model for forecasting oil palm price. In *Proceedings of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and applications* (pp. 13-15).

Persons, W. M. 1919. Indices of business conditions. Review of Economic Statistics

Reghunath, R., Murthy, T. S., & Raghavan, B. R. (2005). Time series analysis to monitor and assess water resources: A moving average approach. *Environmental Monitoring and Assessment*, *109*, 65-72.

Siami-Namini, S., & Namin, A. S. (2018). Forecasting economics and financial time series: ARIMA vs. LSTM. *arXiv preprint arXiv:1803.06386*.

Tilastokeskus. (2022). Energy. Retrieved February 23, 2023, from https://www.stat.fi/tup/suoluk/suoluk_energia_en.html

Tilastokeskus. (2022). Sähkön Hinta Kuluttajatyypeittäin Muuttujina Kuukausi,Hintakomponentti, Sähkön Kuluttajatyyppi Ja Tiedot. PxWeb. Retrieved October 25, 2022, from https://pxdata.stat.fi/PxWeb/pxweb/fi/StatFin/StatFin__ehi/statfin_ehi_pxt_13rb.px/

Tilastokeskus. (2022, October 14). Kuluttajahintaindeksin Vuosimuutos Muuttujina kuukausi ja tiedot. Retrieved October 25, 2022, from https://pxdata.stat.fi/PxWeb/pxweb/fi/StatFin/StatFin__khi/statfin_khi_pxt_122p.px/chart/chartViewColumn/

Young, D. S. (2018). *Handbook of regression methods*. Chapman and Hall/CRC.