Time-Series Analysis of Pulp Prices

Agnes Åkerlund

MITTUNIVERSITETET - MID SWEDEN UNIVERSITY
Högskolepoäng: 30 hp
Termin/år: VT 2020
Handledare: Leif Olsson
Examinator: Aron Larsson
Kurskod/registreringsnummer: IG001A
Utbildningsprogram: Civilingenjör i industriell ekonomi

Abstract

The pulp and paper industry has a significant role in Europe's economy and society, and its significance is still growing. The pulp market and the customers' requirements are highly affected by the pulp market prices and the requested kind of pulp, i.e., Elementary Chlorine Free (ECF) or Total Chlorine Free (TCF). There is a need to predict different market aspects, where the market price is one, to gain a better understanding of a business situation. Understanding market dynamics can support organizations to optimize their processes and production. Forecasting future pulp prices has not recently been done, but it would help businesses to make decisions that are more informed about where to sell their product. The studies existing about the pulp industry and forecast of market prices were completed over 20 years ago, and the market has changed since then in terms of, e.g., demand and production volume. There is a research gap within the pulp industry from a market price perspective.

The pulp market is similar to, e.g., the energy industry in some aspects, and time-series analysis has been used to forecast electricity prices to support decision making by electricity producers and retailers. Autoregressive Integrated Moving Average (ARIMA) is one time-series analysis method that is used when data are collected with a constant frequency and when the average is not constant. Holt-Winters model is a well-known and simple time-series analysis. In this thesis, time-series analysis is used to predict the weekly market price for pulp the three upcoming months, with the research question "With what accuracy can time-series analysis be used to forecast the European PIX price on pulp on a week-ahead basis?". The research method in this thesis is a documentary study leading to a case study where data are collected through the data collection method documents. First, articles are studied to gain understanding within the problem area leading to the use of the artefact time-series analyses and a case study. Then, historical data are collected from the organization FOEX Fastmarkets, where a new market price of pulp has been released every Tuesday since September 1996. The dataset has a total of 1200 data points. After data cleaning, it is merged to 1196 data points that are used for the analysis. To evaluate the results from the time-series analysis models ARIMA and Holt-Winter, Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are used. The software RStudio is used for programming. The results shows that the ARIMA model provides the most accurate results. The mean value for MAE is 16,59 for ARIMA and 44,61 for Holt-Winters. The mean value for MAPE is 1,99% for ARIMA and 5,37% for Holt-Winters.

Keywords: ARIMA, Holt-Winter, time-series analysis, market price, pulp price forecasting.

Contents

List of Abbreviations

AR – Autoregressive

ARIMA – Autoregressive Integrated Moving Average

ARMA – Autoregressive Moving Average

ECF – Elementary Chlorine Free

IOHMM – Input-Output Hidden Markov Model

MA – Moving Average

MAE – Mean Absolute Error

MAPE – Mean Absolut Percentage Error

MTO – Make-to-order

MTS – Make-to-stock

NBSK – Northern Bleached Softwood Kraft

NNs – Neural Networks

TCF – Total Chlorine Free

1 Introduction

Wu et al. (2019) mention that “Production rate is defined as the number of items produced over a given period and should be used as a decision variable for profit maximization.” Producers should strive to maximize profit in the long run and need to understand the market to be able to make the right decisions (Wu et al., 2019). In a make-to-stock (MTS) production, some decisions need to be made regarding what to produce and what volumes to meet future customer requirements, to achieve maximum profit. There exist policies using an MTS production combined with a make-to-order (MTO) production, under which, the left-over production capacity besides the orders produced needs to be determined (Hadj Youssef et al., 2018).

Several factors affect the market situation and an organization's profit for a specific product. Wu et al. (2019) and Hadj Youssef et al. (2018) mention optimal pricing, production rate, and warranty length. The model used for deciding the value of these factors includes the demand and the inventory cost, which is a factor whose value should be decreased to achieve maximum profit. In general, lower pricing of products often generates increased sales volume and reduced profit (Hadj Youssef et al., 2018; Wu et al., 2019). However, there are business situations where the producers and sellers act as price takers based on an exogenous market price. Lübkert et al. (2018) mention a well-known mechanism for predicting usage, where suppliers inform customers about the future predicted market prices, to receive their anticipated future demand at that given market price. That demand is affected by prices shows that the need for a product is dependent on the market price at a specific time. They highlight the value of response to the customer demands and to produce the right amount of their product given a particular market price, for an optimized process. Wu et al. (2019) recommend that future studies should explore the effects of different market variables and to focus on market dynamics.

The pulp industry has a significant role in the economy and society of Europe as the use of paper products continuously increases (González et al., 2009), e.g., because of the phase-out of plastic material. In 2019, the total production of pulp was more than 12 million tons, see Figure 1, and 35% of the volume was sold in the market while the rest were for internal usage at the producing companies. In 2018 more than 4 million tons of northern bleached softwood kraft (NBSK) were produced, one percent increase from 2017. 76% of the NBSK pulp deliveries in 2018 were to Europe (Skogsindustrierna, 2020). SCA Östrand pulp mill located in Timrå, Sweden, produces NBSK pulp among other products with a combined MTS/MTO strategy. This pulp mill has a constant production rate, where other factors affecting the profit. The NBSK pulp consists of two main families, Elementary Chlorine Free (ECF) and Total Chlorine Free (TCF), with different brightness. In 2015, SCA decided to invest SEK 7.8 billion in a new pulp mill at Östrand, where the yearly production of NBSK increased from 430,000 tons to 900,000 tons in 2019, thereby enabling it to become one of the biggest suppliers in the pulp industry (SCA, 2020).

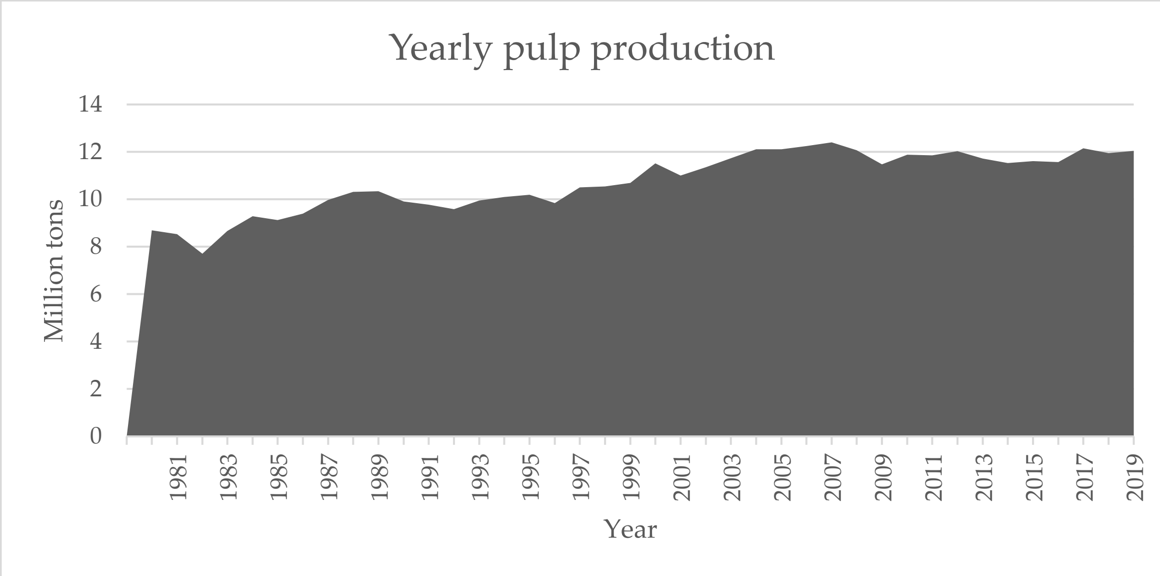

Figure 1 Yearly pulp production (Skogsindustrierna, 2020)

In the pulp industry, uncertain market prices should be taken into account when planning production. Figure 2 demonstrates that the PIX pulp price varies considerably, where the market price is released by an independent organization (Fastmarkets FOEX, 2019). The market price has a significant impact on the demand for different products. In general, if the pulp prices are high then more TCF pulp is requested than if the pulp prices are low. ECF pulp is always requested but is of more interest if pulp prices are low. The price is a signal about which market to sell the product in for maximum profit. In general, if the pulp prices are high, then more pulp is delivered to the Asian market than if the pulp prices are low. The main markets for SCA Östrand pulp mill are the U.S. and Europe. Pulp is often the primary raw material for the buyer and, therefore, is a significant portion of the costs for the organization. This market situation can be found in other market segments as well. Future market prices, not able to be affected by the producing company, can help gain a better understanding of what to produce and where to sell a product to maximize profit. Predicting market prices can also be used in other areas to gain better business insights. Toppinen et al. (1996) forecast the Finnish pulp export price with a dynamic approach where the international stock of pulp is included as a primary factor. However, they cannot manage to forecast any turning points in the pulp price since the model is described as being “too crude”. Future studies are recommended within the area of the pulp industry and its business cycle (Toppinen et al. 1996). Malcolm

(1999) investigates time-series analysis and other methods within the pulp industry, with several factors included besides the market price. The pulp market has changed in terms of market structure, e.g., production volume, customers' demands, and customers' products (Jonsson & Sörensson, 2014; Skogsindustrierna, 2020). The pulp market has evolved since the previous studies made, and therefore, an updated analysis should be considered about predicting market price of pulp.

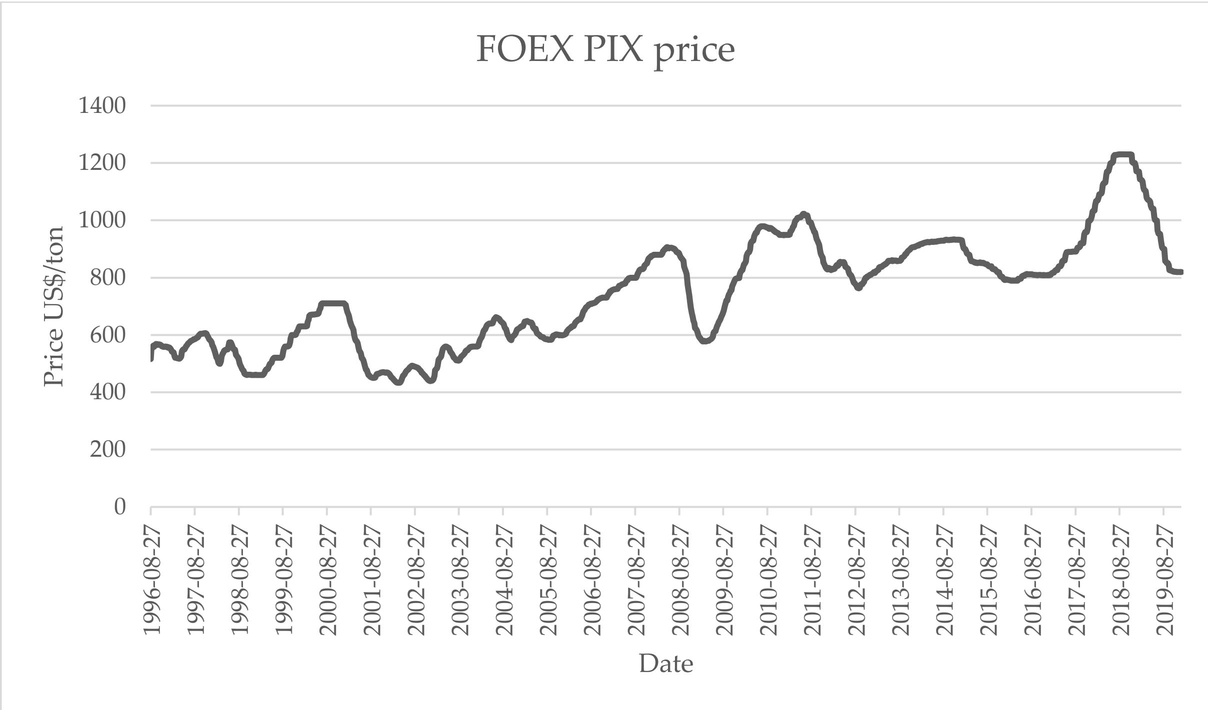

Figure 2 European PIX Price in US$ per ton from August 1996 to August 2019 (Fastmarkets FOEX, 2019)

1.1 Research Question

This thesis aims to investigate the use of time-series analysis to forecast weekly pulp prices for the upcoming three months.

The research question in this thesis is "With what accuracy can time-series analysis be used to forecast the European PIX pulp price on a weekly-ahead basis?"

1.2 Scientific contribution

This thesis contributes to research within the forest industry and especially the pulp market. Time series analysis are applied to a new case with the aim to forecast market prices, where a third party decide market prices based on the market situation. The results could help the forest industry and pulp producers maximize profit as well as businesses with similar market dynamics. Further research within different market aspects are recommended by Wu et al. (2019) and Lübkert et al. (2018). They mention that there is a need to explore different market variables, where the market price is one aspect. Toppinen et al. (1996) recommend future studies within the business cycles of the pulp industry. There are markets who have similar market dynamics within pricing and market value such as the pulp industry. For example within the industry of metals and mining, where the market prices are released from a third party and where it is depending on several market parameters and other external factors. The pulp prices also affect e.g. the paper industry where pulp is the main raw material, the pulp prices affects the profitability at the paper producers as well.

1.3 Delimitations

This study will include only the European PIX Price for NBSK pulp and exclude the U.S. and Chinese prices and other types of pulp. The research is a case study at SCA Östrand pulp mill. The market price is exogenous, where producers and sellers act as price takers. The external factors possibly affecting the pulp market price will be seen as included in the market price, and not handled as separate parameters.

This study should be seen as a proof-of-concept, and the risk of using the forecasted market prices will not be evaluated in this research. The model should be tested and assessed for several time-series and periods before it can be seen as a complete forecasting tool for the pulp market.

2 Scientific Base

2.1 Problem

Based on Wu et al. (2019) and Lübkert et al. (2018), there is a need to explore different market variables, where the market price is one aspect. Toppinen et al. (1996) recommend future studies within the business cycles of the pulp industry. In the pulp industry, the market price is a primal factor affecting, e.g., profit and demand, and it would help pulp producers with their decisions regarding production and delivery. To predict a market price could be a simple way to help suppliers gain a better understanding of the market situation and to increase the profit. At SCA Östrand Pulp Mill, the decision about what to produce is today based on the production planners' experience and expected customer needs. In the case of the pulp market, the market price (PIX) is updated every Tuesday by FOEX (Fastmarkets FOEX, 2019). The pulp mill's capacity is a limitation, and more pulp could be sold and delivered to customers if produced. Since the mill produces full speed, regardless of the number of customer orders or the market situation, the market price is the factor used to decide where to sell and deliver. Improved forecasts for prices would help make the right priorities and accurate decisions in an MTS/MTO production. There are studies forecasting market prices within similar business areas to the pulp industry. Still, there is a lack of peer-reviewed studies within the pulp market price when searching on, e.g., Google Scholar and primo. The studies found by Toppinen et al. (1996) and Malcolm (1999) were completed over 20 years ago, and an updated analysis should be considered since the market has changed in some aspects. Bianchi et al. (1998) recommend future studies in the area of time-series analysis and the ARIMA model where a specific time series should be examined individually to be able to determine if time-series analysis is appropriate. Since there is a need to consider time-series analysis on different time-series and has not been recently done within the business area of pulp, there is a research gap within the pulp industry from a market price perspective. The result could be applicable to other industries similar to the pulp industry such as the industry of metals and mining.

The problem that this thesis address is the uncertainty in the European market price within the pulp market in the area of risk and decision analysis. Time-series analysis is applied as an analysis tool. The results of the time-series analysis could help apply the method to a similar problems. For SCA Östrand, and other pulp producing companies, the result can be used to better forecast customers' needs and to plan deliveries to maximize profit.

2.2 Previous Research

Nogales et al. (2002) highlight the value of an accurate price forecasting to be competitive within the electricity market from the perspective of the producers and consumers. Market prices generated in the electricity industry are public information, but other information can be hard to find, e.g., demand curves. They mention a need to forecast both the demand and price, both in a medium-term horizon and the next-day prices, for participants to maximize their benefits. The paper focuses on the short-term decisions and forecasting of prices, i.e., a 24-price forecast should be calculated for every hour of the upcoming day, based on time-series analysis.

Guo et al. (2016) use time-series analysis to generate scenarios for electricity and gas prices, which are uncertain and often dependent on each other. Energy prices are described as "…high frequency, non-constant variance and mean, weekly seasonality, and high volatility" (Guo et al., 2016, p. 72). There are similarities regarding pulp and energy prices, e.g., regarding the high price volatility (Fastmarkets FOEX, 2019). Nogales et al. (2002) also mention "…time-series analysis has been applied with great success in other areas where the frequency of data is at most weekly…" (p. 342).

In many other markets, as well as the pulp market, there is a need to forecast the market price. Adequate forecasts can provide a basis for organizational planning, production planning, and optimization (Box et al., 2013). There are several forecast techniques, both linear and non-linear models. Artificial intelligence techniques, like neural networks (NNs) that are not linear, have received attention for successful results. However, time-series methods, such as autoregressive integrated moving average (ARIMA), dynamic regression, and Input-Output Hidden Markov Model (IOHMM), are successful techniques to use when the frequency of data points is low. Weekly patterns are considered a low frequency of data (Mandal et al. 2010). Nikolopoulos et al. (2007) mention that multiple linear regression (MLR) is popular to forecast market variables however, when data consist of independent variables. MLR can lead to less accurate forecasts because of a too high complexity or way too simple. Since the model is linear, it cannot present non-linear relations. To avoid these kinds of problems, NNs,ANNs, and simple linear regression models are recommended. (Nikolopoulos et al. 2007) Within forecasting and prediction of future values, linear models such as ARIMA and linear regression has shown better results than nonlinear models such as NNs and ANNs (Dellana & West, 2009; Hastie et al. 2009). Since there are only the date and historical price included in the predictive model, linear models handling this should be included in this study. In linear regression, there are often more than one quantitative inputs used for predicting an output (Hastie et al. 2009).

Guo et al. (2016) and Nogales et al. (2002) motivates why forecasting is essential and how time-series analysis is suitable for the kind of problem addressed in this thesis. Since PIX frequency is weekly, and there exists high price volatility, time-series analysis should be useful to forecast pulp market prices based on historical data. Since time-series analysis has been applied within similar cases to the one this thesis addresses, it will be investigated if it can be successfully used when forecasting pulp prices as well within the area of risk and decision analysis.

An article by Toppinen et al. (1996) used a dynamic forecasting model to predict the market price of Finnish pulp for export, including several parameters, where the international inventories seem to be affecting the price. They recommend further research within the pulp industry and the business cycle, “e.g. to incorporate information on pulp export demand conditions.” Since market prices have been predicted with the usage of only historical market prices in other business areas, it should be investigated if this could be applicable to forecasting pulp market prices as well. Malcolm (1999) forecasted pulp prices with different models, e.g., the time-series analysis method autoregressive moving average (ARMA) in the sample period 1976-1991. He included several factors such as shipments, capacity, investments, costs, and inventories when doing his research (Malcolm, 1999). Since the pulp market has changed in terms of market structure since these previous studies were completed (Jonsson & Sörensson, 2014; Skogsindustrierna, 2020), an updated analysis should be considered.

This thesis is based on the work of Guo et al. (2016) and Nogales et al. (2002) since the market characteristics are similar. In addition, the future recommended studies mentioned by Bianchi et al. (1998) is investigating time-series analysis on different time series and by Toppinen et al. (1996) about investigating the business cycle in the pulp industry. To evaluate results from a time-series analysis in terms of forecasting, Son et al. (2015), Son et al. (2017), and Ramos et al. (2015) use Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) for evaluation and comparison, see Figure 3. The analysis tools for the predictions will be the time series analyses ARIMA and Holt-Winters.

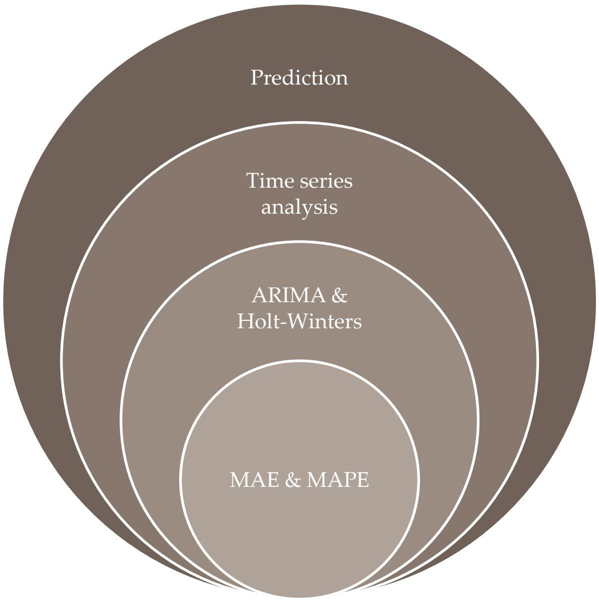

Figure 3 Predictions

2.2.1 Neural Networks and Artificial Neural Networks

Neural Networks (NNs) and Artificial Neural Networks (ANNs) are inspired from the human brains logic and way of thinking. These models uses weights on nodes in a network for processing information, learning and predicting. There are several types of NNs and ANNs, and they are used depending on what the purpose with the usage is. (Shanmuganatan, S. & Samarasinghe, S., 2016). NNs and ANNs are used in previous studies with the purpose to predict future values, among other usage areas, for example Izady et al., (2013), Barello et al., (2014), and Abdelaziz et al., (2014). However, there are situation where linear models have shown better results when predicting future values within more simple predictions (Dellana & West, 2009; Hastie et al. 2009).

2.2.2 Regression analysis

B V, B. & Dakshayini, M., (2018), mention that linear regression and logistic regression are traditional models for prediction. Linear regression models are simple and are often used when the purpose is to forecast an output value bases on a set of input variables by fitting a regression line (Equation 1). Depending on whether the output value is dependent on a single explanatory, or an independent variable, the model is called simple. If the output value is dependent on multiple variables, the regression model is called multiple as well.

$$D = I + S * X + e \tag{1}$$

Where
D = output
X = input
I = Intercept of the regression line
S = Slope of the regression line
e = random error

Tikhonov regularization, also called ridge regression, is a regression model useful when there exists multicollinearity within linear regression. Multicollinearity is when variables in a multiple regression are related linearly (Arnold et al. 2019). The factors included in this thesis is the date with historical prices to forecast the future value.

2.2.3 Time Series analysis

Time-series analysis is a method used to examine trends and cycles in time series. A time series is a set of data sequentially in time, where the data points are equally spaced. The observations are considered dependent on each other, and a time-series analysis includes techniques analyzing this. Time-series analysis is a tool to deal with correlations within data points over time, called autocorrelations. One application area of time-series analysis is to forecast future values based on past values. Many data series in industries show a non-stationary behavior and with no fixed mean value of the data points. There are several time-series analyses, e.g., ARIMA, Holt-Winter, generalized autoregressive conditional heteroscedasticity (GARCH), and Exponential Smoothing (Mitchell, 2010; Box et al., 2013).

GARCH is a model well used within time series and finance. These models includes past variances when forecasting, the variance can vary dependent on the time and are seen as flexible. GARCH is not optimal when there is high volatility within the time series (Chen & Lu, 2013).

2.2.3.1 Holt-Winter´s model and exponential smoothing

Holt-Winter´s model is algorithm is commonly used within time series analysis, and it is based on the time series analysis model exponential smoothing. The model calculates the parameters level and seasonal behavior to forecast future values. (B V, B., & Dakshayini, M., 2018) Holt-Winter's model is a time-series forecasting method that could be an alternative method for this thesis. However, there exist problems regarding forecasting in some specific future periods (Arum, 2017). Holt-Winter is preferred when the data have seasonal patterns (Shaleh et al., 2018). Exponential smoothing is also widely used (Ariyanti et al., 2018). The Holt-Winter and exponential smoothing models do not consider if the data are stationary, to predict values; they use weighted values from the data set and repetitive steps (Syafei et al., 2018). This model could however be used as for comparison to other methods and models.

2.2.3.2 ARIMA Model

One model handling complex aspects within time-series analysis is called ARIMA. ARIMA is based on the ARMA model, where AR stands for autoregressive and MA for moving average. The I, integrated, is complemented in the ARIMA model. The standard notation for ARIMA is ARIMA(*p,d,q*), where *p* is the number of lag observations, or the lag order, in the model. The second parameter, *d*, is the degree of differencing, and *q* is the order of the moving average. AR(*p*) stands for autoregression, which means that the model uses dependent relationships when analyzing observations and lagged observations. AR is a stochastic model that is useful when representing some practically occurring series. AR models characterize how the value of one variable is related to the values of the same variable in previous periods. The time *t* is represented in the time series, equally spaced, as {*t*, *t-1*, *t-2,…, t-p*} where the value of the process, y, is dependent on the time (Equation 2). The variable μ is a constant and γ_p is a coefficient for the lagged observation at time *t-p*, and ϵ_t is the error term for time *t*, also called white noise (Brownlee, 2017; Pinto Moreira de Souza et al., 2017; Box et al., 2013; Katchova, 2013).

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \epsilon_t \tag{2}$$

I, integrated, includes the use of differencing, to make the used time series stationary. Stationary defines as a process that does not change over time, i.e., the variance or expectations are always the same (Nason, 2006). MA(*q*), moving average, uses factors that are dependent and have a moving average value of the observations. The MA model, with *q* lags, have the error term ϵ_t, the variable μ is a constant, and the coefficient for the lagged error term is denoted as θ_q (Equation 3) (Brownlee, 2017; Pinto Moreira de Souza et al., 2017; Box et al., 2013; Katchova, 2013).

$$y_t = \mu + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-1} \tag{3}$$

The ARMA model combines *AR(p)* and *MA(q)* and is called *ARMA(p,q)* (Equation 4). However, the ARMA model requires stationarity, e.g., a fixed mean and variance over the time periods, where the data does not have any trends. Differenced variables are one solution if y_t is not stationary, where *d* is the number of differences needed to make the dataset stationary, denoted *I(d)*, i.e., subtract a data point from the previous data point (Equation 5) (Brownlee, 2017; Katchova, 2013).

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-1} \tag{4}$$

$$\Delta y_t = y_t - y_{t-1} \tag{5}$$

There are different ways to characterize relations between observations in the time series at different lags. To decide the values of the parameters *p*, *d*, and *q*, the Box-Jenkins method can be used. This is a process including three steps: Identification, Estimation, and Diagnostic Checking. Identification includes assessing whether the time series is stationary or which steps are necessary to make it stationary. This can be done through, e.g., a unit root test, where differencing is included, and where it is possible to decide whether the process or time series, is stationary. Other methods are e.g., autocovariance or autocorrelation. To determine the values of AR(p) and MA(q), the data can be plotted, and correlations for observations are shown. The autocorrelation function (ACF) is one plot function, which summarizes the correlations for a data series. Both the lag value and the correlation can be found in this plot. The second step of Box-Jenkins method, estimation, is to use methods that minimize an error or so-called "loss" term. Diagnostic Checking includes investigation about if the model presents results that are overfitted or if residual errors and white noise occurs. When the model is more complex than necessary and random noise are included in data, it is considered as overfitted, this can cause problems with a generalization of the model (Brownlee, 2017).

The ARIMA model and Holt-Winters can be fitted with many tools; one of them is the software RStudio with the library Forecast. This library provides tools used for analyzing and displaying time-series forecasts (R-project, 2020).

Razali et al. (2018) mention that ARIMA showed better results than Holt-Winter when forecasting; however, they also mention that Holt-Winter's method is easy to apply and widely used. Still, Wan Ahmad et al. (2013) show that this method does not produce better results than the ARIMA model for a more extended forecasting period.

2.2.4 Evaluation

To evaluate the results from the model, models for cross-validation should be used. Cross-validation methods, also called out-of-sample testing, are used for evaluating forecasting performance, where data not used in the forecasting are included for comparison and evaluation. (Kaeck et al., 2018; Tashman, 2000; Vadovsky et al., 2016).

Mean Square Error (MSE) is a model often used for evaluation of a linear regression, and it measures the average value of the squared error (Equation 6).

$$MSE = \frac{1}{N}\sum_{i=1}^{n}(actual - predicted)^2 \tag{6}$$

Mean Absolute Percentage Error (MAPE) is a commonly used method to predict errors. MAPE shows a percentage of the error, i.e., the mean value between the difference between the actual and predicted values divided by the actual value (Equation 7). The model has a weakness of not being able to actual handle values equal to zero since it is division included; however, it is possible to handle these kind of problems with data manipulation. A mean value of the data points can be used. (Kaeck et al., 2018; Tashman, 2000; Vadovsky et al., 2016).

$$MAPE = \frac{1}{N}\sum_{i=1}^{n}|\frac{actual - predticed}{actual}| \tag{7}$$

Mean Absolute Error (MAE) is calculated as the mean value between the difference of the predicted and the actual values (Equation 8). This model do not have the weakness as MAPE regarding zero values.

$$MAE = \frac{1}{N}\sum_{i=1}^{n}|predicted - actual| \tag{8}$$

3 Research Method

The research method for this thesis is a documentary study leading to a case-study approach, where data are collected through the data collection method documents, see Figure 4. A case study focuses on one instance intending to provide a deep understanding of relations and events in that specific case. The case should have clear boundaries (Denscombe, 2014). The boundaries in this case are described in Section 1.3 Delimitations and the artefact is applied to a specific business area within a specific time period. In the case of investigating a method on a new business area, a case-study approach is appropriate since “Case studies focus on one (or just a few) instances of a particular phenomenon to provide an in-depth account of events” (Denscombe, 2014, p. 54). A case-study approach is suitable for this thesis since the purpose is to investigate how well time-series analysis could be applied to forecast pulp prices for a given time period.

Figure 4 Research method

A research method provides a framework for how data should be collected and analyzed (Johannesson and Perjons, 2014). For collecting data in this thesis, the data collection method documents are appropriate for answering the research question, and a documentary study is to be done. Documents are when documents are the primary resource of data used in the experiment. Quantitative data can be collected through documents, e.g., statistics, business data, and reports from organizations (Denscombe, 2014). One of the data sources is FOEX Fastmarkets PIX prices with associated dates for this thesis, which could be seen as statistics. Official statistics are included in the public domain as the type of document and are often used for academic research. The data are usually accessible via the internet (Denscombe, 2014). Denscombe (2014) mentions that the access to data, the permanence of data, and that it is cost-effectiveness are some advantages of documentary research. The main issue with documents is the credibility of the data and data source (Johannesson and Perjons, 2014). Besides historical prices, a review of literature within the areas pulp price forecasting, time-series analysis, and markets similar to the pulp market are completed to gain an understanding of the problem. Most of the articles included in this thesis are peer-reviews and full-text from e.g., Elsevier ScienceDirect Journals, ResearchGate, and SpringerLink Journals.

The purpose of analyzing data, such as historical market prices from FOEX Fastmarkets, is to understand it better and to be able to describe, explain, and interpret the data. The part about describing the data involves measurements and components, when the data points are created, who was involved in the data collection, and how often it happens. In other words, the frequency of the data is needed. The second part, to describe the data, includes explaining how, why, and when things happen - the underlying causes and correlations, for example. Interpretation digs in even more rooted in the data collection. Research can be more quantitative or qualitative. Quantitative research analyzes numbers, and qualitative research uses words or images as the analysis unit. Research data can be analyzed through a common process, including five stages, with some differences depending on whether the research is qualitative or quantitative (Denscombe, 2014):

1. Data preparation
    a. Quantitative data: Code before data collection and categorizing the collected data.
    b. Qualitative data: Catalogue and prepare data.
2. Initial exploration
    a. Quantitative data: Check data for trends and/or correlations that are obvious.
    b. Qualitative data: Look for issues or patterns that are obvious. Add comments to the data collected.
3. Analysis
    a. Quantitative data: Test data, for example, with factor or cluster analysis. Connect the analysis and results of the research question(s).
    b. Qualitative data: Code the data and use the categories for comparison. Look for concepts.
4. Presentation
    a. Quantitative data: To present the data, use figures and/or tables and explain the results.
    b. Qualitative data: Write explanations of the results after analysis. Illustrate and use visual models.
5. Validation
    a. Quantitative data: Use benchmarks and/or compare results.
    b. Qualitative data: Use method triangulation for validation or comparison.

3.1 Data preparation

In the case study, the first step, data preparation, is to collect data and data cleansing. Data can be manipulated if needed, and aggregation can be used, which collapses many values into one data point or value (Wickham, 2014). In this case, if needed, a missing data point is created from Equation 9. There is a possibility that more than 52 data points are created; if so, then two data points will be merged into one (Equation 10).

$$x_t = \frac{x_{t-1} + x_{t+1}}{2} \tag{9}$$

$$x_t = \frac{x_t + x_{t+1}}{2} \tag{10}$$

The data used in the case study in this thesis are historical pulp prices released every Tuesday since 16 September 1996. Data was collected from FOEX Fastmarkets in Excel-format including dates and prices (FOEX Fastmarkets). Since the data started to be published 16 September 1996, week 1 year 1997 will be the starting date of the data points, see an example from the data collection in table 1, for the whole data set, see Appendix 1.

Table 1 European Pix price USD from FOEX Fastmarkets

Week/Year	**1997**	**1998**	**1999**	**…**	**…**	**…**	**2017**	**2018**	**2019**
1	559,83	583,19	460	…	…	…	808,94	1003,33	1199,96
2	557,57	579,81	460	…	…	…	808,99	1016,49	1188,7
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
52	584,72	460	610,36	…	…	…	999,63	1200,02	819,95
53	-	-	-	…	…	…	-	-	-

The whole data set consist of 1235 data points, where 4 years have 53 data points in the time period. The frequency of the data points should be 52 each year, because of this some values must be merged, (Equation 10). The data points $x_{(52,2000)}, x_{(52,2005)}, x_{(52,2011)}, x_{(52,2016)}$, $x_{(53,2000)}, x_{(53,2005)}, x_{(53,2011)}, x_{(53,2016)}$ ar e merged as below. The total amount of data points are reduced to 1231 instead of 1235, 23 years including 52 data points, year 1996 including 16 data points and year 2020 including 19 data points. The 16 data points from year 1996 will be excluded, which gives a total of 1219 data points to use in the analysis.

$$x_{(52,2000)} = \frac{x_{(52,2000)} + x_{(53,2000)}}{2} = \frac{710 + 710}{2} = 710$$

$$x_{(52,2005)} = \frac{x_{(52,2005)} + x_{(53,2005)}}{2} = \frac{599{,}21 + 598{,}69}{2} = 598{,}95$$

$$x_{(52,2011)} = \frac{x_{(52,2011)} + x_{(53,2011)}}{2} = \frac{838{,}63 + 833{,}71}{2} = 836{,}1$$

$$x_{(52,2016)} = \frac{x_{(52,2016)} + x_{(53,2016)}}{2} = \frac{808{,}7 + 808{,}83}{2} = 808{,}76$$

Three predictions will be made, including 12 forecasted values each with both the ARIMA model and Holt-Winters. The three last quarters in the data set will be used for forecasting and evaluation. First, 1179 data points with start 16th of September 1996 will be used, and the 12 upcoming Tuesdays will be forecasted. Second, 1191 data points will be used 16th of September 1996 will be used, and the 12 upcoming Tuesdays will be forecasted. And last, 1203 data points will be used. The datapoints not used in each prediction are removed in the excel file containing the data points.

3.2 Method

Time-series analysis can be used as a base for generating scenarios, and Guo et al. (2016) suggest that ARIMA is an appropriate model to use when the price being predicted has the characteristics similar to the energy prices. To investigate if time-series analysis ARIMA is applied to the European pulp market price. In the ARIMA model, historical data are used as a basis to predict future data (Guo et al., 2016). Data regarding market prices are provided by FOEX, which has been functional every Tuesday since 16 September 1996.

Holt-Winters model for predicting values are used for comparison to the ARIMA model and to investigate if other time-series analysis tools could create similar results to the ARIMA model and to increase the reliability within the validation of the results. Holt-Winters model is recommended for these kind of predictions; however, often the data set has a seasonal factor (Shaleh et al., 2018).

First, the data points were plotted to visualize trends. This to conclude that the time-series is non-stationary and has a moving average. This is included in Step 2, Initial Exploration.

The Software RStudio were used for Step 3, Analysis. First, there is code for getting libraries forecast, readxl and tseries including functions needed in Rstudio. A summery of the data set is presented in Rstudio. Then, the function *AUTO.ARIMA* is used. This function identifies the right combination the parameters *AR(p), MA(q),* and the integration parameter *d* in RStudio. The ARIMA model results were used to predict the values of the future data points using the function *FORECAST* in RStudio. Then, the forecast-method *HOLTWINTERS* in Rstudio were used together with the *FORECAST* function to predict future values. In the models, the number of data points to predict were chosen as 12, since the aim of this thesis is to forecast 3 months weekly prices. Three analyses for each model is done in Rstudio, predicting three different periods of year 2020. The results were plotted in RStudio, to be used in the presentation stage and all values predicted and used in the model are summarized.

3.3 Evaluation

MAE and MAPE were used to evaluate the results of the ARIMA and Holt-Winters models created forecast, as cross-validation. The predicted values are the values created in the ARIMA and Holt-Winters model from Rstudio, and the actual values were collected from FOEX Fastmarkets, and then calculated to validate the results in Excel. The evaluation will gain knowledge about the artefact time-series analysis model applied to pulp market prices, more specific the ARIMA model and Holt-Winters.

3.4 Research Ethics

Ethical approval is needed when using questionnaires, interviews, focus groups and/or observations. However, codes of research ethics are guidelines to use in all research. There are some key principles within the code of research ethics (Denscombe, 2014):

i) All participants' interests should be protected and no consequences of involvement in the research should occur. This includes both physical harm and to avoid psychological harm.

ii) Participation is voluntary and based on informed consent. No one should be forced into the research.

iii) The researcher should avoid deception when dealing with participants, and they should be open about their purpose with the study.

iv) The study needs to comply with the laws of the country where the research is located.

3.5 Samhällsetiska aspecter

There also exists some principles to be followed throughout the research for it to be seen as ethical. The research should not cause risk and harm for individuals and society, but maximize its benefits. All individuals and groups rights should be respected. All research should be conducted with integrity. The responsibilities should be defined, and "independence of research should be maintained and where conflicts of interest cannot be avoided they should be made explicit." (Our core principles - Economic and Social Research Council, 2020)

No of the principles mentioned above are being violated in this study. No personal information are revealed and all data are collected from an independent authority.

3.6 Reliability and validity

For high credibility of quantitative data and the methods in use, some criteria should be followed. The validity of the data means the accuracy, the precision of the data points used, and also how appropriate the data are for the research question. The validity also depends on the fact that no errors occur in the data set. Second, to obtain high credibility, the methods of the study must be reliable. One way to ensure reliability is to retest the research instrument and compare results. One external validity factor is the generalizability of the results and findings, i.e., to apply the findings to another case or example. If the study is based on a large data set, then there is a good ground for generalization of the results from the study. Then, the objectivity of the researcher is a factor to have in mind. The researcher should treat the data fairly and be impartial. She should considerate other theories and alternative views. The advantages of a study with quantitative data analysis is that it often appears to be grounded on statistical techniques that are based on mathematical principles. Additional credibility can be added if statistical tests of significance are used. The measurements are often precise and based on measured quantities rather than impressions. A big amount of data can be analyzed quickly compared to qualitative data. When representing the data with help of tables and charts, the communication to others is summarized and effective to use in these terms. However, there are some disadvantages as well, there might be a lack of quality within the data, and the results reliability will then be decreased. There is a risk that the techniques will be more in center than the aim of the research. If too many factors are considered and the complexity of the research increases, then it can be challenging to handle for the researcher. The researcher can in the reality influence the findings of the research from their choices (Denscombe, 2014). Data used in this thesis are collected from an organization that SCA and other pulp producing companies uses when pricing. To evaluate the results of the study, two measurements are used, MAPE and MAE.

4 Results

When plotting the time series year by year, for the years having 52 data points, no seasonal affection can be identified (Figure 5). The data does not follow any particular pattern that also can be seen in figure 2. However, the prices has overall increased since year 1996.

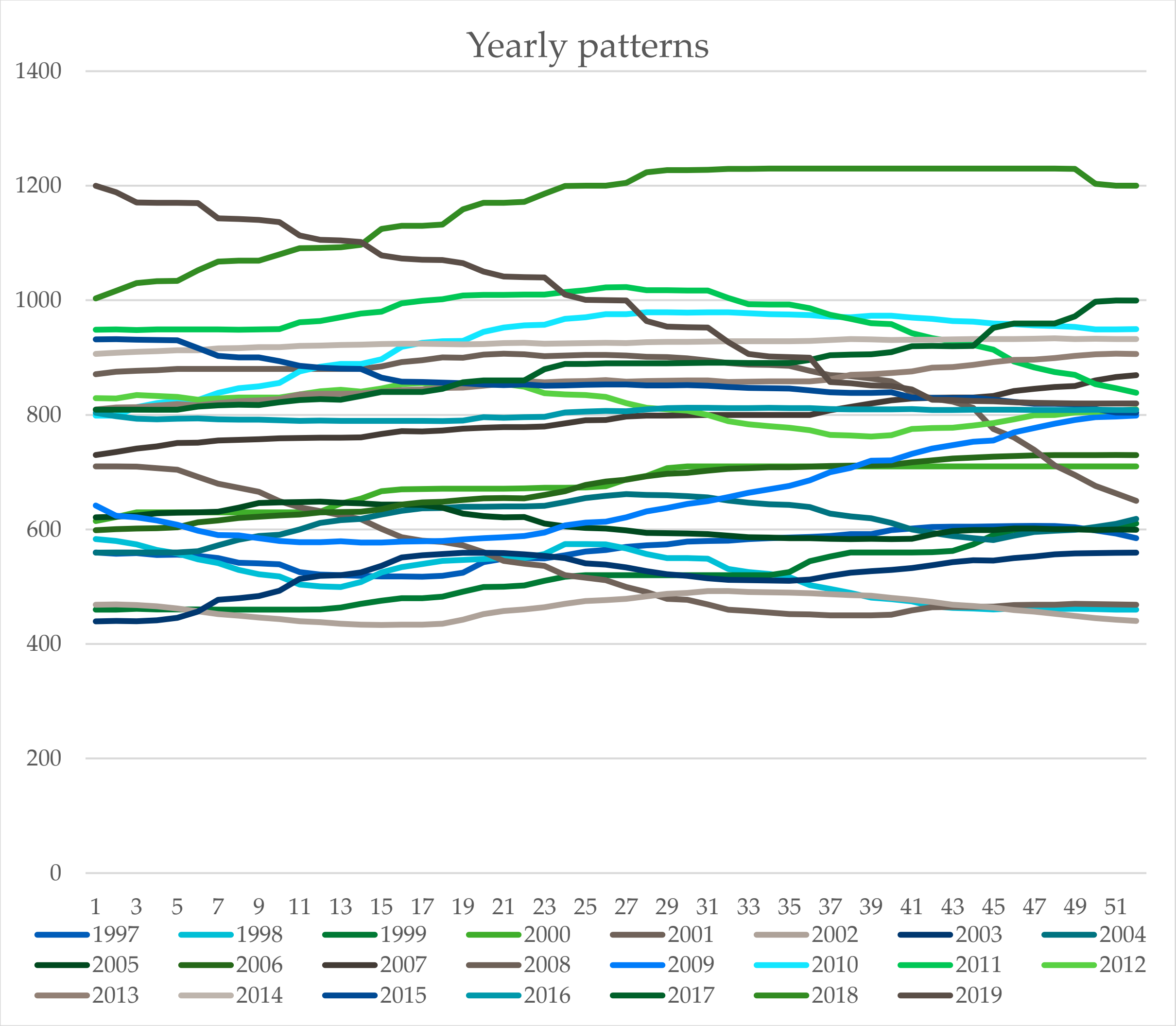

Figure 5 Data points year by year

Summaries of the data sets used are presented, see figure 6-8, from Rstudio. The maximum value, in other words the highest pulp price for the period, is 1230 USD per ton and the minimum price is 433,2 USD per ton for all data sets. The median and mean values are varies between the data sets.

```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
433.2   584.6   766.9   746.7   885.0  1230.0
```

Figure 6 Summary of the first data set

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  433.2   585.0   775.3   747.6   883.0  1230.0
```

Figure 7 Summary of the second data set

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  433.2   586.0   777.4   748.3   881.5  1230.0
```

Figure 8 Summary of the third data set

The results regarding what parameters that are used achieved by AUTO.ARIMA and FORECAST is presented in Figures 9-11. ARIMA(5,1,5) are the result for the first data set, ARIMA(4,1,3) are the result for data set two and three.

```
Model Information:
Series: tsdata
ARIMA(5,1,5)

Coefficients:
          ar1     ar2      ar3     ar4     ar5     ma1     ma2     ma3     ma4      ma5
      -0.7668  0.1395  -0.1925  0.5552  0.8396  1.1932  0.3940  0.4569  0.0610  -0.3028
s.e.   0.0300  0.0358   0.0324  0.0344  0.0305  0.0437  0.0669  0.0626  0.0601   0.0462
```

Figure 9 ARIMA model information for the first data set

```
Forecast method: ARIMA(4,1,3)

Model Information:
Series: tsdata
ARIMA(4,1,3)

Coefficients:
         ar1      ar2     ar3     ar4      ma1     ma2      ma3
      0.8133  -0.7116  0.4562  0.2891  -0.4657  0.6403  -0.2267
s.e.  0.0855   0.1085  0.1058  0.0648   0.0866  0.0762   0.0760
```

Figure 10 ARMA model information for the second data set

```
Model Information:
Series: tsdata
ARIMA(4,1,3)

Coefficients:
         ar1      ar2     ar3     ar4      ma1     ma2      ma3
      0.8003  -0.6919  0.4383  0.2983  -0.4523  0.6236  -0.2143
s.e.  0.0833   0.1051  0.1022  0.0627   0.0847  0.0736   0.0735
```

Figure 11 ARIMA model information for the third data set

The 12 predicted values in USD simulated by the function FORECAST on each dataset and each model can be found in Figures 12 - 17. This presents predicted values for 12 upcoming Tuesdays.

Forecast
881.5500
879.4287
878.6079
881.2761
866.2146
861.6665
860.3021
864.4051
855.8228
848.0675
847.4506
849.6259

Figure 12 results first data set of ARIMA forecast

Forecast
902.3229
904.8579
907.5251
910.3934
912.8747
915.4352
918.4563
921.3402
924.7058
923.8710
923.6969
923.8508

Figure 13 results first data set of Holt-Winters forecast

Forecast
819.1597
815.2315
813.2089
813.6104
813.2690
810.6475
808.3569
808.3197
808.6249
807.0968
804.9578
804.4340

Figure 14 results second data set of ARIMA forecast

Forecast
821.2163
817.8319
815.6448
812.3731
808.1837
803.5513
800.0585
797.0976
791.5417
785.2139
775.4943
770.5084

Figure 15 results second data set of Holt-Winters forecast

Forecast
823.9979
824.4218
825.1414
826.3014
827.1675
827.5000
827.8900
828.6977
829.4783
829.8144
830.0136
830.5236

Figure 16 Results third data set of ARIMA forecast

Forecast
814.8591
807.9395
800.0599
794.4798
789.0635
786.4353
788.3267
796.9042
806.8548
816.4969
827.4910
833.5167

Figure 17 results third data set of Holt-Winters forecast

The results were plotted in Rstudio, see Figures 18 - 23. Where the black line represent the data points from the data set, and the blue line represent the forecasted values.

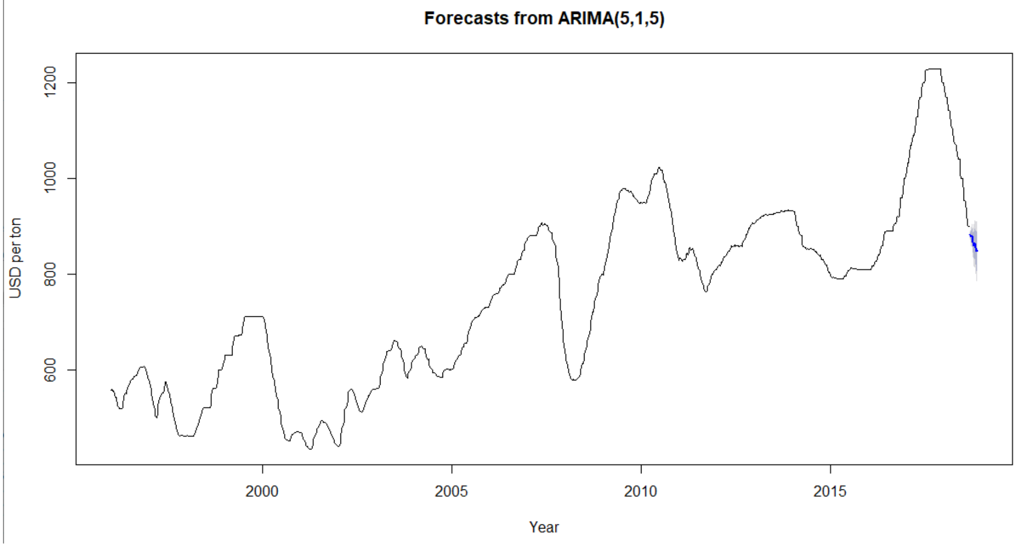

Figure 18 ARIMA model plot with the first data set

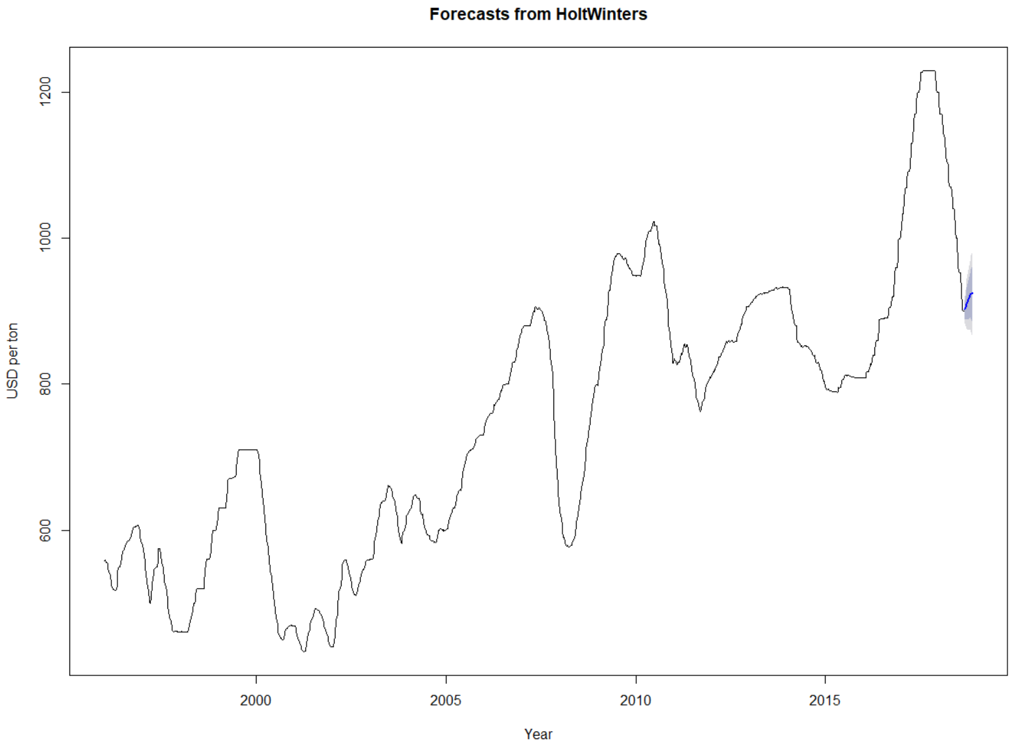

Figure 19 Holt-Winter model plot with the first data set

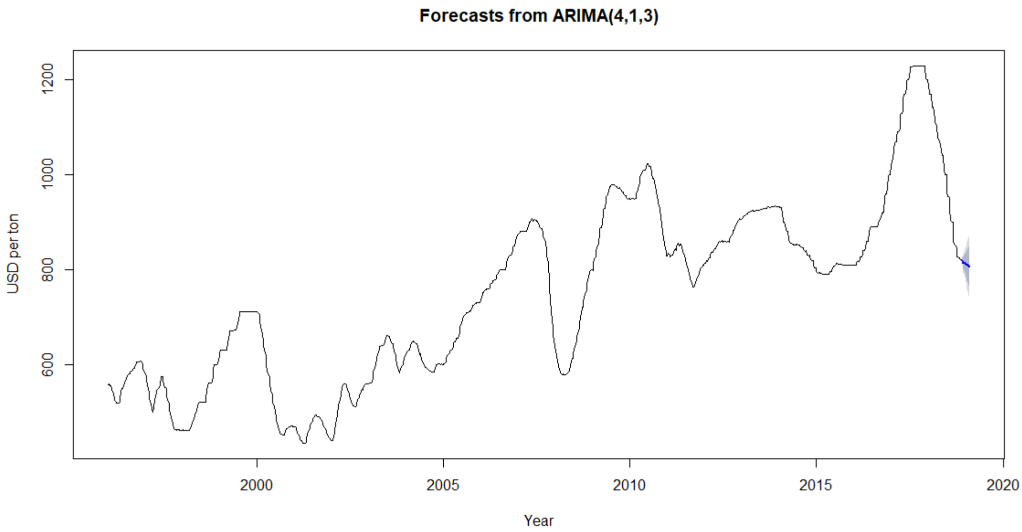

Figure 20 ARIMA model plot with the second data set

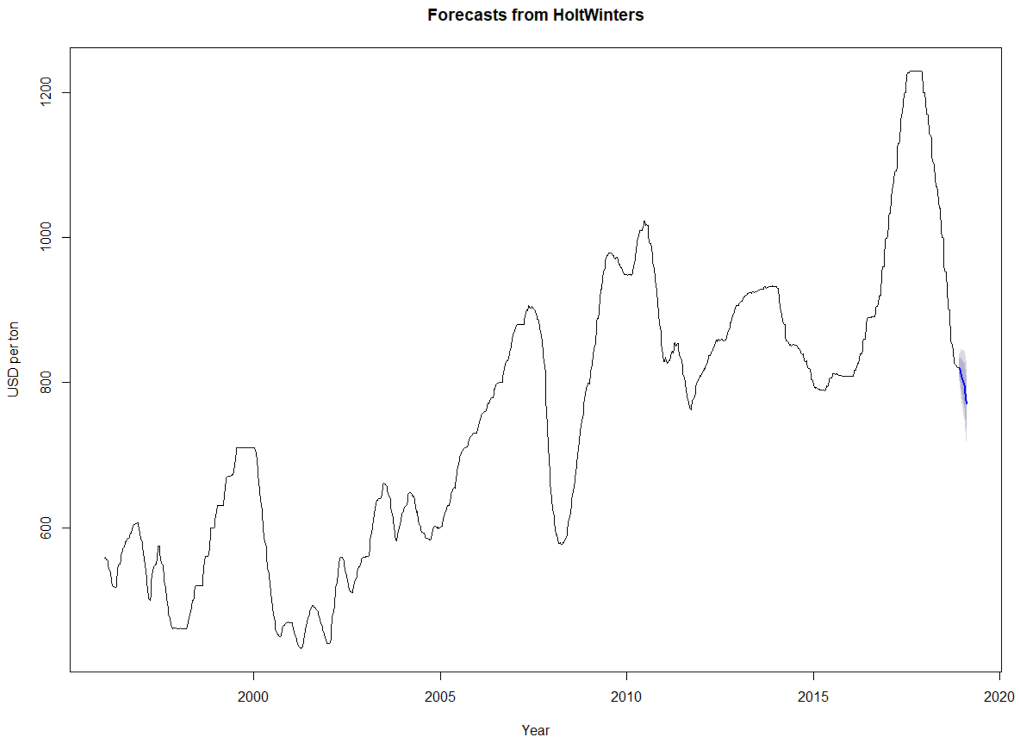

Figure 21 Holt-Winter model plot with the second data set

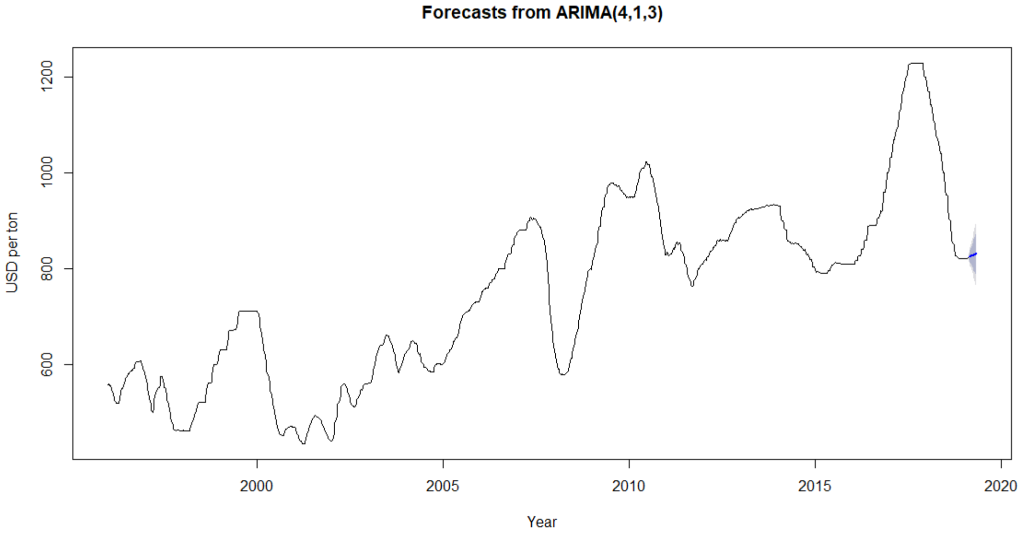

Figure 22 ARIMA model plot with the third data set

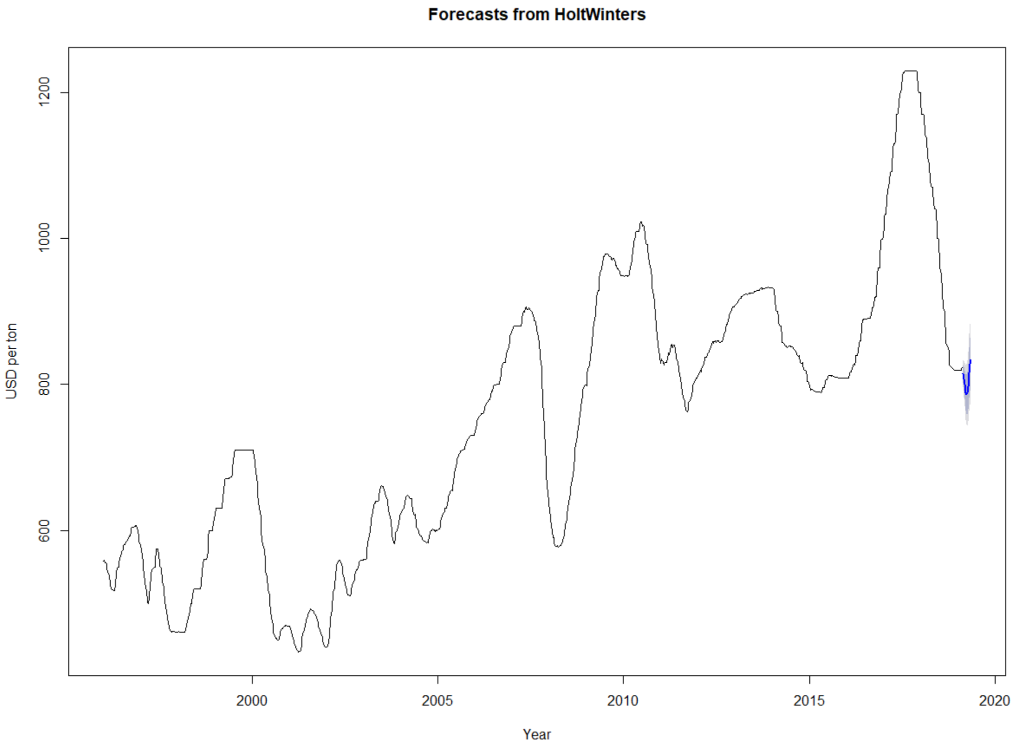

Figure 23 Holt-Winter model plot with the third data set

The results from the evaluation, and comparison to the actual values, can be found in table 2 – 4, for each data set used in the three predications made with the ARIMA model and with the Holt-Winters model in Rstudio. In addition, the validation measurements, MAE and MAPE, can be found in table 2 – 6.

Table 2 Results from the ARIMA model with the first data set

Date	PIX	ARIMA	Absolute value of Actual - Predicted	Absolute value of Actual - Predicted divided by actual				
First data set					MAE	29,34917	MAPE	3,52%
10,9,2019	857,4	881,55	-24,15	-2,82%				
17,9,2019	855,34	879,43	-24,09	-2,82%				
24,9,2019	851,2	878,61	-27,41	-3,22%				
1,10,2019	850,26	881,28	-31,02	-3,65%				
8,10,2019	844,44	866,21	-21,77	-2,58%				
15,10,2019	826,53	861,67	-35,14	-4,25%				
22,10,2019	825,76	860,3	-34,54	-4,18%				
29,10,2019	824,36	864,41	-40,05	-4,86%				
5,11,2019	823,73	855,82	-32,09	-3,90%				
12,11,2019	821,85	848,07	-26,22	-3,19%				
19,11,2019	821,12	847,45	-26,33	-3,21%				
26,11,2019	820,25	849,63	-29,38	-3,58%				

Table 3 Results from the Holt-Winter model with the first data set

Date	PIX	Holt-Winter	Absolute value of Actual - Predicted	Absolute value of Actual - Predicted divided by actual				
First data set					MAE	80,59083	MAPE	9,70%
10,9,2019	857,4	902,32	-44,92	-5,24%				
17,9,2019	855,34	904,85	-49,51	-5,79%				
24,9,2019	851,2	907,53	-56,33	-6,62%				
1,10,2019	850,26	910,39	-60,13	-7,07%				
8,10,2019	844,44	912,87	-68,43	-8,10%				
15,10,2019	826,53	915,44	-88,91	-10,76%				
22,10,2019	825,76	918,46	-92,7	-11,23%				
29,10,2019	824,36	921,34	-96,98	-11,76%				
5,11,2019	823,73	924,71	-100,98	-12,26%				
12,11,2019	821,85	923,87	-102,02	-12,41%				
19,11,2019	821,12	923,7	-102,58	-12,49%				
26,11,2019	820,25	923,85	-103,6	-12,63%				

Table 4 Results from the ARIMA model with the second data set

Date	PIX	ARIMA	Absolute value of Actual - Predicted	Absolute value of Actual - Predicted divided by actual				
Second data set					MAE	9,7825	MAPE	1,19%
3,12,2019	820,13	819,16	0,97	0,12%				
10,12,2019	819,84	815,23	4,61	0,56%				
17,12,2019	819,87	813,21	6,66	0,81%				
27,12,2019	819,95	813,61	6,34	0,77%				
31,12,2019	819,95	813,27	6,68	0,81%				
7,1,2020	820	810,65	9,35	1,14%				
14,1,2020	819,96	808,36	11,6	1,41%				
21,1,2020	820	808,32	11,68	1,42%				
28,1,2020	820	808,62	11,38	1,39%				
4,2,2020	820	807,1	12,9	1,57%				
11,2,2020	821,45	804,96	16,49	2,01%				
18,2,2020	823,16	804,43	18,73	2,28%				

Table 5 Results from the Holt-Winter model with the second data set

Date	PIX	Holt-Winter	Absolute value of Actual - Predicted	Absolute value of Actual - Predicted divided by actual				
Second data set					MAE	20,46667	MAPE	2,49%
3,12,2019	820,13	821,22	-1,09	-0,13%				
10,12,2019	819,84	817,83	2,01	0,25%				
17,12,2019	819,87	815,65	4,22	0,51%				
27,12,2019	819,95	812,37	7,58	0,92%				
31,12,2019	819,95	808,18	11,77	1,44%				
7,1,2020	820	803,55	16,45	2,01%				
14,1,2020	819,96	800,06	19,9	2,43%				
21,1,2020	820	797,1	22,9	2,79%				
28,1,2020	820	791,54	28,46	3,47%				
4,2,2020	820	785,21	34,79	4,24%				
11,2,2020	821,45	775,49	45,96	5,59%				
18,2,2020	823,16	770,51	52,65	6,40%				

Table 6 Results from the ARIMA model with the third data set

Date	PIX	ARIMA	Absolute value of Actual - Predicted	Absolute value of Actual - Predicted divided by actual				
Third data set					MAE	10,64	MAPE	1,27%
25,2,2020	824,73	824	0,73	0,09%				
3,3,2020	824,42	824,42	0	0,00%				
10,3,2020	837,61	825,14	12,47	1,49%				
17,3,2020	839,36	826,3	13,06	1,56%				
24,3,2020	839,73	824,17	15,56	1,85%				
31,3,2020	839,73	827,5	12,23	1,46%				
7,4,2020	839,91	827,89	12,02	1,43%				
14,4,2020	840	828,7	11,3	1,35%				
21,4,2020	840	829,48	10,52	1,25%				
28,4,2020	840	829,81	10,19	1,21%				
5,5,2020	840,1	830,01	10,09	1,20%				
12,5,2020	850,03	830,52	19,51	2,30%				

Table 7 Results from the Holt-Winter model with the thirds data set

Date	PIX	Holt-Winter	Absolute value of Actual - Predicted	Absolute value of Actual - Predicted divided by actual				
Third data set					MAE	32,76667	MAPE	3,91%
25,2,2020	824,73	814,86	9,87	1,20%				
3,3,2020	824,42	807,94	16,48	2,00%				
10,3,2020	837,61	800,06	37,55	4,48%				
17,3,2020	839,36	794,48	44,88	5,35%				
24,3,2020	839,73	789,06	50,67	6,03%				
31,3,2020	839,73	786,43	53,3	6,35%				
7,4,2020	839,91	788,33	51,58	6,14%				
14,4,2020	840	796,9	43,1	5,13%				
21,4,2020	840	806,85	33,15	3,95%				
28,4,2020	840	816,5	23,5	2,80%				
5,5,2020	840,1	827,49	12,61	1,50%				
12,5,2020	850,03	833,52	16,51	1,94%				

5 Discussion

The data collection used in this thesis is both peer-reviewed articles relevant for the problem area, and a data set. The data were collected from a third party that can be seen as increasing for the reliability of the study. Big pulp manufacturers, such as SCA pulp mill, use the data from FOEX Fastmarkets weekly.

The last two predictions made by AUTO.ARIMA and FORECAST created the parameter AR(4), I(1) and MA(3). The first prediction created the parameters AR(5), I(1) and MA(5). The AR(p)-parameters shows how the specific values in the time-series are related to the values in previous periods, the lag observation. The integrated factor I shows that one differencing is needed to make the time series stationary. The MA(q) parameters tells the order of the moving average according to previous periods.

The evaluation done with MAE and MAPE regarding the ARIMA forecasts varying, see Table 8. The first prediction has the highest values of MAE and MAPE. The PIX price changed from 857,4 USD to 820,25 USD over the period, a decrease with 37,15 USD, about 4,3 %, over 12 weeks. The second data set has the lowest MAE and MAPE values. The PIX increased over the period with a total of 3 USD, about 0,37%. However, in this case, PIX decreased and the ARIMA model increased. In the third data set, PIX increased 25,3 USD, about 3%. The ARIMA model predicted increased values as well. This shows that if the price change is large-scaled, then the ARIMA model shows a less good result than if the price change is lower over the specific time period.

The evaluation done with MAE and MAPE regarding the Holt-Winters forecast follows the same pattern as the one regarding the ARIMA predictions. When the ARIMA model estimates that the PIX market price will decrease, so do the Holt-Winters model. In addition, the best results of MAE and MAPE are when the total price change over the period is lowest.

Table 8 MAE and MAPE results

	MAE ARIMA	MAPE ARIMA	MAE Holt-Winters	MAPE Holt-Winters
First data set	29,35	3,52%	80,59	9,7%
Second data set	9,78	1,19%	20,47	2,49%
Third data set	10,64	1,27%	32,77	3,91%

It can be seen in Table 8 that the ARIMA model shows more accurate results than the Holt-Winter model. The comparison can be used to increase the reliability and validity of the results of the ARIMA model. However, since the evaluation shows that the results could be improved, other parameters could be included for an improved forecast. For example, the exchange rate could influence the PIX price on pulp.

In two of three predictions, both the ARIMA model and the Holt-Winters model forecasts that the priced will increased when it actually did. This can tell the producers where the market is heading and help them give a heads up on the future needs even if thee separate prices predicted is not fully accurate.

A mean value for the evaluation MAE and MAPE can be seen in table K. There is a different between the evaluations and the models where MAE for ARIMA shows that the difference is 28,02 USD/tons. The results of MAPE and the different time series analysis shows a difference of 3,38 percentage. There a distinct difference between the results when evaluation them, where the ARIMA model performed a lot better than Holt-Winter.

Table 9 Mean values of the evaluation

	MAE ARIMA	MAPE ARIMA	MAE Holt-Winters	MAPE Holt-Winter
Mean values	16,59	1,99%	44,61	5,37%

The price changes in first and third data set where more distinct than in general, this can be seen as special happenings in the data set and the result could have been affected by this fact.

Because of the differences in the predicted and actual prices forecasted, the forecasted prices per week could be difficult to use for SCA pulp mill. However, the prices could be used for SCA to decide where to sell their products in the future to maximize profit.

6 Conclusion

The ARIMA model predict more accurate results than Holt-Winter for the given time periods used in this thesis. The time-series models applied on pulp market prices needs improvement for increased reliability on the forecasted values. However, the ARIMA model could be a tool for predicting where the market is headed, to use for pulp producers, such as SCA Östrand pulp mill. More tests within other time periods and different of numbers predicted in the model could help gain understanding within the research area. In addition, more comparisons should be done, with other tools than time-series analysis. An application of regression analysis and/or ANNs and NNs should be tested to see if it could create more accurate results than time-series analysis.

7 References

Abdelaziz, F., Amer, M., & El-Baz, H. (2014). An Epsilon Constraint Method for selecting Indicators for use in Neural Networks for Stock Market Forecasting. INFOR: Information Systems and Operational Research, 52(3), 116-125.

Agra, A., Poss, M., & Santos, M. (2018). Optimizing make-to-stock policies through a robust lot-sizing model. International Journal of Production Economics, 200, 302-310.

Ariyanti Yunila Dwi Putri, Mustafid, & Nurhayati Oky Dwi. (2018). Forecasting and Controlling the Food Supply System in Hospital Using Exponential Smoothing. E3S Web of Conferences, 73, 13014.

Arnold, T., Kane, Michael ,. & W Lewis, B. (2019). A computational approach to statistical learning (1st.. ed.).

Arum Handini Primandari. (2017). An Alternative Forecasting Using Holt-Winter Damped Trend for Soekarno-Hatta Airport Passenger Volume. Eksakta: Jurnal Ilmu-Ilmu MIPA, 17(1), 1-10.

Barello, M., Manca, D., Patel, R., & Mujtaba, I. (2014). Neural network based correlation for estimating water permeability constant in RO desalination process under fouling. Desalination, 345, 101.

Bianchi, L., Jarrett, J., & Choudary Hanumara, R. (1998). Improving forecasting for telemarketing centers by ARIMA modeling with intervention. International Journal of Forecasting, 14(4), 497-504.

Box, G., Jenkins, Gwilym M, & Reinsel, Gregory C. (1994). Time series analysis: Forecasting and control (3.rd ed.). Englewood Cliffs: Prentice Hall. González-García, S., Hospido, A., Moreira, M., Romero, J., & Feijoo,

Brownlee, J. (2017). A gentle introduction to the Box-Jenkins method for time-series forecasting. Machine Learning Mastery. https://machinelearningmastery.com/gentle-introduction-box-jenkins-method-time-series-forecasting/

B V, B., & Dakshayini, M. (2018). Performance Analysis of the Regression and Time Series Predictive Models using Parallel Implementation for Agricultural Data. Procedia Computer Science, 132, 198-207.

Chen, L., & Yu, L. (2013). Structural Nonlinear Damage Identification Algorithm Based on Time Series ARMA/GARCH Model. Advances in Structural Engineering, 16(9), 1597-1609.

Chujai, P., Kerdprasop, N., Kerdaprasop, K. (2013). Time-series analysis of household electric consumption with ARIMA and ARMA models. IMECS.

Conejo, A.J., Baringo Morales, L., Kazempour, S.J. & Siddiqui, A.S. (2016). *Investment in Electricity Generation and Transmission Decision Making under Uncertainty*. Cham: Springer International Publishing.

Dellana, S., & West, D. (2009). Predictive modeling for wastewater applications: Linear and nonlinear approaches. Environmental Modelling and Software, 24(1), 96-106.

Denscombe, M. (2014). The Good Research Guide: For Small-Scale Social Research Projects [Elektronisk resurs]. McGraw-Hill.

Fastmarkets FOEX. (2019). *FOEX Fast markets*. Viewed 2020-03-12 from https://www.foex.fi/.

González, S.G., Hospido, A., Moreira, M.T., Romero, J., Feijoo, G. (2009). Environmental impact assessment of total chlorine free pulp from Eucalyptus globulus in Spain. Journal of Cleaner Production, 17(11), 1010-1016.

Guo, X., Beskos, A., & Siddiqui, A. (2016). The natural hedge of a gas-fired power plant. Computational Management Science, 13(1), 63-86.

Hadj Youssef, K., Van Delft, C., & Dallery, Y. (2018). Priority optimization and make-to-stock/make-to-order decision in multiproduct manufacturing systems. International Transactions in Operational Research, 25(4), 1199-1219.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning (Springer Series in Statistics). New York, NY: Springer New York.

Izady, A., Davary, K., Alizadeh, A., Moghaddam Nia, A., Ziaei, A. N., & Hasheminia, S. M. (2013). Application of NN-ARX Model to Predict Groundwater Levels in the Neishaboor Plain, Iran. Water Resources Management, 27(14), 4773.

Johannesson, P. & Perjons, E. (2014). An introduction to design science. Cham: Springer International Publishing.

J. Miah, S., Kerr, D., & Von Hellens, L. (2014). A collective artefact design of decision support systems: Design science research perspective. Information Technology & People, 27(3), 259-279.

Jonsson, A., & Sörensson., R. (2014). Företag inom svensk massa- och pappersindustri. Vinnova – Swedish Governmental Agency for Innovation Systems. VA 2014:08.

Kaeck, A., Rodrigues, P., & Seeger, N. (2018). Model Complexity and Out-of-Sample Performance: Evidence from S&P 500 Index Returns. Journal of Economic Dynamics and Control, 90, 1-29.

Katchova, A. (2013). Time Series ARIMA Models. Econometrics Academy. Viewed 2020-03-29 from https://sites.google.com/site/econometricsacademy/econometrics-models/time-series-arima-models.

Katz, M.H. (2010). Evaluating Clinical and Public Health Interventions: A Practical Guide to Study Design and Statistics. Cambridge University Press.

Lübkert, T. et al., 2018. Understanding price functions to control domestic electric water heaters for demand response. Computer Science - Research and Development, 33(1), pp.81–92.

Malcolm, G.A.. (1999). Time Series Methods for Commodity Price Forecasting: An Application to Market Pulp. CINTRAFOR Working Paper, 76.

Mandal, P., Srivastava, A., Senjyu, T., & Negnevitsky, M. (2010). A new recursive neural network algorithm to forecast electricity price for PJM day-ahead market. International Journal of Energy Research, 34(6), 507-522.

Nason, G. P. (2006). Stationary and non-stationary time series. Statistics in Volcanology. Special Publications of IAVCEI, 1, 000-000.

Nikolopoulos, K., Goodwin, P., Patelis, A., & Assimakopoulos, V. (2007). Forecasting with cue information: A comparison of multiple regression with alternative forecasting approaches. European Journal of Operational Research, 180(1), 354-368.

Nogales, F., Contreras, J., Conejo, A., & Espinola, R. (2002). Forecasting next-day electricity prices by time-series models. IEEE Transactions on Power Systems, 17(2), 342-348.

Our core principles - Economic and Social Research Council. 2020. Viewed 2020-06-07 from https://esrc.ukri.org/funding/guidance-for-applicants/research-ethics/our-core-principles/.

Pinto Moreira de Souza, D.; Da Silva Christo, E.; Rocha Almeida, A. (2017). Location of Faults in Power Transmission Lines Using the ARIMA Method. Energies 2017, 10, 1596.

Ramos, P., Santos, N., & Rebelo, R. (2015). Performance of state space and ARIMA models for consumer retail sales forecasting. Robotics and Computer Integrated Manufacturing, 34(C), 151-163.

Razali, S., Rusiman, M., Zawawi, N., & Arbin, N. (2018). Forecasting of Water Consumptions Expenditure Using Holt-Winter's and ARIMA. Journal of Physics: Conference Series, 995(1), 1-8.

R-Project. *Forecast.* Viewed 2020-03-12 from https://www.r-project.org/nosvn/pandoc/forecast.html .

Sarkodie, S. (2017). Estimating Ghana's electricity consumption by 2030: An ARIMA forecast. Energy Sources, Part B: Economics, Planning, and Policy, 12(10), 936-944.

Shaleh, Rasim, & Wahyudin. (2018). The System of Inventory Forecasting in PT. XYZ by using the Method of Holt Winter Multiplicative. IOP Conference Series: Materials Science and Engineering, 288(1), 1-9.

Shanmuganathan, Subana ; Samarasinghe, Sandhya. (2016). Artificial Neural Network Modelling (Vol. 628, Studies in Computational Intelligence). Cham: Springer International Publishing.

Skogsindustrierna. *Massa – production och handel.* Viewed 2020-04-17. https://www.skogsindustrierna.se/om-skogsindustrin/branschstatistik/massa-produktion-och-handel/.

Son, H., & Kim, C. (2015). Forecasting Short-term Electricity Demand in Residential Sector Based on Support Vector Regression and Fuzzy-rough Feature Selection with Particle Swarm Optimization. Procedia Engineering, 118, 1162-1168.

Son, H., & Kim, C. (2017). Short-term forecasting of electricity demand for the residential sector using weather and social variables. Resources, Conservation & Recycling, 123, 200-207.

Svenska Cellulosa Aktiebolaget SCA. *Massa.* Viewed 2020-03-12 from https://www.sca.com/sv/massa/.

Syafei, D.A., Ramadhan, N., Hermana, J., Slamet, A., Boedisantoso, R., & Assomadi, F.A. (2018). Application of Exponential Smoothing Holt Winter and ARIMA Models for Predicting Air Pollutant Concentrations. Department of Environmental Engineering, Faculty of Civil, Environmental, and Geo Engineering, Institut Teknologi Sepuluh Nopember (ITS), Indonesia.

Tashman, L. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. International Journal of Forecasting, 16(4), 437-450.

Toppinen, A., Laaksonen, S., Hänninen, R. (1996). Forecasting International Pulp Prices With Producers Inventories. Forestry Sciences. 58, 151-158.

Vadovsky, M., Michalik, P., Zolotova, I., & Paralic, J. (2016). Better IT services by means of data mining. 2016 IEEE 14th International Symposium on Applied Machine Intelligence and Informatics (SAMI), 187-192.

Wan Ahmad, W., & Ahmad. (2013). Arima model and exponential smoothing method: A comparison. AIP Conference Proceedings, 1522(1), 1312-1321.

Wickham, H. (2014). Tidy data. Journal of Statistical Software, 59(10), 1-23.

Wu, C.-C., Chou, C.-Y. & Huang, C., 2009. Optimal price, warranty length and production rate for free replacement policy in the static demand market. Omega, 37(1), pp.29–39.

Appendix A – Foex Fastmarkets European Pix Price

Week/Year	1997	1998	1999	2000	2001	2002	2003		2004	2
1	559,83	583,19	460	614,63	710	468,71	439,5		559,46	6
2	557,47	579,81	460	621,91	710	468,89	440		559,91	6
3	558,7	573,96	461,63	630	709,77	468,19	439,78		560	6
4	555,41	564,29	460,21	630	706,88	465,7	441,47		560	6
5	555,86	558,11	460,23	630	704,31	461,95	445,69		560	6
6	553,9	548	460,24	630	692	457,02	456,64		561,81	6
7	549,79	541,25	460	630	679,81	452,3	477,04		572,37	6
8	541,81	529,24	460	630	672,55	449,7	479,81		582,23	6
9	540,85	521,81	460	630	665,42	446,45	483,53		588,5	6
10	539,34	517,86	460	630	649,23	443,56	492,89		590,79	6
11	525,78	503,67	460	630	638,36	439,78	513,42		600,15	6
12	521,05	500,38	460,22	630,83	632,16	438,2	518,75		611,48	6
13	520	499,63	463,62	645,1	624,9	435,11	520		616,14	6
14	518,85	508,14	470,32	653,77	618,04	433,71	524,8		618,61	6
15	518,16	524,74	475,78	666,78	600,98	433,2	536,32		626,08	6
16	517,71	534,38	479,88	670	586,53	433,72	551,12		632,55	6
17	517,42	539,65	480	670,5	580,42	433,9	554,92		637,04	6
18	519,18	544,88	482,86	670,88	577,98	435,19	556,79		637,2	6
19	524,55	546,73	491,49	670,96	572,04	442,41	559,04		639,53	6
20	542,66	548,43	499,2	671,27	560,19	452,12	559,19		639,68	6
21	548,75	550	500	671,09	545,22	457,48	558,82		640	6
22	550,09	550	502,39	671,7	540,19	460,3	556,62		640	6
23	549,56	556,95	510,48	672,6	536,63	464,07	553,76		641,09	6

24	554,9	574,44	517,2	672,88	519,9	470,08	549,71		647,64	6
25	561,14	574,48	520	673,1	515,5	475,12	540,75		655	6
26	563,86	574	520	675,55	511,5	476,79	538,45		658,69	6
27	569,7	566,64	520	687,5	499,55	479,02	533,86		661,68	5
28	572,3	557,21	520	693,15	490,85	483,18	527,37		660,15	5
29	573,88	550,23	520	706,72	478,91	487,32	521,72		659,71	5
30	578,62	550	520	710	477,27	489,07	518,83		658,3	5
31	579,97	548,7	520	710	469,23	492,36	514,47		656,03	5
32	580,57	531	520	710	459,82	492,25	511,79		650,68	5
33	583,19	525,74	520	710	457,58	490,96	511,4		646,63	5
34	584,91	522,5	520	710	455,1	490	510,79		643,96	5
35	585,62	517	525,64	710	452,45	489,46	510,13		642,61	5
36	586,67	502,74	544,52	710	451,67	488,59	512,41		638,96	5
37	588,29	496,27	552,59	710	450	486,78	518,94		627,7	5
38	591,98	489,3	560	710	450	485,57	524,41		622,66	5
39	592,04	481,09	560	710	450	484,17	527,02		619,4	5
40	598,66	478,41	560	710	451,13	480,68	529,32		611,2	5
41	601,66	474,39	560	710	458,89	477,03	532,82		600,08	5
42	604,17	465,7	560,42	710	463,94	473,6	537,7		593,8	5
43	604,63	462,98	562,45	710	464,96	468,49	542,65		588,43	5
44	604,59	462,02	573,86	710	464,48	466,11	546,1		584,49	5
45	605,14	460,35	590,18	710	465,43	463,84	545,69		581,7	5
46	605,69	461,89	599,66	710	467,72	459,13	549,82		589,18	6
47	606,18	461,36	600	710	468,64	456,35	552,73		595,77	6
48	606,06	461,39	600	710	468,67	452,73	556,62		598,03	6
49	604	461,54	600	710	469,92	448,86	557,93		599,63	6
50	598,55	460,64	600	710	469,74	445	558,86		604,49	5
51	592,92	460	600,82	710	469,12	442,57	559,32		609,79	6
52	584,72	460	610,36	710	468,46	440,44	559,48		618,15	5

53				710						5
Week/Year	2006	2007	2008	2009	2010	2011	2012			
1	598,61	729,95	871,16	641,51	798,77	948,92	829,04			
2	600,25	735	875,14	623,97	798,63	949,21	828,83			
3	601,32	740,93	876,73	621,3	813,24	948,2	834,55			
4	602,15	745,07	877,88	615,47	820,66	949,1	832,97			
5	603,98	751,06	880	608,11	824,93	949,37	831,06			
6	612,6	751,45	880	597,84	826,03	949,39	826,46			
7	615,78	755,38	880	590,39	838,6	949,08	828,4			
8	619,81	756,6	880	589,48	846,43	948,53	830,42			
9	622,06	757,68	880	584,48	849,69	949,08	830,1			
10	624,18	758,93	880	579,58	855,82	949,93	830,27			
11	625,92	759,53	880	577,65	875,62	961,81	835,68			
12	629,93	759,97	880	577,89	883,8	963,61	841,01			
13	630,21	760	880	579,18	888,77	970,15	843,85			
14	630,99	760,72	880	577,09	889,08	977,16	840,74			
15	635,21	766,91	884,48	577,34	897,12	980,24	845,73			
16	643,39	771,52	891,89	578,52	918,66	995,04	853,4			
17	647,01	771,18	895,63	579,54	925,76	998,9	855,65			
18	648,23	772,68	900,4	579,83	928,71	1001,76	850,5			
19	651,71	776,01	899,82	582,3	929,21	1008,5	852,14			
20	654,49	777,31	904,93	584,54	944,64	1009,53	853,91			
21	654,82	778,79	906,84	586,36	952,61	1009,42	854,39			
22	654,2	778,65	905,83	588,71	956,43	1009,86	849,39			
23	659,96	779,63	902,32	594,7	957,32	1010	837,84			
24	666,49	785,4	903,36	607,05	967,85	1014,55	835,91			
25	677,38	790,76	904,72	611,87	970,63	1017,82	834,36			

26	683,45	791,25	904,64	613,52	975,7	1022,43	831,2	
27	686,64	796,85	903,7	621,36	975,82	1023,1	820,36	
28	693,04	798,7	901,18	631,25	978,82	1017,61	812,14	
29	697,06	799	900,95	637,64	978,91	1017,87	809,49	
30	698,95	799,26	898,78	644,21	978,75	1017,11	807,31	
31	702,46	799,81	894,86	649,52	979,02	1016,83	799,95	
32	705,63	800	890,4	656,63	978,89	1003,86	789,2	
33	706,95	800	887,87	663,83	977,49	993,16	783,28	
34	708,71	800	887,17	669,83	975,63	992,47	780,33	
35	708,37	800	885,65	676,06	975,37	992,38	777,36	
36	709,69	800	876,72	686	974,19	986,06	773,18	
37	710,66	807,79	869,3	700,14	971,33	974,9	765,11	
38	711,18	814,21	867,46	707,45	970,55	967,75	763,79	
39	711,88	820,07	863,13	720,08	973,12	960,23	762,18	
40	713,15	825,59	858,51	720,56	972,91	958,31	764,37	
41	717,05	828,66	838,35	731,69	969,83	942,91	775,29	
42	720,28	829,86	828,74	741,3	967,7	934,16	776,96	
43	723,93	830	823,27	747	963,81	927,84	777,69	
44	725,23	830	812,29	753,26	963,01	922,03	781,62	
45	727,04	833,13	775,41	755,44	959,31	914,08	785,46	
46	728,03	841,43	760,74	769,44	958,37	893,49	792,32	
47	728,96	845,26	738,81	776,94	956,29	882,79	799,02	
48	729,88	848,78	711,72	784,84	955,13	874,92	800,01	
49	729,92	850,21	694,96	790,96	953,29	869,62	803,9	
50	729,91	860,28	676,18	796,06	949,27	853,33	805,41	
51	730	866,28	662,71	796,88	948,99	846,38	807,58	
52	729,96	869,3	649,92	798,79	949,49	838,63	809,6	
53						833,71		

Week/Year	2013	2014	2015	2016	2017	2018	2019	2020
1	809,37	906,48	932,06	803,36	808,94	1003,33	1199,96	850,03
2	812,76	908,36	932,15	797,54	808,99	1016,49	1188,7	840,10
3	813,41	909,96	931,14	793,08	808,92	1029,98	1171	840,00
4	816,34	911,08	930,61	791,98	808,87	1033,42	1170,24	840,00
5	819,01	912,87	930,4	793,34	809,05	1033,74	1170	840,00
6	816,68	912,55	916,42	793,58	814,51	1052,24	1169,83	839,91
7	820,22	915,99	903,02	792,24	816,67	1067,67	1142,95	839,73
8	823,95	916,55	900,34	791,46	817,69	1068,96	1141,83	839,73
9	825,12	918,02	900,09	791,43	817,41	1069,07	1140,5	839,36
10	827,68	918,08	893,72	790,52	822,35	1080,13	1136,7	837,61
11	835,47	920,48	885,42	789,48	825,67	1090,81	1113,07	824,42
12	836,85	921,17	882,27	789,78	827,41	1091,45	1105,85	824,73
13	837,02	921,9	880,73	789,28	826,26	1092,4	1104,45	823,16
14	837,49	922,59	880,03	789,47	832,84	1097,06	1101,86	821,45
15	840,75	923,62	865,12	789,67	839,88	1124,67	1078,55	820,00
16	843,43	924,05	857,97	789,3	839,9	1130,08	1072,79	820,00
17	844,11	924,25	857,34	789,35	840,09	1130,09	1070,74	820,00
18	846,89	923,85	856,11	789,2	845,31	1132,37	1070,05	819,96
19	847,61	923,13	855,25	789,95	856,73	1158,64	1065,03	820,00
20	851,14	923,93	852,8	795,85	859,92	1170	1050,11	
21	854,74	925,13	851,83	794,82	859,94	1170,17	1041,36	
22	858,88	925,66	852,83	795,95	860,18	1171,85	1040,68	
23	857,02	924,47	850,75	796,34	879,45	1185,93	1040,08	
24	858,26	925	851,33	803,97	888,97	1199,36	1009,82	
25	859,09	925,31	852,35	805,92	888,87	1199,99	1000,95	
26	860,59	925,57	853,24	806,96	889,82	1199,99	1000,31	
27	857,8	925,27	853,06	806,43	889,84	1205,04	999,5	

28	859,27	926,71	851,29	809,79	890,22	1223,56	964,11	
29	859,77	927,25	851,16	811,93	890,25	1227,5	953,95	
30	860,06	927,56	851,9	812,4	890,36	1227,5	952,97	
31	860,24	927,84	850,9	812,23	891,11	1227,69	952,27	
32	857,18	928,77	848,69	811,67	890,96	1229,56	927,1	
33	857,89	928,61	846,96	812,04	890,6	1229,55	906,16	
34	858,66	928,49	846,46	812,11	890,59	1229,86	901,71	
35	858,48	928,8	846,15	811,57	890,59	1229,86	900,59	
36	858,38	929,27	842,74	811,62	895,71	1229,86	899,95	
37	861,54	930,99	839,55	810,41	903,87	1229,93	857,4	
38	869,93	932,09	838,17	809,76	904,99	1229,95	855,34	
39	870,82	932,03	838,35	809,81	905,69	1229,95	851,2	
40	873,4	930,74	839,32	809,75	909,34	1229,96	850,26	
41	875,89	931,14	830,44	809,96	919,91	1230	844,44	
42	882,36	931,52	829,62	808,45	920,16	1230	826,53	
43	883,22	931,66	829,08	808,57	920,07	1230	825,76	
44	887,34	932,3	829,7	809,02	920,15	1230	824,36	
45	892,15	932,44	826,64	809,01	951,68	1230	823,73	
46	895,98	932,26	822,42	809,3	959,37	1230	821,85	
47	896,51	933,14	819,19	808,77	959,5	1230	821,12	
48	899,48	933,68	819,31	808,51	959,57	1230	820,25	
49	902,99	932,5	818,38	808,57	972	1229,32	820,13	
50	906	932,74	809,95	809,05	997,4	1203,33	819,84	
51	906,87	932,57	803,91	808,77	999,61	1200,21	819,87	
52	906,07	932,29	803,86	808,7	999,63	1200,02	819,95	
53				808,83				