More grip on inventory control through improved forecasting
A comparative study at three companies

R.J.I. Basten\textsuperscript{1,}\textsuperscript{*}, E. van Wingerden\textsuperscript{2}, R. Dekker\textsuperscript{3}, and W.D. Rustenburg\textsuperscript{4}

\textsuperscript{1}University of Twente, P.O. Box 217, 7500 AE, Enschede, The Netherlands. r.basten@utwente.nl.
\textsuperscript{2}Eindhoven University of Technology, P.O. Box 513, 5600 MB, Eindhoven, The Netherlands. e.v.wingerden@tue.nl.
\textsuperscript{3}Erasmus University Rotterdam, P.O. Box 1738, 3000 DR, Rotterdam, The Netherlands. rdekker@ese.eur.nl.
\textsuperscript{4}Gordian Logistic Experts, Groenewoudse dijk 63, 3528 BG, Utrecht, The Netherlands. jw.rustenburg@gordian.nl.

October 8, 2012
Report Econometric Institute EI2012–24

Abstract
Inventory control for parts with infrequent demands is difficult since forecasting their demand is problematic. Traditional forecasting methods, such as moving average and single exponential smoothing, are known not to suffice since they do not cope well with periods with zero demands. Croston type methods and bootstrapping methods are more promising. We propose a new bootstrapping method, which we term empirical plus. The added value of this method lies in the fact that it explicitly takes into account that besides the demands, also the supply lead time is stochastic. We compare its performance with a number of methods from all three above-mentioned categories. Opposite to what is done in most comparative studies, we do not focus on performance metrics that are related directly to the forecasting results (e.g., mean squared error), but we focus on the resulting inventory control policy (achieved fill rate and holding costs). We use in our study large data sets from three companies, which we make publicly available. We find that for most parts, the Syntetos Boylan approximation (a Croston type method) performs best. However, if the average inter-demand interval is large and the squared coefficient of variation of the demand size per period is small, the empirical plus method performs better as long as fill rates are not too high. The specific class of parts for which the empirical plus method performs best consists often of the expensive parts, for which forecasting is both difficult, because of the infrequent demands, and important, because of the price. This high price also means that the target fill rate is often not too high. Our findings may therefore lead to large cost reductions in practice.

Keywords: Inventory; spare parts; forecasting

\textsuperscript{*}Corresponding author. Tel.: +31-53-489-4007.
1 Introduction

Companies making, using, or maintaining capital goods typically have large inventories of spare parts. According to the Aberdeen Group (2005) their value exceeds a trillion dollar. Hence, there is a high need to reduce them, also because a lot of these parts may never be used and have to be discarded at some point in time. At the same time, a lack of spare parts can be very costly and high penalties may have to be paid in case of unavailability. The management of spare parts stock is a separate discipline, different from finished product inventory management, as demand is often characterized as being infrequent with sudden high demands. This makes the problem of determining the right amount of stocks an important scientific and practical problem.

Quite some methods have been developed for the inventory control of spare parts, see for example the overview by Kennedy et al. (2002). In many inventory control models a specific demand process, like the Poisson process, is assumed with known parameters. Yet in any real case, neither the demand process nor its parameters are known and they have to be estimated or forecasted.

A recent stream in the forecasting literature investigates models to predict the demand for spare parts. Yet there are few tests of these methods with real data and not always all relevant characteristics of the parts have been revealed. Even less papers have addressed the combination of forecasting and spare parts control, as the typical performance criterion in forecasting, the mean squared error, is not the leading characteristic in inventory control.

Another important practical question is whether one approach can be used for all items or whether we should use different approaches for different types of parts. In other words, what is the best classification of spare parts with respect to their inventory control? This is particularly relevant when developing a decision support system for spare parts inventory control.

In this paper, we compare several spare part demand forecasting methods using real data from three companies. The number of different spare parts they keep is quite large and ranges from some 4,000 to over 250,000 different parts. Still, we consider much less parts for several reasons that we discuss in Section 5.1. One company is involved in the maintenance of passenger trains in the Netherlands, the second supplies equipment for trains and other industrial systems, and the third organization is involved in the maintenance of naval ships. Apart from characterizing their spare parts demand situations, we also provide readers access to the data in a complementary file, which allows them to do new analyses.

The methods we are comparing are simple moving average, single exponential smoothing, double exponential smoothing, two Croston type methods, a so-called MSE method, and two bootstrapping methods. The MSE method is a method that picks the best traditional or Croston type method per part, based on the resulting MSE (mean squared (forecasting) error). The key reason to consider bootstrapping methods is that they immediately forecast the demand over the lead time, opposed to traditional and Croston type methods that forecast the demand per period and then multiply that by the lead time. Hence the bootstrapping methods can incorporate all kind of hidden correlations in demand. One of the two bootstrapping methods
that we use is a new method, in which we explicitly incorporate the fact that not only demands, but also the supply lead time is stochastic.

Using all data, we compare the performance of these methods combined with an inventory control method by establishing trade-off curves between inventory holding costs and service levels. This gives a better view of their performance than comparing costs for one service level target.

Summarizing, we contribute in four ways:

- We propose a new bootstrapping method that explicitly incorporates a stochastic supply lead time.
- We perform a comparative study, focussing on the resulting inventory control, in which we compare traditional, Croston type, and bootstrapping methods.
- We further incorporate in our comparative study the so-called MSE method, through which we show that focussing on the MSE when forecasting may not result in the best inventory control policy.
- For the study we use an extensive data set from three companies, which we make publicly available.

The set-up of the paper is as follows. We discuss the relevant literature in Section 2, including detailed descriptions of the forecasting methods that we will use in our study. We next describe the three companies that supplied us with data sets in Section 3. In Section 4, we then propose the empirical plus method. We explain the set-up of the simulation study in Section 5, and we discuss the results in Section 6. Finally, we give conclusions and recommendations in Section 7.

2 Literature review

The literature on forecasting of spare parts, or slow moving items, is quite extensive, starting with the paper by Croston (1972). A lot of progress has been made from the mid 1990’s onwards and the topic still receives a lot of attention. There exist two relatively recent review papers that include forecasting methods: Boylan and Syntetos (2010) focus on spare parts forecasting, including forecast support systems, and Syntetos et al. (2009) review the literature on forecasting for general inventory control; the sections on forecasting methods for spare parts in these two papers overlap to some extent. Here, we discuss only the most relevant papers in the context of our study. First, since we perform a comparative study using a number of existing forecasting methods and a newly developed forecasting method, we refer to literature on the existing methods (we give the exact calculations for the methods in the appendix). Second, we discuss a couple of recent and related comparative studies, especially those that, as we, consider the resulting inventory control.

We discuss a number of forecasting methods that are commonly used in industry (the traditional methods) and some methods that are reported in the literature to perform well (the
Croston type and bootstrapping methods). Our aim is not to compare (and thus discuss) all forecasting methods, but to compare methods from each of the three categories. The first type of forecasting methods are the traditional methods that are not designed specifically for slow moving items. These methods are discussed in various text books (e.g., Axšäter, 2006) and we consider the simple moving average (SMA; by some authors just called moving average), single exponential smoothing (SES; by some authors just called exponential smoothing), and double exponential smoothing (DES). The downside of these models is that they do not explicitly take into account that there may be periods with zero demands, whereas this is quite common for spare parts.

To cope with this problem, Croston (1972) proposes Croston’s method, an exponential smoothing method that updates forecasts only in periods with positive demands. It forecasts both the demand size in periods with positive demands and the average number of periods between two periods with positive demands, the inter-demand interval. These forecasts are combined to come up with a forecast of the average demand per period. Since Syntetos and Boylan (2001) find that Croston’s method is biased, Syntetos and Boylan (2005) propose an improved version of Croston’s method, the Syntetos Boylan approximation (SBA). Together, we call these two methods the Croston type methods. An interesting new Croston type method is that by Teunter et al. (2011). As the other Croston type methods, this method adapts the expected demand size only in periods with positive demand. However, the expected inter-demand interval is adapted in each period, so that inventory levels may be decreased when demand decreases (e.g., when a component has become obsolete). Since the method has been published after we had performed the main parts of our study, we have not included this method in our study.

The third and final category of forecasting methods are the bootstrapping methods. Probably the most well known method in this category is that of Willemain et al. (2004), but Porras and Dekker (2008) propose another method and, using a data set from a refinery, report that their method slightly outperforms that of Willemain et al.. Since the method of Porras and Dekker is also easier to understand and implement, we discuss (and use) the latter method only, in Section 4. The method is called the empirical method, because it determines an empirical demand distribution by sampling from historical demands.

There also exist other types of methods that take more information into account, e.g., about the changes in the size of the installed base, reliability data as supplied by the manufacturer, data resulting from maintenance planning, or degradation (condition) information from the installed base. In this paper, we have chosen to focus on pure time series analysis methods only. We come back to this point in our recommendations in Section 7.

There exist a number of comparative studies in which various forecasting methods are compared using fictitious or real-life data. Teunter and Duncan (2009) give an overview of such studies upto 2006 and find that most studies use some measure of the resulting forecast error to compare methods, while it would be better to compare the resulting inventory control (achieved service level and holding costs). In our study we do exactly that. Here, we discuss a
number of recent papers with the same focus. First, however, we discuss Syntetos et al. (2005), who classify parts based on two characteristics: the average inter-demand interval, \( p \), and the squared coefficient of variation of the demand sizes, \( CV^2 \), which is calculated by dividing the standard deviation of the demand size by the average demand size and squaring the result. Only positive demand sizes are taken into account in this calculation and notice that \( p \geq 1 \) by definition. Syntetos et al. make the classification by theoretically comparing the mean squared error (MSE) that will result from applying SES, Croston’s method, and SBA. They find that Croston’s method performs best when \( p < 1.32 \) and \( CV^2 < 0.49 \). In all other cases, SBA performs best. They confirm their findings using a set of 3,000 parts from the automotive industry. Boylan et al. (2008) compare the performance of SMA, SES, Croston’s method, and SBA on 16,000 parts from the automotive, aerospace, and chemical industries. They focus on the implications of the forecasting method on the resulting inventory control. Their findings confirm those of Syntetos et al. (2005), but they also find that performance of the total system is robust to the exact choice of the cut-off values. In the current study, we also classify parts, but we include bootstrapping methods in the comparison. Syntetos et al. (2009, p.S154) and Boylan and Syntetos (2010, p.231) mention that it would be useful to have empirical studies comparing also the performance of those methods.

Teunter and Duncan (2009) use data on 5,000 parts of the British Royal Air Force to compare the performance of the zero forecast (forecast is always zero), SMA, SES, Croston’s method, SBA, and a bootstrapping method. They show that the zero forecast outperforms the other methods in some cases, when the performance is compared using forecast errors, e.g., the mean squared error. For that reason, they propose to compare using the resulting inventory control. Doing that, they find that Croston’s method, SBA, and the bootstrapping method have a similar performance. However, all methods achieve lower service levels than the set target. Teunter and Duncan therefore propose a modification for Croston type methods and bootstrapping methods that takes into account the fact that an order in a period is triggered by a demand in that period. We have not used this modification in our study because of the focus of our study: we want to compare well known and (relatively) widely used methods to see what types of methods perform best. We do acknowledge that when implementing forecasting methods in practice, it may be beneficial to include the modification that Teunter and Duncan propose.

### 3 Case companies

We use data from three companies, which we make publicly available\(^1\). The data is on individual order level, time marked with the exact date. This is different from data that many companies can supply, since they store aggregated demand data (e.g., per month) only, which seriously hampers analysis. In our analysis, we exclude parts for which the data is not useful, for instance because the part faced no demand. We give full details of this so-called filtering in Section 5.1.

\(^1\)If you would like to receive the data sets, for non-commercial use, please send an e-mail to research@gordian.nl.
To understand the background of the demand we describe some of the key characteristics and processes of NedTrain Haarlem, Alstom Ridderkerk, and the Naval Maintenance Company in Section 3.1, 3.2, and 3.3, respectively. As all these companies support rather permanent operations, their business is less sensitive to business cycles than, e.g., manufacturing companies.

3.1 NedTrain Haarlem

The first company is NedTrain Haarlem (NTH), part of NedTrain, which is a subsidiary of the Netherlands Railways Group. NTH performs overhaul, conversion, and upgrading of trains/trams/metro cars. We concentrate on the activities of the overhauling of bogies and wheel sets, including age based maintenance, repair, and modernisation. Annually some 1,100 bogies and some 3,000 wheel sets are overhauled. In general terminology, these bogies and wheel sets can be considered line replaceable units, while at NTH they are referred to as main parts. The data we use concerns the usage of parts in the overhauling of bogies and wheel sets. It covers 2,049 parts with the corresponding demand and supply orders between July 4th, 2005 and March 1st, 2011. We further have for each part its price and planned leadtime. From the order and delivery date we can compute an actual leadtime and compare that with the planned leadtime. After filtering 790 parts remain.

Demands for spare parts are triggered by projects and corrective maintenance. The first is to some extent plannable as many main parts of the same type come in for a planned overhaul within a period of two to three years. For an overhaul of each main part some parts are always needed, while some other parts are needed only occasionally. NTH has tried to estimate the probability that parts are needed in the overhaul, but the values obtained so far are not that reliable. So they treat all demand as unplanned. In case of corrective maintenance, the need for parts is only known at the moment the main part has arrived and hence stocks of parts are directly needed. When a spare part is not available when needed, the repair is delayed, which is to be avoided. We also exclude parts that can be quickly manufactured by NTH upon demand, as these are relatively cheap. Table 1 gives an overview of the main characteristics; we show figures of these characteristics in the appendix.

<table>
<thead>
<tr>
<th></th>
<th>Price$^a$</th>
<th>Planned LT$^bc$</th>
<th>Planned realised LT$^bd$</th>
<th>Number of demands</th>
<th>Demand size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of values</td>
<td>790</td>
<td>790</td>
<td>5,133</td>
<td>790</td>
<td>348,313</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.01</td>
<td>10</td>
<td>-763</td>
<td>2</td>
<td>0.00</td>
</tr>
<tr>
<td>10$^{th}$ percentile</td>
<td>1.63</td>
<td>25</td>
<td>-42</td>
<td>6</td>
<td>2.00</td>
</tr>
<tr>
<td>Average</td>
<td>433.01</td>
<td>65</td>
<td>5</td>
<td>90</td>
<td>19.27</td>
</tr>
<tr>
<td>90$^{th}$ percentile</td>
<td>622.70</td>
<td>135</td>
<td>56</td>
<td>196</td>
<td>39.00</td>
</tr>
<tr>
<td>Maximum</td>
<td>54,586.24</td>
<td>375</td>
<td>482</td>
<td>464</td>
<td>1,440.00</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>3,091.82</td>
<td>62</td>
<td>55</td>
<td>82</td>
<td>50.78</td>
</tr>
</tbody>
</table>

Table 1: Characteristics of parts at NedTrain Haarlem

$^a$ In euros $^b$ LT = lead time $^c$ In working days $^d$ In calendar days
Notice that the longest leadtime is 375 working days, which is about 525 calendar days. Hence, the largest deviation can indeed be 482 calendar days! We further like to mention that the most common demand size is 4 (or a multiple of 4, see the appendix), because of the presence of 4 wheels in a bogie. The appendix also shows a scatter plot giving the relation between price and planned leadtime. Although one may expect a strong correlation between price and lead time, this correlation is not evident from the figure.

3.2 Alstom Ridderkerk

Alstom Ridderkerk (ALS) is part of the Alstom Group, which specializes in energy solutions and transport. ALS is responsible for maintenance of electric propulsion and electronic control systems for trains, trams, and metros, modernization of traction systems, and for warehousing spare parts. Some parts are directly ordered by a customer while others are used in a production process for a repair or manufacturing of a main part. The modernization projects allow planning of the need for parts, and the ones solely used in these projects have been left out of our study. For ALS we have the same type of data as for NTH, covering 3,957 parts with non zero prices and positive demand, with orders between April 20th, 2006 and April 20th, 2011. After filtering 748 parts remain. Overall statistics are not that different from NTH, so we only give an overview (see Table 2). Figures with distributions of characteristics are given in the appendix. We omitted a scatter plot as it is similar to that of NTH.

<table>
<thead>
<tr>
<th></th>
<th>Planned</th>
<th>Planned - realised</th>
<th>Number of demands</th>
<th>Demand size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Price(^a)</td>
<td>LT(^bc)</td>
<td>LT(^bc)</td>
<td></td>
</tr>
<tr>
<td>Number of values</td>
<td>748</td>
<td>748</td>
<td>594</td>
<td>748</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.01</td>
<td>1</td>
<td>-536</td>
<td>2</td>
</tr>
<tr>
<td>10(^{th}) percentile</td>
<td>0.43</td>
<td>13</td>
<td>-307</td>
<td>2</td>
</tr>
<tr>
<td>Average</td>
<td>83.90</td>
<td>74</td>
<td>-58</td>
<td>10</td>
</tr>
<tr>
<td>90(^{th}) percentile</td>
<td>240.13</td>
<td>163</td>
<td>70</td>
<td>22</td>
</tr>
<tr>
<td>Maximum</td>
<td>1,900.00</td>
<td>434</td>
<td>333</td>
<td>96</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>189.24</td>
<td>66</td>
<td>149</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 2: Characteristics of parts at Alstom Ridderkerk

\(^a\) In euros \(^b\) LT = lead time \(^c\) In calendar days

3.3 Naval Maintenance Company

The Naval Maintenance Company (NMC) is a service centre responsible for all material within the Defense Material Organization of the Netherlands. NMC maintains, repairs, supplies, and modifies a diversity of systems. These can be propulsion systems, electronic and weapon systems of naval ships, but also guided weapons, night-vision goggles and many other equipment for the defense organization. The data we obtained from NMC consists of demands and supplies for 216,483 parts, covering July 1\(^{st}\), 2005 to April 29\(^{th}\), 2011. Using similar considerations as with NTH and ALS and focusing only on spare parts, we arrived at a selection of 5,191 parts. Table
3 gives an overview of the main characteristics of these parts. Figures with distributions are given in the appendix. We like to note that some high demand sizes, like 10 per occasion, occur relatively often. This can be caused by either a minimum order quantity (MOQ) or a modular order quantity (MOQ), but we did not have these values available. We therefore did not use that information in determining our inventory control policies.

<table>
<thead>
<tr>
<th>Number of values</th>
<th>Planned Price(^a)</th>
<th>Planned (\text{LT}^{bc})</th>
<th>Planned - realised (\text{LT}^{dl})</th>
<th>Number of demands</th>
<th>Demand size</th>
</tr>
</thead>
<tbody>
<tr>
<td>5,191</td>
<td>5,191</td>
<td>25,890</td>
<td>5,191</td>
<td>39,408</td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>0.01</td>
<td>0</td>
<td>-1,265</td>
<td>2</td>
<td>1.00</td>
</tr>
<tr>
<td>(10^{th}) percentile</td>
<td>2.28</td>
<td>2</td>
<td>-8</td>
<td>2</td>
<td>1.00</td>
</tr>
<tr>
<td>Average</td>
<td>4,723.65</td>
<td>4</td>
<td>47</td>
<td>8</td>
<td>19.25</td>
</tr>
<tr>
<td>(90^{th}) percentile</td>
<td>6,498.18</td>
<td>8</td>
<td>122</td>
<td>16</td>
<td>24.00</td>
</tr>
<tr>
<td>Maximum</td>
<td>2,697,070.10</td>
<td>37</td>
<td>1,100</td>
<td>182</td>
<td>15,000.00</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>43,379.69</td>
<td>3</td>
<td>85</td>
<td>11</td>
<td>151.20</td>
</tr>
</tbody>
</table>

Table 3: Characteristics of parts at the Naval Maintenance Company

- \(^{a}\)In euros
- \(^{b}\)LT = lead time
- \(^{c}\)In months
- \(^{d}\)In calendar days

4 Empirical plus method

We explain our new method, which is based on that of Porras and Dekker (2008), in three steps. First, we discuss the method by Porras and Dekker. Second, since we have made a slight change to that method, as mentioned in Section 2, we discuss that change. Third, we explain how our method, the empirical plus method, differs from our implementation of the empirical method. From here on, when we refer to the empirical method, we mean our implementation of that method.

The method that Porras and Dekker (2008) propose functions as follows. There is a fixed lead time \(L\). This lead time may be the average of the historical lead times, the contractual lead time that is agreed upon with the supplier, or any other lead time that the inventory planner finds appropriate. In our source data, this lead time is supplied (termed the planned lead time); it is often, but not always, the contractual lead time. First, the historical demand in periods 1 to \(L\) is summed to obtain the first lead time demand. We say that a window of size \(L\) is placed over the historical demands. The window is moved one period at a time until the end of the historical demand is reached, see Figure 1. If data of five years is available, detailed per day, there are about 1,825 data points and there would be about \(1,825 - L\) lead time demands. In this way, all historical demands over the lead time are covered and an empirical demand distribution can be constructed.

Assuming a base stock level of \(R\) and one-for-one replenishments, we use the following
Figure 1: First three lead time demands, determined by the method of Porras and Dekker (2008)

formula for predicting the part fill rate \( PFR \):

\[
PFR = 1 - \frac{\sum_{x \in X | x > R} (x - R) p(x)}{\sum_{x \in X} x p(x)},
\]

with \( X \) being the set of possible lead time demands and \( p(x) \) being the probability of seeing \( x \in X \), both according to the empirical distribution. It represents the percentage of leadtime demand that exceeds the reorder point \( R \).

Porras and Dekker (2008) use the following formula from Chopra and Meindl (2009) to predict the part fill rate \( PFR \), with \( Q \) the given order size:

\[
PFR = 1 - \frac{\sum_{x \in X | x > R} (x - R) p(x)}{Q}.
\]

This formula however assumes that \( Q \) is that large that only one replenishment order is outstanding at any point in time. For \( Q = 1 \) it may give undesirable results. Both formulas do not take into account that a replenishment cycle may be initiated by a demand larger than one, implying that less than \( R \) units are available for satisfying the demand during the leadtime.

We have adapted the empirical method as follows (as mentioned, we reserve the term empirical method for our implementation of that method). Instead of starting with a window on periods 1 to \( L \) and then moving it one period at a time, we place the window at random over \( L \) consecutive periods and we do this for a fixed number of times (500 times in our simulation study); see Figure 2a for an example. The reason for this change is that our new method, the empirical plus method, then differs from the empirical method in one way only, so that we can clearly show the value of that difference.

The difference is that instead of taking a fixed lead time, so a fixed window size, we sample from the historical lead times each time to determine the lead time and window size. So, for each of the samples (500 in our study), we first randomly pick one of the historical lead times, we use that as the window size, and we then place that window at random over the historical demands. See Figure 2b for an example.
5 Simulation study set-up

We have split the explanation of the set-up of the simulation study into three parts, each having its own section: data, forecasting, and inventory control.

5.1 Data

As explained in Section 3, we use historical data from three different companies: NedTrain Haarlem (NTH), Alstom Ridderkerk (ALS), and the Naval Maintenance Company (NMC). We remove demands if:

- they result from planned maintenance, e.g., at NMC there is a column named Geplande vraag (planned demand, in Dutch) that can have as value either Gepland (planned) or Niet gepland (unplanned), or
- they are negative, meaning that parts have been returned.

We further remove parts if:

- they did not face positive demands after the initialisation phase (of 24 months, see Section 5.2), or
- they observed positive demands during a period of at most six consecutive calendar months (since that is probably a planned project).

5.2 Forecasting

We have mentioned the forecasting methods that we use in Sections 2 and 4, except for the MSE method that we will describe below; we give an overview in Table 4. For the traditional and Croston type forecasting methods, we take an average of the demands in the first 12 (calendar) months to get the first forecast (in month 13). Months 13 to 24 are then used as a warm-up period, and from month 25 on, we can calculate an MSE over the previous 12 months and we
measure the holding costs and achieved part and order line fill rates. SMA always takes the average over the previous 12 months \((N = 12, \text{see the appendix})\) and the trend for DES is initialised at 0 after 12 months \((\hat{b}_{13} = 0, \text{see the appendix})\). For the bootstrapping methods, a warm-up period is not required. From month 25 onwards, 500 samples are taken from the previously observed supply lead times and demands. Notice that the bootstrapping methods are much more time-consuming than the traditional and Croston type methods.

For the traditional and Croston type forecasting methods (except SMA), we need to set one or two smoothing parameters, \(\alpha\) and \(\beta\). The parameters are usually set to values of at most 0.25 in the literature. We therefore choose the best value from the set \(\{0.05, 0.10, 0.15, 0.20, 0.25\}\) as follows. We take one value and start forecasting. In each month (from month 25 onwards), we calculate the MSE and at the end we calculate the average over these MSEs. We do so for all five possible values (25 possible combinations if two parameters are set) and we choose the value (or combination of values) that leads to the lowest average MSE. As a result of our approach, each part has its own combination of parameters for a given forecasting method. The way in which these parameters are determined is not possible in practice, but it does lead to the best performance of each method, which we prefer for the comparison between the methods. Initial tests show that when using the same parameter value for all parts, the performance is not as good as when using our approach. In practice, this may be a drawback since our method will function well only if there is enough historical data available per part to determine its optimal parameter value. But then again, if that is not available, forecasting for parts with infrequent demands will not function well anyway using time series analysis only.

We will also show the results of a method that we term the **MSE method**. In that method, we do not only choose the parameters based on the MSE; we also choose the best forecasting method (out of the traditional and Croston type methods) based on the MSE over the whole data set (not calculated in the first 24 months). It will turn out that this is not a good idea, as we show that SBA generally performs much better.

### 5.3 Inventory control

We use an \((R,nQ)\) inventory control policy with a daily review period and backordering. This means that if the inventory position \((\text{IP} = \text{stock on hand} + \text{outstanding orders} - \text{backorders})\) drops to \(R\) or below, an order is placed of size \(nQ\) \((n \in \mathbb{N})\) such that \(R < \text{IP} \leq R + Q\). If there

<table>
<thead>
<tr>
<th>Method</th>
<th>Abbreviation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple moving average</td>
<td>SMA</td>
<td>Axsäter (2006, pp.11–12) &amp; Appendix</td>
</tr>
<tr>
<td>Single exponential smoothing</td>
<td>SES</td>
<td>Axsäter (2006, pp.12–16) &amp; Appendix</td>
</tr>
<tr>
<td>Double exponential smoothing</td>
<td>DES</td>
<td>Axsäter (2006, pp.16–18) &amp; Appendix</td>
</tr>
<tr>
<td>Croston’s method</td>
<td>Croston</td>
<td>Croston (1972) &amp; Appendix</td>
</tr>
<tr>
<td>Syntetos Boylan approximation</td>
<td>SBA</td>
<td>Syntetos and Boylan (2005) &amp; Appendix</td>
</tr>
<tr>
<td>Empirical method (our implementation)</td>
<td>Empirical plus</td>
<td>Porras and Dekker (2008) &amp; Section 4</td>
</tr>
<tr>
<td>Mean squared error method</td>
<td>MSE method</td>
<td>Section 5.2</td>
</tr>
</tbody>
</table>
is a minimum order quantity specified (MOQ, meaning that the order size should be at least equal to MOQ), \( Q \) is equal to that. Otherwise, if there is a module quantity specified (MOD, meaning that the order size should be a multiple of MOD), \( Q \) is equal to that. \( Q = 1 \) if neither one is specified, which means that the inventory control is equivalent to an \((S - 1, S)\) base stock policy. NTH in principle specifies both the MOD and MOQ, whereas ALS specifies the MOQ only. Still, this does not mean that it is specified for all parts at these companies. At NMC, neither is specified. All in all, \( Q = 1 \) for 198 out of the 790 parts at NTH, for 680 out of the 748 parts at ALS, and for all parts at NMC.

\( R \) is updated at the beginning of each month, looking at past demands only. The empirical plus method does use all lead times, including future ones. This means that although the smoothing parameters are determined using all data, forecasts are made and inventory control is performed in a way that is possible in practice for the traditional and Croston type forecasting methods. Determination of \( R \) for the bootstrapping methods is done by initialising \( R \) at 0, increasing it one at a time, and each time using Equation (1) to determine if the resulting part fill rate is higher than or equal to the target part fill rate. We use the part fill rate because calculating the order line fill rate is not possible for the traditional and Croston type methods when using the fitted Normal distribution only, see below. The difference between the two types of fill rates is that if a demand for 10 parts is observed, whereas only 9 parts are available, the part fill rate is 90%, whereas the order line fill rate is 0%.

For the traditional and Croston type methods we use the same approach, but calculation of the part fill rate is done differently. The forecast methods give a forecast of the mean expected demand per period and we use the square root of the mean squared error (MSE) of the forecasts of the last twelve months as the standard deviation of the expected demand per period (which is quite common in the forecasting literature). We then follow the approach of Axsäter (2006, p.33 & p.85) to get an average and standard deviation of the expected demand over the lead time, to fit that mean and standard deviation to the Normal distribution, and to calculate the fill rate. We have chosen to use the Normal distribution (and not, for instance, the Poisson distribution) for two reasons. The first is that Porras and Dekker (2008) report that this distribution gives good results. The second is that the mean demand is forecasted, whereas it is assumed that the forecast error can be used as an estimator for the standard deviation. In line with many forecasting approaches, we assume that the error is normally distributed with mean equal to zero.

After the initialisation period (of 24 months), we set the starting stock at \( R + Q \). There are multiple ways to set the starting stock and for NTH we have also checked the use of \( R + \lfloor Q/2 \rfloor \): this does not change the overall picture, but both holding costs and service levels are lower on average for all methods. The holding costs are calculated over the stock on hand, using a holding cost rate of 20%. If \( R \) is increased at a review moment, we place an order if necessary; we do not sell stock if \( R \) is decreased.
6 Results

We first give results on the complete set of parts at each of the three companies. It turns out that the Syntetos Boylan approximation generally performs best, but the empirical plus method performs very good as well, albeit mainly at lower fill rates. Therefore, we then focus on the performance of the empirical plus method compared to the Syntetos Boylan approximation. In all cases, we compare the performance of the forecasting methods by establishing trade-off curves between inventory holding costs and achieved fill rates. This gives better insights than comparing at one achieved fill rate. Besides, it is very hard to get exactly the same achieved fill rate for all forecasting methods. Notice that because of their computational burden, we have used less different fill rates for the bootstrapping methods than for the other methods. Figures 3 to 5 show the total holding costs as a function of the achieved order line fill rates at each of the three companies. Interesting here is that:

- At all companies, the traditional methods perform worst, followed by the MSE method. Although it may be surprising that the MSE method does not perform that well, we are not the first to notice that choosing a forecasting method based on its MSE does not necessarily lead to good results. As mentioned in Section 2, Teunter and Duncan (2009) find that the zero forecast outperforms other forecasting methods in many cases, when comparing using the MSE. That is exactly the reason why they propose to (and do) compare forecasting methods using the resulting inventory control.

- At all companies, the performance of SBA is generally better than, but very close to that of Croston’s method. Similarly, the empirical plus method generally performs slightly better
than the empirical method. In fact, the performance is quite similar, but the empirical plus method is usually able to achieve a higher service level than the empirical method.

- At ALS and NMC, the bootstrapping methods perform best for fill rates that are not too high. They do not achieve the really high fill rates and thus the Croston type methods perform better there. The reason for not achieving such fill rates is that the bootstrapping methods give a zero probability to seeing a higher lead time demand than the highest historical lead time demand (the empirical plus method may in fact sometimes give a positive, but very low probability).

- At NTH, the bootstrapping methods perform badly. Unfortunately, we are not able to explain this.

Since SBA and Croston perform similarly, just as empirical plus and empirical, and these methods perform much better than the other methods, we will focus on SBA and empirical plus only from here onwards.

We further see (not shown in a figure) that the optimal smoothing parameters vary a lot and they are in more than 90% of the times at either the lowest or highest possible value (0.05 and 0.25). This holds for both smoothing parameters of SBA at each of the companies.

Figures 6 and 7 show both the achieved part fill rate and the achieved order line fill rate, as a function of the target part fill rate. There are a few interesting things to notice:

- At all companies, the achieved part fill rate is only slightly lower than the achieved order line fill rate. This is due to the fact that demand sizes are generally small. From here onwards, it is not interesting to show both achieved fill rates. Since for companies the order line fill rate is most relevant, we will show that only.
Figure 5: Holding costs versus achieved order line fill rates at the Naval Maintenance Company

Figure 6: Achieved versus target fill rates

- At NTH and NMC, the achieved fill rates are higher than the target for low targets. For higher targets, which are the relevant targets in general, the achieved fill rate is below the target.

- At ALS, the achieved fill rates are (almost) always below the target. As mentioned in Section 2, Teunter and Duncan (2009) also notice that the achieved fill rate is often below its target and they propose a solution to this problem. However, this does not explain the difference that we see between NTH and NMC on the one hand and ALS on the other hand. We think that this can be explained by the fact that at ALS there are many parts with one very high demand. If this exceptional demand occurs in the first two years, then too much inventory will be stocked; if the exceptional demand occurs later, it is not expected and it causes a low fill rate.
We next focus on classification of parts such that using a different forecasting method for each of the classes leads to better results than using one forecasting method for all parts. As mentioned in Section 2, Boylan et al. (2008) propose to use the average inter-demand interval, $p$, and the squared coefficient of variation of the demand sizes, $CV^2$, to classify parts. We find that these two characteristics of parts are the most important ones indeed. We have also looked at the average lead time, the average demand size, and the price. For each of the five characteristics, we have divided per company the whole set of parts into three groups of equal size (same number of parts), with one group scoring highest on the characteristic, one group scoring lowest, and a middle group. We have then plotted for each group at each company the total holding costs as a function of the achieved order line fill rates, similar to Figures 3 to 5. By visually comparing per company and per characteristic the figure for the group ‘highest’ and ‘lowest’, we have checked whether or not the characteristic influences the results. As mentioned, this was the case only for $p$ and $CV^2$. We find that for parts with a high $p$ and a low $CV^2$, it is best to use the empirical plus method, when the fill rate is not too high. If the fill rate is higher, and for all other parts, it is best to use SBA. The specific class of parts for which the empirical plus method performs best consists often of the expensive parts, for which forecasting is both difficult, because of the infrequent demands, and important, because of the price. This high price also means that the target fill rate is often not too high.

Figure 8 gives an example of a part on which the empirical plus method performs better than SBA. Because of the high inter-demand intervals, SBA uses such a low reorder point, that if a demand occurs, chances are high that SBA has too little stock on hand. Furthermore, because the mean squared error (MSE) is quite low, while it is used to fit the Normal distribution, the reorder point is even further reduced from 2 to 1.

For all three companies, we show the difference in performance of empirical plus and SBA when using cut-off points $p = 4$ (months) and $CV^2 = 0.3$ in Figures 9 to 11 (At NTH, only a few parts have $p > 4$ and $CV^2 < 0.3$, making the figure ‘unsmooth’). It can be seen that indeed empirical plus performs better for the class of parts with $p > 4$ and $CV^2 < 0.3$ when fill rates are not too high. It is difficult to determine the optimal cut-off values; results do not change a lot when the cut-off values are changed a little, which Boylan et al. (2008) also find (see Section 2). We have therefore not focussed on determining such optimal cut-off values.
Figure 8: Example of a part on which the empirical plus method performs best

(a) All parts with $p > 4$ and $CV^2 < 0.3$

(b) All parts with $p < 4$ and $CV^2 > 0.3$

Figure 9: Holding costs versus achieved order line fill rates at NedTrain Haarlem

7 Conclusions and recommendations

In this paper we have compared several methods to determine re-order points for spare parts with infrequent demands: three traditional forecasting methods (simple moving average, single exponential smoothing, and double exponential smoothing), two Croston type methods (Croston’s method and the Syntetos Boylan approximation), the MSE method, and two bootstrapping methods. The MSE method picks for each part the traditional or Croston type forecasting method that leads to the lowest MSE. The first bootstrapping method, the empirical method, is a slightly adjusted version of the method proposed by Porras and Dekker (2008), and we

(a) All parts with $p > 4$ and $CV^2 < 0.3$

(b) All parts with $p < 4$ and $CV^2 > 0.3$

Figure 10: Holding costs versus achieved order line fill rates at Alstom Ridderkerk
have proposed the second bootstrapping method, the empirical plus method, which explicitly incorporates the stochasticity of the supply lead time.

The data that we used for comparison originates from three companies, a train maintenance company, a company building and overhauling electric train propulsion systems, and a naval maintenance center. In total, we considered some 6,000 parts with 5 years of demand history (after filtering). The traditional and Croston type forecasting methods produce estimates of the mean and variance of the future demand, which are fitted to a Normal distribution that is then used to calculate reorder points. The bootstrapping methods directly produce distributions of the lead time demand, from which the reorder points are then calculated. Using and updating these reorder points, we simulated the inventory control using the real demand data. This was done for several fill rate targets, resulting in trade-off curves of inventory holding costs and achieved fill rates.

The MSE method performed generally worse than the Croston type methods, indicating that choosing a forecasting method based on its resulting MSE may not lead to good inventory control. The Syntetos Boylan approximation (SBA) performed best among the traditional and Croston type forecasting methods: in most cases it had the lowest cost for a certain fill rate. The empirical plus method performed somewhat better than the empirical method, so we next focussed on a comparison between SBA and the empirical plus method. The results were mixed. When considering individual parts, it appeared that the empirical plus method outperformed the SBA method in case of long times between demands and a low variability in demand sizes. This specific class of parts consists often of the expensive parts, for which forecasting is both difficult, because of the infrequent demands, and important, because of the price. The empirical plus method may therefore be very helpful in practice; it may lead to large cost reductions. The SBA performed better in the opposite cases.

There are a few limitations that we need to mention. Our focus was on comparing forecasting methods from different categories, not on comparing all possible forecasting methods. This means that there may be other methods that perform even better. Teunter et al. (2011) have proposed a new Croston type method, which may perform better than SBA. This may mean that the performance of the best Croston type method could be improved. On the other hand, the empirical plus method may also be improved, since it had difficulties obtaining high fill
rates. Although this may not be a problem for the category of parts for which the method performed best, because these parts are expensive, it may still be interesting to try and improve on this issue. It may be possible to combine the idea of sampling from the historical lead times, as we do in the empirical plus method, with the idea of ‘jittering’ that is key to the method of Willemain et al. (2004). Further, for both types of methods the ideas of Teunter and Duncan (2009) may be incorporated. They propose a modification that takes into account the fact that an order in a period is triggered by a demand in that period.

Taking into account other information into a forecasting method may be very beneficial as well. This could be information on the changes in the size of the installed based, condition monitoring information, reliability data, or information on maintenance planning. It is hard to get this data in practice, which makes an empirical study difficult. However, showing the value that such information may have, may induce companies to collect this data and make it available. Dekker et al. (2011) discuss the usage of information on the size of the installed based and also the practical problems of obtaining such information. We are further aware of some first attempts by, e.g., Lin et al. (2012) to use condition (degradation) information to improve spare parts supply, but without combining this with time series analysis based forecasting, and, e.g., Hua et al. (2007) and Romeijnders et al. (2012) to use maintenance information in a forecasting method. Still, we think that there is a lot to gain here.

Acknowledgements

The authors would like to thank Joost van Gennip of NedTrain Haarlem, Robert van Twillert of Alstom Ridderkerk, and Hennie de Vries of the Naval Maintenance Company for their support and for allowing us to use and publish some of their companies’ data. The authors further greatly appreciate the help of Gordian’s consultants, e.g., with extracting and cleaning of data. Specifically: Maarten Driessen, Marcel Prakken, Ruud Speekenbrink, and Stijn Wouters. Finally, the first author gratefully acknowledges the support of The Lloyd’s Register Educational Trust, an independent charity working to achieve advances in transportation, science, engineering and technology education, training and research worldwide for the benefit of all. This paper is based on the Master’s thesis of the second author (Van Wingerden, 2011).

A The traditional and Croston type forecasting methods

When forecasting demand, there are two basic models: the constant model that assumes that the expected demand per period does not change, and the trend model that assumes that there is some trend over time. If we consider periods \( t \in \mathbb{N} \) and denote by \( x_t \) the observed demand in period \( t \), then the constant model assumes that there is a constant \( a \) and an error \( \epsilon_t \) such that \( x_t = a + \epsilon_t \), whereas the trend model further assumes a trend \( b \) so that \( x_t = a + bt + \epsilon_t \). We then define \( \hat{a}_t \) to be the estimate of \( a \) after observing the demand in period \( t \), \( \hat{b}_t \) to be the estimate of \( b \) after observing the demand in period \( t \), and \( \hat{x}_{t,t+k} \) to be the forecast of the demand in period
after observing the demand in period $t$. We further define $\alpha$ and $\beta$ to be two
smoothing constants with $0 \leq \alpha, \beta \leq 1$.

The traditional forecasting methods then forecast as follows.

- Simple moving average (SMA): \( \hat{x}_{t,t+k} = \hat{a}_t = 1/N \sum_{n=t-N+1}^{t} x_n \), with $N$ being the number
  of periods over which an average is taken and $t \geq N$.

- Single exponential smoothing (SES): \( \hat{x}_{t,t+k} = \hat{a}_t = (1-\alpha)\hat{a}_{t-1} + \alpha x_t \) ($t > 1$).

- Double exponential smoothing (DES): For $t > 1$, \( \hat{a}_t = (1-\alpha)(\hat{a}_{t-1} + \hat{b}_{t-1}) + \alpha x_t \) and
  \( \hat{b}_t = (1-\beta)\hat{b}_{t-1} + \beta(\hat{a}_t - \hat{a}_{t-1}) \). Then, \( \hat{x}_{t,t+k} = \hat{a}_t + k\hat{b}_t \).

Initialisation of each of the methods above and below is discussed in Section 5.

For the Croston type methods, we define the following additional variables, all defined as
after observing the demand in period $t$: \( k_t \) as the interval since the previous positive demand
\( (k_t \geq 1) \), \( \hat{k}_t \) as the estimate of the average inter-demand interval \( (\hat{k}_t \geq 1) \), and \( \hat{d}_t \) as the estimate
of the average size of a positive demand \( (\hat{d}_t > 0) \). If \( x_t = 0 \), then \( \hat{k}_t = \hat{k}_{t-1} \) and \( \hat{d}_t = \hat{d}_{t-1} - 1 \).
Otherwise, \( \hat{k}_t = (1-\alpha)\hat{k}_{t-1} + \alpha k_t \) and \( \hat{d}_t = (1-\beta)\hat{d}_{t-1} + \beta x_t \). For Croston’s method, \( \hat{x}_{t,t+k} = \hat{a}_t = \hat{d}_t/\hat{k}_t \), while for the Syntetos Boylan approximation (SBA), \( \hat{x}_{t,t+k} = \hat{a}_t = (1-\alpha/2)\hat{d}_t/\hat{k}_t \).

B Characteristics of the three case companies

B.1 NedTrain Haarlem

#### Figure 12: Prices of parts (in euros) at NedTrain Haarlem
Figure 13: Planned lead times of parts (in working days) at NedTrain Haarlem

Figure 14: Planned - realised lead times (in calendar days) at NedTrain Haarlem

Figure 15: Number of demands per part at NedTrain Haarlem
Figure 16: Demand sizes at NedTrain Haarlem

Figure 17: Relation between the price (in euros) and the planned lead time (in working days) at NedTrain Haarlem
B.2 Alstom Ridderkerk

Figure 18: Prices of parts (in euros) at Alstom Ridderkerk

Figure 19: Planned lead times of parts (in calendar days) at Alstom Ridderkerk

Figure 20: Planned - realised lead times (in calendar days) at Alstom Ridderkerk
Figure 21: Number of demands per part at Alstom Ridderkerk

Figure 22: Demand sizes at Alstom Ridderkerk
B.3 Naval Maintenance Company

Figure 23: Prices of parts (in euros) at the Naval Maintenance Company

Figure 24: Planned lead times of parts (in months) at the Naval Maintenance Company

Figure 25: Planned - realised lead times (in calendar days) at the Naval Maintenance Company
Figure 26: Number of demands per part at the Naval Maintenance Company

Figure 27: Demand sizes at the Naval Maintenance Company
References


