**Intermittent Demand Forecasting Using XGBoost Method under Linearly Decreasing Process**

**Master Thesis**

Data Science and Marketing Analytics

Erasmus School of Economics

Erasmus University Rotterdam

| | |
|---|---|
| Submitted by: | Chang, Chin Chung |
| Student no.: | 577151 |
| Submitted on: | 13.08.2021 |
| Supervisor: | Prof. dr. ir. Rommert Dekker |
| Co-reader: | Dr. Erjen van Nierop |

2021

# Contents

**Abstract**

Intermittent demand forecasting is a significant issue in the industries. Predicting trends and demands could be difficult, resulting in a high stock cost and losing business opportunities. Therefore, an accurate forecasting method is a must to prevent corporations from suffering from the loss. This research compares the XGBoost model and other conventional methods under multiple trends and item patterns with both the simulated and empirical datasets. Accuracy rate (MASE and sMAPE) and inventory control (tradeoff curve) are applied to test the performance of different methods. The results indicate that XGB outperforms other methods under multiple trends. The XGBoost method also shows its superiority in predicting the category of Intermittent and Lumpy.

# 1 Introduction

## 1.1 Motivation of experiment

The lumpiness and intermittency result from irregular demand size and demand intervals have considerably impacted specific industries. According to Gallagher et al. (2005), the spare parts market has been a business that reached 400 billion per year. Beata Ślusarczyk (2018) also claim that it is incredibly vital that the spare parts are provided in time for air transport because the delay of specific materials raises amounts of costs. Kim et al. (2016) assert that spare parts availability is also essential to customer satisfaction and lowering inventory costs.

Thus, an accurate prediction for spare parts demand is crucial in business operation. The prediction higher than actual demand leads to high stock cost while that lower than the actual makes companies lose business opportunities since the variation in non-zero demand size and zero demand that seemed to occur randomly make the prediction more difficult.

The mainstream forecasting methods includes Simple Exponential Smoothing method (SES) (Brown (1960)), the Croston method (Croston (1972)) , and its advanced version(Syntetos, Boylan (2005); Teunter et al. (2011)), and the bootstrapping methods (Willemain et al. (1994); Porras, Dekker (2008)). As the machine learning techniques keep developing rapidly, they are applied to forecast intermittent spare parts differently from traditional methods, giving credit to faster and stronger computing capability nowadays. Smyl (2020) argues that the success of machine learning algorithms lies in identifying the non-linearity of demand without a hypothesis distribution. Yet, the comparative studies that include machine learning methods are not the majority. Rather, most of the studies that compare multiple techniques focus only on the Croston method and its alterations.

What is more, despite the prevalence of the Neural Network and support vector machine (SVM) model in forecasting intermittent demand, the other machine learning methods, such as XGBoost, are less mentioned in this field. Originated from the gradient tree boosting machine (GBM), XGBoost(eXtreme Gradient Boosting) is a state-of-the-art tree-based method in forecasting, which can be applied as in classification or regression. It has been proven successful in applications in many industries. For example, this ensemble tree-based technique has been proven to prevent models from overfitting and has better performance and efficiency than SVM in forecasting wind turbine fault detection ( Zhang et al. (2018)).

The applications of the machine learning method have been conducted in recent years. Makridakis et al. (2018) compares the several classic machine learning methods except for the XGBoost technique with eight traditional statistical ones using the data from M3 competition. According to Makridakis et al. (2018), the accuracy of the machine learning method compared to that of standard statistical techniques come under question. The performance of statistical ones dominates among accuracy measures. However, none of these eight statistical methods include the Croston method, its variants, and bootstrapping methods.

Even though Kiefer et al. (2021) mention XGBoost and the Croston method, there is no comparison between XGBoost and other statistical methods. Thus, there is a vacuum left between XGBoost and those methods commonly used in predicting intermittent spare parts demand. This paper mainly applies traditional statistical methods and general XGBoost in intermittent spare parts forecasting to find out the insights by comparing them.

In addition, it is noteworthy that most studies forecasting using XGBoost are only evaluated by accuracy rate instead of inventory control. Inventory performance measures should also be considered a vital measure in spare parts forecasting because it is not only more practical for firms in stock control but also the accuracy measure cannot ensure an excellent stock performance. The latter reason is mentioned in several studies Syntetos, Boylan (2006); Teunter, Duncan (2009); Syntetos et al. (2010).

At last, a decreasing trend or obsolescence of items is also worth considering because this is quite common in business operation and inventory management. Pince, Dekker (2010) mention that the dead stock that accounts for 5 percent of all inventory in aerospace industries causes 1 million dollars in total value in aerospace industries. Sugiono, Alimbudiono (2020) also note that the stocks brought by slow-moving items are a big issue in ceramic tiles industries. Therefore, taking this obsolescence into account,Teunter et al. (2011) propose a new method based on the Croston method. One of the purposes is to evaluate if the XGBoost way captures the decreasing trends, leading to better performance.

## 1.2 Research questions

Thus, according to the motivation in the previous section, the research contribution leads to two research questions as follows:

**Research Question 1: Does XGBoost have a better prediction result than traditional statistical methods in forecasting intermittent demand, especially under the evaluation of inventory control?**

Regarding research question 1, multiple methods are applied to forecast intermittent items in simulated and empirical datasets, and are tested by forecasting accuracy and inventory performance in the following chapters. It is noteworthy that we offer a comparative study between a machine learning method and other conventional techniques. Specifically, it is emphasized that inventory control is introduced to measure performance for a machine learning method, which is less frequent used than accuracy measurement is.

**Research Question 2: Under which demand patterns does XGBoost perform better than conventional forecast models?**

Regarding research question 2, we consider a decreasing trend to test the performance of the XGBoost method and other conventional ones under such a pattern.

**Research Question 3: What is the relationships between the forecasting**

**performance and computation cost for XGBoost in prediction?**

Regarding research question 3, it is no doubt that the tradeoff between forecasting performance and time of model training exists and is discussed in our research.

Overall, we summarize the contribution of our research below:

1. Since XGBoost is rarely mentioned in this domain, we evaluate the performance of XGBoost in intermittent forecasting.
2. A comparative study between a machine learning method and other conventional techniques is provided.
3. Less frequently used than accuracy measurement, the measurement of inventory control is introduced to test the performance for a machine learning method.
4. A decreasing trend in demands is considered to test the performance of XGBoost.

## 1.3 Research Design

Figure 1 indicates that the flow of the forecasting process and presents a clear picture of this paper. First, we generate simulated datasets divided into training data and test data with several patterns using a hybrid method simply combining Syntetos et al. (2005) classification scheme and linearly decreasing demand processes in two cases. These patterns include lumpy, erratic, intermittent, and smooth, and along with the decreasing demand process, are discussed in detail in the following section. On the other side, the empirical datasets are collected and also be identified if there's any pattern contained. Second, both simulated datasets and empirical datasets are analyzed by five forecasting methods mentioned in Chapter 3, including SBA, TSB, the EMP ( the empirical method proposed by Porras and Dekker), WSS, and XGBoost. Finally, we compare the outcome of different methods evaluated through both accuracy measures and inventory performance.

More specifically, the structure of this thesis is organized as follows. In Chapter 2, forecasting methods are introduced, and we further describe the XGBoost model in detail.

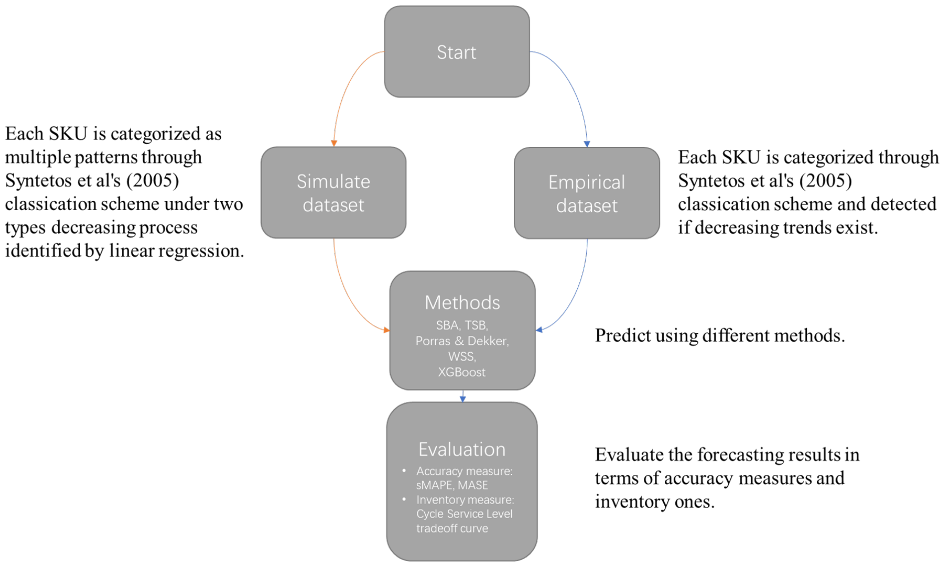



Figure 1: The flow of experiment design

In Chapter 3, a brief related work of forecasting methods are provided. In Chapter 4, we introduce the data descriptions, especially the data classification scheme classifying SKUs (stock keeping units), the empirical data, and the method that simulates a certain decreasing pattern of data. In Chapter 5, evaluation methods are proposed. In Chapter 6, the results from the comparisons between statistical methods and the XGBoost approach are presented for both simulated data and empirical data. Also, the cost of time training a model along with model performance are investigated. In Chapter 7, the conclusions and limitations are provided.

## 2 Brief Literature Review

Spare parts demand forecasting is always tricky due to the bumpiness. According to Kennedy et al. (2002), the irregularity comes from several aspects, one of which is production characteristics. Thus, the intermittency and lumpiness can be viewed as an issue related to the variability of demand and sporadicity of the interval period. In other words, the variation for both demand size and interval length can be an uncertainty followed by a parametric distribution. Those two factors, demand size and demand interval, result in different demand patterns, increasing the prediction complexity. Subsequently, the data classification scheme is incorporated into this paper to offer further practical insights. The categorization includes the smooth, intermittent, erratic, and lumpy proposed by Syntetos et al. (2005). In some circumstances, the item of the spare part involves decreasing demand processes. According to Pinçe Çerag, Dekker (2011), this decreasing trend brought by obsolescence indicates that spare parts items phase out over time, leading to a risk of high inventory costs due to the excess holding stocks resulting from the obsolete forecast. Subsequently, the thing with the trend finally becomes very lumpy (slow-moving). Since the different data patterns complicate prediction, the accuracy of prediction is significant for those companies that try to lower their inventory cost or avoid the loss of business opportunities when facing the substantial operational impact brought by intermittency and lumpiness.

Conventional methods such as simple exponential smoothing (SES) and moving average approach are frequently applied to predict spare parts demands for most companies for decades. Yet, these traditional methods are not accurate compared to other methods since Croston (1972) introduces another alternative method, an approach based on exponential smoothing, separating the forecast into two elements that estimate demand size and inter-demand intervals. The recent period under the SES application has heavier weights because the simple exponential method focuses only on the non-zero demand period, doesn't consider periods with zero demands, and treats them equally, leading to bias. Croston (1972) assumes that the demand intervals follow the geometric distribution and that the demand size follows the normal distribution. However, Willemain et al.

(1994) challenge this assumption because the correlations and distribution in industrial data show that this is not the case. Subsequently,Syntetos, Boylan (2005) offer a modified version of the Croston method by adding a bias correction coefficient. Furthermore, Teunter–Syntetos–Babai method, also known as TSB (Teunter et al. (2011)) , is introduced to be an improvement based on the Croston method and taking obsolescence into account.

In addition to the methods above, which are categorized as parametric methods, the mainstream methods also include non-parametric bootstrapping models. Willemain et al. (2004) propose an approach based on classic bootstrapping, using a two-state Markov process and a jittering procedure. Syntetos et al. (2015) argue that the Willemain method (WSS) has superior performance than SBA because the inter-demand interval is erratic and with short lead times. Yet, this case is not the other way around. The SBA outperforms WSS when demand is more erratic with longer lead times. Porras, Dekker (2008) introduce a new non-parametric method in which the histogram of the lead time demand is constructed through a series of successive demand windows over the data's historical horizon. In general, those bootstrapping methods show good performance when applied to the empirical data, compared to parametric methods.

The earliest study using the machine learning technique in predicting time series dataset is proposed by Gutierrez et al. (2008). They point out that the neural network technique solves the shortcomings of the conventional forecasting method, including the unable to capture nonlinear patterns in data occasionally and the bias brought by outliers. According to the result, the neural network model outperforms the traditional methods (SES, Croston, and SBA) in predicting lumpy demand from a real-world data series. Another method based on machine learning technology is Support Vector Machine (SVM) in predicting intermittent demand. Kaya, Turkyilmaz (2018) compare three machine learning methods, including artificial neural networks, support vector machines, and decision tree models. They claim that the support vector machine has the most appropriate performance with the lowest error rate. Jiang et al. (2020) argue that adaptive univariate SVM, an extension of the SVM model, achieves a statistically significant accuracy improvement and better inventory performance for the SKU with non-smooth demand series.

Proposed by Chen, Guestrin (2016), the XGBoost method is applied in some papers regarding time series prediction in recent days. WANG et al. (2017) propose the XGBoost approach based on Discrete Wavelet Transform that outperforms DWT-SVR (support vector regression based on Discrete Wavelet Transform) and DWT-ANN (artificial neural network based on Discrete Wavelet Transform) in electricity consumption dataset. Alim et al. (2020) use a hybrid model containing XGBoost and ARIMA methods to improve the predictive performance of stock market volatility. Alim et al. (2020) compare the ARIMA model and XGBoost model to predict human brucellosis in mainland China and show the XGBoost model's performance is better than that of the ARIMA model. Apart from the papers mentioned above, the related research in intermittent demand forecasting has come along recently. Kiefer et al. (2021) compares methods from statistical, machine learning, and deep learning in predicting Intermittent and lumpy demand, using SPEC (Stock-keeping-oriented Prediction Error Costs) metric and MASE (Mean Absolute Scaled Error) to evaluate performance. The result indicates that the XGBoost method ranks 6th out of 9 approaches when comparing by MASE, better than the Croston method. As we discuss in the last chapter because there is a gap between the XGBoost algorithm and other conventional techniques, one of the objects of this paper is to compare them both numerically and empirically.

# 3 Forecasting Methods

In our work, we discuss different methods, including parametric methods, bootstrapping methods, and machine learning methods. These methods are applied to both simulated data and the empirical part of this thesis work. After presenting the notation and equation, we introduce the discussions of these methods. Apart from these methods mentioned above, the zero-forecast process is also considered a baseline, i.e., a benchmark forecasting method.

## 3.1 Parametric Methods

Parametric methods don't assume that the demand follows a specific distribution. Mainstream parametric methods include the Croston method and its modifications. Before the Croston method is proposed, exponential smoothing is widely used in forecasting spare parts demands. However, the shortcoming is evident due to the forecasts made only in periods with non-zero demand.

### 3.1.1 Notation

The notation is applied in the remainder of this section of parametric methods:

$z_t$: The actual demand size in $t$ period

$z'_t$: The demand forecast for the next period in period $t$ .

$s_t$: The actual inter-demand intervals in period $t$

$s'_t$: The estimate of inter-demand intervals in period $t$ for period $t+1$

$Y$ : The actual mean demand per period made in period t for period $t+1$

$Y'_t$: The estimate of average demand per period made in period t for period $t+1$

$p_t$: The probability of non-zero demand occurs in period $t$ for period t

$p'_t$: The estimate of the probability of non-zero demand occurs in period $t$ for period t

$\alpha$: The smoothing parameter that is less than 0

$\beta$: The smoothing parameter that is less than 1

### 3.1.2 The Croston method

Croston (1972) introduce the Croston method correcting this drawback through the forecast of two elements, $z$, demand size, and, $s$, inter-demand intervals. Subsequently, non-zero demand size and inter-demand intervals are both modeled through exponential smoothing and only updated when demand size is positive or otherwise remain the same $s'_t$ and $z'_t$ in last period $s'_{t-1}$ and $z'_{t-1}$. The forecast is written as the quotient of demand size forecast and inter-demand intervals forecast. This method is given by:

$$\text{if } z_t = 0 : \ s'_t = s'_{t-1},\ z'_t = z'_{t-1}$$
$$\text{if } z_t \neq 0 : \ s'_t = \alpha s_t + (1-\alpha)s'_{t-1},\ z'_t = \alpha z_t + (1-\alpha)z'_{t-1}$$

$$Y'_t = \tfrac{z'_t}{p'_t}$$

where the $z_t$ denotes the actual demand size in $t$ period, the $z'_t$ denotes the demand forecast for the next period in $t$ period. $s_t$ is the actual inter-demand intervals, $s'_t$ the estimate of inter-demand intervals for period $t+1$ . $Y'_t$ denotes the estimate of average demand per period made in period t for period $t+1$. $\alpha$ is the smoothing parameter.

Even though Croston is always used as a benchmark, in most cases, Croston's modifications have more superior performance than Croston itself. For this reason, we exclude the Croston method in comparison in this thesis and only keep its modification methods as followed.

### 3.1.3 The SBA method

Syntetos–Boylan approximation, known as the SBA method, is proposed by Syntetos, Boylan (2005). Pointing out that the Croston method is biased, the authors developed this SBA method by providing a bias correction. Incorporating this bias approximation, SBA method is described as follows:

$$\text{if } z_t = 0 : \; s'_t = s'_{t-1}, \; z'_t = z'_{t-1}$$
$$\text{if } z_t \neq 0 : \; s'_t = \alpha s_t + (1-\alpha)s'_{t-1}, \; z'_t = \alpha z_t + (1-\alpha)z'_{t-1}$$

$$Y'_t = \left(1 - \tfrac{\alpha}{2}\right)\tfrac{z'_t}{p'_t}$$

where $1 - \frac{\alpha}{2}$ is the correction coefficient with a constant alpha, the $Y'_t$ the new demand forecast in the end of period $t$. Yet, the modified equation with separate smoothing parameters initially proposed by Schultz (1987) are introduced in some researches ( Teunter et al. (2011); Babai et al. (2019)). The equation modified is provided below:

$$\text{if } z_t = 0 : \; s'_t = s'_{t-1}, \; z'_t = z'_{t-1}$$
$$\text{if } z_t \neq 0 : \; s'_t = \alpha s_t + (1-\alpha)s'_{t-1}, \; z'_t = \beta z_t + (1-\beta)z'_{t-1}$$

$$Y'_t = \left(1 - \tfrac{\beta}{2}\right)\tfrac{z'_t}{p'_t}$$

The two different smoothing parameters are used to achieve the best performance for the modifications of the Croston method.

### 3.1.4 The TSB method

Teunter et al. (2011) develop a method called TSB, a new method originating from Croston by replacing inter-demand interval forecast with demand probability forecast while remaining part of demand size forecast. Considering the obsolescence in inventory control, TSB updates the demand probability right after zero demand happens. Also, Teunter et al. (2011) claim that this method is unbiased compared to the Croston method and argue that the result of the TSB method indicates good performance in linking intermittent forecasting and inventory obsolescence. The TSB method is written:

$$\begin{aligned}
&\text{if } p_t = 0: \; p'_t = p'_{t-1} + \beta(0 - p'_{t-1}), \; z'_t = z'_{t-1} \\
&\text{if } p_t = 1: \; p'_t = p'_{t-1} + \beta(1 - p'_{t-1}), \; z'_t = z'_{t-1} + \alpha(z_t - z'_{t-1})
\end{aligned}$$

$$Y'_t = p'_t z'_t$$

where $p_t$ and $z_t$ denotes the probability and demand size respectively. The probability is either 0 or 1 at period $t$ and the forecast probability at time $t$ can be written by the exponential smoothing. The demand size part follows what the Croston method does, only updated when the positive probability occurs. The final forecast $Y_t$ is the product of probability and demand size. Teunter et al. (2011) argue that the result of TSB method indicate a good performance in linking the intermittent forecasting and inventory obsolescence.

## 3.2 Non-parametric Methods

Unlike parametric approaches, non-parametric methods do not follow a specific distribution; instead, non-parametric methods obtain the lead-time demand directly from a dataset. Amongst non-parametric forms, the bootstrapping techniques are one of the most common ones. Hasni et al. (2019) assert the bootstrapping techniques use resampling with replacement to draw a virtual copy of the real population. In this paper, we consider two bootstrapping methods, including the method introduced by Willemain et al. (2004) and Porras, Dekker (2008).

### 3.2.1 The WSS method

Willemain et al. (2004) create a Markov process to estimate transition probabilities based on the historical data, and thus this process captures and models autocorrelation. Subsequently, according to transition probabilities, a forecast of zero and non-zero values are generated over the forecast horizon. Each non-zero value marked in the set of forecasts is replaced by a numerical value sampled randomly from non-zero historical demands. Consequently, the predicted non-zero demand size are obtained after jittering and summing over the horizon. Finally, the distribution of lead time demand is generated based on the above procedure. The WSS approach has been proven to be more accurate than exponential smoothing and Croston's Method as an advanced bootstrapping method.

### 3.2.2 The EMP method

Porras, Dekker (2008) introduce an empirical model (EMP) that estimates the LTD distribution, a histogram of demands over lead time. The function without sampling is simpler than bootstrapping and is explained as follows. First of all, a fixed lead time is obtained from the historical data. Moreover, we move the lead time range one period at a time, and the sum of the lead time demand is calculated after each move. Finally, the empirical distribution of lead time demands is constructed by repeating the method moving this range. Porras, Dekker (2008) claim that the empirical method has superior performance to the WSS method.

## 3.3 Machine learning Methods

Based on the gradient boosting machine introduced by Friedman (2001), Chen, Guestrin (2016) propose the XGBoost that is categorized as a supervised learning method in forecasting. XGBoost is also a tree ensemble model, which denotes that the sum of the predictions of each leaf in multiple trees equals the prediction for each example:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F$$

where $F$ denotes the regression trees space, $f$ the function in $F$, and $K$ means the trees numbers.

XGBoost can be applied to not only regression but also classification issues. The difference lies in the objective function that allows the model to fit the training data by minimizing the function itself. The loss function and regularization term compose the objective function together.

As a function for smoothing weight that controls the model complexity, the regularization term prevents the model from overfitting. In light of the regression issue, the error rate, such as root mean square error (RMSE), instead of binary classification error rate, is used to be a loss function in this paper.

Furthermore, the prediction in $t$-th iteration(tree) can be described by the following equation:

$$\hat{y_i}^{(t)} = \sum_{k=1}^{t} f_k(x_i) = \hat{y_i}^{(t-1)} + f_t(x_i)$$

The final objective function at the $t$-th iteration of trees can be denoted as follows:

$$obj^{(t)} = \sum_{j=1}^{T}[G_j w_j + \frac{1}{2}(H_j + \lambda)w_j^2] + \gamma T$$

where $G_j = \sum_{i \in I_j} g_i$, $H_j = \sum_{i \in I_j} h_i$, and $\gamma$ is a penalty defined by user to encourage pruning, and $T$ denotes the number of tree leaves.

The way to prune the tree is expressed by the equation below:

$$Gain = \frac{1}{2}\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} - \gamma$$

where $L$ denotes the left node, $R$ the right node. As long as the $Gain$ is negative, the pruning takes place.

As a tree-based model, XGBoost develops a unique tree splitting rule. The greedy algorithm allows the model to detect the best split for constructing an XGBoost tree. For every single tree, the node with the most significant scores is selected to be split. Yet, this traditional tree splitting method, the exact greedy algorithm, is inefficient due to numerous thresholds caused by multiple features and examples in the dataset. Therefore, an approximate algorithm is introduced and shows its efficiency by dividing data into quantiles splitting observations. Subsequently, a weighted quantile sketch is presented to propose candidate split points. Unlike the usual percentile of a feature that evenly distributes candidates, the weighted quantile method maintains that each quantile has an equal sum of weight.

Learning rate and subsampling ratio are two techniques to avoid overfitting in XG-Boosting. The learning rate, between 0 to 1, prevents the model from overfitting by controlling bias reduction. Yet, in each iteration, it takes a small step of prediction towards the actual value. Consequently, the final prediction equals the sum of the learning rate multiplied by prediction in each tree across all iterations. XGBoost can choose a portion of both features and samples to construct a tree, like the random forest method. Furthermore, column subsampling and training instance subsampling is determined by a subsampling ratio.

## 3.4 Benchmark technique

The zero forecast method is a simple method, by which each forecasted value is given zero over the forecast horizon. (function) Chatfield, Hayya (2007) argue that comparing to Croston, SES (simple exponential smoothing), and MA (moving average methods), the zero forecast method has the best performance on both accuracy error and inventory cost for the demands with high lumpiness which denotes the degree of probability of non-zero demand. However, there is only one advanced forecasting method, the Croston method, that compares with zero forecasts, indicating insufficient proof of its superiority over other methods in this paper. According to Teunter, Duncan (2009), the zero forecast method is perceived as the poorest method because there is no prediction from inventory control. Despite the zero forecast method having low accuracy error, it is not doubted that the zero forecast method performs worst in inventory control measures (ibid). Our work considers the zero forecast method as a baseline method assumed to have the worst performance against other methods since it does not predict anything.

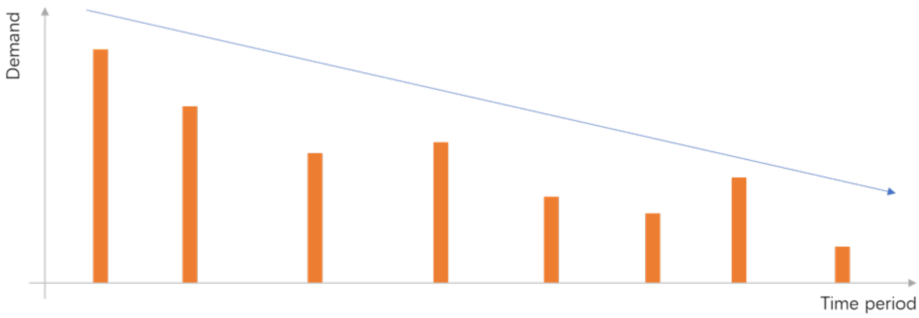
# 4 Data Description

This chapter discusses the dataset to which forecasting methods are applied in this thesis work. To make our research more robust, we include two types: simulated and empirical. We will investigate whether the results from the simulated dataset are in line with those in the empirical one. Besides, generating data for simulated datasets, especially the simulation in decreasing trends, is provided. Moreover, we examine the data descriptions and their traits for four empirical datasets from different industries. Lastly, the classification scheme is presented, which is necessary since different forecasting methods might have good performance for specific item categorization.

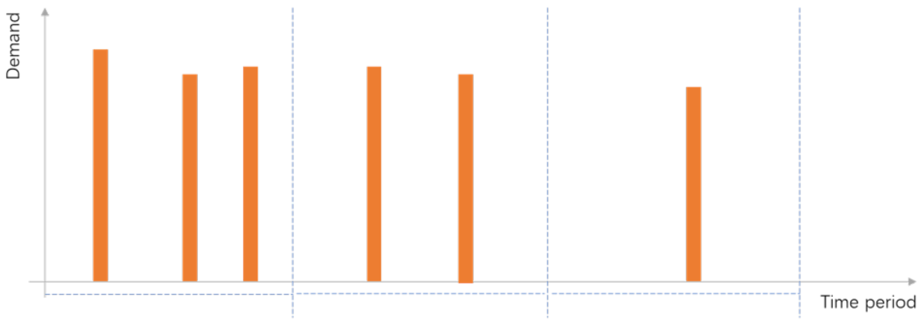
## 4.1 Simulated data and settings

For the investigation of forecasting, a simulated dataset of 200 months, along with different patterns, is generated. As a part of the R package "tsintermitten" proposed by Kourentzes (2014), the simulated function contains four inputs, i.e., the number of SKUs, the observation in time series for each SKU, the average inter-demand interval, and the squared coefficient of variation of the non-zero demands. We use the latter two arguments of the inputs for the function to generate time-series data in two parts: (1) the non-zero demand interval following a binomial distribution (2) the non-zero demand size following a negative binomial distribution.

In addition, those two arguments also construct the classification scheme that creates four patterns: smooth, intermittent, erratic, and lumpy. From this point, we simulate the patterns using the R package "tsintermittent". In terms of combining the latter two arguments, we create intermittent data with all those four patterns.




(a) The decreasing trend for demand size




(b) The decreasing trend for demand probability

Figure 2: Two cases of decreasing trend

Besides these four patterns, a decreasing trend in the demand process is also considered. The figure 2 shows two cases: (1) a decreasing trend for demand size but in demand probability. (2) a decreasing trend for demand probability but demand size.

In the first case, in other words, the demand size decreases linearly without the probability decreases. We firstly simulate data $D_O$ as that there is no trend and extract the non-zero demand $D'_O$ from $D_O$. Furthermore, we manufacture another decreasing dataset that equals the product of a random series of numbers following a normal distribution and a descending arithmetic sequence of which the common difference is 1. In addition, the result is rounded. The formula is provided as follows:

$$D_{new} = S \cdot N$$

where $D_{new}$ denotes the new dataset, $S$ the vector contains a descending arithmetic sequence where the largest element equals to the total observations of non-zero demand $D'_O$, with a common difference that is 1, $N$ the vector follows a random normal distribution with mean that equals 3 and standard deviation that is 1. The mean, which is 3, prevents the negative numbers occur. Furthermore, the descending sequence $S$ usually guarantees a linearly decreasing trend after the multiplication because the trend in $S$ is more extreme in the $N$ following a normal distribution. Figure 4 offers an example by breaking down the process and giving detail.

| S | N | D |
|---|---|---|
| 10 | 3.7 | 37 |
| 9 | 2.7 | 25 |
| 8 | 1.4 | 11 |
| 7 | 1.5 | 10 |
| 6 | 1.4 | 8 |
| 5 | 2.5 | 12 |
| 4 | 1.5 | 6 |
| 3 | 3.7 | 11 |
| 2 | 5.1 | 10 |
| 1 | 1.7 | 2 |

Figure 3: The new data $D_{new}$ is constructed by multiplying vector $S$ with vector $N$.

We then reorder the non-zero demand in the original data simulated initially with the sequence followed by the new linearly decreasing series just created.

Lastly, building a linear regression model using $S$ as the independent variable and using the non-zero demand size $D'_O$ as a dependent variable to test whether the linear trend is significant. Thus, if the p-value is lower than 0.05, a linearly decreasing series is established.

The steps of the simulation in the first case are presented as follows:

1. Create a dataset $D_O$ with no trend and build another dataset $D'_O$ composed by non-zero demand in $D_O$.

2. Construct a new dataset using formula: $D_{new} = S \cdot N$.

3. Apply the order of $D_{new}$ to $D'_O$, and replace the original sequence of non-zero demand in the dataset $D_O$.

4. Build a linear regression model for examination, keep the observations with a

negative coefficient, and remove the ones with an insignificant linear model.

In the second case, contrary to the first case, the demand sizes remain the same while the probabilities decrease through all periods. We divide the data into five buckets (usually the number is five[1]), and for each bucket, the probabilities of demand sizes drop by a certain level, which is the initial probability divided by 5. For example, the initial probability is 0.9, and the probability in each following bucket is 0.72, 0.54, 0.36, 0.18, respectively, with a fixed reduction. In other words, the frequency of non-zero demand in each block is decreasing. However, in this case, the probability doesn't drop to zero because we only simulate a decreasing trend in demand probability instead of complete obsolescence. A similar technique that dividing the whole demand history into several blocks is also used in the paper proposed by Babai et al. (2014). Instead of 5 blocks, 3 blocks containing 8 periods in each block and 4 blocks containing 21periods in each block are used. Choosing five non-overlapping blocks is arbitrary yet it guarantees 40 periods in each block for simulated data with 200 periods in total.

In addition, like the first case, the second one is also applied by a linear model to check the significance of the linear model. Constructing a linear regression model of which the dependent variable is the frequency of non-zero demand within each block and the independent variable is the number of non-overlapping blocks.

The simulation in the second case is provided in the following steps:

1. Break the entire periods into $n$ non-overlapping periods. If the periods with each block are lower than 10 periods, change $n$ to make the period number equal to 10.

2. Compute the frequency of non-zero demand in each block.

3. Build a linear regression model for examination, keep the observations with a negative coefficient, and remove the ones with an insignificant linear model.

According to the result, note that the demand pattern is quite lumpy, resulting from the high average zero demand interval led by the low probability of zero demand. The

[1]the number of blocks depends on the number of periods, generally with ten periods minimum.

spare parts simulated are either lumpy or intermittent.

The method to identify this decline in empirical data for either non-zero demand size or non-zero demand probabilities is explained in the next section. What is more, for the simplicity of the analysis, the lead time remains one period for all SKUs in all cases. We may further advance this research by incorporating different lead times as an argument to validate forecasts in our future work.

Unlike the lead time and price of SKUs are both given in the empirical dataset, the lead time and the price in the simulated dataset are assumed to be 3 and 1, respectively, for all SKU for the reason of simplicity. These are designed in such a way because the lead time and price are difficult to simulate, hardly to find an appropriate distribution.

After the process mentioned above, we finally simulated a dataset including 270 observations and 120 periods.

## 4.2 Empirical data and settings

In this paper, the empirical data includes four datasets. The number of datasets and abbreviations are given in table 1. The first dataset contains the information of spare parts from a large refinery oil company with 56 months. The second one comes from the automotive dataset used by Syntetos, Boylan (2005) with 24 months periods and 3000 SKUs. The third one is a dataset from a manufacturing firm in the Netherlands with 46 weeks and 3451 SKUs. The fourth one is the RAF (Royal Air Force), first described by Teunter, Duncan (2009), and it contains seven years periods (84 months) with 5000 SKUs. Except for the third dataset from the manufacturing firm, the lead time and price for each item are available. Both lead time and price are assumed to be one in the only exception.

Table 2 to Table 4 describe the statistics that include the average demand sizes, inter-demand intervals, and demand per period. Subsequently, the table provides the information regarding the pattern for each SKU by evaluating the minimum, median,

Table 1: The abbreviation of datasets

| No. | Dataset | Abbreviation | References |
|---|---|---|---|
| 1 | Refinery oil company | RO | Porras, Dekker (2008) |
| 2 | Manufacturing firm | MF | Syntetos, Boylan (2006) |
| 3 | Royal Air Force | RAF | Teunter, Duncan (2009), |

and maximum number of three metrics. All these three datasets are quite lumpy or slow, with a minimum number of demand intervals equal to 2 and 3.82, respectively, meaning it takes a long time to fulfill a demand. Finally, Table 5 of the classification demonstrates the characteristics of each dataset.

Moreover, the price and lead time are offered in four datasets except for Dataset MF from the manufacturing industry. We then assume that the lead time equal to 1, and the price for each SKU remains the same. Also, just as we do to simulate a decreasing trend in simulation data, two decreasing patterns in the demand process plan to be identified in the empirical data.

Empirical datasets may contain the obsolescence items in which the decreasing trend is identified by a linear model. For both decreasing cases mentioned in the last section, the independent variable of the linear model is the time period while the dependent variables are different due to the characteristics of decreasing patterns. In the first case where only non-zero demand decrease, it is obvious that the non-zero demand is the dependent variable and the periods number is the independent variable. In the second case where only probabilities of non-zero demand decline, we assign the frequency of non-zero demand as the dependent variable and the blocks number from 1 to n[2] as the independent variable. In addition, the items with decreasing patterns are selected on the condition that the coefficient of the independent variable, namely the slope, is negative and that the p-value of the model is smaller than 0.05, demonstrating a significant linear relationship.

[2]The number of blocks is five except for the dataset MF due to its entire periods only 46 periods. Therefore, the n equals 4 in this case

[3]All means and standard deviations in the table are computed for each SKU in each dataset unless otherwise stated.

Table 2: The summary of Dataset 1 –RO

| | Demand size | | Demand interval | | Demand per period | |
|---|---|---|---|---|---|---|
| percentile | Mean[3] | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. |
| 0%(min) | 1 | 0 | 2 | 0 | 0.04 | 0 |
| 25% | 1 | 0 | 7.14 | 4.37 | 0.15 | 0.1 |
| 50% | 1.56 | 0.58 | 10.67 | 7.23 | 0.37 | 0.35 |
| 75% | 4 | 2.81 | 16.33 | 12.02 | 0.94 | 1.01 |
| 100%(max) | 1600 | 1612.31 | 28 | 38.18 | 465.38 | 667.42 |

Table 3: The summary of Dataset 2 – MF

| | Demand size | | Demand interval | | Demand per period | |
|---|---|---|---|---|---|---|
| percentile | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. |
| 0%(min) | 0.08 | 0 | 1.11 | 0 | 0 | 0 |
| 25% | 3.07 | 1.89 | 4.93 | 3.81 | 0.54 | 0.59 |
| 50% | 8.5 | 6.56 | 10.92 | 8.51 | 2.05 | 2.43 |
| 75% | 22.05 | 17.99 | 23.46 | 20.81 | 6.08 | 7.81 |
| 100%(max) | 10780.44 | 7358.02 | 74 | 101.82 | 1523.85 | 1516.14 |

Table 4: The summary of Dataset 3 – RAF

| | Demand size | | Demand interval | | Demand per period | |
|---|---|---|---|---|---|---|
| percentile | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. |
| 0%(min) | 1 | 0 | 3.82 | 0 | 0.04 | 0 |
| 25% | 1.56 | 0.81 | 7.27 | 5.43 | 0.38 | 0.37 |
| 50% | 3.83 | 3.06 | 9 | 6.93 | 0.84 | 0.97 |
| 75% | 11.33 | 9.31 | 11.57 | 8.63 | 2.67 | 3.44 |
| 100%(max) | 668 | 874.42 | 24 | 16.46 | 232.06 | 363.6 |

## 4.3 Classification scheme

In this paper, the classification of SKUs is taken into account by aggregating items according to their characteristics. Applying the classification scheme is to detect the best forecasting methods for a specific item categorization. Johnston, Boylan (1996) suggest

Table 5: The number of SKUs for different patterns in multiple datasets.

| | SKUs number | | | |
|---|---|---|---|---|
| Pattern | RO | MF | RAF | Simulated |
| Intermittent | 7967 | 640 | 2565 | 29 |
| Lumpy | 2770 | 974 | 2434 | 133 |
| Smooth | 6 | 2 | 1 | 77 |
| Erratic | 0 | 26 | 0 | 31 |

that when the mean inter-demand interval is higher than 1.25, the size/interval method, known as Croston's intermittent demand estimation procedure, outperforms the weighted moving average approach (EWMA) comparing these two methods. Then Syntetos et al. (2005) propose an alternative classification scheme based on the research of Johnston, Boylan (1996) by comparing the Croston method, SBA, and EWMA, in which mean squared errors (MSE) are taken as the evaluation measures. This method is also known as the SBC classification scheme.

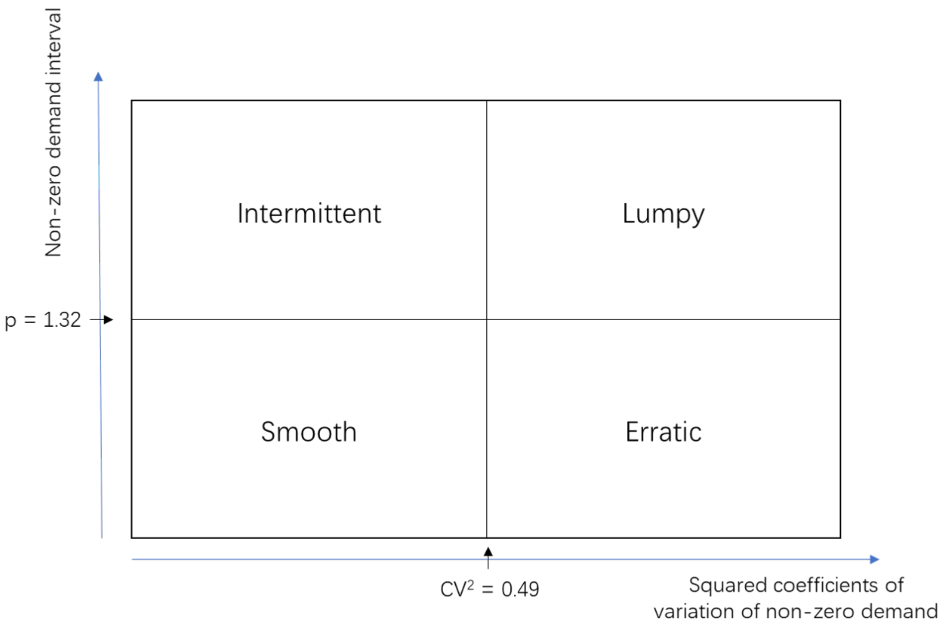



Figure 4: SBC categorization scheme

As the metrics classifying all spare parts, the squared coefficient of variation of non-zero demands and the mean demand interval are computed for all SKUs. They take the average demand interval p that equals 1.32 and the squared coefficient of variation $CV^2$, which is 0.49 as the cut-off point for average demand interval and squared coefficient of variation. This classification scheme is eventually built by those two cut-off points that construct a matrix with four demand patterns: smooth ($p < 1.32$ and $CV^2 < 0.49$), intermittent ($p >= 1.32$ and $CV^2 < 0.49$), erratic ($p < 1.32$ and $CV^2 >= 0.49$), and lumpy ($p >= 1.32$ and $CV^2 >= 0.49$). Figure 5 explains how this categorization works: smooth item has comparatively low non-zero demand interval and low squared coefficient of variation for non-zero demand; intermittent demand has a low squared coefficient of variation of non-zero demand and high demand interval in average; the items of erratic and lumpy both have high variability in average demand interval yet the former one has low demand timing while that of the latter one is relatively high. With the widespread application in the industry, we decide to apply this classification scheme in this paper to specify the type of spare parts both in simulated datasets and empirical datasets.

# 5 Evaluation metrics

To evaluate those five methods mentioned in the previous chapter, we apply forecasts accuracy and inventory performance, both of which are the leading performance measures in forecasting intermittent spare parts.

## 5.1 Accuracy measures

Forecast accuracy is considered as an "error measure", which means the gap between forecasts value and actual value. Although there are many forecast accuracy measures, the measure perfectly designed for the intermittent dataset is few. Prestwich et al. (2014) claim that existing error measures do not help predict intermittent demand. Therefore, more than one accuracy measure is applied in a paper for most of the cases.

Pinçe Çerağ et al. (2021) suggests that, among all papers related to intermittent spare parts forecasting, the most popular six forecast accuracy are ME, MAE, MSE, MAPE, MASE, and RMSE. Despite their prevalence, Prestwich et al. (2014) argue that scale-dependent measures, such as ME, MAE, MSE, and RMSE, are not accepted as valuable methods to compare several series due to their scale dependency. It is reasonable to use these measures only if the forecasting methods applied to single time series or multiple time series that share the same units.

Apart from those four methods above and for simplicity and comparison with the results in other papers, MAPE ranked 4th amongst those accuracy measures may be a candidate for the accuracy measure. Unlike scale-dependent measures, it allows us to compare the series from different SKUs due to its benefit of unit-free. Thus, it is used to compare forecasting methods between different datasets. However, the drawback of MAPE is still apparent. The method categorized as percentage errors, such as MAPE, is undefined once the actuals, the denominator in the equation, equals zero. Also, Syntetos, Boylan (2005) claim that removing the period with zero demand or summing a small number to avoid zero demand has no significant effect in enhancing this MAPE's confidence measure. Therefore, it is not advisable that many software exclude the periods with actual that is zero. Moreover, MAPE is asymmetric, namely more penalty on the negative errors (forecasts higher the actual) than that on positive ones (forecasts lower than the actual) since the ratio cannot over 100 percent if the estimates are much higher than the actuals. Therefore, a modified version of MAPE, sMAPE (symmetric MAPE), is considered because the issue brought by the zero actuals value in MAPE is eliminated and expands its upper range from 100 to 200.

Since every measure has its disadvantage, the second measure is introduced to validate each other. MASE, Mean Absolute Scaled Error, is proposed and recommended by Hyndman, Koehler (2006) due to its ability to be independent of the scale of time series and to deal with the occurrence of infinite and undefined values. According to the argument of Hyndman, Koehler (2006), MASE is considered to outperform several methods in forecasting competitions. What is more, according to Prestwich et al. (2014), MASE exceeds other methods specifically in forecasting competitions and more valid for non-stationary data, such as demand with trends or seasonality.

In the end, two accuracy measures are calculated, including:

- Symmetric Mean absolute percentage error: $sMAPE = mean\left(\frac{200|e_t|}{Y_t+\hat{Y}_t}\right)$

- Mean Absolute Scaled Error: $MASE = mean\left(\frac{|e_s|}{\frac{1}{t-1}\sum_{i=2}^{t}|Y_i-Y_{i-1}|}\right)$

where $t$ denotes the period, and $e_t$ equals $Y_t-\hat{Y}_t$. According to the equations, it is obvious that for $sMAPE$ the less the error (close to zero) is, the better in forecasting and that for $MASE$ the greater the scaled error is than one, the worse the forecast performance is.

Considered multiple items, we calculate the mean of accuracy rate across all SKUs.

## 5.2 Inventory performance

On the other side, inventory performance is also vital to measurement because Syntetos, Boylan (2006) assert that a high accuracy method doesn't guarantee a high performance of inventory investment. We adopt an order-up-to-level policy that replenishment is triggered to make the stock return to the reorder point once stock is below it. The order-up-to-level stock can be computed by the demand size, lead time, and cycle service level CSL (i.e., the probability of no out-of-stock). The main focus of this approach is to find out the minimum stock on a condition of a certain service level. The CSL is described as follows:

$$CSL = P(LTD <= S)$$

where $P$ denotes the probability, LTD the lead time demand, and S the reorder point. Order up to a level stock that achieves a certain service level is then calculated as follows:

$$S = \mu_{LTD} + \sigma_{LTD} \times \Phi^{-1}(CSL)$$

where $\mu_{LTD}$ denotes the average of the lead time demand, $\sigma_{LTD}$ the deviation of lead time demands, $\Phi^{-1}$ the inverse of cumulative distribution function over lead time, and $CSL$ is the target service level that we expect to achieve. It is noteworthy that in this paper $\Phi$ denote a cumulative function the standard normal distribution for parametric methods, while a cumulative function of empirical distribution for non-parametric techniques. Moreover, in this equation, we can know that order up to level $S$ is also known as reorder point and $\sigma_{LTD} \times \Phi^{-1}(CSL)$ means the safety stock, and service level is given as the equation above: $CSL = p(LTD <= S)$. Furthermore, this paper assumes that the demand is uncertain and the lead time itself is stable. The mean of lead time demand can be described as $\mu_{LTD} = \mu_{LT} \times \mu_D$ and the variance of demand $\sigma_{LTD}$ is described as $\sigma_{LTD} = \sqrt{\mu_{LT}} \times \sigma_D$ , where $\mu_{LT}$ denotes the average lead time, $\mu_D$ the mean of demand size per period, and $\sigma_D$ one standard deviation of demand size per period. Note that, in some papers, it mentions RMSE (Root mean squared error) which is identical with standard deviation mathematically.

Subsequently, we consider the trade-off curve between achieved cycle service level and Inventory holding cost; these trade-off curve measures are applied in several papers (Syntetos, Boylan (2006); Syntetos et al. (2015); Babai et al. (2014)). To construct this trade-off curve, we first set a series of target CSL. In this paper, we use forecasting methods to meet the seven CSL targets: 0.8, 0.85, 0.9, 0.92, 0.95, 0.97, 0.99. Then, for each SKU in the dataset, the base stock can be computed using the total forecast demand over the lead time to meet CSL targets. Moreover, the achieved CSL can be obtained in the same way by using the actual demand instead of forecast demand. Finally, we calculate the inventory investment by multiplying the base stock and price for each SKU.

The trade-off curve is constructed as following steps:

1. Set a series of CSL target.

2. For normal distribution, the equation $S = \mu_{LTD} + \sigma_{LTD} \times \Phi^{-1}(CSL)$ calculate inventory cost using forecast demand. For empirical distribution, we change the equation and let $S = \Phi^{-1}(CSL)$

3. For each cycle, once the safety stock is less than thee actual demand, then this cycle is define as a shortage cycle. Calculate the number of shortage cycle and determine the achieved CSL.

4. For each target CSL and inventory cost, the achieved CSL can be found, constituting several dots that finally form a curve by connecting them in lines, and the trade-off curve is then constructed.

In this paper, two types of distribution of lead time demands are considered: the normal distribution and the empirical distribution. For parametric methods (SBA, TSB) and machine learning method (XGB), the normal distribution are applied, while for non-parametric methods (EMP and WSS), the lead time demand are are assumed to follow empirical distribution. Teunter, Duncan (2009) mention that the mean of normal distribution equals the product of per period forecast and lead time, converting period forecasts into lead time forecasts.

Finally, we build a the tradeoff curve with its x-axis the inventory investment and y-axis the achieved CSL. By summing across all SKUs in the dataset, we multiply the unit price by the amount calculate and compute the inventory investment. Note that the achieved CSL is averaged across all SKUs. Also, a curve for comparison between achieved CSL and target CSL is constructed to show to what extend target CSL can reach to the achieved one.

# 6 Model Setups

In this research, we construct a model for each SKUs across multiple forecasting methods. For this purpose, this chapter mainly introduce the detail setups in sampling and forecasting, the settings of different measures, and the parameters tuning in parametric methods and the XGBoost method, except for the non-parametric methods since they don't contain any parameters. Moreover, the application of the XGBoost method in time series data is provided using a feature engineering technique to generate multivariate data based on univariate data.

## 6.1 Sampling and forecasting

To evaluate the performance of forecasting methods, we divide both the simulated datasets and empirical ones into training data and test data. The training data provides input for the model and optimizes the parameters of parametric methods, while the purpose of test data is to report the performance by comparing it with the actual value. Within the dataset, 70% of the demand history is selected to be the training data and 30 % for test data.

Besides, a rolling method is applied to the training and test process. For example, a forecasts $Y_t'$ at period $t$ is based on the periods from 1 to $t-1$ from training data. We keep updating the training data after obtaining each forecast.

Note that since the forecast is designed to be over the lead time, the final forecast equals the prediction result of a one-period unit that multiplies lead time, making it to be lead time forecasts.

## 6.2 Settings for the variants of the Croston method

For parametric methods mentioned in Chapter 3, we take the first non-zero demand and the first interval as the initial demand and initial interval. Furthermore, the model parameters are specified to be 2, namely that the demand and interval parameters are different in fitting the model.

In addition, we optimize smoothing parameters over training data using the Nelder-Mead method (Nelder, Mead (1965) ) from the R package “tsintermittent", which minimizes the cost function by varying its parameters. The reason why we don’t apply the popular grid search is that local search such as the Nelder-Mead method is much more efficient in finding optimized parameters. It is more effective in saving time since each empirical dataset contains thousands of observations in this research along with thousands of models needed to be built.

Besides, we used the accuracy measures (sMAPE and MASE) mentioned in Chapter 5 to obtain the optimized smoothing constants seperately. The parameters with the smallest error provided by sMAPE and MASE are chosen for each model based on each SKU.

## 6.3 Settings for the non-parametric methods

The procedure of fitting the non-parametric models (EMP and WSS) are basically the same as that mentioned in Chapter 2. According to same replications mentioned by Willemain et al. (2004), we jitter the non-zero demand 1000 times to get a robust result when fitting the WSS model. Through the bootstrapping process, a distribution of lead time demand is obtained under these two methods, and the forecast over the lead time equals the mean from this distribution.

## 6.4 The application of XGBoost in demand forecasting

Inder (1984) use an approach to apply the lag of the response data to time series forecasting. In other words, the lag features as dependent variable are values from previous periods. To elaborate that, Figure 5 demonstrates the input dataset for the XGBoost method and explains the way that the function used to establish the model by creating a matrix of features based on lagged terms of each demand period, and we further feed that matrix as the variables for the XGBoost model. As an example, the number of lagged periods in Figure 5 is four. However, considered that variables are not enough for training a model, the rule for lagged period number is made. In this research, the rule that determines the number of the lagged period comes from the R package “forecastxgb" introduced by Ellis (2017), and it is described as follows:

$$N_{lagged} = \max(8, 2 \cdot Freq)$$

where $N_{lagged}$ denotes the number of max lagged periods, and $Freq$ means the frequency of a time series dataset. $Freq$ is higher than one if the dataset is cyclical or it equals one, For example, if the basic period of a dataset with periodicity is month, then the frequency might be 12 because of 12 months each year. The equation above indicates that $N_{lagged}$ is simply but the max value between eight and two times the dataset frequency, meaning at least there are 9 variables (one original period as a dependent variable and eight lagged period as independent variables) as an input in the XGBoost method. The number of eight is a given volumn according to Ellis (2017). Assuming that there is neither seasonality nor any other periodicities in the dataset, we then assign the lagged periods with 8, which means that we take eight lagged times as variables for our model. Thus, a multivariate data is generated by creating lagged periods that originates from univariate time series data.

The cross-validation technique is also applied in constructing a preferable model. By doing so, a good iteration number, namely the number of the tree, is computed. When tuning the hyperparameters of the XGBoost model, we obtain the optimized parameter using 10-fold cross-validation and max iterations that is up to 100. In order to keep this model simple and efficient when applying it to thousands of spare parts with hundreds of time periods, we only select one parameter, which is the number of iterations, for model training in such a way that sMAPE and MASE are selected to be the accuracy measure.

| Year | Demand |
|---|---|
| 1990 | 1 |
| 1991 | 2 |
| 1992 | 3 |
| 1993 | 5 |
| 1994 | 6 |
| 1995 | 7 |
| 1996 | 1 |
| 1997 | 2 |
| 1998 | 3 |
| 1999 | 5 |
| 2000 | 2 |

| Year | Demand | lag1 | lag2 | lag3 | lag4 |
|---|---|---|---|---|---|
| 1994 | 6 | 5 | 3 | 2 | 1 |
| 1995 | 7 | 6 | 5 | 3 | 2 |
| 1996 | 1 | 7 | 6 | 5 | 3 |
| 1997 | 2 | 1 | 7 | 6 | 5 |
| 1998 | 3 | 2 | 1 | 7 | 6 |
| 1999 | 5 | 3 | 2 | 1 | 7 |
| 2000 | 2 | 5 | 3 | 2 | 1 |

Figure 5: The example of input data for XGBoost in intermittent demand forecasting. The left table is the original datasets while the right one is processed as a result of adding variables of which value from previous periods. Note that this case is merely an example with four lagged periods without a rule applied for selecting a minimum lagged number because the number in this research is at least eight.

## 6.5 An example of the experiment

To better illustrate the experiment in this research, we apply item No.16 from simulated datasets to the procedure mentioned above. Figure 6 demonstrates the monthly demand for item No. 16.

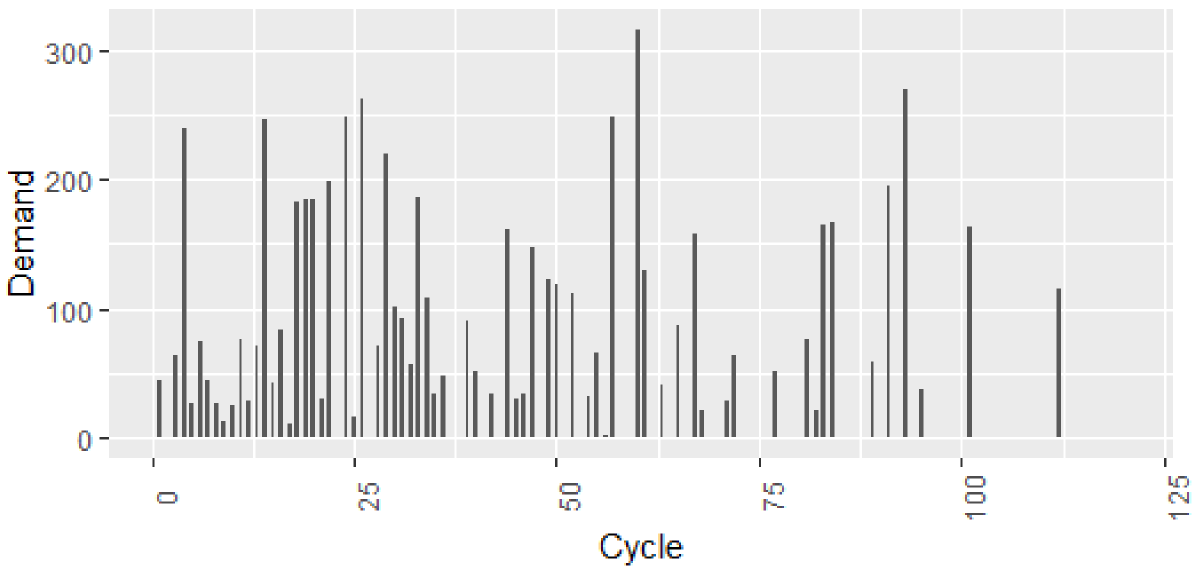



Figure 6: Monthly demand of item No. 16 from simulated dataset with a probability decreasing trend.

First, we determine the trend type and pattern type based on the characteristics of this SKU. For this item, the coefficient of the linear relationships between the non-zero demand and frequency of the demand occurrence is -4.9, and the p-value of this regression model equals 0.0019. Thus, a probability decreasing trend is undoubted with this SKU. The pattern type is determined by the average demand $p$ and $CV^2$, which is 1.70 and 0.60 respectively, and thus classified as pattern Lumpy.

Moreover, we apply each forecasting method to build the model and predict the demand. For multiple items, we average the accuracy rate across all SKUs.

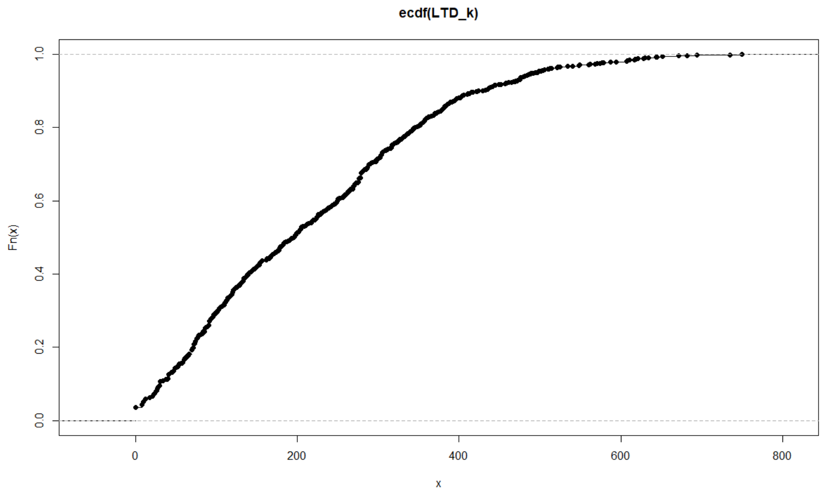



Figure 7: The empirical cumulative distribution of lead time demand based on the WSS methods for item No. 16.

Lastly, the evaluation is executed under the measures of accuracy rate and inventory performance. Applying the accuracy rate techniques, we compute the sMAPE and MASE under different patterns and trends. As for the inventory performance, we determine the lead time demand distribution, compute the safety stock, and then calculate the achieved CSL. Figure 7 demonstrates the empirical cumulative distribution of WSS. For item No. 16, target CSL$=$ 0.92, safety stock $=$ 458, achieved CSL $=$ 0.917. Target CSL $=$ 0.95, safety stock $=$ 495, achieved CSL $=$ 0.944. By varying the target CSL, we can compute a series of achieved CSL. For parametric methods, we apply the normal distribution. Regarding multiple items, we average the achieved CSL by forecasting methods.

# 7 Results

This section introduces the results from the simulated dataset and the dataset in the real world discussed in Chapter 4, where we introduce three empirical datasets, RO (Refinery oil company), MF (Manufacturing firm), and RAF (Royal Air Force). For each part, there are two types of measurement mentioned in Chapter 5: the forecasting accuracy and the inventory performance under different demand trends. Both measurements include further information on different patterns. Therefore, this section presents the forecasting accuracy and inventory investment separately.

Furthermore, the results from different patterns discussed in Chapter 3 are provided in the section on forecasting accuracy. Table 5 in Chapter 4 indicates that most of the items are intermittent and lumpy in the three datasets. For the section on inventory investment, the results based on different types are not included because of the limitations of this research space.

Due to the same reason, the part of accuracy rate cannot adequately address the results from all the datasets. Yet, the tradeoff curve is discussed for each empirical dataset.

## 7.1 Forecasting accuracy

According to the previous chapter, two accuracy measures, MASE and sMAPE, are considered to evaluate forecasting accuracy across different methods. These two measures are both smaller, the better.

Table 6 demonstrates the comparison of forecasting accuracy between multiple models under MASE and sMAPE in the simulated dataset. For each row in Table 6, the bold number indicates the methods with the best performance.

[4] N denotes the normal demand without trends, PD the demands with decreasing probabilities, and SD the demands with decreasing size.

Table 6: A comparison of the accuracy of forecasting methods in simulated dataset is provided. N is the abbreviation of the normal demand without trends, PD the abbreviation of demands with decreasing probabilities, and SD the abbreviation of demands with decreasing size. The method of best performance is highlighted for each row.

| | | Parametric | | Non-parametric | | ML | Benchmark |
|---|---|---|---|---|---|---|---|
| Measure | Trend[4] | SBA | TSB | EMP | WSS | XGB | ZF |
| MASE | N | 0.87 | 0.87 | 0.87 | 0.88 | **0.79** | 0.99 |
| | PD | 1.90 | 1.88 | 1.89 | 1.98 | 0.96 | **0.58** |
| | SD | 1.78 | 1.46 | 1.54 | 1.90 | **0.92** | 1.20 |
| sMAPE | N | 1.24 | 1.24 | 1.24 | **1.23** | 1.33 | 2.00 |
| | PD | **1.77** | 1.78 | **1.77** | **1.77** | 1.85 | 2.00 |
| | SD | **1.21** | 1.23 | 1.26 | 1.28 | 1.26 | 2.00 |

Note that under MASE measurement, XGB and ZF outperform other methods for three types of trends. First, the XGB performs best under the pattern N and SD. However, when applying sMAPE, this is not the case. The error rates of traditional methods are lower than those of XGB and zero forecasts. Significantly, the ZF has the worst performance of which value equals 2 using sMAPE as the measure. Thus, a contradiction appears.

We believe that ZF has the worst performance when applied to sMAPE may because the error $|e_t|$ is always equal to $Y_t$ yet $\hat{Y}_t$ is zero all the time according to the formula $sMAPE = mean\left(\frac{200|e_t|}{Y_t+\hat{Y}_t}\right)$. Compared to sMAPE, the MASE owns a denominator from in-sample, bring no such issue under sMAPE. To solve this contradiction between the results of two measurements, we introduce a method to evaluate the models. We start with the max-min normalization adopted to calculate the scaled score of each model for each row in Table 7. The smaller number after scaled is preferable since it's a transformation as the accuracy rate. Moreover, we average the value with the same trend pattern. Finally, Table 7 summarizes the results, which can be thought that the accuracy rate of the target method is how far away from that of the best method.

[5]The time is measured by second.

Table 7: The comparison of the accuracy of forecasting methods in the simulated and empirical dataset is presented by averaging the max-min normalization of MASE and sMAPE. The larger the percentage, the worse the performance. The method of best performance is highlighted.

| | | Parametric | | Non-parametric | | ML | Benchmark |
|---|---|---|---|---|---|---|---|
| Trend | Data Type | SBA | TSB | EMP | WSS | XGB | ZF |
| N | Simulated | 20% | 20% | 20% | 23% | **6%** | 100% |
| | Empirical | **49%** | **49%** | 50% | 54% | 52% | 50% |
| PD | Simulated | 48% | 48% | 47% | 50% | **31%** | 50% |
| | Empirical | 48% | 48% | 48% | 51% | **29%** | 50% |
| SD | Simulated | 44% | 29% | 35% | 54% | **3%** | 64% |
| | Empirical | 47% | 47% | 45% | 52% | **40%** | 50% |
| Time[5] | Total | 1.4s | 1.4s | 0.1s | 5.3s | 9.7s | 0.04s |

After taking the average of the accuracy results using max-min normalization, from Table 7, we conclude Finding S1 as follows:

**Finding 1.** XGB outperforms the traditional methods under three types of trends, significantly PD and SD.

Regarding Finding 1, it is evident that XGB dominates under three trends in simulated data. Compared to the condition in simulated data, the performance of XGB in empirical data is the best except for Trend N.

Note that TSB exceeds the others in trend SD among traditional methods, showing the sensitivity to demand obsolescence due to the update of the occurrence of nonzero demand. But, surprisingly, WSS has relatively higher numbers (low performance) than others, even than the non-parametric method EMP in trend SD.

In addition, Table 7 also includes the time that each forecasting methods spend in building the models. The numbers denote the average time for each SKU. As expected, it takes the longest for XGB to compute since the machine learning technique requires training and validation to determine the optimized parameters. There is no doubt that the second time-consuming method is WSS because of the 1000 times of bootstrapping generating the lead time demand distribution. Thus, we can conclude another finding:

**Finding 2.** XGB and WSS are the most and second most time-consuming methods, respectively, among all techniques.

Table 8 shows that under which pattern the forecasting methods outperform others. Similarly, as the case containing two opposite results between the conventional methods and XGB (and ZF) mentioned previously, we calculate the mean of the max-min normalization again for each method [7]. Finally, more details of results regarding the patterns are summarized by Table 8 as follows:

**Finding 3.** For Intermittent and Lumpy under the trend of simulated data, XGB has the best performance. Yet, XGB performs worse than others in empirical data.

**Finding 4.** The preferable method for trend PD is XGB, no matter under what pattern it is. This results in simulated data and empirical data are consistent.

**Finding 5.** XGB beats other methods in the pattern Erratic and Lumpy under trend SD for both simulated and empirical data.

Regarding Finding 3, for the pattern under the trend N, the reason the conventional methods win over XGB could be that the forecast variation of conventional methods is smaller than that of XGB. This fact results in a small error rate in the pattern Smooth and Erratic, under which the non-zero demand interval is lower than that under Intermittent and Lumpy.

Moreover, Finding 4 and Finding 5 are following Finding 1. According to Finding 4 and Finding 5, it is proved that XGB performs well under trend PD and SD, even for the performance for the SKUs divided into multiple patterns.

[6]NA means no type under this demands trend.

[7]For more details of the results under different patterns in MASE and sMAPE, Table 9 in the appendix describes the plots.

Table 8: A comparison of the accuracy of forecasting methods under different pattern in simulated and empirical datasets using the max-min normalization is provided. The method with best performance is highlighted.

| | | | Parametric | | Non-parametric | | ML | Benchmark |
|---|---|---|---|---|---|---|---|---|
| Trend | Data Type | Pattern | SBA | TSB | EMP | WSS | XGB | ZF |
| N | Simulated | Smooth | **0%** | **0%** | **0%** | **0%** | 8% | 100% |
| | | Erratic | 1% | **0%** | **0%** | 3% | 16% | 100% |
| | | Intermittent | 49% | 49% | 49% | 50% | **21%** | 50% |
| | | Lumpy | 46% | 46% | 46% | 50% | **27%** | 50% |
| | Empirical | Smooth | **37%** | 39% | 41% | 40% | 50% | 50% |
| | | Erratic | 12% | 25% | **0%** | 4% | 22% | 100% |
| | | Intermittent | **39%** | **39%** | **39%** | 40% | 52% | 50% |
| | | Lumpy | **46%** | **46%** | 48% | 50% | 47% | 50% |
| PD | Simulated | Smooth | NA[6] | NA | NA | NA | NA | NA |
| | | Erratic | NA | NA | NA | NA | NA | NA |
| | | Intermittent | 49% | 50% | 48% | 50% | **33%** | 50% |
| | | Lumpy | 46% | 44% | 46% | 50% | **26%** | 50% |
| | Empirical | Smooth | NA | NA | NA | NA | NA | NA |
| | | Erratic | NA | NA | NA | NA | NA | NA |
| | | Intermittent | 46% | 46% | 43% | 48% | **39%** | 50% |
| | | Lumpy | 45% | 44% | 46% | 51% | **29%** | 50% |
| SD | Simulated | Smooth | **0%** | 6% | 13% | 21% | 5% | 100% |
| | | Erratic | 24% | 34% | 58% | 67% | **0%** | 62% |
| | | Intermittent | 47% | 46% | 48% | 50% | **23%** | 50% |
| | | Lumpy | 41% | 41% | 48% | 52% | **14%** | 50% |
| | Empirical | Smooth | NA | NA | NA | NA | NA | NA |
| | | Erratic | 11% | 13% | 21% | 53% | **6%** | 98% |
| | | Intermittent | 48% | **47%** | 48% | 48% | 50% | 50% |
| | | Lumpy | 41% | 40% | 41% | 50% | **34%** | 50% |

## 7.2 Inventory performance

Chapter 5 discusses the measure that applies the tradeoff curve between achieved CSL and inventory cost to evaluate the extent to which the circle service level can achieve under the same inventory cost. Obviously, the higher the curve, the better the performance because at the same inventory cost, the curve far away from the x-axis achieves a higher circle service level. The tradeoff curve shows that different methods under three types of trends where the achieved CSL (y-axis) correspond to the target CSL i.e. 0.8, 0.85, 0.9, 0.92, 0.95, 0.97, 0.99. To demonstrate the curve more clearly, ZF is not considered in the following figures. Without a doubt, it has the worst performance due to the zero forecasts for each cycle, which leads to a low safety stock level, making the actual demands levels higher than that. Note that the price remains the same in the simulated dataset, and thus the inventory investment (x-axis) equals the stock amounts.

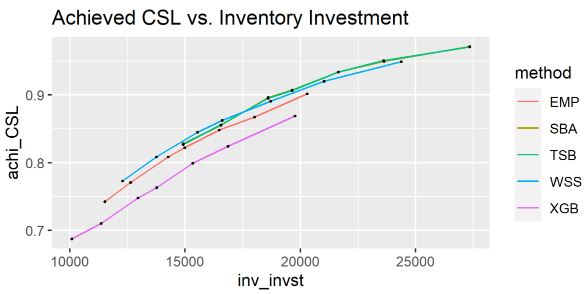
### 7.2.1 Numerical Investigation

Figure 8 describes the tradeoff curve between average achieved CSL and inventory investment. Note that SBA and TSB in Figure 8 sometimes overlap with each other. The main findings are listed as follows:

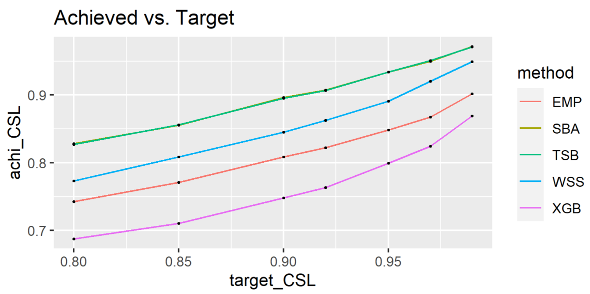
**Finding 6.** According to Figures 8a, 8c, and 8e, the performance of XGB falls behind the others. The gap between XGB and other methods under SD is even more significant than those of the other two trends. The result is also in line with that in Figures 8b, 8d, and 8f.

**Finding 7.** Figure 8 indicates that SBA and TSB outperform WSS or EMP in the higher target CSL point.

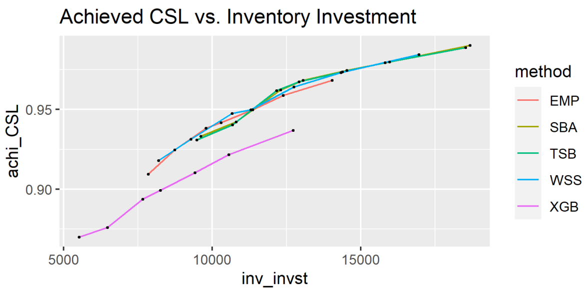
Regarding Finding 8, even though XGB has a low error rate presented in the previous section, it doesn't mean XGB has a good inventory performance. Considering the opposite result between accuracy rate and tradeoff curve, we can conclude that XGB may have a good forecasting demand averagely but is weak in inventory control, bringing some



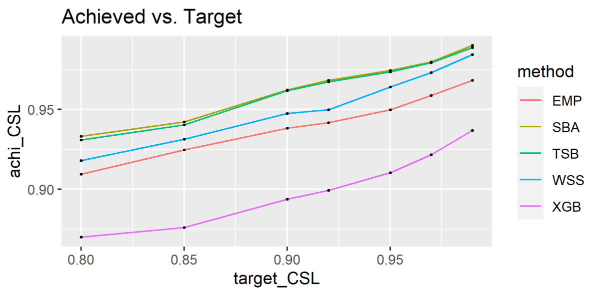

(a) Trend N: A normal demands without decreasing trends



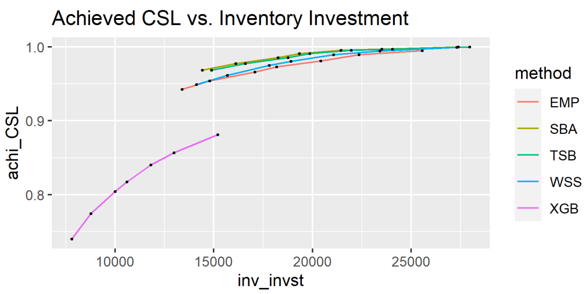

(b) Trend N: A normal demands without decreasing trends



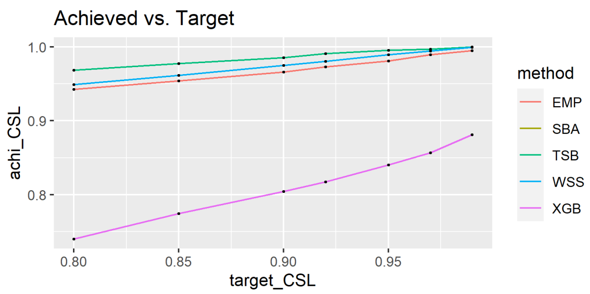

(c) Trend PD: A decreasing demands probabilities




(d) Trend PD: A decreasing demands probabilities



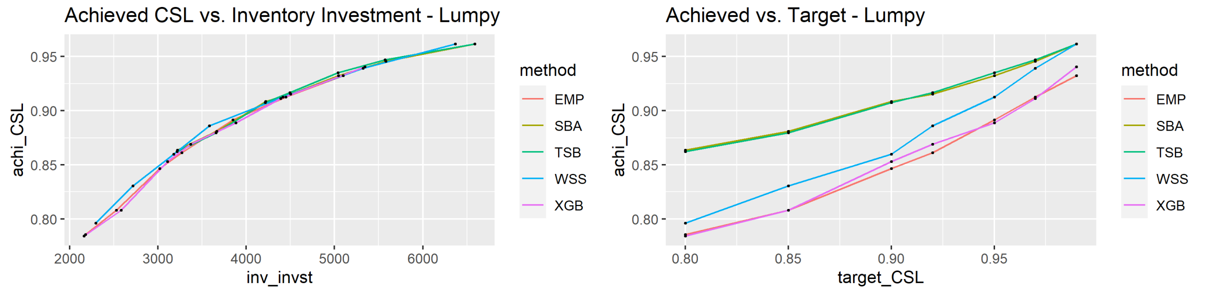

(e) Trend SD: A decreasing demands size




(f) Trend SD: A decreasing demands size

Figure 8: The tradeoff curve of simulated data: average achieved CSL vs. inventory investment.

shortages in the cycle period.




(a) The tradeoff curves in the lumpy pattern under trend N

(b) The achieved CSL and target CSL in the lumpy pattern under trend N

Figure 9: The result from pattern Lumpy under trend N in simulated data.

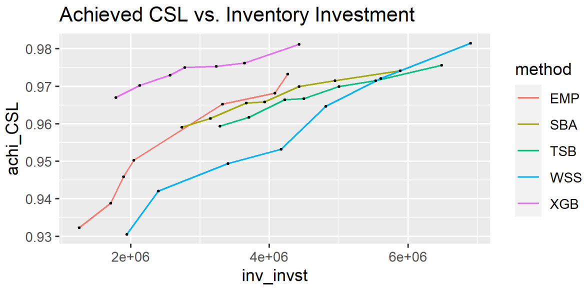
As for item patterns under three trend types, XGB also performs worse than others in most scenarios except the performance in pattern Lumpy in trend N. Despite the exception, XGB only wins over EMP, according to Figure 9.

### 7.2.2 Empirical Investigation

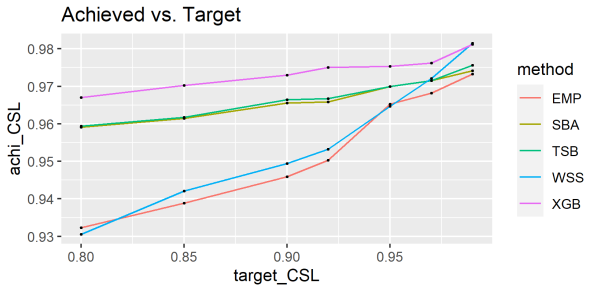
Figure 10, Figure 11, and Figure 12 describe the relationships between average achieved CSLs and inventory investment in three empirical datasets, representing the dataset RO, MF, and RAF, respectively.

Sometimes, the figures on the left side are not totally in line with those on the right side. For example, the lines in Figure 11a, Figure 11c, and Figure 11e, and those in Figure 11b, Figure 11d, and Figure 11f don't always have the same position at the same level. This difference may be because the empirical distributions of EMP and WSS are discrete. Therefore, we give priority to the left side of the figure since it refers to the inventory investment. However, if it is difficult to compare each line in the figure, we can only compare different methods through the plots on the right side. For example, in Figure 12, we notice that the tradeoff curve of EMP and WSS is too left to compare with the other three methods. Thus, we take Figure 12b, Figure 12d, and Figure 12f as the primary basis for comparison due to the precise position provided.

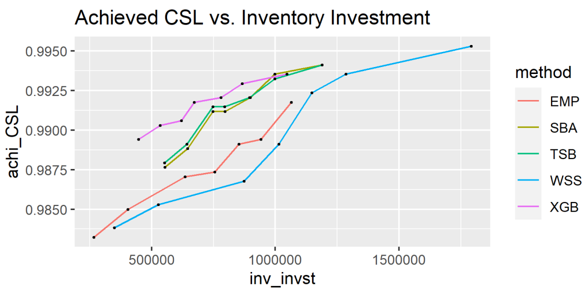
From Figure 10, Figure 11, and Figure 12, the results of three empirical datasets are



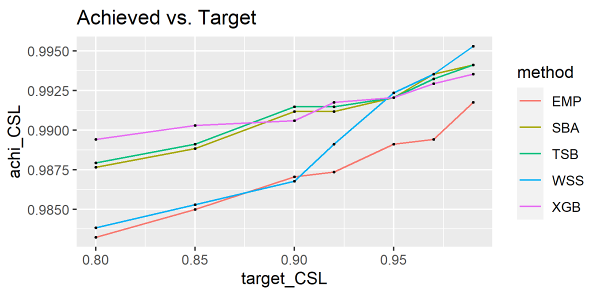

(a) Trend N: A normal demands without decreasing trends



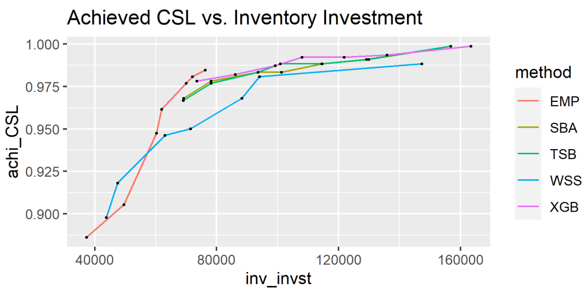

(b) Trend N: A normal demands without decreasing trends



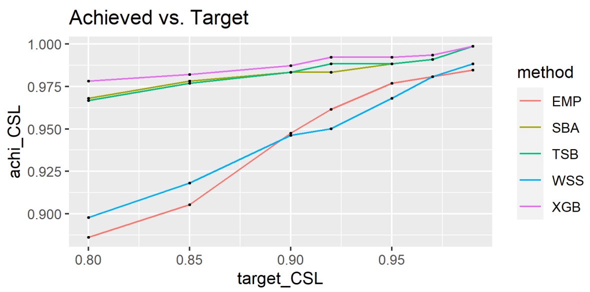

(c) Trend PD: A decreasing demands probabilities




(d) Trend PD: A decreasing demands probabilities

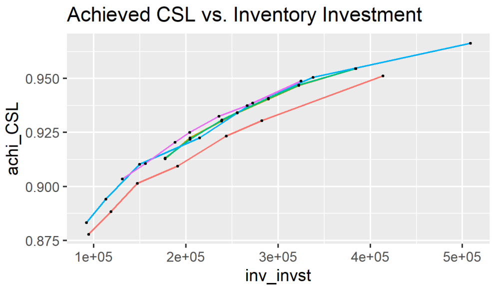


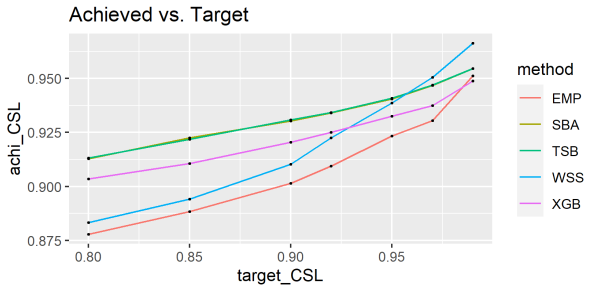

(e) Trend SD: A decreasing demands size




(f) Trend SD: A decreasing demands size

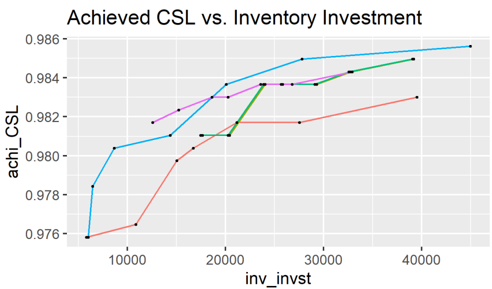
Figure 10: Dataset RO: The figure on the left side shows the tradeoff curve of average achieved CSL and inventory investment; The figure on the right side presents the tradeoff curve between average achieved CSL and target CSL

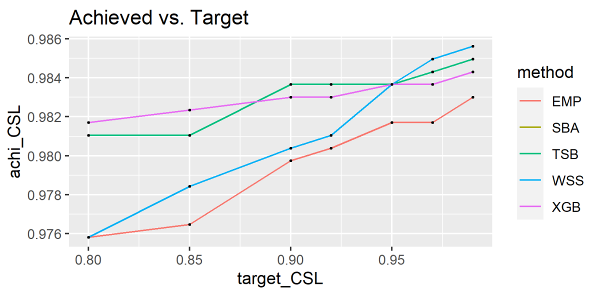



(a) Trend N: A normal demands without decreasing trends

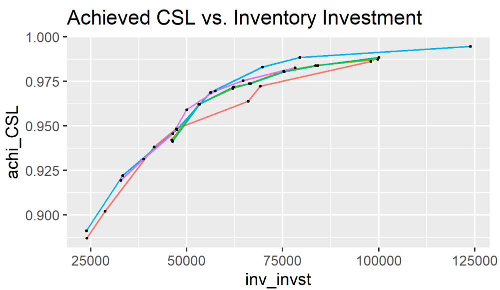
(b) Trend N: A normal demands without decreasing trends

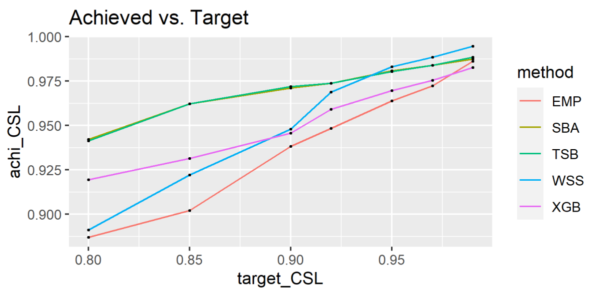



(c) Trend PD: A decreasing demands probabilities

(d) Trend PD: A decreasing demands probabilities

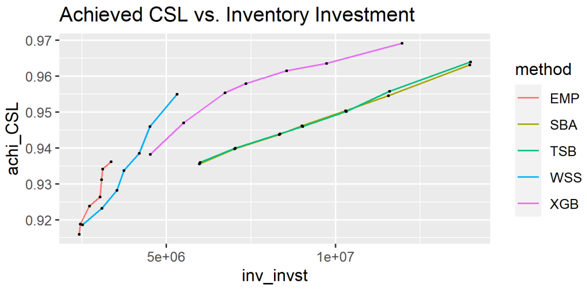


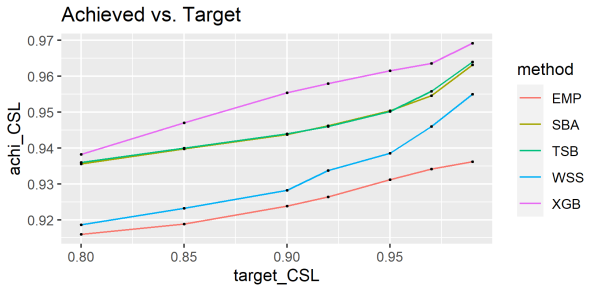

(e) Trend SD: A decreasing demands size

(f) Trend SD: A decreasing demands size

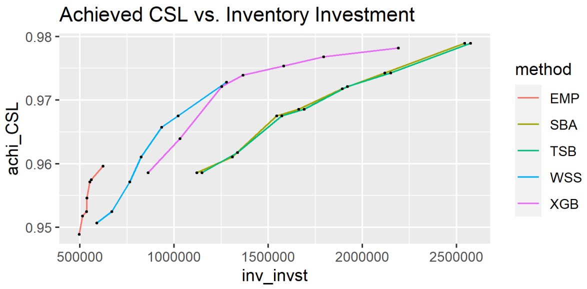
Figure 11: Dataset MF: The figure on the left side shows the tradeoff curve of average achieved CSL and inventory investment; The figure on the right side presents the tradeoff curve between average achieved CSL and target CSL



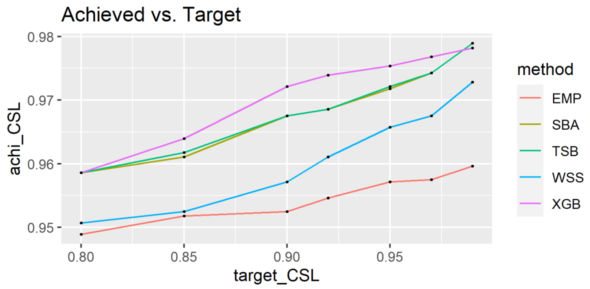

(a) Trend N: A normal demands without decreasing trends



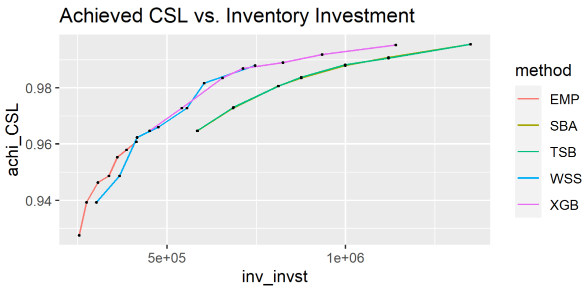

(b) Trend N: A normal demands without decreasing trends



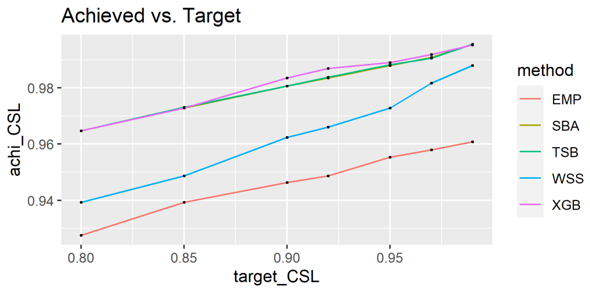

(c) Trend PD: A decreasing demands probabilities




(d) Trend PD: A decreasing demands probabilities




(e) Trend SD: A decreasing demands size




(f) Trend SD: A decreasing demands size

Figure 12: Dataset RAF: The figure on the left side shows the tradeoff curve of average achieved CSL and inventory investment; The figure on the right side presents the tradeoff curve between average achieved CSL and target CSL

presented below:

**Finding 8.** XGB shows competitive inventory performance under three trends in empirical datasets.

Regarding Finding 8, XGB in Figure 10 and Figure 12 achieves better inventory performance under three trends. Despite that XGB is not the best method according to Figure 11 for dataset MF, the line of XGB still slightly runs above those of others.

However, the result in empirical datasets is not following that in the simulated dataset. This may be due to the number of SKUs with different patterns. In empirical datasets, according to Table 5, the items of Intermittent and Lumpy are the majority, while in the simulated dataset, the distribution of the pattern is more even. Thus, we suppose XGB presents better inventory performance in Intermittent and Lumpy.

**Finding 9.** Overall, the Croston-based methods are better than EMP and WSS in the empirical dataset.

Regarding Finding 9, WSS has a better performance than the Croston-based methods, SBA and TSB. Yet, for most of the cases, SBA and TSB outperform the non-parametric methods in inventory control. This result is in line with Finding 7 from simulated datasets.

# 8 Conclusions

In this chapter, the summary of the findings above is presented, corresponding to the research questions brought up in Chapter 1. The findings include the simulated results and empirical study. In addition, the contribution of this paper is brought up. Furthermore, we discuss the limitations of our research.

## 8.1 Summary

The forecasting accuracy from simulated datasets and empirical ones are basically in line with each other. Considering the consistent results from simulated and empirical datasets, we can conclude that the XGBoost outperforms other methods under trend PD and SD. Furthermore, the XGBoost method dominates in pattern Intermittent and Lumpy under trend PD and pattern Erratic and Lumpy under trend SD. Overall, the XGBoost method gains its superiority in forecasting accuracy when the non-zero demand trend decreases for demand probabilities and demand size. The result in accuracy rate is consistent with the results of Kiefer et al. (2021) that shows the XGBoost outperforms the Croston method.

Furthermore, even though the performance of the XGBoost method is excellent. Yet, it takes the longest computing time to train and validate the model. Thus, the tradeoff between time and accuracy still needs to be considered when applied to forecasting in the industry.

According to the results of inventory investment, two out of three empirical datasets show the superiority of XGB for three trends. Thus, even though XGB doesn't dominate other methods in dataset MF, it is still better or equal to other methods. Yet, the results in simulated datasets are quite the opposite: XGB has the worst inventory performance under all trends. There are two differences between simulated and empirical datasets. First, the simulated datasets' price equals 1, and the lead time is assumed three periods. Second, the empirical datasets contain much fewer items of Smooth and Erratic than the simulated datasets do. Figure 7 shows that Lumpy is the only pattern where XGB has a slight advantage, yet it is still not dominated. Therefore, differences between the outcome of simulated data and empirical data could be because of the price of the items, the lead time, or other factors such as variation of demands or demands intervals, other than what the simple categorization scheme can explain. Although more research is needed to explain this gap, the empirical results are more convincing since the datasets come from the real world. In addition, we believe that the XGBoost has a better result because the XGBoost method applies the lag of the response data to forecasting, thus enabling to catch the pattern of the demand.

The contribution of this study can be concluded as follows. The XGBoost method has been tested to be an effective technique in forecasting intermittent SKUs in this research, providing a new technique to forecast intermittent items. Moreover, as a comparative study, this research that compares the XGBoost method with other conventional statistic techniques also fills the gap between this particular technique and other traditional methods under the prevalence of neural networks in this area. Also, identified in this study, an innovative trend of obsolescence in which the probability decreases, offers a new condition to consider. Another contribution of this study is also indicated by the application of inventory performance to the XGBoost method, leading to a more realistic evaluation of the actual world.

## 8.2 Limitations

Overall, XGB has demonstrated a method that performs well in decreasing trends using forecasting accuracy and inventory investment as measurement (except for the results in inventory investment in the simulated dataset). However, this paper includes some assumptions and restrictions, and thus some limitations are identified.

Firstly, the normal distribution of lead time demand for SBA, TSB, and XGB is assumed. Yet, the results might be different if another distribution is introduced, such as negative binomial distribution(Teunter, Duncan (2009)) or lognormal distribution(Teunter, Duncan (2009))

Secondly, the simulated datasets assume that the price is one and the lead time demand is three for each SKU. That could lead to a different tradeoff curve of inventory investment. Therefore, a more diverse simulated dataset is needed to be more similar to the actual world data.

Lastly, this paper only includes one machine learning method as a result of lacking another benchmark method. Other machine learning techniques can also be incorporated in future work.

# A Accuracy rate

Table 9 describes the original accuracy rate before taking max-min normalization.

Table 9: The comparison of accuracy of forecasting methods for MASE an sMAPE. The accuracy rate of XGB and the method of best performance are highlighted.

| | | | Parametric | | Non-parametric | | ML | Benchmark |
|---|---|---|---|---|---|---|---|---|
| Measure | Trend | Pattern | SBA | TSB | EMP | WSS | XGB | ZF |
| MASE | N | Smooth | **0.72** | **0.72** | **0.72** | 0.73 | 0.84 | 1.93 |
| | | Erratic | **0.73** | **0.73** | **0.73** | 0.75 | 0.80 | 1.04 |
| | | Intermittent | 0.99 | 0.99 | 0.99 | 1.00 | 0.80 | **0.73** |
| | | Lumpy | 0.89 | 0.89 | 0.89 | 0.92 | 0.73 | **0.67** |
| | PD | Smooth | NA[8] | NA | NA | NA | NA | NA |
| | | Erratic | NA | NA | NA | NA | NA | NA |
| | | Intermittent | 1.99 | 2.00 | 1.96 | 2.03 | 1.05 | **0.61** |
| | | Lumpy | 1.75 | 1.68 | 1.78 | 1.89 | 0.82 | **0.54** |
| | SD | Smooth | **0.92** | 1.09 | 1.29 | 1.51 | 1.00 | 2.69 |
| | | Erratic | 1.79 | 2.12 | 3.01 | 3.32 | **0.93** | 1.52 |
| | | Intermittent | 1.28 | 1.27 | 1.30 | 1.33 | 0.89 | **0.76** |
| | | Lumpy | 1.94 | 1.93 | 2.12 | 2.22 | 0.91 | **0.71** |
| sMAPE | N | Smooth | **0.48** | **0.48** | **0.48** | 0.48 | 0.57 | 2.00 |
| | | Erratic | **0.76** | **0.76** | **0.76** | 0.77 | 0.88 | 2.00 |
| | | Intermittent | 1.61 | 1.61 | 1.61 | **1.60** | 1.67 | 2.00 |
| | | Lumpy | 1.56 | 1.56 | 1.56 | **1.55** | 1.68 | 2.00 |
| | PD | Smooth | NA | NA | NA | NA | NA | NA |
| | | Erratic | NA | NA | NA | NA | NA | NA |
| | | Intermittent | **1.77** | **1.77** | **1.77** | **1.77** | 1.85 | 2.00 |
| | | Lumpy | **1.78** | **1.78** | **1.78** | **1.78** | 1.85 | 2.00 |
| | SD | Smooth | **0.40** | 0.44 | 0.48 | 0.53 | 0.49 | 2.00 |
| | | Erratic | 0.86 | 0.93 | 1.08 | 1.13 | **0.69** | 2.00 |
| | | Intermittent | 1.51 | 1.51 | **1.50** | **1.50** | 1.61 | 2.00 |
| | | Lumpy | **1.48** | **1.48** | 1.49 | 1.50 | 1.56 | 2.00 |

[8] No type under this demands trend.

# References

*Alim Mirxat, Ye Guo-Hua, Guan Peng, Huang De-Sheng, Zhou Bao-Sen, Wu Wei.* Comparison of ARIMA model and XGBoost model for prediction of human brucellosis in mainland China: a time-series study // BMJ open. Dec 07, 2020. 10, 12. e039676.

*Babai M. Z., Dallery Y., Boubaker S., Kalai R.* A new method to forecast intermittent demand in the presence of inventory obsolescence // International journal of production economics. Mar 2019. 209. 30–41.

*Babai Mohamed Zied, Syntetos Aris, Teunter Ruud.* Intermittent demand forecasting: An empirical study on accuracy and the risk of obsolescence // International journal of production economics. Nov 2014. 157, 1. 212–219.

*Beata Ślusarczyk.* Importance of Spare Parts Supplies in Air Transport // International journal of innovation, management and technology. 2018. 169–173.

*Brown Robert.* Statistical forecasting for inventory control // Journal of the Royal Statistical Society. 1960. 348–349.

*Chatfield Dean C., Hayya Jack C.* All-zero forecasts for lumpy demand: a factorial study // International journal of production research. Feb 15, 2007. 45, 4. 935–950.

*Chen Tianqi, Guestrin Carlos.* XGBoost: A Scalable Tree Boosting System // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016.

*Croston J. D.* Forecasting and Stock Control for Intermittent Demands // The Journal of the Operational Research Society. Sep 01, 1972. 23, 3. 289–303.

*Ellis Peter.* forecastxgb-r-package Kernel Description. 05 2017. Accessed: 2017-05-17.

*Friedman J.* Greedy function approximation: A gradient boosting machine. // Annals of Statistics. 2001. 29. 1189–1232.

*Gallagher Tim, Mitchke M., Rogers M.* Profiting from spare parts // The McKinsey Quarterly. Mar 2, 2005.

*Gutierrez Rafael S., Solis Adriano O., Mukhopadhyay Somnath.* Lumpy demand forecasting using neural networks // International journal of production economics. 2008. 111, 2. 409–420.

*Hasni M., Aguir M. S., Babai M. Z., Jemaï Zied.* Spare parts demand forecasting: a review on bootstrapping methods // International Journal of Production Research. 2019. 57. 4791 – 4804.

*Hyndman Rob J., Koehler Anne B.* Another look at measures of forecast accuracy // International journal of forecasting. Oct 2006. 22, 4. 679–688.

*Inder B.A.* Finite-sample power of tests for autocorrelation in models containing lagged dependent variables // Economics Letters. 1984. 14, 2. 179–185.

*Jiang Peng, Huang Yibin, Liu Xiao.* Intermittent demand forecasting for spare parts in the heavy-duty vehicle industry: a support vector machine model. -11-11 2020. 1.

*Johnston F. R., Boylan J. E.* Forecasting for Items with Intermittent Demand // The Journal of the Operational Research Society. Jan 01, 1996. 47, 1. 113–121.

*Kaya Gamze Ogcu, Turkyilmaz Ali.* INTERMITTENT DEMAND FORECASTING USING DATA MINING TECHNIQUES // Applied Computer Science. Jun 01, 2018. 14, 2. 38–47.

*Kennedy W. J., Patterson J. Wayne, Fredendall Lawrence D.* An overview of recent literature on spare parts inventories // International journal of production economics. 2002. 76, 2. 201–215.

*Kiefer Daniel, Grimm Florian, Bauer Markus, Van Dinther.* Demand Forecasting Intermittent and Lumpy Time Series: Comparing Statistical, Machine Learning and Deep Learning Methods. 2021.

*Kim Thai Young, Dekker Rommert, Heij Christiaan.* Spare part demand forecasting for consumer goods using installed base information. Feb 25, 2016.

*Kourentzes Nikolaos.* On intermittent demand model optimisation and selection // International journal of production economics. Oct 2014. 156. 180–190.

*Makridakis Spyros, Spiliotis Evangelos, Assimakopoulos Vassilios.* Statistical and Machine Learning forecasting methods: Concerns and ways forward // PLoS ONE. 2018. 13.

*Nelder J. A., Mead R.* A Simplex Method for Function Minimization // The Computer Journal. 01 1965. 7, 4. 308–313.

*Pince Cerag, Dekker Rommert.* ERIM REPORT SERIES RESEARCH IN MANAGEMENT ERIM Report Series reference number An Inventory Model for Slow Moving Items Subject to Obsolescence. -08 2010.

*Pinçe Çerag, Dekker R.* An Inventory Model for Slow Moving Items Subject to Obsolescence // Eur. J. Oper. Res. 2011.

*Pinçe Çerağ, Turrini Laura, Meissner Joern.* Intermittent demand forecasting for spare parts: A Critical review // Omega. 2021. 105. 102513.

*Porras Eric, Dekker Rommert.* An inventory control system for spare parts at a refinery: An empirical comparison of different re-order point methods // European journal of operational research. 2008. 184, 1. 101–132.

*Prestwich Steven, Rossi Roberto, Tarim S. Armagan, Hnich Brahim.* Mean-based error measures for intermittent demand forecasting // International journal of production research. Nov 17, 2014. 52, 22. 6782–6791.

*Schultz Carl R.* Forecasting and Inventory Control for Sporadic Demand Under Periodic Review // Journal of the Operational Research Society. 1987. 38. 453–458.

*Smyl Slawek.* A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting // International Journal of Forecasting. 2020. 36. 75–85.

*Sugiono N.K., Alimbudiono R.S.* Slow Moving and Dead Stock: Some Alternative Solutions // Proceedings of the 17 th International Symposium on Management (INSYMA 2020). 2020. 330–335.

*Syntetos A., Boylan J.* On the stock control performance of intermittent demand estimators // International Journal of Production Economics. 2006. 103. 36–47.

*Syntetos A., Nikolopoulos K., Boylan J.* Judging the judges through accuracy-implication metrics: The case of inventory forecasting // International Journal of Forecasting. 2010. 26. 134–143.

*Syntetos A. A., Boylan J. E., Croston J. D.* On the categorization of demand patterns // The Journal of the Operational Research Society. May 01, 2005. 56, 5. 495–503.

*Syntetos Aris A., Babai M. Zied, Gardner Everette S.* Forecasting intermittent inventory demands: simple parametric methods vs. bootstrapping // Journal of business research. Aug 2015. 68, 8. 1746–1752.

*Syntetos Aris A., Boylan John E.* The accuracy of intermittent demand estimates // International journal of forecasting. Apr 2005. 21, 2. 303–314.

*Teunter R. H., Duncan L.* Forecasting intermittent demand: a comparative study // The Journal of the Operational Research Society. Mar 01, 2009. 60, 3. 321–329.

*Teunter Ruud H., Syntetos Aris A., Babai M. Zied.* Intermittent demand: Linking forecasting to inventory obsolescence // European journal of operational research. 2011. 214, 3. 606–615.

*WANG WEIZENG, SHI YULIANG, LYU GAOFAN, DENG WANGHUA.* Electricity Consumption Prediction Using XGBoost Based on Discrete Wavelet Transform // DEStech Transactions on Computer Science and Engineering. Oct 31, 2017. aiea.

*Willemain T., Smart C. N., Shockor Joseph H., DesAutels P.* Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method // International Journal of Forecasting. 1994. 10. 529–538.

*Willemain Thomas R., Smart Charles N., Schwarz Henry F.* A new approach to forecasting intermittent demand for service parts inventories. -07 2004. 375.

*Zhang Dahai, Qian L., Mao Baijin, Huang Can, Huang Bin, Si Yulin.* A Data-Driven Design for Fault Detection of Wind Turbines Using Random Forests and XGboost // IEEE Access. 2018. 6. 21020–21031.