European Journal of Operational Research 261 (2017) 169–181

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Production, Manufacturing and Logistics

An improved method for forecasting spare parts demand using extreme value theory

Sha Zhu[a,b,*], Rommert Dekker[a], Willem van Jaarsveld[c], Rex Wang Renjie[a,b], Alex J. Koning[a]

[a] Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, Rotterdam 3000 DR, The Netherlands
[b] Erasmus Research Institute of Management, P.O. Box 1738, Rotterdam 3000 DR, The Netherlands
[c] Industrial Engineering & Innovation Sciences, Technische Universiteit Eindhoven, P.O. Box 513, Eindhoven 5600 MB, The Netherlands

ARTICLE INFO

Article history:
Received 24 November 2015
Accepted 31 January 2017
Available online 5 February 2017

Keywords:
Forecasting
Inventory
Spare parts
Extreme value theory
Semi-parametric

ABSTRACT

Inventory control for spare parts is essential for many organizations due to the trade-off between preventing high holding cost and stockouts. The lead time demand distribution plays a central role in inventory control. The estimation of this distribution is problematic as the spare part demand is often intermittent, and as a consequence often only a limited number of non-zero data points are available in practice. The well-known empirical method uses historical demand data to construct the lead time demand distribution. Although it performs reasonably well when service requirements are relatively low, it has difficulties in achieving high target service levels. In this paper, we improve the empirical method by applying extreme value theory to model the tail of the lead time demand distribution. To make the most out of a limited number of demand observations, we establish that extreme value theory can be applied to lead time demand periods computed over overlapping intervals. We consider two service levels: the expected waiting time and cycle service level. Our experiments show that our method improves the inventory performance compared to the empirical method and is competitive with the WSS method, Croston's method and SBA for a range of demand distributions.

1. Introduction

The supply of aftermarket parts is an important source of profit for companies that sell durable equipment, see Gallagher, Mitchke, and Rogers (2005). Findings from Deloitte's Global Service and Parts Management Benchmark Survey show that in 2006 the service business accounted for an average of nearly 26% of revenues across the industries, see Koudal (2006). After-sales networks operate in an unpredictable marketplace because demands for repairs crop up intermittently, see Kennedy, Patterson, and Fredendall (2002), Cohen, Agrawal, and Agrawal (2006), Syntetos, Babai, and Altay (2012).

An essential element in spare parts inventory control is the forecasting of the lead time demand, as the lead time is the period in which a stockout may occur when demand is larger than foreseen. Differences in the monetary values of the stockholdings between lead time demand forecasting methods can be substantial; Eaves and Kingsman (2004) report a case in which the use of an inferior forecasting method leads to an additional investment of 13.6% of the total value of the inventory. Unfortunately, estimating the lead time demand distribution is especially difficult for slow-moving spare part types as typically only limited positive demand data points are available in practice.

The demand distribution may be estimated either parametrically or nonparametrically. Parametric methods have the advantage of being relatively simple while still showing decent empirical performance (Syntetos, Babai, & Gardner, 2015). However, as parametric estimators are derived from assumptions, they may turn out to be severely biased in case these assumptions do not hold. Therefore, nonparametric estimators are preferred, as the traditional parametric estimators have problems dealing with intermittent demand and particular patterns. Popular nonparametric approaches are the bootstrap method, see Willemain, Smart, and Schwarz (2004), which we refer to as the WSS method in this paper, and the empirical method, see Porras and Dekker (2008) and van Wingerden, Basten, Dekker, and Rustenburg (2014). Nonparametric estimators only provide relevant information for demand levels in the scope of the historical demand data of say N positive data points, and basically break down in case of extrapolation beyond this scope. In particular, the largest data point would be

* Corresponding author at: Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.

E-mail addresses: zhu@ese.eur.nl, zhushashasha@gmail.com (S. Zhu), rdekker@ese.eur.nl (R. Dekker), W.L.v.Jaarsveld@tue.nl (W. van Jaarsveld), rwang@ese.eur.nl (R.W. Renjie), koning@ese.eur.nl (A.J. Koning).

http://dx.doi.org/10.1016/j.ejor.2017.01.053
0377-2217/

expected to lie at the $N/(N+1)$th percentile, and thus achieving service levels beyond this percentile may prove difficult using empirical methods. Thus, we resort to semi-parametric estimators in the tail for high service levels.

In this paper, we propose the empirical-EVT method, which applies extreme value theory (EVT, see Coles (2001), Beirlant, Goegebeur, Segers, and Teugels (2004), Reiss and Thomas (2007)) to the tail part of the distribution and handles the remainder of the distribution (the non-tail part, say) via the empirical method. The empirical distribution is used as starting point to ensure that the structure of the non-extreme part of the data is preserved. Application of EVT allows us to closely approximate the tail of a distribution using one single parameter, the extreme value index, see de Haan and Ferreira (2006). Only the largest historical demands are used to estimate the extreme value index, the other historical demands are input to the empirical method. The new method inherits the advantage of non-parametric approaches without losing the ability to achieve high service levels. As EVT allows dependence (more precisely, β-mixing dependence) between successive lead time demands, we establish that the empirical-EVT method may be applied to lead time demands computed over *overlapping* time periods. With this result, our method makes the most out of a limited demand history.

A simulation study is conducted to assess the performance of the empirical-EVT where we estimate the lead time demand distributions based on samples from a known distribution. Instead of one period forecasting error measures such as mean squared error, we employ service levels to evaluate performance, as advocated in Syntetos and Boylan (2006) and Teunter and Duncan (2009).

The experiments show that the new method improves the inventory performance under the expected waiting time and cycle service level for a range of demand generating processes and service targets, when compared with the empirical method. Moreover, the new method is competitive with methods such as WSS, Croston's method (1972), and the Syntetos-Boylan approximation (SBA) (2005), in the sense that it sometimes outperforms these methods, and sometimes is outperformed by these methods, depending on the demand data and other parameters.

We use in our empirical study the automotive dataset described by Syntetos and Boylan (2005) and the aircraft aircraft component repair dataset by Romeijnders, Teunter, and van Jaarsveld (2012). Both Croston's method and WSS perform well in the automotive dataset, followed by empirical-EVT. The limited training data leads to the unsatisfactory performance of empirical-EVT since it is difficult to estimate the tail of lead time demand distribution based on 13 periods observations. However, Empirical-EVT method performs best in the aircraft component repair dataset where demand is available in 84 periods.

The paper is organized as follows. Section 2 gives an overview of the relevant literature. Section 3 briefly describes EVT theory and how to use it in our study. In Section 4 a simulation study and an empirical study give insights into the differences between the empirical-EVT method, the empirical method, WSS, Croston's method and SBA. The last section presents the final conclusions.

2. Literature

In Section 2.1, we review forecasting methods for slow-moving items. In Section 2.2, we review extreme value theory.

2.1. Intermittent demand forecasting

Demand forecasting is a key issue in the field of spare parts management. For an overview on spare parts demand forecasting research, we refer to Boylan and Syntetos (2010).

Traditional forecasting methods such as simple moving average (SMA) and simple exponential smoothing (SES) fail to perform well for intermittent demand, see Syntetos and Boylan (2005). Croston's method (CR, see Croston (1972)) isolates periods with positive demands, is "robustly superior" to SES, see Willemain, Smart, Shockor, and DeSautels (1994), and is biased, see Syntetos and Boylan (2001). The Syntetos-Boylan approximation (SBA, see Syntetos & Boylan (2005)), the Syntetos method (SY, see Syntetos (2001)) and the Teunter-Syntetos-Babai method (TSB, see Teunter, Syntetos, & Babai (2011)) are bias-corrected modifications of the CR method. According to Syntetos and Boylan (2006), the SBA method outperforms the SMA, SES and CR methods. Teunter and Sani (2009) prefers the SY over the SB method, as the latter actually overcompensates the bias. Evidence in Babai, Syntetos, and Teunter (2014) suggests that TSB does not outperform CR, SBA and SY unless the degree of intermittence is low and demand is decreasing. For other modified CR methods, see Johnston and Boylan (1996), Snyder (2002), Shale, Boylan, and Johnston (2006). The variance of SES, CR, SY and SBA intermittent demand estimates are discussed in Syntetos and Boylan (2010). Though these methods have been widely used, they have the disadvantage of assuming a particular parametric structure of the demand distributions.

Bootstrapping is a non-parametric resampling technique, which builds the lead time demand distribution by repeated sampling from observations. The WSS modified bootstrapping method, see Willemain et al. (2004), resamples from past data using a Markov chain approach to switch between no demand and demand periods. Teunter and Duncan (2009) finds that bootstrapping performs equally well as the CR and SBA method, but is more difficult to implement. Syntetos et al. (2015) concludes that the WSS modified bootstrapping method does have advantages over the SES, CR and SBA methods, but questions whether WSS is worth the added complexity.

The empirical method, proposed in Porras and Dekker (2008), is a far less complex non-parametric method which uses the empirical cumulative distribution function to estimate the lead time demand distribution for fixed lead times. The empirical method was slightly extended in van Wingerden et al. (2014) so as to cover variable lead times as well. As the empirical cumulative distribution function only provides information for demand levels in the scope of the historical demand data, the empirical method basically breaks down for high service levels. Syntetos et al. (2015) mentions poor performance of the empirical method.

Other forecasting methods supplement historical demand data with additional information. The use of installed base information is discussed in Jalil, Zuidwijk, Fleischmann, van Nunen and Jo (2011) and Dekker, Pinçe, Zuidwijk, and Jalil (2013). Information on component repairs is first considered in Romeijnders et al. (2012). Topan et al. (2016) assesses the value of the imperfect demand information and proposes a lost-sales inventory model with a general representation of demand information to find the ordering policy minimizing total inventory holding, shortage and ordering cost under imperfect information.

2.2. Extreme value theory

Extreme value theory is a branch of statistics modeling the tail behavior of a distribution. The most prominent application is the estimation of an unknown upper quantile value corresponding to a small given exceedance probability. However, EVT also covers the estimation of an unknown exceedance probability. The determination of a safe height for the North Sea dikes in the Netherlands acted as an important driver behind the development of EVT, see de Haan (1990). Nowadays, EVT is widely used in financial risk management to estimate downward risk measures such as value at risk and expected shortfall. To our knowledge, the only previous application of EVT to inventory problems is in Kogan and Rind (2011), where the design of an inventory for critical equipment is considered. As critical equipment is characterized by *infinitely large* underage costs, they require that no stockout occurs in the coming years with high probability. Based on this requirement, they develop rules for determining inventory based on EVT, homogeneous Poisson processes, and Chebyshev's inequality.

3. Theory

3.1. Extreme value theory

Loosely speaking, EVT theory is built upon the idea that the tail behavior of many uncertain quantities that are encountered in practice can be modeled using the Generalized Pareto Distribution (GPD). We explain EVT in more detail in Sections 3.1.1 and 3.1.2, 3.1.3.

3.1.1. Tail approximation

Let X be a random variable with cumulative distribution function $F(x) = P\{X \le x\}$, and let $x^* = \sup\{x : F(x) < 1\}$ denote the endpoint of the support of F. Note that x^* may be either finite or infinite.

Throughout this paper, we shall assume the existence of a positive function f such that

$$\lim_{\tau \uparrow x^*} \frac{1 - F(\tau + x f(\tau))}{1 - F(\tau)} = (1 + \gamma x)^{-1/\gamma} \tag{1}$$

for all x for which $1 + \gamma x > 0$, see condition 4 of Theorem 1.1.6 at p. 10 in de Haan and Ferreira (2006). The assumption allows the approximation of the tail of the distribution of X for a sufficiently large threshold τ,

$$1 - F(x) \approx (1 - F(\tau))\left\{1 - H_\gamma\left(\frac{x - \tau}{f(\tau)}\right)\right\} \tag{2}$$

for all $x > \tau$, see p. 67 in de Haan and Ferreira (2006). Here H_γ denotes the cumulative distribution function of the GPD:

$$H_\gamma(x) = \begin{cases} 1 - (1 + \gamma x)^{-1/\gamma} & \text{for } \gamma \neq 0, \\ 1 - e^{-x} & \text{for } \gamma = 0. \end{cases} \tag{3}$$

The parameter γ of the GPD plays a central role in EVT, and hence is referred to as the extreme value index; it acts as a shape parameter of the GPD approximating the tail of the distribution. The support of the GPD is $[0, \infty)$ if γ is non-negative, and $[0, -1/\gamma]$ if γ is negative. For any distribution such that (1) holds, we say that the distribution belongs to the domain of attraction of $H_\gamma(x)$.

Assumption (1) is not restrictive because it only pertains to the tail behavior of the distribution. In that sense, it is very much weaker than (for example) stating that demand follows a negative binomial distribution, or a normal distribution, or any other specific distribution, precisely because such an assumption specifies the entire distribution.

Assumption (1) is satisfied by a very wide range of continuous distributions (Balkema & de Haan, 1974; Pickands, 1975). This reflects that tail behavior of many distributions allow approximate modeling by means of the GPD: this explains why the GDP occurs in the assumption. This generality is in fact one of the strongest points of EVT, and the main reason why it has been applied to a very wide range of problems: extreme sea-levels (for dike height determination), insurance losses, market risk, environmental loads on structures, etc. Moreover, according to Shimura (2012), many discrete distributions can be regarded as a discretization of a continuous distribution satisfying (1), and hence satisfy (1) themselves.

An interesting approach to test the applicability of EVT for a specific real-life scenario in which the demand distribution is unknown, would be a goodness-of-fit test. Unfortunately, goodness-of-fit in EVT is not yet fully developed, see Section 2.3 in Beirlant, Caeiro, and Gomes (2012). An Anderson-Darling type test of (1) based on the tail empirical process is proposed in Drees, de Haan, and Li (2006). Several tests of (1) for $\gamma > 0$ are found in Koning and Peng (2008).

3.1.2. The basic model

The basic model on which EVT in its original form rests, assumes that $X_1, X_2, \ldots, X_n$ is a random sample of size n drawn from the distribution given by F; in other words, the observations $X_1, X_2 \ldots, X_n$ are *independent* copies of the random variable X. Let $X_{(1)} \le X_{(2)} \le, \ldots \le X_{(n)}$ be the ordered sample. Choose $0 \le k < n$, and use $X_{(n-k)}$ as an empirical threshold. That is, replace τ by $X_{(n-k)}$. To allow the derivation of asymptotic results for n tending to infinity, we assume that the choice k depends on n, and satisfies $k \to \infty$ and $k/n \to 0$ as $n \to \infty$. That is, as n grows large, k grows large but nevertheless vanishes relative to n.

Since $1 - F(X_{(n-k)}) \approx k/n$, we obtain, see Equation (3.1.4) in de Haan and Ferreira (2006, p. 68),

$$1 - F(x) \approx \frac{k}{n}\left\{1 - H_\gamma\left(\frac{x - X_{(n-k)}}{\alpha}\right)\right\} \tag{4}$$

for all $x > X_{(n-k)}$, with $\alpha = f(X_{(n-k)})$. In particular, this approximation continues to hold even beyond the sample maximum $X_{(n)}$.

In recent EVT literature $\alpha = f(X_{(n-k)})$ is usually replaced by $\alpha = a(k/n)$. We also do this in our study. The functions f and a are related through the equation $f(x) = a(1/(1 - F(x)))$, see Theorem 1.1.6 at p. 11 in de Haan and Ferreira (2006); note that $1/(1 - F(X_{(n-k)})) \approx n/k$. As the parameters γ and α appearing in (4) are unknown, we simply replace them by estimators. We use the moment estimators proposed in Dekkers, Einmahl, and de Haan (1989). Summarize the k largest order statistics via the first two "moments" $M_n^{(1)}$ and $M_n^{(2)}$ defined by

$$M_n^{(j)} := \frac{1}{k}\sum_{i=0}^{k-1}\left(\log X_{(n-i)} - \log X_{(n-k)}\right)^j \tag{5}$$

for $j = 1, 2$. Then, the moment estimators of the extreme value index γ and the scale α are given by

$$\hat{\gamma} := M_n^{(1)} + 1 - \frac{1}{2}\left(1 - \left(M_n^{(1)}\right)^2\left(M_n^{(2)}\right)^{-1}\right)^{-1}, \tag{6}$$

$$\hat{\alpha} := \tfrac{1}{2}X_{(n-k)}M_n^{(1)}\left(1 - \left(M_n^{(1)}\right)^2\left(M_n^{(2)}\right)^{-1}\right)^{-1}. \tag{7}$$

Until now, we have only assumed that the number $k = k(n)$ of used largest order statistics satisfies $k \to \infty$ and $k/n \to 0$ as $n \to \infty$. In practice, the choice of k is made by trading off bias and variance. For small k, only a limited amount of the information contained in the data is used, and hence the variance of the moment estimator is relatively large. Although selecting a larger value of k will reduce this variance, it is typical that the bias of the moment estimator will increase at the same time as the relevance of (1) diminishes. (See de Haan and Ferreira (2006).)

Three different ways of choosing k have appeared in literature: moment estimator plot (Hill, 1975), bootstrap method (Danielsson, de Haan, Peng, & de Vries, 2001); (Draisma, de Haan, Peng, & Pereira, 1999) and unbiased moment estimator plot (HaanMercadierZhou2014). For the purpose of applying EVT to inventory control, we use an automated method for threshold selection, see Appendix D in the supplementary material.

3.1.3. Relaxing the independence assumption

The basic model assumed in the previous paragraph required that $X_1, X_2, \dots$ is a sequence of independent and identically distributed (i.i.d.) random variables. In practice, the independence assumption may turn out to be too restrictive. As we intend to apply EVT to the lead time demands directly, the independence assumption also becomes an issue in our approach to LTD estimation. We may view lead time demands as "moving sums" of demands in subsequent time intervals (for instance, days, weeks, months or year). Let D_j denote the demand in time interval j. If the window size L is fixed, then we may express the demands over window size as

$$X_i^{[L]} = \sum_{j=i}^{i+L-1} D_j, \quad \text{for } i = 1, 2, \dots. \tag{8}$$

Besides constant L, multiple levels of aggregation are also applicable to our forecasting method. Various aggregation window sizes may lead to different inventory performance, see Rostami-Tabar, Babai, Syntetos, and Ducq (2013), Rostami-Tabar, Babai, Syntetos, and Ducq (2014), Petropoulos and Kourentzes (2015). For a discussion of the optimal choice of aggregation window size, see Nikolopoulos, Syntetos, Boylan, Petropoulos, and Assimakopoulos (2011). For the present paper, we choose to restrict the window size to the leadtime L. We believe combining the empirical or empirical-evt approach approach with a temporal aggregation approach may be an interesting topic for further research. Whenever allowed, we shall drop the superscript $[L]$, and use the short hand notation X_i rather than the full one $X_i^{[L]}$.

Typically, the demands $D_1, D_2, \dots$ are assumed to be i.i.d., see Croston (1972) for instance. The consecutive lead time demands $X_1, X_2, \dots$ become dependent. In fact, $X_1/L, X_2/L, \dots$ is a moving average process of order L, that is, an ARMA(0, L) process. To avoid this dependence, one has resorted to considering "non-overlapping" lead time demands, $X_1, X_{L+1}, X_{2L+1}, \dots$ say, see Nikolopoulos et al. (2011). (In the context of time series analysis, the construction of non-overlapping lead time demands $X_1, X_{L+1}, X_{2L+1}, \dots$ is referred to as "temporal aggregation".)

Fortunately, the independence assumption has been relaxed in EVT, see Resnick and Starica (1998), Drees (2003), Rootzén (2009). In essence, we should require that the sequence $X_1, X_2, \dots$ is β-mixing instead. The β-mixing dependence condition (also known as absolute regularity or weak Bernoulli condition) was proposed in Volkonskiĭ and Rozanov (1959), and is thoroughly discussed in Bradley (2005). Loosely speaking, β-mixing precludes long range dependence. Many random sequences in practice – among which Harris chains, ARMA, ARCH and GARCH processes – are β-mixing, see Athreya and Pantula (1986), Mokkadem (1988), Carrasco and Chen (2002), Fryzlewicz and Rao (2011). Since consecutive lead time demands $X_1, X_2, \dots$ constructed from i.i.d. demands via (8) behave as an ARMA(0, L) process up rescaled by a fixed factor L, it follows that these lead time demands are indeed β-mixing (the rescaling does not affect the dependence structure).

In addition, we may relax the independence between the demands, as long as we end up with β-mixing lead time demands. For example, one may show that the moving sum of an ARMA process is also an ARMA process, see Granger and Morris (1976); thus, if the demands are not independent but form an ARMA process instead, the lead time demands are still β-mixing. Recall that an ARMA process is stationary.

Autocorrelated demands have been found in intermittent industrial datasets, see Willemain et al. (1994). The simulation experiment in Altay, Litteral, and Rudisill (2012) shows that the forecast accuracy and stock control performance of the SES and SBA methods are vulnerable to autocorrelated demands. The theoretical work on relaxing the independence assumption in EVT suggests that EVT-based methods are to some extent robust with respect to stationary autocorrelated demands. As EVT-based methods implicitly assume that the extreme value index stays constant over time, we do not expect EVT-based methods to work well for non-stationary demand.

After using ad-hoc arguments to conclude that the assumption of stochastic independence of lead time demands "looks highly plausible as a first approximation", results in Kogan and Rind (2011) are derived under the basic EVT model. The validity of these ad-hoc arguments is difficult to assess. Our discussion above shows that it possible to avoid ad-hoc arguments by using a comprehensive and rigorous theoretical argument. In fact, we believe that relaxing the independence assumption is essential for any application of EVT to inventory control.

3.2. How to apply EVT based on the empirical LTD forecasting method

In this subsection we detail the application of the empirical-EVT method using three different service levels: expected waiting time $\mathbb{E}$WT, fill rate β and cycle service level CSL. These three service levels have in common that a larger base stock level yields a better service level (a smaller expected waiting time, a larger fill rate or a larger cycle service level). Thus, $\mathbb{E}$WT, β and CSL are in fact monotonic functions $\mathbb{E}\mathrm{WT}(S)$, $\beta(S)$ and $\mathrm{CSL}(S)$ of the base stock level S.

Our aim is to determine the smallest base stock level $S_{\min}$ such that the corresponding performance is at least as good as some given critical service level. However, we are unable to achieve this aim since the exact relation between base stock level and service level is unknown. The best we can do is to estimate this relation using historical demand data. As a consequence, we arrive at an estimated smallest base stock level $\hat{S}_{\min}$ rather than $S_{\min}$ itself.

The empirical-EVT method deals with two regions: the non-tail and the tail regions, separated by an unknown threshold τ which is estimated by the empirical threshold $X_{(n-k)}$. We handle the non-tail region non-parametrically using the empirical method, and the tail region semi-parametrically by EVT. The computation of $\hat{S}_{\min}$ involves the following steps.

Step 1: obtain the LTD sample Obtain a sample $X_1, X_2, \dots, X_n$ of lead time demands. Typically, these lead time demands are obtained by summing demands over given time periods, as in (8); let $\bar{D} = n^{-1}\sum_{i=1}^{n} D_i$ denote the mean demand during the data collection period.
However, we do leave open the possibility that the sample $X_1, X_2, \dots, X_n$ was obtained by some other data generating process (DGP), as long as EVT is still applicable; that is, the DGP should yield lead time demands which are β-mixing. (See Bradley (2005) for β-mixing).

Step 2: construct the ordered sample Sort the sample $X_1, X_2, \dots, X_n$ in ascending order. This yields the ordered sample $X_{(1)} \le X_{(2)} \le X_{(3)} \le \dots \le X_{(n)}$. We shall refer to $X_{(i)}$ as the ith order statistic. Remark that the information contained in the ordered sample is sufficient to construct the empirical distribution function $\hat{F}_n(x)$ appearing in paragraph 4.1.2 of Porras and Dekker (2008), because we may write

$$\hat{F}_n(x) = \sum_{i=1}^{n} 1_{\{X_i \le x\}} = \sum_{i=1}^{n} 1_{\{X_{(i)} \le x\}} \quad \text{for all } x \in \mathbb{R}, \tag{9}$$

Moreover, as the empirical distribution function is a step function which jumps at the order statistics, we may reconstruct the ordered sample from the empirical distribution function. Thus, the ordered sample $X_{(1)} \le X_{(2)} \le X_{(3)} \le \dots \le X_{(n)}$ carries exactly the same information as the empirical distribution function $\hat{F}_n(x)$.

Step 3: select the empirical threshold $X_{(n-k)}$ Choose k as described in paragraph 3.1.2.

Step 4: estimate the parameters of GPD function Use the moment estimators $\hat{\gamma}$ and $\hat{\alpha}$ defined by (6) and (7) to estimate the extreme value index γ and the scale $\alpha = f(\tau)$.

Step 5: estimate the relation between S and service level This step depends on the service level used.

1. *Expected waiting time* If $\hat{\gamma} < 1$, estimate the expected waiting time $\mathbb{E}\mathrm{WT}(S)$ by means of the estimator $\widehat{\mathbb{E}\mathrm{WT}}(S)$ defined by

$$\widehat{\mathbb{E}\mathrm{WT}}(S) = \frac{1}{\bar{D}}\left\{n^{-1}\sum_{i=1}^{n-k}\left(X_{(i)} - S\right)^{+} + \frac{k}{n}\hat{\mu}_{\mathrm{tail}}(S)\right\} \tag{10}$$

with

$$\hat{\mu}_{\mathrm{tail}}(S) = \left(X_{(n-k)} - S\right)^{+} + \hat{\Psi}\left(S \vee X_{(n-k)}\right) - \hat{\Psi}\left(X_{(n-k)} - \frac{\hat{\alpha}}{\hat{\gamma}}\right)1_{\{\hat{\gamma}<0\}} \tag{11}$$

and

$$\hat{\Psi}(x) := \begin{cases} \dfrac{\hat{\alpha}}{1-\hat{\gamma}}\left\{1 + \hat{\gamma}\left(\dfrac{x - X_{(n-k)}}{\hat{\alpha}}\right)\right\}^{1-\frac{1}{\hat{\gamma}}} & \text{if } \hat{\gamma} \neq 0, \\ \hat{\alpha}\exp\left(\dfrac{-x + X_{(n-k)}}{\hat{\alpha}}\right) & \text{if } \hat{\gamma} = 0, \end{cases} \tag{12}$$

For the derivation of this estimator, see Appendix B in the supplementary material.

Set the estimated expected waiting time to infinite if $\hat{\gamma} \geq 1$. In this case, the waiting time distribution is classified as extremely heavy-tailed.

Although the expected waiting time $\mathbb{E}\mathrm{WT}$ is a widely used service level, it has the problematic feature of becoming infinite if $\gamma \geq 1$, or becoming extremely high for γ slightly lower than 1. This feature, which is a consequence of the fact that mathematical expectation does not exist for heavy tailed distribution, is inherited by its estimator: $\widehat{\mathbb{E}\mathrm{WT}}$ is infinite if $\hat{\gamma} \geq 1$, or extremely high for $\hat{\gamma}$ slightly lower than 1, see Appendix A.

In short, the expected waiting time does not combine well with heavy tailed waiting time distributions.

2. *Fill rate* The fill rate $\beta(S)$ is closely related to the expected waiting time. Recall that, according to (8), we may view X_i as just short hand notation for $X_i^{[L]}$. Let $\mathbb{E}\mathrm{WT}^{[L]}(S)$ be the expected waiting time for $X^{[L]}$, and let $\mathbb{E}\mathrm{WT}^{[L-1]}(S)$ be the expected waiting time for $X^{[L-1]}$. Then, $\beta(S) = 1 - \bar{D}^{-1}\{\mathbb{E}\mathrm{WT}^{[L]}(S) - \mathbb{E}\mathrm{WT}^{[L-1]}(S)\}$.

Now suppose that in Step 1 we have not only collected $X_1^{[L]}, X_2^{[L]}, \ldots, X_n^{[L]}$ but $X_1^{[L-1]}, X_2^{[L-1]}, \ldots, X_n^{[L-1]}$ as well. Note that if the demands are non-negative random variables, then $X_i^{[L]}$ is stochastically larger than $X_i^{[L-1]}$ for every i. Let $\widehat{\mathbb{E}\mathrm{WT}}^{[L]}(S)$ be the expected waiting time for $X^{[L]}$, and let $\widehat{\mathbb{E}\mathrm{WT}}^{[L-1]}(S)$ be the expected waiting time for $X^{[L-1]}$. Then, $\hat{\beta} = \hat{\beta}(S) = 1 - \bar{D}^{-1}\{\widehat{\mathbb{E}\mathrm{WT}}^{[L]}(S) - \widehat{\mathbb{E}\mathrm{WT}}^{[L-1]}(S)\}$ is an estimator of $\beta(S)$.

The fill rate suffers from the same problems as the expected waiting time for large values of $\hat{\gamma}$. Moreover, it also does not combine well with heavy tailed waiting time distributions. Although $X_i^{[L]}$ is stochastically larger than $X_i^{[L-1]}$ for every i, this does not imply that $X_i^{[L]} - X_{(n-k)}^{[L]}$ is stochastically larger than $X_i^{[L-1]} - X_{(n-k)}^{[L-1]}$ for every i since there is no obvious relation between $X_{(n-k)}^{[L]}$ and $X_{(n-k)}^{[L-1]}$. As a consequence, even though $\widehat{\mathbb{E}\mathrm{WT}}^{[L]}(S)$ and $\widehat{\mathbb{E}\mathrm{WT}}^{[L-1]}(S)$ decrease with increasing base stock level S, $\widehat{\mathbb{E}\mathrm{WT}}^{[L]}(S) - \widehat{\mathbb{E}\mathrm{WT}}^{[L-1]}(S)$ and $\hat{\beta}(S)$ fail to change monotonically with the increasing S, which contradicts common sense.

3. *Cycle service level* The cycle service level $\mathrm{CSL}(S) := P(X \leq S)$ is a service level which, in contrast to the expected waiting time and fill rate, is able to accommodate heavy tailed distributions. It is estimated by means of $\widehat{\mathrm{CSL}}(S)$ defined by

$$\widehat{\mathrm{CSL}} = \begin{cases} \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq S\}} & \text{if } S \leq \tau, \\ 1 - \frac{k}{n}\left\{1 - H_{\hat{\gamma}}\left(\frac{S - X_{(n-k)}}{\hat{\alpha}}\right)\right\} & \text{if } S < \tau, \end{cases} \tag{13}$$

For the derivation of this estimator, see Appendix C in the supplementary material.

- *Step 6: estimate the smallest base stock level* Let $\mathbb{E}\mathrm{WT}_{\mathrm{obj}}$, β_{obj} and $\mathrm{CSL}_{\mathrm{obj}}$ denote the target waiting time, the target fill rate and the target cycle service level, respectively. Now set the estimator $S_{\min}$ equal to the smallest S satisfying one of the following service level requirements: $\widehat{\mathbb{E}\mathrm{WT}}(S) \leq \mathbb{E}\mathrm{WT}_{\mathrm{obj}}$; or $\hat{\beta}(S) \geq \beta_{\mathrm{obj}}$; or $\widehat{\mathrm{CSL}}(S) \geq \mathrm{CSL}_{\mathrm{obj}}$.

Based on the discussion under Step 5 above, we recommend to use EVT with the CSL rather than $\mathbb{E}\mathrm{WT}$ and/or fill rate.

Table 1
Historical demands of 60 periods.

Period n	Demand									
period 1–10	0	0	0	0	0	0	6	0	0	0
period 10–20	0	0	0	0	0	0	0	0	0	0
period 21–30	0	0	0	0	0	0	1	0	0	0
period 31–40	0	10	0	0	0	0	4	0	0	0
period 41–50	6	0	0	0	0	0	0	3	0	0
period 51–60	0	0	0	0	0	0	0	0	0	0

Table 2
Computation of EDF $\hat{F}_n$.

LTD	Frequency	Proportion	$\hat{F}_n$
0	32	32/56	32/56
1	5	5/56	37/56
4	4	4/56	41/56
6	9	9/56	50/56
10	6	6/56	1

3.3. Example - applying EVT to the empirical LTD forecasting method

Two examples under expected waiting time and cycle service level are provided in this subsection. Example 1 shows that we can decrease the expected waiting time by applying extreme value theory. However, when the extreme value index γ is relatively large, the expected waiting time service level breaks down, due to the fat tail of the lead time demand distribution. The cycle service level does not suffer from this problem. This is illustrated in Example 2.

Example 1

Table 1 shows demand samples during 60 months. We set the expected waiting time target $\mathbb{E}\mathrm{WT}_{obj}$ equal to 0.03.

Next we tally the LTD sample and compute $\hat{F}_n$ according to (9), see Table 2. According to (6), we construct estimators $\hat{\gamma}_k$ using Table 2, see Fig. 1, for $k = 6, \ldots, 24$. $\hat{\gamma}$ does not exist for

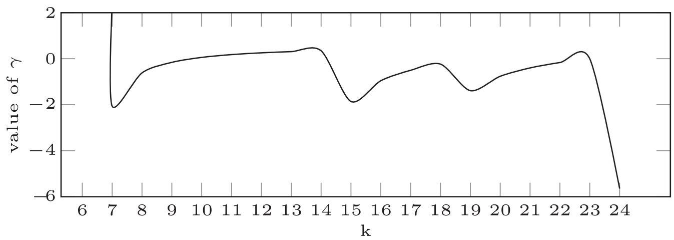

Fig. 1. Select threshold from moment estimator plot.

Table 3
Results of example 1.

Service level	S_{EVT}	S_{emp}
EWT	16	10
CSL	14	10

Table 4
Historical demands of 60 periods.

Period n	Demand									
period 1–10	0	0	0	0	0	0	0	0	5	0
period 10–20	0	0	0	0	19	0	0	0	0	0
period 21–30	0	0	0	0	0	0	0	0	0	0
period 31–40	0	0	0	5	0	0	0	0	0	0
period 41–50	0	0	0	0	0	0	0	0	0	0
period 51–60	0	0	0	0	0	0	0	1	0	5

Table 5
Results of example 2.

Service level	S_{EVT}	S_{emp}
EWT	43,213	19
CSL	34	19

$k = 1, \dots, 5$ since the denominator in (6) equals to zero. We do not consider $k > 24$ in order to keep at least one positive observation in the non-tail part. The figure shows that the value of $\widehat{\gamma}$ is relatively stable when $8 \leq k \leq 14$, and we select the threshold position k equal to 10. Once the threshold is determined, we can obtain the estimates $\widehat{\gamma} = 0.056$ and $\widehat{\alpha} = 2.299$. Now (10) allows us to calculate $\widehat{\mathbb{E}\text{WT}}(S)$ for any given S, which in turn enables us to determine $S_{\text{EVT}}^{\text{EWT}} = S_{\min}$, where $S_{\min}$ is the smallest S which satisfies $\widehat{\mathbb{E}\text{WT}}(S) \leq \mathbb{E}\text{WT}_{\text{obj}}$. Finally, this yields $S_{\text{EVT}}^{\text{EWT}} = 16$.

The procedure above could be followed for other service levels as well. For instance, we obtain $S_{\text{EVT}}^{\text{CSL}} = 14$ under cycle service level $\text{CSL}_{\text{obj}} = 0.99$, see (13). For reference, we remark that in this example the empirical method proposed in Porras and Dekker (2008) yields $S_{\text{emp}}^{\text{EWT}} = 10$, $S_{\text{emp}}^{\text{CSL}} = 10$, see Table 3 for the results.

Example 2

This example produces an illustration to the remark in Step 5 in Section 3.2 that the expected waiting time does not combine well with heavy tailed distributions. Table 4 shows another demand sample during 60 months. We use the same target and lead time as in Example 1. Following the same procedure we obtain the results in Table 5.

Theoretically, the tail becomes fatter when the extreme value index γ is closer to 1. As the extreme value index estimate 0.67 is already rather large, $S_{\text{EVT}}^{\text{EWT}}$ becomes enormous. As long as $\widehat{\gamma}$ is moderate, the expected waiting time produces reasonable results. However, if the estimated extreme value index is rather large (we have observed $\widehat{\gamma} > 0.5$), then the expected waiting time yields unrealistically high base stock levels. To avoid such extremely high base stock levels, it is better to use the EVT method with the cycle service level. Alternatively, we may impose an upper bound on $S_{\text{EVT}}^{\text{EWT}}$ to avoid the extreme cases. One could also opt to limit the parameter space of $\widehat{\gamma}$, but the upper limit needs to depend on $\hat{\alpha}$, which would make such an approach cumbersome.

4. Experiments

In this section, we perform experiments comparing the relative performance of the empirical-EVT method and several alternative methods. Section 4.1 discusses experiments where demand is generated using Monte Carlo simulation, and Section 4.2 discusses experiments based on real demand data.

4.1. Simulation

4.1.1. Setup

We apply a base stock policy with periodic review and full backordering. In the simulation, demand for each time period for both the *training set* and *test set* is generated according to a probability distribution that will be specified in Section 4.1.2. The test set is independent of the training set. Lead time demands for the training set are constructed according to (8). Given a target service level, the training set is used to estimate the base stock level $S_{\min}$ using various methods: empirical-EVT, the empirical method, WSS (Willemain & Smart, 2001; Willemain et al., 2004), Croston (Croston, 1972) and SBA (Syntetos & Boylan, 2005). We choose smoothing constant $\alpha = 0.2$ for Croston's method and SBA. We further use the test set to obtain the estimated service level $\widehat{\text{EWT}}^*$ and $\widehat{\text{CSL}}^*$ under such base stock level $S_{\min}$. Here * denotes that the estimator is obtained from the training set.

Thus, the setting of our experiment conforms to the setting faced by companies in real life: forecasts and inventory levels must be set based on some past period (our training period), while the resulting base-stock levels are applied for some future period (our test period), and the quantity of interest is the performance of the base-stock level in this future period. We will vary the length of the training set, because the amount of data present may affect performance in practice. The length of the test set is fixed to 1000 periods, and the simulation is replicated for 5000 times for each parameter setting. We report the average performance over the test set for all replications.

We have three designs in our simulation experiments.

- **Design A**: In order to explore the influence of training set length n on the performance of the methods for setting base-stock levels, we set $L = 5$, fix the target service level $\mathbb{E}\text{WT}_{obj} = 0.03$ or $\text{CSL}_{obj} = 0.97$ and increase n from 60 to 500 time periods. This includes periods with both positive demand and zero demand and corresponds to 12–100 positive demand observations, since we will have positive demand in roughly 1 in 5 periods, see Section 4.1.2. Results are given in Figs. 2, 5 and 7.
- **Design B**: We consider a training set length $n = 60$, $L = 5$ and different target service levels (CSL_{obj} varies from 0.85 to 0.99, $\mathbb{E}\text{WT}_{obj}$ from 0.01 to 0.15) to explore the impact of target service level on performance, see Figs. 3, 6 and 8.
- **Design C**: We vary the lead time L from 2 to 6 given target service level $\mathbb{E}\text{WT}_{obj} = 0.03$ or $\text{CSL}_{obj} = 0.97$ and training set length $n = 60$ to explore the influence of the lead time on the performance of the various methods, see Fig. 4.

In Example 2 in Section 3.3, we have seen that the estimated expected waiting time may become extremely high for larger estimated extreme value indexes $\widehat{\gamma}$. To avoid this issue, we set an upper limit where $S_{\text{EVT}} = 1.5 \cdot S_{\text{emp}}$. We also use S_{emp} as a lower limit on S_{evt}. Note that very large S_{evt} does not occur when using the cycle service level (see Step 5 in Section 3.2).

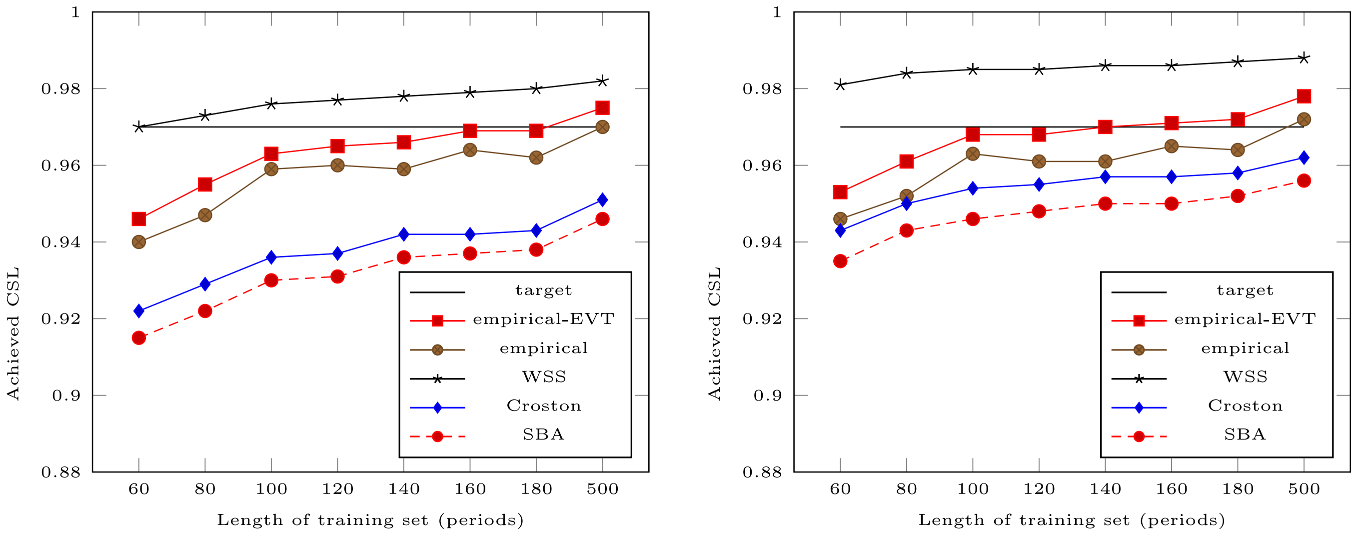

Fig. 2. Simulation results of achieved CSL under fixed target 0.97 and different number of observations (including positive and zero demand). Only CM demand is considered. The underlying positive demand distribution is compound Poisson (left)/folded normal (right). Each result shows the average of 5000 simulations.

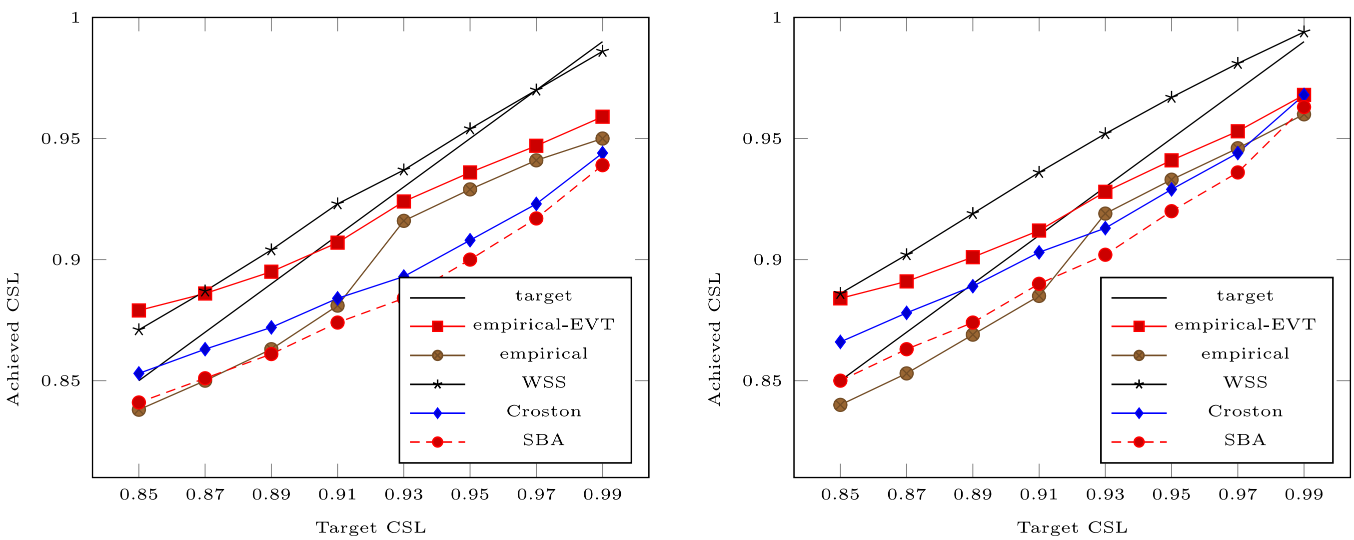

Fig. 3. Simulation results of achieved CSL under fixed training set length 60 (including positive and zero demand) and different targets. Only CM demand is considered. The underlying positive demand distribution is compound Poisson (left)/folded normal (right). Each result shows the average of 5000 simulations.

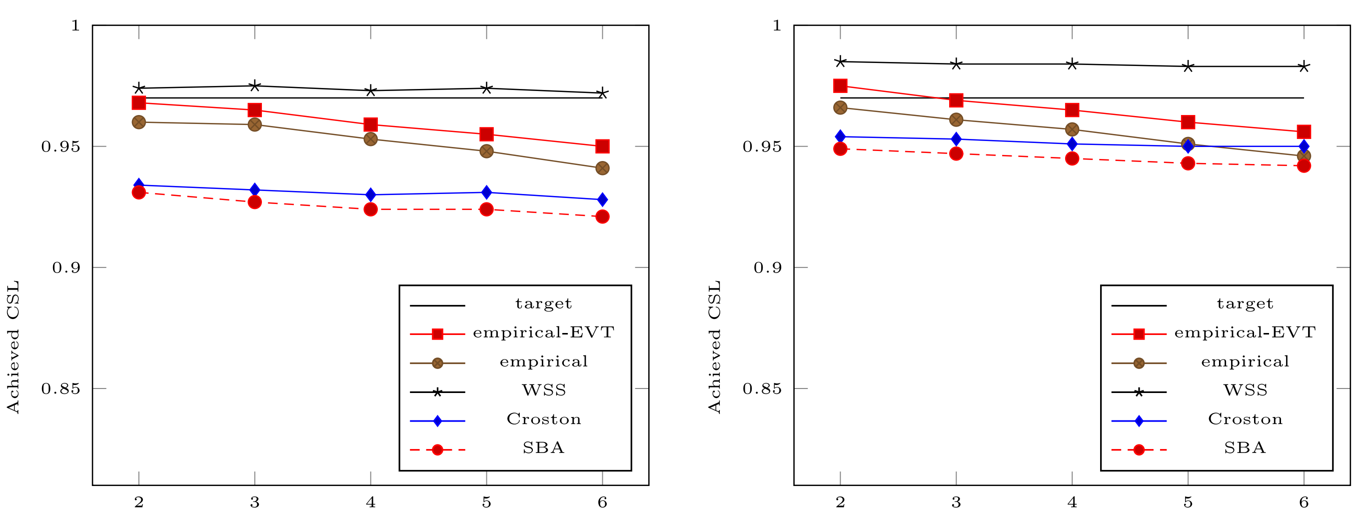

Fig. 4. Simulation results of achieved CSL under fixed training set length 80 (including positive and zero demand), fixed target 0.97 and different lead time. Only CM demand is considered. The underlying positive demand distribution is compound Poisson (left)/folded normal (right). Each result shows the average of 5000 simulations.

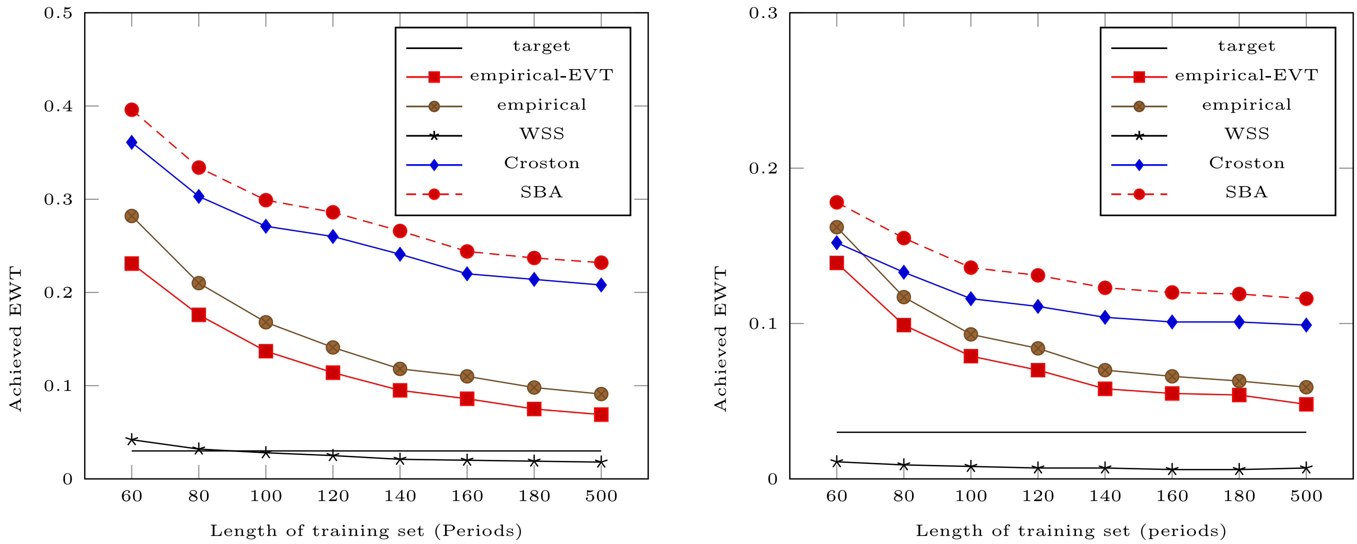

Fig. 5. Simulation results of achieved $\mathbb{E}$WT under fixed target 0.03 and different number of observations (including positive and zero demand). Only CM demand is considered. The underlying positive demand distribution is compound Poisson (left)/folded normal (right). Each result shows the average of 5000 simulations.

4.1.2. Demand process

In our simulations, i.i.d intermittent demand is generated as follows. First, we consider demand generated for corrective maintenance (CM). In each time period, a positive CM demand is generated with probability $p_{nonzero} = 0.2$ and zero demand with probability $p_{zero} = 1 - p_{nonzero} = 0.8$. Next we choose one of the following distributions to represent the positive integer CM demand: (I). Geometric based compound Poisson distribution with $p = 0.5$ and $\lambda = 2.5$; (II). Truncated normal distribution with $u = 5, \sigma^2 = 3$, where we set negative values to zero. See Lengu, Syntetos, and Babai (2014) for a detailed discussion on compound Poisson distributions and their fit to spare parts demand. It is well-known that the truncated normal distribution satisfies (1). Note that Shimura (2012, p. 304 example 1) establishes applicability of EVT for the geometric distribution. Moreover, as the tail behavior of the compound Poisson distribution is related with the right tail of the compounding distribution (Willmot, 1990, p. 147), this also establishes applicability of EVT for compound Poisson demand with a geometric compounding distribution.

We also consider cases in which this CM is augmented with demand stemming from preventive maintenance (PM). PM in general may have a relatively large value. Positive integer PM demand is generated once every 12 time periods and has truncated normal distribution with $u = 12, \sigma^2 = 2$. In simulations, we thus either consider CM demand only, or we consider CM and PM together by using positive PM demand to replace the corresponding CM demand in the same period.

4.1.3. Results

We compare the accuracy of the various methods by evaluating the differences between the targets and the achieved service levels. We discuss underperformance (real performance does not reach target performance) as well as overperformance (real performance exceeds the target).

Figs. 2–6 focus on situations with CM demand only. The empirical-EVT consistently outperforms the empirical method. Both the empirical-EVT and the WSS method achieve real cycle service level that is quite close to the target. Here, WSS is closer to the target for relatively short training set length (60–80 periods - 12–16 positive demand observations), while empirical-EVT is closer to the target for more periods ($>$ 20 observations). This would give the WSS an edge in practice because the number of demand observations is typically limited there. Overperformance of WSS is observed when the positive demand is normally distributed, see Figs. 2 and 3. The empirical method, Croston's method and SBA have difficulties in reaching the target. The achieved cycle service level of each approach increases as the training set length increases. Fig. 4 shows that the empirical method and empirical-EVT are the most sensitive to the lead time, while the WSS method seems very robust to changes in the lead time.

Figs. 5 and 6 show the performance of each approach under the expected waiting time target. The performance of each approach improves with the increase of training set length. Empirical-EVT outperforms the empirical method, Croston's method and SBA. WSS however performs better than empirical-EVT. It has very good performance for compound Poisson demand, but it again overperforms when demand is normally distributed.

We will see in Section 4.2 that Croston's method and SBA attain a much higher performance for empirical datasets when compared to their results in Figs. 2–6 for CM only demand. A partial explanation may be given by sudden large demands in the empirical data, which may result from planned/preventive maintenance (PM) actions. The limited information on the empirical data can neither confirm nor rule out the role of PM in the large demands. To test the effect of sudden large demands in a more controlled environment, we report on simulation experiments in which CM demand is augmented with PM demand, as discussed in Section 4.1.2.

The results are shown in Figs. 7 and 8. For each demand distribution in which PM demand is considered alongside CM we find that Croston's method and SBA attain a much higher CSL than in the CM only case. Additionally, SBA reduces the overstocking by Croston's method. Moreover, the cycle service level of the empirical-EVT is quite close to the target, and the most accurate of all methods for higher targets. The empirical method achieves a lower cycle service level while Croston's method, SBA and WSS result in overperformance. When the target is relatively low, overperformance happens under all approaches except for the empirical method.

From Fig. 7 we observe that SBA and Croston have a CSL that is too high, whereas our empirical-EVT method has a CSL that is below the target. An interesting idea then is to *combine* SBA and empirical-EVT by taking the simple average of the base-stock levels that result from the two methods. Admittedly, averaging the

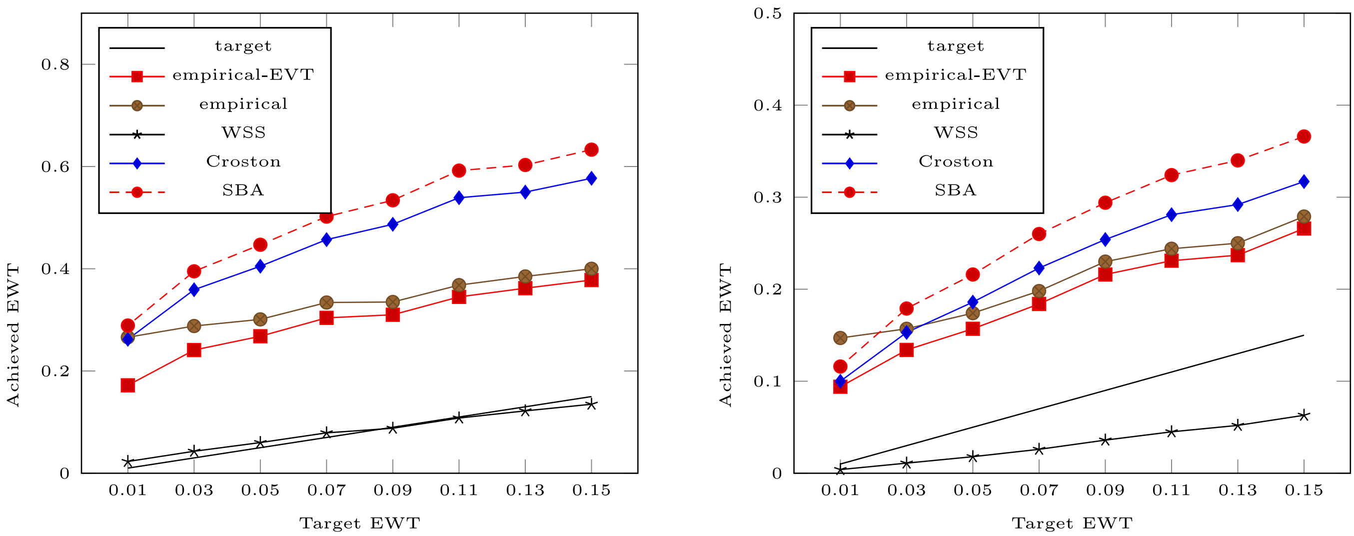

Fig. 6. Simulation results of achieved 𝔼WT under fixed training set length 60 (including positive and zero demand) and different targets. Only CM demand is considered. The underlying positive demand distribution is compound Poisson (left)/folded normal (right). Each result shows the average of 5000 simulations.

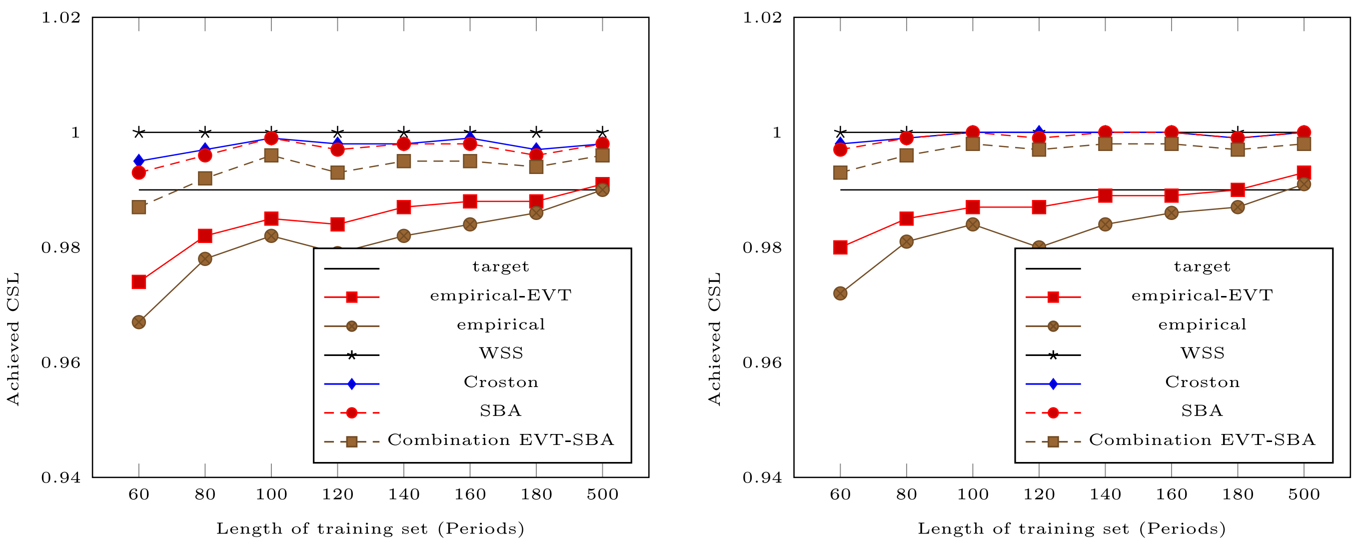

Fig. 7. Simulation results of achieved CSL under fixed target 0.99 and different number of observations (including positive and zero demand). Both CM demand and PM demand are considered. The underlying distribution of positive CM demand is compound Poisson (left)/folded normal (right). Each result shows the average of 5000 simulations.

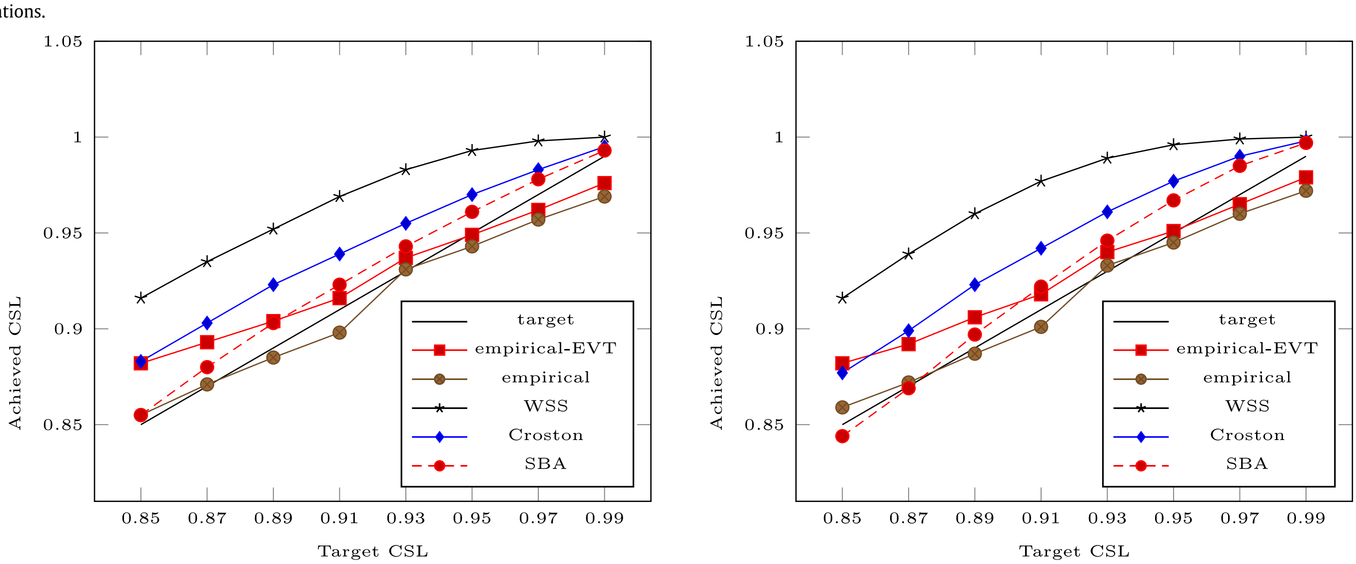

Fig. 8. Simulation results of achieved CSL under fixed training set length 60 (including positive and zero demand) and different targets. Both CM demand and PM demand are considered. The underlying distribution of positive CM demand is compound Poisson (left)/folded normal (right). Each result shows the average of 5000 simulations.

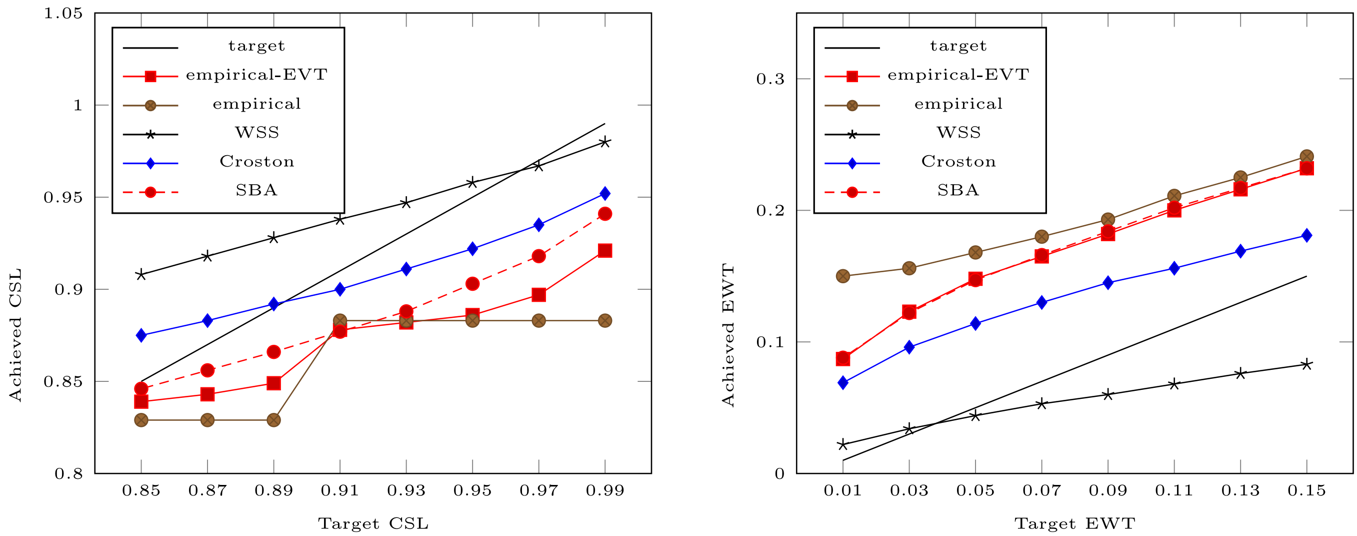

Fig. 9. Empirical results of achieved CSL (left graph) and $\mathbb{E}$WT (right graph) for the automotive case.

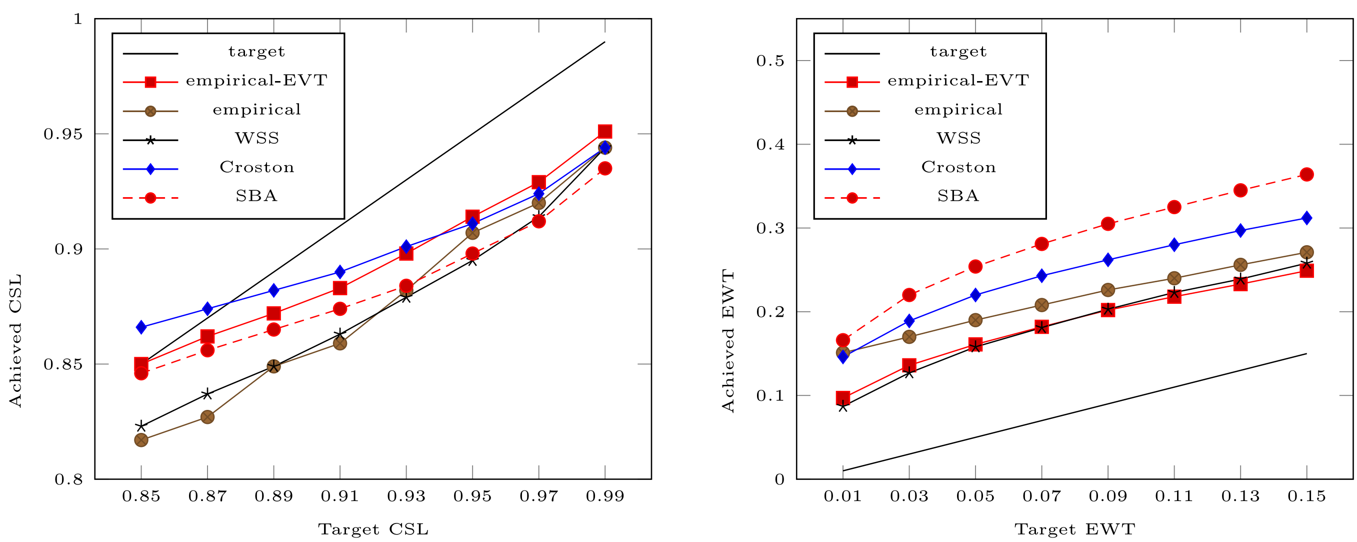

Fig. 10. Empirical results of achieved CSL (left graph) and $\mathbb{E}$WT (right graph) for the aircraft component repair case.

methods like this is a bit ad hoc, and it may be difficult in practice to identify which combinations work well. For the case at hand, the resulting achieved CSL of this combined method is also tabulated in Fig. 7, and it performs very well, especially for the compound Poisson demand.

4.2. Empirical study

4.2.1. Setup and parameters

To demonstrate empirical results of the proposed approach, we conduct a study based on real data. We use the automotive dataset described by Syntetos and Boylan (2005) from an automotive industry and data on component repairs from Romeijnders et al. (2012). The automotive industry dataset records intermittent demand of 3000 Stock Keeping Units (SKUs) over 23 time periods. The training set includes demand in the first 13 time periods and the last 10 time periods are classified as the test set.

The aircraft component repair database used in Romeijnders et al. (2012) gives spare parts usage in component repairs. We ignore the component level data and focus only on the parts used in all component repairs together. The database tracks the demand history of 16,903 types of spare parts over 122 months. We set the first 84 months as the training period for each item and test the resulting base-stock levels in the last 38 periods. The dataset contains many very slow moving items that might not be stocked in a real life setting. Therefore, we will consider only the parts which were used in at least 14 months in the 7 years of our training period, corresponding to at least 2 usages annually. Parts with less usage may not be stocked at all, and therefore the performance for these parts is less relevant. This results in 2549 parts that will be used in further analysis.

Except for using a different demand model, the setup of our experiments is in line with the approach in Section 4.1. That is, we use the data in the training set to estimate the required base-stock level to achieve a certain service, and then use the test set for determining the real service level associated with that base-stock level. This approach is applied to individual each SKU or spare part and we obtain the average service level over 3000 SKUs or 2549 spare parts. We consider a constant $L = 3$ for the automotive industry data and $L = 5$ for the component repair data.

4.2.2. Results

Figs. 9 and 10 show the empirical result of each forecasting method for the automotive case and the aircraft component case respectively. In the automotive case, Croston's method, SBA and WSS perform well. WSS results in an overperformance in case of low target CSL or high target $\mathbb{E}$WT. SBA has a lower achieved CSL (or higher achieved $\mathbb{E}$WT) than Croston's method since it uses the smoothing constant to adjust the estimator of mean demand. The empirical-EVT performs slightly worse than SBA. The empirical method has difficulties in reaching the target, when compared to the other approaches. In the aircraft component repair case, the empirical-EVT method performs the best for the $\mathbb{E}$WT, while it is the joint winner for the CSL target. In general, all methods have difficulties in achieving the target. Croston's method is competitive when CSL is considered, followed by SBA, the empirical method and WSS.

4.3. Analysis and discussion

Comparing Figs. 7 and 8 with Figs. 2 and 3, we found the performance of empirical-EVT is more stable than WSS, Croston's method and SBA in different situations. E.g., with the introduction of PM demand, the underperformance turns into overperformance for Croston's method and SBA. WSS as well leads to severe overperformance in the simulation in which PM demand is considered as well as CM demand. We provide two partial explanations. In general, PM demand has larger values than CM demand. WSS may select these large values repeatedly to forecast for each period and sum them up to estimate the lead time demand. The repeated selection gives rise to overestimations and hence overperformance. Besides, as the jittering process in WSS approach provides more variation around larger numbers than around smaller ones, PM results in large generated demands. As a result, the jittering process exacerbates the overperformance.

Overperformance of WSS is also observed in the situation of considering only CM demand and folded normal distributed positive demand. It results from the fact that positive demand values in this situation are relatively stable and the jittering process, in this case, loses its advantage by increasing the estimated base stock level $S_{\min}$ unnecessarily. That the overperformance in case of compound Poisson distributed positive demand is much less than the overperformance for folded normal supports this explanation: the compound Poisson has double the standard deviation of the folded normal distribution for our parameters.

Lead time does not have much effect on the performance of Croston's method, SBA and WSS. The accuracy of the empirical-EVT and the empirical method decreases with lead time. These latter methods obtain lead time demand history through summing up the values within time windows of lead time length. Thus, larger lead times lead to more samples with the same value as a large proportion of the demand series is zero. This results in less variation in the sample used as input for the empirical method and the empirical-EVT method. Thus larger lead time lead to the decrease in inventory performance of empirical-EVT, when keeping the demand history fixed.

The performance of WSS is highly related to the dataset. It has overperformance in case of automotive industry and underperformance in the aircraft component repair case. Performance of the empirical-EVT is influenced by the limitation of lead time demand. As the training set from the automotive industry gives demand in 13 time periods and provides only 11 lead time demand under the empirical-EVT approach, too few lead time demand above the threshold is available to estimate the tail. The less accurate estimation of the lead time demand tail limits the performance of the empirical-EVT approach. Data from the aircraft component repair case allow us to approximate the tail based on history in 84 periods.

In summary, based on Figs. 2, 5 and 7, empirical-EVT gives reasonably accurate estimates when the training set includes more than 16–24 positive demands (in our simulation, this corresponds to a training period longer than 80–120 periods). It has better accuracy than the benchmarks for such cases. For shorter demand histories, empirical-EVT has much lower accuracy in absolute terms. Its accuracy may still be better than some benchmarks, but this depends on the precise setting.

5. Conclusions

LTD forecasting is essential to spare parts inventory control but difficult as the demand has the feature of irregularity and lumpiness. Non-parametric approaches, like the empirical method, are suitable for spare parts since they can represent the erratic and lumpy demand behavior. A limited number of observations prevents the empirical method from achieving high performance.

We propose a semi-parametric LTD forecasting method for spare parts. It is applicable for forecasting the lead time demand and determining the inventory control parameters of spare parts. The empirical-EVT method is a combination of non-parametric empirical method and EVT extrapolation. It samples LTD from actual data and uses EVT to model the distribution above a high threshold so that it can predict possible extreme values. The new method can represent the demand behavior as well as achieve high target service levels.

We build models for different service levels and analyse their applications. Simulation shows that the empirical-EVT method has a relative good performance and avoids overperformance which regularly happens under WSS, Croston's method and SBA. Still, the empirical-EVT has performance issues with limited demand histories, and may be outperformed by WSS, and even by simpler methods such as Croston's and SBA. The empirical study based on datasets from two companies demonstrates that accuracy of WSS highly depends on the dataset. Moreover, the test shows that the empirical-EVT struggles to perform well when demand history consists of only very few periods. In contrast, performance of empirical-EVT is better in cases where only relatively few demand points are available, but over many periods, as shown for our second empirical test. In those cases, the method is rather competitive. This should be taken into account when considering to apply the method in practice.

Our theoretical treatise indicates that the empirical-EVT method has a problem in estimating the fill rate. The fill rate fails to change monotonously with the increase of base stock level when applying EVT independently for the LTD with lead time L and the LTD with lead time $L-1$. Another issue arises for the expected waiting time, which can only be estimated when the extreme value index is not bigger than or close to 1. This problem is solved by considering the cycle service level instead of expected waiting time. The empirical-EVT method in combination with the cycle service level works well. However, the issues related to applying EVT with expected waiting time of fillrate may be a limiting factor when applying it.

Future research should focus on three related problems. Firstly, we observed that the empirical-EVT method might overestimate the base stock level in the case of large training set length, and it could be interesting to further investigate this convergence issue. A second issue is finding ways to apply EVT to estimate the fill rate, in order to avoid the lack of monotonicity identified in Section 3.2. Lastly, it could be interesting to apply EVT to other forecasting methods. E.g., in WSS it could be used to replace the jittering process in order to general lead time demand which has not been observed in the history. It is not immediately clear how to use the EVT approach with parametric methods such as Croston's method or SBA, and this too is an interesting subject for further research.

Acknowledgment

We thank two anonymous reviewers for their valuable input. We thank an anonymous reviewer for his suggestion to combine EVT and Croston/SBA.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.ejor.2017.01.053.

References

Altay, N., Litteral, L. A., & Rudisill, F. (2012). Effects of correlation on intermittent demand forecasting and stock control. *International Journal of Production Economics, 135*(1), 275–283. doi:10.1016/j.ijpe.2011.08.002. Advances in Optimization and Design of Supply Chains.

Athreya, K., & Pantula, S. (1986). A note on strong mixing of arma processes. *Statistics and Probability Letters, 4*, 187–190. doi:10.1016/0167-7152(86)90064-7.

Babai, M. Z., Syntetos, A., & Teunter, R. (2014). Intermittent demand forecasting: an empirical study on accuracy and the risk of obsolescence. *International Journal of Production Economics, 157*(0), 212–219. doi:10.1016/j.ijpe.2014.08.019.

Balkema, A., & de Haan, L. (1974). Residual life time at great age. *The Annals of Probability, 2*(5), 792–804. doi:10.1214/aop/1176996548.

Beirlant, J., Caeiro, F., & Gomes, I. M. (2012). An overview and open research topics in statistics of univariate extremes. *REVSTAT, 10*(1). URL https://www.ine.pt/revstat/pdf/rs120101.pdf.

Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. (2004). Statistics of extremes: theory and applications. *Wiley Series in Probability and Statistics*. Chichester: John Wiley and Sons Ltd,.

Boylan, J., & Syntetos, A. (2010). Spare parts management: a review of forecasting research and extensions. *Journal of Management Mathematics, 21*, 227–237. doi:10.1093/imaman/dpp016.

Bradley, R. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys, 2*, 107–144. doi:10.1214/154957805100000104.

Carrasco, M., & Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory, 18*, 17–39. doi:10.1017/S0266466602181023.

Cohen, M. A., Agrawal, N., & Agrawal, V. (2006). Winning in the aftermarket. *Harvard Business Review, 84*(5), 129–138.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. London: Springer-Verlag.

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

Danielsson, J., de Haan, L., Peng, L., & de Vries, C. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis, 76*(2), 226–248. doi:10.1006/jmva.2000.1903.

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

Dekkers, A., Einmahl, J., & de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics, 17*(4), 1833–1855. doi:10.1214/aos/1176347397.

Draisma, G., de Haan, L., Peng, L., & Pereira, T. (1999). A bootstrap-based method to achieve optimality in estimating the extreme-value index. *Extremes, 2*(4), 367–404. doi:10.1023/A:1009900215680.

Drees, H. (2003). Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli, 9*(1), 617–657. doi:10.3150/bj/1066223272.

Drees, H., de Haan, L., & Li, D. (2006). Approximations to the tail empirical distribution function with application to testing extreme value conditions. *Journal of Statistical Planning and Inference, 136*(10), 3498–3538. doi:10.1016/j.jspi.2005.02.017.

Eaves, A. H. C., & Kingsman, B. G. (2004). Forecasting for the ordering and stockholding of spare parts. *The Journal of the Operational Research Society, 55*(4), 431–437. Special Issue: OR in Defence. URL http://www.jstor.org/stable/4101997.

Fryzlewicz, P., & Rao, S. (2011). Mixing properties of ARCH and time-varying ARCH processes. *Bernoulli, 17*(1), 320–346. doi:10.3150/10-BEJ270.

Gallagher, T., Mitchke, M. D., & Rogers, M. C. (2005). Profiting from spare parts. *The McKinsey Quarterly, 2*, 1–4.

Granger, C. W. J., & Morris, M. J. (1976). Time series modelling and interpretation. *Journal of the Royal Statistical Society. Series A (General), 139*(2), 246–257. URL http://www.jstor.org/stable/2345178.

de Haan, L. (1990). Fighting the arch-enemy with mathematics. *Statistica Neerlandica, 44*(2), 45–68. doi:10.1111/j.1467-9574.1990.tb01526.x.

de Haan, L., & Ferreira, A. (2006). *Extreme value theory: an introduction*. New York: Springer.

Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics, 3*(5), 1163–1174. doi:10.1214/aos/1176343247.

Jalil, M., Zuidwijk, R., Fleischmann, M., van Nunen, A., & Jo (2011). Spare parts logistics and installed base information. *Journal of the Operational Research Society, 62*(3), 442–457. doi:10.1057/jors.2010.38. Special Issue: Supply Chain Forecasting and Planning.

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

Kennedy, W., Patterson, J. W., & Fredendall, L. D. (2002). An overview of recent literature on spare parts inventories. *International Journal of Production Economics, 76*(2), 201–215. doi:10.1016/S0925-5273(01)00174-8.

Kogan, V., & Rind, T. (2011). Determining critical power equipment inventory using extreme value approach and an auxiliary poisson model. *Computers & Industrial Engineering, 60*(1), 25–33. doi:10.1016/j.cie.2010.09.013.

Koning, A. J., & Peng, L. (2008). Goodness-of-fit tests for a heavy tailed distribution. *Journal of Statistical Planning and Inference, 138*(12), 3960–3981. doi:10.1016/j.jspi.2008.02.013.

Koudal, P. (2006). The service revolution in global manufacturing industries. *Deloitte Research*: vol. 2. doi:10.13140/2.1.2631.9042.

Lengu, D., Syntetos, A. A., & Babai, M. Z. (2014). Spare parts management: linking distributional assumptions to demand classification. *European Journal of Operational Research, 235*(3), 624–635.

Mokkadem, A. (1988). Mixing properties of ARMA processes. *Stochastic Processes and their Applications, 29*(2), 309–315. doi:10.1016/0304-4149(88)90045-2.

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

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

Pickands, J., III (1975). Statistical inference using extreme order statistics. *The Annals of Statistics, 3*(1), 119–131. doi:10.1214/aos/1176343003.

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

Reiss, R.-D., & Thomas, M. (2007). *Statistical analysis of extreme values with applications to insurance, finance, hydrology and other fields* (3rd). Basel: Birkhäuser.

Resnick, S., & Starica, C. (1998). Tail index estimation for dependent data. *The Annals of Applied Probability, 8*(4), 1156–1183. doi:10.1214/aoap/1028903376.

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

Rootzén, H. (2009). Weak convergence of the tail empirical process for dependent sequences. *Stochastic Processes and their Applications, 119*(2), 468–490. doi:10.1016/j.spa.2008.03.003.

Rostami-Tabar, B., Babai, M., Syntetos, A., & Ducq, Y. (2013). Demand forecasting by temporal aggregation. *Naval Research Logistics, 60*(6), 479–498. doi:10.1002/nav.21546.

Rostami-Tabar, B., Babai, M., Syntetos, A., & Ducq, Y. (2014). A note on the forecast performance of temporal aggregation. *Naval Research Logistics, 61*(7), 489–500. doi:10.1002/nav.21598.

Shale, E., Boylan, J., & Johnston, F. (2006). Forecasting for intermittent demand: the estimation of an unbiased average. *Journal of the Operational Research Society, 57*(5), 588–592. URL http://www.jstor.org/stable/4102455.

Shimura, T. (2012). Discretization of distributions in the maximum domain of attraction. *Extremes, 15*, 299–317. doi:10.1007/s10687-011-0137-7.

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

Syntetos, A. (2001). *Forecasting for Intermittent Demand*. Buckinghamshire Chilterns University College. Ph.D. thesis.

Syntetos, A., Babai, M., & Altay, N. (2012). On the demand distributions of spare parts. *International Journal of Production Research, 50*(8), 2101–2117. doi:10.1080/00207543.2011.562561.

Syntetos, A., & Boylan, J. (2001). On the bias of intermittent demand estimates. *International Journal of Production Economics, 71*, 457–466. doi:10.1016/S0925-5273(00)00143-2.

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

Syntetos, A., & Boylan, J. (2010). On the variance of intermittent demand estimates. *International Journal of Production Economics, 128*, 546–555. doi:10.1016/j.ijpe.2010.07.005.

Syntetos, A. A., Babai, M. Z., & Gardner, E. S., Jr. (2015). Forecasting intermittent inventory demands: simple parametric methods vs. bootstrapping. *Journal of Business Research, 68*(8), 1746–1752. doi:10.1016/j.jbusres.2015.03.034. Special Issue on Simple Versus Complex Forecasting.

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

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

Teunter, R., & Sani, B. (2009). On the bias of Croston's forecasting method. *European Journal of Operational Research, 194*, 177–183. doi:10.1016/j.ejor.2007.12.001.

Teunter, R., Syntetos, A., & Babai, M. (2011). Intermittent demand: linking forecasting to inventory obsolescence. *European Journal of Operational Research, 214*, 606–615. doi:10.1016/j.ejor.2011.05.018.

Topan, E., Tan, T., van Houtum, G. J. J. A. N., & Dekker, R. (2016). Using imperfect advance demand information in lost-sales inventory systems. *BETA publicatie : working papers*: 512. Eindhoven: Technische Universiteit Eindhoven.

Volkonskiĭ, V. A., & Rozanov, Y. A. (1959). Some limit theorems for random functions. *Theory of Probability and Its Applications, 4*(2), 178–197. doi:10.1137/1104015.

Willemain, T. R., & Smart, C. N. (2001). System and method for forecasting intermittent demand. U.S Patent, (6,205,431 B1). Filed Oct 29, 1998, and issued Mar 20, 2001.

Willemain, T. R., Smart, C. N., & Schwarz, H. F. (2004). A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting, 20*(3), 375–387. doi:10.1016/S0169-2070(03)00013-X.

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

Willmot, G. E. (1990). Asymptotic tail behaviour of Poisson mixtures with applications. *Advances in Applied Probability, 22*(1), 147–159. URL http://www.jstor.org/stable/1427602

van Wingerden, E., Basten, R., Dekker, R., & Rustenburg, W. (2014). More grip on inventory control through improved forecasting: a comparative study at three companies. *International Journal of Production Economics, 157*, 220–237. doi:10.1016/j.ijpe.2014.08.018.