After-Market Spare Parts Forecasting at Sandvik Stationary Crushing & Screening

2021/2022

LUND UNIVERSITY

MIOM05 - Degree Project In Production Management
Faculty of Engineering LTH

Division of Production Management

By:

Artur Jusopov & Arian Marofkhani

Supervisor:

Gudrun Kiesmüller

List of Abbreviations

AM - After-market
SKU - Stock Keeping Unit
WIP - Work In Progress
ADI - Average Inter-Demand Interval
CV - Coefficient of Variation for a time series with non-zero demand.
SES - Simple Exponential Smoothing Method
SBA - Syntetos-Boylan Approximation Method
HL - Holt’s Linear Method
HLDAMP - Holt’s Linear Damped Method
SNAIVE - Seasonal Naive Method
MASE - Mean Absolute Scaled Error
MAD - Mean Absolute Deviation
MAPE - Mean Absolute Percentage Error
CRPS - Continuous Ranked Probability Score

Acknowledgements

This master thesis concludes the mechanical engineering program at The Faculty of Engineering, Lund University. It has been a bumpy road with many ups and some downs. For the past five years the faculty has challenged us to the limit and for that we are grateful.

We would like to share our gratitude towards Sandvik for enabling this master thesis. To write a master thesis in collaboration with a prominent industrial company has been rewarding for both of us. Many thanks to Macarena Ribalta, Erik Videsson and Mathias Fransson for your continuous support throughout the project.

We would also like to thank Gudrun Kiesmüller, supervisor of this master thesis, for guiding us through the thesis with important insights. We are grateful for the expertise you have provided us with.

Finally, we would like to thank our friends and families for their unconditional support. The completion of this thesis would not have been possible without you.

Lund, March 2022

Arian Marofkhani & Artur Jusopov

Abstract

Title: After-Market Spare Parts Forecasting at Sandvik Stationary Crushing & Screening.

Authors: Arian Marofkhani and Artur Jusopov

Supervisors: Professor Gudrun Kiesmüller, Lund University, Faculty of Engineering, Division of Production Management.
Macarena Ribalta, Planning & Logistics Manager, Sandvik Stationary Crushing & Screening.

Examiner: Professor Johan Marklund, Lund University, Faculty of Engineering, Division of Production Management.

Background: Sandvik Stationary Crushing & Screening in Svedala is planning to fully roll out a forecasting system called Voyager and needs guidance in their forecasting process. Sandvik wants to explore how different forecasting methods could help to improve forecast accuracy on a SKU level as well as on a SKU/Stockroom/Customer Cluster level.

Purpose: The purpose of the master thesis is to identify and propose quantitative forecasting methods with the aim to improve forecasting accuracy on a SKU and SKU/Stockroom/Customer Cluster level.

Methodology: The research approach aims to fulfill the purpose of the study by performing a single case study research. The study adopts an exploratory, explanatory and a descriptive focus to gain deep insight into the research area as well as to understand the current situation of the case company - Sandvik SRP AB. The study incorporates an empirical, data driven approach to collect and filter historical data as well as apply quantitative forecasting methods.

Results: The best forecasting method was determined for all ABC-XYZ classes. The best method for each ABC-XYZ class performed better than Moving Average 12. Simple Exponential Smoothing yielded the best forecast accuracy for classes AX, AY, BX, BY, CX and CY on both level 1- and 3 with an exception of class CX on level 1. SBA and Croston's method yielded the best forecast accuracy for classes AZ, BZ and CZ on both level 1- and 3. The reliability of point forecasts seems to increase with a lower coefficient of variation in time-series.

Recommendations: It is recommended that Sandvik classifies their products according to an ABC-XYZ classification, where Simple Exponential Smoothing is the recommended forecasting method for classes AX, AY, BX, BY and CY on both level 1 and 3. It is also recommended that Simple Exponential Smoothing should be used for class CX on level 3. Furthermore, it is recommended that SBA and Croston's method should be used for classes AZ, BZ and CZ on both level 1 and 3. Finally, it is recommended that the forecasting accuracy is monitored with the help of a tracking signal to ensure tolerable forecasting accuracy.

Sammanfattning

Titel: Prognostisering av reservdelar på eftermarknaden hos Sandvik Stationary Crushing & Screening

Författare: Arian Marofkhani and Artur Jusopov

Handledare: Professor Gudrun Kiesmüller, Lunds Universitet, Lunds Tekniska Högskola (LTH), Avdelningen för Produktionsekonomi.
Macarena Ribalta, Planning & Logistics Manager, Sandvik Stationary Crushing & Screening.

Examinator: Professor Johan Marklund, Lunds Universitet, Lunds Tekniska Högskola (LTH), Avdelningen för Produktionsekonomi.

Bakgrund: Sandvik Stationary Crushing & Screening i Svedala ämnar att implementera ett nytt prognostiseringssystem som heter Voyager och efterfrågar hjälp inom sin prognostiseringsprocess. Sandvik vill undersöka hur olika prognostiseringsmetoder kan hjälpa till att förbättra prognosernas noggrannhet på artikel- och artikel/varuhus/kundkluster nivå.

Syfte: Syftet med examensarbetet är att identifiera och föreslå kvantitativa prognostiseringsmetoder med målet att förbättra prognosernas noggrannhet på artikel- och artikel/varuhus/kundkluster nivå.

Metod: Forskningsmetodikens syfte är att uppfylla ändamålet med studien genom att genomföra en fallstudie. Studien ansätter ett explorativt, deskriptivt och förklarande fokus för att få en djupare insikt inom forskningsområdet samt för att förstå den nuvarande situationen hos fallföretaget Sandvik SRP AB. Studien integrerar en empirisk, datadriven metodik för att samla och filtrera historisk data samt för att tillämpa kvantitativa prognosmetoder.

Resultat: Den bästa prognostiseringsmetoden bestämdes för samtliga ABC-XYZ klasser. Den bästa metoden för varje enskild ABC-XYZ klass presterade bättre jämfört med 12 månaders glidande medelvärde. Enkel exponentiell utjämning gav upphov till den bästa prognosnoggrannheten för samtliga X och Y-klasser på både nivå 1 och 3 med klass CX som undantag. SBA och Crostons metod gav upphov till den bästa prognosnoggrannheten för samtliga Z-klasser på både nivå 1 och 3. Punktprognosernas pålitlighet verkar öka med minskad variationskoefficient i tidsserierna.

Rekommendationer: Det är rekommenderat att Sandvik klassificerar sina produkter enligt en ABC-XYZ klassifikation där Enkel Exponentiell Utjämning är den rekommenderade prognosmetoden för AX, AY, BX, BY och CY på både nivå 1 och 3. Enkel Exponentiell Utjämning är också den bästa prognosmetoden för klass CX på nivå 3. Det är även rekommenderat att SBA och Crostons metod används för klassifikationerna AZ, BZ och CZ på både nivå 1 och 3. Vidare är det rekommenderat att prognosnoggrannheten övervakas med hjälp av en styrsignal för att säkerställa acceptabel prognosnoggrannhet.

Table of Contents

1. Introduction

The following section aims to provide the reader with background information to understand the driving forces behind this master thesis. Moreover, the case company will be described and the purpose of this report will be outlined. The scope and delimitations of the report will conclude the introduction chapter.

1.1 Background

Supply chain management described in its core is to plan, control and manage the flow of physical goods from supplier to customer. Physical goods is an umbrella term for raw materials, WIP and finished goods, which are also known as inventories. For manufacturing companies, a considerable amount of capital is tied into inventories. Therefore, managing inventories is one of the main priorities for top management (Axsäter, 2006). Setting sufficient inventory levels is one of the most critical tasks of an AM supply chain. Component availability requirements in the AM are high as the effects of stock outs may be financially harmful for a company (Hua, Zhang, Yang, Tan, 2007). SKUs in the AM often experience volatile demand patterns. This imposes uncertainty in terms of determining sufficient stock levels to keep at all nodes in a supply chain. Companies struggle with the trade-off between keeping too much stock, which imposes hefty inventory holding costs, and keeping too little stock, which leads to insufficient fill rate levels, lost sales and unsatisfied customers. (Pince, Turrini, Meissner, 2021; Axsäter, 2006). At the heart of the problem is the difficulty of forecasting demand with high accuracy. Due to this difficulty, efforts in the academic world are continuously being put into developing quantitative forecasting methods that capture the nature of the actual demand and foresee its size before arrival.

To mitigate the challenge of demand forecasting, quantitative forecasting methods are chosen based on demand patterns. Demand pattern classification is commonly based on two parameters; the ADI as well as CV^2. The ADI is defined as the average inter-demand interval for a demand time series. The CV^2 is defined as the coefficient of variation squared for a demand series, excluding the demand points in the series that equal zero (Johnston, Boylan, 1996; Syntetos, Boylan, Croston, 2005; Costantino, Di Gravio, Patriarca, Petrella, 2017). Traditional threshold values of these parameters suggest whether a demand pattern should be classified as Smooth, Erratic, Intermittent or Lumpy. Given the classification, the search for an optimal forecasting method can be narrowed down. Examples of traditional forecasting methods used for time series that exhibit data points with few zero demands are Simple Exponential Smoothing, Holt's Linear method and Holt-Winters method. For data that is considered lumpy or intermittent other methods are necessary. Methods such as Croston's method and Syntetos-Boylan approximation (SBA) are generally used (Sanguri, Mukheerje, 2021; Pince, Turrini, Meissner, 2021).

When forecasts are to be used for different hierarchical levels, it is generally not optimal to forecast on all levels independently. The problem arises when the forecasted quantity is not consistent throughout the hierarchical levels, i.e., if the forecasted end customer demand for different locations does not add up to the total forecasted demand. One solution to the problem is to use Hierarchical Forecasting. Hierarchical Forecasting allows for disaggregation of the generated initial forecasts of the total demand. E.g., performing a forecast on the total demand followed by disaggregating the forecasts down to specific warehouses and end customers. This method is called a "Top-down" approach. The main benefits with this approach is that demand patterns are usually easier to identify on higher hierarchical levels. If one can find a way to measure historical proportions of each node at the bottom level accurately, the disaggregation will be accurate and yield satisfying results.

The "Bottom-up" approach is used when forecasts are performed at the bottom level and then aggregated upwards. The strength of this approach is that no assumptions have to be made when aggregating. However, time series tend to be quite noisy at the lowest level of the hierarchy which make them difficult to forecast accurately.

There are also so-called “optimal” methods when forecasts on different hierarchical levels are reconciled in the hope of obtaining smaller deviations on each hierarchical level. The common theme for all hierarchical forecasting techniques is that they are consistent, i.e, the forecasted sum has to be equal on all hierarchical levels (Hyndman et al., 2011).

The mentioned methods will be used in order to distribute the forecast to all hierarchical levels. Hopefully, insights regarding the feasibility of the different techniques can be obtained.

1.2 Company Description

Sandvik Group is a Swedish multinational engineering company founded in 1862. Sandvik specializes in rock processing, metal cutting and materials technology. It is a publicly traded company listed on the Stockholm stock exchange and part of the prestigious index OMXS30, indicating that the stock is one of the thirty most traded in Sweden. Sandvik had a revenue corresponding to 86.4 billion SEK in 2020 and an operating profit of 11.18 billion SEK during the same calendar year.

Sandviks division *Stationary Crushing & Screening* (SC&S) is part of the business area *Rock Processing Solutions* (RPS). Industrial products such as breakers, demolition tools as well as both stationary and mobile crushers and screens are part of the product portfolio of RPS. The division of SC&S innovates, produces, sells and distributes equipment used for size reduction and compositional sorting of rocks. Whenever the original equipment is worn out, customers order AM parts in order to keep their operations going on full capacity. The customers of SC&S are based in all corners of the world and mainly operate within the mining and construction industry.

As Sandvik has a decentralized organization, each division within the RPS business area has full control over their own supply chain. This implies that all production and inventory related policies are determined within the SC&S division. An organizational overview of Sandviks business areas, divisions and AM product families is displayed in the figure below.

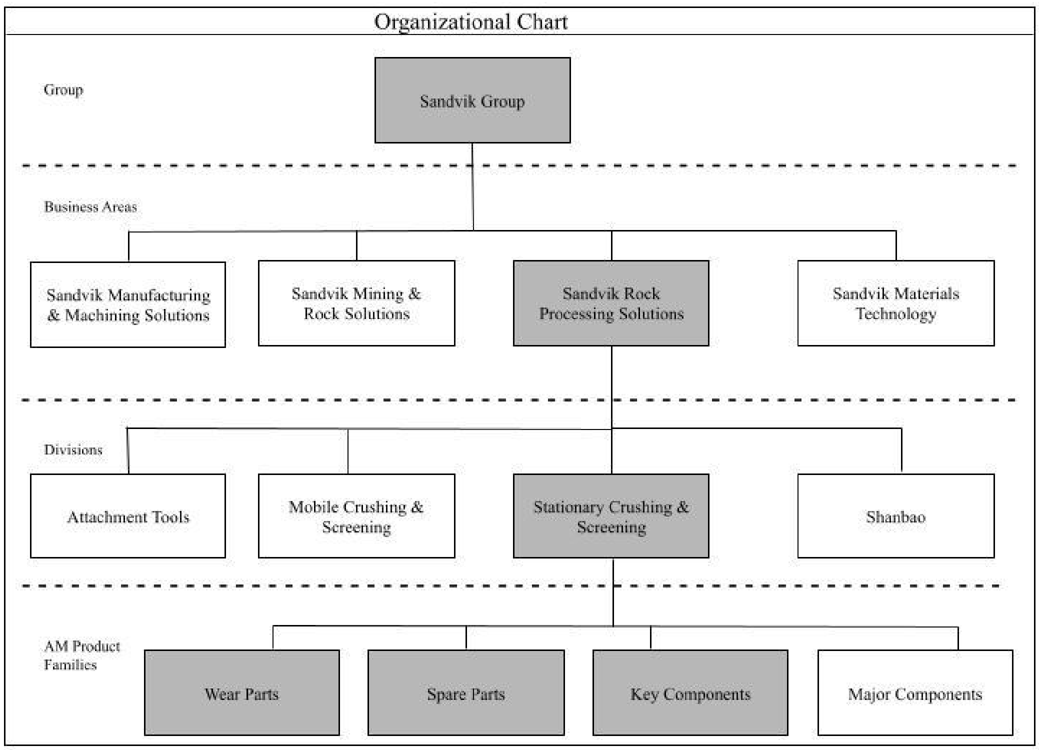

Figure 1. Sandviks organizational chart. The grey boxes are those in scope of this master thesis.

1.3 Problem Formulation

Currently, SC&S are using separate ERP systems for each entity in the supply chain. All entities are only able to view the demand one step downstream in the supply chain. The current information and material flow is illustrated below.

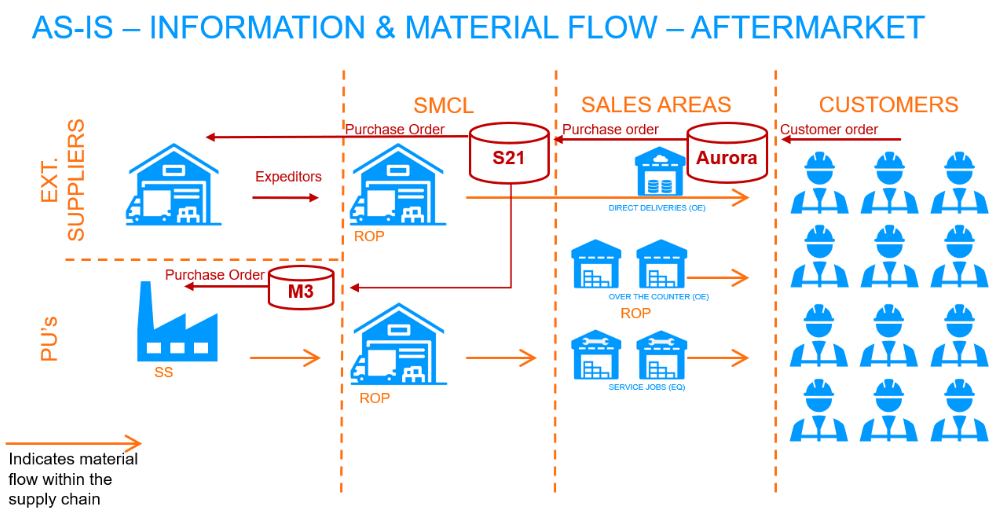

Figure 2. Illustration of the current material and information flow in the supply chain.

One crucial order winner for SC&S aftermarket is item availability for the end customers, i.e service level measured in the proportion of demand satisfied directly from stock on hand at the final echelon in the supply chain. Due to insufficient integration of existing ERP systems, difficulties in predicting final customer demand have emerged. All nodes function as isolated entities and only plan to fulfill the demand one step downstream in the supply chain. This adds to the famous bull-whip effect in each step upstream in the supply chain, which leads to the production units producing more than needed in order to fulfill the end customer demand. This is a common phenomena in supply chains where full transparency is lacking (Syntetos, Babai, Boylan, Kolassa, Nikolopoulos, 2016). The inaccurate forecasting of end customer demand has led to inadequate fill rate levels on important items, leading to lost sales in some cases due to material shortages and excess inventory in cases of overproduction.

SC&S has ambitions of fully rolling out a new forecasting software system called Voyager. The benefits with this system is that all entities within the supply chain will have full insight into the end customer demand. Therefore, the possibilities to set adequate production- and inventory levels will increase since the bull-whip effect will be reduced. The “to be” information and material flow is illustrated below.

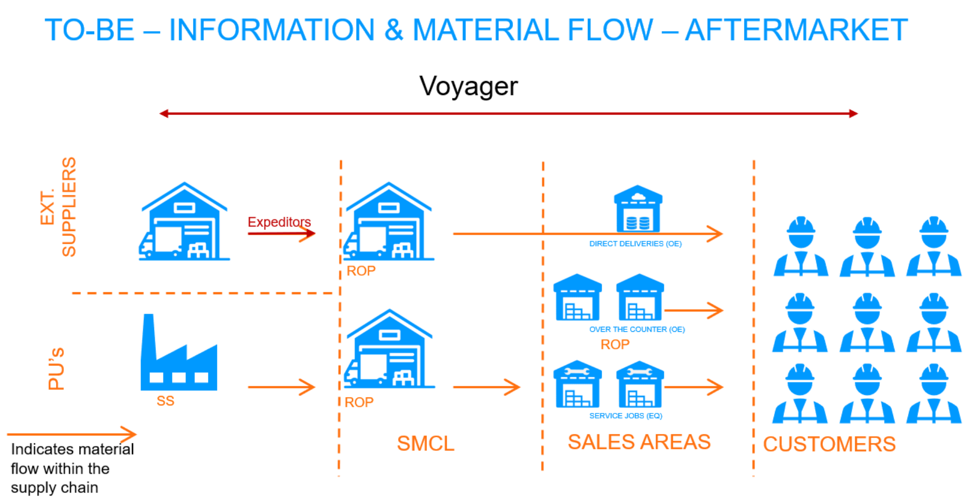

Figure 3. Illustration of the “to be” material and information flow in the supply chain.

C&S has currently chosen 30 test items for Voyager in order to validate the usability of the software as a forecasting tool. The software forecasts the demand for each month, 12 months in advance. By the end of a month, when the total demand for a SKU is logged in the system, the forecast for the upcoming 12 months will be updated in the system. However, as good as the intentions are for the system, there is an intraorganizational lack of understanding on what theoretical methods the demand forecasting is relying on. Also, the company does not seem to think that the software uses suitable forecasting methods for a plurality of the SKUs in the test set. For example, Moving Average 12 is the method that is currently being used when forecasting the absolute majority of the SKUs. It is a simple method where the future forecast is calculated using the average value of the last 12 historical observations. Due to the forecast inaccuracy of Moving Average 12 in many cases, extensive manual adjustments have been required for the forecasting of the test items in the system.

Due to the difficulties related to the current forecasting process, it seems meaningful to explore how the forecasting process within the organization could be improved. Reducing the forecast error for AM SKUs has the potential to yield improved operational efficiency.

1.4 Master Thesis Description and Purpose

1.4.1 Master Thesis Description

The ambition of this thesis is to investigate the potential in clustering on a SKU level as well as SKU/Stockroom/Customer level according to historical demand patterns. Then formulate a suitable quantitative forecasting method for each cluster on both hierarchical levels in order to improve the forecasting accuracy compared to Voyager. The formulated methods should then be used as guidance on how to forecast on level 1 and 3 in Voyager henceforth. The level of forecasting in the scope of the master thesis can be viewed below.

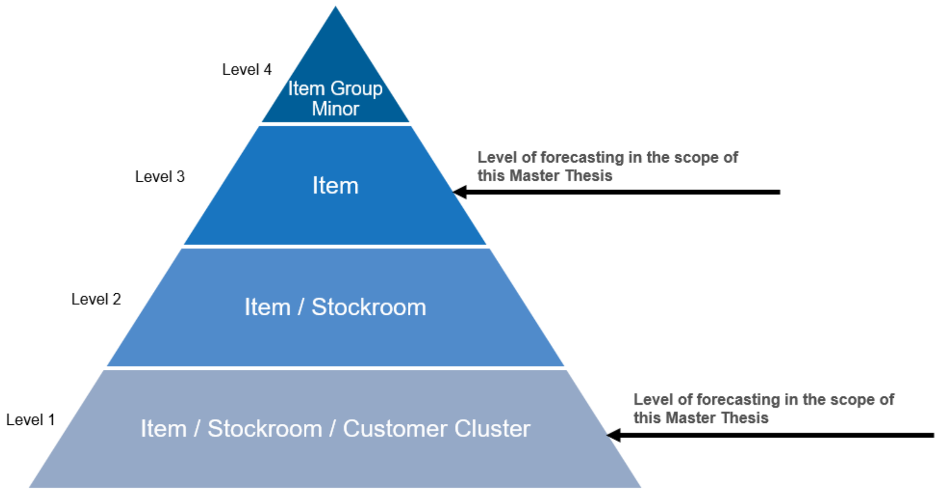

Figure 4. Level of forecasting in the scope of this masters thesis.

The forecast on SKU level, i.e level 3, will be used as a foundation for production planning for the external suppliers and the internal production units. The forecast on SKU/Stockroom/Customer Cluster level, i.e level 1, will be used as a foundation when C&S decides **what** to stock **where** and to **whom,** i.e set adequate stock levels at the last echelon of the supply chain. The reason behind why it is important to forecast on customer cluster level is due to specific service level agreements between Sandvik and their customer clusters. SKU A, stored in stockroom B, might have different service level agreements with customer clusters X and Y. This is the rationale behind forecasting on this disaggregated level.

1.4.2 Master Thesis Purpose

The purpose of the master thesis is to identify and propose quantitative forecasting methods with the aim to improve forecasting accuracy on a SKU and SKU/Stockroom/Customer Cluster level.

1.5 Delimitations

The expected duration of this master thesis is 20 weeks of full time studies. Hence, it is of high importance to point out delimitations and potential constraints in early stages of the project in order to ensure completion of the project in the given time frame.

This study was conducted at Sandvik, focusing on how to improve forecasting accuracy for AM parts. The proposed forecasts should predict the demand size for each individual month, five months ahead in time. In order to determine sufficient forecasting methods, analysis in *R* had to be conducted. Out of the four AM product families illustrated in *figure 1*, the focus will be on proposing suitable forecasting methods for items within *wear parts, spare parts and key components.* The reasoning behind leaving *major components* out of the analysis is because they require lots of manual adjustments from top management due to their hefty unit prices. Thus, it does not make sense to create a baseline forecast for such items.

Furthermore, there are some limitations relating to the raw data that Sandvik is able to provide us with. Their system cannot provide demand data that is older than three years. Meaning that the provided data will contain no more than 36 data points for each SKU (One data point is equivalent to the monthly demand of a SKU). Also, how one chooses to define demand radically impacts the raw data that will serve as the cornerstone for the forecast. One could, and probably should claim that demand occurs whenever a customer arrives and tries to place an order, independent if this actually leads to a sale (Axsäter, 2006). However, the raw demand data provided to us is equivalent to the historic sales data. Consequently, there is an unknown amount of actual demand that has not been registered in the raw data that has been given to us. Hence, there is a need to understand the inherent uncertainty of the raw data that constitutes the foundation of which the forecast relies upon.

2. Methodology

The following section aims to describe the chosen research- methodology and framework used to conduct the study. The research approach, the steps in the chosen research framework and the reasoning behind the choices will be motivated.

2.1 Research Approach

This master thesis can be considered to follow a deductive as well as an inductive approach. According to Kim (2021), there exists two main approaches for conducting research, namely the *deductive* and *inductive* methods. The deductive approach is most often related to the collection and analysis of *quantitative* data, whereas the inductive approach usually entails the handling of *qualitative* data. Both the methods follow a different approach, and the research study can vary significantly based on which of the two is chosen. However, as Saunders et al. (2007) argues, the two approaches can be used exclusively but also synchronously, as they can complement each other when conducting a research study.

The deductive approach is explained by Saunders et al. (2007) as starting by developing a theory and hypothesis followed by designing a research strategy to test the hypothesis. In the inductive approach, the data collection and analysis of the data comes first followed by developing a theory. Combining the methods has proven to give a positive impact on observed studies, and the nature of this research study has been deemed to necessitate the use of both. Going back and forth between the two would serve the best interest of conducting the study. To perform the statistical analysis, qualitative data collection and analysis is necessary. As well as using relevant literature within the field of demand forecasting to form a foundation in competence to be able to properly conduct the study. This would in turn relate to the deductive approach, where a hypothesis and a theory is formed before the data collection and analysis. However, there is a need to understand the current situation at the company, to understand the processes in place before attempting to improve them. This relates to an inductive approach, with qualitative data collection, where investigation of the company software takes place to gain an understanding of the current procedures.

2.2 Research Strategy

Building a theoretical foundation as well as knowing what type of data to collect is an important part of the procedures within a study. Therefore it is pivotal to have a research strategy in place to be able to answer the research questions to full extent. In this research study, a single case study research strategy will be adopted in order to gain the relevant information required to conduct the study. According to Saunders et al. (2007), a case study is a strategy that involves an empirical investigation of a specific contemporary occurrence within its context by using numerous sources of evidence. Eisenhardt (1989) further explains that case studies generally combine the evidence in the form of data collection, e.g qualitative- and quantitative data. To start off the case study research, it is required to have the research questions set, in order to narrow down the focus of the study and search for relevant data.

Case study research allows for explanatory and exploratory studies. According to Saunders et al. (2007), there exists three main purposes of a study, namely the descriptive, explanatory and exploratory. An explanatory study puts the focus on explaining the relationship between variables in a certain system or a problem. Exploratory research puts emphasis on seeking insights into phenomena by asking questions and reviewing them in a reconceptualizing manner. The aim of descriptive research is defined as the research conducted to describe a population; the persons, events or situations. Descriptive studies can be used as a component in the explanatory and exploratory studies. This study aims to understand the current procedures of demand forecasting at the company and present improved alternative methods. Therefore a case study strategy with an explanatory, exploratory as well as a descriptive focus is deemed to be the best fitting approach. For understanding the current situation, the descriptive focus will be followed. The explanatory focus will be followed to understand the relationship between each time series and their demand patterns. Finally the exploratory focus will be adopted to revise and conceptualize new solutions.

2.3 Research Framework

While conducting the master thesis, it is necessary to adopt a research methodology. Having a clear path makes for more efficient research by finding the most suitable way to perform the research. This is done by defining clear steps that need to be taken from formulating research questions all the way to developing a design for the study. This study will be conducted following the five steps of the forecasting process described by Sanders (2017).

1. Decide what should be forecasted
2. Collect and filter demand data
3. Time-series classification
4. Choose forecasting method
5. Measurement of accuracy

The first step in Sanders' forecasting process is to decide what should and can be forecasted. It is important to understand that some time-series simply can not be forecasted due to the lack of data or has a low forecastability due to volatile demand patterns. Following this, the necessary data needs to be prepared by collecting and filtering based on criterias such as relevance, errors or availability. When the data has been collected, it should then be classified based on e.g. the historical patterns of the demand as a first step before a relevant forecasting method is appointed. For this study, a categorization of the SKUs based on the underlying demand patterns will be performed as a first step for deciding which forecasting methods would be suitable for each category. This is done for a more efficient process by eliminating methods that are not appropriate for certain categories. After the categorization has been made, the forecasts are then produced and the performance is evaluated.

Each step in Sanders forecasting processes and their adaptation to this research study will be further detailed in depth in section 3.1.

2.4 Data Collection

There are a myriad of suitable approaches for data collection and they all contribute in different ways depending on what kind of information that is desired from the researcher. There is often a correlation between the choice of methodology and the optimal approach. One key feature for a good methodology is that it has clear guidelines on how the research should be conducted. Hence, by going through these guidelines the researcher gets a clear view of what type of data is needed before proceeding with an approach. (Björklund, Paulsson, 2012).

2.4.1 Primary and secondary data

Primary data refers to data that is collected explicitly for the purpose of answering the research questions. An example of this could be raw data provided by the case company. It could also be the exchange of information through interviews or questionnaires. Secondary data on the other hand could potentially support the research but is not necessarily intended to. It is data that originated due to other reasons but could provide implicit support for the conducted research (Björklund, Paulsson, 2012). Literature studies are good examples of secondary data. They refer to every piece of written material that could support the purpose of the research. However, literature studies come with the risk of propagating some kind of agenda and therefore it is of high importance to complement key statements in books, research papers etc with multiple sources.

2.5 Research Quality

Assuring research quality can be done by using four different logical tests. This is due to the fact that any form of research design represents a set of logical statements, case studies included. The four steps are defined below as according to Yin (2008):

- **Construct validity:** identifying correct operational measures for the concepts being studied.
- **Internal validity:** seeking to establish a causal relationship, whereby certain conditions are believed to lead to other conditions, as distinguished from spurious relationships.
- **External validity:** defining the domain to which a study’s findings can be generalized.
- **Reliability:** demonstrating that the operations of a study, such as the data collection procedures, can be repeated with the same results.

These four logical tests will be described in the following subsections as described by Yin (2008).

2.5.1 Construct Validity

The first step in assuring research quality is to identify the correct measures for conducting the study. Yin (2008) proposes three tactics that are used to construct validity. The first tactic is to use multiple sources of evidence in a way that promotes convergent lines of inquiry. The second tactic is to establish a chain of evidence. The last tactic is to have the draft of the report reviewed by key informants. The first two steps are relevant during the data collection whereas the third plays an important role further down near the end of the study.

2.5.2 Internal Validity

The researchers should, with the study, be able to explain the relationship between certain conditions and why certain events affect others. The issues that arise when attempting to address internal validity surface when the researchers infer an explanation to why unobservable events occur, based on evidence collected beforehand. If the inferences are not thoroughly examined on whether they actually represent reality then it might weaken the internal validity. Questions like if all the possibilities or alternative explanations have been addressed or if all the evidence converges into one explanation are typically the beginning of addressing internal validity. Although it is difficult to state how to build internal validity, there are some tactics that are noteworthy. The tactics are patterns matching, explanation building, addressing rival explanations or using logic models. It is important to note however that internal validity is only applicable for explanatory research, which is one of the selected research purposes for this study and therefore will be tested Yin (2008).

2.5.3 External validity

Building external validity deals with the question on whether the results of the study case are applicable on other similar cases. For case studies, the goal is to generalize a set of results. This is called analytical generalization and has been criticized to have major barriers when applying the results to other cases. However, as Yin (2008) implies, it is generally a misinterpretation by critiques done due to a flawed comparison between a survey research and a case study research. To build external validity, Yin (2008) suggests that using theory and replicating the study by applying the theory to other similar case studies is a sufficient approach.

2.5.4 Reliability

The final logical test, reliability, is performed to minimize errors and the bias in the conducted study. To test the reliability of a study, it must be done in such a way that other researchers can follow the same processes and perform the same case again. There is a clear distinction between this step and the previous. Reliability is about doing the same case study again whereas external validity is specifically about replicating the results on other, similar case studies. Therefore it is important to document the procedures during the whole study, to enable building reliability.

3. Theoretical Framework

The following section will present the theoretical framework that this study relies on. Generic clustering techniques will be presented to the reader. Moreover, forecasting methods and models will be presented thoroughly as well as common accuracy measurements. The fundamentals of hierarchical forecasting will also be outlined. The theoretical frameworks are necessary for the reader to understand in order to interpret the results of the study.

3.1 Forecasting

In a supply chain context forecasting can be viewed as an attempt to predict future demand. With high forecasting accuracy an organization is enabling higher operational efficiency when planning their production and inventory levels. There are multiple forecasting methodologies that exist. However, there is not a single forecasting methodology that is applicable for all situations due to the different demand patterns that time series experience. There are established forecasting processes that aim to clearly outline the different steps for practitioners. One of those processes is described in (Sanders, 2017) and illustrated in figure 5 below.

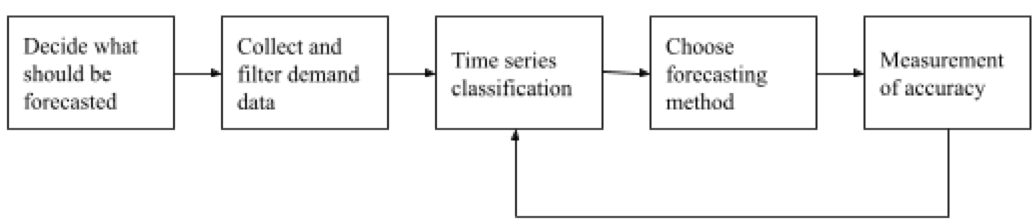

Figure 5. Sanders forecasting process.

3.1.1 Decide What Should be Forecasted

The first step for an organization is to agree upon what should be forecasted. An important part of this step of the process is to identify the purpose of the forecast. The time horizon of the forecast and how often the forecast should be updated are also crucial for the organization to fully understand before initiating the process (Sanders, 2017).

Many organizations offer a wide range of SKUs to their customers. Investigating optimal forecasting methods by going through each SKU individually is rather time consuming. Therefore, this approach is not considered to be particularly resource efficient. An alternative to this approach would be to classify the SKUs to different clusters and assign appropriate forecasting method(s) to each cluster.

Another important aspect to consider before initiating the forecasting process is to have reasonable expectations on the forecasts. The forecastability of a time-series is important to understand before having any expectations on forecasting accuracy. Forecastability of a time-series is strongly correlated with the variation it experiences. Some time-series have very low variation, meaning that the probability mass of the distribution is compact around the mean value of the series. In these cases, the forecastability is high and it is reasonable to have high expectations on a deterministic point forecast. Other time-series experience high variation with wide tails for the probability distribution. In such cases, forecastability is low and a deterministic point forecast is not reliable. Having a solid understanding of the forecastability of the SKU assortment eases the decision of what should- and what should not be forecasted.

3.1.2 Collect and Filter Demand Data

According to Hyndman and Athanasopoulos (2018), there exists two types of data. Statistical data and the expertise of the people that manage the data and will be using the generated forecasts. The available data is often not enough to fit a statistical model to yield forecasts with desirable accuracy. Generally, the more data available the better the chances of fitting a good statistical model. But old data can provide outdated information, then it can be more useful to use more recent data. Armstrong (2003) highlights the effect of inaccurate data. Errors, missing data or changes to the system that is to be forecasted can lead to forecasting errors. These uncertainties can arise due to lost sales, when the demand that has been documented only takes the actual sales into consideration while the actual demand might have been higher.

3.1.3 Time-series Classification

Historical time-series data can exhibit different underlying demand patterns. The various patterns can affect the forecastability in such a way that some time-series can prove to be impossible to accurately forecast. Heinecke et al. (2011) proposes a method for classification based on two coefficients, the average interarrival time between demand points and the variation of demand quantities. The classifications are made to get a better understanding of the behavior of the demand patterns and as such provide an idea of which methods and models are suitable for certain time-series. The four classes mentioned by Heinecke et al. (2011) are as follows,

- Intermittent
- Lumpy
- Smooth
- Erratic

Intermittent demand exhibits demand patterns where the quantities have little variation but the intervals between each demand point shows a high variation. Lumpy patterns are similar to intermittent patterns but differ in demand quantities, where they display a high variation and the most difficult of the four classifications when it comes to producing reliable forecasts. Erratic demand patterns have relatively stable interdemand times, however the quantities show a large variation. Smooth patterns make for the most forecastable time-series where the interdemand times are commonly occuring and quantities show little variation.

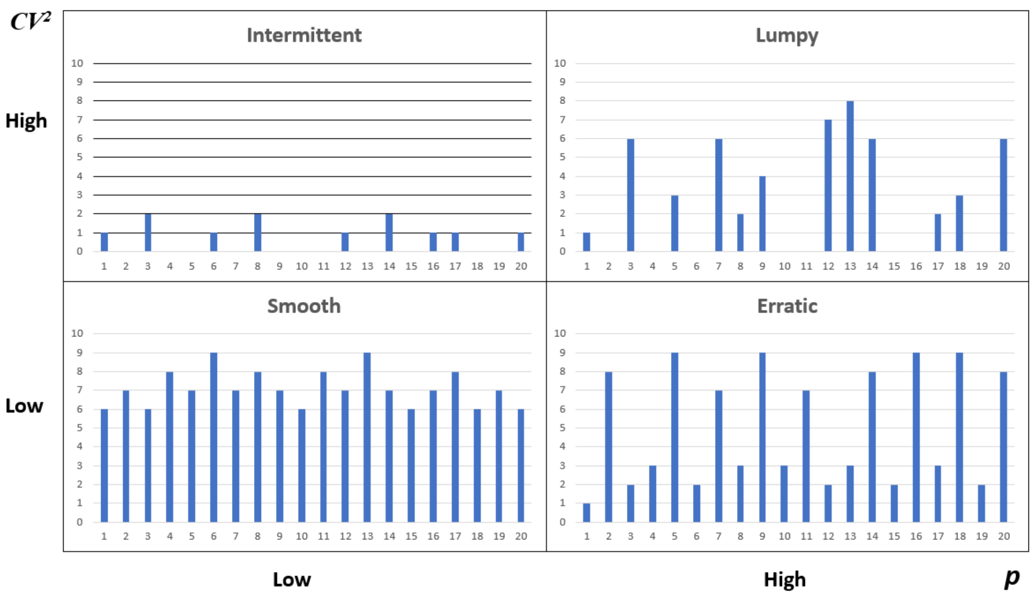

Figure 6. The four demand patterns based on their interrrival times and demand quantities.

As mentioned, the classification is done by determining two coefficients,

$CV^2 = \frac{\sigma^2}{\mu^2}$, the squared coefficient of variation which is the squared standard deviation (σ^2) divided by the squared mean (μ^2). The measurement is a unitless quantity that describes the dispersion of the data around the mean.

p , the average of the interval between two non zero demand points. This metric is a measure of the regularity of the demand in time.

Heinecke et al. (2011) further elaborates on the thresholds for each category,

- ❖ Smooth - $p < 1.32$, $CV^2 < 0.49$
- ❖ Erratic - $p < 1.32$, $CV^2 \geq 0.49$
- ❖ Lumpy - $p \geq 1.32$, $CV^2 \geq 0.49$
- ❖ Intermittent - $p \geq 1.32$, $CV^2 < 0.49$

which can also be visualized in a graph as shown in figure 7.

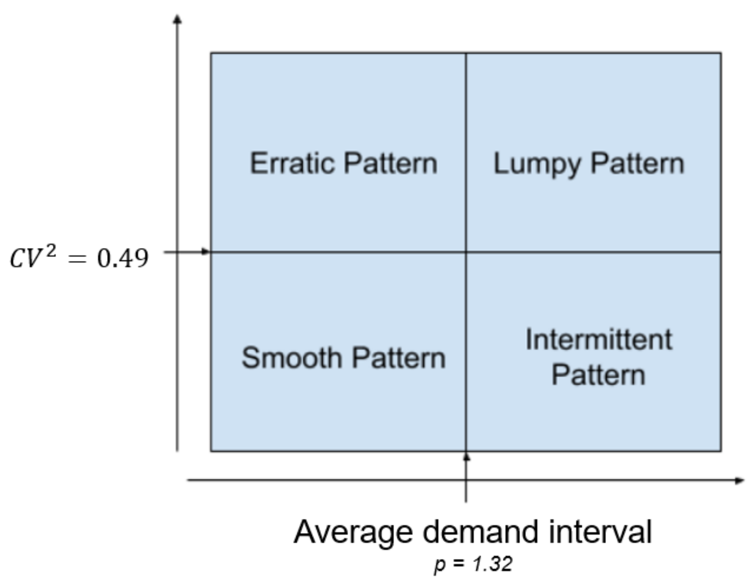

Figure 7. The threshold values for the different classifications.

3.1.3.1 ABC-Analysis

ABC-analysis is conducted to classify time-series according to the Pareto principle, which is based on the observation that a small number of elements make up the lion share of achieved results (Stojanovicm, Regodic, 2017). The Pareto principle implies the same concept as the 80:20 rule, which means that 20% of sold articles contribute 80% to the revenue. The ABC-analysis in combination with the Pareto rule establishes three groups. Group A contains around 20% of the products that contribute to 80% of the total sales value. Group B contains products that contribute to 15% of the total sales value. Group C contains products that contribute to 5% of the total sales value (Buliński, Waszkiewicz, Buraczewski, 2013 The graphic representation of this classification is illustrated in figure 8 below.

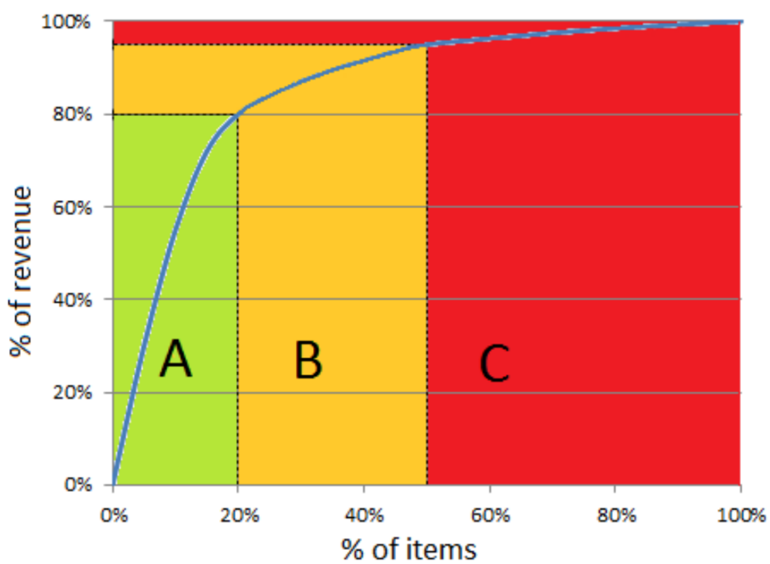

Figure 8. Illustration of the ABC-classification.

3.1.3.2 XYZ-Analysis

XYZ-analysis is conducted in order to classify time-series according to demand variation. This is a necessary process for organizations that want to distinguish time-series by demand uncertainty. The XYZ-analysis establishes three groups according to the coefficient of variation for a time-series. Group X contains time-series with continuous demand and are characterized with small demand fluctuations over time. Group Y contains time-series with discontinuous demand and are characterized to have intermediate demand fluctuations over time. Group Z contains time-series where demand arrives very sporadically and these are extremely difficult to forecast accurately. The threshold values for the coefficient of variation are discussed in (D'Alessandro, Baveja, 2000; Stojanovicm, Regodic, 2017). The authors suggest that elements are separated with the help of a reference value of the coefficient of variation according to:

$Group\ X\ -\ CofV\ < 0.52$
$Group\ Y\ -\ 0.52 \leq CofV\ \leq 1.0$
$Group\ Z\ -\ CofV\ > 1.0$

The definition of $CofV$ follows below

$$\mu = \frac{1}{T}\sum_{t=1}^{T} D_t \quad (1)$$

$$\sigma = \sqrt{\frac{1}{T}\sum_{t=1}^{T} (D_t - \mu)^2} \quad (2)$$

$$CofV = \frac{\sigma}{\mu} \quad (3)$$

Where:
μ − mean demand between periods *1* and *T*.
σ −standard deviation of demand between periods *1* and *T*.

D_t −Demand for period t.

3.1.3.3 ABC/XYZ-Analysis

If the ABC and XYZ classifications are merged, 9 groups are created that time-series can be classified as. It is useful to make this classification as it eases the process to provide generic forecasting strategies for each group. For example, if forecasting method(s) are significantly represented as the most appropriate for a certain group, the organization just has to look at the classification for a time-series in order to know what forecasting method(s) are appropriate to use. Characteristics for each group are presented in the table 1 below.

Table 1. The characteristics for each group in an ABC/XYZ classification.

	A	**B**	**C**
X	High value. High predictability. Continuous demand.	Medium value. High predictability. Continuous demand.	Low value. High predictability. Continuous demand.
Y	High value. Medium predictability. Fluctuating demand.	Medium value. Medium predictability. Fluctuating demand.	Low value. Medium predictability. Fluctuating demand.
Z	High value. Low predictability. Irregular demand.	Medium value. Low predictability. Irregular demand.	Low value. Low predictability. Irregular demand.

3.1.4 Choose Forecasting Method

Forecasting methodologies can be divided into two main categories, *qualitative and quantitative.* It is therefore important to point out that this report will solely focus on quantitative forecasting methods. In contrast to qualitative forecasting, quantitative forecasting relies on mathematical statistics. Mathematics is the necessary foundation to capture the nature of historical demand and henceforth compute a forecast (Axsäter, 2006). By using a quantitative approach, a lot of subjective assumptions can be eliminated from the forecasting process. In the following section a selection of well known quantitative forecasting models and methods will be presented.

3.1.4.1 Moving Average Methods

The moving average method is a common forecasting method for time-series analysis. The common trait for all time-series methodologies is that they account for what has happened in the past and try to give a prediction of the future. However, the methodologies will yield different forecasts based on how they weight the historical sequence in the time-series. The simple moving average method creates a forecast by computing the average demand of the last N periods in the time-series. Parameter N is chosen depending on the smoothness of the historical demand. Many observations in the computation of the forecast will yield a stable forecast. In contrast, less observations in the computation of the forecast will yield a forecast that is more sensitive to changes in demand (Axsäter, 2006). The benefits with having a moving average is that old data will be filtered out as the forecast is updated on a periodic basis. The equation of simple moving average and its associated definitions are presented in table 2 below.

Table 2.

$$F_{t+1} = \frac{D_t + D_{t-1} + D_{t-2} + ... + D_{t-N+1}}{N} \qquad (4)$$

Where:

F_{t+1} - Forecast for period *t+1*

D_t - Observed demand for period *t*.

N - Number of observations accounted for when computing the forecast for period *t+1*.

3.1.4.2 Naive Methods

The naive forecasting method is commonly used in the industry due to its simplicity. The method implies that the forecast for all upcoming periods is equal to the last observed demand. Due to the low complexity of the method it might come as a surprise that it has displayed higher forecasting accuracy for one-step forecasts compared to forecasting methods with higher complexity. Naturally, the method is well suited for time-series that experience low demand volatility and randomness (Krajewski et al., 2013).

There is also a customized version that is derived from the naive forecast. It is well suited for time-series with strong seasonal patterns. This version is also known as the *seasonal naive method*. The rationale behind this method is similar to the naive method. The forecast for the next period is equal to the last observed demand from the same period during last season. If the periodicity of the time-series is monthly, the forecast for the next period will be equal to the last observed demand for the same month from last year (Hyndman & Athanasopoulos, 2018).

3.1.4.3 Exponential Smoothing Methods

Exponential smoothing has proved to be a powerful method for forecasting time-series. The method was formulated a couple of decades ago and has since become a cornerstone method among practitioners (Hyndman & De Gooijer, 2006). There are three common variants of the exponential smoothing method; simple exponential smoothing, Holt's linear method and Holt-Winters method (Hyndman et al., 2008).

The simple exponential smoothing method creates a forecast that relies on the demand and forecast for the previous period. The value of the smoothing parameter α affects how sensitive the forecast is for changes in demand. The value of the smoothing parameter has a fundamental impact on how the forecast acts and assumes values in the range $0 \leq \alpha \leq 1$. A value close to 0 is of stable character and a value close to 1 is highly reactive to recent changes in the demand.

One challenge when applying simple exponential smoothing is to find a suitable optimization criterion to determine the value of the smoothing parameter α. Usually it is determined by minimizing the average error between the forecasts and observed values in a time-series. This error can be expressed in MAD, MAPE, MASE and other accuracy measurement procedures that will be further explained in chapter *3.1.5*. The equation for simple exponential smoothing and its associated definitions are presented in table 3 below.

Table 3.

$$F_{t+1} = \alpha D_t + (1 - \alpha)F_t \qquad (5)$$

Where:

F_{t+1}- Forecast for period *t+1*.

F_t - Forecast for period *t*. D_t - Observed demand for period *t*. $0 \leq \alpha \leq 1$

Holt's linear method is originally derived from simple exponential smoothing with the difference that it contains a trend component (Holt, C. E., 1957). The method can be divided into two parts. Equation (7) computes a forecast according to simple exponential smoothing but with a trend component b_{t-1} added to the last smoothed value of the demand. Equation (8) on the other hand describes where the trend component b_t emerges from. Trend component b_t is updated periodically based on the trend parameter β and the difference between the last two smoothed values of the demand. The sum of these equations yield Holt's linear one-step ahead forecast. There are cases where the trend component b_t might increase or decrease exponentially which might yield misleading forecasts on long-term horizons. In order to mitigate this issue, a third smoothing parameter can be used to dampen the trend (Hyndman & Athanasopoulos, 2018). Smoothing parameters α, β and ϕ assumes values between 0 and 1.

The equations that describe Holt's linear method and Holt's damped linear method and their associated definitions are presented in table 4 below.

Table 4.

Forecast equation: $F_{t+1} = l_t + b_t$ (6) Level Equation: $l_t = \alpha D_t + (1 - \alpha)(l_{t-1} + b_{t-1})$ (7) Trend Equation: $b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}$ (8)

Where:
F_{t+1} - One step ahead forecast from period *t*.
l_t - Exponentially smoothed demand in period *t*.
b_t - Exponentially smoothed trend in period *t*.
D_t - Observed demand for period *t*.
$0 \leq \alpha \leq 1$
$0 \leq \beta \leq 1$

Forecast equation: $F_{t+1} = l_t + \phi b_t$ (9)

Level Equation: $l_t = \alpha D_t + (1 - \alpha)(l_{t-1} + \phi b_{t-1})$ (10)

Trend Equation: $b_t = \beta(l_t - l_{t-1}) + (1 - \beta)\phi b_{t-1}$ (11)

Where:
F_{t+1} - One step ahead forecast from period *t*.
l_t - Exponentially smoothed demand in period *t*.
b_t - Exponentially smoothed trend in period *t*.
D_t - Observed demand for period *t*.
$0 \leq \alpha \leq 1$
$0 \leq \beta \leq 1$
$0 \leq \phi \leq 1$

If there is a presence of both trend and seasonality in a time-series, Holt-Winters method usually performs the best out of the exponential smoothing methods. Holt-Winters method is an extension of Holt's linear method (Winters, P. R., 1960). In addition to a level- and trend equation there is also a seasonal equation in Holt-Winters method. In this seasonal equation a seasonal index s_t is computed with the help of a seasonal smoothing parameter γ

(Hyndman & Athanasopoulos, 2017). When applying this method one has to be aware of what kind of seasonality the time-series is displaying. If the seasonality is constant over time, an additive seasonality model is preferred and the seasonal equation is added to the level- and trend equations. If the seasonality displays variations over time, a multiplicative seasonality is preferred and the seasonal equation is multiplied with the level- and trend equations. Holt-Winters method with an additive seasonal component and its associated definitions are presented in table 5.

Table 5.

Forecast Equation: $F_{t+1} = l_t + b_t + s_{t+1-m(k+1)}$ (12)
Level Equation: $l_t = \alpha(D_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1})$ (13)
Trend Equation: $b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}$ (14)
Seasonal Equation: $s_t = \gamma(D_t - b_{t-1} - l_{t-1}) + (1 - \gamma)s_{t-m}$ (15)
Where:
F_{t+1} - One step ahead forecast from period *t*.
l_t- Exponentially smoothed demand in period *t*.
b_t- Exponentially smoothed trend in period *t*.
D_t- Observed demand for period *t*.
s_t- Seasonal index for period *t*.
m - Frequency of seasonality. For example m=12 for monthly data.
k - Integer part of $\frac{h-1}{m}$. Ensures that the estimates of the seasonal indices used for forecasting come from the final year of the sample.
$0 \leq \alpha \leq 1$
$0 \leq \beta \leq 1$

$$0 \leq \gamma \leq 1$$

3.1.4.4 Croston's Method

When a time-series contains consecutive periods without demand it is referred to as intermittent. Exponential smoothing is not considered to be a useful forecasting method for time-series like this. After a large demand size has arrived the exponential smoothing method will overestimate the demand for the upcoming periods and hence yield significant forecast errors. However, the method is still used for intermittent time-series with a low value of smoothing value α, making the forecast less reactive to recent changes in demand. This still does not mitigate the issue properly and other methods such as Croston's method are usually preferred for intermittent time-series (Croston, 1972). Croston's method contains two separate time-series. The first time-series, equation (17), forecasts the size of the demand for period *t+1*. The second time-series, equation (18), forecasts the interarrival time between two non-zero demand points. The final forecast (19) is equivalent to the ratio between the forecast in equation (17 and 18). The equation for Croston's method and its associated definitions are presented in table 6 below.

Table 6.

$If\, D_t \neq 0$ then,

$$d_{t+1} = (1 - \beta)D_t + \beta d_t \quad (17)$$

$$k_{t+1} = (1 - \alpha)K_t + \alpha k_t \quad (18)$$

$$F_{t+1} = \frac{d_{t+1}}{k_{t+1}} \quad (19)$$

$If\, D_t = 0$ then,

$$d_{t+1} = d_t$$

$k_{t+1} = k_t$
$F_{t+1} = F_t$

Where:

F_{t+1} - Forecast for period *t+1*.
d_{t+1} - Forecast of the size of demand for period *t+1*.
k_{t+1} - Forecast of the interarrival time for period *t+1*.
D_t - Demand observation for period *t*.
K_t - Interarrival observation for period *t*.
$0 \leq \alpha \leq 1$
$0 \leq \beta \leq 1$

3.1.4.5 SBA Method

The SBA method is closely tied to Crostons methodology. It contains two separate time-series where one estimates the demand size and the other estimates interarrival frequency. The method was proposed as an attempt to remove the inherent bias in Crostons methodology (Syntetos, Boylan, 2005; Doszyn, 2020). The bias in Croston's method arises if one assumes that the demand size and interarrival time are independent of each other. In this case the following relationship must hold,

$$E\left(\frac{d_{t+1}}{k_{t+1}}\right) = E(d_{t+1})E\left(\frac{1}{k_{t+1}}\right) \qquad (20)$$

But,

$$E(\frac{1}{k_{t+1}}) \neq \frac{1}{E(k_{t+1})}$$

And therefore, Croston's method is biased. To cope with this bias, Syntetos & Boylan showed that by multiplying the forecast with the factor $(1 - \frac{\beta}{2})$, the bias is eliminated. Therefore they proposed this method that is named after themselves. The equation for SBAs method and its associated definitions are presented in table 7 below.

Table 7.

If $D_t \neq 0$ then,
$d_{t+1} = (1 - \beta)D_t + \beta d_t$ (21)
$k_{t+1} = (1 - \alpha)K_t + \alpha k_t$ (22)
$F_{t+1} = \frac{d_{t+1}}{k_{t+1}}(1 - \frac{\beta}{2})$ (23)
If $D_t = 0$ then,
$d_{t+1} = d_t$
$k_{t+1} = k_t$
$F_{t+1} = F_t$
Where:
F_{t+1}- Forecast for period *t+1*.
d_{t+1}- Forecast of the size of demand for period *t+1*.
k_{t+1}- Forecast of the interarrival time for period *t+1*.
D_t- Demand size for period *t*.

K_t - Interarrival observation for period *t*. $0 \leq \alpha \leq 1$ $0 \leq \beta \leq 1$

3.1.4.6 Exponential Smoothing - ETS State-Space Models

The different methods that rely on exponential smoothing have been stated earlier in this chapter. What they all have in common is that they produce deterministic point forecasts for upcoming periods. However, there are models that take stochasticity into account when computing forecasts and these models are known as *ETS State-Space Models*. An ETS state-space model computes a point forecast the same way as the underlying exponential smoothing method but also takes the forecast error distribution into account. Hence, prediction intervals where the forecast is within are yielded when forecasting with these models (Hyndman, Athanasopoulos, 2018).

Forecasts that use state-space models can be found in the *ETS Package* in *R*. The package can be used in order to determine an optimal state-space model for a particular time-series. As default, the package uses the Akaike Information Criterion (AIC_c) as an optimization criterion. AIC_c is a corrected version of *AIC* that aims to avoid overfitting. Overfitting a model to a time-series is common for small sample sizes. The formulas of *AIC* and AIC_c are presented in table 8 below.

Table 8.

$AIC = -2log(L) + 2k$ (24) $AIC_c = AIC + k(k+1)T - k - 1$ (25)

Where:

L - Likelihood function of the model. The Likelihood function describes the joint probability of the observed data as a function of the parameters of the chosen statistical model.
k - Number of parameters in the model.
T - Mean Absolute Deviation between point forecast and actual demand.

Low values are desired for both AIC and AIC_c. Thus, we can see from the equations in (24) and (25) that goodness of fit is rewarded (assessed by the likelihood function). It can also be noted that AIC and AIC_c penalize an increasing amount of parameters in a model. The penalty discourages overfitting, which is desirable since an increasing number of parameters almost always improves the goodness of fit.

The forecast of the ETS state-space models relies on three different components. These components are *Error, Trend and Seasonality.* The trend and seasonality components can be none, additive, multiplicative and for the trend component even damped. Depending on the trend and seasonality components, the method is determined. The 12 possible methods are displayed in figure 9 below.

Trend component	Seasonal component		
	N (none)	A (additive)	M (multiplicative)
N (none)	NN	NA	NM
A (additive)	AN	AA	AM
M (multiplicative)	MN	MA	MM
D (damped)	DN	DA	DM

Figure 9. The trend and seasonal component determines the 12 displayed methods.

However, for each method, there is an error component which is either additive or multiplicative. This means that for each of the 12 methods, there are 2 possible models. In total, this yields 24 different models. For example, if the *ETS* package suggests that a time-series has an additive trend component but no seasonal component, Holt's linear method will be chosen. If an additive error component is chosen, then the yielded model equates to (E,T,S)=(A,A,N) (Hyndman et al., 2002).

3.1.5 Measurement of Accuracy

An integral part of the forecasting process is to evaluate forecasting accuracy. Without measurement of the forecasting accuracy it is not possible to make quantitative comparisons between methods. This quantitative comparison is what enables one to draw a conclusion around a certain method's feasibility (Sanders, 2017).

3.1.5.1 Point Forecast Measurements

There are a myriad of different measurements of point forecasting accuracy to choose from depending on what information one wants to obtain from the forecast. The most recognized and used accuracy measurement is the Mean Absolute Deviation (MAD) (Axsäter, 2006; Armstrong, 2003). It is widely used due to its intuitive definition as well as being regarded as the industry standard way of measuring accuracy. However, there are drawbacks with using an absolute scale because it makes comparisons of accuracy between time-series less meaningful.

Accuracy measurements that are scale-dependent are suitable when comparing different methods for time-series with the same scale. Mean Squared Error

(MSE) have historically been the most recognized scale-dependent measurements used in academic research and among industry practitioners. However, there are also drawbacks with the scale-dependent measurements. First and foremost, the interpretation of the values they yield are not particularly intuitive. Also, these measurements are generally more sensitive to outliers which increase the odds that strange values are obtained for a significant number of time-series (Hyndman, De Gooijer, 2006).

Accuracy measurements that are expressed in percentage form are popular due to the fact that they mitigate the issues tied to both absolute and scale measurements. With the use of percentage based measurements, it is meaningful to compare results between time-series and the definition of the measurement is easy to interpret. Mean Absolute Percentage Error (MAPE) is the most known percentage based measurement. However, the major drawback with MAPE is that it is not suitable for a plurality of time-series, especially those with high intermittency. If MAPE is used in cases with zero-demand periods, the MAPE value either diverges towards infinity or is undefined (Hyndman, De Gooijer, 2006).

In their publication (Hyndman, Koehler, 2006), the duo argues that scaled measurements should be adopted as industry standard as they are compatible with almost all types of time-series. It is common that the absolute- and percentage-based measurements either yield uncomparable or undefined values, a problem that is mitigated by the use of the Mean Absolute Scaled Error (MASE). The MASE is equal to the ratio between the MAD and a scale factor. Since the ratio is unit free, it is possible to compare the MASE value for different time-series. The scale factor that is commonly used is the one-step naive forecast error. The MASE value is hence computed as the ratio between the MAD for the used forecasting method and the MAD for the naive method. A MASE value larger than one implies that the naive method performs better than the used forecasting method. A MASE value smaller than one implies that the naive method performs worse than the used forecasting method.

MASE was used as the accuracy measurement throughout this report. It is therefore expressed in table 9 below.

Table 9.

$$MASE = \left(\frac{\frac{1}{T}\sum_{t=1}^{T}\left|F_t - D_t\right|}{\frac{1}{T-1}\sum_{t=2}^{T}\left|D_t - D_{t-1}\right|} \right) \qquad (26)$$

Where:

F_t- Forecast for period *t*.

D_t- Observed demand for period *t*.

T- Length of time-series

3.1.5.2 Distributional Forecast Measurements

Forecasting methods presented in this thesis provide point forecasts. Usually, point forecasts are convenient as a basis for organizations to measure forecast accuracy due to its intuitive nature. However, a point forecast is just a numerical value that is bound to be incorrect. In order to evaluate the trustworthiness of a point forecast one has to evaluate the accuracy of its associated distributional forecast (Gneiting, Katzfuss, 2014).

When using the ETS forecast models presented in chapter *3.1.4.6*, point forecasts equivalent to the models underlying method are yielded. However, since there is an error component associated with a degree of stochasticity, a 95% prediction interval of the forecast is also generated.

In order to understand the concept of how a prediction interval is yielded for distributional forecasts, one has to grasp the idea of residual diagnostics in a time series. Each observation in the training set of a time series can naturally be forecasted using all previous observations. These forecasts are known as fitted values. The residual is equivalent to what is left over after fitting a model, i.e the difference between the observation and its corresponding fitted value. Residual analysis is useful when evaluating how well a model has captured the information in a data set. One crucial property of the residuals is that it has a mean value equal to or in the close vicinity of zero, otherwise the forecast can be considered to be biased. However, it is the standard deviation of the residuals that is interesting when computing a prediction interval of the forecast. In this thesis, only forecast models with additive error components were studied. Henceforth, the residuals can be considered to be normally distributed around the mean residual value of 0. A distributional forecast with normally distributed residuals is expressed as viewed in *table 10*.

Table 10.

$F_t \pm c\sigma_{residual}$ (27)

Where:

F_t - Point Forecast for period t.

c - Factor of coverage probability. When talking about prediction intervals of 95%, c is equal to 1.96.

$\sigma_{residual}$ - The standard deviation of the residuals after fitting the model in the training data set.

The accuracy measurements presented thus far in this chapter only measure point forecast accuracy. In order to evaluate the accuracy of distributional forecasts, there is a need to use other measurements.

When computing a distributional forecast, a prediction interval which the demand observation is expected to be within is created. The lower limit of this prediction interval gives the 0.025 quantile of the forecast distribution. Therefore, one should expect that the actual demand observation is below this limit 2.5% of the time and above it 97.5% of the time. Suppose one has an interest to compute the quantile forecast with probability *p* at future time *t,* this quantile forecast is denoted $f_{p,t}$. This means, the demand observation D_t is expected to be less than $f_{p,t}$ with probability *p*. If D_t is a demand observation at time *t,* the quantile score, $Q_{p,t}$, can be expressed as presented in equation * below.

$$Q_{p,t} = \begin{cases} 2(1-p)(f_{p,t} - D_t), & if\, D_t < f_{p,t} \\ 2p(D_t - f_{p,t}), & if\, D_t \geq f_{p,t} \end{cases} \qquad (28)$$

Generally there is an interest in the whole forecast distribution rather than a specific quantile. For these cases, all quantile scores are averaged over all possible values of *p* to yield the *CRPS* (Continuous Ranked Probability Score). A low value of the CRPS implies that the prediction interval is narrow which in turn implies that the reliability of the point forecast is high. Conversely, a high value of the CRPS implicates a wide prediction interval with low trustworthiness of the point forecast. Important to note is that the point forecast is equivalent to the forecast of the 0.5 quantile ($f_{0.5,t}$), which is the mean value of the forecast distribution. For instance, if one would like to compute the quantile score for all percentiles within a 95% prediction interval forecast, the CRPS formula would be expressed as in equation * below.

$$CRPS = \frac{\sum_{p=2.5}^{97.5} Q_{0.01p,t}}{95} \qquad (29)$$

However, CRPS is an absolute measurement. As with point forecasts, it is useful to compare the accuracy of distributional forecasts in a scale-free measurement. With point forecasts, the MASE measurement served this purpose. For distributional forecasts, skill scores can be computed (Gneiting, Katzfuss, 2014). With skill scores, a forecast accuracy measurement relative to some benchmark method is computed. As with MASE, the naive method is used as the benchmark method. If method X would be used as the method, its skill score would be expressed as displayed below.

$$Skill\ Score\ =\ \frac{CRPS_{Model\ X}}{CRPS_{Naive}} \quad (30)$$

3.1.6 Monitoring Point Forecasts

The demand for a SKU can naturally shift over time and this has an effect on the demand pattern. A method that is considered to be optimal at one point can be obsolete at another point in the future. In order to mitigate the risk of using a method that systematically generates inaccurate forecasts over time, it is important to keep track of the generated forecasts with the help of a signal. The function of the signal is that it relates the absolute forecast deviation of the current period with the average historical forecast deviation (Axsäter, 2006; Olhager, 2000). When the forecast error deviates from the tolerable ratios, the practitioner should be notified and examine if the current forecasting method is feasible or not. The monitoring equation is viewed below.

$$TSD_t = \frac{\left|D_t - F_t\right|}{MAD_t} \qquad (31)$$

Where:

TSD_t − Tracking signal during period *t*.

MAD_{t-1} −Mean absolute deviation during period *t-1*.

F_t- Point Forecast for period *t* .

D_t −Demand observed during period *t*.

3.2 Hierarchical Forecasting

This report aims to forecast demand on both SKU level as well as on SKU/Stockroom/Customer Cluster level. Therefore it is of high importance to understand the interdependence between forecasts on these different levels. To understand how forecasts can be translated to other hierarchical levels it is essential to study the theory of hierarchical forecasting.

Compared to traditional univariate time-series forecasting it is more challenging to forecast multivariate time-series, more known as hierarchical time-series. The underlying reason is that hierarchical time-series impose aggregation constraints that have to be accounted for when performing the forecast. The implication of the aggregation constraints is that the generated forecasts have to be consistent throughout the hierarchical structure. Coherency in the forecast brings value to practitioners as it relates time-series on high and low levels. (Mirectic, Rostami-Tabar, Nikolicic, Marinko, 2021; Hyndman, Athanasopoulos, 2018). In an organization there are different functions with forecasting scopes that vary. Where production planners focus on aggregated forecasts at high levels, regional planners tend to zoom in on different geographical areas and customers. For such purposes, disaggregated forecasts on low levels are necessary in order to be valuable (Hyndman, Athanasopoulos, 2018).

3.2.1 Hierarchical Time Series

A hierarchical time-series is a nested time-series that can be disaggregated based on different attributes. A simple example would be time-series that can be divided into smaller categories depending on geographical locations. If the total demand of a product is known; given that other attributes of the product are also known, the total demand of the product can then be further split into individual time-series with different attributes. To exemplify, If product A is sold in several countries, the total demand of A can be disaggregated into the amount of demand that comes from each individual country. These can then be further disaggregated into smaller categories such as customers or even smaller locations. Starting from the bottom most level, each category is nested into larger groups of categories, which defines the hierarchical structure of the time-series.

Below, a figure of a simple two-level hierarchical structure is presented. In total there are three types of nodes also distinguished as levels that range from 0-2, where level 0 is the total aggregated values of the observations and level 2 is the bottom disaggregated level. The aggregated top of the hierarchy, which is level 0, is denoted as D_t where y is the observed value at any given period t. The series at level 0 are then disaggregated into finer categories at level 1. The nodes in level 1 are denoted as $D_{i,t}$, where D is the value of the series in node i at any given period t. Lastly, the bottom nodes at level 2 are the completely disaggregated time-series. To exemplify, $D_{A,t}$ corresponds to the value of node A on level 1 for any period *t,* while $D_{AA,t}$ corresponds to the value of node AA on level 2 for any period t (Athanasopoulos et al. 2019). The categorization for the bottom two levels can vary depending on the time-series. The top level signifies the total aggregate series, whilst the bottom two levels are commonly differentiated by geographical locations or other attributes pertaining to the available data.

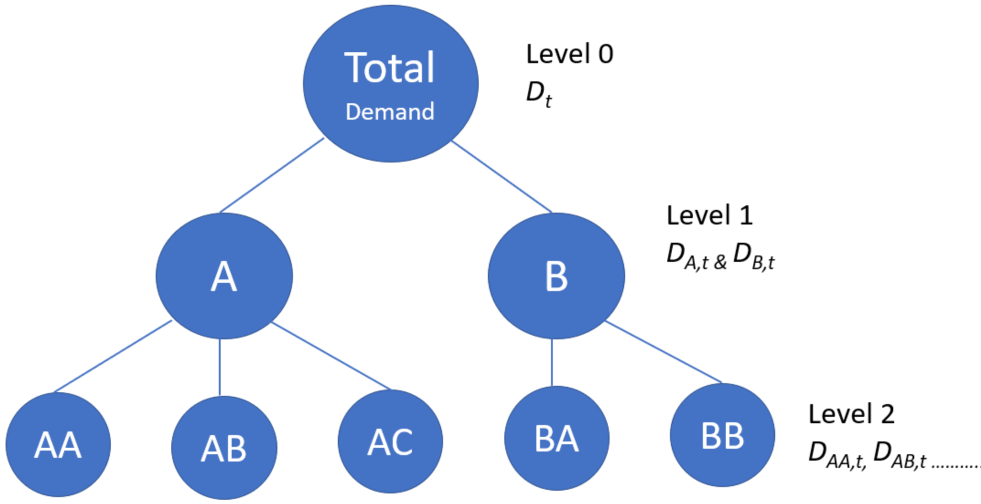

Figure 10. An example of a two-level hierarchical structure.

The aggregation constraints rule that the sum of the values of the bottom series (at level 2) add up to the sum of the series on level 1 which in turn add up to the sum of the series at level 0. As such, for any given period *t*, the following relationships must hold for a coherent structure:

$$D_t = D_{AA,t} + D_{AB,t} + D_{AC,t} + D_{BA,t} + D_{BB,t} \quad (32)$$

$$D_{A,t} = D_{AA,t} + D_{AB,t} + D_{AC,t} \quad (33)$$

$$D_{B,t} = D_{BA,t} + D_{BB,t} \quad (34)$$

Substituting (33) and (34) into (32) also yields: $D_t = D_{A,t} + D_{B,t}$

Let $n = 8$ denote the total number of nodes in the hierarchy and $m = 5$ denote the number of nodes at the bottom of the hierarchy. Then S is the $n \ x \ m$ summation matrix that explains the linear constraints when aggregating. With the summation matrix, a relationship between the total observations D_t and the total observed values at the bottom nodes can be achieved by the following relationship:

$$\widehat{D}_t = S\widehat{B}_t \tag{35}$$

which is also presented in its full form by figure 11.

$$\begin{bmatrix} D_t \\ D_{A,t} \\ D_{B,t} \\ D_{AA,t} \\ D_{AB,t} \\ D_{AC,t} \\ D_{BA,t} \\ D_{BB,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} D_{AA,t} \\ D_{AB,t} \\ D_{AC,t} \\ D_{BA,t} \\ D_{BB,t} \end{bmatrix}$$

Figure 11. The relationship between the total observations and the observed values at the bottom of the hierarchy.

In (35), $\widehat{D}_t$ is a vector that contains all time-series in the hierarchy and $\widehat{B}_t$ is a vector that contains all time-series in the bottom nodes, at level 2:

$$\widehat{D}_t = \left(D_t, D_{A,t}, D_{B,t}, D_{AA,t}, D_{AB,t}, D_{AC,t}, D_{BA,t}, D_{BB,t}\right)$$
$$\widehat{B}_t = \left(D_{AA,t}, D_{AB,t}, D_{AC,t}, D_{BA,t}, D_{BB,t}\right)$$

3.2.2 Point Forecasting with Single-Level Approaches

The most common approaches for hierarchical forecasting begin by generating base forecasts on either the top most aggregated level of the hierarchy or the bottom most disaggregate level. When the base forecasts are generated, they are then either aggregated upwards or downwards the hierarchy, depending on the chosen method. There are two methods that are most widely used, these include the top-down and bottom-up approaches (Athanasopoulos et al. 2019). However, Wickramasuriya et al. 2019 proposes an alternative method that, according to their study, performs better than the conventional methods. These three approaches will be explained in detail in the following subsections.

3.2.3 Bottom-Up Approach

The most simple way of producing coherent point forecasts is to generate base forecasts on the bottom level of the hierarchy. The forecasts are then summed up and aggregated upwards to the desired level in structure in a relatively simple manner. The advantage of using the Bottom-up method is that no information is lost and the absence of bias when aggregating the base forecast. The available data on the bottom level of the hierarchy can however prove to be very difficult to forecast due to its potential volatile nature (Athanasopoulos et al. 2019). This would especially prove to be problematic in cases where the data displays intermittency.

Let $\widehat{b}_{T+1|T}$ be a vector containing the one-step-ahead base forecasts generated for the bottom nodes, where T signifies the latest observed time period. Then the base forecasts for the bottom of the hierarchy in figure 10 would be:

$\widehat{b}_{T+1|T} = \left(F_{AA,T+1|T}, F_{AB,T+1|T}, F_{AC,T+1|T}, F_{BA,T+1|T}, F_{BB,T+1|T}\right)$, where each element in the vector corresponds to the forecasted value of all the nodes on level 2. Given this, the forecasts for the entire hierarchy is:

$\widehat{F}_{T+1|T} = S\widehat{b}_{T+1|T}$, or in its full form:

$$\begin{bmatrix} F_{T+1|T|} \\ F_{A,T+1|T|} \\ F_{B,T+1|T|} \\ F_{AA,T+1|T|} \\ F_{AB,T+1|T|} \\ F_{AC,T+1|T|} \\ F_{BA,T+1|T|} \\ F_{BB,T+1|T|} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} F_{AA,T+1|T|} \\ F_{AB,T+1|T|} \\ F_{AC,T+1|T|} \\ F_{BA,T+1|T|} \\ F_{BB,T+1|T|} \end{bmatrix}$$

Figure 12. The forecasts for the entire hierarchy when using the Bottom-up approach.

3.2.4 Top-Down Approaches

The Top-down approaches begin at the opposite end of the hierarchy, that is, the base forecasts are created on the top level of the total series D_t. The forecasts are then disaggregated downwards in the hierarchy based on a series of proportions for each node in the lowest level. The proportions represent the share of the base forecast of the total series that will be distributed to each node at level 2. According to Gross & Sohl (1990), one approach is by calculating the average historical proportions. If F_t is the forecasted value for the total series at level 0 at any given time period, then p_i is the proportion of the forecast F_t that each series at the bottom level 2 is appointed. Using the same hierarchical time-series as in figure 10, the forecasts for the bottom level series can be represented as in table 10:

Table 10.

Forecasts for the bottom level series $F_{AA,t} = p_1 F_t$ $F_{AB,t} = p_2 F_t$ $F_{AC,t} = p_3 F_t$ $F_{BA,t} = p_4 F_t$ $F_{BB,t} = p_5 F_t$ where the proportions p_i are calculated with the following formula: $p_i = \frac{1}{T}\sum_{t=1}^{T} \frac{F_{i,t}}{F_t}$, where p_i is the mean of the proportions for the bottom series i based on historical data over all periods t and T is the latest observed period.

Following this, the Top-down forecast for the entire hierarchy can be represented as:

$\widehat{F}_{T+1|T} = S\widehat{p}\,\widehat{F}_{t,T+1|T}$, where S is the summing matrix used to aggregate the forecasts to the rest of the hierarchy, $\widehat{F}_{t,T+1|T}$ is the total aggregated forecast at level 0 and p is a vector containing all the mean historical proportions of all the bottom level series$\left(p = \left(p_1, p_2, ..., p_5\right)\right)$.

The advantage of using a Top-down approach is the data used to generate the base forecasts. In its most aggregate form, the data generally displays less volatile behavior and leads to increased forecastability. However, the disadvantage is that there is a loss of information when disaggregating. Information such as seasonality, special events etc. (Athanasopoulos et al. 2019).

3.2.5 The Optimal Reconciliation Method

The Optimal Reconciliation Method is different from the previous two methods in the sense that the point forecasts are not generated on any single level of the hierarchy. Instead, forecasts are generated on every level of the hierarchy independently, without taking the aggregation constraints found in the matrix *S* into account. These forecasts are generally not coherent throughout the hierarchy and the next step in the process is therefore to adjust the forecasts into coherency (Athanasopoulos et al. 2019). Wickramasuriya et al. (2019) proposes an approach for forecast reconciliation that, according to their studies, shows an improvement of forecasting accuracy over the other existing methods.

Let $\widehat{F}_{T+1|T}$ denote a vector containing the one-step-ahead base forecasts for all the series on all levels of the hierarchy. Let then these forecasts be assembled in the same structure as figure 10. The forecasts on each level would then not be coherent because they ignore any aggregation constraints and are produced independently. This implies that generally the forecasts on level 2 would not add up to the sum of the forecasts on level 1 and so on, unless a very simple method is used such as the Naive method. The coherent forecasts for any hierarchical structure can then be formulated as follows:

$$\widehat{F}^{\star}_{T+1|T} = SG\widehat{F}_{T+1|T} \quad (36)$$

Where $\widehat{F}^{\star}_{T+1|T}$ corresponds to the coherent forecasts on all levels of the hierarchy, *S* is the summing matrix that distributes the newly generated forecasts to other levels in the hierarchy and *G* is a matrix that maps the base forecasts to the lowest level in the hierarchy. In other words, *G* generates new forecasts that are aggregated upwards with the summing matrix *S* which transforms the independent forecasts $\widehat{F}_{T+1|T}$ into the coherent forecasts $\widehat{F}^{\star}_{T+1|T}$, across the hierarchy. To avoid bias when forecasting using the optimal method, it is important that $SGS = S$, which gives a constraint on the mapping matrix G.

The next step is to identify the G matrix that gives the least error in variances of the forecasts. According to Wickramasuriya et al. (2019), the G matrix that yields least variance in the errors of the forecast compared to the actual values must follow the following structure:

$G = \left(S'W_{T+h|T}^{-1}S\right)^{-1} S'W_{T+h|T}^{-1}$, and thus the reconciled, coherent forecasts can then be written as following,

$$\widehat{F}_{T+1|T}^{\star} = S\left(S'W_{T+1|T}^{-1}S\right)^{-1} S'W_{T+1|T}^{-1}\widehat{F}_{T+1|T} \quad (37)$$

Where $W_{T+1|T}$ is the variance-covariance matrix of the errors present in the base forecasts. For the purposes of this thesis, the Minimum Trace (shrink) method was chosen to identify $W_{T+1|T}$, as it performed the best out of the four according to Athanasopoulos et al. (2019). However, it is an exceptionally challenging task, particularly when the forecasting horizon is larger than 1.

4. Analysis

The following section will provide a deeper analysis on how the theoretical framework was applied during this study. Moreover, results will be presented in three different sections. The first section deals with results associated with forecasting accuracy on SKU level. The second section deals with results associated with forecasting accuracy on SKU/Stockroom/Customer Cluster level. Finally, the third section will deal with the findings of the hierarchical forecasting analysis.

4.1 Procedures

Most of the calculations and data preparation steps were performed exclusively in Excel and the programming software *R*. The historical data and the characteristics of each time-series that formed the foundation for the analysis was provided in the form of Excel sheets by the company.

4.1.1 Data Preparation

The available data consisted of 36 monthly historical data points for each time-series. To prepare the data before performing a statistical analysis and forecasting future demands, it was filtered based on certain criterias that were deemed appropriate. One of the measures taken was to divide the data into training and testing periods, 31 and 5 respectively, as recommended by Hyndman & Athanasopoulos (2021). The reason being, after a forecast has been made, the accuracy should be measured based on real data (test data). Whereas the method parameters should be estimated using the training data. The test data should not affect the estimation of the parameters, because by doing so would yield bias in the forecasts. To validate the generated forecasts, a comparison has to be made between data that was not part of the training data set and the forecasts that are based on the training set.

Another process in the filtering of the demand data was the exclusion of all time-series that displayed less than four non-zero demand points during the training period. This was done in order to remove the time-series with insufficient amount of data points to fit a method. Another criterion used to filter was the requirement of having at least one non-zero demand point during the last five periods. This was decided together with the company spokespersons with the reasoning that any series with only zero demand during the last five months should be treated as outdated and no longer being sold.

4.1.2 Demand Classification

Before the forecasts were generated, all the time-series were placed into the categories defined in section 3.1.3. This was done for two reasons; in order to decide which forecasting methods to exclude for certain time-series and to visualize the demand patterns that are present in the data. As Croston's and SBA's methods are only fit for time-series that display lumpy and intermittent demand patterns, they were not used to forecast the series that exhibit smooth and erratic behaviors.

4.1.3 ABC-XYZ Classification

Another classification was made where the time-series were divided into nine categories as defined in section 3.1.3.3. This was done in order to illustrate the most impactful categories with regards to total historical revenue and the number of time-series present. The simplicity of performing an ABC-XYZ analysis makes for an appropriate method for classifying the time-series. As such, this was chosen as the approach for cataloging and assigning a suitable forecasting method for each series.

4.1.4 Point Forecasting

After the demand classification was made, and all the time-series were allocated to four different clusters, the forecasting methods described in section 3.1 were used to produce point forecasts for each series. It was decided that for the smooth and erratic time-series, only Simple Exponential Smoothing, Holt's Linear, Holt's Linear Damped, Naive and Seasonal Naive methods would be used as Croston's and SBA are not suitable for time-series within these two clusters. For the lumpy and intermittent clusters, Croston's, SBA, and Simple Exponential Smoothing methods were used. The forecasts were generated by first using the software R to fit the methods for each time-series and producing a table in which the method parameters were calculated. As pointed out in section *3.1.4.6*, the ETS State-Space models used the AIC_c criterion to choose the best fit model. However, to optimize the ingoing parameters the methods were optimized based on minimizing the MAD during the training period. With the given optimized parameters, the one-step-ahead forecasts were then produced for the test period.

Several methods were used to produce forecasts for each time-series. Therefore, after obtaining the forecasts, the MASE values were calculated as described in section 3.1.5.1. The method that yielded the lowest MASE value was therefore chosen as the most suitable forecasting method for each individual time-series.

4.1.5 State-Space Models and Their Use-Case

Beyond producing point forecasts for each time-series, the statistical models that have an underlying Exponential Smoothing method were investigated. As mentioned in section 3.1.4.6, the ETS state-space models generate the same point forecasts as the methods but offer further insight into the reliability of the forecasts. For all the time-series that were appointed an Exponential smoothing method as the best fit, a best fit ETS model was also studied. The statistical models generate a prediction interval that enables measuring the reliability of the point forecast. The models are then evaluated using the CRPS values calculated based on the forecast distribution. The CRPS values can in turn be used to compute the corresponding Skill Scores.

4.1.6 Hierarchical Forecasting

Exploring the possibility of using hierarchical forecasting methods is one of the objectives of the study. As the forecasting packages in the programming tool *R* only allows for one point forecasting method at a time when using hierarchical forecasting, SES was chosen as the method to examine. Moreover, the available methods are limited and as such reinforced the decision making for limiting the investigation to only one method.

The data was filtered on the most disaggregated level, namely, SKU/Stockroom/Customer Cluster level. The filtering was done by only keeping the time-series which were appointed SES as the most suitable method from section 4.1.4. Thereafter, the hierarchical forecasting methods described in section 3.2 were used to generate and aggregate/disaggregate the point forecasts on all levels. The accuracy of the forecasts were then measured using MASE. As hierarchical forecasting produces forecasts on both level 1 and 3, the MASE values were compared to the independant point forecasts from section 4.1.4 that used SES.

4.2 Results on Level 3

After training the methods during the training period, one-step ahead forecasts were produced five consecutive times during the test period. Point forecast accuracy expressed in MASE was computed for the 1632 examined SKUs.

4.2.1 Point Forecast Accuracy of Methods on Level 3

An overview of the best and worst performing methods based on the MASE values are presented in table 11 below.

Table 11. Overview of method performance on level 3.

Best Method	**Percentage of SKUs**
Naive Method	10.3%
Seasonal Naive Method	12.8%
Simple Exponential Smoothing	33.3%
Holt’s Linear Method	10.8%
Holt’s Linear Damped Method	9.7%
Croston's Method	11.8%
SBA Method	11.3%

The absolute distribution of the SKUs amongst the best methods is also illustrated in figure 13 below.

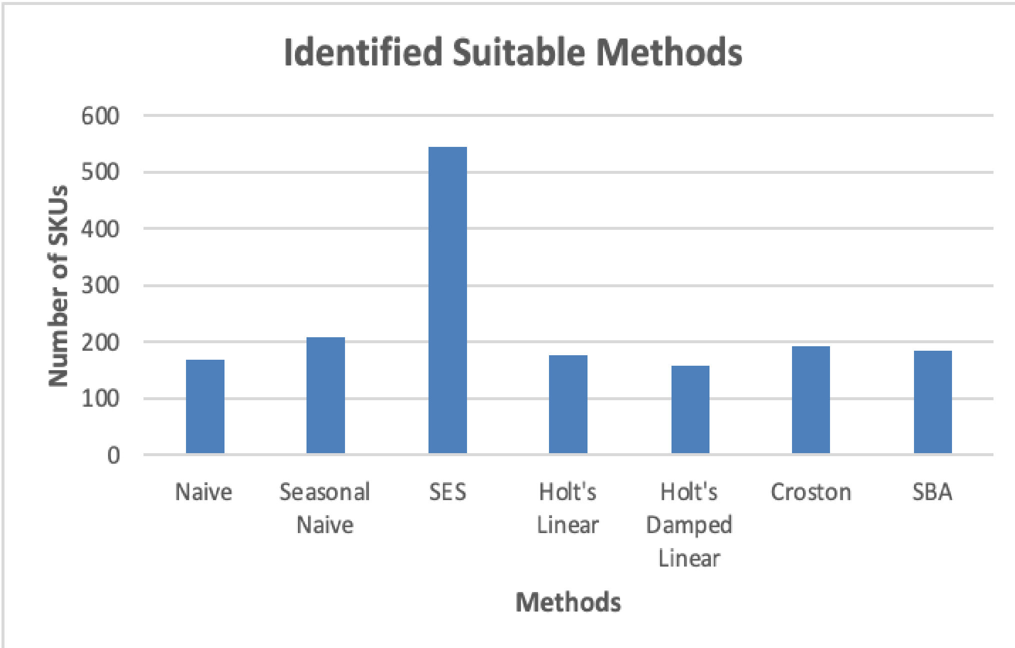

Figure 13. The absolute distribution of the SKUs amongst the best methods on level 3.

Exponential Smoothing was the method that yielded the best forecast accuracy for most SKUs, roughly 33.3% of them.

It is of high interest to visualize the distribution of the MASE values for each method. Therefore, a box plot of the MASE values was created for all methods. The black line within each box is equivalent to the median MASE value for the method. The lower- and upper bound of the boxes represent the 25th and 75th percentile of the distribution, meaning that 50% of the MASE values exist within this interval. The lower- and upper spikes could possibly deviate $\pm$ 50% from the median value of the distribution. Values that deviate by more than $\pm$ 50% from the median are considered to be outliers and are denoted with a black dot. The boxplots are illustrated in figure 14 below.

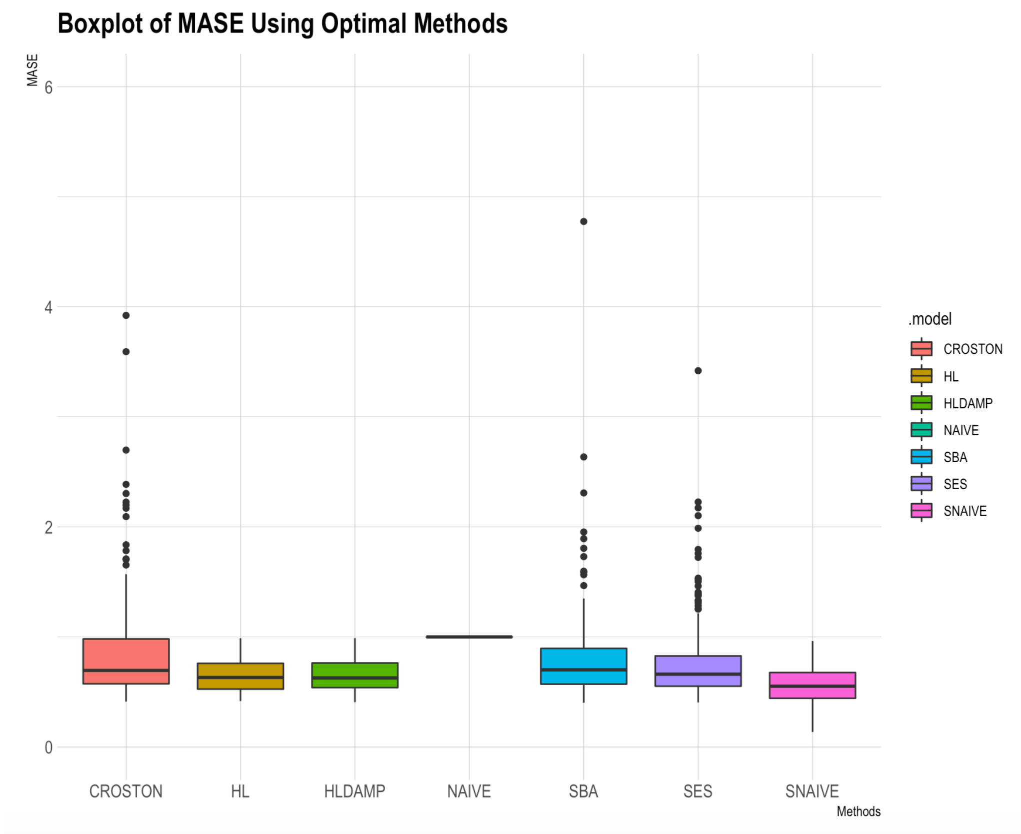

Figure 14. Box plot visualization of MASE distribution for all methods on level 3.

The box plot reveals some interesting findings. The Seasonal Naive method has the lowest average MASE value for those SKUs where it was deemed to be the best method. This is interesting as it yielded the worst MASE value for the largest number of SKUs amongst all methods. Hence, its performance seems to be very sensitive to the specific nature of the time series it is forecasting. The MASE value for Croston seems to have the largest spread around its mean value, indicating that its values are not particularly compact around the mean MASE value. The boxplot also reveals other findings. The methods that were only considered suitable for the smooth and erratic demand patterns are; Naive, Holt's Linear, Holt's Linear Damped and Seasonal Naive. What these methods have in common is that they do not display any MASE values greater than 1. For the Naive method the reason is clear, the definition of MASE implies that its value is equal to 1. For the remaining methods it is also clear why they don't display MASE values greater than 1. Had they done so, they would not be considered as the best method due to their inferiority to the Naive method. The methods that were only suitable for the lumpy and intermittent time-series were Croston's and SBA. The reason why these methods display MASE values greater than 1 is because they were not compared to the Naive methods which they probably should have been. Approximately 15% of the time-series were SBA and Croston's method was deemed the optimal would have performed better with the Naive forecast. Since Simple Exponential Smoothing was used to forecast time-series in all four demand patterns, which is the reason why the boxplot shows MASE values greater than 1 as well.

As stated in chapter *1.3*, Sandvik is currently using MA12 as the forecasting method for all of their 30 test SKUs in Voyager. Therefore, it is of high interest to compare forecast accuracy performance between MA12 and the identified best methods. What would the distribution of the MASE values look like if MA12 would have been used instead of the identified best methods? The boxplots are illustrated in figure 15 below.

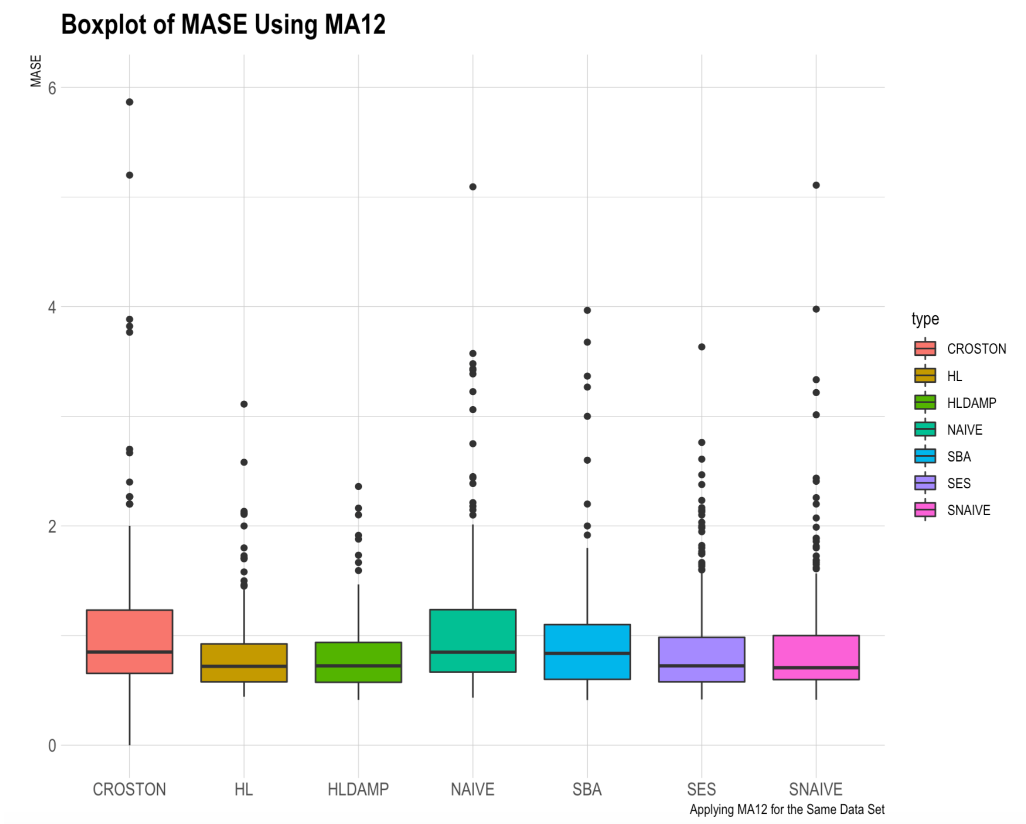

Figure 15. Box plot visualization of MASE distribution if MA12 would have been used instead of the identified best methods on level 3.

The box plot in figure 15 reveals that the MASE value for all methods, except the Naive method, performs better on both in terms of the mean and median value compared to if MA12 would have been used. This is naturally a good sign since it indicates that appointing suitable methods yields better point forecast accuracy on an average compared to using MA12 for all time series. Below in table 12, the mean and median value of the MASE distribution is displayed both for the best method and for MA12.

Table 12. Mean and median MASE values for best method and MA12 on level 3.

Best Method	Number of SKUs	Mean MASE Value Using Best Method	Mean MASE Value Using MA12	Mean MASE Value Improvement Using the Best Method	Median MASE Value Using Best Method	Median MASE Value Using MA12	Median Value Improvement Using the Best Method
Croston	192	0.87	1.18	+0.31	0.69	0.85	+0.16
SBA	184	0.80	1.05	+0.25	0.70	0.84	+0.14
SES	545	0.77	0.84	+0.07	0.67	0.73	+0.06
Holt's Linear	176	0.74	0.85	+0.11	0.65	0.72	+0.07
Holt's Linear Damped	158	0.73	0.82	+0.09	0.65	0.73	+0.08
Naive	168	1.00	0.92	-0.08	1.00	0.85	-0.15
Seasonal Naive	209	0.67	0.91	+0.24	0.58	0.70	+0.12
Total	**1632**						

4.2.2 Point Forecast Accuracy of Methods Across ABC-XYZ Classes on Level 3

The 1632 SKUs that were examined in this project were also classified according to an ABC-XYZ-analysis. Two types of ABC-XYZ charts were created. The first chart displays the distribution of SKUs across the ABC-XYZ classes. The second chart displays the historic revenue distribution of these SKUs across the ABC-XYZ classes.

The distribution of the SKUs across the ABC-XYZ classes is illustrated in figure 16 below.

	X	Y	Z
A	6.13 %	8.15 %	3.00 %
B	3.37 %	9.87 %	7.29 %
C	6.00 %	24.02 %	32.17 %

Figure 16. The distribution of the SKUs across the ABC-XYZ classes.

The distribution of the historic revenue across the ABC-XYZ classes is illustrated in figure 17 below.

	X	Y	Z
A	36.39 %	32.31 %	11.26 %
B	2.59 %	7.32 %	5.12 %
C	0.72 %	1.79 %	2.50 %

Figure 17. The distribution of the historic revenue generated from SKUs across the ABC-XYZ classes.

Figure 16 and 17 gives interesting insights into how skewed the distribution of revenue driving SKUs is. Merely 17.28% of the SKUs are part of the A-class that is generating 80% of the total revenue. Moreover, 62.19% of the SKUs are part of the C-class that only generates 5% of the total revenue. The insights that these numbers provide are valuable to have when evaluating the impact of improved point forecast accuracy for each class.

How the different forecasting methods performed within each ABC-XYZ class was also examined. An overview of the result for each class is presented in table 13 below.

Table 13. The two best forecasting methods for each ABC-XYZ class on level 3.

Classes	**Number of SKUs**	**Best Method (% of SKUs)**	**Second Best Method (% of SKUs)**
AX	100	Simple Exponential Smoothing (48%)	Holt's Linear Damped (26%)
AY	132	Simple Exponential Smoothing (36%)	Seasonal Naive (25%)
AZ	45	Croston (32%)	SBA (30%)
BX	55	Simple Exponential Smoothing (37%)	Naive (22%)
BY	160	Simple Exponential Smoothing (30%)	Seasonal Naive (18%)
BZ	116	Croston (31%)	SBA (29%)
CX	98	Simple Exponential Smoothing (33%)	Naive (18%)
CY	391	Simple Exponential Smoothing (28%)	Seasonal Naive (20%)
CZ	535	Croston (27%)	SBA (26%)
Total	**1632**		

The results show that the Exponential Smoothing method generates the most accurate point forecast for class AX, AY, BX, BY, CX and CY. It should also be noted that the naive and seasonal naive method performed quite well for the X and Y classes. Moreover, it can be ascertained that Croston's and SBAs methods are dominant within class AZ, BZ and CZ where the forecastability is lower compared to the other classes.

Within class X, simple exponential smoothing was deemed to be the best method for 39% of the SKUs. In class X, the smoothing parameter α assumed values in a range between 0.00 and 0.21. 20% of the SKUs displayed some kind of trend where one of Holt's methods was deemed to be the best fit. In the cases where a trend component was found, trend parameter β only assumed values between 0.00 and 0.06, indicating that the trend slope was very flat and that simple exponential smoothing would work just fine as an alternative method. In cases when Holt's Linear Damped method was deemed to be the best, the damping parameter γ assumed values between 0.88 and 0.98. This implies that the influence of the damping parameter is not particularly impactful and that it would yield similar results as Holt's Linear Method, which in turn yielded similar results as simple exponential smoothing.

Within class Y, simple exponential smoothing was deemed to be the best method for 30% of the SKUs. In class Y, the smoothing parameter α assumed values in a range between 0.00 and 0.39. The wider range of α compared to class X indicates that there are more time-series which have higher reactivity to recent demand occurrences. In other words, some forecasts in this class are not as smooth as they are in class X. 13% of the SKUs displayed some kind of trend where one of Holt's methods was deemed to be the best fit. In the cases where a trend component was found, trend parameter β assumed values between 0.00 and 0.11. Even in this case, the trend slope is quite flat and simple exponential smoothing would yield a similar forecast. In cases when Holt's Linear Damped method was deemed to be the best, the damping parameter γ assumed values between 0.83 and 0.98.

Within class Z, Croston and SBAs methods constituted the best method for 55% of the SKUs. For Croston's method in class Z, the smoothing parameter α assumed values in a range between 0.00 and 0.29 and smoothing parameter β assumed values in a range between 0.00 and 0.23. For SBAs method in class Z, the smoothing parameter α assumed values in a range between 0.00 and 0.32 and smoothing parameter β assumed values in a range between 0.00 and 0.23. The fact that these two methods performed the best within class Z is not surprising. All time series within class Z show a coefficient of variation greater than 1. SBA and Croston are two methods that seem to capture variability in time-series in a better way compared to the other examined methods.

It is of interest to visualize the distribution of the MASE values for each class. Therefore, a box plot of the MASE values was created for all classes. The black line within each box is equivalent to the median MASE value for the class. The lower- and upper bound of the boxes represent the 25th and 75th percentile of the distribution, meaning that 50% of the MASE values exist within this interval. The lower- and upper spikes could possibly deviate $\pm\ 50\%$ from the median value of the distribution. Values that deviate by more than $\pm\ 50\%$ from the median are considered to be outliers and are denoted with a black dot. The boxplots are illustrated in figure 18 below.

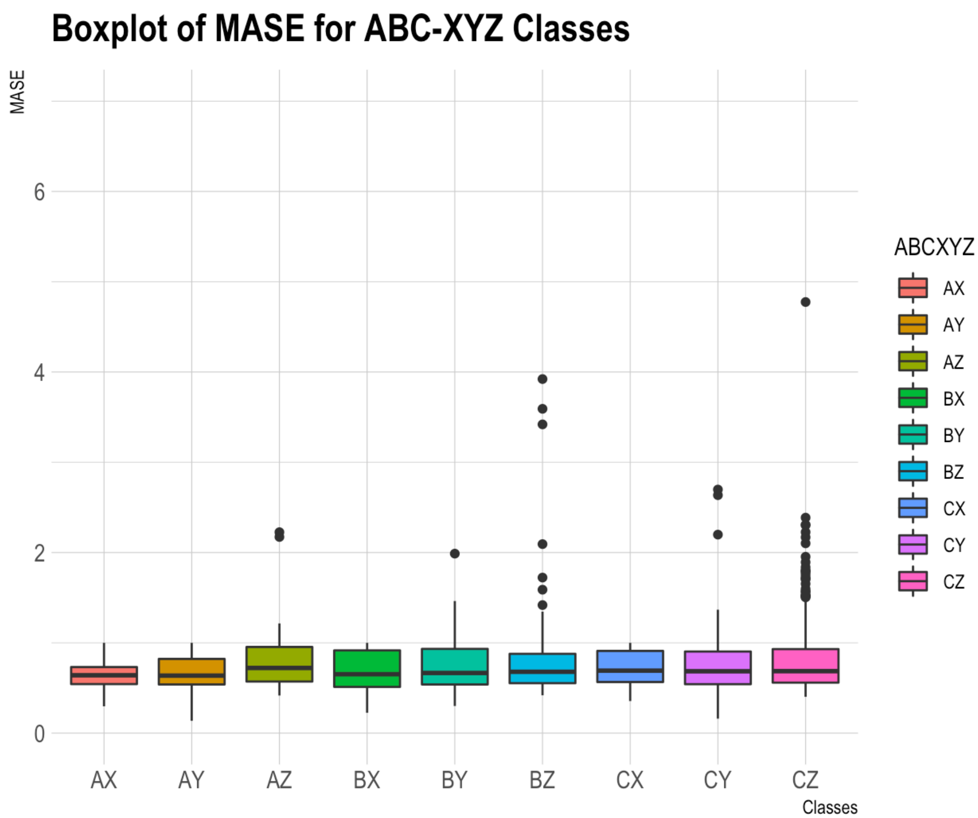

Figure 18. Box plot visualization of MASE distribution for all ABC-XYZ classes on level 3.

Some interesting insights are provided when looking at the box plot in figure 18. It seems like the median MASE value is the lowest for X-classes and highest for the Z-classes. Also, the X-classes seem to have MASE values that are more compact around their median value compared to the Z-classes. Finally, the outliers seem to be more extreme for the Y and particularly Z-classes compared to class X. This could be explained by the fact that class Y and Z make out a larger portion of the SKUs and hence the probability for extreme outliers increase. However, it could also be explained by the fact that forecastability is lower for the Y and Z-classes and therefore extreme MASE values are yielded in some instances. Generally, it seems like there is some correlation between MASE values and the XYZ-classification whereas the MASE-values seem indifferent to the ABC-classification. This insight is intuitive. Forecastability of a SKU has an impact on point forecast accuracy whereas the monetary value contribution of an SKU does not.

4.2.3 Distributional Forecast Accuracy of ETS State-Space Models Across XYZ Classes on Level 3

As mentioned in chapter *3.1.5.2*, it is also interesting to examine the accuracy of the distributional forecasts. For the three methods in the ETS State-Space models; Simple Exponential Smoothing, Holt's Linear and Holt's Linear Damped, point forecasts equivalent to the underlying methods were yielded. However, since these models also incorporated additive errors, a distributional forecast with 95% prediction intervals was created. A boxplot for the X, Y and Z classes was created to investigate the distribution of the skill scores for each class. The boxplot can be viewed in figure 19 below.

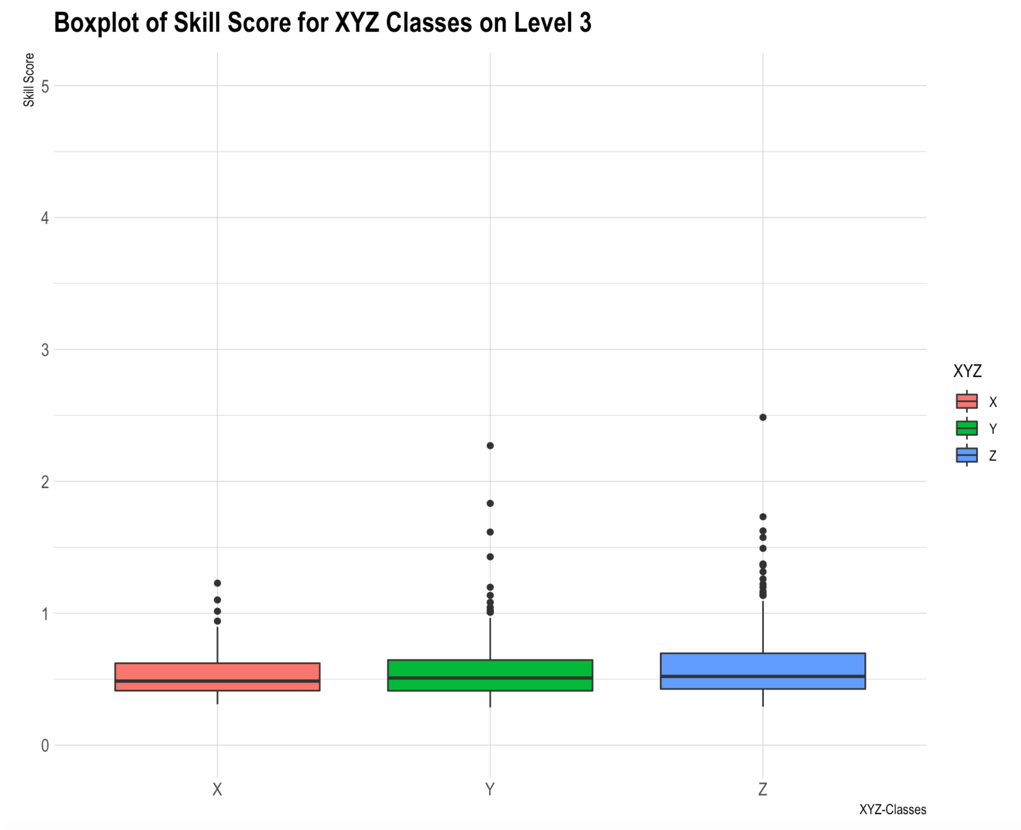

Figure 19. Boxplot of Skill Score for XYZ classes on Level 3.

It is interesting to see the discrepancy in the skill score distributions between the classes. Judging from figure 19 it looks like the median skill scores across the classes are similar. However, the increase of outliers as the coefficient of variation increases must be considered. This phenomena strengthens the notion that time-series with low coefficient of variation yield relatively compact distributional forecasts and hence produce low skill scores. The interpretation of a low skill score is that it strengthens the reliability of the point forecast.

4.2.4 Benefits of Improved Forecast Accuracy on Level 3

Improved forecasting accuracy on level 3 enables Sandvik to plan their production in a more cost efficient manner. The provided ABC-XYZ classification of Sandviks SKUs provide a solid overview of which classes that have constituted the lion share of the revenue. Class AX and AY only make up 14.28% of the SKUs but make out 68.70% of the revenue. Simple Exponential Smoothing was deemed to be the best method for 41% of the SKUs in these classes. For SKUs where Simple Exponential Smoothing was the best method, it yielded an average MASE value of 0.77. For the same SKUs, Moving Average 12 yielded an average MASE value of 0.84. Hence, with an implementation of Simple Exponential smoothing, there is potential in improving forecasting accuracy slightly. Even if the relative forecasting accuracy improvements are small, it has the potential to achieve noticeable monetary effects. The takeaway is that small forecasting improvements in important classes could have significant impacts on efficiency in production planning.

The most potential in improving relative forecasting accuracy is within the Z-classes where Croston and SBAs methods are dominant. The SKUs in the Z-classes have a coefficient of variance greater than 1, which implies difficulties in forecasting demand accurately. The Skill scores presented in chapter 4.2.3 also cemented this notion. It is also within these classes where most of the SKUs are allocated, mainly in the CZ category. However, the historical revenue within this class is, due to the C classification, significantly lower relative to the other categories. Therefore, the monetary impact of improving forecast accuracy within class CZ would be small. Under the condition that there is enough capacity in production, it would be a good idea to produce large batches of SKUs within class CZ and push the finished goods downstream close to the customers in order to ensure high product availability. The takeaway is that forecasting improvements in less important classes don’t have significant impacts on efficiency in production planning.

4.3 Results on Level 1

The same procedure that was carried out for time-series on level 3 was also conducted for time-series on level 1. After training the methods, one-step ahead forecasts were produced five consecutive times. Point forecast accuracy expressed in MASE was computed for the 5668 SKU/Stockroom/Customer Cluster time-series.

4.3.1 Point Forecast Accuracy of Methods on Level 1

An overview of the best and worst performing methods based on the MASE values are presented in table 14 below.

Table 14. Overview of method performance on level 1.

Best Method	**Percentage of SKUs**
Naive Method	2.0%
Seasonal Naive Method	2.1%
Simple Exponential Smoothing	28.1%
Holt’s Linear Method	2.0%
Holt’s Linear Damped Method	1.7%
Croston's Method	31.3%
SBA Method	32.8%

The absolute distribution of the SKU/Stockroom/Customer Cluster time-series amongst the best methods is also illustrated in figure 20 below.

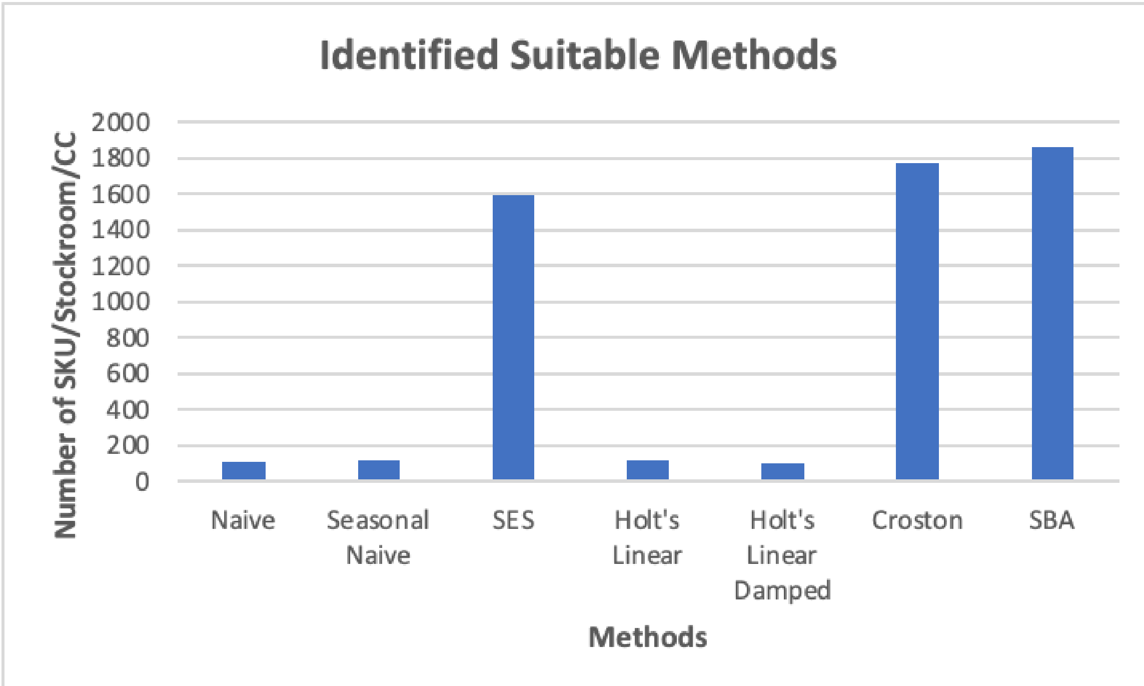

Figure 20. The absolute distribution of the SKU/Stockroom/Customer Clusters amongst the best methods on level 1.

SBA was the method that yielded the best forecast accuracy for most time-series, roughly 32.8% of them.

Moreover, the MASE distribution for each method was computed in the same way as in chapter *4.1* where level 3 time-series was analyzed. The boxplot of MASE-values for each forecasting method on level 1 is displayed in figure 21 below.

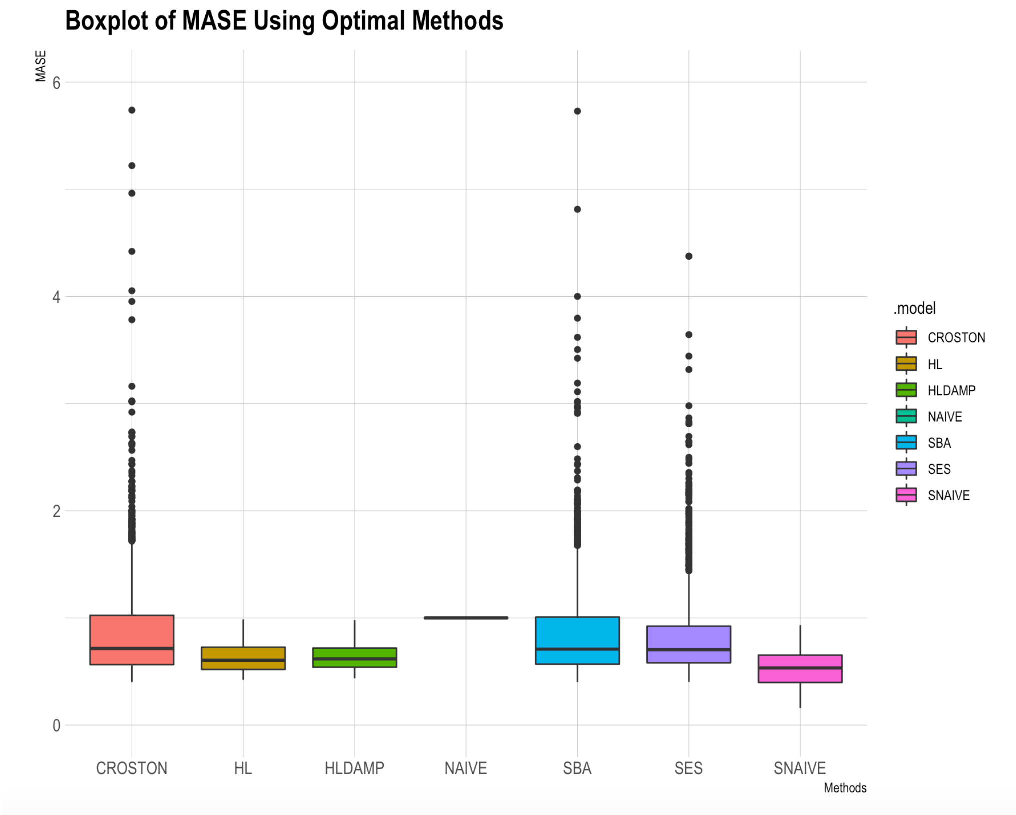

Figure 21. Box plot visualization of MASE distribution for all methods on level 1.

The boxplot does reveal some interesting findings. Croston, SBA and simple exponential smoothing is deemed to be the best method for an overwhelming majority of the time-series on level 1. However, the boxplot of these methods imply that they fail to forecast demand with high accuracy for a plurality of the time-series. This is evident due to the significant amount of outliers displayed in the plot. What this indicates is that the forecastability is generally lower for time-series on level 1 compared to level 3. Even if a method is deemed to be the best available it does not necessarily imply that it will do a good job predicting future demand. The mean MASE values as well as the spread around the mean MASE values are generally greater on level 1 compared to level 3. This further strengthens the notion that forecastability is lower on level 1 compared to level 3.

As stated in chapter *1.3*, Sandvik is currently using MA12 as the forecasting method for all of their 30 test SKUs in Voyager on a disaggregated level as well. Therefore, it is of high interest to compare forecast accuracy performance between MA12 and the identified best methods. What would the distribution of the MASE values look like if MA12 would have been used instead of the identified best methods? The boxplots are illustrated in figure 22 below.

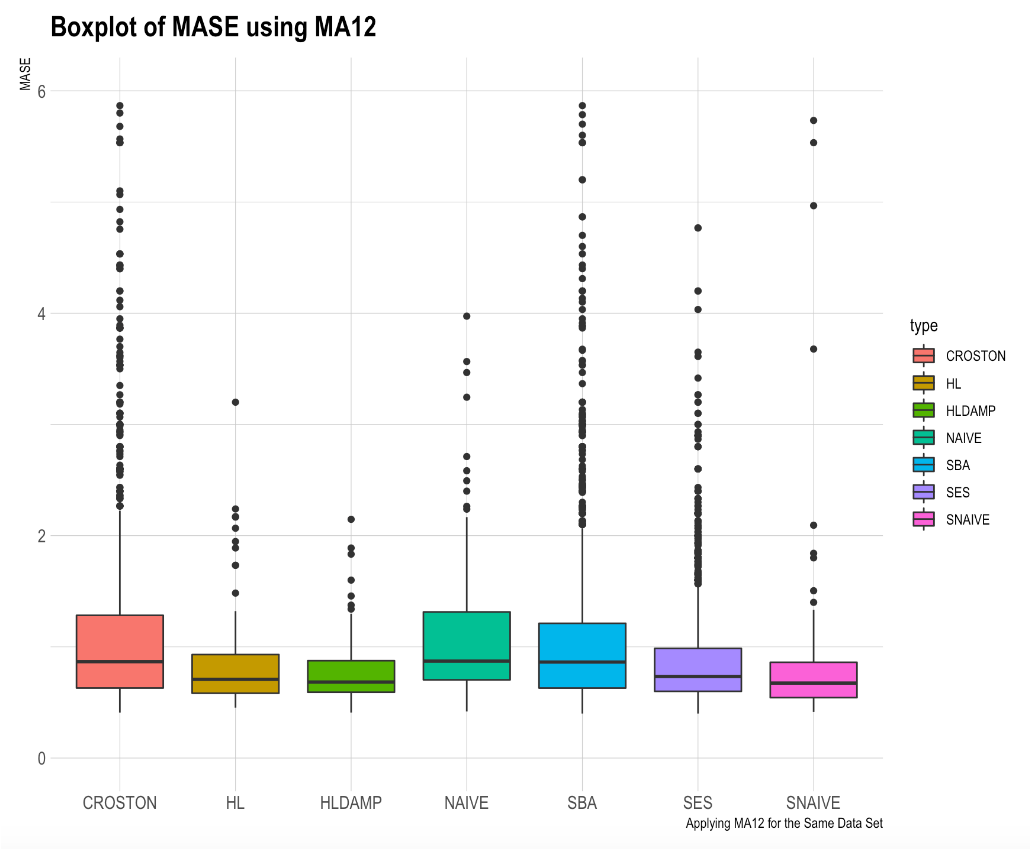

Figure 22. Box plot visualization of MASE distribution if MA12 would have been used instead of the identified best methods on level 1.

Similarly to the results on level 3, the box plot in figure 22 reveals that the MASE value for all methods, except the naive method, performs better on mean and median compared to if MA12 would have been used. Below in table 15, the mean and median value of the MASE distribution is displayed both for the best method and for MA12.

Table 15. Mean and median MASE values for best method and MA12 on level 1.

Best Method	**Number of Time-Series**	**Mean MASE Value Using Best Method**	**Mean MASE Value Using MA12**	**Mean MASE Value Improvement Using the Best Method**	**Median MASE Value Using Best Method**	**Median MASE Value Using MA12**	**Median Value Improvement Using the Best Method**
Croston	1775	0.87	1.20	+0.33	0.71	0.87	+0.16
SBA	1859	0.86	1.18	+0.32	0.71	0.87	+0.16
SES	1592	0.85	0.95	+0.10	0.71	0.74	+0.03
Holt's Linear	114	0.72	0.85	+0.13	0.62	0.71	+0.09
Holt's Linear Damped	96	0.71	0.79	+0.08	0.61	0.68	+0.07
Naive	113	1.00	0.98	-0.02	1.00	0.87	-0.13
Seasonal Naive	119	0.68	0.90	+0.22	0.52	0.67	+0.15
Total	**5668**						

4.3.2 Point Forecast Accuracy of Methods Across ABC-XYZ Classes on Level 1

The time-series on level 1 were also classified according to an ABC-XYZ-analysis. Two types of ABC-XYZ charts were created. The first chart displays the distribution of SKU/Stockroom/Customer Cluster time-series across the ABC-XYZ classes. The second chart displays the historic revenue distribution of these time-series across the ABC-XYZ classes.

The distribution of SKU/Stockroom/Customer Cluster time-series across the ABC-XYZ classes is illustrated in figure 23 below.

	X	Y	Z
A	0.58 %	4.66 %	10.72 %
B	0.16 %	2.39 %	17.29 %
C	0.25 %	4.90 %	59.05 %

Figure 23. The distribution of SKU/Stockroom/Customer Cluster time-series across the ABC-XYZ classes.

The distribution of the historic revenue across the ABC-XYZ classes is illustrated in figure 24 below.

	X	Y	Z
A	6.40 %	27.69 %	45.90 %
B	0.13 %	1.89 %	12.98 %
C	0.03 %	0.53 %	4.45 %

Figure 24. The distribution of the historic revenue generated from SKU/Stockroom/Customer Cluster time-series across the ABC-XYZ classes.

Figure 23 and 24 gives interesting insights into how skewed the distribution of revenue driving SKUs is. Merely 15.96% of the time-series are part of the A-class that is generating 80% of the total revenue. Moreover, 64.20% of the time-series are part of the C-class that only generates 5% of the total revenue. The insights that these numbers provide are valuable to have when evaluating the impact of improved point forecast accuracy for each class.

How the different forecasting methods performed within each ABC-XYZ class was also examined. An overview of the result for each class is presented in table 16 below.

Table 16. The two best forecasting methods for each ABC-XYZ class on level 1.

Classes	Number of SKU/Stockroom/Customer Cluster time-series	Best Method (% of SKU/Stockroom/Customer Cluster time-series)	Second Best Method (% of SKU/Stockroom/Customer Cluster time-series)
AX	33	Simple Exponential Smoothing (57 %)	Seasonal Naive (27 %)
AY	264	Simple Exponential Smoothing (36 %)	Seasonal Naive (15 %)
AZ	608	SBA (35 %)	Croston (30 %)
BX	9	Simple Exponential Smoothing (33 %)	Seasonal Naive (33 %)
BY	135	Simple Exponential Smoothing (38 %)	Naive (13 %)
BZ	980	SBA (35 %)	Croston (34 %)
CX	14	Holt's Damped Linear (29 %)	Seasonal Naive (29 %)
CY	278	Simple Exponential Smoothing (35 %)	Holt's Linear (16%)
CZ	3347	SBA (38 %)	Croston (36 %)
Total	**5668**		

The displayed results in table 16 convey that there are three methods that are noticeably prominent among classes on level 1; Simple exponential smoothing, Croston and SBA.

With the clear dominance from this trinity in mind, it should still be noted that seasonal naive is the second best method for class X. Within class X, simple exponential smoothing was deemed to be the best method for 47% of the time-series. Smoothing parameter α assumed values in a range between 0.00 and 0.26. Low demand fluctuations is a characteristic trait for the time-series in class X, hence it does not come as a surprise that simple exponential smoothing performs consistently well in this class.

Within class Y, simple exponential smoothing also yields the best forecast accuracy. It is the best method for 36 % of the time-series. The methods associated smoothing parameter α assumed values in a range between 0.00 and 0.46. Methods with any kind of trend are nowhere to be found in this class. The naive method seemed to perform quite well as it ended up as the runner-up method in all Y-classes.

Within class Z, an overwhelming majority of the time-series are to be found. The historic revenue from the time-series on level 1 is heavily concentrated in this class, mainly in class AZ. Croston and SBAs methods constituted the best method for 72% of the time-series. For Croston's method in class Z, the associated smoothing parameter α assumed values in a range between 0.00 and 0.42. Its associated smoothing parameter β assumed values in a range between 0.00 and 0.35. For SBAs method in class Z, the associated smoothing parameter α assumed values in a range between 0.00 and 0.41. It's associated smoothing parameter β assumed values in a range between 0.00 and 0.33. The fact that two methods suited for intermittent and lumpy time-series performed the best for class Z is very much in line with the expectations. In order to visualize the distribution of the MASE values for each class, a box plot was created. It can be viewed in figure 25 below.

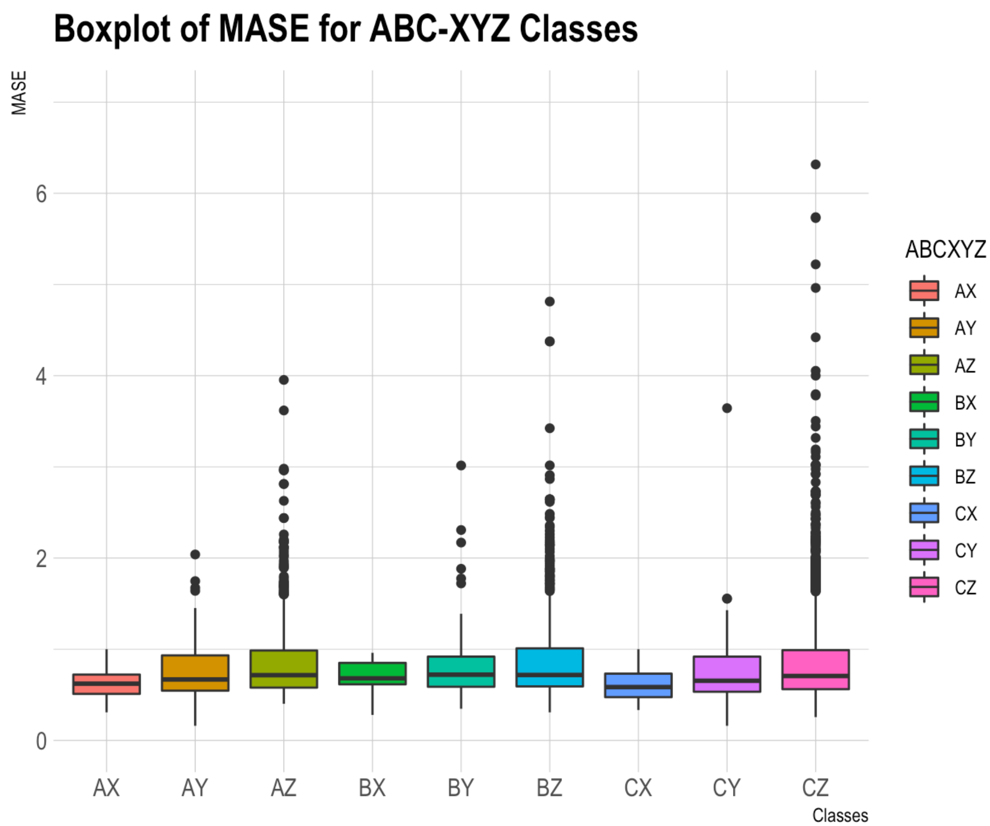

Figure 25. Box plot visualization of MASE distribution for all ABC-XYZ classes on level 1.

Some of the notions that emerged in chapter *4.1* are further cemented when looking at figure 25. Forecastability of a time-series, i.e the degree of variance it experiences, has a significant effect on the accuracy measurement outcome. It is clear that the time-series in class X are easier to forecast with high accuracy compared to time-series in class Z. The degree of forecastability for time-series in class Y is naturally in between class X and class Z.

4.3.3 Distributional Forecast Accuracy of ETS State-Space Models Across XYZ Classes on Level 1

A boxplot for the skill score on level 1 was also computed in order to visualize its distribution across the XYZ-classes. The boxplot can be viewed in figure 26 below.

Figure 26. Boxplot of Skill Score for XYZ classes on level 1.

It is interesting to see the discrepancy in the skill score distributions between the classes. Judging from figure 26 it looks like the median skill scores across the classes are similar. However, the significant increase of outliers as the coefficient of variation increases is hard to neglect. This could partly be explained by the fact that the majority of the time-series on level 1 fall into the Z-class. However, it is coherent with theory that skill scores are significantly worse for time-series that experience large variance. Time-series with high coefficient of variation yield distributional forecasts with wide tails. Therefore it seems reasonable that the accuracy of the distributional forecasts performs worse for these time-series. The interpretation of a bad skill score is that it reduces the reliability of the point forecast.

4.3.4 Benefits of Improved Forecast Accuracy on Level 1

Improved forecasting accuracy on level 1 enables regional planners at Sandvik to plan their stock levels in a more cost efficient manner. The one-step ahead forecasts provided on level 1 serves as an input variable when the regional planners determine sufficient reorder points in the supply chain. The provided ABC-XYZ classification of Sandviks SKU/Stockroom/Customer Cluster time-series provide a solid overview of which classes that have constituted the lion share of the revenue and can serve as a foundation when deciding strategies for stock keeping. Class AZ only makes up 10.72% of the time-series on level 1 but accounts for 45.90% of the total revenue. Therefore, it is of high importance to keep stock levels as low as possible without missing service level agreements towards the customer clusters. Croston and SBAs method was deemed to be the best method for 74% of the time-series in class AZ. For time-series where Croston's method was the best method, it yielded an average MASE value of 0.87. For the same time-series, Moving Average 12 yielded an average MASE value of 1.20. A similar discrepancy in forecasting accuracy was displayed when comparing SBA and Moving Average 12. Henceforth, with an implementation of Croston and SBAs method in class AZ, there is substantial potential in improving point forecasting accuracy. However, one should remember that distributional forecast accuracy for class Z is low. Naturally, this puts the reliability of the point forecast under a question mark. The takeaway is that an implementation of Croston and SBAs method on class AZ could improve Sandviks stock levels compared to forecasting class AZ with Moving

Average 12. However, one should be aware that there is an inherent uncertainty when forecasting time-series that experience high variance.

The most potential in improving relative forecasting accuracy is within the Z-classes where Croston and SBAs methods are prominent. The time-series in these categories have a coefficient of variation greater than 1, which implies difficulties in forecasting demand accurately. It is also within these classes where the vast majority of the time-series are allocated, mainly in the CZ category. However, the historical revenue within this class is, due to the C classification, significantly lower relative to the other categories. Therefore, the monetary impact of improving forecast accuracy within class CZ would be minimal even though most time-series are found here.

Under the condition that there is enough inventory capacity in the warehouses, it would be a good idea to stock up on the SKUs in class CZ to ensure high product availability. It is also convenient that these products, that hold a small share of the historical revenue, also are mostly small components that take up relatively small spaces. The takeaway is that forecasting improvements in less important classes don't have significant impacts on monetary gains. Hedging against demand uncertainties with excessive stock levels for these classes is a wise way to ensure high product availability towards the customer clusters to a relatively low holding cost.

4.4 Results of Hierarchical Forecasting

As mentioned, this study explores the possibility of using hierarchical forecasting. As such, the time-series that were appointed a SES method as the best fit on level 1 were chosen as samples for testing. The aim being to compare the individual forecasting methods in the two previous sections with the point forecasts produced by utilizing the various hierarchical forecasting methods.

The results of the different hierarchical forecasting approaches and the comparison with the independent forecasts are presented below in table 17**.**

Table 17: Mean MASE values of the hierarchical forecasting results compared to the independent approach.

Level	Independent SES	Top-Down	Bottom-up	Optimal Method
Level 3: SKU	**0.770**	**0.978**	**0.997**	**0.991**
Level 1: SKU/Stockroom/Customer Cluster	**0.850**	**0.871**	**0.851**	**0.870**

Looking at the results, all three hierarchical forecasting approaches using SES as the forecasting method performed similarly. There were no significant differences when comparing the average MASE values neither on level 3 nor 1. It is however apparent that all three performed significantly better on the bottom most disaggregated level.

The top-down approach yielded a slightly better forecasting accuracy on the top most aggregate level compared to the bottom-up and optimal method. The bottom-up approach performed better on the bottom level. This is an expected result as each respective approach should generate better forecasts within their starting levels compared to the other approaches. However, the fact that the top-down approach performed significantly worse on level 3 compared to level 1 was highly unexpected. Each method should according to literature produce better forecasting accuracy on their starting levels. More specifically, the bottom-up approach should generate forecasts with a higher accuracy on the bottom level compared to the aggregated forecasts on the top-level and likewise the top-down approach should perform better on the top level. This is not the case after evaluating the accuracy of the top-down approach.

The optimal method performed slightly worse than the top-down approach on level 3 and better than the bottom-up approach. On level 1 it performed worse than the bottom-up approach and almost the same as the top-down approach. The results indicated that the bottom-up approach produces slightly more accurate results than the other two approaches.

Comparing all three approaches with the independent forecasting with SES, they are outperformed on the top most aggregated level. On the bottom level however, they perform fairly similarly with the independent method having a slight edge. This shows that the independent forecasting method is generally a more reliable forecasting approach.

As the bottom-up approach produces forecasts on the bottom level, the average MASE value should correspond to the same as the independent, likewise the top-down approach should provide the same average MASE value on the top level. It is worth noting that the forecast accuracy on level 3 from the top-down approach does not at all correspond to the MASE value of the forecasts produced independently on the same level. This could mainly be due to the fact that the α-parameters are not optimized in the same way as they were when forecasting on each level independently. There was no optimization criterion in the hierarchical forecasting packages in *R* and the discrepancy in accuracy on level 3 between the top-down approach and the independent forecasting could be explained by this. The same goes for the small discrepancy between the average MASE values on level 1 between the bottom-up approach and the independent forecasting.

Another unexpected result from the hierarchical forecasting is the poor performance of the Minimum Trace Shrink optimal reconciliation method. According to literature, and most importantly Wickramasuriya et al. (2019), it is the method that generally performs the best out of the three and in some cases better than the independent forecasting methods, especially on the bottom levels. However our results show that it was in fact not the best overall method for the time-series at our disposal.

5. Conclusions & Recommendations

In this chapter, the conclusions of the study will be presented with respect to the research purpose presented in chapter 1.4.2. Following the conclusions, the final recommendations of the study will be presented for the case company.

5.1 Conclusions

The purpose of the study was to identify and propose quantitative forecasting methods with the aim to improve forecasting accuracy on a SKU and SKU/Stockroom/Customer Cluster level at Sandvik. This objective was approached by analyzing historical demand data provided by Sandvik. The study resulted in an ABC-XYZ classification where an optimal method was determined for each class.

Sanders' forecasting process constituted the foundation for the master thesis. The five steps in the process mapped out the necessary activities to help improve the forecasting process. The forecasting methods included in the thesis were chosen based on frequency of appearance in relevant literature. The chosen methods included the Naive, Seasonal Naive, Simple Exponential Smoothing, Holt's Linear, Holt's Linear Damped, SBA and Croston's method. These methods were deemed to be comprehensive enough to cover the varying demand patterns that the time-series displayed.

Both point- and distributional forecast accuracy was measured in the study. Literature provided meaningful insights when sufficient accuracy measurements were determined. For the point forecast measurement the choice fell on MASE. For the distributional forecast measurement the choice fell on Skill Scores. The common theme between the two measurements is the scaled nature which is deemed appropriate when comparing forecasting accuracy between time-series.

The results showed that all the chosen forecasting methods with the exception of the Naive method improved point forecast accuracy on both level 1 and 3 compared to Moving Average 12. The results also indicated that the distributional forecast accuracy is best for class X and worst for class Z. In other words, the reliability of the point forecasts decrease as the forecastability of time-series decreases.

The time-series on both level 1 and 3 were classified according to an ABC-XYZ classification. The rationale behind the classification was to differentiate the time-series based on value impact (ABC) and forecastability (XYZ). The simplicity of performing this analysis makes for an appropriate method for classifying the time-series. As such, this was chosen as the approach for cataloging each time-series and assigning a suitable forecasting method for each class.

The results from chapter 4 showed that Simple Exponential Smoothing was the best method for the classes AX, AY, BX, BY, CX and CY on both level 1 and 3 with the exception of class CX on level 1. For classes AZ, BZ and CZ, Croston's and SBA were the most prominent forecasting methods on both level 1 and 3.

The results from the hierarchical forecasting showed no improvement in forecasting accuracy compared to forecasting on level 1 and 3 independently. Therefore, the three approaches were not deemed fit for the case company.

To conclude, Croston's, SBA and Simple Exponential Smoothing were deemed the most appropriate forecasting methods for the time-series on both level 1 and 3 and will thus be recommended to the case company.

5.2 Recommendations

The study resulted in a couple of recommendations. The first pertaining to the measurement of accuracy for the chosen methods. MASE is the accuracy measurement that should be used when evaluating the forecasts. Its scaled nature makes it appropriate when measuring accuracy performance for a whole ABC-XYZ class. We do not recommend that Sandvik continues using MAPE and/or SMAPE as their main accuracy measurements. These methods are not suitable for intermittent time-series which accounts for a plurality of the time-series.

Another recommendation is that Sandvik should classify their time-series according to an ABC-XYZ classification. Differentiation of the time-series in this manner provides a better overview of value and forecastability.

The method that yielded the best forecast accuracy for each ABC-XYZ class was determined. We recommend that Simple Exponential Smoothing is appointed as the forecasting method for classes AX, AY, BX, BY and CY on both level 1 and 3. In addition, it was also the best method for class CX on level 3. Furthermore, we recommend that SBA and Croston's methods are appointed as forecasting methods for classes AZ, BZ and CZ.

As new demand arrives in the future, the demand characteristics for time-series can change and hence current ABC-XYZ classifications become obsolete. To be able detect changes in demand characteristics, it is necessary to monitor the forecasts with a tracking signal. When the forecast error deviates from the tolerable ratios, the practitioner should be notified and examine if the current forecasting method is feasible or not. Should a time-series deviate from the tolerable ratios, it should then be re-classified according to the ABC-XYZ classification described in section 3.1.3.1.

References

Hua, Z. S. Zhang, B. Yang, J. and Tan, D. S. (2007). *A New Approach of Forecasting Intermittent Demand for Spare Parts Inventories in the Process Industries.* The Journal of the Operational Research Society. Vol. 58, No. 1 (Jan., 2007), pp. 52-61 (10 pages). http://www.jstor.org/stable/4622667

Pinçe, Ç. Turrini, L. and Meissner, J. (2021). *Intermittent demand forecasting for spare parts: A Critical review.* Omega. Volume 105, December 2021, 102513. https://www-sciencedirect-com.ludwig.lub.lu.se/science/article/pii/S0305048321001225?via%3Dihub

Johnston, F. R. and Boylan, J. E. (1996). *Forecasting intermittent demand: A comparative evaluation of croston's method. Comment.* International Journal of Forecasting, 1996, vol. 12, issue 2, 297-298 https://www.sciencedirect.com/science/article/abs/pii/0169207095006427

Syntetos, M. Boylan, J. and Croston, J. D. (2005). *On the categorization of demand patterns.* Journal of the Operational Research Society 56(5). DOI: https://doi.org/10.1057/palgrave.jors.2601841

Costantino, F. Gravio, G. D. Patriarca, R. and Petrella, L. (2018). Spare parts management for irregular demand items. Omega Volume 81, December 2018, Pages 57-66. https://doi.org/10.1016/j.omega.2017.09.009

Axsäter, S. (2006). *Inventory Control, 2nd edition*. Springer.

Saunders, M. Lewis. and P. Thornhill, A. (2007). *Research Methods for Business Students.* 4th edition. Pearson Education. https://www-vlebooks-com.ludwig.lub.lu.se/vleweb/Product/Index/2026340?page=0

Kim, S.M. (2021). *Inductive or deductive? Research by maxillofacial surgeons.* Journal of the Korean Association of Oral and Maxillofacial Surgeons. 2021; 47(3): 151~152.

Hyndman, R. J. Gooijer, J. G. (2006). *25 years of time series forecasting.* International Journal of Forecasting. Volume 22, Issue 3, 2006, Pages 443-473. https://doi.org/10.1016/j.ijforecast.2006.01.001

Eisenhardt, K. M. (1989). *Building theories from case study research.* Academy of Management Review. Vol. 14 Issue 4, p532-550.

Olhager, J. (2000). *Produktionsekonomi.* Studentlitteratur AB.

Holt, Charles C. (1957). *Forecasting seasonals and trends by exponentially weighted moving averages.* International Journal of Forecasting. Volume 20, Issue 1, January–March 2004, Pages 5-10. https://doi.org/10.1016/j.ijforecast.2003.09.015

Yin, Robert K. (2008). Case Study Research. Design and Methods. Fourth edition. Applied Social Research Methods Series Volume 5. https://media.oiipdf.com/pdf/c0090627-41e0-4d8a-b497-f29adb6af69d.pdf

Mukherjee, K. Sanguri, K. (2021). *Forecasting of intermittent demands under the risk of inventory obsolescence.* Journal of Forecasting. Volume 40, Issue 6 September 2021. https://onlinelibrary-wiley-com.ludwig.lub.lu.se/doi/full/10.1002/for.2761

Doszyn, M. (2020). *Accuracy of Intermittent Demand Forecasting Systems in the Enterprise.* European Research Studies Journal, European Research Studies Journal, vol. 0(4), pages 912-930. https://ideas.repec.org/a/ers/journl/vxxiiiy2020i4p912-930.html

Hyndman, R. J. Koehler, A. B. and Snyder, R. D. (2002). *A State Space Framework for Automatic Forecasting Using Exponential Smoothing Methods.* International Journal of Forecasting 18(3):439-454. DOI: 10.1016/S0169-2070(01)00110-8

Hyndman, R. J. Koehler, A. B. (2006). *Another look at measures of forecast accuracy.* International Journal of Forecasting. Volume 22, Issue 4, October–December 2006, Pages 679-688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Gneiting, T. Katzfuss, M. (2014). *Probabilistic Forecasting.* Annual Review of Statistics and Its Application. Vol. 1:125-151. https://doi.org/10.1146/annurev-statistics-062713-085831

Hyndman, R. J. Ahmed, A. Athanasopoulos, G. and Shang, H. Lin. (2011). *Optimal combination forecasts for hierarchical time series.* Computational Statistics & Data Analysis. Volume 55, Issue 9, 1 September 2011, Pages 2579-2589. https://www-sciencedirect-com.ludwig.lub.lu.se/science/article/pii/S0167947311000971?via%3Dihub

Athanasopoulos, G. Gamakumara, P. Panagiotelis, A. Hyndman, R. J. & Affan, M. (2019). *Hierarchical Forecasting.* Department of Econometrics and Business Statistics Working Paper 02/19. . https://www.monash.edu/business/ebs/our-research/publications/ebs/wp02-2019.pdf?fbclid=IwAR3rogBluZCXSLduN_d3zN-Z5t5B87bAnaSmCzMS71NHV0KdVV0TzL0Tb1g

Heinecke, G. Syntetos, A.A. Wang, W. (2011). *Forecasting-based SKU classification.* International Journal of Production Economics. Volume 143, Issue 2, June 2013, Pages 455-462. https://www-sciencedirect-com.ludwig.lub.lu.se/science/article/pii/S092552731100483X?via%3Dihub

Stojanović, M. Regodić, D. (2017). *The Significance of the Integrated Multicriteria ABC-XYZ Method for the Inventory Management Process.* Acta Polytechnica Hungarica 14(5):29-48.

DOI: 10.12700/APH.14.5.2017.5.3

Buliński, J., Waszkiewicz, C., & Buraczewski, P. (2013)*Utilization of ABC/XYZ analysis in stock planning in the enterprise*. Annals of Warsaw University of Life Sciences – SGGW Agriculture No 61 (Agricultural and Forest Engineering) pp. 89-96.

Winters, P. R. (1960). F*orecasting sales by exponentially weighted moving averages*. Management Science. Vol. 6, No. 3, pp. 324-342. https://www.jstor.org/stable/2627346

Croston, J.D. (1972) *Forecasting and Stock Control for Intermittent Demands*. Operational Research Quarterly, 23, 289-303. http://dx.doi.org/10.2307/3007885

Gross, C. W., & Sohl, J. E. (1990). *Disaggregation methods to expedite product line forecasting*. Journal of Forecasting, 9, 233–254.

Mircetic, D. Rostami-Tabar, B. Nikolicic, S. and Maslaric, M. (2021). *Forecasting hierarchical time series in supply chains: an empirical investigation.* International Journal of Production Research. DOI: https://doi.org/10.1080/00207543.2021.1896817

Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J. (2019). *Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization*. Journal of the American Statistical Association, 114(526), 804–819. https://doi.org/10.1080/01621459.2018.1448825

Armstrong, J. S. (2002). *Principles of Forecasting. A Handbook for Researchers and Practitioners.* Springer.

Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia.
https://otexts.com/fpp2/

Hyndman, R.J. and Athanasopoulos, G. (2021) *Forecasting: principles and practice, 3rd edition*. OTexts: Melbourne, Australia. https://OTexts.com/fpp3

D'Alessandro, A.J. and Baveja, A., 2000. Divide and conquer: Rohm and Haas' response to a
changing specialty chemicals market. Interfaces, 30(6), pp.1-16.

Syntetos, Aris A. Babai, Z. Boylan, John E. Kolassa, S. and Nikolopoulos, K. (2016). *Supply chain forecasting: Theory, practice, their gap and the future.* European Journal of Operational Research.
Volume 252, Issue 1, 1 July 2016, Pages 1-26.
https://www-sciencedirect-com.ludwig.lub.lu.se/science/article/pii/S0377221715010231?via%3Dihub

Björklund, M. and Paulsson, U. (2012). *Seminarieboken: att skriva, presentera och opponera.* Studentlitteratur AB.

Krajewski, L. J. Malhotra, M. J. and Ritzman, L. P. (2013). *Operations Management. PROCESSES AND SUPPLY CHAINS.* Twelfth Edition. Pearson.