

On Intermittent Demand Model Optimisation and Selection

Nikolaos Kourentzes^{1,*}

Lancaster University Management School, Lancaster, LA1 4YX, UK

Abstract

Intermittent demand time series involve items that are requested infrequently, resulting in sporadic demand. Croston's method and its variants have been proposed in the literature to address this forecasting problem. Recently other novel methods have appeared. Although the literature provides guidance on the suggested range for model parameters, a consistent and valid optimisation methodology is lacking. Growing evidence in the literature points against the use of conventional accuracy error metrics for model evaluation for intermittent demand time series. Consequently these may be inappropriate for parameter or model selection. This paper contributes to the discussion by evaluating a series of conventional time series error metrics, along with two novel ones for parameter optimisation for intermittent demand methods. The proposed metrics are found to not only perform best, but also provide consistent parameters with the literature, in contrast to conventional metrics. Furthermore, this work validates that employing different parameters for smoothing the non-zero demand and the inter-demand intervals of Croston's method and its variants is beneficial. The evaluated error metrics are considered for automatic model selection for each time series. Although they are found to perform similarly to theory driven model selection schemes, they fail to outperform single models substantially. These findings are validated using both out-of-sample forecast evaluation and inventory simulations.

Keywords: Intermittent demand, Croston's method, SBA method, TSB method, Forecasting, Optimisation, Model selection

*Correspondance: N Kourentzes, Department of Management Science, Lancaster University Management School, Lancaster, Lancashire, LA1 4YX, UK. Tel.: +44-1524-592911
Email address: n.kourentzes@lancaster.ac.uk (Nikolaos Kourentzes)

1. Introduction

Spare parts are typically demanded in a sporadic or intermittent fashion. This makes such time series different to conventional ones, due to the presence of several periods with zero demand. Some examples where intermittent demand can appear are listed by Willemain et al. (2004), including cases of demand of heavy machinery and respective spare parts, aircraft service parts, electronics, maritime spare parts, etc. Johnston et al. (2003) identified that in many cases such items can account for up to 60% of the total stock value. Due to their slow moving nature, such items are at greatest risk of obsolescence. This can have substantial impact on the operations of organisations, which tie resources in stocking items of this nature (Boylan and Syntetos, 2010). Companies often hold more than the necessary stock, to account for poor demand forecasts (Ghobbar and Friend, 2003), thus making accurate forecasts important.

Croston (1972) first argued that traditional time series methods, like exponential smoothing, do not produce reliable forecasts for intermittent demand time series and instead proposed an alternative method. Several studies have verified the good forecasting accuracy and inventory performance of this method; for example see Willemain et al. (1994), Johnston and Boylan (1996), Willemain et al. (2004) and Syntetos and Boylan (2006). The literature has identified several improvements of the original method, correcting it for bias (Syntetos and Boylan, 2005), or proposing new methods for intermittent demand able to overcome structural limitations of the original method, such as obsolescence issues (Teunter et al., 2011). However, most of the literature uses ad-hoc model parameters in applying these methods on intermittent demand problems.

The lack of a valid and consistent optimisation methodology complicates the application of Croston's method in real problems. The common approach in the literature is to apply a set of different parameters across all time series of a dataset (for some examples see Syntetos and Boylan, 2005; Teunter and Duncan, 2009). However, because of the nature of intermittent data, evaluating which parameter is better is not trivial. Using the conventional time series approach, that is minimising the error of some in-sample fit, is contentious, as the use of conventional errors has been challenged in the literature (Syntetos and Boylan, 2005; Wallström and Segerstedt, 2010). On the

other hand, using out-of-sample criteria for identifying which parameter performs better requires availability of adequate sample and may be impractical, if not impossible, for individual time series parameter selection. Additionally using conventional error metrics may be inappropriate, as already mentioned. Nonetheless, in the forecasting literature it is widely accepted that using the optimal model parameters for each time series is best practice, leading to accuracy improvements (Fildes et al., 1998; Hyndman et al., 2002; Gardner, 2006). This paper aims to explore this question: given the limitations of existing error metrics – as optimisation cost functions – for Croston’s method and its variants, how should we optimise such models?

In the literature there are some examples of optimised intermittent demand models (Eaves and Kingsman, 2004; Petropoulos et al., 2008; Boylan et al., 2008; Petropoulos et al., 2013; Teunter et al., 2010), however these have not focused on exploring the validity and performance of using conventional time series optimisation for intermittent data. This work explores the performance of different error metrics as cost functions for optimising forecasting models for intermittent demand. Furthermore, different optimisation setup options, such as optimising the initialisation values of the models or not, are evaluated. The conventional error metrics are found lacking and two novel ones are proposed. These are found to produce better results in terms of forecasting and inventory performance and to automatically select parameters with values that are in agreement with the suggestions of the literature.

A closely related topic to parameter selection is model selection, as both are conventionally based on evaluating the quality of the fit of models to time series. Recently there has been a surge in researching alternative intermittent demand forecasting methods, raising the issue of model selection. Assuming that for each method optimal parameters can be identified, an important problem is which model to use. Although there have been some theoretical advancements for selecting between a subset of intermittent demand methods, depending on the time series characteristics (Syntetos et al., 2005; Kostenko and Hyndman, 2006; Heinecke et al., 2011), these cannot provide generic model selection guidance. This paper investigates whether the different error metrics explored here can facilitate and automate model selection for intermittent demand data. The findings are contrasted with the literature, highlighting problems in model selection for intermittent demand time series.

The rest of the paper is structured as follows: section 2 frames the pa-

parameter and model selection problem for intermittent demand drawing from the literature. Section 3 presents the forecasting methods that this paper focuses on. Section 4 discusses the applicability of established metrics as optimisation cost functions and introduces two novel ones that overcome the limitation of existing ones. Section 5 presents the experimental setup and the results of the empirical evaluation, while section 6 concludes with a short discussion of the findings.

2. Background research

Croston (1972) proposed a forecasting method specific for intermittent demand problems. Croston's method derives a non-zero demand and an inter-demand interval time series from the original intermittent data. These new time series are then smoothed and forecasted independently using single exponential smoothing, employing the same smoothing parameter α for both. Dividing the resulting estimates produces the final forecast, which is used to predict the average future demand per time period. Since then, Croston's method has been widely researched, applied in practice, and attracted some criticism. The theoretical grounding of the method has been questioned by Snyder (2002), Shenstone and Hyndman (2005) and Snyder et al. (2012). Furthermore, the method assumes that the demand size and the inter-demand intervals are independent, something that has been questioned by Willemain et al. (1994). Kourentzes (2013) showed that modelling such dependence explicitly is beneficial. Nonetheless, several studies have verified the good forecasting accuracy and inventory performance of this method (for example see Willemain et al., 1994; Johnston and Boylan, 1996; Willemain et al., 2004), with Shenstone and Hyndman (2005) arguing that these limitations and issues do not "*mean that Croston's method itself is not useful*". The availability of the method in several established forecasting packages, such as ForecastPro, SAS and SAP APO, is indicative of its widespread use in practice by organisations. These factors make it important to explore how to optimally parametrise and use such models.

Syntetos and Boylan (2001) showed that the original Croston's method is biased and proposed a modified version that corrected the problem (Syntetos and Boylan, 2005), demonstrating improved accuracy. Teunter and Sani (2009b) and Wallström and Segerstedt (2010) provide evidence that this modification can still be biased when the intermittency of a series is quite low. Shale et al. (2006) showed that if the orders arrive as a Poisson process then

a different modification is more appropriate. Levén and Segerstedt (2004) proposed an alternative modification to Croston’s method to avoid the bias of the original method, yet this was shown to be biased in a different manner (Boylan and Syntetos, 2007). Teunter et al. (2011) observed that although the previous work dealt with the bias of Croston’s method, it was still unsuitable to deal with obsolescence issues. To address this, they proposed a new method that updates the probability of demand continuously, in contrast to Croston’s method that updates its estimates only when non-zero demand is observed. Recently, several other methods have appeared in the literature, such as bootstrapping based methods (Willemain et al., 2004; Porras and Dekker, 2008), neural networks (Gutierrez et al., 2008; Kourentzes, 2013) and aggregation based methods (Nikolopoulos et al., 2010; Babai et al., 2012); however, these have not been widely used yet. The reader is referred to Boylan and Syntetos (2010) and Bacchetti and Saccani (2012) for a more detailed review of the area.

Focusing on Croston’s method and its variants, it is evident from the literature that the selection of the parameters is mostly done ad-hoc (for examples see: Syntetos and Boylan, 2005; Teunter and Duncan, 2009; Wallström and Segerstedt, 2010; Romeijnders et al., 2012). Syntetos and Boylan (2005) argue that optimising the parameters of Croston’s method is not straightforward, due to the limited non-zero observations. The short demand history causes the initialisation of the method to be carried forward into the forecasts. In practice intermittent demand datasets are often short. Arguably, this problem may be overcome by optimising both model parameters and initial values, which is standard practice for exponential smoothing on fast moving items (Hyndman et al., 2002; Ord and Fildes, 2012). Croston (1972) suggested that the parameters should be between 0.1 and 0.3, while Syntetos and Boylan (2005) advised for values between 0.05 and 0.2. Babai et al. (2011) investigated parameters up to 0.3 and found that depending on the levels of intermittency and lumpiness different smoothing parameters achieved minimum bias, thus demonstrating the need to tune the parameters according to the series at hand.

Similarly, the method proposed by Teunter et al. (2011) was not optimised, although they provide guidelines on how to select the parameters. Romeijnders et al. (2012) that employed the same method used ad-hoc parameters, although they provide additional details on how to initialise the method.

Conventional exponential smoothing optimisation is typically done on

squared or absolute errors (Gardner, 2006). Wallström and Segerstedt (2010) point out that in the case of intermittent demand, an optimiser will focus on the many zero demand periods, thus biasing the forecast to be lower than the actual demand. They found that mean absolute error was particularly vulnerable to this, something highlighted by Teunter and Duncan (2009) as well. This issue has been investigated in the literature in a related context, when trying to identify an appropriate evaluation metric for intermittent series. Such error measures will favour zero-demand forecasts, while using different error measures is not practical due to the zeros included in the time series. Special metrics have appeared in the literature (Syntetos and Boylan, 2005; Hyndman and Koehler, 2006). Although these can be calculated for the majority of circumstances, they still do not solve the potential bias in favour of a zero-forecast. Wallström and Segerstedt (2010) proposed a series of new metrics based on cumulative errors that mitigate to an extent these problems and can be seen as a proxy to inventory metrics. On the other hand, inventory metrics have been long discussed as a better evaluation alternative for intermittent demand experiments (for example see Sani and Kingsman, 1997; Syntetos and Boylan, 2006; Strijbosch et al., 2011; Kourentzes, 2013). Therefore, applying conventional exponential smoothing optimisation for Croston’s method and its variants is not trivial, explaining the large body of literature that prefers to use ad-hoc parameter selection.

There are a few examples of optimised Croston’s method in the literature. For example, Eaves and Kingsman (2004) aggregated the demand data and optimised their methods based on mean absolute percentage errors. Although this approach will remove the zero-demand periods that make the optimisation problematic, the characteristics of the time series are altered, which can be undesirable (Spithourakis et al., 2012). Petropoulos et al. (2008, 2013) optimised Croston’s method by mean squared error and found good results in terms of bias, even though there is little theoretical support for this result due to the chosen error metric. A similar approach was followed by Boylan et al. (2008). Teunter et al. (2010) also reported using optimal parameters, however they provided limited information how this was done.

Therefore a valid and consistent optimisation methodology for Croston’s method and its variants is lacking. This complicates the application of such forecasting methods in real problems. Once optimal parameters are estimated for each method for a time series, one has to identify which is the most appropriate model. There is evidence in favour of selecting the appropriate model per time series (Tashman and Kruk, 1996; Hyndman et al.,

2002), however this is an unresolved question for intermittent demand data. Syntetos et al. (2005) proposed an item classification framework based on the average inter-demand intervals and the coefficient of variation of the demand that indicated when it was advisable to use the original Croston’s method or the Syntetos-Boylan approximation. Kostenko and Hyndman (2006) refined this framework, but also recognised its limitations due to the weaknesses in Croston’s method and its variants. Nonetheless, Heinecke et al. (2011) found that the proposed refinements provided improved forecasting performance over the original classification. However, this framework is unable to support the different non-Croston based methods that have been proposed in the literature, thus limiting its power further. Ideally, in analogy to fast moving items, some valid and reliable in-sample fitting criteria would be desirable.

3. Intermittent Demand Methods

Conventional time series methods are not ideal for intermittent demand data. The standard method for forecasting such data is Croston’s method, as proposed originally by Croston (1972) and later corrected by Rao (1973). The main idea behind this method is instead of forecasting an item in the conventional way, to separate the time series into two components; the non-zero demand size z_t and the inter-demand intervals x_t . Consequently, both z_t and x_t are modelled and forecasted using single exponential smoothing. The estimated \hat{z}_t and \hat{x}_t are only updated when demand occurs and remain constant otherwise. The forecast \hat{y}_t is given by:

$$\hat{y}_t = \hat{z}_t / \hat{x}_t. \tag{1}$$

The multi-step ahead forecast is a constant with value equal to \hat{y}_t . If demand occurs at every period, Croston’s method becomes identical to single exponential smoothing.

Conventionally, both vectors are smoothed using the same parameter. Petropoulos et al. (2008, 2013) and Teunter et al. (2010) have considered separate parameters for smoothing each of z_t and x_t . Similar arguments exist in Schultz (1987) and Snyder (2002). Such a setup allows for increased flexibility on how fast each vector is updated. However, it also introduces additional complexity on selecting the appropriate parameters. Hereafter, this method is referred to as *CRO*.

Syntetos and Boylan (2001) showed that Croston’s method is biased, due to the division in eq. (1), and suggested a modified version (Syntetos and Boylan, 2005):

$$\hat{y}_t = (1 - \alpha_x/2) \hat{z}_t/\hat{x}_t, \quad (2)$$

where α_x is the smoothing parameter for x_t . This modified version has demonstrated good empirical performance, superior to the original method (Syntetos and Boylan, 2005, 2006).

Teunter et al. (2011) observed that Croston’s method is unsuitable to deal with obsolescence issues, because of the limited updates of \hat{z}_t and \hat{x}_t , which happen only in periods of non-zero demand. They proposed a new method to address this issue. Similarly to *CRO*, this method does not model the intermittent time series directly. Instead, two vectors are considered; the demand probability d_t and the demand size z_t . The demand probability is either 0 or 1 depending whether demand occurs at time t or not and is modelled using single exponential smoothing. The demand probability estimate, \hat{d}_t , is updated at the end of every period. The demand size is also modelled with a single exponential smoothing, but is updated only when a positive demand occurs. Each smoothing model uses a separate parameter, recognising the need for different rates of updates. Teunter et al. (2011) argue that the smoothing parameter of demand probability should be smaller than the smoothing parameter of demand size. The forecast of this method is:

$$\hat{y}_t = \hat{d}_t \hat{z}_t. \quad (3)$$

The authors report good performance of the method, but alert to the importance of choosing the two smoothing parameters carefully. This method will be referred to as *TSB*.

Even though the application of conventional time series models has been criticised in the literature (Willemain et al., 1994; Johnston and Boylan, 1996), single exponential smoothing is a popular benchmark for intermittent demand methods (Willemain et al., 1994; Syntetos and Boylan, 2005; Teunter and Duncan, 2009; Teunter et al., 2011). The forecasts are calculated as:

$$\hat{y}_t = \alpha y_t + (1 - \alpha) \hat{y}_{t-1}, \quad (4)$$

where α is the smoothing parameter (Ord and Fildes, 2012). Note that this method does not make any special considerations for periods of zero

demand. Nonetheless there is some evidence in the literature of its good performance for intermittent time series (for example, see Wallström and Segerstedt, 2010). Typically its smoothing parameter is kept in the same area as Croston’s method (Syntetos and Boylan, 2005; Teunter and Duncan, 2009). Hereafter this method will be referred to as *SES*.

4. Cost Functions

Cost or loss functions are used in optimisation to measure the fit of a model to the real data and are often based on some error metric. Considering conventional time series optimisation, one of the most popular cost functions is the mean squared error (MSE). Given a time series y_t and forecasts \hat{y}_t at time t , the MSE is defined as:

$$MSE_n = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (5)$$

where n is the number of in-sample observations. For the case of exponential smoothing, which is the basis of all *CRO*, *SBA* and *TSB* methods, MSE is the most widely used optimisation criteria (Gardner, 2006; Ord and Fildes, 2012). As it is based on squared errors, it is more susceptible to extreme values and the mean absolute error (MAE) has been discussed as an alternative Gardner (2006):

$$MAE_n = n^{-1} \sum_{i=1}^n |y_i - \hat{y}_i|. \quad (6)$$

By minimising either MSE or MAE one can identify the parameters of a model, conditional on a time series. As it was discussed in section 2, the applicability of such metrics for intermittent time series has been strongly contested. They tend to bias forecasts in favour of the zero-demand forecast and are not a good proxy for the corresponding inventory decisions associated with the forecasts. Nonetheless, there is some evidence of their good performance (Boylan et al., 2008; Petropoulos et al., 2008).

Wallström and Segerstedt (2010) proposed a new metric to address some of these issues. They introduced the *Periods in Stock* (PIS) metric, which measures the total number of periods a unit of the forecasted item has been in stock or out-of-stock. This measure assumes a “*fictitious*” stock, from which the over- or under-stocking is measured. This is calculated as:

$$PIS_n = - \sum_{i=1}^n \sum_{j=1}^i (y_j - \hat{y}_j). \quad (7)$$

Essentially, it tracks how the cumulative error of the forecast evolves through time. A good forecast will result in a small PIS over the assessment period. A positive number implies some stock is left over the assessment period, while a negative number is the cumulative out-of-stock. Although this metric is not immune to preferring the zero-demand forecasts for very intermittent time series, it is more robust to conventional error metrics due to the double cumulative summation involved. Even though the objective of PIS is to be used as a bias metric, considering the absolute PIS one can use it as a cost function. The minimum absolute PIS will also mean that a model provides the minimum over- or under-stocking over the in-sample assessment period, which is desirable. Nonetheless, as PIS is a bias measure, it is expected to provide minimum fitting bias, rather than minimum fit errors.

Croston's method and its variants do not provide an expected demand as a forecast, rather a "*demand rate*", as illustrated in eq. (1). The output is the average expected demand in each future period. For example a forecast of 0.1 should be seen as a demand of 1 unit over 10 periods, a demand rate of 0.1. Note that in this sense, Croston's method does not violate the integer valued nature of intermittent demand time series. Consequently, measuring the difference of such a demand rate forecast from the raw time series data is not meaningful, as they have different units (Kourentzes, 2013). Both MSE and MAE are problematic in this aspect. Although PIS overcomes this, due to its cumulative error basis, we can use this property of the forecasts to derive a series of new cost functions. At each point *CRO* provides an average expected demand over each time period. Instead of comparing this with the realised demand, we can compare it with the in-sample mean demand over time, i.e. including both zero and non-zero demand periods. This way both values are of the same nature and thus comparable. Furthermore, in order to track the evolution of the demand for instationary time series the cumulative mean is considered as in:

$$r_i = \hat{y}_i - i^{-1} \sum_{j=1}^i y_j. \quad (8)$$

From r_i we can construct a squared and an absolute version of the metric, namely the mean squared rate (MSR) and mean absolute rate (MAR) errors:

$$MSR_n = \sum_{i=1}^n r_i^2, \quad (9)$$

$$MAR_n = \sum_{i=1}^n |r_i|. \quad (10)$$

To make this cost function more robust it is advisable to winsorise the first periods of r_i due to the limited available sample to estimate the cumulative mean. The intuition behind both MSR and MAR is to fit the intermittent demand forecasts on the average demand per period, rather than the realised demand. This way the timing of the non-zero demand does not impact on the errors, while the demand probability and size are considered.

In the following section these cost functions are used to optimise the methods outlined in section 3, assessing their performance.

5. Empirical Evaluation

The objective of the empirical evaluation is to assess the performance of the five different cost functions: MSE, MAE, PIS, MSR and MAR. Furthermore, a set of different optimisation designs are evaluated. For *CRO* and *SBA* optimising a single parameter for both z_t and x_t or separate parameters is investigated. For all *CRO*, *SBA* and *TSB* the impact of optimising the initial estimates is explored, as well as imposing parameter upper bounds lower than 1. Subsequently, different alternatives for model selection based on these metrics are analysed. The evaluation is done both on accuracy and inventory metrics.

5.1. Datasets

The different setups are evaluated on 3,000 real time series of automotive spare parts. Each time series is 24 months long. This dataset originates from the study by Syntetos and Boylan (2005). The observed average non-zero demand ranges from 1 to 194 units, while the inter-demand interval from 1.04 to 2 periods. Following the classification by Syntetos et al. (2005) the dataset contains the following series: 441 erratic, 314 lumpy, 1271 smooth and 974 intermittent. Three different lead times are investigated, $t + 1$, $t + 3$ and $t + 5$. From each time series the last 5 observations are withheld as a holdout test set, providing 3000 out-of-sample forecasts for each lead time, ample to allow for valid comparisons. A simulated dataset is used to support the inventory metrics comparison. This contains 5,000 time series. The real time series are broken into pairs of demand size and intervals, which are bootstrapped to produce the new time series. This way the properties of the original dataset are retained.

5.2. Accuracy Metrics

Following the discussions in Hyndman and Koehler (2006) and Wallström and Segerstedt (2010) two different error measures are used to assess the performance of the models. First we calculate the mean absolute scaled error (MASE) across the 3,000 time series. This error has desirable statistical properties and can be calculated for intermittent time series. For each time series and forecast horizon the absolute scaled error is calculated as:

$$ASE_h = |y_h - \hat{y}_h| / \left(n^{-1} \sum_{k=2}^n |y_k - y_{k-1}| \right), \quad (11)$$

where h is the horizon and n the number of in-sample observations. Essentially this is the conventional out-of-sample absolute error, scaled by the in-sample 1-step ahead in-sample error to allow for summarising across time series. The individual ASE are averaged across all time series to provide the reported MASE value.

The originally proposed PIS is a scale depended measure of bias. To summarise it across series its absolute value is considered, which is consequently scaled by the mean in-sample demand of each item. This makes the error measure scale-independent and therefore it is now possible to simply aggregate across series. The modified scaled absolute periods in stock (sAPIS) measure is calculated as:

$$sAPIS_h = n \left| - \sum_{i=1}^h \sum_{j=1}^i (y_j - \hat{y}_j) \right| / \sum_{k=1}^n y_k, \quad (12)$$

The advantage of this metric is that it will penalise adequately zero forecasts, which otherwise may appear to perform well, due to the number of zero-demand periods in the out-of-sample period. Note that because of the absolute in the calculation the direction of the bias is lost, but its size is retained. Therefore, when summarising across time series the various measured biased will not cancel out.

The error measures above focus on point forecasts. Snyder et al. (2012) suggest to use prediction distribution based measures instead, the latter being more informative. However this is not done in this paper as direct calculation of inventory metrics is performed.

The focus of the accuracy metrics experiments is to identify whether any of the optimisation alternatives performs better. To this purpose considering the mean rank of each model in terms of PIS is useful. The model with the lowest mean rank performs the best. Given the sample size of the forecasts,

it is possible to assess the significance of the differences. The Friedman test, which is the non-parametric analogous to ANOVA, is used to identify whether at least one model is significantly different (Hollander and Wolfe, 1999). Consequently, the post-hoc Nemenyi test is used to assess the detailed ranking of the different models. The latter was shown to be a generalisation of the MCB test that was used to identify significantly better models in the M3 forecasting competition (Koning et al., 2005; Hibon et al., 2012).

5.3. Inventory Simulation and Metrics

The use of accuracy metrics to evaluate the performance of methods in intermittent demand data has been contested in the literature (for examples see: Gardner, 1990; Teunter and Duncan, 2009; Syntetos et al., 2010; Kourentzes, 2013). Inventory metrics are directly related to the operations and decision making process of the organisations. For this reason an inventory simulation is performed. The common in practice order-up-to policy (T, S) is used (Teunter and Sani, 2009a). Target service levels of 0.80, 0.90, 0.95 and 0.99 are considered and the inventory review period is set to one month. The initial stock is assumed to be equal to the order-up-to level S . In each period the actual demand is subtracted from the holding stock H . If the stock falls below S , then an order $S - H$ is placed. Lead times of one, three and five periods are considered. To simplify the simulation, it is assumed that any out-of-stocks are serviced by competitors and therefore are considered lost. The simulation follows the setup described by Kourentzes (2013).

The first simulation uses the real dataset. The first 13 observations are used to initialise the simulation and the remaining 11 are used for the empirical evaluation. This data split was suggested by Syntetos and Boylan (2006) using the same dataset, as well as the stock initialisation assumption. A second simulation is done on the 5,000 constructed time series. In this case there are no sample limitations. The first 96 observations of the simulated time series are used for fitting the models and initialising the simulation, the next 1,000 observations are used as a *burn-in* period and the last 240 observations are used to measure the inventory performance of the methods. The purpose of the *burn-in* period is to remove the impact of any stock initialisation values and advance the simulation to a realistic state regarding level of stock and orders. Inventory metrics are not measured during the *burn-in* period.

The advantage of using the real data is clear. On the other hand, with the limited history of the real data it is not possible to use a *burn-in* period, thus the results are sensitive to the simulation initialisation assumptions. The advantage of using simulated data is that this limitation is overcome and it is trivial to collect large samples. Results from both datasets will be presented, as both setups have their own merits.

The holding stock and out-of-stocks are recorded for the out-of-sample simulation period. These are used to create trade-off curves between stock-holding and out-of-stock. Both are made scale independent by dividing them with the average in-sample demand to allow summarising the results across time series. Furthermore, service level α is tracked.

5.4. Optimisation Results

Here the out-of-sample results of the different optimisation setups are presented. Table 1 provides the mean ranks of the different cost functions across all methods ‘Overall’, *CRO*, *SBA*, *CRO* and *SBA* together ‘*CRO*, *SBA*’, *TSB* and *SES* according to PIS. Mean ranks based on MASE are very similar. The mean ranks are calculated across all time series and forecast horizons, resulting in 9,000 forecasts. The lower the value of the reported mean rank the better the performance of the method. The best performance for each method is highlighted in boldface. For all methods apart from *TSB* the MAR cost function is the best. For *TSB* the MSR is marginally better than MAR.

Table 1: Mean rank per cost function and forecasting model

	Overall	CRO	SBA	CRO, SBA	TSB	SES
MSE	3.012	2.990	3.045	3.018	2.989	2.988
MAE	3.042	2.979	3.073	3.026	3.106	2.965
PIS	3.215	3.335	3.117	3.226	3.172	3.275
MSR	2.870	2.854	2.890	2.872	2.864	2.910
MAR	2.861	2.841	2.876	2.858	2.869	2.862

Figure 1 provides a summary of the statistical tests for 5% significance level. For all methods the results of the cost functions are significantly different according to the Friedman test. For any results joined by a vertical line there is not enough evidence of significant differences according to the Nemenyi test. For example, in figure 1 for *CRO* there is no evidence of differ-

ences between MAR and MSR, as well as between MAE and MSE. The mean rank for each cost function is also provided in the figure for easy reference.

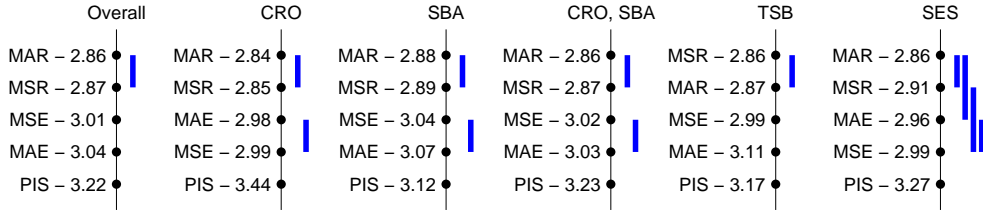


Figure 1: Mean rank and Nemeneyi test results for different cost functions per method.

We can see that in all cases MAR and MSR are in practice similar, but perform significantly different from other cost functions. For all *CRO*, *SBA* and ‘*CRO*, *SBA*’ the use of MSE and MAE results in similar performance, while this is not true for *TSB*. This is reflected in the *Overall* results as well. The behaviour of *SES* is quite distinct. Although MAR again ranks highly and significantly different from all other cost functions but MSR, the latter performs similarly to MAE. The PIS metric, although intuitively appealing since it uses cumulative errors, it does not perform well as a cost function for any of the methods. However, it is useful to keep in mind that minimising the absolute PIS, will not provide the minimum error fit, as it is a bias measure.

The similarity of MSR and MAR can be explained by the nature of r_i . The rate error is already smooth, without extreme outliers, and therefore considering either squared or absolute errors provides similar performance. We can conclude that the proposed MAR and MSR are useful for optimising the investigated methods. Hereafter, to avoid clutter in the results, only the results based on MAR optimisation will be presented.

Table 2: Mean rank of optimisation setup alternatives

Type	No. of parameters	Optimise \mathbf{y}_0	CRO, SBA	CRO	SBA	TSB
1 0	1	No	2.510	2.528	2.493	-
1 1	1	Yes	2.467	2.476	2.458	-
2 0	2	No	2.560	2.550	2.571	1.623
2 1	2	Yes	2.462	2.446	2.478	1.377

Table 2 provides results on the remaining of the optimisation setups, i.e. the number of parameters to be optimised and whether to optimise the vector of initialisation values \mathbf{y}_0 . The best mean PIS rank for each method is

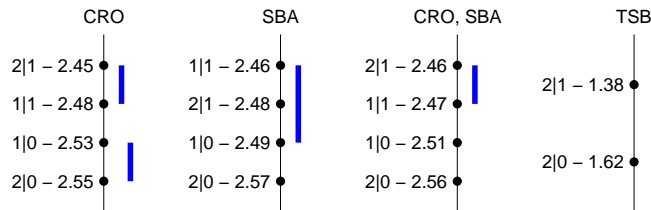


Figure 2: Mean rank and Nemenyi test results for different optimisation setups per method.

highlighted in boldface. In all cases optimising \mathbf{y}_0 is beneficial. This is in agreement with conventional exponential smoothing modelling for fast moving items (Hyndman et al., 2008). Considering that intermittent demand series are typically short and that low smoothing parameters are advisable, the forecasts can be substantially influenced by the initialisation values. Optimising them avoids using ad-hoc values, which may not be representative of the series and can lead to poor forecasts.

For *CRO* and ‘*CRO, SBA*’ optimising the non-zero demand and the inter-demand intervals separately is beneficial. However, consulting figure 2 we can see that this is not significantly different from optimising a single parameter, based on the results of the Nemenyi test. For *SBA* not using optimised initial values and a single smoothing parameter is not significantly different as well. Figure 2 follows the same naming convention as column ‘*Type*’ in table 2.

Given that from a theoretical point of view there is no motivation in updating both \hat{x}_t and \hat{z}_t at the same rate (Snyder, 2002), it makes sense to prefer the flexibility of using separate smoothing parameters. Therefore, based on these results, it is suggested to use two smoothing parameters and optimal initialisation values for all intermittent demand methods considered.

In the literature the smoothing parameters of intermittent demand methods are typically small (Syntetos and Boylan, 2005; Teunter et al., 2011). It might be beneficial to translate this into upper bounds for parameter optimisation. In figure 3 this option is explored. The parameters are optimised with upper bounds from 0.1 up to 1. The average sAPIS across all forecast horizons is plotted, across all time series, for *CRO*, *SBA* and *TBS* separately. The results suggest that for both *CRO* and *SBA* imposing lower upper bounds, for the optimisation of the smoothing parameters, does not provide any advantages. For *TBS* the opposite is revealed. Very low upper bounds can harm the accuracy of the method substantially. Based on these

results, it is suggested to use only the natural upper bound of 1 for parameter optimisation.

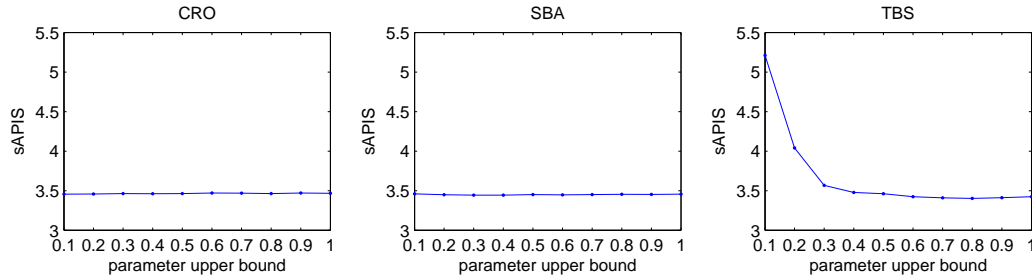


Figure 3: sAPIS for optimal parameters with upper bounds from 0.1 to 1.

To further understand this result, figure 4 provides boxplots for the optimal parameters of *CRO*, *SBA* and *TBS*, as identified by MAR. Most of the parameters are within the region suggested by the literature, with only some outlying values. The first and third quartiles of the boxplots broadly match the range of identified best parameters by Babai et al. (2011) and Petropoulos et al. (2013), who used the same dataset. Note that in these studies the same parameters were applied to all time series in the dataset. The parameters for *CRO* and *SBA* demonstrate a very similar behaviour, while *TBS* parameters are generally lower with less outliers. The reasonable optimal parameters occur only in the case of MAR and MSR. For the latter the parameter boxplots are very similar to figure 4. Conventional cost functions, such as MSE and MAE result in substantially higher smoothing parameters, as illustrated in figure 5 for MAE. MSE parameter values are similar. These do not follow the recommendations of the literature. The smoothing parameters are often quite high, resulting in very sensitive models that do not perform well for intermittent data. This explains the poor ranking of such metrics demonstrated in table 1 and figure 1. The absolute PIS cost function resulted in even higher smoothing parameters.

This illustrates the success of the proposed MAR and MSR cost functions for the case of intermittent demand forecasting. Due to their nature, the resulting parameters are low, in agreement with the literature, and consequently the different forecasting methods perform well. This is found to be impossible with conventional cost functions, as well as with PIS.

When comparing the optimal parameters with a-priori selected parameters the good performance of MAR and MSR holds. All methods are simu-

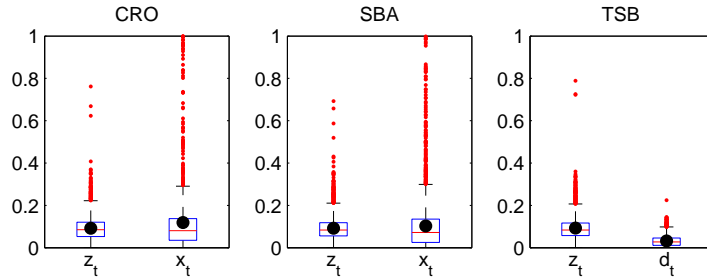


Figure 4: Boxplots of the optimal parameters for updating z_t and x_t for the different methods based on MAR. The average in each case is plotted with a dot (\bullet).

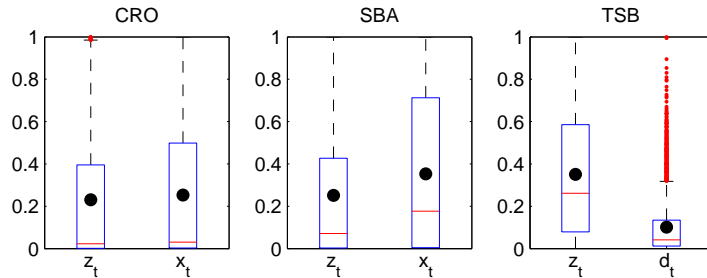


Figure 5: Boxplots of the optimal parameters for updating z_t and x_t for the different methods based on MAE. The average in each case is plotted with a dot (\bullet).

lated using fixed parameters of 0.05, 0.10, 0.15 and 0.20, following the suggestions by Syntetos and Boylan (2005), who used the same dataset. The comparisons indicate that for *CRO* and *TSB* the optimal parameters are significantly better in all cases, while for *SBA* parameters of 0.15 and 0.20 provide less accurate forecasts, yet not significantly different. Detailed results are not provided due to space considerations. Parameter optimisation achieved significantly better performance for both *CRO* and *TSB*, and equivalent performance for *SBA*.

The inventory metrics results for the real and the simulated datasets are summarised in figures 6 and 7 respectively. The figures refer to the results for the lead time of three periods. The results for one and five periods are similar with less or more exaggerated differences respectively and are not reported to conserve space. The different forecasting methods are reported separately in each column. The target and realised service level α are provided on the top

row of plots. The trade-off curves between scaled holding stock and scaled backlog are provided in the lower row of plots. To facilitate the matching of the target service levels with the trade-off curves the different levels are indicatively marked for one of the curves. Ideally a method should achieve the target service level with the minimal amount of cost. In other words, the trade-off curves that are closer to the origin of the axes dominate other trade-off curves, as they achieve lower stock-outs while requiring less holding stock.

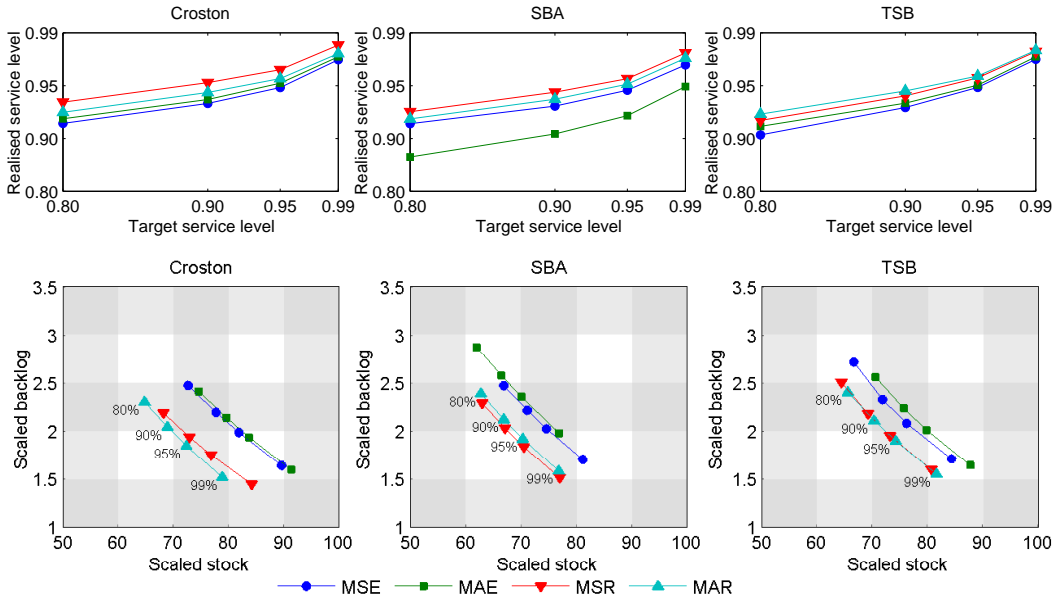


Figure 6: Service level and scaled stock-backlog trade-off curves for real dataset.

Considering the inventory simulation results on the real dataset (figure 6), for all *CRO*, *SBA* and *TSB* the MAR and MSR behave similarly. They consistently achieve higher service levels than MSE and MAE, while requiring less stock. Their trade-off curves dominate those of MSE and MAE. The differences between MSE and MAE are small, with the exception of the poor service levels achieved by MAE in the case of *SBA*.

Considering the results from the simulated time series in figure 7 the good performance of MAR and MSR is verified again. They consistently, across all models, achieve higher service levels while using less stock. Furthermore the differences between MAR and MSR are very small. Interestingly, all three *CRO*, *SBA* and *TSB* perform very similarly when optimised using MAR. On

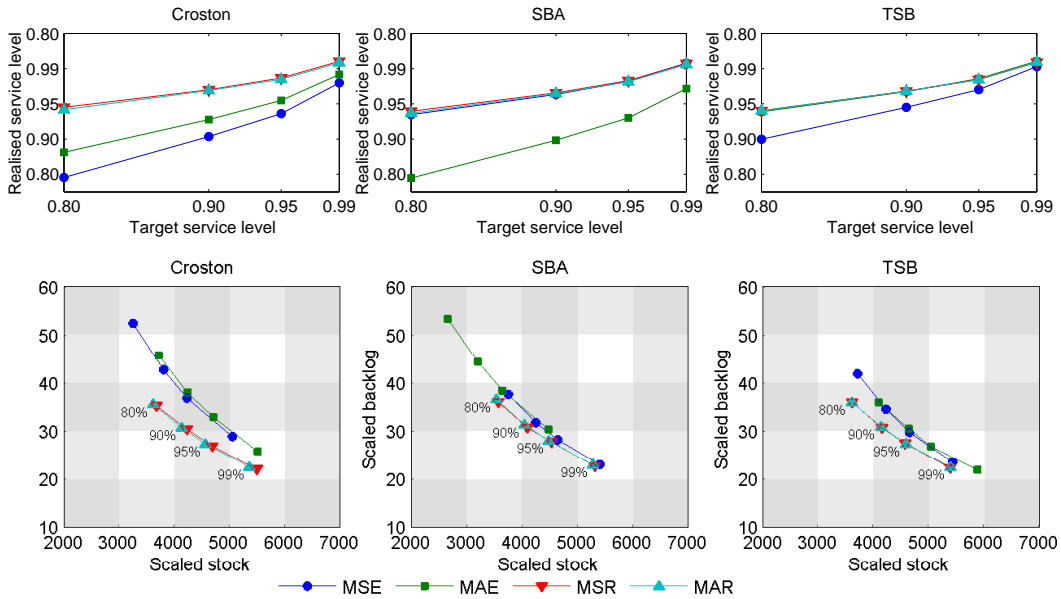


Figure 7: Service level and scaled stock-backlog trade-off curves for simulated dataset.

the other hand, MSE and MAE have rather unstable behaviour, reflecting the findings in figure 5. The trade-off curves of MSR and MAR dominate their MSE and MAE counterparts in all cases. It is interesting to observe that MAE performs particularly poorly in optimising *SBA* again.

Apart from better accuracy and inventory performance, another advantage of the optimally selected parameters, using the proposed cost functions, is that model parametrisation for intermittent demand is simplified considerably. There is no need to try several ad-hoc parameters and evaluate them using out-of-sample bias and accuracy metrics or inventory simulations to pick the best. This has direct implications for improving the automation of Croston’s method and its variants in forecasting systems and software.

5.5. Model Selection Results

Based on the findings in section 5.4 we can investigate whether it is possible to select between the different models for each time series, based on the outlined error metrics. All models reported hereafter are optimised using MAR. The in-sample fit of the optimised models using MSE, MAE, PIS, MSR and MAR is recorded and the model that gives the lowest error is selected for each time series. Four scenarios are considered. In the first one the model

pool includes only *CRO* and *SBA*. The second one includes additionally *TSB* and the third one *SES*, while the fourth one includes all methods.

Table 3: sAPIS and MASE results for the different in-sample error model selection

Error	sAPIS				MASE			
	t+1	t+3	t+5	Rank	t+1	t+3	t+5	Rank
Scenario 1: CRO, SBA								
MSE	0.747	3.105	6.529	3.002	0.851	0.880	0.912	3.008
MAE	0.747	3.103	6.531	3.012	0.851	0.880	0.912	2.985
PIS	0.747	3.102	6.535	2.994	0.851	0.882	0.913	3.004
MSR	0.747	3.098	6.513	2.996	0.851	0.880	0.912	3.007
MAR	0.747	3.097	6.514	2.997	0.851	0.879	0.912	2.995
Scenario 2: CRO, SBA, TSB								
MSE	0.743 [†]	3.083	6.488	2.991	0.847 [†]	0.878	0.910	2.981
MAE	0.743 [†]	3.087	6.498	3.001	0.848	0.879	0.909 [†]	2.979
PIS	0.743 [†]	3.081	6.489	2.988	0.848	0.879	0.911	3.005
MSR	0.745	3.084	6.483	2.998	0.849	0.878	0.911	3.013
MAR	0.746	3.090	6.490	3.022	0.850	0.878	0.911	3.022
Scenario 3: CRO, SBA, SES								
MSE	0.746	3.099	6.524	3.015	0.850	0.879	0.911	3.005
MAE	0.745	3.097	6.522	2.997	0.849	0.879	0.911	2.977
PIS	0.745	3.089	6.514	2.979	0.849	0.880	0.911	3.008
MSR	0.746	3.089	6.498	2.991	0.850	0.878	0.911	3.001
MAR	0.747	3.095	6.515	3.018	0.851	0.879	0.912	3.009
Scenario 4: CRO, SBA, TSB, SES								
MSE	0.743 [†]	3.085	6.496	2.998	0.848	0.878	0.910	2.992
MAE	0.743 [†]	3.087	6.500	3.006	0.847 [†]	0.878	0.910	2.979
PIS	0.745	3.082	6.499	3.009	0.849	0.880	0.911	3.030
MSR	0.744	3.079 [†]	6.475 [†]	2.982	0.848	0.877 [†]	0.910	2.988
MAR	0.745	3.086	6.482	3.006	0.849	0.877 [†]	0.911	3.010

[†] Best result across all model pools for that horizon.

The results of the evaluation, for each horizon, for both sAPIS and MASE are provided in table 3. The selection criteria for each forecast horizon that provides the best accuracy for each scenario is highlighted in boldface. The result that is best across all model pools is denoted with a [†]. The striking feature of the results are the small differences. Mean PIS and MASE ranks across all forecast horizons are provided. These highlight further the small

magnitude of the differences. The Friedman test indicates that for each case there are no statistically significant differences, therefore in practice there is no evidence to prefer any of the selection criteria.

The number of models considered in the pool does not alter this finding. On the other hand, the type of models considered in the pool seems to affect the overall accuracy, albeit marginally again. Using *TSB* and *SES* seems to be better than using just *CRO* and *SBA*. However, these results do not indicate whether model selection is beneficial at all.

Table 4 presents the sAPIS and MASE results, across the different forecast horizons, for individual methods and pools of methods, as selected by MAR. The most accurate results are highlighted in boldface. In addition to the in-sample model selection schemes, as a benchmark, the item classification proposed by Syntetos et al. (2005) as refined by Kostenko and Hyndman (2006) is used. The refined classification was shown to perform better than the original (Heinecke et al., 2011). The classification is based on the average inter-demand interval p of each time series and the squared coefficient of variation of the demand v . Although the original classification was identifying four regions, the refined version separates the space in only two regions. Any item with $v > 2 - (3/2)p$ should be modelled with *SBA* and the rest with *CRO*. For this dataset this indicates that 616 items are appropriate for *CRO* and that 2,384 should be modelled with *SBA*. The resulting model selection that follows from this classification is referred to as ‘*KH select*’.

Table 4: Individual and model selection sAPIS and MASE results

Model	sAPIS			MASE		
	t+1	t+3	t+5	t+1	t+3	t+5
CRO	0.749	3.110	6.563	0.852	0.882	0.914
SBA	0.746	3.100	6.521	0.851	0.879	0.911
CRO, SBA	0.747	3.097	6.514	0.851	0.879	0.912
KH select	0.747	3.101	6.523	0.852	0.880	0.912
TSB	0.742	3.070	6.457	0.847	0.876	0.909
CRO, SBA, TSB	0.746	3.090	6.490	0.850	0.878	0.911
SES	0.742	3.076	6.494	0.845	0.878	0.911
CRO, SBA, SES	0.747	3.095	6.515	0.851	0.879	0.912
CRO, SBA, TSB, SES	0.745	3.086	6.482	0.849	0.877	0.911

The results between sAPIS and MASE are similar. The model selection that uses *CRO* and *SBA* is marginally better than ‘*KH select*’, but fails to

demonstrate significant gains over modelling all time series with either *CRO* or *SBA*. Once *TSB* is included in the model pool, the differences in the results increase, though not significantly. However, using only *TSB* to model all the time series performs the best. This is in agreement with the reported good performance of *TSB* by Teunter et al. (2011). If instead of *TSB*, *SES* is introduced in the model pool, smaller improvements are observed, explained by the poorer performance of *SES* on its own compared to *TSB*. On the other hand, *SES* outperforms both *CRO* and *SBA*. This is in contrast to the findings by Syntetos and Boylan (2005) and can be attributed to the optimal *SES* parameters that are used in this study. Wallström and Segerstedt (2010) also found evidence of good performance of *SES*, based on the PIS metric. Finally, if all four methods are considered in the model pool, the results of the model selection improve, but still ranks second, though closely, to *TSB*.

Table 5: Types of selected models

Model	% of models in the results			
	CRO	SBA	TBS	SES
CRO	100.00%	0.00%	0.00%	0.00%
SBA	0.00%	100.00%	0.00%	0.00%
CRO, SBA	48.87%	51.13%	0.00%	0.00%
KH select	20.53%	79.47%	0.00%	0.00%
TSB	0.00%	0.00%	100.00%	0.00%
CRO, SBA, TSB	24.60%	23.93%	51.47%	0.00%
SES	0.00%	0.00%	0.00%	100.00%
CRO, SBA, SES	39.40%	43.43%	0.00%	17.17%
CRO, SBA, TSB, SES	21.80%	21.50%	47.70%	9.00%

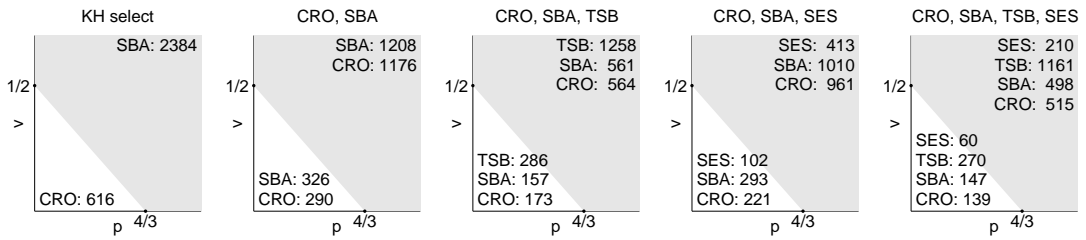


Figure 8: Selected models classified according to the KH scheme.

To further investigate the differences between the different model selection schemes table 5 provides the percentage contribution of each type of model

selected under each model pool. Comparing *CRO*, *SBA* with ‘*KH select*’ it is evident that the selected models differ significantly. Once *TSB* becomes available in the model pool then the majority of time series are modelled with it. Given its good performance this result makes sense. Inclusion of *SES* in the model pool does not alter the percentages significantly, indicating that from an in-sample fitting perspective *SES* does not score high.

Figure 8 identifies the number of time series that were modelled with each method, for each model selection case, when the time series are classified according to ‘*KH select*’. It is notable that the similarly performing ‘*CRO*, *SBA*’ has a substantially different distribution of models to ‘*KH select*’. In particular, more time series are modelled with *SBA* than with *CRO*, where the classification supports the latter. *TSB* and *SES*, when included in the model pool, are selected proportionally in both regions of the classification.

Considering the inventory metrics no important differences are observed between the different model pools. The same is true for ‘*KH select*’. In figures 6 and 7 we have already seen that once the different intermittent demand models are optimised by MAR they have similar inventory performance. These small differences naturally are retained when the model selection schemes are evaluated. Therefore, it is concluded that model selection using error metrics offers no significant differences to ‘*KH select*’, which in turn is only marginally different to the individual optimised models. This finding holds irrespective of which model selection criteria was used.

6. Conclusions

This paper investigated different alternatives for optimising intermittent demand methods, in particular the Croston’s method and the *SBA* approximation, as well as the more recent *TSB* method. It also explored different options for fitting error driven model selection for intermittent demand.

Two new cost functions, MSR and MAR, were proposed. They are based on the idea of using demand rates over time to calculate the error, rather than realised demand. The proposed cost functions were found to outperform MSE and MAE error metrics, as well as a cost function version of PIS. The resulting optimal parameters were found to be in agreement with the guidelines of the literature. This was not true for conventional error metrics, which demonstrated unstable behaviour, often resulting in large smoothing parameters. It is important to note that the proposed measures were developed having Croston’s method and its variants in mind, which provide a demand

rate forecast. Alternative forecasting methods that do not provide such forecasts may not benefit from MSR and MAR. This was reflected in the *SES* results.

Significant gains were identified by optimising the initialisation values of all methods. For both *CRO* and *SBA* optimising the non-zero demand and inter-demand interval estimates separately was found to be beneficial. These findings were verified using both out-of-sample accuracy and inventory metrics.

This work has useful implications for practitioners. Although there have been significant developments in intermittent demand methods, the discussion on how to best select model parameters has been limited, often based on ad-hoc selection. This is reflected in established forecasting and ERP systems, such as SAP Advanced Planner and Optimiser, which require the user to input the appropriate model parameters. This work provides easy to implement recommendations addressing this issue, allowing practitioners to identify optimal parameters for each of their time series. This can also help in achieving increased automation of the forecasts. Future work should explore the implications for the forecasting process in organisations and potential changes in the perceived quality of the forecasts by experts. This in turn may drive them to change their behaviour in adjusting the forecasts (Syntetos et al., 2009).

The results on model selection were not equally clear. No preference was identified for any of the in-sample fitting criteria for model selection purposes. All of them demonstrated similar performance. Their forecasting accuracy was marginally better than the ‘*KH select*’ item classification, when only *CRO* and *SBA* were used. Notably, the models selected for each time series were substantially different, raising questions regarding the practical usefulness of either selection schemes. Although classifying series as erratic, lumpy, smooth and intermittent is helpful for communicating their properties, the model selection results did not support clear differences in their modelling needs. This has been somewhat reflected in the ‘*KH select*’ classification, where only two classes are identified. Notably, none of the model pools managed to outperform substantially the individual models.

Although this paper contributes to the optimisation of intermittent demand methods, it does not provide a clear suggestion on model selection, as no superior approach was identified. The topic of model selection will become increasingly relevant with the recent surge of intermittent demand forecasting research. As new methods appear in the literature that are able

to outperform Croston's method and its derivatives, as well as overcome their limitations, a robust and valid model selection methodology is needed.

References

- Babai, M. Z., Ali, M. M., Nikolopoulos, K., 2012. Impact of temporal aggregation on stock control performance of intermittent demand estimators: Empirical analysis. *Omega* 40 (6), 713–721.
- Babai, M. Z., Syntetos, A. A., Teunter, R., 2011. Intermittent demand estimators: Empirical performance and sensitivity to the smoothing constants used. Working paper.
- Bacchetti, A., Saccani, N., 2012. Spare parts classification and demand forecasting for stock control: Investigating the gap between research and practice. *Omega* 40 (6), 722–737.
- Boylan, J., Syntetos, A., 2007. The accuracy of a modified Croston procedure. *International Journal of Production Economics* 107 (2), 511–517.
- Boylan, J. E., Syntetos, A. A., 2010. Spare parts management: a review of forecasting research and extensions. *IMA Journal of Management Mathematics* 21 (3), 227–237.
- Boylan, J. E., Syntetos, A. A., Karakostas, G. C., 2008. Classification for forecasting and stock control: a case study. *Journal of the Operational Research Society* 59 (4), 473–481.
- Croston, J. D., 1972. Forecasting and stock control for intermittent demands. *Operational Research Quarterly (1970-1977)* 23 (3), pp. 289–303.
- Eaves, A. H. C., Kingsman, B. G., 2004. Forecasting for the ordering and stock-holding of spare parts. *The Journal of the Operational Research Society* 55 (4), 431–437.
- Fildes, R., Hibon, M., Makridakis, S., Meade, N., 1998. Generalising about univariate forecasting methods: further empirical evidence. *International Journal of Forecasting* 14 (3), 339–358.
- Gardner, E. S., 2006. Exponential smoothing: The state of the art—part II. *International Journal of Forecasting* 22 (4), 637–666.

- Gardner, Jr., E. S., 1990. Evaluating forecast performance in an inventory control system. *Management Science* 36, 490–499.
- Ghobbar, A. A., Friend, C. H., December 2003. Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model. *Computers & Operations Research* 30, 2097–2114.
- Gutierrez, R. S., Solis, A. O., Mukhopadhyay, S., 2008. Lumpy demand forecasting using neural networks. *International Journal of Production Economics* 111 (2), 409–420.
- Heinecke, G., Syntetos, A., Wang, W., 2011. Forecasting-based SKU classification. *International Journal of Production Economics* forthcoming.
- Hibon, M., Crone, S. F., Kourentzes, N., 2012. Statistical significance of forecasting methods: An empirical evaluation of the robustness and interpretability of the MCB, ANOM and Nemenyi test. *Proceedings of the 32nd International Symposium on Forecasting*. <http://www.forecasters.org/proceedings12/HibonMicheleISF2012.pdf>.
- Hollander, M., Wolfe, D. A., 1999. *Nonparametric statistical methods*. Wiley-Interscience.
- Hyndman, R. J., Koehler, A. B., 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting* 22 (4), 679–688.
- Hyndman, R. J., Koehler, A. B., Ord, J. K., Snyder, R. D., 2008. *Forecasting with Exponential Smoothing: The State Space Approach*. Springer Verlag, Berlin.
- Hyndman, R. J., Koehler, A. B., Snyder, R. D., Grose, S., 2002. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting* 18 (3), 439–454.
- Johnston, F. R., Boylan, J. E., 1996. Forecasting for items with intermittent demand. *The Journal of the Operational Research Society* 47 (1), 113–121.
- Johnston, F. R., Boylan, J. E., Shale, E. A., 2003. An examination of the size of orders from customers, their characterisation and the implications for inventory control of slow moving items. *The Journal of the Operational Research Society* 54 (8), 833–837.

- Koning, A. J., Franses, P. H., Hibon, M., Stekler, H. O., 2005. The M3 competition: Statistical tests of the results. *International Journal of Forecasting* 21 (3), 397–409.
- Kostenko, A. V., Hyndman, R. J., 2006. A note on the categorization of demand patterns. *Journal of the Operational Research Society* 57, 1256–1257.
- Kourentzes, N., 2013. Intermittent demand forecasts with neural networks. *International Journal of Production Economics* 143, 198–206.
- Levén, E., Segerstedt, A., 2004. Inventory control with a modified Croston procedure and Erlang distribution. *International Journal of Production Economics* 90 (3), 361–367.
- Nikolopoulos, K., Syntetos, A., Boylan, J., Petropoulos, F., Assimakopoulos, V., 2010. An aggregate–disaggregate intermittent demand approach (adida) to forecasting: an empirical proposition and analysis. *Journal of the Operational Research Society* 62 (3), 544–554.
- Ord, J. K., Fildes, R., 2012. *Principles of Business Forecasting*. South-Western Cengage Learning, Mason, Ohio.
- Petropoulos, F., Nikolopoulos, K., Assimakopoulos, V., 2008. An evaluation of the accuracy of intermittent demand forecasting approaches. *Proceedings of the 28th International Symposium on Forecasting*. <http://www.forecasters.org/submissions08/PetropoulosFotiosISF2008.pdf>.
- Petropoulos, F., Nikolopoulos, K., Spithourakis, G., Assimakopoulos, V., 2013. Empirical heuristics for improving intermittent demand forecasting. *Industrial Management and Data Systems* 113 (5), 683–696.
- 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.
- Rao, A., 1973. A comment on: Forecasting and stock control for intermittent demands. *Operational Research Quarterly* 24, 639–640.
- 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.

- Sani, B., Kingsman, B. G., 1997. Selecting the best periodic inventory control and demand forecasting methods for low demand items. *Journal of the Operational Research Society* 48 (7), 700–713.
- Schultz, C. R., 1987. Forecasting and inventory control for sporadic demand under periodic review. *Journal of the Operational Research Society* 38 (5), 453–458.
- Shale, E. A., Boylan, J. E., Johnston, F. R., 2006. Forecasting for intermittent demand: the estimation of an unbiased average. *The Journal of the Operational Research Society* 57, 588–592.
- Shenstone, L., Hyndman, R. J., 2005. Stochastic models underlying croston’s method for intermittent demand forecasting. *Journal of Forecasting* 24 (6), 389–402.
- Snyder, R., 2002. Forecasting sales of slow and fast moving inventories. *European Journal of Operational Research* 140 (3), 684–699.
- Snyder, R. D., Ord, J. K., Beaumont, A., 2012. Forecasting the intermittent demand for slow-moving inventories: A modelling approach. *International Journal of Forecasting* 28 (2), 485–496.
- Spithourakis, G., Petropoulos, F., Nikolopoulos, K., Assimakopoulos, V., 2012. A systemic view of adida framework. *IMA Management Mathematics* forthcoming.
- Strijbosch, L. W. G., Syntetos, A. A., Boylan, J. E., Janssen, E., 2011. On the interaction between forecasting and stock control: the case of non-stationary demand. *International Journal of Production Economics* 133 (1), 470–480.
- Syntetos, A. A., Boylan, J. E., May 2001. On the bias of intermittent demand estimates. *International Journal of Production Economics* 71 (1-3), 457–466.
- Syntetos, A. A., Boylan, J. E., 2005. The accuracy of intermittent demand estimates. *International Journal of Forecasting* 21 (2), 303–314.
- Syntetos, A. A., Boylan, J. E., 2006. On the stock control performance of intermittent demand estimators. *International Journal of Production Economics* 103 (1), 36–47.

- Syntetos, A. A., Boylan, J. E., Croston, J. D., 2005. On the categorization of demand patterns. *Journal of the Operational Research Society* 56 (5), 495–503.
- Syntetos, A. A., Nikolopoulos, K., Boylan, J. E., 2010. Judging the judges through accuracy-implication metrics: The case of inventory forecasting. *International Journal of Forecasting* 26 (1), 134–143.
- Syntetos, A. A., Nikolopoulos, K., Boylan, J. E., Fildes, R., Goodwin, P., 2009. The effects of integrating management judgement into intermittent demand forecasts. *International Journal of Production Economics* 118 (1), 72–81.
- Tashman, L. J., Kruk, J. M., 1996. The use of protocols to select exponential smoothing procedures: A reconsideration of forecasting competitions. *International Journal of Forecasting* 12 (2), 235–253.
- Teunter, R., Sani, B., 2009a. Calculating order-up-to levels for products with intermittent demand. *International Journal of Production Economics* 118 (1), 82–86.
- Teunter, R., Sani, B., April 2009b. On the bias of croston’s forecasting method. *European Journal of Operational Research* 194 (1), 177–183.
- Teunter, R., Syntetos, A., Babai, M., 2010. Determining order-up-to levels under periodic review for compound binomial (intermittent) demand. *European Journal of Operational Research* 203 (3), 619–624.
- Teunter, R. H., Duncan, L., 2009. Forecasting intermittent demand: a comparative study. *The Journal of the Operational Research Society* 60, 321–329.
- Teunter, R. H., Syntetos, A. A., Babai, M. Z., 2011. Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research* 214 (3), 606–615.
- Wallström, P., Segerstedt, A., 2010. Evaluation of forecasting error measurements and techniques for intermittent demand. *International Journal of Production Economics* 128 (2), 625–636.

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.

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.