A Comparative Study on Spare Part Demand Forecasting Methods

Thesis by

İrem Kasapoğlu (638151)

Supervisor

Prof. dr. ir. R. Dekker

Second Assessor

Prof. dr. Martijn G. de Jong

Erasmus School of Economics

Erasmus University of Rotterdam

Rotterdam, the Netherlands

ABSTRACT

Benchmarking Spare Part Demand Forecasting Methods
İrem Kasapoğlu

Spare part demand forecasting presents unique challenges due to intermittent demand patterns and the high costs associated with inventory management. This study aims to enhance forecasting accuracy and inventory control performance by evaluating diverse forecasting methods and investigating the impact of outlier detection and handling procedures as a part of data preprocessing on both artificial and real data sets, categorized based on inherent data characteristics. Nine forecasting methods, including parametric, bootstrapping, and machine learning approaches, are assessed: Croston's method, Simple Exponential Smoothing (SES), Syntetos-Boylan approximation (SBA), the Teunter-Syntetos-Babai method (TSB), Willemain's bootstrapping method (WSS), Multi-Layer-Perceptron (MLP), LightGBM, Random Forest (RF), and Support Vector Regression (SVR). Outliers are detected using the Local Outlier Factor (LOF) and mean imputation is employed for outlier handling. Results indicate that the Syntetos-Boylan approximation (SBA) achieves the highest forecasting accuracy overall, while there's no consistent superior performance of any method in stock control management over all data sets. The study reveals that outlier detection and handling procedures enhance forecasting accuracy for most methods, but have no significant impact on inventory control performance. Additionally, the study emphasizes the importance of considering total execution duration and required expertise when selecting a method. Overall, the research highlights the influential role of outlier detection and handling as data preprocessing steps, as well as the impact of data characteristics on forecasting method performance.

Contents

Abstract **2**

1 Introduction **5**

2 Literature Review **8**

2.1 Recent Literature on Spare Part Demand Forecasting 8

2.1.1 Time Series Methods 9

2.1.2 Contextual Methods 18

2.1.3 Comparative Studies 21

2.2 Recent Literature on Outlier Detection Methods 25

3 Methodology **31**

3.1 Experimental Design 31

3.2 Selected Forecasting Methods 32

3.3 Utilized Forecasting Accuracy Measures 36

3.4 Utilized Inventory Control Measures 37

3.5 Utilized Outlier Detection Method 38

3.6 Demand Pattern Identification and Data Splitting Rule 39

4 Data Description and Classification **41**

4.1 Industrial Data Sets 41

4.2 Artificial Data Sets 43

4.3 Data Pre-processing 44

4.4 Classification of Data Sets 46

5 Results **48**

5.1 Exemplary Results of Forecasting Methods 48

5.2 Exemplary Results of Performance Measures 53

5.3 Main Results 55

6 Conclusion and Discussion **66**

6.1 Findings 66

6.2 Comparison of the Findings with the Existing Body of Literature 69

6.3 Conclusion 71

6.4 Discussion 72

Bibliography **73**

Appendix **83**

Chapter 1

Introduction

In the field of supply chain management, most companies require demand forecasts as they can represent a competitive advantage in the decision-making process. Given that precise forecasts serve as inputs for various functions, including production, sourcing, inventory planning, and supply chain planning, their accuracy is of utmost importance (Ivanov et al., 2017). Precise forecasting guarantees to have a sufficient amount of products to meet customer demand without holding excessive inventory that could result in tied-up cash for businesses. However, the quest for reliable demand forecasting remains one of the biggest challenges in supply chain management since observed inventory holding costs can vary with a range of 5-45% in the industry and with an average of 25% (Durlinger and Paul, 2012). Therefore, there is still a need to develop dependable demand forecasting models to make better and more accurate predictions.

Spare parts are essential components that are used to replace worn-out or damaged parts, ensuring that equipment and systems continue to operate in their intended way. Certain industries, such as aerospace, automotive, manufacturing, commercial aviation, and military, maintain a wide array of spare parts in inventory, which has notable consequences for both availability and inventory holding. The management of these parts is thus a crucial task (Boylan and Syntetos, 2008). And the importance of spare parts lies in their ability to minimize downtime, reduce maintenance costs, and increase the longevity of equipment.

Spare part demand forecasting is the process of predicting the future demand for replacement parts that are required to maintain or repair equipment, machines, and products. Accurate forecasting of spare parts demand is essential for organizations to ensure the availability of parts when needed, minimize inventory holding costs, and avoid stock shortages that could disrupt operations. Specifically, if predictions considerably exceed actual demand, it will lead to producing or stocking too many products that cannot be sold and, in the end, lead to increased costs and tied-up capital. On the other hand, lower service levels caused by longer lead times result in prolonged downtime which can be very costly as well as loss of business opportunities when forecasts fall short of actual demand. Thus, managing demand uncertainty has always been a major concern in manufacturing planning and control systems. Accordingly, diverse methods and techniques have been introduced to address the unpredictability of demand. (Bartezzaghi et al., 1999).

Spare parts are frequently characterized by intermittent demand patterns, indicating that the demand for such parts is marked by extended periods of zero demand and being infrequent, also the demand sizes may vary greatly (Wang and Syntetos, 2011). Thus, spare part demand requires specialized methods. In addition, items with intermittent demand can prove to be highly costly for businesses, suggesting that such parts may account for as much as 60 percent of the total investment in-stock inventory, according to Johnston et al. (2003). Therefore, the evolution of spare part demand forecasting methods continues to advance by the need of organizations in managing these items, improving accuracy, reducing costs, and enhancing operational efficiency. As technology continues to advance and data availability continues to increase, newer spare part demand forecasting methods and modifications to the existing ones emerge to meet the specific needs of the organizations. However, forecasting methods with variable performance results across different data sets or under different conditions have been reported. Therefore the first research question will attempt to answer better why and when a particular forecasting method performs better.

As an alternative strategy to improve forecasting performances, data pre-processing (e.g. outlier detection, handling missing values) may lead to an improvement in forecasting performance. It is difficult to accurately identify abnormal demand and correct it due to the intermittent and highly variable demand occurrences of spare parts. These occurrences may arise due to the high demand for either preventive maintenance which is performed regularly and in high amounts or corrective maintenance which is unplanned, both of which result in significant spikes in demand and should be consequently labeled as outliers. Both the presence of missing values and outliers in demand can lead to a decrease in the accuracy of a forecast. Although there are various academic papers on outlier detection methods, a comprehensive study is lacking that utilizes statistical and/or machine learning methods for outlier detection in the spare parts demand forecasting field, how an outlier detection method is conducted as well as how it affects the forecasting performance. Thus, this study aims to address this gap with the help of the second research question. With all these taken into consideration, the main questions of this research can be listed as:

"*Which spare part demand forecasting method is best for which type of data?*" and

"*Does the performance of spare part demand forecasting method depend on the data pre-processing*?"

The steps that are planned to be taken in this paper are as follows. Firstly relevant and comparative spare part demand forecasting studies will be reviewed to answer which method performs better and under which circumstances it does so as well as to identify potentially good methods. Secondly, an outlier detection method as a data pre-processing step will be applied to the selected data set(s). At the end of this step, there will be first raw data sets to which no outlier detection method is applied and second pre-processed data sets. After the pre-processing, all the data sets will be characterized to identify similarities or differences according to a classification framework. As the last step, selected forecasting methods will be applied on all data sets. Next, a comparative analysis will be made in terms of the methods' performance results. In addition, two different versions of a data set will be compared to answer whether data pre-processing methods affect the forecasting method performance.

Chapter 2

Literature Review

2.1 Recent Literature on Spare Part Demand Forecasting

The scholarly framework concerning the forecasting of spare part demand can be classified into three primary branches, accompanied by multiple subcategories, which demonstrate the advancement and progression of the discipline over time. The three main branches are as follows: time series methods, contextual forecasting methods, and comparative studies which provide performance benchmarks for various methods. The aforementioned literature framework is based on Pinçe et al. (2021).

First of all, time series methods analyze historical data to identify patterns and trends inherent in the data. In practice, time series forecasting methods are widely used because of their ease of usage. However, they do not consider external factors such as expert opinions, product features, or maintenance information to determine the drivers of spare parts demand. The time series methods can be subsumed under three branches, parametric methods as the first one assume a specific mathematical distribution for the data and use this assumption to estimate the future demand, whereas non-parametric methods do not make any assumptions about the underlying distribution of the data and instead rely on the historical data itself to make predictions. As the third branch, forecast improvement strategies focus on categorizing demand data by identifying the underlying demand characteristics (data classification) or grouping data with similar demand patterns (data aggregation). As the second major category, contextual methods employ both contextual information (equipment life, maintenance scheduling, working conditions, etc.) and statistical techniques. It can be divided very broadly into installed base forecasting which incorporates installed base information and judgemental forecasting which incorporates forecasters' opinions and experience into the prediction process. Lastly, the third major category from this framework is the comparative studies that benchmark the performance of methods by conducting comprehensive empirical evaluations using either industrial or simulated data sets and assessing the methods' performance based on either forecast accuracy or inventory performance measures. In the following sections, these three main categories outlined in the literature review framework on spare part demand forecasting by Pinçe et al. (2021) will be discussed in detail. These categories include time series methods, contextual methods, and comparative studies, respectively.

2.1.1 Time Series Methods

Parametric Approaches

The initial major branch within time series forecasting methods is known as the parametric approach. These approaches refer to a collection of methods that depend on a distributional assumption of demand lead time and the determination of the mean and variance of demand via the utilization of a suitable forecasting procedure. (Syntetos et al., 2011). Following the establishment of the demand distributional assumption, the determination of the inventory system's decision parameters (such as the re-order point or the order-up-to-level) is performed by utilizing the demand and variance estimates. The main goal of these methods is to forecast future demand by taking all these into consideration. Many parametric approach models have been developed to forecast spare parts demand. Error-trend-seasonality (ETS) framework covers different exponential smoothing methods such as single exponential smoothing (SES) and Holt-Winter's method, as it uses trend and seasonality components to compute the error. However, ETS framework models fail to perform well when there is more than one seasonality (Naim et al., 2018), and computationally heavy when there is large-scale time series data (Qi et al., 2022). SES from the ETS framework is the first time series forecasting model. However, since it is designed for continuous demand, the predictions yield inaccurate results when there are long zero-demand periods, such as intermittent or erratic demand. Hence ad-hoc parametric approaches have been developed to solve this problem. Croston (1972) was the first one to investigate the problem, as a result, he developed a model that is called *Croston* or *Croston's method* which estimates the non-zero demand and demand interval with exponential smoothing respectively, resulting in more accurate results with lower safety stock levels. As a modification to Croston's method, Syntetos and Boylan (2001) showed that Croston's method is positively biased and developed a method called *Syntetos-Boylan Approximation (SBA)* by defining a bias correction coefficient Syntetos et al. (2005).

In the field of spare part demand forecasting, there is another important problem called demand obsolescence. It means that demand levels for a spare part gradually or suddenly decrease and reduce to zero over time. Teunter et al. (2011) point out that Croston and its modifications are slow to adjust to new demand levels when demand gradually decreases or suddenly becomes zero. Hence a new method referred to as *TSB* has been developed by Teunter et al. (2011) to address the issue of demand obsolescence. The difference between the new method and Croston and SBA is that demand probability forecasts are used in combination with demand size forecasts instead of demand interval forecasts, additionally, the demand probability is updated in every period (Pinçe et al., 2021). Since demand probability is updated every period, when there is no demand it leads to a downward adjustment in forecast results, thus making it easier to detect obsolescence. When scenarios with stationary demand, linearly decreasing demand, and sudden obsolescence are observed, TSB outperforms other methods (SES, Croston, and SBA) in both bias and variance. However, Babai et al. (2014) noted in an empirical paper that TSB does not yield significantly more accurate results in comparison to other methods when there is obsolescence. Taking this into consideration, building on TSB, Babai et al. (2019) developed a new method called *modified SBA* or *mSBA* to address the same issue, demand obsolescence. This new method can be defined as a combination of two methods, SBA and TSB. When there is demand obsolescence, the forecast updates yield similar results to TSB, on the other hand, when there is positive demand, the updates yield similar results to SBA. Thus they noted that the modified SBA outperforms all other methods aforementioned in terms of mean error(bias) and mean absolute error (Babai et al., 2019). As an alternative approach to address demand obsolescence, SD et al. (2014), developed a new method referred to as *hyperbolic exponential smoothing (HES)*. When there is zero demand, forecast updates are adjusted downward similar to TSB, however, the updates decay hyperbolically instead of exponentially. Though SD et al. (2014) also show that the TSB method yields more accurate results when HES is easier to apply in practical situations.

There are some dynamic time series models that have been developed. Pennings et al. (2017) introduce a new model that covers the positive cross-correlation between interarrival times and demand sizes in order to predict future spare parts demand. They compare their newly introduced method with traditional methods by utilizing industrial data sets, although

SBA outperforms all other methods including the new method in terms of forecast accuracy, the proposed method outperforms all the traditional ones in terms of inventory measures if a strong positive cross-correlation between the interarrival time and the demand size is observed. Secondly, Snyder et al. (2012) introduced a new method that models the demand distribution means with several different distributions (such as Poisson, negative binomial, and hurdle-shifted Poisson) to incorporate the potential random changes in the mean of the demand distribution. The mean gets updated by smoothing the previous period's mean and the actual values of the demand. Although these dynamic methods show gain in terms of accuracy, they require advanced statistics skills and thus are less intuitive when compared to traditional methods (such as Croston, SBA, etc.). After Snyder et al. (2012), Jiang et al. (2020) introduced a new technique for forecasting by fitting a mixed zero-truncated Poisson hurdle model that takes into account variations in the underlying demand process. By analyzing data from an electric power company, they found that their method yielded better forecast accuracy results compared to other techniques such as Croston, the Poisson model, etc. However, more research needs to be conducted for a better performance assessment of both methods, Snyder et al. (2012) and Jiang et al. (2020).

As the last branch of parametric approaches, parametric bootstrapping is a simulation technique that allows the generation of new samples by simulating data from a probability distribution that is assumed to represent the population. It is a powerful resampling method that can estimate various distribution parameters. In parametric bootstrapping, a model is first fitted to the original data set, and the estimated parameters of the model are then used to simulate new samples to fill the gaps. The simulated samples are drawn from the assumed probability distribution, using the estimated parameters as the input. By repeating this process many times, a large number of new samples can be generated, and statistical properties of interest can be estimated from these samples. Snyder (2002) proposes two different parametric bootstrapping methods, namely the log-space adaptation (LOG) and the adaptive variance version (AVAR). These methods leverage demand history to compute least-squares estimates of the mean and standard deviation of demand, as well as the smoothing parameter of SES. According to Bookbinder and Lordahl (1989), as the bootstrap is a "distribution-free" approach, it is more likely to capture the skewness of intermittent lead-time demand distributions effectively than its parametric counterpart. Thus, as one of the initial bootstrapping methods they introduce a new model which can be viewed as an extension of Efron's bootstrap method, featuring two additional steps, first estimating the mean and standard deviation of the bootstrap samples' lead-time demand values, and second deriving a theoretical density function using the empirical mean and standard deviation of these values. In general, the bootstrap technique provides good service levels at a lower cost, except when the simulated distribution has a standard shape and a positive skewness (Hasni et al., 2019).

Non-Parametric Approaches

Non-parametric approaches are the second major branch under time series forecasting methods. When the lead-time demand can not be characterized by a parametric distribution, parametric methods may underperform. In addition, if the distribution is assumed incorrectly, it would lead to a misleading estimation of the demand (Hasni et al., 2019). As a result, these non-parametric approach methods are developed as an alternative to parametric methods. Although non-parametric models require more computational power than parametric models, they make fewer assumptions regarding the time series data and they solely retain the empirical structure of the observed time series data (Gautam and Singh, 2020).

Viswanathan and Zhou (2008) introduce a non-parametric bootstrapping method, referred to as *VZ*, which involves creating a lead-time demand distribution by sampling demand sizes and demand intervals separately and in a progressive manner. If the demand is moderately intermittent and the historical data is limited, the VZ method outperforms other methods (SBA and WSS) only when the lead time is long (Hasni et al., 2019). Next, Porras and Dekker (2008) propose another method named *the empirical method*. The empirical method is a simpler non-parametric approach that utilizes the empirical cumulative distribution function to estimate the demand distribution for fixed lead times. As an extension of the empirical method, Van Wingerden et al. (2014) also propose covering random lead times. Based on an empirical study with different industrial data sets, the proposed model yields slightly better inventory performance than that of the empirical method. Similarly, the empirical method is extended by Zhu et al. (2017) in which the lead time demand distribution is constructed using historical demand data by employing extreme value theory to model the tail of the distribution. Their study shows that by conducting both empirical research and simulations, the proposed approach results in reduced expected waiting times, increased cycle service levels, and improved target service levels in comparison to the empirical method. Lastly, one of the most popular non-parametric approaches is developed by Willemain et al. (2004) and can be referred to as Willemain's bootstrapping method, *WSS* in short. They introduce a modified bootstrap method that addresses three challenging characteristics of intermittent demand: autocorrelation, frequent repetition of values, and short time series (Willemain et al., 2004). It resamples based on historical data by using a Markov chain approach that alternates between periods of zero demand and periods of demand equal to 1. A jittering procedure is then applied to create demand sizes by modifying the sampled demand values with random variation, overall with the aim of reducing the irregularities in the data. Through an extensive empirical study by the authors using nine industrial data sets, their results indicate that their method leads to higher accuracy in the demand distribution with a fixed lead time when compared to both exponential smoothing and Croston. In addition, one of the main advantages of the bootstrapping method proposed is its capability to generate demand values that have not been observed historically and it does not require any assumptions on the distribution of data (Syntetos et al., 2015).

Machine Learning Models

Another subcategory of non-parametric methods is machine learning (ML) models. Traditional time-series models may fail to capture the non-linear patterns in data. Therefore, ML modeling can be used as an alternative to address these issues. ML techniques that are used in the field of spare part demand forecasting belong to the supervised learning methods category, which learns demand patterns directly from the data itself. Under the subcategory of ML models, neural network models have been widely used in the literature. Gutierrez et al. (2008), conducted one of the earliest studies in this area, in which they utilized an electronic product distributor's data to compare their neural network model with SES, Croston, and SBA. They found that the neural network model generally produced more accurate forecasts than the other methods unless there was a significant difference between the average demand sizes of the training and test data sets. Next, Mukhopadhyay et al. (2012) used the same data to compare a modified version of the method described in Gutierrez et al. (2008) study with traditional methods, including SES, Croston, SBA, and weighted moving average. Their analysis shows that the modified neural network method outperforms the others in terms of the MAPE and the median relative absolute error (MdRAE). In a generalized version of Gutierrez et al. (2008) by Kourentzes (2013), the proposed method incorporates three different network settings and the

Levenberg-Marquardt algorithm (Marquardt, 1963) to improve the training speed. In contrast to the findings of Gutierrez et al. (2008), Kourentzes' study shows that the neural network method yields worse results regarding accuracy measures, such as ME and MAE, but better in terms of service levels than the Croston-type methods. In their study, Lolli et al. (2017) introduce an *extreme learning machine*, a neural network with a simpler and faster learning algorithm. They evaluate its performance by comparing it with the aforementioned methods using an automotive data set. Their findings indicate that although the aforementioned neural network models using back-propagation perform better than the extreme learning machine in terms of MAPE, the proposed model is easier and faster to implement. Next, Guo et al. (2017) merge a genetic neural network model with three exponential smoothing variants and a hierarchical forecasting method. They evaluate the effectiveness of this combined method using aircraft spare parts data and demonstrate that it generates superior results in terms of accuracy than the forecasts of each method. On top of Pinçe et al. (2021)'s framework, recently Wei et al. (2022) propose an improved sparrow search algorithm optimized back-propagation (CGSSA-BP) neural network method to improve the accuracy and stability for intermittent spare part demand. The authors conduct a comparison study with sparrow search algorithm (SSA)-BP neural network and traditional BP neural network models, the results show that the newly introduced model outperforms other neural network models.

Forecast Improvement Strategies

Forecast improvement strategies are the last major category under time series forecasting methods. Different estimation approaches are necessary to account for the varying demand characteristics of spare parts. It can be beneficial to comprehend these unique features as the performance of a particular method may depend on them. In this section, methods that will be covered are divided into two main sections, demand classification, and data aggregation methods. Papers that belong to the demand classification subcategory aim to identify a forecasting technique that yields the best performance for a given demand characteristic, on the other hand, papers that focus on improving forecasts through data aggregation aim to decrease the variability in spare parts demand that can arise from extended periods of zero demand and highly variable demand sizes.

Demand classification methods

Williams (1984) conducted one of the initial studies on demand classification, wherein by breaking down the variance in lead-time demand, demand is categorized as sporadic, slow-moving, or smooth, and subsequently, a specific distribution is used to estimate reorder levels for each demand category. Johnston and Boylan (1996) propose a forecasting technique that is based on estimates of the mean and variance of demand size, along with the average intervals between demand periods, as a means of providing evidence for Willemain et al. (1994)'s assumption on the Croston's method not being a significant improvement over SES when data exhibits either too little or too much intermittency. Also, they establish a parameter called average inter-demand interval (p). After that, Syntetos and Boylan (2005) not only define an additional parameter for demand classification, the squared coefficient of variation of demand (CV^2) but also carry out a comparative analysis of the mean squared errors (MSEs) for various methods, and use this information to determine threshold values for p and CV^2. Subsequently, these established threshold values are used as a classification framework based on two dimensions of demand in order to identify the forecasting method that provides the highest level of accuracy. Kostenko and Hyndman (2006) criticize the validity of this framework's cutoff values and introduce a different cutoff value scheme. Expanding on the aforementioned studies, Boylan et al. (2008) demonstrate that the number of zero-demand periods can serve as an alternative parameter for classification by investigating the stock control inferences of the demand classification scheme. Nevertheless, the forecasting methods recommended under this framework fail to meet the target service levels for lumpy demand. In a similar vein, Syntetos et al. (2011) expand on this research area by conducting an empirical analysis of the connection between the classification parameters (p and CV^2) and the compatibility between practical demand distributions and industrial data sets. Based on their findings, they form heuristic guidelines for selecting the optimal theoretical demand distribution for inventory control. There are several other studies that propose alternative methods for addressing the demand classification problem. Lengu et al. (2014) introduces a demand classification framework that categorizes SKUs according to the mode and variability of the observed demand sizes and presents an empirical study that reveals that the compound Poisson distributions considered by the classification scheme tend to be well-suited for intermittent demand items. In their study, Petropoulos et al. (2014) examine the primary factors that contribute to forecast accuracy, focusing on eight distinct characteristics (seasonality, trend, cycle, randomness, number of observations, p, CV^2, and forecasting horizon). They also introduce a selection method that takes these characteristics into account to identify the most suitable forecasting technique and create a match in between.

Data aggregation methods

As an alternative approach to enhance the effectiveness of spare parts demand forecasting methods, data aggregation essentially involves combining data with comparable demand patterns, whether that be over time or across different time series, in order to reduce the occurrences of zero-demand periods and thus increase the accuracy of forecasts. Willemain et al. (1994) are one of the first studies to examine the impact of temporal data aggregation on Croston's method and show that the aggregation of daily data into weekly data yields higher accuracy. Next, Nikolopoulos et al. (2011) introduce a temporal aggregation framework referred to as *ADIDA* (aggregate-disaggregate intermittent demand approach) which involves generating forecasts using temporally aggregated data, which are then allocated back into the original (disaggregated) time series. Through empirical evaluations with an industrial data set, the authors demonstrate that ADIDA can substantially enhance the accuracy of both the naïve and SBA methods. Building upon the study by Nikolopoulos et al. (2011), Babai et al. (2012) dive further into the effectiveness of ADIDA by examining the inventory performance of three forecasting methods (Croston, SBA, and SES) using the same industrial data set. They determine that the use of ADIDA leads to higher service levels than those obtained through the utilization of disaggregated data for these forecasting methods. Similarly, Mohammadipour and Boylan (2012) suggest a temporal aggregation framework for the integer auto-regressive moving average (INARMA) process, when they apply this approach to two industrial data sets and reveal that, in most instances, the forecasts generated by the aggregation method yield lower MSEs compared to the cumulative forecasts obtained through summing up h-step ahead estimates. Next, Petropoulos et al. (2016) formulate a new aggregation framework, referred to as *iADIDA (inverse ADIDA)*, that aggregates demand volumes rather than time to reduce the variability of demand. Through their empirical evaluation, the authors show that iADIDA enhances forecast accuracy and is particularly effective for data sets with high degrees of data-volume variance. Boylan and Babai (2016) perform a statistical analysis to compare the performance of non-overlapping and overlapping temporal aggregation methods in which they demonstrate with an industrial data set that aggregating time series with overlapping time buckets generally outperforms the non-overlapping approach, with the exception of scenarios where the demand history is limited or the demand changes slowly. These findings support Porras and Dekker (2008) empirical method perspective which yields that overlapping time buckets enhance forecast accuracy.

Moreover, *cross-sectional/hierarchical* forecasting is an alternative forecast improvement approach to temporal data aggregation methods. This alternative approach enhances forecasting by grouping items based on similar characteristics and then predicting their total demand as a whole. In their study, Moon et al. (2012) evaluated direct and hierarchical forecasting methods, where direct forecasting methods relied on SES and hierarchical forecasts were generated by employing item and group-level direct forecasts. The authors discover that the most effective approach for minimizing forecasting errors and inventory costs is a combination of SES models, specifically involving aggregated quarterly data at the group level and monthly data at the item level. Li and Lim (2018) conduct a more recent study that showcases the benefits of using hierarchical forecasting in intermittent demand forecasting. They introduce a new hierarchical approach for this purpose, which employs aggregated and disaggregated forecasts through *a greedy aggregation-decomposition method (GAD)*. The study's findings indicate that the authors' proposed method outperform commonly used techniques like Croston and SBA, as well as ADIDA and iADIDA. Additional widely recognized techniques to enhance forecast accuracy are combining forecasts obtained from various alternative methods (Petropoulos and Kourentzes, 2015) that can be referred to as *Intermittent Multiple Aggregation Prediction Algorithm (iMAPA)* and improving outlier detection (Zhu et al. (2017), Romeijnders et al. (2012)). Firstly, the authors' analysis of combining forecasts derived from alternative methods concludes that combining forecasts based on modified frequencies, whether from one or multiple methods, can enhance the accuracy of predictions. Secondly, Pinçe et al. (2021) indicate that classifying demand spikes as outliers can be an alternative solution since these demand spikes come from preventive maintenance tasks. Thus, by detecting these significant demand points and consequently removing them from the data set as outliers, the performance of the forecasting methods may be improved. However, after detecting these demand spikes, removing them directly from the data set may not be the case for this paper. First, these data points should be compared with the installed base information (if there is any), then it needs to be decided whether removing these demand values from the data set is appropriate.

2.1.2 Contextual Methods

It is very well known that the management of spare parts is difficult since they have a high obsolescence risk because of their specific features and their demand is highly intermittent. Although there are various time series forecasting methods that take historical demand into account, they can be improved in such a way as to adapt rapidly to demand changes caused by external factors. The demand for spare parts frequently changes and reflects the life cycle of the equipment that they are used in. In addition, the other external factors that contain contextual information affecting the spare parts demand in a dynamic way are mainly based on maintenance schedules, the age of the equipment as well as its operating conditions. Hence there would be significant benefits if the demand for spare parts could be predicted by taking contextual knowledge into account. A practical example is from the energy industry, where the contextual data gathered to offer suggestions to consumers improves forecast results for energy management in buildings (Jozi et al., 2022). To overcome the limitations of time series techniques, the use of methods that incorporates contextual information has increased recently. Depending on the available contextual knowledge, this research stream can be separated into two subcategories. The articles categorized as the first group examines the effects of judgmental interventions on statistical forecasting techniques, while those classified as the second group concentrate on creating new forecasting methods that integrate data on the installed base.

Judgmental Forecasting

Both academic literature and standard industry practices suggest that human interventions are frequently used in the area of demand forecasting. Judgmental forecasting can be described as producing a forecast with the incorporation of human judgment and "gut feeling". According to Perera et al. (2019), judgmental forecasting refers to 3 different approaches. The first approach can be defined as pure judgmental forecasting, which involves creating predictions solely through unassisted human judgment, cognitive abilities, and various types of accessible information and insights. The second approach can be described as a combination of forecasts, which involves generating distinct statistical and/or judgmental forecasts, which are subsequently merged or averaged to obtain a final forecast using either additional human judgment or a structured averaging methodology. Recently, integration methods have been widely used in practice and implemented in numerous ways (Franses, 2014). The Third and final approach is the most commonly used one, which involves using human judgment to modify forecasts derived from statistical models. One of the main challenges of judgmental forecasting is reaching a consensus among evaluators who may have varying opinions because of their unique experiences, background knowledge, and perspectives (Salehzadeh et al., 2020). Pinçe et al. (2021) point out that little research attention has been given to understanding the impact of the judgmental intervention on items with intermittent demand over the years. Syntetos et al. (2009) provide the first academic attempt to explore the impact of judgment in intermittent demand forecasting. Their research shows that adjusted forecasts for intermittent demand items are more precise than the forecasts generated by the system, marking the first evidence of the effectiveness of judgmental interventions in this context. Also, the findings from the study indicate that negative adjustments are likely to enhance precision, which supports Fildes et al. (2009)'s conclusions. In a similar way, Boutselis and McNaught (2019) present a study on forecasting spare parts demand for military equipment where the actual demand can vary substantially due to changes in the context of military operations. To tackle this challenge, the authors introduce three Bayesian network models to generate forecasts of spare parts demand for a single period. Their findings suggest that the Bayesian network models are more accurate than the expert-adjusted SES and logistics regression methods. The research results conducted in different contexts are inconclusive regarding the effectiveness of making judgmental adjustments to statistically generated forecasts in improving accuracy and operational performance. Thus, additional research is necessary to extend the initial discoveries, considering the practical importance of judgmental forecasting in the field of intermittent demand.

Installed Base Forecasting

The need for spare parts comes from the replacement of parts in machines that are already in use, either as a preventive or corrective measure. The information that regulates the spare part demand for spare parts, such as the failure rate of components and maintenance policy, can be referred to as the installed base information (Dekker et al., 2013). Consequently, installed base forecasting is contextual forecasting where this information is fed into the forecasting procedure to increase prediction accuracy. According to the literature review Auweraer et al. (2019) conducted, there are three primary sources of information that influence the demand for spare parts related to the installed base: (1) the status of the spare part and the size and condition of the installed base, (2) the maintenance policy that determines when parts need to be replaced, and (3) environmental factors that affect the reliability of the part.

Although the concept of installed base forecasting is not extensively discussed in the field of operations literature, the idea of utilizing installed base information on demand forecasting for spare parts is not novel. In order to enhance the accuracy of spare parts demand forecasts at IBM, Cohen et al. (1990) propose integrating part failure rates and the number of machines installed into exponential smoothing. The logistics software, named SPARTA II, was developed by Petrovic and Petrović (1992). It employs fuzzy set theory and a Bayesian algorithm, along with installed base information, to estimate the likelihood of satisfying spare parts demand and to determine inventory levels. Next, in their study Aronis et al. (2004) utilize a Bayesian approach to estimate the failure rates of parts in telecommunication systems, later these estimation results are used to predict the demand for new parts that have no prior failure history. Ghobbar and Friend (2002) examine the correlation between demand lumpiness and installed base information for the aircraft industry by analyzing data from an airline operator using a general linear model, their findings indicate that the two primary sources of the observed lumpiness in demand patterns are maintenance (condition-based) and utilization information (flying hours and the number of landings) for the aircraft. Later, Jalil et al. (2010) present a paper to emphasize the possible economic benefits of utilizing installed base data in the field of spare parts logistics, also examine several data quality concerns related to the use of installed base data demonstrating that demand planning of spare parts depends on the quality factors as well. Next, Dekker et al. (2013) present a summary of installed base management and suggest various ways in which information installed base information can be utilized for improving forecasts, in addition, they conclude by reviewing some models that the prediction power of spare part demand forecast can be improved with the help of installed base information.

On the other hand, there are some academic papers aiming to fill the gap in the comparison of the proposed methods against widely used spare part demand forecasting methods. Hua et al. (2007) have been one of the first to do so. They introduce a new approach that takes the plant and equipment overhaul information into account with logistic regression for forecasting the intermittent demand of spare parts, and by utilizing petrochemical enterprise data, they demonstrate that their approach generates more precise lead time demand predictions when compared to SES, Croston's method, and Markov's bootstrapping method. Following Hua et al. (2007), Wang and Syntetos (2011) introduce a new technique that focuses on maintenance information to predict intermittent demand, their simulation results show that their approach performs exceptionally better when compared to SBA for almost all cases. Next, Romeijnders et al. (2012) introduce the first method that uses component repairs information, the method is evaluated in a comparative study using data from a service provider in the aviation industry with the findings indicating that the two-step method is highly accurate and outperforms Croston's method. The two-step approach can leverage data on planned maintenance and repairs to reduce forecasting errors by up to 20 percent (Romeijnders et al., 2012). Subsequently, Zhu et al. (2020) propose a forecasting mechanism that estimates the spare part demand distribution by analyzing the maintenance plan, and then they create a dynamic inventory control method that depends on these predictions. To evaluate the proposed approach, they compare it to other time series forecast methods, utilizing data from two major maintenance organizations, and their findings reveal that their technique can result in significant cost savings.

2.1.3 Comparative Studies

The literature contains numerous reports of comparisons made between various forecasting methods such as this paper. These comparative studies practically use performance benchmarks for forecasting methods that are utilized in predicting spare part demand. However, forecasting spare parts demand is particularly challenging due to its intermittent nature, which makes it difficult to estimate the lead-time demand distribution and find the right distribution if there is any for lead-time demand (Willemain et al., 2004). Mostly, comparative studies evaluate the performance of the different methods based on multiple forecast accuracy and inventory performance measures (Pinçe et al., 2021). This allows for a comprehensive assessment of the strengths and weaknesses of each forecasting method in handling spare part demand forecasting. The measurement of forecast accuracy aims to determine the difference between the actual demand observed in the past and the forecasts made. At the same time, inventory performance measures evaluate the effectiveness of a given forecasting method concerning various aspects of inventory management, such as service levels, on-hand inventory, instances of stock-outs, and total inventory costs. Although calculating the 90-95th percentiles of the forecasted demand distribution is a common practice in most inventory applications, it is essential to forecast the distribution of the tail to develop an effective inventory management system.

By considering both forecast accuracy and inventory performance measures, a more complete picture of the effectiveness of different forecasting methods can be obtained. However, when compared to forecast accuracy measures, there are only a small number of studies that have assessed how effective a particular forecasting method is by using an inventory control measure. Attaining a high level of accuracy in forecasting spare parts demand does not guarantee an equivalent level of performance in inventory management(Teunter and Duncan, 2009). Syntetos and Boylan (2006) suggest that evaluating the effectiveness of a forecasting method should be done based on inventory performance. Teunter and Duncan (2009) also demonstrate that forecast accuracy measures are not suitable for intermittent demand, despite being frequently employed in academic research. Therefore, it is crucial to assess a forecasting method by considering its impact on inventory management and service level. The most prevalent metrics for assessing inventory performance are the service level and tradeoff curve measures according to Pinçe et al. (2021). Most studies adopt cycle service level or fill rate as the primary criterion for assessing service level, and employ tradeoff curves to elucidate the interplay between inventory costs/volumes and attained service levels or back order volumes, where tradeoff curves are considered as pragmatic and realistic measures of inventory performance (Syntetos et al., 2015). In addition, a large number of studies utilize measure combinations for analyzing the impact of a forecasting method on spare parts inventory. Hereafter, insights from several comparative studies will be provided in the following subsections.

Insights from Comparative Papers

This section of the literature review focuses on comparative studies, in which critical benchmarking of various forecasting techniques can be observed. Willemain et al. (1994)'s paper is one of the first comparative studies, which compares SES and Croston's method while violating Croston's normality and independence assumptions, they find that Croston performs better than SES for both industrial and simulated data. By comparing five different forecasting techniques, Sani and Kingsman (1997) show that the moving average achieves the best inventory performance, followed by Croston and both provide higher inventory performance levels than SES for intermittent demand. In a similar vein, adding SBA on top of earlier mentioned methods, Syntetos and Boylan (2006) conclude that SBA outperforms all other methods in terms of inventory performance. With a focus on aircraft spare parts demand, Ghobbar and Friend (2003) compare thirteen well-known methods using MAPE, as a result, they find that the weighted moving average outperforms all other methods. However, both Eaves and Kingsman (2004) and

Teunter and Duncan (2009) criticize the inappropriate use of accuracy measures for forecasting intermittent demand. Moreover, the number of comparative papers including bootstrapping methods is surprisingly low (Pinçe et al., 2021). Syntetos et al. (2015) are one of the few ones to do so. They conclude that although Willemain's bootstrapping method (Willemain et al., 2004) outperforms the traditional ones, parametric methods still may be preferable because of their simplicity in practical applications.

The aim of Pinçe et al. (2021) is to provide a more comprehensive summary of the latest spare parts demand forecasting literature by analyzing the related studies' findings. To do so, they use 56 papers, which conduct a forecasting method comparison study, from the literature. For the quantitative analysis, they count how many times the method of interest is superior to the other methods as a result they come up with "better performance" scores. It is important to note that they classify the papers according to both data and the selected performance measure types. They divide the better performance score by the total number of comparisons, which turns into a "percentage better" score. Across data types and performance measures, they average these percentage better scores and obtain an "average percentage better" score. In this way, they provide where the method of interest stands among all other methods in such papers that use different data types and performance metrics but carry out similar comparisons. Their analysis starts with the comparison of Croston and SBA as the two benchmark methods. Next, they extend their work to include the comparison of the benchmark methods with traditional methods and newer forecasting methods. On top of these, they include the comparison of parametric and nonparametric methods as well as contextual methods. Their main findings are as follows, SBA is superior to Croston 87% of the time in terms of accuracy measures, however, it is inconclusive most of the time when the main concern is the inventory performance measures. In addition, inventory measures should be used alongside the accuracy measures to provide a more complete picture, the forecasting techniques using installed base information and deep learning algorithms show promise.

Academic literature suggests that Machine Learning (ML) techniques can be used as alternatives for statistical methods in the prediction of time series since both try to find the best possible combination of minimizing downtime and maximizing cost efficiency for organizations (Spiliotis et al., 2020). Although the goal of ML techniques is identical to that of statistical methods, their performances differ in terms of forecasting accuracy, inventory control efficiency, and computing efficiency, thus comparison between these methods becomes vital as the number of technological improvements has been increasing recently. Furthermore, although Pinçe et al. (2021)'s framework includes machine learning methods, it only considers Neural Network (NN) techniques under machine learning methods which can be considered a shortcoming. Considering the tradeoff between prediction performance and computational costs, deep learning approaches are computationally more demanding in general than pure machine learning methods. Thus, on top of the framework mentioned above, the literature review of other ML models (such as Support Vector Regression (SVR), Regression Trees, Random Forest (RF), etc.) has been conducted in a similar vein by including comparative papers. Firstly, Ahmed et al. (2010) showcase a comprehensive evaluation study of 8 different time series forecasting machine learning models applied to M3 competition monthly data. According to their findings, MLP and Gaussian processes (GP) are the top two models, and the radial basis function is the worst-performing method in the study. Next, Makridakis et al. (2018) present a comparative paper including 8 ML methods and 14 statistical ones. Their study shows that statistical methods (such as SES, Holt) are the best-performing methods in terms of sMAPE and are the least computationally demanding ones compared to ML methods. In a more recent study, Spiliotis et al. (2020) present a comprehensive comparative paper utilizing 11 statistical methods (e.g. SES, Croston, SBA, and TSB) as well as 7 ML methods (e.g. MLP, RF, and SVR) on daily SKU demand that can be characterized as intermittent and erratic. Their results show that the ML techniques such as Gradient Boosting Trees (GPT), RF, SVR, and K-Nearest Neighbor Regression (KNNR) are the top-performing methods in terms of absolute mean scaled error (AMSE) and root mean squared scaled error (RMSSE). However, regarding the tradeoff between accuracy and computational costs, they highlight that ML methods take 4 times longer to compute on average. Similarly, Kiefer et al. (2021) compare statistical, ML, and deep learning methods using mean absolute scaled error (MASE) and a new metric called stock-keeping-oriented prediction error costs (SPEC) on the M5 competition data set. In terms of the SPEC, Croston produces the most favorable outcomes, on the other hand in terms of MASE, Long short-term memory (LSTM) is superior to all other methods. In addition, they divide the time series based on intermittent and lumpy classes. They show that in intermittent time series Croston again ranks first while in lumpy time series Croston, LSTM and RF are the three top models. Moreover, M5 competition results show the forecasting superiority of ML methods when applied to intermittent time series in terms of weighted RMSSE. Specifically, four out of five winning submissions use a variation of the LightGBM algorithm (Makridakis et al., 2022). As a result, marking the first time in all M competitions, in the M5 competition all of the ML methods outperform all of the statistical benchmarks and their combinations. Most recently, Hendricks (2022) conduct a comparative study utilizing 4 machine learning and 6 statistical techniques, aiming to identify the most effective approach for predicting the demand for spare parts in simulators and weapon ranges that are used in non-manufacturing settings. The findings reveal that Support Vector Regression (SVR) outperforms all other methods for predicting the demand for spare parts required for scheduled and non-scheduled maintenance activities for a span of 52 weeks. In addition, they reveal that ML methods employed such as SVR, RF, and MLP perform better in terms of accuracy for repairable parts needed for maintenance actions that span 2 and 4 years. However, when predicting the demand for maintenance activities over 6 years, TSB shows superior performance compared to some of the ML methods.

2.2 Recent Literature on Outlier Detection Methods

As an alternative strategy to improve forecasting performances, data pre-processing (e.g. outlier detection, handling missing values) may lead to an improvement in forecasting performance. It is difficult to accurately identify abnormal demand and correct it due to the intermittent and highly variable demand occurrences of spare parts. These occurrences may arise due to the demand for either preventive maintenance which is planned or scheduled maintenance which is unplanned, both of which result in significant spikes in demand and are consequently labeled as outliers. Both the presence of missing values and outliers in demand can lead to a decrease in the accuracy of a forecast as it is suggested by Haan (2021b) that using a data preprocessing method improves the forecast performance. Although there are various academic papers on outlier detection methods, a comprehensive study is lacking that utilizes statistical and/or machine learning methods for outlier detection in the spare parts demand forecasting field, how an outlier detection method is conducted as well as how it affects the forecasting performance. Outliers are defined as data instances that exhibit a remarkable deviation from the established norms of a data set or anticipated patterns of behavior (Smiti, 2020). While the presence of these outliers may sometimes mislead analytical results and it is best to remove them, conversely, including the outliers can provide meaningful insights and, thus, their retention may produce a better performance. Therefore, this paper will investigate whether addressing these demand spikes can improve the performance of a forecasting model.

Multiple frameworks exist for classifying outlier detection techniques. They can be classified into two categories, parametric (statistical) methods and nonparametric (model-free) methods (Ben-Gal, 2006). Statistical methods generate assumptions about the underlying distribution of the observations or estimate the parameters of the distribution. They identify outliers as observations that deviate from the assumed model. However, statistical outlier detection methods have several drawbacks such as they often can not handle high-dimensional data and data with no prior knowledge of its underlying distribution very well (Papadimitriou et al., 2003). Thus, these approaches can not be used if the distribution is unfamiliar or unknown, which makes the use of ML methods convenient. Since this paper's focus will be on ML outlier detection methods, detailed information on statistical outlier detection methods will not be covered (please see Ben-Gal (2006) for detailed information).

Machine Learning (ML) Methods

Detecting outliers in extensive real-world databases presents significant challenges to the current methods being used, which has led the ML methods to emerge. Outlier detection is an example of how machine learning can be applied, as it employs techniques to identify observations that deviate greatly from the rest. In the field of demand forecasting, it is being explored whether machine learning can be used to identify unusual patterns in the demand for spare parts. There are two branches of machine learning approaches for outlier detection purposes, supervised and unsupervised methods. Unsupervised methods develop a model without any given information about which observations are outliers. On the other hand, supervised methods are trained on a subset of the data in which the outliers have been identified and labeled. Since there are no outliers identified or labeled in any of the data sets that will be utilized in this paper, the subsequent discussion covers only the unsupervised outlier detection methods. In the following, firstly tree-based methods will be discussed, secondly, k-nearest neighborhood (KNN) based methods will follow, and lastly, a discussion of a density-based method will be held.

Decision tree (tree-based) algorithms are highly recognized and extensively utilized among all ML techniques (Salzberg, 1994). As one of the first and most commonly used tree-based methods, Quinlan (1993) presents an algorithm, called *C4.5* which is created as an extension of the ID3 algorithm. It constructs a decision tree through a recursive partitioning of subsets of the training data set. Following the construction of the tree, certain nodes are identified and eliminated based on their impact, measured by the decrease in error that results from their removal. These pruned nodes are subsequently designated as outliers and their influence is eliminated from the model. Next, John (1995) suggests the presence of non-informative records at a local level does not contribute to the algorithm's ability to detect patterns at a global level and develops a more robust method, called *ROBUST-C4.5*. It involves a continued application of removal of observations, whereby the process is re-executed on a reduced training set, leading to the creation of newer decision trees with subsequent node pruning, and this sequence is repeated iteratively until no further nodes can be pruned. Later on, the author presents a comparative study on C4.5 and ROBUST-C4.5, which shows that ROBUST-C4.5 leads to slightly higher accuracy and can generate trees that are 29 percent smaller compared to C4.5's trees, without compromising on the accuracy. As another tree-based method, Liu et al. (2008) propose a method that utilizes the concept of isolation, which can be referred to as *Isolation Forest (iForest)*. The Isolation Forest algorithm employs a collection of randomly created trees to evaluate the level of "outlierness" of each observation. This is accomplished by computing the average length of the path for each observation and filtering out the observations with the shortest average length. The iForest algorithm is effective in dealing with problems that involve a high number of irrelevant attributes, even in situations where the training data does not include any outliers, in addition, it performs well in high-dimensional problems (Liu et al., 2008). However, Fan et al. (2023) conduct an empirical study on intermittent time series based on the real demand data for after-sale components from two major manufacturing companies and conclude that iForest does not perform well in terms of accuracy for intermittent demand.

In outlier detection algorithms, distance-based methods known as nearest neighbor (NN) techniques have gained popularity recently (Su and Tsai, 2011). In distance-based methods, if an observation is positioned at a distance beyond a preset threshold value from a specified proportion of other observations in the data set, then it can be classified as an outlier. When dealing with higher dimensional space, distance-based methods tend to perform better and can be computed more efficiently than statistical methods (Malik et al., 2014). However, a primary challenge to the practical implementation of NN techniques lies in the computationally intensive task of calculating the distance between every pair of data points (Su and Tsai, 2011). K-nearest neighbor (KNN) can be seen as one of the most widely used distance-based techniques due to its simplicity and conventionality, which is a non-parametric approach that is used for classification purposes. Its algorithm estimates the distances between different points on the input vectors based on a distance measure e.g. the Euclidean distance. Subsequently, it allocates the unclassified point to the class of its K nearest neighbors Omar et al. (2013). Moreover, Ramaswamy et al. (2000) introduce a partition-based algorithm to speed up the KNN algorithm. This novel algorithm initially divides the input data into separate subsets, followed by pruning complete partitions once it is established that they can not contain outliers. Consequently, this approach significantly reduces the computational costs involved in the process. In a more recent paper, Bandaragoda et al. (2018) introduce a novel approach, referred to as *iNNE*, for detecting anomalies using isolation, which integrates decision trees and KNN. The authors develop this method to overcome the weaknesses of the iForest algorithm as it constructs spherical boundaries around a set of observations using their proximity to their closest neighbors. The size of the sphere determines the likelihood of an observation being identified as an anomaly. It performs multiple rounds of this process and identifies the observation with the largest sphere in each round, adding it to a list. Subsequently, it assigns an isolation score to every observation on the list, and the top-ranked observations are considered the most probable anomalies. The benefits of utilizing this approach include its ability to effectively process multi-dimensional data sets and its linear time complexity. Nevertheless, due to its recent development, there is limited usage of this method apart from the reference provided in this paper.

The density-based approach for identifying outliers is introduced as a solution to the limitations of distance-based global outlier detection techniques. As a density-based approach, Local outlier factor (LOF) is known as a state-of-art unsupervised algorithm that focuses on local outliers. Also, the concept of local anomalies was initially introduced with this method which is defined by Breunig et al. (2000). It differs from other methods in such a way that it attributes a score of outlierness to each object, referred to as the local outlier factor (LOF). This degree is determined locally based on how isolated the object is relative to its surrounding data points. LOF shows strong potential, according to Breunig et al. (2000), in detecting significant local outliers that have gone unnoticed using previous techniques. Once the LOF algorithm is defined, many more methods emerged following the same logic. Thus, variations are primarily attributed to determining the local neighborhood and calculating outliers. While Tang et al.

(2002) propose a redefinition of outliers by connectivity-based outlier factor (COF) algorithm, Ni et al. (2008) introduce a local entropy-based weighted subspace outlier mining algorithm, referred to as SPOD, which can be seen as an enhancement to the LOF algorithm. In a similar vein, Hu and Qin (2010) introduce the density-based local outlier finder (DLOF) method, which utilizes information entropy to identify outlier characteristics for each data point. Later on, Tang and He (2017) develop a new scoring system to compute the local outlierness, called a relative density-based outlier score (RDOS). The authors conduct an extensive empirical study using both synthetic and real-life data sets, their results illustrate that their approach is more efficient in identifying outliers than existing methods.

Conclusion on the Recent Literature Review

In summary, time series methods aim to capture intermittent spare parts demand. Traditional methods like SES, Croston, SBA, and TSB are reliable benchmarks. Balancing forecasting accuracy with inventory performance and obsolescence, determining lead-time demand distribution, and performing well on different data sets are the most common challenges. Overall, forecasting results are heavily dependent on the data, the method, and the performance measure used. Each method has its strengths and performs well under specific conditions. Contextual methods incorporate external information to enhance forecasting, while judgemental forecasting uses managerial knowledge, and installed base forecasting utilizes product/maintenance characteristics. Contextual methods improve precision but their reliance on intuition can harm forecasting accuracy.

The academic literature compares various forecasting techniques, including traditional methods like SES, Croston, and SBA, as well as newer methods like deep learning algorithms. Evaluating a forecasting method should consider both forecasting accuracy and inventory performance measures. A comprehensive benchmarking framework by Pinçe et al. (2021) focuses on neural network techniques but lacks a review of other ML models. Recent studies suggest that ML techniques like RF and SVR may outperform traditional methods in specific conditions, but their computational costs should be considered. Future research should explore ML techniques for intermittent demand forecasting, considering both accuracy and inventory performance measures.

Addressing missing values and outliers in demand through data preprocessing can improve forecasting performance. This study explores whether identifying significantly different demand values enhances forecasting. Existing literature lacks a comprehensive investigation of outlier detection using statistical and/or ML techniques for spare part demand forecasting. Outlier detection techniques can be classified as parametric and nonparametric methods, with a focus on ML-based nonparametric methods in this paper. ML methods offer advantages over statistical techniques, such as handling high-dimensional data and data with unknown distributions. The use of ML for outlier detection in spare part demand forecasting is an emerging field requiring further research to determine the most effective approach.

Chapter 3

Methodology

In this section of the paper, a graphical representation of the experimental design is presented. Following that, the selected forecasting methods employed to address the first research question are introduced and described in technical detail. Furthermore, the outlier detection method utilized to tackle the second research question is described similarly. Next, the paper will provide details about the forecasting accuracy and inventory control measures used. Subsequently, the methodology used for classifying demand based on the four categories established by Boylan et al. (2008) will be elucidated. This methodology is employed to address the first research question of this paper. As the last step, the information on how the data sets are split into train and test sets will be provided.

3.1 Experimental Design

The experimental design, as visually presented in Figure 3.1, is grounded on the research questions presented in this paper's introduction. These questions focus on the diverse outcomes that can arise from utilizing data pre-processing and spare part demand forecasting methods on various types of data sets. Initially, the industrial data sets that have been provided to me require some data wrangling, and subsequently, data pre-processing including an outlier detection method will be applied. In the following step, the data sets are classified based on their demand characteristics to address the initial research question of this paper which focuses on examining the impact of different data types. The chosen forecasting methods are then applied, followed by a comparison of these methods based on their forecasting accuracy and inventory control measures. Additionally, a comparison will be conducted between two versions of industrial data sets: one with applied outlier detection procedure and the other without, it aims to address the second research question in this paper. Furthermore, the paper discusses the variations in the outcomes due to differences in the data sets.

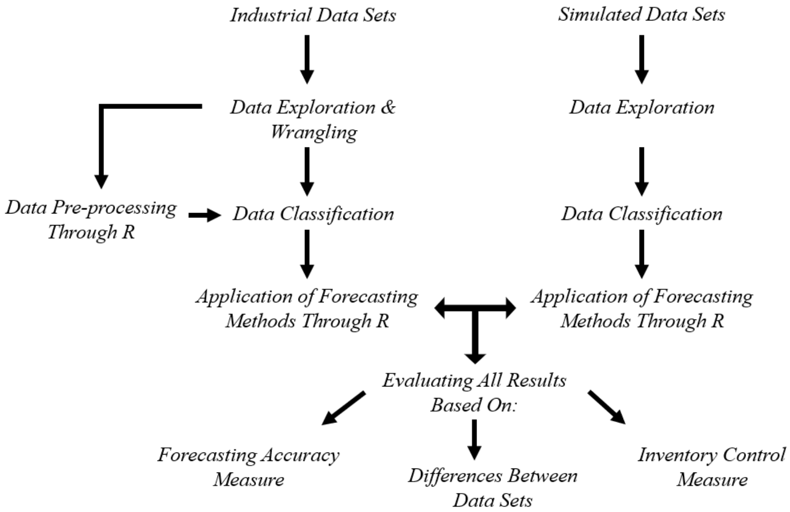


Figure 3.1: The flow of the experimental design

3.2 Selected Forecasting Methods

The first forecasting method to be employed in this paper is Single Exponential Smoothing (SES) which will serve as a benchmark. Although it is known to perform poorly with intermittent demand, it is one of the most commonly used methods in forecasting which is why this method is included (Monfared et al., 2014). The weighted average predictions produced by SES decrease over time and are smoothed using the smoothing parameter α, which is determined by the cost function. According to Syntetos and Boylan (2005), this smoothing parameter is usually set between 0.1 and 0.3 in settings with intermittent demand. The formula of SES is as follows:

$$\hat{y}_t = \alpha y_t + (1 - \alpha)\hat{y}_{t-1} \tag{3.1}$$

As another benchmark method, Croston's method, which is developed to outperform SES, will be employed for comparison in this paper. Croston's method predictions are based on two components, namely the demand size z_t and the inter-demand interval p_t. Since predictions are updated only when demand occurs, the demand size z_t has to be non-zero in at least two periods. The formula for Croston's predictions is given as follows:

$$\hat{y}_t = \frac{\hat{z}_t}{\hat{p}_t} \tag{3.2}$$

The first observation in the series is used as the initial values of z_t and p_t. SES is used to predict both z_t and p_t, with a smoothing parameter optimized by a cost function, as recommended by Kourentzes (2014). The average estimated demand for each time period in the forecasting horizon is the ultimate output from Croston.

As a modification to Croston's method, Syntetos and Boylan (2005) showed that Croston's method is positively biased and developed SBA, which is included as the third forecasting method. In their estimation, they incorporated a smoothing parameter α, which is used to reduce bias and smooth out the inter-demand interval p_t. In a similar vein to Croston, the initial values of z_t and p_t are derived from the first observation, and α is set to be 0.1. SBA's formula is as follows:

$$\hat{y_t} = (1 - \frac{\alpha}{2})\frac{\hat{z_t}}{\hat{p_t}} \tag{3.3}$$

As the next method TSB, which has been developed by Teunter et al. (2011) to address the issue of demand obsolescence, is utilized in this paper. When dealing with intermittent demand accompanied by obsolescence, both of them together could lead to delayed identification of obsolescence. TSB's forecasting is conducted based on SES, too. When addressing the issue of obsolescence, TSB modifies Croston's method by replacing the inter-demand interval p_t with d_t which is the demand probability. d_t becomes 1 when demand occurs, and 0 otherwise. The formula for the predictions made by TSB is as follows:

$$\hat{y_t} = \hat{d_t}\hat{z_t} \tag{3.4}$$

Kourentzes's (2014) R-package called "tsintermittent" is utilized for Croston and three other methods (SES, SBA, and TSB), which requires the selection of various parameters, including a cost function, for optimization purposes. The cost function options that are discussed include mean squared error (MSE), mean absolute error (MAE), mean squared rate (MSR), and mean absolute rate (MAR), in addition, all of the parameters are extensively discussed in Kourentzes (2014)'s work. Their empirical research shows that the MAR cost function is the most effective for Croston, SES, and SBA, while TSB performs better with the MSR cost function. Therefore, the MAR cost function is employed for Croston, SES, and SBA, and the MSR cost function is employed for TSB in this paper.

Before diving into machine learning (ML) techniques, an alternative spare parts forecasting method is Willemain's bootstrapping (WSS) which is introduced by Willemain et al. (2004) and known as one of the most popular non-parametric approaches. In addition, as mentioned before, it addresses challenging characteristics of intermittent demand as well as shows superior accuracy results when compared to both benchmark methods, namely SES and Croston, according to Willemain et al. (2004). Taking all into consideration, this method is decided to be employed in this paper. Utilizing a two-stage Markov chain process, WSS is a technique that calculates non-zero demand points and resamples demand using historical data. The incorporation of a jittering process provides greater variability in the method's predictions and avoids replication of prior values. The technique pre-calculates a sequence of zero and non-zero demands, based on transition matrix probabilities, to prepare for the next step.

According to Spiliotis et al. (2020), ML methods have the ability to replace statistical methods in predicting time series since they share the same goal of determining the best possible combination that minimizes downtime and maximizes cost efficiency. Thus, Spiliotis et al. (2020) developed a method called Multi-Layer Perceptron (MLP), also known as a single hidden layer neural network, which has been trained using Smyl (2020)'s approach. This method involves using a rolling input and output window of constant size to predict future data points. The MLP is built using the R-package RSNNS, as described by Bergmeir and Benítez (2012). During the training process, Grid Search Cross-Validation (CV) is used for one-SKU hyperparameter optimization. The following parameters are selected to optimize the model by determining the lowest RMSE values which provide the highest accuracy: maximum number of iterations to learn (*maxit*), learning function which defines how learning takes place in the network (*learnFunc*), learning rate (*learnFuncParams*). When applying the MLP model, the data needs to be scaled due to its use of a non-linear activation function, and the learning speed increases with scaling Zhang et al. (1998). The scaling for the linear transformation of the data to be between 0 and 1 is conducted according to the following formula:

$$\hat{y_t} = \frac{y_t - y_{min}}{y_{max} - y_{min}} \tag{3.5}$$

After generating predictions, the scaling needs to be reversed to obtain final forecasts and assess forecasting accuracy.

As the second ML algorithm, LightGBM method is employed by using the code for constructing the predictive algorithm from Kailex (2020), which is used to enter the M5 competition. This particular method is chosen for various reasons, such as LightGBM serving as the foundation for many of the top-performing methods in the competition as well as the code for this method being available in R. Similarly to the MLP, the LightGBM method uses a rolling input and output window, as well as lag variables to generate predictions. The model is trained using a Poisson loss function, and the hyper-parameters for the LightGBM algorithm are also adapted from Kailex (2020). In addition, the following parameters are tuned using Grid Search CV for one-SKU hyper-parameter optimization of the LightGBM model by determining the lowest RMSE value which provides the highest accuracy: the limit of the max depth for the tree (*max_depth*), the maximum number of leaves in one tree (*num_leaves)* and the learning rate (*learning_rate*). Before the application of the LightGBM algorithm, the data needs to be scaled in a similar vein with equation 3.5. After training the model, it gets evaluated at every 400 iterations to determine the optimal stopping point. Once training is finished, lag variables are utilized to generate forecasts. For a detailed understanding of its functionality and the R implementation, please see Microsoft (2021).

This paper utilizes Random Forest (RF) as the third ML method. By constructing multiple decision trees, RF overcomes the limitations associated with single decision trees as it becomes more robust to noise and is less likely to over-fit on the training data (Breiman, 2001). According to Spiliotis et al. (2020), RF is one of the best-performing methods on daily SKU demand, which is described as erratic and smooth, in terms of both AMSE and RMSSE. That's why this method is included in this study. Similarly to the other ML models, the RF method uses the same logic of rolling input and output windows to generate predictions. The implementation of RF is conducted using the randomForest R package Liaw and Wiener (2002). Before the application of the RF algorithm, the data needs to be scaled in a similar vein with equation 3.5. The optimal values of the following parameters are obtained through hyper-parameter optimization based on a single SKU and by performing a grid search cross-validation: the number of non-pruned trees (*ntree*), the minimum size of terminal nodes (*nodesize*), and the number of variables randomly sampled as candidates at each split (*mtry*). This process involves identifying the lowest RMSE value that ensures the highest level of accuracy.

Support Vector Regression (SVR) is the last forecasting and ML technique that is employed in this paper since it outperforms the rest of the methods besides RF in the study of Spiliotis et al. (2020). By minimizing the overall error, SVR generates predictions through the identification of a hyperplane that maximizes the margin between two classes (Schölkopf and Smola, 2001). Before the application of the RF algorithm, the data needs to be scaled in a similar vein with equation 3.5. Similarly to the other ML models, SVR method uses the same logic of rolling input and output windows to generate predictions. It is implemented by using e1071 R package (Meyer et al., 2019). As stated by Spiliotis et al. (2020), ν-regression is selected as it constructs a small number of support vectors and therefore simplifies the computations. The kernel used for both training and predicting is chosen from the linear, polynomial, radial basis, and sigmoid alternatives. The optimal values of the following hyper-parameters are determined through a combination of one-SKU based hyper-optimization and a grid search cross-validation (CV) process which involves identifying the lowest RMSE value: tolerance of the termination criterion (*tolerance*) and ν (*nu*), epsilon and cost (C).

3.3 Utilized Forecasting Accuracy Measures

There are numerous performance measures available in the literature to compare the predictive power of various forecasting methods. Since there is not a generally accepted forecast accuracy measure, most research papers utilize multiple metrics to provide a comprehensive understanding of the overall performance of a specific method. Thus, two different forecasting accuracy metrics will be employed in this paper. The forecasting accuracy metrics allow us to obtain insights into the extent of the difference between the projected and realized levels of demand. They can be categorized into two groups, namely absolute and relative accuracy measures (Pinçe et al., 2021). Absolute accuracy measures evaluate the performance of a particular forecasting technique for a specific time series. In line with Pinçe et al. (2021), mean absolute scaled error (MASE) is a prevalent absolute accuracy metric, and it is also recognized by Koehler and Hyndman (2006) as the most promising metric for intermittent demand. Henceforth, this study has opted to employ MASE. The formula for $MASE_t$ is as follows:

$$MASE_t = \frac{1}{t}\sum_{s=1}^{t} \frac{|e_s|}{\frac{1}{t-1}\sum_{i=2}^{t}|Y_i - Y_{i-1}|} \tag{3.6}$$

where e_s can be defined as forecast error and equals to $e_s = Y_s - \hat{Y}_s$. In this study, the second metric to be utilized is the root mean squared scaled error (RMSSE), which is a scaled version of RMSE and a modified form of MASE. Koehler and Hyndman (2006) first introduce it and reveal its superiority in handling data sets with intermittent demand. Moreover, it has also been employed in the M5 Competitions (Makridakis et al., 2022). $RMSSE_t$ formula is as follows, in which n indicates the number of observations within the training data, y_t and $\hat{y_t}$ signify the true and forecasted values of the time series at time t and h denotes the prediction horizon (equal to the duration of the test data set):

$$RMSSE = \sqrt{\frac{\frac{1}{h}\sum_{t=n+1}^{n+h}(y_t - \hat{y_t})^2}{\frac{1}{n-1}\sum_{t-2}^{n}(y_t - y_{t-1})^2}}. \tag{3.7}$$

In terms of metric interpretation, superior results are indicated by lower values for both measures.

3.4 Utilized Inventory Control Measures

Attaining a high level of accuracy in forecasting spare parts demand does not always guarantee an equivalent level of performance in inventory management (Teunter and Duncan, 2009). In a similar vein, Dekker et al. (2013) support the idea of integrating inventory control metrics results in enhanced management of the supply chain. Consequently, the intention of this paper is to provide a more comprehensive assessment of the efficacy of diverse forecasting methods by incorporating inventory performance measures in addition to forecasting accuracy measures. As suggested by Pinçe et al. (2021), tradeoff curves and service levels are the most frequently employed stock control measures in the current literature on spare parts demand, and therefore have been chosen for this paper.

Firstly, trade-off curves between inventory holding costs and achieved fill rates will be established since this approach provides more comprehensive insights according to Van Wingerden et al. (2014). Holding cost can be found by the multiplication of the average inventory level by the holding cost percentage. Since some data sets in this study do not provide necessary variables for evaluation, industry-accepted proxy measures will be used to approximate the holding cost variable. For instance, taking a holding cost between 20% to 30% of the total inventory value is reasonable, where 25% is selected as a middle-ground approximation (McCue (2022), Durlinger and Paul (2012)). The inventory policy approach outlined by Van Wingerden et al. (2014) is used in this paper in order to assess achieved fill rates. It involves determining a base stock level (R) based on the historical demand patterns and updating inventory position (IP) on a periodic basis with back orders being permitted as necessary. To determine the Part Fill Rate (PFR), the current IP is divided by the current demand size and then the resulting value is multiplied by 100. For instance, if the demand size is 20 and the IP is 10, the achieved PFR would be 50%. This calculation is repeated for each demand moment in the series and then averaged across all these moments to determine the achieved PFR for each forecasting method. After all, a target fill rate will be predetermined, followed by the computation of the corresponding R using the selected forecasting method. In this study, fill rate targets of 75%, 80%, 85%, 90%, 95%, 99%, 99.99% will be utilized. The method's fill rate performance will then be evaluated in comparison to the target fill rate, taking into account the holding costs associated with each achieved fill rate. The described procedure will be iterated for each forecasting method, leading to the generation of the tradeoff curves. Lastly, the item fill rate, which follows a normal distribution, rather than the order fill rate as is utilized by Van Wingerden et al. (2014) will be employed in this paper.

3.5 Utilized Outlier Detection Method

Local Outlier Factor (LOF) method is an unsupervised density-based outlier detection algorithm, which is first proposed by Breunig et al. (2000) and has led the way in developments of many more density-based techniques (Tang et al. (2002), Ni et al. (2008), and etc.). There are 3 steps to take when calculating the LOF score, which is the resulting outlierness degree of each object. According to Goldstein and Uchida (2016), firstly each record x requires determining its k nearest neighbors (N_k), and if there is a tie-in distance for the kth neighbor, more than k neighbors are utilized. Secondly, the local reachability density (LRD) is computed by determining the density of a record's nearby k-nearest-neighbors (N_k). As the final step, by evaluating the LRD of a record against the LRDs of its k neighbors, the LOF score is calculated which represents the proportions of local densities of records. Normal records, which are not computed to be outliers, are assigned to have a similar level of density to their neighbors with an approximate score of 1.0. On the other hand, outlier records are attributed to lower levels of density which results in higher LOF scores (higher than 2 as a rule of thumb).

For this algorithm, the only and the most significant task is to decide the value of k, the number of neighbors we are interested in. The rule of thumb for k is to be between 10 and up to 50 (Goldstein and Uchida, 2016). In the context of evaluating the Local Outlier Factor (LOF) method against alternative outlier detection algorithms, such as the partition-based algorithm proposed by Ramaswamy et al. (2000), an additional parameter known as "top n" is introduced, which requires optimization. It is the responsibility of the user to specify the desired value for this parameter in the partition-based algorithm. The reason for selecting the LOF method rather than the partition-based algorithm for implementation in this paper is partly attributed to the fact that it involves optimizing a single parameter, which simplifies the optimization process. The implementation is done by using the R-package called DDoutlier with k values selected according to the rule of thumb. Lastly, the outlier detection procedure is applied only to the industrial data sets in this study.

Finally, it is noteworthy to mention that when making decisions regarding the choice of outlier detection methods, a total of three distinct methods were considered, the LOF, the partition-based algorithm, and, the interquartile range (IQR) method. The rationale for excluding the utilization of the partition-based algorithm is elaborated above. In addition, Interquartile Range (IQR) is a statistical technique that entails segmenting the data set into quartiles. The IQR is calculated as the disparity between the third quartile (Q3) and the first quartile (Q1). Detection of outliers involves identifying data points that reside below Q1 - 1.5 times the IQR or exceed Q3 + 1.5 times the IQR. The decision to refrain from implementing this approach comes from the limited occurrence of non-zero demand values, which shows the intermittent characteristics of the data sets. Thus, even minor demand values are classified as outliers by this method, implying that the IQR necessitates a larger number of demand observations to determine the outlierness of a record accurately.

3.6 Demand Pattern Identification and Data Splitting Rule

This section of the paper aims to describe the methodology for the demand classification framework as well as the rule that's been followed to split each data set into train and test splits. To begin with demand classification, there have been many attempts in the literature to classify demand (Williams (1984), Johnston and Boylan (1996), etc.), and demand classification has been recognized as a forecast improvement strategy as well. Boylan et al. (2008) emphasize that items with different SKUs have distinct demand characteristics that call for unique approaches. Hence, to effectively address the first research question, which seeks to explore suitable methods for different demand patterns, it is essential to categorize SKUs. In this study, Boylan et al.

(2008)'s framework is adopted to identify demand patterns. They categorize demand patterns into four groups according to two metrics, the mean inter-demand interval (p) and the squared coefficient of variation of the demand sizes (CV^2). These four demand categories are erratic, lumpy, smooth, and intermittent. Their threshold values can be seen in 3.6. Items with erratic

Demand pattern	p	CV^2
erratic	< 1.32	≥ 0.49
lumpy	≥ 1.32	≥ 0.49
smooth	< 1.32	< 0.49
intermittent	≥ 1.32	< 0.49

Table 3.1: Demand Patterns and Cut-off Values

demand are characterized by demand values that greatly fluctuate. Meanwhile, lumpy demand items have rare occurrences of demand but highly variable demand values, whereas items with intermittent demand have also rare occurrences of the same demand values. Lastly, items that are categorized as smooth have regular demand patterns with low variability in demand sizes. In order to identify demand patterns for each item in the following chapter, p and CV^2 values need to be calculated. p can be computed by dividing the total number of periods by the number of demand occurrences that have non-zero values. CV^2 values can be computed by first dividing the standard deviation of demand occurrences that have non-zero values by the mean of demand occurrences that have non-zero values and then taking the square of the resulting number.

Lastly, each data set is split into train and test splits for forecasting methods according to the 70-30% rule, where the first 70% of each data is used for training purposes and the last 30% is labeled as test data, which is used for validation purposes. On the other hand, the separation of data sets into train and test is not a prerequisite for the outlier detection method.

Chapter 4

Data Description and Classification

In this section, a concise description of the main attributes of data sets, both artificial and industrial, is provided. In addition, data pre-processing steps including an outlier detection and handling process are elucidated. Given that this paper employs an outlier detection method, it results in two different versions of industrial data sets. Therefore, the classification of data sets is performed correspondingly in the subsequent part. Lastly, the R script for data wrangling, description, and demand classification is adapted from the GitHub repository Khue (2023) which originates from Haan (2021a) and is adjusted according to the needs of this paper.

4.1 Industrial Data Sets

This paper will utilize four data sets sourced from industrial companies. The number of distinct items originally within the data sets, the overall duration they represent, and the industry they belong to can be seen in 4.1. In addition to the provided information in the table, these data sets encompass additional variables. The initial data set, referred to as *AUTO*, lacks information regarding price and lead time. The same data set was employed by Syntetos and Boylan (2005). The second data set, referred to as *MAN*, encompasses various variables including prices, inventory costs (equivalent to 20% of the product cost), lead time, demand frequency, demand size data, minimum order quantity (MOQ), as well as fixed order costs (Haan, 2021b). It is important to note that there are demand values that are not integers in the MAN data set. The third data set, referred to as *BRAF*, was also employed by Teunter and Duncan (2009). *BRAF* encompasses information on demand size, demand frequency, prices, and lead time; nevertheless, lacks information on inventory costs (Haan, 2021b). The fourth data set, referred to as *OIL*, was utilized in the studies of Porras and Dekker (2008). It also contains information on prices and lead times. Additionally, it offers the minimum and maximum advised stock levels for each item and indicates the classification of the system where an item is installed in terms of its impact on operations, categorizing it as low (L), medium (M), or highly critical (H).

Data Set	# of Items (SKUs)	Total Amount of Periods	Industry
AUTO	3000	24 months	Automotive
MAN	3451	150 weeks	Manufacturing
BRAF	5000	84 months	Aircraft
OIL	14523	56 months	Petroleum

Table 4.1: Industrial Data Sets

Data set	Average Monthly Sales of Items				Product Price (€)				RPS
	Min	Mean	Max	SD	Min	Mean	Max	SD	
AUTO	0.542	4.450	129.167	7.573	25.075	727.851	5979.315	1053.357	163.569
MAN	0.007	24.224	4599.653	139.294	0.033	35.964	2669.700	101.824	1.485
BRAF	0.036	1.442	65.083	3.617	0.001	102.321	9131.992	373.334	70.943
OIL	0.036	1.077	228.571	4.114	0.010	450.338	82562.590	1453.955	418.280

Table 4.2: Descriptive Statistics of Industrial Data Sets

In 4.1, we compute the average monthly sales for each item across all time periods within each data set. Subsequently, we calculate the minimum, maximum, mean, and standard deviation values for each data set, encompassing all of its items. This table utilizes the data set versions that have undergone data pre-processing steps, as detailed in the following subsection called "Data pre-processing". There are two important things to note, firstly the minimum price of 0.001 Euros can be interpreted as a result of multiple units of the same product being sold at a reduced price, leading to an actual price below one cent. Secondly, weekly sales data is aggregated over a 4-week period to obtain monthly sales figures for the MAN data set.

Computation of Individual Prices

In 4.1 above, the descriptive statistics of industrial data sets are given including the ratio of price and sales (RPS) values. The computation of individual prices for the AUTO is adapted from the studies of Haan (2021b). This analysis aims to facilitate the calculation of the inventory control performance by incorporating pricing information into the AUTO. Hereafter, the steps that are taken for this analysis are explained. Initially, RPS values are computed for the MAN, BRAF, and OIL data sets. This is achieved by dividing the average product price by the average monthly product sales. The RPS for the AUTO data set is established as the average of the RPS values of the other data sets, which is equal to 163.569. Next, the mean price per item is computed by multiplying the monthly sales by the RPS, which results in 727.851 in **??**. Despite the comparatively lower average monthly item sales, this high value can be explained by the existence of specific items that have high frequencies of demand, exerting a substantial influence on the overall average product price. Next, the ratio of monthly sales (RMS) is calculated. Individual item prices can be more accurately assigned by keeping in mind that items with high frequencies typically exhibit lower prices. Thus, RMS is computed by dividing the monthly sales values of each item by the average monthly item sales. As the last step, individual prices for AUTO are calculated by dividing the mean product price by the RMS values.

4.2 Artificial Data Sets

The generation of artificial data enables the incorporation of specific attributes to examine their influence on the performance of the selected forecasting methods. Thus, relating to the first research question, artificial data sets are generated to evaluate the potential impact of different data types. It is important to note that all artificial data sets are generated and provided by Haan (2021b). Four simulated data sets in this study were generated using the R package 'tsintermittent'. According to Petropoulos et al. (2014), the package utilizes a Bernoulli distribution to model non-zero demand arrivals, and a negative binomial distribution is employed to generate non-zero demands. Furthermore, creating a data set with this package involves specifying certain mandatory input parameters (such as the number of time series, the number of observations, etc.). Considering that the average count of time series which means the average number of items in the industrial data sets was 6493.5, the value of 'n' for the artificial data sets was set to 6500. The data configuration emulated monthly sales data, with the number of periods set to 60 months (equivalent to a span of five years). Next, the average demand size, CV^2, and p were intentionally established in a manner that ensured each of the four simulated data sets fell into one of the four distinct categories outlined in section 2.6. These data sets are denoted as SIM1, SIM2, SIM3, and SIM4 respectively. The average demand size for all the artificial data sets was set to an arbitrary value of 10. The specific configurations can be found in Table 4.3. As can be seen, when the inter-demand interval (p) exceeds 1, the mean monthly demand decreases. Moreover, the average ratio of price and sales (RPS) of 163.569 was employed to determine the average product price, consistent with the industrial data sets. Subsequently, the individual product prices are determined following the process outlined in section 3.1.

Data set	Demand Category	CVˆ2	p	Demand (monthly)		Product Price (€)	
				mean	SD	mean	SD
SIM1	Erratic	0.75	1.00	10.01	1.12	1637.97	189.28
SIM2	Lumpy	0.80	1.50	6.66	1.12	1089.67	196.00
SIM3	Smooth	0.30	1.05	9.50	0.74	1553.29	122.36
SIM4	Intermittent	0.25	1.45	6.90	0.81	1128.10	138.61

Table 4.3: Artificial Data Sets Configurations

4.3 Data Pre-processing

Given that the majority of forecasting techniques depend on the characteristics of the data, the process of data pre-processing becomes highly significant as it is also suggested in Haan (2021b). In light of this, a detailed explanation of the data pre-processing steps will be provided hereafter, beginning with a description of the data wrangling process, and subsequently elaborating on the procedure for detecting and handling outliers.

The data sets utilized in this study represent a panel structure, characterized by the inclusion of multiple items over a time frame. Specifically, the items refer to distinct SKUs (spare parts) as observations, while the columns (variables) pertain to time periods in the beginning. Firstly, separate price vectors are generated from the MAN, BRAF, and OIL data sets for subsequent utilization, as AUTO and artificial data sets lack price information. Moreover, the variable representing the total installed base information in the OIL data set is also extracted. Next, adhering to the approach described in Haan (2021b), all lead times are uniformly established as one since some of them lack lead time information and also for the ease of computations. Next, the AUTO and OIL data set columns are reversed to ensure that they begin from the nearest period. As a result, all the data sets now commence from the earliest period. Subsequently, a modification is made to the variable names by replacing them with the corresponding number of periods. Following that, for industrial data sets, any negative demand values which can be considered as returns are replaced with zeros. To facilitate computations, any missing demand values (NAs) are substituted with zeros in industrial data sets. Next, a transpose operation is performed on all the data sets, thereby rearranging the panel structure such that while the observations correspond to time periods, the variables represent distinct spare parts from here on. A selection process is carried out solely for the industrial data sets to eliminate items that lack multiple occurrences of demand. As a result, a total of 2059 items from the MAN, and 5676 items from the OIL data set are omitted. The final number of items for MAN and OIL is 1392 and 8847 items respectively, AUTO and BRAF data sets remain unchanged. Lastly, all the data sets are split into train and test sets according to the 70-30% rule.

Outlier Detection and Handling

In this study, the Local Outlier Factor (LOF) technique is employed during the data preprocessing stage to detect outliers in industrial data sets. Subsequently, if outliers are identified, the mean imputation is utilized as a strategy for handling these outlier values. The LOF detects row-based outliers which translates to time periods in this case. The application of the LOF begins with the determination of suitable values for the parameter "k", which is the number of neighbors. By the rule of thumb for selecting the k values ($10 < k < 50$), various options, namely 20, 30, and 40, were explored to optimize this parameter. As the value of k is altered, the number of detected outlier rows also fluctuates, highlighting the dependency between the k value and the identification of outlier rows. Consequently, a value of 40 is chosen for "k", except in the case of the AUTO data set where a value of 20 is selected. This decision for AUTO data set is influenced by the fact that the it comprises only 24 months, thereby limiting the maximum selection of k to that specific value. The decision for selecting the k value equal to 40 is made with the help of total installed base information that was extracted from the OIL data set. This information is used to analyze if those rows, which are detected as outliers when the k value is 40, include any demand values that are higher than the total installed base. In the process of calculating LOF scores, it is necessary to establish a threshold that determines the classification of a row as an outlier. In general, normal instances tend to receive a score around 1 and outliers yield higher scores distinguishing them from the normal instances (Goldstein and Uchida, 2016). Hence, a threshold value of 2 is established for the LOF score. If it is higher than two, rows are identified as outliers. By this rule, no outliers are detected in the AUTO data set. On the contrary, the BRAF and OIL data sets each present two months as outliers, whereas the MAN demonstrates an exceptional 33-week outlier count, significantly surpassing the others. This difference in the number of outliers can be attributed to variations in time period units; the BRAF and OIL data sets employ months, while the MAN data set employs weeks.

During the outlier handling phase, the initial approach was the median imputation. Regarding the OIL and BRAF data sets, imputing the median values to the two identified outlier rows led to both rows being filled with zeros, as the computed median values were all zero.

Hence, instead of employing the median imputation, we used the mean imputation in this paper. Therefore, firstly column-wise mean values per item over all time were computed, and then these values were subsequently imputed into the respective outlier rows identified by the LOF technique. Upon completing the mean imputation process, the MAN, BRAF, and OIL data sets got labeled as their subsequent versions, MAN2, BRAF2, and OIL2 correspondingly. Lastly, a re-evaluation was conducted to identify items within these data sets that possess a single occurrence of demand. A total of 14 items were identified and subsequently dropped from the MAN2 data set (1378 items exist as a result). However, no alterations were made to the number of items within the BRAF2 and OIL2 data sets.

4.4 Classification of Data Sets

As mentioned earlier in Section 2.6, the framework of Boylan et al. (2008) is followed to identify demand patterns in this study. The distinct spare parts are categorized into four groups (erratic, lumpy, smooth, and intermittent) according to the mean inter-demand interval (p) and the squared coefficient of variation of the positive demand sizes (CV^2). The threshold values of 1.32 and 0.49 for the identification of demand patterns are shown in 3.6. Hereafter, two distinct tables are presented. The initial table consists of data sets that have not yet addressed outliers, while the second table comprises data sets where outliers have been addressed as per the steps outlined in Section 3.3.

To create the first table (4.4), the initial step is undertaken by computing these two classification parameters for each data set. After careful calculations, spare parts in the data sets are classified into four categories. The BRAF and OIL data sets comprise intermittent and lumpy items, with a notable prevalence of intermittent ones (58% and 77% respectively). The MAN data set exhibits a similar composition characterized by lumpy and intermittent items, accounting for 58% and 40% respectively, accompanied by a smaller proportion of erratic items of 2%. The AUTO data set is predominantly categorized as smooth with 41%; however, it also displays a considerable variation across other classifications (intermittent, erratic, and lumpy with 36%, 13%, and 10% respectively). Regarding the artificial data sets, SIM1 predominantly exhibits an erratic nature, accounting for 95% of its categorization. Similarly, SIM2 can be classified as lumpy, with an 86% representation. On the other hand, SIM3 demonstrates a predominantly smooth pattern, constituting 99% of its categorization. Lastly, SIM4 is primarily characterized as intermittent, representing 88% of its composition.

Data Sets	CV^2	p	Erratic	Lumpy	Smooth	Intermittent
AUTO	0.41	1.32	378	307	1241	1074
MAN	0.92	16.41	23	806	1	562
BRAF	0.63	11.14	0	2095	0	2905
OIL	0.57	15.15	0	2040	0	6807
SIM1	0.75	1.00	6198	0	302	0
SIM2	0.80	1.50	410	5614	25	451
SIM3	0.30	1.05	36	0	6464	0
SIM4	0.25	1.45	1	7	786	5706

Table 4.4: Demand Pattern Identifications

To create the second table (4.4), the classification parameters are computed for the altered versions of three industrial data sets, namely MAN2, BRAF2, and OIL2, excluding the AUTO data set since it does not consist of any outliers. After the outlier handling process, the MAN2 data set exhibits a higher composition of lumpiness, accounting for 97%, accompanied by a smaller proportion of erratic items of 3%. When compared to the MAN data set, The MAN2 data set displays a reduced intermittent pattern. Moreover, transitioning from MAN to MAN2 led to a significant increase in the CV^2 value and a substantial reduction in the p value. This could be attributed to imputing mean values for those outlier rows, which were previously populated mostly with zeros. As a result, the increased CV^2 metric value reflects heightened variation in demand quantities, and the decreased the p metric value indicates a decrease in the average interval between two demands, both due to the presence of non-zero demand values post-imputation. The BRAF2 data set, on the other hand, can be described as exhibiting both lumpiness (68%) and intermittency (32%). Upon comparison with the BRAF data set, it is once again evident that BRAF2 displays a reduced intermittent pattern. The OIL2 data set is predominantly categorized as lumpy with 69%; however, it also displays a considerable amount of intermittent items (31%). Similarly, there has been a decrease in the intermittent pattern observed when transitioning from the OIL data set to the OIL2 data set.

Data Sets	CV^2	p	Erratic	Lumpy	Smooth	Intermittent
MAN2	4.76	3.18	37	1336	1	4
BRAF2	0.95	8.79	0	3400	0	1600
OIL2	0.81	10.23	0	6130	0	2717

Table 4.5: Demand Pattern Identifications Second Versions

Chapter 5

Results

This chapter is structured into two primary sections. The initial main section, referred to as "exemplary results" offers thorough insights into the various algorithms employed, encompassing the selection and optimization of hyperparameters, the training process, and the generation of predictions. The subsequent main section focuses on evaluating the attained level of forecasting accuracy and the performance of inventory control for each method, as outlined in the exemplary results section. In total, this study utilizes nine distinct forecasting techniques applied to eleven different data sets. Hereafter, these two main sections will be discussed respectively.

5.1 Exemplary Results of Forecasting Methods

This section provides comprehensive elucidations in regard to all algorithms employed, encompassing the selection and optimization of hyperparameters, the training process, and the generation of predictions. A randomly selected item from the SIM4 data set, referred to as ts.12, will be utilized to exemplify and clarify the processes in greater detail. Its descriptive statistics as well as the demand pattern are given in Table 5.1. This series, ts.12, consists of demand data spanning 60 periods, which is shown in 5.2. To facilitate the implementation of various methods, it is separated into two: a training and a test set, with a ratio of 70% for the training and 30% for the test set. As a result, the forecasting horizon is set to be 18 periods which is defined by the duration of the test set. Hereafter, a step-by-step description of each method will be presented, followed by the presentation of exemplary results showcasing the forecasting accuracy and stock control measures.

Data	CVˆ2	p	Demand				Price	Classification
			Min	Mean	Median	Max		
ts.12	0.19	1.82	0	7.28	7	23	1068.22	Intermittent

Table 5.1: Descriptive Statistics of Part ts.12

Period	Demand																			
	ts.12																			
1	12	9	0	0	0	0	8	0	14	13	7	8	0	6	23	19	10	13	0	4
21	18	0	0	17	7	11	0	0	4	0	6	14	0	16	15	0	4	7	11	9
41	7	7	10	11	0	2	23	0	11	11	17	2	8	5	0	13	6	6	4	9

Table 5.2: Demand Values of Part ts.12

SES, Croston, SBA and TSB

To begin with, SES requires the input of several parameters, such as the forecast horizon, smoothing parameters, initial values for demand, interval size, and the selected cost function for optimization. The input data for SES begins with the 42 values from the ts.12 training data. In this case, the forecast horizon is set to 1, resulting in lead-time demand (LTD) forecasts. The optimization of the smoothing parameters is carried out using the Mean Absolute Rate (MAR) cost function, given in the appendix as equation (1), which has been identified by Kourentzes (2014) as the optimal choice for Croston's method and is used as the cost optimization function for SES, Croston, and SBA methods. However, as explained in Kourentzes (2014), TSB diverges from the others by employing the Mean Squared Rate (MSR) as the most suitable cost function. The initial demand for SES is derived from the first observed non-zero demand, while the initial interval is set to the value of the first interval. Given the comparable approach and simplicity of SES, Croston, SBA, and TSB, the process configurations remain consistent across these methods. The function iterates over each item and forecasts the next period's value using these methods based on the available historical data in the training set. The loop then increments the period used for prediction by 1 and continues to the next iteration to forecast the subsequent period. This process is repeated until all periods for all items in the test data are predicted.

Willemain's Bootstrapping Method (WSS)

Willemain's bootstrapping methodology (Willemain et al., 2004) commences by converting the training data into a binary format, where the absence of demand is denoted by 0, while periods exhibiting positive demand are represented by 1. The method proceeds with the computation of the transition probability matrix, which characterizes the likelihood of the series transitioning between distinct demand states. In the case of ts.12, the transition probabilities can be summarized as follows: If the preceding demand was zero, there is a 35.7% probability that the subsequent demand will also be zero, and a 64.3% probability of positive demand. If the preceding demand was positive, there is a 33.3% probability of a demand-free period and a 66.7% probability of another positive demand. Utilizing this transition matrix, a sequence comprising zeroes and non-zeroes is generated to cover the forecast horizon, which corresponds to the lead time. Subsequently, the method involves the prediction of the lead-time demand (LTD) by leveraging the transition matrix and the last observation from the training data. This prediction process is iteratively performed, repeated 1000 times, with the objective of obtaining a comprehensive distribution of projected LTD values. Next, all positive demand values are substituted with random positive demand values drawn from the training data. Subsequently, a jittering process is applied to both the predicted and substituted values, simulating natural variations in demand. Consecutively, the predicted and jittered values are accumulated over the forecast horizon, leading to a singular LTD value. Lastly, the mean and standard deviation of the ensemble of LTD values is derived.

Multi-Layer Perceptron (MLP)

The initial step in the neural network approach involves normalizing the data within the range of 0 to 1 according to the equation (3.5). It is worth noting that this normalization process is the same across all ML methods employed. Next, a rolling window approach is applied to the training data, as suggested by Haan (2021b) and Smyl (2020), a window size of 5 periods as input is selected for several reasons, such as it provides sufficient information for ML methods to capture underlying dependencies, but choosing a larger window could lead to underperformance of the methods. Thus, the first 5 periods of the training data are saved as input data, while the 6th period serves as the first output data point. The decision to have an output window equal to 1 is not only due to matching the lead time but also to ensure a comparison between the MLP method and the statistical methods (Haan, 2021b). The window is then shifted by one period, with periods 2-6 becoming the input data and period 7 as the output data. This process continues until no additional input data is available, creating a new data set for each item. In this new data set, the first 5 columns represent the input, while the 6th column represents the corresponding output.

The subsequent stage involves training the NN model using the newly constructed data. This study utilizes the MLP by comprising 6 hidden layers (*size*), consistent with the work conducted by Haan (2021b). In order to identify the optimal values for specific hyperparameters, as described in Chapter 3, the grid-search CV method is employed. The optimal value for the maximum number of iterations (*maxit*) is determined to be 100 from the given set of values (100, 250, 500, 1000), which aligns with Spiliotis et al. (2020). This indicates that setting the parameter *maxit* to 100 results in the lowest RMSE value. The optimal value for the learning rate is determined to be 0.1, from the given set of values (0.01, 0.1, 0.2) and the maximum output difference value is set to the default, which is 0. Finally, the default activation function for the hidden layer, *Act_Logistics*, is utilized, while for the learning function, the standard backpropagation method is chosen as the optimal option to estimate the network weights (*learnFunc*). This decision is made after considering alternative approaches such as scaled conjugate gradient (SCG) and weight decay backpropagation aligning with the work of Spiliotis et al. (2020). Using the identified optimal hyperparameter values, the NN model is trained, and predictions are generated accordingly. As the final step, denormalization is applied to these predictions in order to transform them into representative demand values. The resulting prediction values from the MLP model can be seen in Table 5.3 for ts.12.

LightGBM

To begin the LightGBM method, the data is first adjusted to fit within the 0 to 1 range. Next, a rolling window technique is utilized on the training data, where the initial five periods of the training data are preserved as input, and the sixth period is taken as the output data point. There are three hyperparameters that are optimized by the grid-search CV for this algorithm. The learning rate (*learning_rate*) is determined from a set of values as 0.001, 0.01, 0.05, 0.1, and 0.5, while the maximum tree depth *max_depth*) ranges from 20 to 100, increasing by 10. Additionally, the maximum number of leaves for each weak learner (*num_leaves*) ranges from 10 to 130, increasing by 10. The optimal parameter values are as follows: the learning rate is set to 0.001, the maximum depth is set to 20, and the number of leaves is set to 20. The rest of the parameter configurations are made according to Kailex (2020). The algorithm undergoes multiple training iterations until it achieves the minimum RMSE value. Once the model is trained, it becomes capable of forecasting the next period based on the preceding five periods and the process of generating subsequent period predictions continues until the forecasting horizon is reached. Lastly, the forecasted values are denormalized and the resulting predictions can be seen in 5.3.

The forecasted values for the last 18 periods for ts.12 with the lightGBM algorithm exhibited the same pattern. Through several iterations of diverse lightGBM models employing different parameter configurations, it has been observed that training the algorithm with a single item consistently generates identical predictions for all periods. However, when multiple items are used, the predictions for the last 18 periods do not consistently yield the same value. Also, during the evaluation of the forecasting accuracy performance, the occurrence of identical predictions is not considered problematic when compared with other techniques in terms of forecasting accuracy measures.

Random Forest (RF)

The initial step in the RF model is again normalizing the data within the range of 0 to 1. In a similar vein to the previous ML algorithms, the rolling window approach is employed. Thus, the algorithm trains itself with the initial five columns, while the sixth column acts as the corresponding output for the given input values. There are several hyperparameters that are optimized by using the grid-search CV. The selection process involved determining the appropriate number of non-pruned trees (*ntree*) from a set of options from 100 to 1000 increasing by 100. Similarly, the minimum size of terminal nodes (*nodesize*) is chosen from alternatives including 5, 10, 100, 250, and 500. Additionally, the number of variables randomly sampled as potential candidates at each split (*mtry*) is selected from 1,2,3,4,5. These parameter choices are in accordance with the values chosen by Spiliotis et al. (2020) in their comparative study. The hyperparameter values that result in the most accurate predictions, in terms of RMSE, are as follows: *ntree* is set to 1000, *nodesize* is set to 100 even though the 100, 250, and 500 provides the equally lowest value and *mtry* is set to 1 since all values provide the equally lowest value for *mtry*. Despite the optimal *ntree* value being 1000, the default value of 500 is employed due to the excessively lengthy execution duration of the RF algorithm. Once the model is trained with the optimal hyperparameters, it is employed to forecast the subsequent period by leveraging the information derived from the first five periods. Consequently, the predictions are generated for the subsequent periods until the conclusion of the forecasting horizon and they get denormalized in a similar way. The resulting prediction values can be seen in Table 5.3 for ts.12.

Support Vector Regression (SVR)

The Support Vector Regression (SVR) model begins by normalizing the data within the range of 0 to 1 and then employs the rolling window technique in a similar way for the training process.

As suggested by Spiliotis et al. (2020), nu-regression is employed to enhance computational efficiency and reduce complexity for the SVR model. The optimal values of hyperparameters are found by the grid-search CV. During the optimization phase, the cost (C) is assigned values of 0.001, 0.01, 0.1, 1, and 10, while the epsilon is set to 0.001, 0.01, 0.1, and 1, where the optimal values for both are determined to be 1. The termination criterion tolerance *(tolerance)* is chosen from a set of values including 0.001, 0.01, and 0.1. Additionally, the value of nu (*nu*) ranged between 0.3 and 0.7. The optimal configuration for the most accurate model entails setting the nu to 0.3 and the tolerance to 0.001. These sets of values are all in alignment with the study of Spiliotis et al. (2020). The SVR method offers various options for the kernel function (*kernel*) during the training and prediction processes. These include linear, polynomial, radial basis, and sigmoid kernels. The polynomial kernel provides the most accurate results in terms of RMSE in this case. When training the model with the optimal hyperparameters, it utilizes the data from the initial five periods to forecast the subsequent period, and it again continues until the end of the forecasting horizon, generating predictions that are subsequently denormalized in a similar manner. The resulting prediction values can be seen in Table 5.3 for part ts.12.

5.2 Exemplary Results of Performance Measures

Once the training procedure is complete for all forecasting techniques, the predictions are generated and denormalized, yielding forecasted demand values for the ts.12 series, which are presented in Table 5.3. The table shows that each method produces 18 predictions, with the initial row displaying the actual demand values from the last 18 periods of the part ts.12 data. The presented table illustrates that despite the wide range of demand sizes observed in the test set of ts.12, ranging from 0 to 23, the predicted demand values derived from all the methods employed in this paper exhibit a narrower range of variations, generally falling within the range of 5 to 9. Hereafter, the performance of all the forecasting methods will be examined according to two different aspects: initially forecast accuracy measures and subsequently stock control measures.

Exemplary Results of Forecasting Accuracy Measures

The forecasting accuracy measures, including Mean Absolute Scaled Error (MASE), and Root Mean Squared Scaled Error (RMSSE), are utilized to conduct a comparative analysis for this study. The results, presented in Table 5.4, reveal the performance of each method in relation to these metrics. A lower value indicates a better performance for both of the measures. The

Methods	Period																	
	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
ts.12	10	11	0	2	23	0	11	11	17	2	8	5	0	13	6	6	4	9
Croston	8.04	8.09	8.33	8.20	6.53	8.36	8.36	7.71	8.09	8.90	8.21	8.18	7.78	7.75	7.56	7.39	7.20	6.88
SES	7.40	7.78	8.24	7.05	6.31	8.54	7.46	7.90	8.28	9.11	8.34	8.30	7.96	7.14	7.74	7.58	7.43	7.11
SBA	8.03	8.12	7.94	7.94	6.12	7.70	7.70	7.13	7.45	8.10	7.86	7.89	7.70	7.65	7.30	7.27	7.20	7.07
TSB	7.78	8.13	8.56	7.18	6.60	8.60	7.46	7.91	8.31	9.06	8.45	8.44	8.14	7.27	7.81	7.69	7.56	7.26
WSS	7.82	7.83	8.27	7.51	7.99	8.22	7.55	7.79	8.06	7.77	8.14	8.16	8.07	7.68	8.09	7.93	8.02	8.70
MLP	7.55	7.55	7.52	7.57	7.64	7.49	7.51	7.60	7.44	7.55	7.49	7.60	7.54	7.66	7.55	7.52	7.59	7.55
LightGBM	7.51	7.51	7.51	7.51	7.51	7.51	7.51	7.51	7.51	7.51	7.51	7.51	7.51	7.51	7.51	7.51	7.51	7.51
RF	5.16	7.14	8.29	8.46	7.51	6.79	8.26	9.42	3.99	9.14	5.09	5.74	4.13	9.07	8.82	8.32	8.87	4.58
SVR	7.50	7.53	7.50	7.50	7.54	7.54	7.48	7.56	7.32	7.81	7.44	7.68	7.45	7.59	7.49	7.47	7.50	7.46

Table 5.3: Comparison of Test Set of Part ts.12 and Predicted Demand Values

method that performs the best for each metric is highlighted in bold fonts. In terms of MASE and RMSSE, the MLP method outperforms the others by providing the lowest values for ts.12 series.

Methods	MASE	RMSSE
Croston	0.708	2.377
SES	0.725	2.403
SBA	0.702	2.375
TSB	0.724	2.398
WSS	0.693	2.286
MLP	**0.686**	**2.262**
LightGBM	0.687	2.267
RF	0.752	2.440
SVR	0.691	2.276

Table 5.4: Forecasting Accuracy Measures Exemplary Results of Part ts.12

Exemplary Results of Stock Control Measures

The subsequent assessment focuses on the stock control performance of each method. Initially, the predictions made by each method are used to determine the base stock levels R for each target fill rate. Secondly, the calculation of holding costs and achieved fill rates associated with these base stock levels follows. Figure 5.1 presents the results corresponding to each target fill rate based on the predictions made for ts.12. It is essential to highlight that these tradeoff curves are specifically for a single item (ts.12). It is observed that the achieved fill rate surpasses the target fill rate for all utilized methods, which reveals that these methods tend to overestimate the anticipated demand quantity and in the end, this leads to too much stock in hand. Among all Random Forest (RF) tends to overestimate the anticipated demand the least with the lowest achieved fill rates for each target fill rate, and it is followed by the LightGBM model. Regarding the inventory holding costs associated with each achieved fill rate, all methods exhibit a similar trend of gradual cost increase as the fill rates increase.

Figure 5.1: Tradeoff Curves of Stock Control Performance Measures for ts.12 Predictions

5.3 Main Results

According to the configurations outlined in the exemplary results section above, all the forecasting methods are utilized on every data set. To be precise, a total of 9 forecasting techniques are applied to 11 distinct data sets. Furthermore, the execution times for each method applied on all data sets are measured and recorded in Table 5.5, allowing for an assessment of their ease of implementation, which will be discussed in the next chapter. For the execution times, the elapsed time which means the total duration elapsed in real life, is used. However, it is worth noting that the running time of these methods is contingent upon the computational capabilities of the computer employed, and all the computations for all methods, except the TSB, of this paper have been conducted using a standard Windows computer with Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz, 8GB RAM. *For the TSB method, the computer used has the following configurations, Apple M2 Pro, 16GB RAM. Hereafter, in this main results section, the discussion will begin with the level of forecasting accuracy performance achieved, and continue with evaluating the results of the inventory control performance for each model.

Forecasting Accuracy Main Results

By employing all the methods on full data sets, the comprehensive outcomes of forecasting accuracy are derived, relying upon the chosen accuracy metrics, namely MASE and RMSSE. Table 5.7 provides these accuracy findings for each method and for all data sets. Additionally, in the last column, the average Percentage Better score is computed following the guidelines outlined in Pinçe et al. (2021). This metric quantifies the frequency with which a particular

Method	Total Elapsed Run Time (min)	Total Elapsed per Item (sec)
SES	9.62	0.010
Croston	114.41	0.115
SBA	135.63	0.137
TSB*	36.20	0.037
WSS	1857.82	1.875
MLP	318.70	0.322
LightGBM	92.12	0.093
RF	1402.38	1.415
SVR	393.53	0.397
Total	**4360.41 = 72.67 hours**	
*Computer configurations are different for TSB, as explained in the text.		

Table 5.5: Total Elapsed Run Time for Each Forecasting Method on All Data Sets

method outperformed another, expressed as a percentage of the total number of comparisons made.

To begin with the simulated data sets, SBA exhibits superior performance over all other methods in terms of both metrics when considering the SIM3 data set, which showcases a smooth demand pattern. Furthermore, in the case of the SIM1 data set, which presents an erratic demand pattern, SBA outperforms other methods in terms of the MASE metric. Conversely, when analyzing SIM2 and SIM4 data sets, characterized by mainly lumpy and intermittent demand patterns respectively, MLP demonstrates a superior performance across both metrics. In addition, it provides the most accurate model in terms of RMSSE with SIM1 data.

For the initial versions of the industrial data sets (AUTO, MAN, BRAF, and OIL), where no outlier detection and handling method is implemented, three methods exhibit superior performance. Primarily, for the AUTO data set, SBA surpasses other methods, except for SES, in terms of both metrics, by demonstrating higher accuracy values. Furthermore, SES exhibits superior performance specifically for the MAN data set when assessed using the RMSSE metric. Lastly, the SVR algorithm emerges as the most effective method across both metrics for the BRAF and OIL data sets. Additionally, when considering the MAN data set, SVR outperforms other methods in terms of the MASE metric. Despite benchmarking methods being specifically tailored to address intermittent demand, it can be seen that SVR displays a superior performance for most of the real data sets characterized as intermittent and lumpy with the incorporation of hyper-parameter optimization.

For the second version of real data sets (referred to as MAN2, BRAF2, OIL2), wherein the outlier detection and mean imputation for outlier handling are implemented, SVR emerges as the top-performing technique for both the BRAF2 and OIL2 data sets, which both exhibit mainly lumpy and intermittent pattern. However, in the case of the MAN2 data that is mainly lumpy, SBA and SES exhibit better performance than other methods in terms of the MASE and RMSSE metrics, respectively. Moreover, an assessment is carried out, as presented in Table 5.6, to determine whether the application of an outlier detection and handling procedure has led to enhanced accuracy for each method. The evaluation compares the forecasting performance of each method with the two different versions of the data sets, for example, MAN is compared with MAN2, and BRAF is compared with BRAF2. If one outperforms the other, it is indicated with a "check" symbol. Consequently, all methods with the exception of WSS and SVR for the MAN and BRAF data sets, have demonstrated improvements when detected outliers are handled by the mean imputation, thus the MAN2 and BRAF2 data sets exhibit better performance than the first versions of themselves respectively across both measures. Specifically, the WSS method does not exhibit improvements in terms of the MASE metric for the same data sets, while MAN2 does not display a higher accuracy for the SVR method in terms of the MASE. For the comparison of the OIL and OIL2 data sets, only three methods, specifically MLP, LightGBM, and RF, demonstrate improved performance based on the RMSSE metric. It is shown that the OIL2 data set has lower RMSSE values with these methods, indicating enhanced accuracy in comparison to the OIL data set, however, when considering the MASE metric, no method demonstrates any improvements in accuracy when the outlier detection and handling procedure is applied to the OIL data set.

Lastly, the Percentage Better results are computed by comparing the performance of each method with the rest of the given methods for each metric and data set. This comparison is based on the achieved forecasting accuracy and is conducted column-wise. Subsequently, row-wise averages are calculated to evaluate the overall performance of each method. SBA outperforms the other methods on average across both measures. Conversely, the LightGBM method consistently exhibits the poorest performance overall. It is noteworthy to state that these average percentage better results are consistent with the conclusions drawn by Haan (2021b) despite the inclusion of two other ML methods in this study.

Method	Measure	Data Set					
		MAN	MAN2	BRAF	BRAF2	OIL	OIL2
SES	MASE		✓		✓	✓	
	RMSSE		✓		✓	✓	
Croston	MASE		✓		✓	✓	
	RMSSE		✓		✓	✓	
SBA	MASE		✓		✓	✓	
	RMSSE		✓		✓	✓	
TSB	MASE		✓		✓	✓	
	RMSSE		✓		✓	✓	
WSS	MASE	✓		✓		✓	
	RMSSE		✓		✓	✓	
MLP	MASE		✓		✓	✓	
	RMSSE		✓		✓		✓
LightGBM	MASE		✓		✓	✓	
	RMSSE		✓		✓		✓
RF	MASE		✓		✓	✓	
	RMSSE		✓		✓		✓
SVR	MASE	✓			✓	✓	
	RMSSE		✓		✓	✓	

Table 5.6: Forecasting Performance Comparison for Two Different Versions of Data Sets

Inventory Control Performance Main Results

An evaluation of inventory performance is conducted, and to illustrate the performance level, trade-off curves are constructed individually for each data set. These curves present two distinct trade-offs. Firstly, they demonstrate the trade-off between the average achieved fill rates over all items by the method and the associated holding costs, denoted by label (a). Secondly, the other trade-off curve illustrates the trade-off between the achieved fill rate and the predetermined target fill rate, labeled as (b). A higher fill rate results in increased holding costs, with the figures reaching their maximum at the 99.99% fill rate level, which approximates a 100% fill rate for the purposes of this study. In addition, the second curves display a gradual rise in the achieved fill rate until it reaches the 99.99% target fill rate once again. The separate analysis of trade-off curves is performed to evaluate the performance of the employed forecasting method within the specific characteristics of each data set. It is worth acknowledging that the trade-off curves of some data sets, which yield similar results, are placed in the appendix. To better capture the inventory control performances of different methods for each data set, Table 5.8 has been generated, which shows the ranking of average achieved fill rate values over all target fill rates of all methods for each data set, from 1 to 9, 1 being the highest fill rate and 9 being the lowest one. It can be seen that the WSS achieves the highest average fill rates in 10 out of 11 data sets.

Method	Measure	Data Set											Percentage
		SIM1	SIM2	SIM3	SIM4	AUTO	MAN	BRAF	OIL	MAN2	BRAF2	OIL2	Better (Avg)
SES	MASE	0.673	1.028	0.488	0.783	0.780	2.335	1.996	1.410	1.703	1.966	1.836	71.59%
	RMSSE	2.722	3.358	1.883	2.424	**1.710**	**5.215**	3.283	1.404	**4.046**	3.257	1.651	72.73%
Croston	MASE	0.673	1.027	0.487	0.780	0.788	2.499	2.080	1.799	1.667	2.069	2.158	64.77%
	RMSSE	2.722	3.346	1.878	2.412	1.721	5.329	3.300	1.651	4.071	3.270	1.752	57.95%
SBA	MASE	**0.664**	1.012	**0.484**	0.778	**0.777**	2.439	2.001	1.659	**1.664**	1.974	2.058	**80.68%**
	RMSSE	2.712	3.337	**1.874**	2.409	**1.710**	5.304	3.289	1.622	4.070	3.254	1.729	**76.14%**
TSB	MASE	0.675	1.032	0.490	0.785	0.792	2.321	1.956	1.420	1.698	1.904	5.591	60.23%
	RMSSE	2.731	3.363	1.887	2.430	1.728	5.223	3.287	1.471	4.053	3.254	2.522	52.27%
WSS	MASE	0.689	1.043	0.498	0.783	0.906	2.537	2.307	2.046	3.408	2.341	2.682	20.45%
	RMSSE	2.737	3.353	1.889	2.420	1.875	5.319	3.376	1.536	4.347	3.344	1.831	31.82%
MLP	MASE	0.669	**0.999**	0.518	**0.773**	0.820	3.055	2.274	1.879	2.307	2.264	2.436	46.59%
	RMSSE	**2.701**	**3.308**	2.069	**2.386**	1.735	5.468	3.347	1.976	4.296	3.330	1.774	47.73%
LightGBM	MASE	0.896	1.367	0.714	0.994	1.156	3.258	2.389	1.964	2.493	2.381	2.376	5.68%
	RMSSE	3.561	4.421	2.756	3.162	2.652	6.369	3.464	1.992	5.231	3.459	1.800	2.27%
RF	MASE	0.682	1.039	0.491	0.777	0.808	2.926	2.327	1.931	1.905	2.324	2.337	40.91%
	RMSSE	2.712	3.338	1.876	2.394	1.724	5.500	3.380	1.944	4.260	3.340	1.759	51.14%
SVR	MASE	0.694	1.075	0.510	0.815	0.840	**1.953**	**1.174**	**0.767**	1.988	**1.147**	**1.095**	56.82%
	RMSSE	2.730	3.365	2.031	2.515	1.768	5.391	**3.254**	**1.295**	4.281	**3.229**	**1.571**	53.41%

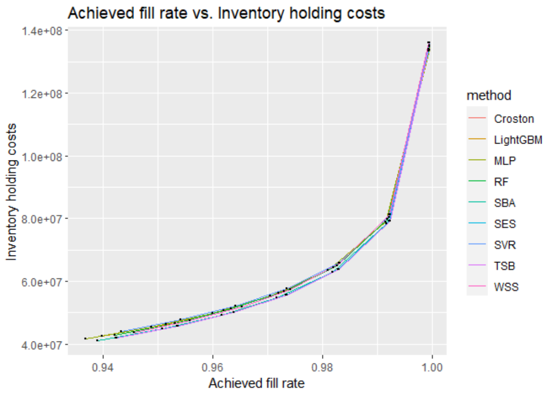
Table 5.7: Forecasting Accuracy Measure Results for Each Method and Each Data Set

Data	Methods								
	Croston	SES	SBA	TSB	WSS	MLP	LightGBM	RF	SVR
SIM1	5	4	8	3	1	9	7	6	2
SIM2	4	5	7	3	1	9	8	6	2
SIM3	3	4	6	2	1	8	9	5	7
SIM4	4	5	7	3	1	9	8	6	2
AUTO	4	6	8	3	1	2	9	5	7
MAN	3	2	5	4	1	8	7	6	9
BRAF	2	5	6	8	1	7	4	3	9
OIL	2	3	4	5	1	8	7	6	9
MAN2	4	3	5	2	1	7	8	6	9
BRAF2	2	7	6	8	1	5	4	3	9
OIL2	3	4	5	1	2	6	8	7	9

Table 5.8: Average Achieved Fill Rate Ranking over All Target Fill Rates for Each Data Set

a. Stock Control Performance of SIM1

The trade-off curves for the SIM1 data set are given in Figure 5.2. The figure's part (a) demonstrates closely clustered outcomes across all methods for inventory holding costs. Part (b) states that the WSS method achieves the highest fill rate across all target fill rate levels. The MLP method exhibits the lowest achieved fill rates for each target fill rate, followed by the SBA method. However, as the target fill rate increases to 99%, the SVR method achieved higher fill rates, while the MLP method continues to perform lower achieved fill rates for every target fill rate. Overall, due to fluctuations in performance, it is not possible to identify a single best-performing approach among the methods for the SIM1 data. In terms of forecasting accuracy, the SBA and MLP methods demonstrated superiority based on the MASE and RMSSE measures, respectively. These findings of the performance comparisons prove that relying solely on forecasting accuracy measures for practical relevance is not sufficient.

(a)


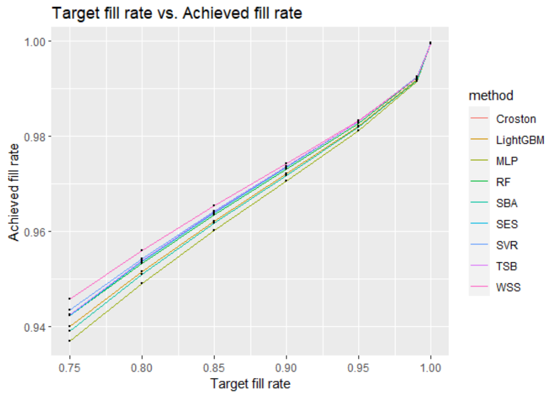
(b)


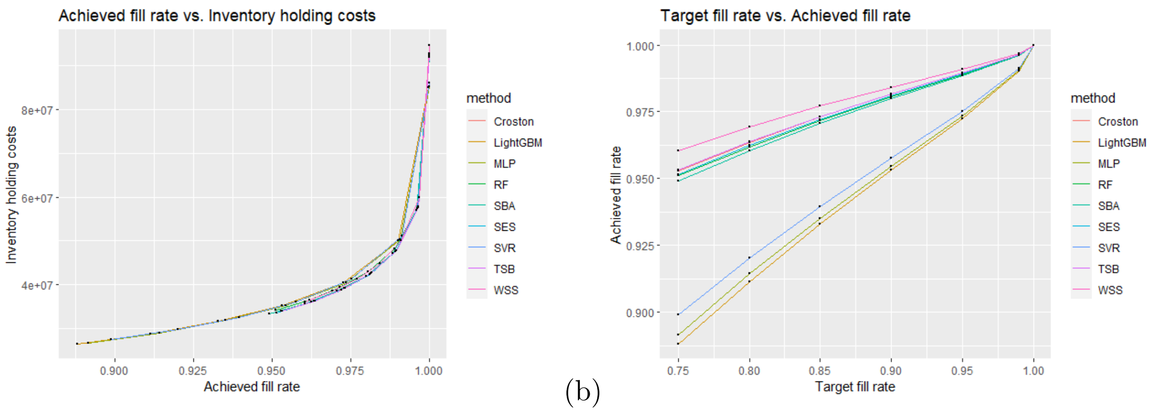
Figure 5.2: Tradeoff Curves of Stock Control Performance Measures for SIM1 Predictions

a. Stock Control Performance of SIM2

The SIM2 data set follows mainly a lumpy demand pattern. Trade-off curves for the SIM2 data are given in the appendix in Figure 1 since it presents similar outcomes to the SIM1 data. The first trade-off curve (part a) demonstrates clustered outcomes across all methods when assessing inventory holding costs. In part (b), the WSS method consistently achieves the highest fill rates with a big margin at every target fill rate and is followed by the SVR method. Conversely, the MLP exhibits the lowest achieved fill rates across all target fill rates and is followed by the LightGBM. The inclusion of zero values and the increased variability in demand pattern, when compared to the SIM1 data set, does not appear to have significantly impacted the stock control performance of the WSS and MLP methods for both SIM1 and SIM2 data sets. In order to address the first research question, although both demonstrate similar outcomes in terms of stock control, their demand patterns do not show similar characteristics. Lastly, the MLP method demonstrates superior performance in terms of the MASE and RMSSE measures, the current findings reveal its deficiency in achieving high fill rates across all target fill rates.

a. Stock Control Performance of SIM3

In Figure 5.3 below, the trade-off curves for the SIM3 data set are given. In part (a), various methods are relatively similar in terms of the holding costs and the achieved fill rates. In the second curve, there is a distinct segregation, which illustrates that the WSS achieves the highest fill rates for each target fill rate. The TSB and Croston methods follow closely as second and third respectively, while the remaining methods are grouped closely. Conversely, the LightGBM shows the lowest achieved fill rates for each target fill rate and is followed by the MLP and the SVR respectively, which means they overestimate the anticipated demand to a lesser extent. The presence of a smooth pattern, indicating a limited degree of demand variability and low occurrence of intermittency, appears to have influenced the stock control performances of the TSB, Croston, LightGBM, and SVR methods. The superior performance of SBA in terms of forecasting accuracy for the SIM3 data conflicts with the inventory control outcomes. This indicates that these two types of measures hold independent significance to determine the most suitable method.



(a) (b)

Figure 5.3: Tradeoff Curves of Stock Control Performance Measures for SIM3 Predictions

a. Stock Control Performance of SIM4

The trade-off curves for the SIM4 data showcase similar outcomes to the SIM2 data, thus the curves are given in Figure 2 in the appendix. In part (a), all the methods demonstrate a minor disparity in terms of the holding costs for achieved fill rates. In part (b), the WSS method shows again the highest achieved fill rates for each target fill rate. Similarly to the SIM2 results, the MLP provides the lowest achieved fill rates. Although SIM2 and SIM4 data sets' stock control performances are in alignment, their demand characteristics show different patterns (lumpy and intermittent respectively). Thus, there is no consistent behavior in the methods' performances to indicate which method is best for which type of data in this case. In terms of forecasting accuracy, the MLP method had superior performance across all metrics. However, this outcome does not correlate with the inventory control performance in this case.

a. Stock Control Performance of AUTO

The AUTO data set's trade-off curves show similarity to the inventory control performance of the SIM3 data, thus these curves are given in Figure 3 in the appendix. In part (a), the LightGBM method exhibits the worst performance in terms of high holding costs for every achieved fill rate, while the remaining methods display relatively similar performance. In part (b), the LightGBM achieves the lowest achieved fill rates across all target fill rates similar to the SIM3 data set. In contrast, the WSS method demonstrates higher achieved fill rates with each target fill rate, leading to overestimating the demand to a bigger extent. Since the LightGBM achieves the lowest fill rates and higher holding costs for each target fill rate, it shows the poorest performance. Thus, in addressing the first research question, it becomes evident that despite the predominantly smooth demand patterns in both the AUTO and SIM3 data sets, no consistent performance of the methods emerges in terms of inventory control. Regarding the forecasting accuracy results, the SBA model emerged as the top-performing method based on both metrics. However, this outcome contradicts the stock control performance results of the SIM3 data.

a. Stock Control Performance of MAN

In Figure 5.4 below, the trade-off curves for the MAN data set are given. The first trade-off curve reveals a consistent pattern where the utilization of the SVR method is at first linked to lower holding costs at the expense of lower achieved fill rates. This trend is also evident in the second curve, where the SVR method occupies the lowest position compared to other methods, indicating its tendency to achieve lower fill rates for each target fill rate. The remaining methods, however, are closely grouped together in part (a). In contrast, the WSS model occupies the highest position on the curve, indicating its ability to achieve higher fill rates for every target fill rate and having average inventory holding costs for each achieved fill rate. These findings are in contrast with the forecasting accuracy metrics, as the SVR and SES models provided the highest levels of accuracy in terms of the MASE and RMSSE measures, respectively.

a. Stock Control Performance of BRAF

The BRAF data set consists of mainly intermittent and lumpy items, demonstrating a high mean duration between demands (p:11.14) along with both substantially and slightly fluctuating demand sizes. The trade-off curves of the BRAF data exhibit similarities to the MAN data set and thus are placed in the appendix in Figure 4, which is as expected since both of them showcase similar demand characteristics. In part (a), all the methods are clustered close

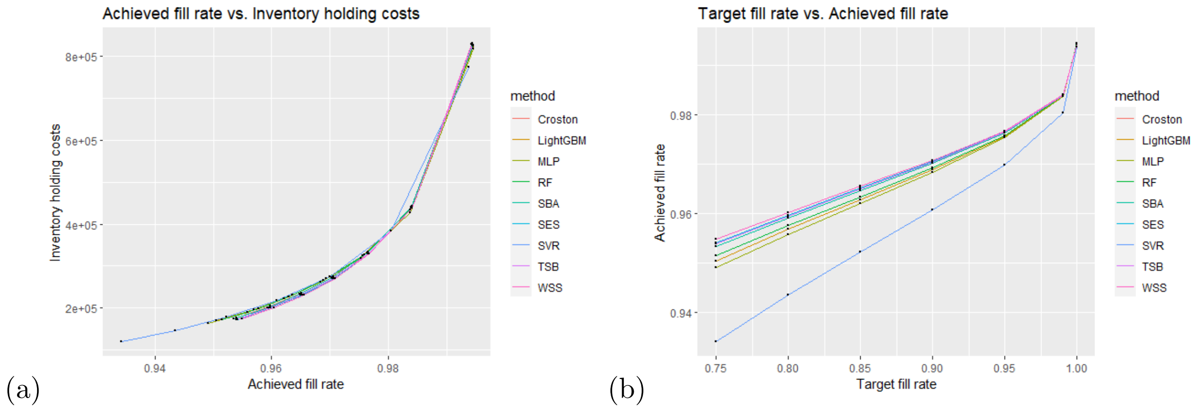


Figure 5.4: Tradeoff Curves of Stock Control Performance Measures for MAN Predictions

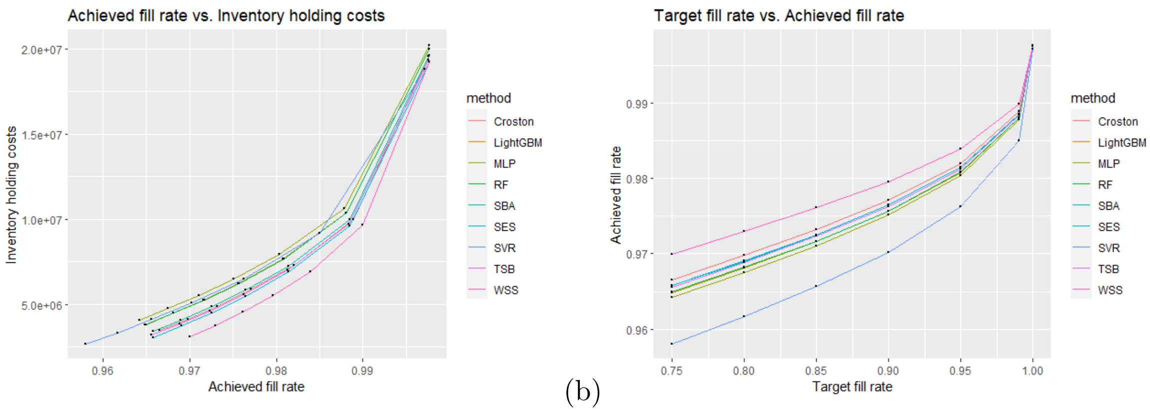
to each other except the SVR which is linked to the lowest holding costs for every achieved fill rate. In part (b) the WSS is connected to higher achieved fill rates for every target fill rate, which means overestimating the anticipated demand to a higher extent. Conversely, the SVR overestimates the demand to a lesser extent by achieving the lowest achieved fill rates for each target fill rate, but still overpassing every target rate. In addressing the first research question, it can be concluded that for mainly lumpy and intermittent demand patterns, the SVR method shows similar inventory control performance. Lastly, these findings do not fully align with the forecasting accuracy results, where the SVR was the best-performing method for all accuracy measures.

a. Stock Control Performance of OIL

In Figure 5.5 below, the trade-off curves for the OIL data set are given. In both trade-off curves, a clearer separation of the method performances can be seen. It is shown that the WSS method achieves lower holding costs with higher achieved fill rates, also achieving higher fill rates for each target fill rate. Thus, it outperforms other methods in terms of inventory control for this data set. On the other hand, the SVR is associated with lower achieved fill rates for relatively high holding costs. It is followed by the MLP method, which exhibits lower achieved fill rates for each target fill rate and higher holding costs for each achieved fill rate. Consequently, the WSS method emerges as the superior approach, while the SVR and MLP methods show the poorest performances among other utilized methods. Therefore, the superior performance of the more straightforward methods shows that in data sets characterized by high intermittency and lumpiness, the straightforward estimation approach tends to outperform the understanding of underlying dependencies offered by ML techniques, specifically the SVR and the MLP methods.

a. Stock Control Performance of MAN2

Transitioning to the second version of industrial data sets, the MAN2 data set is characterized by mainly lumpy behavior. The trade-off curves for the MAN2 data set are given in Figure 5 in the appendix since the trade-off curves of both the MAN and MAN2 data sets show similarity. It can be seen from the curves that while the WSS method outperforms all techniques, achieving higher fill rates and lower inventory holding costs for each target fill rate, the SVR method is the poorest-performing one as achieving lower fill rates and higher inventory holding costs for each target fill rate. Although the MAN and MAN2 data sets have similar demand patterns (both are mainly lumpy), the WSS and the SVR methods' performances showcase differences in terms of inventory control. When comparing the forecasting accuracy, the SBA and SES methods outperform others in terms of both measures. Thus, there is no alignment between the findings of forecasting accuracy and inventory control performance for the MAN2 data set.



(a) (b)

Figure 5.5: Tradeoff Curves of Stock Control Performance Measures for OIL Predictions

a. Stock Control Performance of BRAF2

Figure 6 in the appendix presents the trade-off curves of the BRAF2 data, which can be described as mainly lumpy. It can be seen that the WSS method achieves higher fill rates to be associated with higher holding costs for every target fill rate. In addition, the SVR method exhibits the lowest holding costs to be associated with the lowest achieved fill rates across all target fill rates. Taking into account the implementation of the outlier detection and handling procedure, resulting in an increased CV^2 and a decreased p value, it appears that this procedure has not had a substantial impact on the outcomes of inventory control performance. In addressing the first research question, when there are mainly intermittent and lumpy items, the inventory control performance of the SVR method yields similar results. Lastly, the stock control performance findings contradict the forecasting accuracy results of the BRAF2 data set, which displayed that the SVR method outperformed all the other methods for both accuracy measures.

a. Stock Control Performance of OIL2

Lastly, in Figure 7 in the appendix, the trade-off curves for the OIL2 data set are given. The OIL2 data set is categorized as both lumpy with 69% and intermittent with 31%. There has been a decrease in the intermittent pattern and an increase in the lumpy pattern when transitioning from the OIL data set to the OIL2 data set. The TSB method achieves the highest holding costs to be associated with the highest achieved fill rates for each target fill rate for the OIL2 data set. On the other hand, the SVR method shows the lowest holding costs that are linked with the lowest achieved fill rates for every target fill rate. It can be concluded that the SVR method's inventory performance results yield similar outcomes when data sets exhibit mainly intermittent and lumpy demand patterns. Lastly, the stock control performance findings again contradict the forecasting accuracy results, which displayed that the SVR method outperformed all the other methods for both accuracy measures for the OIL2 data set.

Chapter 6

Conclusion and Discussion

In this section, firstly a series of findings will be introduced and discussed thoroughly. Next, these findings will be compared with the results reported in the existing academic literature. After presenting all of the findings, a connection will be established between the findings and the research questions that are listed in the paper's introduction. Lastly, a comprehensive conclusion will be provided, accompanied by discussion points on the present paper and recommendations for future research endeavors.

6.1 Findings

Finding 1: *The implementation feasibility of the utilized and evaluated methods varies in terms of the overall execution duration and the level of expertise necessary.*

The execution times for each method are measured and recorded in Table 5.5, in terms of the total elapsed run time from the beginning of the execution until the end in minutes and the total elapsed run time per item in seconds. As can be seen, the SES exhibited the shortest execution duration, whereas the WSS displayed the longest, primarily attributable to the iterative nature of the bootstrapping technique for 1000 times in this case. The RF model, surprisingly for a machine learning approach, exhibits the second longest execution duration. RF is a methodology that entails the consolidation of numerous complex decision trees to address the challenge of overfitting associated with individual decision trees. The long execution duration can be attributed to this consolidation process, specifically in our case, involving 500 decision trees, to yield a singular result. The remaining methods showcase an average execution time of 6.5 hours or less for all data sets. However, when it comes to the level of expertise needed for the implementation of the methods, there are major differences. First of all, the Croston, SES, SBA, and TSB methods are extensively documented and possess a straightforward application process. On the other hand, the WSS method, despite being well-documented, necessitates strong mathematical knowledge such as Markov chains, thus it makes this method harder to implement. Secondly, all the ML methods offer various configurations for their hyperparameters, signifying that while the methods are relatively easy to implement, achieving optimal performance is challenging. In the case of all machine learning (ML) methods, the hyperparameter optimization process is carried out, leading to the generation of multiple versions of each model. These versions are subsequently compared based on the root mean square error (RMSE) metric. Therefore, the training process for hyperparameter optimization requires a profound understanding of data science, highlighting the requirement for a high level of knowledge in this domain. Moreover, the implementation of the LightGBM method is also challenging due to its limited availability of implementation examples in comparison to other methods. As a result, both the total execution duration and the required level of expertise should play a pivotal role in the selection of a method.

Finding 2: *SBA outperforms the other methods on average across both accuracy measures. Conversely, the LightGBM method consistently exhibits the poorest performance overall according to the average Percentage Better score.*

In Table 5.7, the SBA method demonstrated an overall superior forecasting accuracy, achieving the highest average Percentage Better scores for both accuracy measures. Furthermore, the SBA method exhibited favorable characteristics in terms of ease of implementation and average total execution duration compared to the other methods. Therefore, it holds significant potential as an overall method for practical applications to be prioritized when forecasting accuracy is the key metric. Nevertheless, these favorable results in forecasting accuracy did not extend to the inventory control assessment for the SBA method. Conversely, the LightGBM method exhibited the lowest performance according to the average Percentage Better score, highlighting its poor performance when accurate forecasts are required, even with one-SKU based hyper-parameter optimization implementation. Lastly, it should be noted that this finding is in full alignment with the results of Haan (2021b).

Finding 3: *The process of outlier detection and handling has shown enhancements in the results of forecasting accuracy. However, this process did not lead to significant changes in the performance of inventory control.*

Except for the WSS and SVR methods, all other approaches demonstrate improvements when transitioning from MAN and BRAF to their second versions respectively for both forecasting accuracy measures when outliers are handled using mean imputation. It is worth noting that the WSS method performs poorly in terms of the MASE metric for both data sets, while SVR exhibits lower accuracy in terms of only the MASE measure for the MAN2 data set. Regarding the OIL2 data set, only 4 out of 9 methods (namely SES, MLP, LightGBM, and RF methods) show improved performance in the RMSSE metric. However, none of the methods demonstrate enhanced accuracy when considering the MASE measure. As a result, in the context of three distinct data sets for both measures as well as nine distinct techniques, there has been an enhancement in the accuracy of forecasting outcomes in 46 out of 63 instances, thereby yielding a noteworthy improvement rate of 73.02%. On the other hand, when considering the application of the outlier detection and handling procedure to the MAN2, BRAF2, and OIL2 data sets, which led to changes in CV^2 and p values, it is evident that this procedure has not significantly influenced the results of inventory control performance. The visualizations of both trade-off curves reveal that the initial versions of the data sets (MAN, BRAF, and OIL) closely resemble their respective second versions (MAN2, BRAF2, OIL2). This can be attributed to the fact that both the original data sets (MAN, BRAF, OIL) and their revised versions predominantly consist of lumpy and intermittent items.

Finding 4: *Willemain's method (WSS) exhibits higher achieved fill rates for each target fill rate across almost all examined data sets and the Support Vector Regression (SVR) exhibits lower achieved fill rates.*

The WSS emerges as showing the highest achieved fill rates for all target fill rates in 10 out of 11 data sets. However, consistently surpassing the target fill rates results in an overestimation of the anticipated demand, leading to an excess stock in hand for the majority of scenarios. Consequently, achieving higher fill rates by this means has not substantiated any superiority over alternate methodologies. This way of behavior in inventory control can be attributed to its unique approach. It assesses the likelihood of the next demand occurrence being positive or zero, thus acquiring knowledge from the available data. Thus, as mentioned in Haan (2021b), exploring the potential of combining this method's approach with a more accurate forecasting technique for estimating expected demand sizes could be a promising avenue for future research. Conversely, the SVR method exhibits the lowest achieved fill rates for each target fill rate in 6 (all industrial) out of 11 data sets. Despite its effectiveness in terms of accuracy measures, as evidenced by its performance in 4 out of 7 industrial data sets with hyper-parameter optimization, the SVR method fails to deliver superiority in terms of inventory control. When the hyper parameter optimization process is conducted for all the ML methods, different parameters leading to different models are subsequently compared based on only the RMSE metric. Thus, there exists a potential avenue for future investigation, wherein the hyper-parameter optimization process integrates an inventory control measure to augment the performance of the SVR method for inventory management applications. This approach may exhibit promising results for enhancing the methods' effectiveness, specifically in terms of improving inventory control performance.

6.2 Comparison of the Findings with the Existing Body of Literature

Pinçe et al. (2021) Review

This paper's literature review is based on the framework of Pinçe et al. (2021), which aims to provide a comprehensive summary of the latest spare parts demand forecasting literature by analyzing the related studies' findings. Thus, their framework is being revisited to compare the outcomes of this research paper with the corresponding findings of their study. In the framework, they utilize 56 papers from the literature where they systematically examined and quantified the outcomes of those papers in their review without actually implementing them. Their analysis initiates by comparing Croston and SBA as the two benchmark methods. The findings indicate that in terms of accuracy measures, SBA outperforms Croston approximately 87% of the time. For this research paper, there are a total of 22 comparisons between Croston and SBA, and SBA outperforms Croston 100% of the time. Pinçe et al. (2021) noted in their analysis that the sole instance where Croston outperformed SBA was observed when applied to a data set from the fashion industry. This could explain the higher percentage of superiority observed in the present paper's comparison since exclusively spare parts demand data was utilized.

On the other hand, when considering inventory performance measures as the primary concern, the results are indicated as often inconclusive by Pinçe et al. (2021). In this study, with the stock control performance results given in the appendix for Croston and SBA, it is evident that Croston consistently attains lower fill rates across all target fill rates. However, the disparities observed are of such negligible magnitude that one could interpret these findings as inconclusive similar to Pinçe et al. (2021). The authors explain this dissimilarity in stock control performance by the inherent bias present in Croston. As Croston exhibits a positive bias, it consistently yields higher average inventory levels and service levels compared to the SBA method (Pinçe et al., 2021). Due to the utilization of distinct aggregated groups of methods that are not employed within the scope of this paper, it is not feasible to conduct a further comparison with the quantitative results presented by Pinçe et al. (2021).

Spiliotis et al. (2020) Review

The framework proposed by Pinçe et al. (2021) primarily concentrates on Neural Network (NN) techniques within the domain of machine learning (ML), presenting a limitation in terms of methodological diversity. To address this limitation, our paper incorporates a comparative analysis study conducted by Spiliotis et al. (2020), which encompasses 11 statistical and 7 ML methods. The methodology employed in our study not only aligns with the selection of methods utilized in their research but also incorporates the process of hyper-parameter optimization, whereby parameter values are chosen in accordance with them. However, some points from the study of Spiliotis et al. (2020) do not align with our paper, which can be listed as first they did not take into account inventory control performance, and second, they applied the methods to a single data set, which was predominantly characterized as smooth and erratic, similar to the demand patterns of SIM1, SIM3, and AUTO data sets of our paper.

In their study, four ML techniques, namely Gradient Boosting Tree (GBT), RF, SVR, and kNN Regression (kNNR) stood out as the top-performing methods when evaluated using the RMSSE metric. Considering the similar hyper-parameter optimization approach, while the RF did not exhibit exceptional performance in our paper, the SVR model's results align with their findings as it outperforms other methods utilized in terms of the RMSSE across a majority of the industrial data sets. Nevertheless, their investigation revealed that the forecasting accuracy of the MLP approach, as evaluated by the (RMSSE), was inferior to that of the statistical methods employed. This partially contradicts this paper's findings, since, in 3 out of 4 artificial data sets, the MLP is superior to the statistical methods for both measures. Moreover, SBA was identified as the most accurate statistical method in terms of forecasting accuracy by their study. This finding aligns with the results of our study, where SBA consistently demonstrates superior performance. Finally, the researchers emphasize that on average, ML methods require four times longer computational time compared to other methods, which can also be observed in and aligns with our results.

6.3 Conclusion

This section aims to establish a link between the previously discussed findings and the two research questions addressed in this paper. The first research question is as follows: *" Which spare part demand forecasting method is best for which type of data?"*

This paper evaluates multiple methods, specifically 9 different methods, as well as multiple types of methods, including parametric, bootstrapping, and machine learning approaches. The selection of a suitable method based on the data type exhibits variability as established by the main results and the comprehensive literature review. The findings indicate that none of these methods consistently outperformed others in terms of both forecasting accuracy and inventory control. Specifically, the SBA method, which is parametric, demonstrated the highest forecasting accuracy, while for the inventory control performance, the superior method is not consistent over all data sets. Consequently, to answer which method is best for which type of data, there has been no consistency of superiority to decide which type of method is best for which type of data over all data sets.

The second research question of this study is as follows: *" Does the performance of spare part demand forecasting method depend on the data pre-processing?*

Outlier detection and handling as a data pre-processing procedure has shown enhancements in the results of forecasting accuracy. Specifically, with the exception of the WSS and SVR, all other approaches exhibit improvements in both the MAN2 and BRAF2 data sets when outliers are handled through mean imputation. However, the WSS method underperforms after outlier detection handling in terms of the MASE metric for both data sets, while SVR shows lower accuracy in the MASE measure for only the MAN2 data set. For the OIL2 data, only MLP, LightGBM, and RF methods demonstrate enhanced performance after the outlier detection and handling process for the RMSSE metric among the 9 methods evaluated, however, none of the methods exhibit improved accuracy when considering the MASE measure. On the other hand, this outlier detection and handling process did not lead to significant changes in the performance of inventory control. The trade-off curve visualizations indicate a similarity between the initial versions of the data sets (MAN, BRAF, and OIL) and their respective second versions.

6.4 Discussion

In this section, several possibilities are listed to elucidate the potential directions for future research and to highlight the aspects that were not included within the scope of this paper. First, in the current approach, the hyper-parameter optimization process is carried out for all ML methods but with a model comparison based solely on the RMSE metric. However, a potential area for further research lies in incorporating an inventory control measure into the hyper-parameter optimization, specifically aimed at enhancing the performances of the ML methods in the field of inventory management applications. This approach may hold promise for improving the methods' efficiencies for practical applications in the supply chain domain.

Lastly, there exist certain opportunities for modifying the data pre-processing choices made in this study, as well as potential adjustments for future research to explore their implications. One notable change would involve incorporating actual lead times instead of setting all to 1. In addition, segmenting the data sets based on classifications, investigating demand patterns such as seasonality or substantial trends, and incorporating installed base forecasting as also suggested by Haan (2021b).

Bibliography

N.K. Ahmed, A.F. Atiya, N.E. Gayar, and H. El-Shishiny. An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5–6):594–621, 2010. URL https://doi.org/10.1080/07474938.2010.481556.

K. Aronis, I. Magou, R. Dekker, and G. Tagaras. Inventory control of spare parts using a bayesian approach: A case study. *European Journal of Operational Research*, 154(3):730–739, 2004. URL https://doi.org/10.1016/s0377-2217(02)00837-8.

S. Auweraer, P.M. Pardalos, and A. Syntetos. Forecasting spare part demand with installed base information: A review. *International Journal of Forecasting*, 35(1):181–196, 2019. URL https://doi.org/10.1016/j.ijforecast.2018.09.002.

M.Z. Babai, M. Ali, and K. Nikolopoulos. Impact of temporal aggregation on stock control performance of intermittent demand estimators: Empirical analysis. *Omega*, 40(6):713–721, 2012. URL https://doi.org/10.1016/j.omega.2011.09.004.

M.Z. Babai, A. Syntetos, and R.H. Teunter. Intermittent demand forecasting: An empirical study on accuracy and the risk of obsolescence. *International Journal of Production Economics*, 157:212–219, 2014. URL https://doi.org/10.1016/j.ijpe.2014.08.019.

M.Z. Babai, Y. Dallery, S. Boubaker, and R. Kalai. A new method to forecast intermittent demand in the presence of inventory obsolescence. *International Journal of Production Economics*, 209:30–41, 2019. URL https://doi.org/10.1016/j.ijpe.2018.01.026.

T. Bandaragoda, K.M. Ting, D.W. Albrecht, F. Liu, Y. Zhu, and J.C.K. Wells. Isolation-based anomaly detection using nearest-neighbor ensembles. *Computational Intelligence*, 34 (4):968–998, 2018. URL https://doi.org/10.1111/coin.12156.

E. Bartezzaghi, R. Verganti, and G. Zotteri. A simulation framework for forecasting uncertain lumpy demand. *International Journal of Production Economics*, 59(1-3):499–510, 1999. URL https://doi.org/10.1016/S0925-5273(98)00012-7.

I. Ben-Gal. *Outlier Detection*. Springer EBooks, 2006. URL https://doi.org/10.1007/0-387-25465-x_7.

C. Bergmeir and J.M. Benítez. Neural networks in r using the stuttgart neural network simulator: Rsnns. *Journal of Statistical Software*, 46(7):1–26, 2012. URL https://doi.org/10.18637/jss.v046.i07.

J.H. Bookbinder and A.E. Lordahl. Estimation of inventory re-order levels using the bootstrap statistical procedure. *IIE Transactions*, 21(4):302–312, 1989. URL https://doi.org/10.1080/07408178908966236.

P. Boutselis and K.R. McNaught. Using bayesian networks to forecast spares demand from equipment failures in a changing service logistics context. *International Journal of Production Economics*, 209:325–333, 2019. URL `https://doi.org/10.1016/j.ijpe.2018.06.017`.

J.E. Boylan and M. Babai. On the performance of overlapping and non-overlapping temporal demand aggregation approaches. *International Journal of Production Economics*, 181:136–144, 2016. URL `https://doi.org/10.1016/j.ijpe.2016.04.003`.

J.E. Boylan and A.A. Syntetos. *Forecasting for Inventory Management of Service Parts*. Springer, London EBooks, 2008. URL `https://doi.org/10.1007/978-1-84800-011-7_20`.

J.E. Boylan, A.A. Syntetos, and G.C. Karakostas. Classification for forecasting and stock control: a case study. *Journal of the Operational Research Society*, 59(4):473–481, 2008. URL `https://doi.org/10.1057/palgrave.jors.2602312`.

L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001. URL `https://doi.org/10.1023/A:1010933404324`.

M. Breunig, H. Kriegel, R.T. Ng, and J. Sander. Lof. *Sigmod Record*, 29(2):93–104, 2000. URL `https://doi.org/10.1145/335191.335388`.

M.A. Cohen, P.V. Kamesam, P.R. Kleindorfer, H.L. Lee, and A. Tekerian. Optimizer: Ibm's multi-echelon inventory system for managing service logistics. *Interfaces*, 20(1):65–82, 1990. URL `https://doi.org/10.1287/inte.20.1.65`.

J.D. Croston. Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, 23(3):289–303, 1972. URL `https://doi.org/10.2307/3007885`.

R. Dekker, Ç. Pinçe, R. Zuidwijk, and M.A. Jalil. On the use of installed base information for spare parts logistics: A review of ideas and industry practice. *International Journal of Production Economics*, 143(2):536–545, 2013. URL `https://doi.org/10.1016/j.ijpe.2011.11.025`.

P. Durlinger and I. Paul. Inventory and holding costs. In *Durlinger Consultant*, page 1. 2012.

A.H.C. Eaves and B.G. Kingsman. Forecasting for the ordering and stock-holding of spare parts. *Journal of the Operational Research Society*, 55(4):431–437, 2004. URL `https://doi.org/10.1057/palgrave.jors.2601697`.

L. Fan, J. Zhang, W. Mao, and F. Cao. Unsupervised anomaly detection for intermittent sequences based on multi-granularity abnormal pattern mining. *Entropy*, 25(1):123, 2023. URL `https://doi.org/10.3390/e25010123`.

R. Fildes, P. Goodwin, M.S. Lawrence, and K. Nikolopoulos. Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, 25(1):3–23, 2009. URL `https://doi.org/10.1016/j.ijforecast.2008.11.010`.

P.H. Franses. *Expert Adjustments of Model Forecasts*. Cambridge University Press, 2014.

A. Gautam and V. Singh. Parametric versus non-parametric time series forecasting methods: A review. *Journal of Engineering Science and Technology Review*, 13(3):165–171, 2020. URL `https://doi.org/10.25103/jestr.133.18`.

A.A. Ghobbar and C. Friend. Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model. *Computers and Operations Research*, 30(14): 2097–2114, 2003. URL `https://doi.org/10.1016/s0305-0548(02)00125-9`.

A.A. Ghobbar and C.H. Friend. Sources of intermittent demand for aircraft spare parts within airline operations. *Journal of Air Transport Management*, 8(4):221–231, 2002. URL `https://doi.org/10.1016/s0969-6997(01)00054-0`.

M. Goldstein and S. Uchida. A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PLOS ONE*, 11(4):0152173, 2016. doi: 10.1371/journal.pone.0152173.

F. Guo, J. Diao, Q. Zhao, D. Wang, and Q. Sun. A double-level combination approach for demand forecasting of repairable airplane spare parts based on turnover data. *Computers and Amp; Industrial Engineering*, 110:92–108, 2017. URL `https://doi.org/10.1016/j.cie.2017.05.002`.

R.S. Gutierrez, A.O. Solis, and S. Mukhopadhyay. Lumpy demand forecasting using neural networks. *International Journal of Production Economics*, 111(2):409–420, 2008. URL `https://doi.org/10.1016/j.ijpe.2007.01.007`.

D.B. de Haan. Spare part demand forecasting, 2021a. URL `https://github.com/danieldehaan96/spdf`.

D.B. de Haan. Benchmarking spare parts demand forecasting methods. Master's thesis, 2021b. URL `http://hdl.handle.net/2105/60771`.

M. Hasni, M.S. Aguir, M.Z. Babai, and Z. Jemai. Spare parts demand forecasting: a review on bootstrapping methods. *International Journal of Production Research*, 57(15-16):4791–4804,, 2019. doi: 10.1080/00207543.2018.1424375.

C.T. Hendricks. Comparison of statistical and machine learning methods for improving demand forecasting of spare parts in simulators and weapon ranges - proquest, 2022. URL `https://www.proquest.com/dissertations-theses/comparison-statistical-machine-learning-methods/docview/2609298463/se-2`.

C. Hu and X. Qin. A density-based local outlier detecting algorithm. *Jisuanji Yanjiu yu Fazhan/Computer Research and Development*, 47:2110–2116, 12 2010.

Z. Hua, B. Zhang, J. Yang, and D.T. Tan. A new approach of forecasting intermittent demand for spare parts inventories in the process industries. *Journal of the Operational Research Society*, 58(1):52–61, 2007. URL `https://doi.org/10.1057/palgrave.jors.2602119`.

D. Ivanov, A. Tsipoulanidis, and J. Schönberger. *Global Supply Chain and Operations Management*. In Springer texts in business and economics. Springer International Publishing, 2017. URL https://doi.org/10.1007/978-3-319-24217-0.

M.A. Jalil, R. Zuidwijk, M. Fleischmann, and J. Nunen. Spare parts logistics and installed base information. *Journal of the Operational Research Society*, 62(3):442–457, 2010. URL https://doi.org/10.1057/jors.2010.38.

A. Jiang, KL Tam, X. Guo, and Y. Zhang. A new approach to forecasting intermittent demand based on the mixed zero-truncated poisson model. *J Forecast*, 39(1):69–83, 2020.

G. John. Robust decision trees: removing outliers from databases. *Knowledge Discovery and Data Mining*, page 174–179, 1995. URL https://aaaipress.org/Papers/KDD/1995/KDD95-044.pdf.

F R Johnston, J E Boylan, and E A Shale. An examination of the size of orders from customers, their characterisation and the implications for inventory control of slow moving items. *Journal of the Operational Research Society*, 54(8):833–837, 2003. URL https://doi.org/10.1057/palgrave.jors.2601586.

F.R. Johnston and J.E. Boylan. Forecasting for items with intermittent demand. *The Journal of the Operational Research Society*, 47(1):113–121, 1996. URL https://doi.org/10.2307/2584256.

A. Jozi, T. Pinto, and Z. Vale. Contextual learning for energy forecasting in buildings. *International Journal of Electrical Power and Energy Systems*, 136:107707, 2022. URL https://doi.org/10.1016/j.ijepes.2021.107707.

Kailex. *M5 ForecasteR v2*. Kaggle, 2020. URL https://www.kaggle.com/code/kailex/m5-forecaster-v2/report.

N. Khue. Spare part demand forecasting, 2023. URL https://github.com/KhueNguyenTK/Spare-Part-Demand-Forecasting.git.

D. Kiefer, F. Grimm, M. Bauer, and C. Dinther. Demand forecasting intermittent and lumpy time series: Comparing statistical, machine learning and deep learning methods. In *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2021. URL https://doi.org/10.24251/hicss.2021.172.

A.E. Koehler and R.J. Hyndman. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006. URL https://doi.org/10.1016/j.ijforecast.2006.03.001.

A.V. Kostenko and R.J. Hyndman. A note on the categorization of demand patterns. *Journal of the Operational Research Society*, 57(10):1256–1257, 2006. URL https://doi.org/10.1057/palgrave.jors.2602211.

N. Kourentzes. Intermittent demand forecasts with neural networks. *International Journal of Production Economics*, 143(1):198–206, 2013. URL https://doi.org/10.1016/j.ijpe.2013.01.009.

N. Kourentzes. On intermittent demand model optimisation and selection. *International Journal of Production Economics*, 156:180–190, 2014. URL https://doi.org/10.1016/j.ijpe.2014.06.007.

D. Lengu, A. Syntetos, and M. Babai. Spare parts management: Linking distributional assumptions to demand classification. *European Journal of Operational Research*, 235(3):624–635, 2014. URL https://doi.org/10.1016/j.ejor.2013.12.043.

C. Li and A. Lim. A greedy aggregation–decomposition method for intermittent demand forecasting in fashion retailing. *European Journal of Operational Research*, 269(3):860–869, 2018. URL https://doi.org/10.1016/j.ejor.2018.02.029.

A. Liaw and M.C. Wiener. Classification and regression by randomforest. *R News*, 2002. URL https://cogns.northwestern.edu/cbmg/LiawAndWiener2002.pdf.

F. Liu, K.M. Ting, and Z. Zhou. Isolation forest. In *International Conference on Data Mining*, 2008. URL https://doi.org/10.1109/icdm.2008.17.

F. Lolli, R. Gamberini, A. Regattieri, E. Balugani, T. Gatos, and S. Gucci. Single-hidden layer neural networks for forecasting intermittent demand. *International Journal of Production Economics*, 183:116–128, 2017. URL https://doi.org/10.1016/j.ijpe.2016.10.021.

S. Makridakis, E. Spiliotis, and V. Assimakopoulos. Statistical and machine learning forecasting methods: Concerns and ways forward. *PLOS ONE*, 13(3):0194889, 2018. URL https://doi.org/10.1371/journal.pone.0194889.

S. Makridakis, E. Spiliotis, and V. Assimakopoulos. M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38(4):1346–1364, 2022. URL https://doi.org/10.1016/j.ijforecast.2021.11.013.

K.A. Malik, H. Sadawarti, and K.G. S. Comparative analysis of outlier detection techniques. *International Journal of Computer Applications*, 97(8):12–21, 2014. URL https://doi.org/10.5120/17026-7318.

D.W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963. URL https://doi.org/10.1137/0111030.

I. McCue. *Inventory Carrying Costs: What It Is and How to Calculate It.* Oracle NetSuite, 2022. URL https://www.netsuite.com/portal/resource/articles/inventory-management/inventory-carrying-costs.shtml.

D.A. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch. Misc functions of the department of statistics, probability theory group, 2019. URL https://mirrors.tuna.

tsinghua.edu.cn/CRAN/web/packages/e1071/index.html. Formerly: E1071), TU Wien [R package e1071 version 1.7-4].

Microsoft. Welcome to lightgbm's documentation! — *LightGBM*, 3(2.1.99), 2021. URL https://lightgbm.readthedocs.io/en/latest/index.html. documentation.

M. Mohammadipour and J.E. Boylan. Forecast horizon aggregation in integer autoregressive moving average (inarma) models. *Omega*, 40(6):703–712, 2012. URL https://doi.org/10.1016/j.omega.2011.08.008.

M. Monfared, R. Ghandali, and M. Esmaeili. A new adaptive exponential smoothing method for non-stationary time series with level shifts. *Journal of Industrial Engineering, International*, 10(4):209–216, 2014. URL https://doi.org/10.1007/s40092-014-0075-5.

S. Moon, C. Hicks, and A. Simpson. The development of a hierarchical forecasting method for predicting spare parts demand in the south korean navy—a case study. *International Journal of Production Economics*, 140(2):794–802, 2012. URL https://doi.org/10.1016/j.ijpe.2012.02.012.

S. Mukhopadhyay, A.O. Solis, and R.S. Gutierrez. The accuracy of non-traditional versus traditional methods of forecasting lumpy demand. *J. Forecast*, 31:721–735, 2012. URL https://doi.org/10.1002/for.1242.

I. Naim, T. Mahara, and A.R. Idrisi. Effective short-term forecasting for daily time series with complex seasonal patterns. *Procedia Computer Science*, 132:1832–1841, 2018. URL https://doi.org/10.1016/j.procs.2018.05.136.

W. Ni, G. Chen, J. Lu, Y. Wu, and Z. Sun. Local entropy based weighted subspace outlier mining algorithm. *Jisuanji Yanjiu yu Fazhan/Computer Research and Development*, 45:1189–1194, 07 2008.

K. Nikolopoulos, A. Syntetos, J.E. Boylan, F. Petropoulos, and V. Assimakopoulos. An aggregate–disaggregate intermittent demand approach (adida) to forecasting: an empirical proposition and analysis. *Journal of the Operational Research Society*, 62(3):544–554, 2011. URL https://doi.org/10.1057/jors.2010.32.

S. Omar, M.A. Ngadi, and H.H. Jebur. Machine learning techniques for anomaly detection: An overview. *International Journal of Computer Applications*, 79(2):33–41, 2013. URL https://doi.org/10.5120/13715-1478.

S. Papadimitriou, H. Kitagawa, P.B. Gibbons, and C. Faloutsos. Loci: fast outlier detection using the local correlation integral. 2003. doi: 10.1109/icde.2003.1260802.

C. Pennings, J. Dalen, and E. Laan. Exploiting elapsed time for managing intermittent demand for spare parts. *European Journal of Operational Research*, 258(3):958–969, 2017. URL https://doi.org/10.1016/j.ejor.2016.09.017.

H.N. Perera, J. Hurley, B. Fahimnia, and M. Reisi. The human factor in supply chain forecasting: A systematic review. *European Journal of Operational Research*, 274(2):574–600, 2019. URL https://doi.org/10.1016/j.ejor.2018.10.028.

F. Petropoulos and N. Kourentzes. Forecast combinations for intermittent demand. *Journal of the Operational Research Society*, 66(6):914–924, 2015. URL https://doi.org/10.1057/jors.2014.62.

F. Petropoulos, S. Makridakis, V. Assimakopoulos, and K. Nikolopoulos. Horses for courses' in demand forecasting. *European Journal of Operational Research*, 237(1):152–163, 2014. URL https://doi.org/10.1016/j.ejor.2014.02.036.

F. Petropoulos, N. Kourentzes, and K. Nikolopoulos. Another look at estimators for intermittent demand. *International Journal of Production Economics*, 181:154–161, 2016. URL https://doi.org/10.1016/j.ijpe.2016.04.017.

D. Petrovic and R. Petrović. Sparta ii: Further development in an expert system for advising on stocks of spare parts. *International Journal of Production Economics*, 24(3):291–300, 1992. URL https://doi.org/10.1016/0925-5273(92)90141-s.

Ç. Pinçe, L. Turrini, and J. Meissner. Intermittent demand forecasting for spare parts: A critical review. *Omega*, 105:102513, 2021. URL https://doi.org/10.1016/j.omega.2021.102513.

E. Porras and R. Dekker. An inventory control system for spare parts at a refinery: An empirical comparison of different re-order point methods. *European Journal of Operational Research*, 184(1):101–132, 2008. URL https://doi.org/10.1016/j.ejor.2006.11.008.

L. Qi, X. Li, Q. Wang, and S. Jia. fetsmcs: Feature-based ets model component selection. *International Journal of Forecasting*, 2022. URL https://doi.org/10.1016/j.ijforecast.2022.06.004.

J.R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1993.

S. Ramaswamy, R. Rastogi, and K. Shim. Efficient algorithms for mining outliers from large data sets. *Sigmod Record*, 29(2):427–438, 2000. URL https://doi.org/10.1145/335191.335437.

W. Romeijnders, R.H. Teunter, and W. Jaarsveld. A two-step method for forecasting spare parts demand using information on component repairs. *European Journal of Operational Research*, 220(2):386–393, 2012. URL https://doi.org/10.1016/j.ejor.2012.01.019.

R. Salehzadeh, R.A. Tabaeeian, and F. Esteki. Exploring the consequences of judgmental and quantitative forecasting on firms' competitive performance in supply chains. *Benchmarking: An International Journal*, 27(5):1717–1737, 2020. URL https://doi.org/10.1108/bij-08-2019-0382.

S.L. Salzberg. C4.5: Programs for machine learning by j. ross quinlan. *Machine Learning*, 16 (3):235–240, 1994. URL https://doi.org/10.1007/bf00993309.

B. Sani and B.G. Kingsman. Selecting the best periodic inventory control and demand forecasting methods for low demand items. *The Journal of the Operational Research Society*, 48(7): 700–713, 1997. URL `https://doi.org/10.2307/3010059`.

B. Schölkopf and A.J. Smola. Learning with kernels. In *The MIT Press eBooks*. The MIT Press, 2001. URL `https://doi.org/10.7551/mitpress/4175.001.0001`.

Prestwich SD, Tarim SA, Rossi R, and Hnich B. Forecasting intermittent demand by hyperbolic-exponential smoothing. *Int J Forecast*, 30(4):928–33, 2014.

A. Smiti. A critical overview of outlier detection methods. *Computer Science Review*, 38:100306, 2020. URL `https://doi.org/10.1016/j.cosrev.2020.100306`.

S. Smyl. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting*, 36(1):75–85, 2020. URL `https://doi.org/10.1016/j.ijforecast.2019.03.017`.

R. Snyder. Forecasting sales of slow and fast moving inventories. *European Journal of Operational Research*, 140(3):684–699, 2002. URL `https://doi.org/10.1016/S0377-2217(01)00231-4`.

RD. Snyder, Ord JK, and Beaumont A. Forecasting the intermittent demand for slow-moving inventories: a modelling approach. *Int J Forecast*, 28(2):485–96, 2012.

E. Spiliotis, S. Makridakis, A. Semenoglou, and V. Assimakopoulos. Comparison of statistical and machine learning methods for daily sku demand forecasting. *Operational Research*, 22(3): 3037–3061, 2020. URL `https://doi.org/10.1007/s12351-020-00605-2`.

X. Su and C. Tsai. Outlier detection. *Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery*, 1(3):261–268, 2011. URL `https://doi.org/10.1002/widm.19`.

A. Syntetos and J.E. Boylan. The accuracy of intermittent demand estimates. *International Journal of Forecasting*, 21(2):303–314, 2005. URL `https://doi.org/10.1016/j.ijforecast.2004.10.001`.

A. Syntetos and J.E. Boylan. On the stock control performance of intermittent demand estimators. *International Journal of Production Economics*, 103(1):36–47, 2006. URL `https://doi.org/10.1016/j.ijpe.2005.04.004`.

A. Syntetos, K. Nikolopoulos, J.E. Boylan, R. Fildes, and P. Goodwin. The effects of integrating management judgement into intermittent demand forecasts. *International Journal of Production Economics*, 118(1):72–81, 2009. URL `https://doi.org/10.1016/j.ijpe.2008.08.011`.

A. Syntetos, M.Z. Babai, and E.S. Gardner. Forecasting intermittent inventory demands: simple parametric methods vs. bootstrapping. *Journal of Business Research*, 68(8):1746–1752, 2015. URL `https://doi.org/10.1016/j.jbusres.2015.03.034`.

A.A. Syntetos, J.E. Boylan, and J.D. Croston. On the categorization of demand patterns. *Journal of the Operational Research Society*, 56(5):495–503, 2005. URL `https://doi.org/10.1057/palgrave.jors.2601841`.

A.A. Syntetos, M.Z. Babai, D. Lengu, and N. Altay. Distributional assumptions for parametric forecasting of intermittent demand. *Service Parts Management*, page 31–52, 2011. URL https://doi.org/10.1007/978-0-85729-039-7_2.

B. Tang and H. He. A local density-based approach for outlier detection. *Neurocomputing*, 241: 171–180, 2017. URL https://doi.org/10.1016/j.neucom.2017.02.039.

J. Tang, Z. Chen, A.W. Fu, and D.W. Cheung. Enhancing effectiveness of outlier detections for low density patterns. In *Lecture Notes in Computer Science*, page 535–548. Springer Science+Business Media, 2002. URL https://doi.org/10.1007/3-540-47887-6_53.

R.H. Teunter and L. Duncan. Forecasting intermittent demand: a comparative study. *Journal of the Operational Research Society*, 60(3):321–329, 2009. URL https://doi.org/10.1057/palgrave.jors.2602569.

R.H. Teunter, A.A. Syntetos, and M.Z. Babai. Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research*, 214(3):606–615, 2011.

E. Van Wingerden, R. Basten, R. Dekker, and W. Rustenburg. More grip on inventory control through improved forecasting: A comparative study at three companies. *International Journal of Production Economics*, 157:220–237, 2014. URL https://doi.org/10.1016/j.ijpe.2014.08.018.

S. Viswanathan and C.X. Zhou. A new bootstrapping based method for forecasting and safety stock determination for intermittent demand items. 2008.

W. Wang and A. Syntetos. Spare parts demand: Linking forecasting to equipment maintenance. *Transportation Research Part E-Logistics and Transportation Review*, 47(6):1194–1209, 2011. URL https://doi.org/10.1016/j.tre.2011.04.008.

Z. Wei, G. Chen, L. Xu, and C. Gao. Prediction of maintenance spare parts based on bp neural network optimized by improved sparrow search algorithm. In *2022 IEEE 5th International Conference on Information Systems*, 2022. URL https://doi.org/10.1109/iciscae55891.2022.9927513. and Computer Aided Education (ICISCAE).

T.R. Willemain, C.N. Smart, J.H. Shockor, and P.A. DeSautels. Forecasting intermittent demand in manufacturing: a comparative evaluation of croston's method. *International Journal of Forecasting*, 10(4):529–538, 1994. URL https://doi.org/10.1016/0169-2070(94)90021-3.

T.R. Willemain, C.N. Smart, and H.F. Schwarz. A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting*, 20(3):375–387, 2004. URL https://doi.org/10.1016/s0169-2070(03)00013-x.

T.M. Williams. Stock control with sporadic and slow-moving demand. *Journal of the Operational Research Society*, 35(10):939–948, 1984. URL https://doi.org/10.1057/jors.1984.185.

G. Zhang, B.E. Patuwo, and M.S. Hu. Forecasting with artificial neural networks. *International Journal of Forecasting*, 14(1):35–62, 1998. URL https://doi.org/10.1016/s0169-2070(97)00044-7.

S. Zhu, R. Dekker, W. Jaarsveld, R.W. Renjie, and A.J. Koning. An improved method for forecasting spare parts demand using extreme value theory. *European Journal of Operational Research*, 261(1):169–181, 2017. URL https://doi.org/10.1016/j.ejor.2017.01.053.

S. Zhu, W. Jaarsveld, and R. Dekker. Spare parts inventory control based on maintenance planning. *Reliability Engineering and System Safety*, 193:106600, 2020. URL https://doi.org/10.1016/j.ress.2019.106600.

APPENDIX

$$MAR_n = \sum_{i=1}^{n} |r_i| \tag{1}$$

$$r_i = \hat{y}_i - i^{-1} \sum_{j=1}^{i} y_j \tag{2}$$

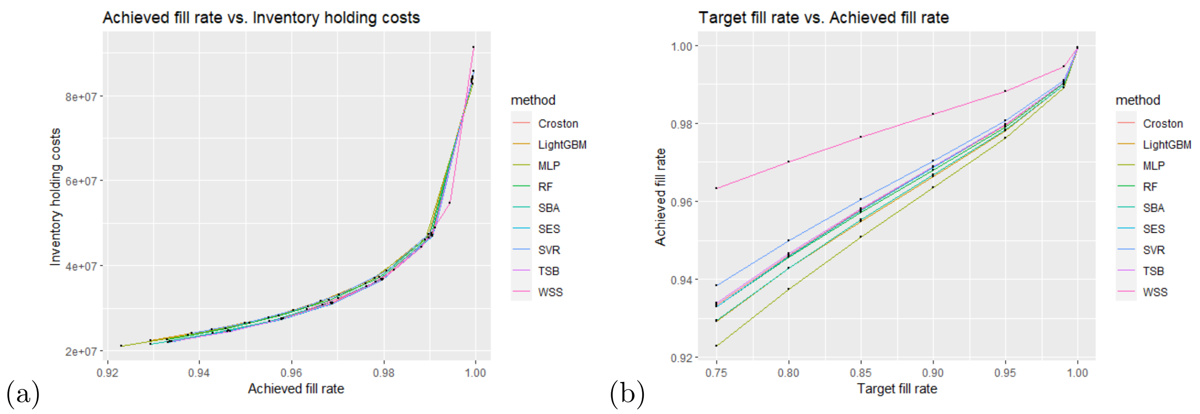


Figure 1: Tradeoff Curves of Stock Control Performance Measures for SIM2 Predictions

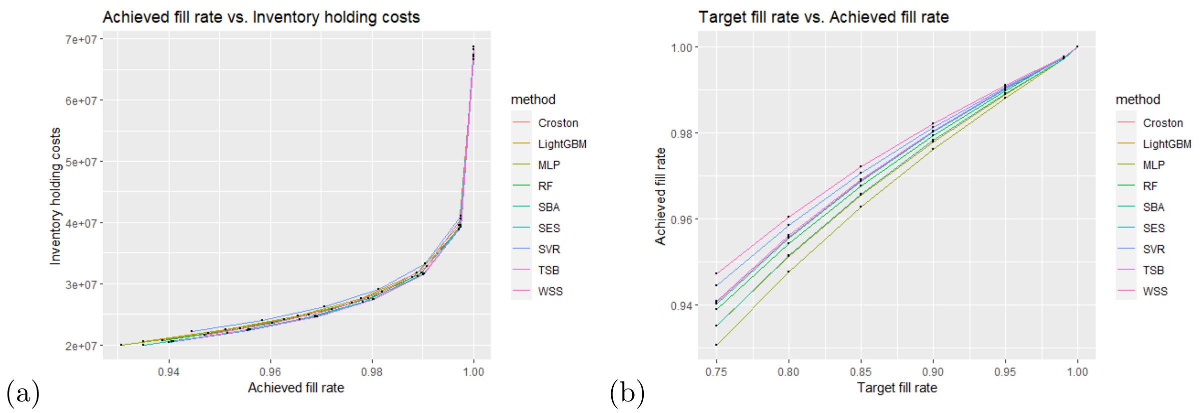


Figure 2: Tradeoff Curves of Stock Control Performance Measures for SIM4 Predictions

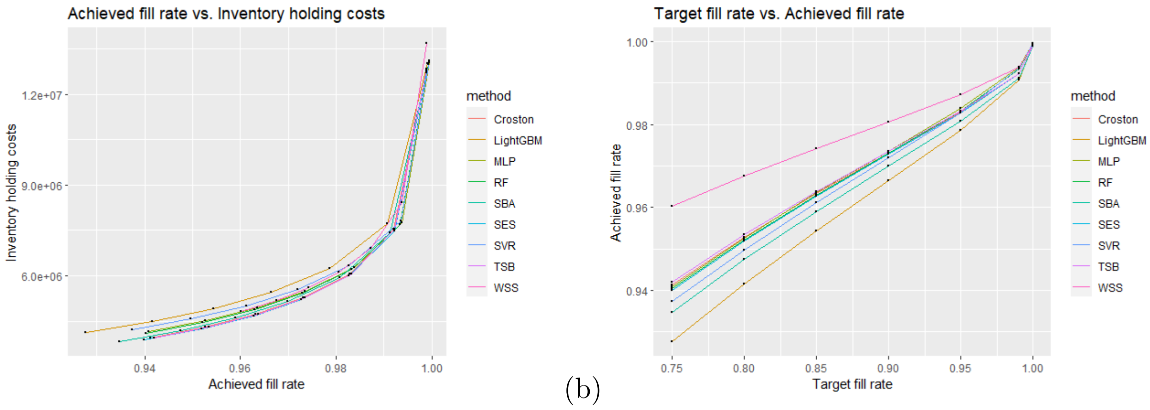


Figure 3: Tradeoff Curves of Stock Control Performance Measures for AUTO Predictions

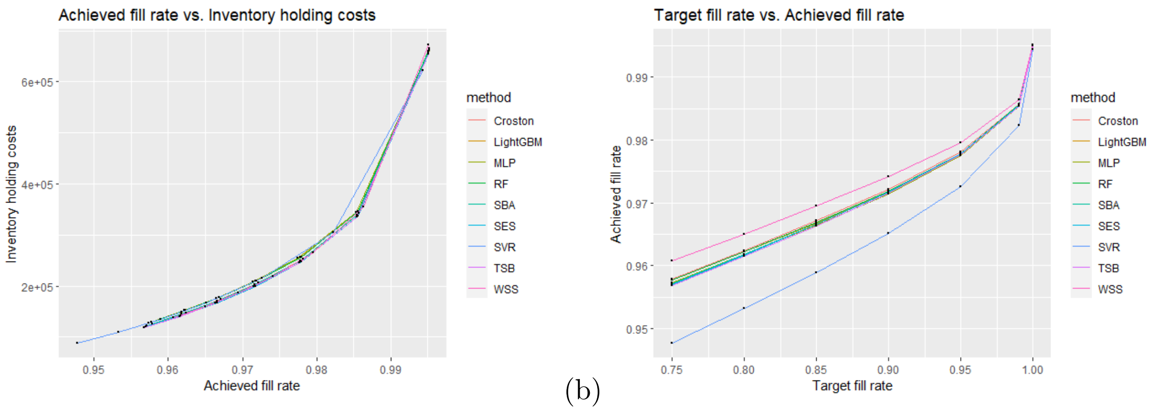


Figure 4: Tradeoff Curves of Stock Control Performance Measures for BRAF Predictions

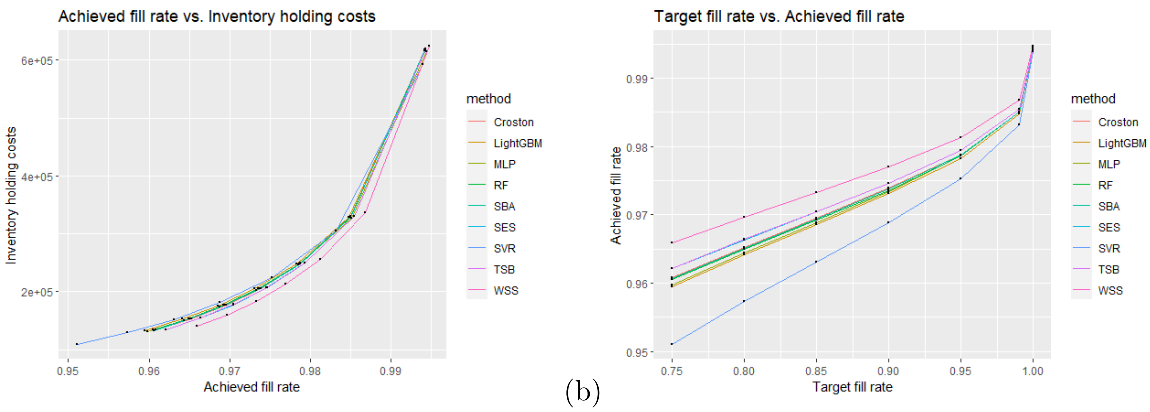
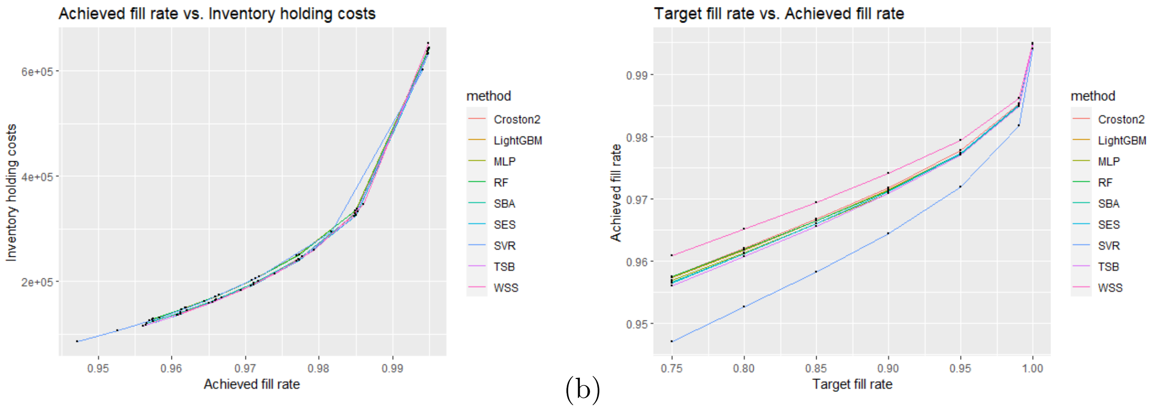
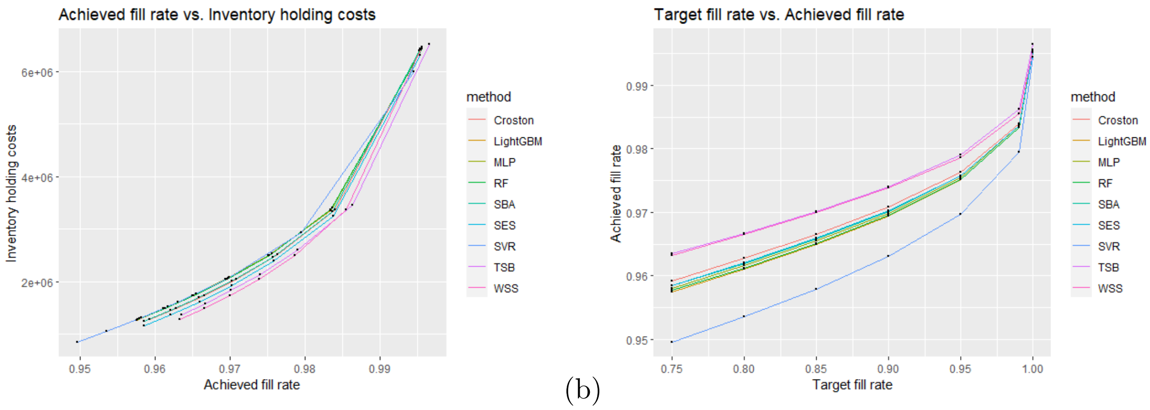


Figure 5: Tradeoff Curves of Stock Control Performance Measures for MAN2 Predictions



(a) (b)

Figure 6: Tradeoff Curves of Stock Control Performance Measures for BRAF2 Predictions



(a) (b)

Figure 7: Tradeoff Curves of Stock Control Performance Measures for OIL2 Predictions

Table 1: Stock Control Performance Results for ts.12

Achieved Fill Rate	Holding Costs	Target Fill Rates	Method
0.9552169	3269.031	0.7500000	Croston
0.9646240	3558.657	0.8000000	Croston
0.9718086	3896.251	0.8500000	Croston
0.9808486	4321.022	0.9000000	Croston
0.9892217	4950.597	0.9500000	Croston
0.9999034	6131.573	0.9900000	Croston
1.0000000	10337.508	0.9999999	Croston
0.9541853	3241.362	0.7500000	SES
0.9640351	3530.988	0.8000000	SES
0.9712198	3868.583	0.8500000	SES
0.9802597	4293.354	0.9000000	SES
0.9889714	4922.928	0.9500000	SES
0.9996532	6103.905	0.9900000	SES
1.0000000	10309.840	0.9999999	SES

0.9522493	3189.437	0.7500000	SBA
0.9629301	3479.063	0.8000000	SBA
0.9701147	3816.657	0.8500000	SBA
0.9791547	4241.428	0.9000000	SBA
0.9885018	4871.002	0.9500000	SBA
0.9991835	6051.979	0.9900000	SBA
1.0000000	10257.914	0.9999999	SBA
0.9555813	3278.806	0.7500000	TSB
0.9648320	3568.432	0.8000000	TSB
0.9720167	3906.026	0.8500000	TSB
0.9810566	4330.797	0.9000000	TSB
0.9893101	4960.371	0.9500000	TSB
0.9999918	6141.348	0.9900000	TSB
1.0000000	10347.283	0.9999999	TSB
0.9563447	3299.280	0.7500000	WSS
0.9652677	3588.906	0.8000000	WSS
0.9724524	3926.500	0.8500000	WSS
0.9814923	4351.271	0.9000000	WSS
0.9894953	4980.845	0.9500000	WSS
1.0000000	6161.822	0.9900000	WSS
1.0000000	10367.757	0.9999999	WSS
0.9522284	3188.877	0.7500000	MLP
0.9629181	3478.503	0.8000000	MLP
0.9701028	3816.098	0.8500000	MLP
0.9791427	4240.868	0.9000000	MLP
0.9884967	4870.443	0.9500000	MLP
0.9991784	6051.419	0.9900000	MLP
1.0000000	10257.355	0.9999999	MLP
0.9517242	3175.355	0.7500000	LightGBM
0.9625227	3464.981	0.8000000	LightGBM
0.9698150	3802.576	0.8500000	LightGBM
0.9788550	4227.346	0.9000000	LightGBM
0.9883744	4856.921	0.9500000	LightGBM
0.9990561	6037.897	0.9900000	LightGBM
1.0000000	10243.833	0.9999999	LightGBM
0.9481482	3079.444	0.7500000	RF
0.9589467	3369.070	0.8000000	RF
0.9677739	3706.664	0.8500000	RF
0.9768138	4131.435	0.9000000	RF
0.9875069	4761.009	0.9500000	RF

0.9981886	5941.986	0.9900000	RF
1.0000000	10147.921	0.9999999	RF
0.9517936	3177.215	0.7500000	SVR
0.9625921	3466.841	0.8000000	SVR
0.9698546	3804.435	0.8500000	SVR
0.9788945	4229.206	0.9000000	SVR
0.9883912	4858.780	0.9500000	SVR
0.9990730	6039.757	0.9900000	SVR
1.0000000	10245.692	0.9999999	SVR

Table 2: Stock Control Performance Results for SIM1

Achieved Fill Rate	Holding Costs	Target Fill Rates	Method
0.9423585	41819748	0.7500000	Croston
0.9536177	45598991	0.8000000	Croston
0.9638693	50004160	0.8500000	Croston
0.9734651	55546869	0.9000000	Croston
0.9829830	63761998	0.9500000	Croston
0.9923018	79172214	0.9900000	Croston
0.9994298	134054221	0.9999999	Croston
0.9423605	41820204	0.7500000	SES
0.9536194	45599447	0.8000000	SES
0.9638708	50004616	0.8500000	SES
0.9734658	55547324	0.9000000	SES
0.9829838	63762454	0.9500000	SES
0.9923022	79172670	0.9900000	SES
0.9994298	134054677	0.9999999	SES
0.9391145	41065621	0.7500000	SBA
0.9509275	44844863	0.8000000	SBA
0.9616884	49250032	0.8500000	SBA
0.9717933	54792741	0.9000000	SBA
0.9818950	63007870	0.9500000	SBA
0.9917740	78418086	0.9900000	SBA
0.9993818	133300093	0.9999999	SBA
0.9424799	41826962	0.7500000	TSB
0.9537280	45606205	0.8000000	TSB
0.9639666	50011374	0.8500000	TSB
0.9735429	55554083	0.9000000	TSB
0.9830374	63769212	0.9500000	TSB
0.9923294	79179428	0.9900000	TSB

0.9994307	134061435	0.9999999	TSB
0.9457122	43781185	0.7500000	WSS
0.9559397	47560428	0.8000000	WSS
0.9653732	51965597	0.8500000	WSS
0.9742698	57508306	0.9000000	WSS
0.9832380	65723435	0.9500000	WSS
0.9922077	81133651	0.9900000	WSS
0.9993823	136015658	0.9999999	WSS
0.9368958	41583873	0.7500000	MLP
0.9489600	45363116	0.8000000	MLP
0.9601023	49768285	0.8500000	MLP
0.9705729	55310994	0.9000000	MLP
0.9810996	63526123	0.9500000	MLP
0.9915729	78936339	0.9900000	MLP
0.9995633	133818346	0.9999999	MLP
0.9399531	42567986	0.7500000	LightGBM
0.9515047	46347229	0.8000000	LightGBM
0.9620881	50752398	0.8500000	LightGBM
0.9720740	56295107	0.9000000	LightGBM
0.9820413	64510236	0.9500000	LightGBM
0.9919435	79920452	0.9900000	LightGBM
0.9994945	134802459	0.9999999	LightGBM
0.9422362	42966518	0.7500000	RF
0.9533040	46745761	0.8000000	RF
0.9634735	51150930	0.8500000	RF
0.9730432	56693639	0.9000000	RF
0.9826601	64908768	0.9500000	RF
0.9922432	80318984	0.9900000	RF
0.9995787	135200991	0.9999999	RF
0.9434603	43886202	0.7500000	SVR
0.9542600	47665445	0.8000000	SVR
0.9642020	52070614	0.8500000	SVR
0.9735708	57613323	0.9000000	SVR
0.9830047	65828452	0.9500000	SVR
0.9924621	81238668	0.9900000	SVR
0.9996449	136120675	0.9999999	SVR

Table 3: Stock Control Performance Results for SIM2

Achieved Fill Rate	Holding Costs	Target Fill Rates	Method

0.9335466	22149588	0.7500000	Croston
0.9462346	24676450	0.8000000	Croston
0.9577866	27621816	0.8500000	Croston
0.9686105	31327759	0.9000000	Croston
0.9794861	36820524	0.9500000	Croston
0.9904060	47124039	0.9900000	Croston
0.9991610	83819020	0.9999999	Croston
0.9330821	21924840	0.7500000	SES
0.9459867	24451702	0.8000000	SES
0.9577026	27397068	0.8500000	SES
0.9686360	31103011	0.9000000	SES
0.9796170	36595777	0.9500000	SES
0.9905371	46899292	0.9900000	SES
0.9991855	83594272	0.9999999	SES
0.9294388	21514286	0.7500000	SBA
0.9429361	24041148	0.8000000	SBA
0.9551942	26986514	0.8500000	SBA
0.9667105	30692456	0.9000000	SBA
0.9782731	36185222	0.9500000	SBA
0.9898503	46488737	0.9900000	SBA
0.9991142	83183718	0.9999999	SBA
0.9339912	22116268	0.7500000	TSB
0.9466470	24643129	0.8000000	TSB
0.9581647	27588495	0.8500000	TSB
0.9689514	31294438	0.9000000	TSB
0.9797794	36787204	0.9500000	TSB
0.9905895	47090719	0.9900000	TSB
0.9991832	83785700	0.9999999	TSB
0.9633567	29717637	0.7500000	WSS
0.9701476	32244498	0.8000000	WSS
0.9763617	35189864	0.8500000	WSS
0.9822698	38895807	0.9000000	WSS
0.9882753	44388573	0.9500000	WSS
0.9944378	54692088	0.9900000	WSS
0.9995112	91387069	0.9999999	WSS
0.9229127	21094845	0.7500000	MLP
0.9375312	23621707	0.8000000	MLP
0.9508894	26567073	0.8500000	MLP
0.9634632	30273016	0.9000000	MLP
0.9761566	35765782	0.9500000	MLP

0.9889994	46069297	0.9900000	MLP
0.9992472	82764277	0.9999999	MLP
0.9293700	22391107	0.7500000	LightGBM
0.9427618	24917969	0.8000000	LightGBM
0.9549211	27863335	0.8500000	LightGBM
0.9664287	31569278	0.9000000	LightGBM
0.9780978	37062043	0.9500000	LightGBM
0.9898597	47365558	0.9900000	LightGBM
0.9992667	84060539	0.9999999	LightGBM
0.9330259	22678118	0.7500000	RF
0.9456275	25204980	0.8000000	RF
0.9571634	28150346	0.8500000	RF
0.9680733	31856289	0.9000000	RF
0.9791229	37349054	0.9500000	RF
0.9903377	47652569	0.9900000	RF
0.9993297	84347550	0.9999999	RF
0.9382988	24015438	0.7500000	SVR
0.9498188	26542300	0.8000000	SVR
0.9603919	29487666	0.8500000	SVR
0.9704049	33193609	0.9000000	SVR
0.9806041	38686375	0.9500000	SVR
0.9910860	48989889	0.9900000	SVR
0.9994529	85684870	0.9999999	SVR

Table 4: Stock Control Performance Results for SIM3

Achieved Fill Rate	Holding Costs	Target Fill Rates	Method
0.9529304	33928996	0.7500000	Croston
0.9635260	36333908	0.8000000	Croston
0.9729595	39137126	0.8500000	Croston
0.9815064	42664214	0.9000000	Croston
0.9895083	47891891	0.9500000	Croston
0.9964036	57698143	0.9900000	Croston
0.9999114	92622169	0.9999999	Croston
0.9515647	33637083	0.7500000	SES
0.9625023	36041995	0.8000000	SES
0.9722031	38845213	0.8500000	SES
0.9809908	42372302	0.9000000	SES
0.9892443	47599978	0.9500000	SES
0.9963257	57406230	0.9900000	SES

0.9999094	92330256	0.9999999	SES
0.9491452	33293511	0.7500000	SBA
0.9604985	35698423	0.8000000	SBA
0.9706288	38501641	0.8500000	SBA
0.9798348	42028730	0.9000000	SBA
0.9885058	47256406	0.9500000	SBA
0.9960173	57062658	0.9900000	SBA
0.9999017	91986684	0.9999999	SBA
0.9531562	33934464	0.7500000	TSB
0.9637397	36339376	0.8000000	TSB
0.9731461	39142594	0.8500000	TSB
0.9816550	42669683	0.9000000	TSB
0.9896148	47897359	0.9500000	TSB
0.9964527	57703611	0.9900000	TSB
0.9999123	92627637	0.9999999	TSB
0.9604708	36094001	0.7500000	WSS
0.9692171	38498913	0.8000000	WSS
0.9770498	41302131	0.8500000	WSS
0.9841738	44829219	0.9000000	WSS
0.9908982	50056896	0.9500000	WSS
0.9967843	59863148	0.9900000	WSS
0.9999190	94787174	0.9999999	WSS
0.8916617	26575122	0.7500000	MLP
0.9143356	28980033	0.8000000	MLP
0.9351624	31783252	0.8500000	MLP
0.9545192	35310340	0.9000000	MLP
0.9734371	40538017	0.9500000	MLP
0.9906716	50344268	0.9900000	MLP
0.9998458	85268294	0.9999999	MLP
0.8880512	26424401	0.7500000	LightGBM
0.9115215	28829312	0.8000000	LightGBM
0.9330197	31632530	0.8500000	LightGBM
0.9531232	35159619	0.9000000	LightGBM
0.9725720	40387295	0.9500000	LightGBM
0.9902008	50193547	0.9900000	LightGBM
0.9998036	85117573	0.9999999	LightGBM
0.9511892	34171031	0.7500000	RF
0.9619259	36575943	0.8000000	RF
0.9716288	39379161	0.8500000	RF
0.9805078	42906250	0.9000000	RF

0.9889304	48133926	0.9500000	RF
0.9963961	57940178	0.9900000	RF
0.9999613	92864204	0.9999999	RF
0.8989494	27390812	0.7500000	SVR
0.9201745	29795724	0.8000000	SVR
0.9396063	32598942	0.8500000	SVR
0.9576537	36126031	0.9000000	SVR
0.9753146	41353707	0.9500000	SVR
0.9913827	51159959	0.9900000	SVR
0.9998655	86083985	0.9999999	SVR

Table 5: Stock Control Performance Results for SIM4

Achieved Fill Rate	Holding Costs	Target Fill Rates	Method
0.9405261	20539226	0.7500000	Croston
0.9556926	22449083	0.8000000	Croston
0.9688045	24675255	0.8500000	Croston
0.9801479	27476287	0.9000000	Croston
0.9900675	31627839	0.9500000	Croston
0.9974096	39415460	0.9900000	Croston
0.9999881	67150325	0.9999999	Croston
0.9401541	20410779	0.7500000	SES
0.9555360	22320636	0.8000000	SES
0.9687782	24546807	0.8500000	SES
0.9802245	27347840	0.9000000	SES
0.9901690	31499392	0.9500000	SES
0.9974705	39287013	0.9900000	SES
0.9999878	67021878	0.9999999	SES
0.9351092	19987256	0.7500000	SBA
0.9515103	21897113	0.8000000	SBA
0.9657588	24123285	0.8500000	SBA
0.9781383	26924317	0.9000000	SBA
0.9889960	31075869	0.9500000	SBA
0.9971155	38863490	0.9900000	SBA
0.9999856	66598355	0.9999999	SBA
0.9409140	20504690	0.7500000	TSB
0.9560939	22414547	0.8000000	TSB
0.9691693	24640719	0.8500000	TSB
0.9804654	27441751	0.9000000	TSB
0.9902873	31593303	0.9500000	TSB

0.9974912	39380924	0.9900000	TSB
0.9999878	67115789	0.9999999	TSB
0.9471504	21637420	0.7500000	WSS
0.9604652	23547277	0.8000000	WSS
0.9720167	25773449	0.8500000	WSS
0.9820849	28574481	0.9000000	WSS
0.9908986	32726033	0.9500000	WSS
0.9975715	40513654	0.9900000	WSS
0.9999878	68248519	0.9999999	WSS
0.9306181	19927974	0.7500000	MLP
0.9477223	21837831	0.8000000	MLP
0.9628047	24064002	0.8500000	MLP
0.9761413	26865035	0.9000000	MLP
0.9879609	31016587	0.9500000	MLP
0.9971367	38804208	0.9900000	MLP
0.9999968	66539073	0.9999999	MLP
0.9351113	20605354	0.7500000	LightGBM
0.9512897	22515211	0.8000000	LightGBM
0.9654593	24741382	0.8500000	LightGBM
0.9778589	27542415	0.9000000	LightGBM
0.9888149	31693967	0.9500000	LightGBM
0.9971335	39481587	0.9900000	LightGBM
0.9999923	67216453	0.9999999	LightGBM
0.9388560	20691370	0.7500000	RF
0.9541767	22601227	0.8000000	RF
0.9676148	24827399	0.8500000	RF
0.9793677	27628431	0.9000000	RF
0.9896857	31779983	0.9500000	RF
0.9975420	39567604	0.9900000	RF
0.9999966	67302469	0.9999999	RF
0.9444886	22129773	0.7500000	SVR
0.9584976	24039630	0.8000000	SVR
0.9706437	26265802	0.8500000	SVR
0.9812478	29066834	0.9000000	SVR
0.9905353	33218386	0.9500000	SVR
0.9974948	41006007	0.9900000	SVR
0.9999856	68740872	0.9999999	SVR

Table 6: Stock Control Performance Results for AUTO

Achieved Fill Rate	Holding Costs	Target Fill Rates	Method
0.9412319	3954331	0.7500000	Croston
0.9527760	4318403	0.8000000	Croston
0.9632684	4742773	0.8500000	Croston
0.9730921	5276727	0.9000000	Croston
0.9828719	6068128	0.9500000	Croston
0.9921214	7552664	0.9900000	Croston
0.9989867	12839698	0.9999999	Croston
0.9399131	3884876	0.7500000	SES
0.9518332	4248947	0.8000000	SES
0.9627332	4673317	0.8500000	SES
0.9728111	5207272	0.9000000	SES
0.9827768	5998673	0.9500000	SES
0.9921830	7483209	0.9900000	SES
0.9990225	12770243	0.9999999	SES
0.9347902	3820762	0.7500000	SBA
0.9474310	4184833	0.8000000	SBA
0.9590509	4609203	0.8500000	SBA
0.9699490	5143158	0.9000000	SBA
0.9808249	5934559	0.9500000	SBA
0.9912250	7419095	0.9900000	SBA
0.9989223	12706129	0.9999999	SBA
0.9419402	3949093	0.7500000	TSB
0.9533809	4313165	0.8000000	TSB
0.9638365	4737535	0.8500000	TSB
0.9735879	5271489	0.9000000	TSB
0.9832378	6062890	0.9500000	TSB
0.9923279	7547426	0.9900000	TSB
0.9990240	12834460	0.9999999	TSB
0.9602320	4807963	0.7500000	WSS
0.9675362	5172035	0.8000000	WSS
0.9742376	5596405	0.8500000	WSS
0.9806563	6130359	0.9000000	WSS
0.9872344	6921760	0.9500000	WSS
0.9936833	8406296	0.9900000	WSS
0.9989983	13693330	0.9999999	WSS
0.9409011	4152962	0.7500000	MLP
0.9526591	4517033	0.8000000	MLP
0.9635234	4941403	0.8500000	MLP
0.9735608	5475358	0.9000000	MLP

0.9837912	6266759	0.9500000	MLP
0.9937946	7751295	0.9900000	MLP
0.9995043	13038329	0.9999999	MLP
0.9276593	4120984	0.7500000	LightGBM
0.9415389	4485056	0.8000000	LightGBM
0.9543275	4909426	0.8500000	LightGBM
0.9663761	5443380	0.9000000	LightGBM
0.9785904	6234781	0.9500000	LightGBM
0.9907714	7719317	0.9900000	LightGBM
0.9991882	13006351	0.9999999	LightGBM
0.9403515	4099411	0.7500000	RF
0.9520938	4463483	0.8000000	RF
0.9629561	4887853	0.8500000	RF
0.9729902	5421807	0.9000000	RF
0.9831799	6213208	0.9500000	RF
0.9933384	7697744	0.9900000	RF
0.9993775	12984778	0.9999999	RF
0.9373744	4211868	0.7500000	SVR
0.9496352	4575940	0.8000000	SVR
0.9612008	5000310	0.8500000	SVR
0.9718917	5534264	0.9000000	SVR
0.9827779	6325665	0.9500000	SVR
0.9935748	7810202	0.9900000	SVR
0.9995814	13097235	0.9999999	SVR

Table 7: Stock Control Performance Results for MAN

Achieved Fill Rate	Holding Costs	Target Fill Rates	Method
0.9540260	178090.9	0.7500000	Croston
0.9597055	204826.9	0.8000000	Croston
0.9651039	235990.9	0.8500000	Croston
0.9705284	275202.4	0.9000000	Croston
0.9765834	333319.7	0.9500000	Croston
0.9841197	442338.2	0.9900000	Croston
0.9944322	830596.9	0.9999999	Croston
0.9541098	173629.6	0.7500000	SES
0.9597806	200365.6	0.8000000	SES
0.9651952	231529.6	0.8500000	SES
0.9706384	270741.1	0.9000000	SES
0.9766585	328858.4	0.9500000	SES

0.9841634	437876.9	0.9900000	SES
0.9944388	826135.6	0.9999999	SES
0.9534522	175222.9	0.7500000	SBA
0.9592366	201958.9	0.8000000	SBA
0.9647304	233122.9	0.8500000	SBA
0.9702344	272334.4	0.9000000	SBA
0.9763753	330451.7	0.9500000	SBA
0.9839944	439470.2	0.9900000	SBA
0.9944038	827728.9	0.9999999	SBA
0.9539155	173596.0	0.7500000	TSB
0.9595782	200332.0	0.8000000	TSB
0.9650056	231496.0	0.8500000	TSB
0.9704527	270707.5	0.9000000	TSB
0.9765093	328824.8	0.9500000	TSB
0.9840406	437843.3	0.9900000	TSB
0.9943942	826102.1	0.9999999	TSB
0.9548842	174406.0	0.7500000	WSS
0.9603462	201141.9	0.8000000	WSS
0.9655865	232306.0	0.8500000	WSS
0.9708429	271517.5	0.9000000	WSS
0.9767664	329634.8	0.9500000	WSS
0.9841305	438653.3	0.9900000	WSS
0.9943709	826912.0	0.9999999	WSS
0.9490991	163873.7	0.7500000	MLP
0.9558412	190609.7	0.8000000	MLP
0.9621750	221773.8	0.8500000	MLP
0.9684255	260985.2	0.9000000	MLP
0.9753769	319102.6	0.9500000	MLP
0.9837214	428121.0	0.9900000	MLP
0.9946067	816379.8	0.9999999	MLP
0.9504997	169697.5	0.7500000	LightGBM
0.9568894	196433.5	0.8000000	LightGBM
0.9629246	227597.5	0.8500000	LightGBM
0.9689218	266809.0	0.9000000	LightGBM
0.9755912	324926.4	0.9500000	LightGBM
0.9837919	433944.8	0.9900000	LightGBM
0.9946000	822203.6	0.9999999	LightGBM
0.9514922	172997.3	0.7500000	RF
0.9576611	199733.3	0.8000000	RF
0.9634680	230897.3	0.8500000	RF

0.9693353	270108.8	0.9000000	RF
0.9758717	328226.2	0.9500000	RF
0.9839000	437244.6	0.9900000	RF
0.9945880	825503.4	0.9999999	RF
0.9341203	120333.7	0.7500000	SVR
0.9435108	147069.7	0.8000000	SVR
0.9522719	178233.8	0.8500000	SVR
0.9608121	217445.2	0.9000000	SVR
0.9699025	275562.6	0.9500000	SVR
0.9805230	384581.0	0.9900000	SVR
0.9937597	772839.8	0.9999999	SVR

Table 8: Stock Control Performance Results for BRAF

Achieved Fill Rate	Holding Costs	Target Fill Rates	Method
0.9578326	126231.37	0.7500000	Croston
0.9624342	148124.70	0.8000000	Croston
0.9671440	173644.05	0.8500000	Croston
0.9722001	205753.23	0.9000000	Croston
0.9781653	253343.85	0.9500000	Croston
0.9856840	342615.95	0.9900000	Croston
0.9950540	660549.99	0.9999999	Croston
0.9569366	120141.35	0.7500000	SES
0.9616763	142034.68	0.8000000	SES
0.9665379	167554.03	0.8500000	SES
0.9717553	199663.20	0.9000000	SES
0.9778757	247253.83	0.9500000	SES
0.9855711	336525.93	0.9900000	SES
0.9950869	654459.97	0.9999999	SES
0.9571103	123117.16	0.7500000	SBA
0.9617683	145010.49	0.8000000	SBA
0.9665544	170529.84	0.8500000	SBA
0.9716928	202639.01	0.9000000	SBA
0.9777584	250229.64	0.9500000	SBA
0.9854313	339501.74	0.9900000	SBA
0.9950077	657435.77	0.9999999	SBA
0.9567939	119180.67	0.7500000	TSB
0.9615098	141074.00	0.8000000	TSB
0.9663631	166593.35	0.8500000	TSB
0.9715714	198702.52	0.9000000	TSB

0.9776890	246293.15	0.9500000	TSB
0.9854090	335565.25	0.9900000	TSB
0.9950061	653499.28	0.9999999	TSB
0.9607159	138632.46	0.7500000	WSS
0.9650520	160525.79	0.8000000	WSS
0.9694592	186045.14	0.8500000	WSS
0.9741274	218154.31	0.9000000	WSS
0.9795305	265744.94	0.9500000	WSS
0.9863493	355017.04	0.9900000	WSS
0.9950781	672951.08	0.9999999	WSS
0.9572978	128084.45	0.7500000	MLP
0.9617898	149977.78	0.8000000	MLP
0.9664326	175497.13	0.8500000	MLP
0.9714246	207606.30	0.9000000	MLP
0.9775011	255196.93	0.9500000	MLP
0.9853723	344469.03	0.9900000	MLP
0.9951321	662403.07	0.9999999	MLP
0.9576885	130036.45	0.7500000	LightGBM
0.9621824	151929.78	0.8000000	LightGBM
0.9668214	177449.13	0.8500000	LightGBM
0.9718135	209558.30	0.9000000	LightGBM
0.9778337	257148.93	0.9500000	LightGBM
0.9855804	346421.03	0.9900000	LightGBM
0.9951765	664355.07	0.9999999	LightGBM
0.9577485	130132.88	0.7500000	RF
0.9622396	152026.21	0.8000000	RF
0.9668761	177545.57	0.8500000	RF
0.9718558	209654.74	0.9000000	RF
0.9778733	257245.36	0.9500000	RF
0.9856015	346517.46	0.9900000	RF
0.9951830	664451.50	0.9999999	RF
0.9476785	88149.92	0.7500000	SVR
0.9532511	110043.25	0.8000000	SVR
0.9589565	135562.60	0.8500000	SVR
0.9651584	167671.77	0.9000000	SVR
0.9726136	215262.40	0.9500000	SVR
0.9823069	304534.50	0.9900000	SVR
0.9943331	622468.53	0.9999999	SVR

Table 9: Stock Control Performance Results for OIL

Achieved Fill Rate	Holding Costs	Target Fill Rates	Method
0.9665717	3465440	0.7500000	Croston
0.9697738	4126540	0.8000000	Croston
0.9732208	4897133	0.8500000	Croston
0.9771034	5866715	0.9000000	Croston
0.9820266	7303781	0.9500000	Croston
0.9888610	9999478	0.9900000	Croston
0.9976286	19599944	0.9999999	Croston
0.9657696	3069932	0.7500000	SES
0.9690380	3731032	0.8000000	SES
0.9725426	4501624	0.8500000	SES
0.9764711	5471206	0.9000000	SES
0.9814701	6908272	0.9500000	SES
0.9885016	9603969	0.9900000	SES
0.9976217	19204436	0.9999999	SES
0.9657764	3417629	0.7500000	SBA
0.9690150	4078729	0.8000000	SBA
0.9725105	4849322	0.8500000	SBA
0.9764481	5818903	0.9000000	SBA
0.9814575	7255969	0.9500000	SBA
0.9884612	9951666	0.9900000	SBA
0.9975912	19552133	0.9999999	SBA
0.9655907	3194752	0.7500000	TSB
0.9688495	3855852	0.8000000	TSB
0.9723536	4626445	0.8500000	TSB
0.9762795	5596027	0.9000000	TSB
0.9812983	7033093	0.9500000	TSB
0.9883591	9728790	0.9900000	TSB
0.9975626	19329256	0.9999999	TSB
0.9699824	3099716	0.7500000	WSS
0.9729735	3760816	0.8000000	WSS
0.9760970	4531409	0.8500000	WSS
0.9795943	5500991	0.9000000	WSS
0.9839398	6938057	0.9500000	WSS
0.9899118	9633754	0.9900000	WSS
0.9976517	19234220	0.9999999	WSS
0.9642840	4078289	0.7500000	MLP
0.9675306	4739389	0.8000000	MLP
0.9710644	5509982	0.8500000	MLP

0.9751403	6479563	0.9000000	MLP
0.9803786	7916629	0.9500000	MLP
0.9878198	10612326	0.9900000	MLP
0.9976471	20212793	0.9999999	MLP
0.9648658	3822261	0.7500000	LightGBM
0.9680998	4483361	0.8000000	LightGBM
0.9716163	5253953	0.8500000	LightGBM
0.9756473	6223535	0.9000000	LightGBM
0.9808001	7660601	0.9500000	LightGBM
0.9880753	10356298	0.9900000	LightGBM
0.9976902	19956765	0.9999999	LightGBM
0.9649491	3819005	0.7500000	RF
0.9681783	4480105	0.8000000	RF
0.9716795	5250697	0.8500000	RF
0.9757027	6220279	0.9000000	RF
0.9808406	7657345	0.9500000	RF
0.9881044	10353042	0.9900000	RF
0.9976940	19953509	0.9999999	RF
0.9580213	2664360	0.7500000	SVR
0.9616694	3325460	0.8000000	SVR
0.9656236	4096052	0.8500000	SVR
0.9702389	5065634	0.9000000	SVR
0.9762528	6502700	0.9500000	SVR
0.9849641	9198397	0.9900000	SVR
0.9971200	18798864	0.9999999	SVR

Table 10: Stock Control Performance Results for MAN2

Achieved Fill Rate	Holding Costs	Target Fill Rates	Method
0.9607984	134107.7	0.7500000	Croston
0.9652132	153956.9	0.8000000	Croston
0.9695271	177093.5	0.8500000	Croston
0.9738679	206204.6	0.9000000	Croston
0.9788012	249351.8	0.9500000	Croston
0.9850397	330288.6	0.9900000	Croston
0.9941704	618537.2	0.9999999	Croston
0.9621387	134519.2	0.7500000	SES
0.9663492	154368.4	0.8000000	SES
0.9704312	177505.0	0.8500000	SES
0.9746131	206616.1	0.9000000	SES

0.9793592	249763.3	0.9500000	SES
0.9854266	330700.1	0.9900000	SES
0.9943039	618948.7	0.9999999	SES
0.9606371	133325.1	0.7500000	SBA
0.9650803	153174.2	0.8000000	SBA
0.9694209	176310.8	0.8500000	SBA
0.9737856	205422.0	0.9000000	SBA
0.9787475	248569.1	0.9500000	SBA
0.9850143	329505.9	0.9900000	SBA
0.9941670	617754.6	0.9999999	SBA
0.9621101	134929.5	0.7500000	TSB
0.9663718	154778.7	0.8000000	TSB
0.9704843	177915.3	0.8500000	TSB
0.9746520	207026.4	0.9000000	TSB
0.9793732	250173.5	0.9500000	TSB
0.9854105	331110.3	0.9900000	TSB
0.9942759	619359.0	0.9999999	TSB
0.9659234	140536.5	0.7500000	WSS
0.9696704	160385.6	0.8000000	WSS
0.9732917	183522.3	0.8500000	WSS
0.9769742	212633.4	0.9000000	WSS
0.9812893	255780.5	0.9500000	WSS
0.9867441	336717.3	0.9900000	WSS
0.9947260	624966.0	0.9999999	WSS
0.9598025	130536.7	0.7500000	MLP
0.9643880	150385.9	0.8000000	MLP
0.9687945	173522.5	0.8500000	MLP
0.9733397	202633.6	0.9000000	MLP
0.9785522	245780.7	0.9500000	MLP
0.9850632	326717.5	0.9900000	MLP
0.9943775	614966.2	0.9999999	MLP
0.9594519	132727.3	0.7500000	LightGBM
0.9641028	152576.5	0.8000000	LightGBM
0.9685434	175713.1	0.8500000	LightGBM
0.9731107	204824.2	0.9000000	LightGBM
0.9782540	247971.3	0.9500000	LightGBM
0.9847469	328908.1	0.9900000	LightGBM
0.9941704	617156.8	0.9999999	LightGBM
0.9605543	133530.5	0.7500000	RF
0.9649519	153379.7	0.8000000	RF

0.9692067	176516.3	0.8500000	RF
0.9735904	205627.4	0.9000000	RF
0.9786494	248774.6	0.9500000	RF
0.9850164	329711.4	0.9900000	RF
0.9942609	617960.0	0.9999999	RF
0.9510407	108955.4	0.7500000	SVR
0.9573325	128804.6	0.8000000	SVR
0.9630658	151941.2	0.8500000	SVR
0.9688106	181052.4	0.9000000	SVR
0.9752300	224199.5	0.9500000	SVR
0.9831752	305136.3	0.9900000	SVR
0.9939351	593384.9	0.9999999	SVR

Table 11: Stock Control Performance Results for BRAF2

Achieved Fill Rate	Holding Costs	Target Fill Rates	Method
0.9574853	123532.5	0.7500000	Croston2
0.9620382	144700.3	0.8000000	Croston2
0.9667303	169373.9	0.8500000	Croston2
0.9717694	200418.9	0.9000000	Croston2
0.9777347	246432.3	0.9500000	Croston2
0.9852644	332745.8	0.9900000	Croston2
0.9948358	640143.1	0.9999999	Croston2
0.9564361	117024.6	0.7500000	SES
0.9611344	138192.3	0.8000000	SES
0.9659675	162865.9	0.8500000	SES
0.9711737	193911.0	0.9000000	SES
0.9773130	239924.4	0.9500000	SES
0.9850873	326237.9	0.9900000	SES
0.9948516	633635.1	0.9999999	SES
0.9566271	119870.6	0.7500000	SBA
0.9612469	141038.3	0.8000000	SBA
0.9660161	165711.9	0.8500000	SBA
0.9711430	196756.9	0.9000000	SBA
0.9772289	242770.3	0.9500000	SBA
0.9849707	329083.8	0.9900000	SBA
0.9947729	636481.1	0.9999999	SBA
0.9560356	115137.9	0.7500000	TSB
0.9607322	136305.6	0.8000000	TSB
0.9655758	160979.2	0.8500000	TSB

0.9707999	192024.3	0.9000000	TSB
0.9769767	238037.7	0.9500000	TSB
0.9848335	324351.2	0.9900000	TSB
0.9947489	631748.4	0.9999999	TSB
0.9608181	136684.1	0.7500000	WSS
0.9650920	157851.9	0.8000000	WSS
0.9694281	182525.5	0.8500000	WSS
0.9740146	213570.5	0.9000000	WSS
0.9793440	259583.9	0.9500000	WSS
0.9860885	345897.4	0.9900000	WSS
0.9949016	653294.7	0.9999999	WSS
0.9569176	125361.2	0.7500000	MLP
0.9613639	146529.0	0.8000000	MLP
0.9659692	171202.6	0.8500000	MLP
0.9709327	202247.6	0.9000000	MLP
0.9770040	248261.0	0.9500000	MLP
0.9849183	334574.5	0.9900000	MLP
0.9949119	641971.8	0.9999999	MLP
0.9573767	127925.3	0.7500000	LightGBM
0.9618161	149093.0	0.8000000	LightGBM
0.9664122	173766.6	0.8500000	LightGBM
0.9713773	204811.7	0.9000000	LightGBM
0.9773776	250825.1	0.9500000	LightGBM
0.9851547	337138.6	0.9900000	LightGBM
0.9949656	644535.8	0.9999999	LightGBM
0.9575004	128170.2	0.7500000	RF
0.9618920	149337.9	0.8000000	RF
0.9664501	174011.6	0.8500000	RF
0.9713503	205056.6	0.9000000	RF
0.9773460	251070.0	0.9500000	RF
0.9851391	337383.5	0.9900000	RF
0.9949702	644780.7	0.9999999	RF
0.9470923	85318.5	0.7500000	SVR
0.9526118	106486.3	0.8000000	SVR
0.9582867	131159.9	0.8500000	SVR
0.9644645	162204.9	0.9000000	SVR
0.9719274	208218.3	0.9500000	SVR
0.9817076	294531.8	0.9900000	SVR
0.9940549	601929.0	0.9999999	SVR

Table 12: Stock Control Performance Results for OIL2

Achieved Fill Rate	Holding Costs	Target Fill Rates	Method
0.9591945	1286573.2	0.7500000	Croston
0.9627143	1497313.7	0.8000000	Croston
0.9665239	1742957.4	0.8500000	Croston
0.9708028	2052033.4	0.9000000	Croston
0.9762578	2510130.4	0.9500000	Croston
0.9839949	3369444.4	0.9900000	Croston
0.9952868	6429808.6	0.9999999	Croston
0.9584980	1164858.2	0.7500000	SES
0.9620624	1375598.6	0.8000000	SES
0.9659166	1621242.4	0.8500000	SES
0.9702511	1930318.3	0.9000000	SES
0.9757876	2388415.4	0.9500000	SES
0.9837043	3247729.3	0.9900000	SES
0.9952819	6308093.5	0.9999999	SES
0.9584252	1244245.3	0.7500000	SBA
0.9619803	1454985.8	0.8000000	SBA
0.9658419	1700629.5	0.8500000	SBA
0.9701834	2009705.4	0.9000000	SBA
0.9757452	2467802.5	0.9500000	SBA
0.9836565	3327116.4	0.9900000	SBA
0.9952418	6387480.6	0.9999999	SBA
0.9634562	1377243.6	0.7500000	TSB
0.9666424	1587984.1	0.8000000	TSB
0.9701038	1833627.8	0.8500000	TSB
0.9740090	2142703.7	0.9000000	TSB
0.9790328	2600800.8	0.9500000	TSB
0.9862464	3460114.8	0.9900000	TSB
0.9965396	6520479.0	0.9999999	TSB
0.9632704	1282248.2	0.7500000	WSS
0.9665237	1492988.6	0.8000000	WSS
0.9699528	1738632.4	0.8500000	WSS
0.9738104	2047708.3	0.9000000	WSS
0.9786499	2505805.4	0.9500000	WSS
0.9854830	3365119.3	0.9900000	WSS
0.9954594	6425483.5	0.9999999	WSS
0.9580828	1315282.6	0.7500000	MLP
0.9616183	1526023.0	0.8000000	MLP
0.9654459	1771666.8	0.8500000	MLP

0.9698468	2080742.7	0.9000000	MLP
0.9755083	2538839.8	0.9500000	MLP
0.9836492	3398153.7	0.9900000	MLP
0.9955941	6458517.9	0.9999999	MLP
0.9575057	1274718.7	0.7500000	LightGBM
0.9610686	1485459.1	0.8000000	LightGBM
0.9649410	1731102.9	0.8500000	LightGBM
0.9693992	2040178.8	0.9000000	LightGBM
0.9751109	2498275.9	0.9500000	LightGBM
0.9833241	3357589.8	0.9900000	LightGBM
0.9955017	6417954.0	0.9999999	LightGBM
0.9576830	1280751.6	0.7500000	RF
0.9612459	1491492.1	0.8000000	RF
0.9651013	1737135.8	0.8500000	RF
0.9695357	2046211.8	0.9000000	RF
0.9752197	2504308.8	0.9500000	RF
0.9834015	3363622.8	0.9900000	RF
0.9955151	6423987.0	0.9999999	RF
0.9495205	849141.2	0.7500000	SVR
0.9535429	1059881.7	0.8000000	SVR
0.9579280	1305525.4	0.8500000	SVR
0.9630431	1614601.4	0.9000000	SVR
0.9697118	2072698.4	0.9500000	SVR
0.9794829	2932012.4	0.9900000	SVR
0.9944677	5992376.6	0.9999999	SVR