Intermittent demand forecasting: an empirical study on accuracy and the risk of obsolescence

(To appear in the *Int. Journal of Production Economics*)

Mohamed Zied Babai [1], Aris Syntetos [2], Ruud Teunter [3]
[1]Kedge Business School, France; [2]Cardiff University, UK; [3]University of Groningen, The Netherlands

Abstract

Intermittent demand items account collectively for considerable proportions of the total stock value of any organization. Forecasting the relevant inventory requirements constitutes a very difficult task and most work in this area is based on Croston's estimator that relies upon exponentially smoothed demand sizes and inter-demand intervals. This method has been shown to be biased and a number of variants have been introduced in the literature, including the recently proposed TSB method that updates the demand probability instead of the demand interval and in doing so reacts faster to decreasing demand. The TSB has been shown theoretically to be unbiased (for all points in time), but its empirical performance has not been investigated yet and this constitutes one of the objectives of our work. More generally, we explore the empirical performance of forecasting methods used in an intermittent demand context, paying particular attention to the effects and implications of the smoothing constant values employed for updating purposes. We do so by means of experimentation on large empirical datasets from the military sector and automotive industry. The results enable insights to be gained into the sensitivity of the various methods' forecasting and stock control performance to the smoothing constant values used. The paper concludes with an agenda for further research.

Keywords: Intermittent Demand, Forecasting, Smoothing Constants, Obsolescence, Empirical Analysis.

[1] Corresponding author : Dr. Mohamed Zied Babai
Kedge Business School
680 cours de la Libération, 33405 Talence Cedex, France
Tel: +33 5 56 84 55 55, e-mail: mohamed-zied.babai@kedgebs.com

1. Introduction

Intermittent demand for a product or spare part appears sporadically, with some time periods showing no demand at all. When demand occurs, the demand size may be constant or variable, perhaps highly so. Intermittent demand items are usually amongst the slower movers in any organisational setting. However, despite the comparatively low contribution to the total turnover, these slow moving Stock Keeping Units (SKUs) may constitute up to 60% of the total stock value (Johnston *et al.* 2003; Quintana and Leung 2007). In addition, spare parts that have become ubiquitous in modern societies, are almost invariably 'intermittently moving'. This is true both for engineering spares and service parts kept at the wholesaling and retailing level. Consequently, the stock-bases in the military context, process industries, aerospace, automotive and IT sectors are dominated by such items.

Most work on intermittent demand forecasting is based on Croston's influential article (Croston 1972) which for many years was neglected but in the last 15 years has seen 163 citations (Publish or Perish, PoP accessed May 26, 2014 - see also Fildes *et al.* 2008). The method was claimed to be unbiased but despite its theoretical superiority modest benefits were recorded in the literature when it was compared with simpler forecasting techniques, such as Simple Exponential Smoothing, SES (Willemain *et al.* 1994). Some empirical evidence even suggested losses in performance (Sani and Kingsman 1997). An earlier paper (Syntetos and Boylan 2001) showed Croston's method to be biased and since then some modifications to his method have been put forward in the academic literature.

Syntetos (2001) proposed an *exactly unbiased* modification of Croston's method. This procedure was further evaluated on real data (Teunter and Sani 2009) and it was found to perform very well. Syntetos and Boylan (2005) put forward an *approximately unbiased* procedure that nevertheless is characterised by a smaller parametric estimation error (variance of the estimates) than that associated with the previous method. This estimator is known in the literature as the SBA method (after Syntetos-Boylan Approximation). SBA has been independently assessed on 18,750 SKUs from the Royal Air Force (RAF, UK) and found to outperform Croston's estimator (Eaves and Kingsman 2004). Independent evidence in support of this method's utility has also been provided by Gutierrez *et al.* (2007) and Petropoulos *et al.* (2008). Teunter and Duncan (2008) discussed the application of the zero forecast method in an intermittent demand context (the forecasts of the demand are always equal to zero) (see also Chatfield and Hayya 2007). Although such an estimator makes little sense in a stock control context, it may indeed be used as a forecasting benchmark. Similarly, the Naïve forecast may also be used for benchmarking purposes; in this case the last actual demand (zero or not) becomes the forecast for the next time period. A review of developments on intermittent demand forecasting is given by Boylan and Syntetos (2010).

Teunter *et al.* (2011) recently suggested a new intuitively appealing forecasting procedure (called TSB after Teunter, Syntetos and Babai) that links naturally to the issue of inventory obsolescence. The performance of this method was assessed through an extensive simulation study on theoretically generated data and it was shown to compare very well to the other methods discussed in the literature. However, this method has never been evaluated on real data and this constitutes one of the objectives of this paper.

The purpose of this paper is two-fold: first, we empirically evaluate the performance of intermittent demand estimates (along the lines discussed by Syntetos and Boylan, 2005), considering also the TSB method that has never been tested before on real data. Second, we analyse the effect of the smoothing constants on the performance of different estimation procedures by means of an empirical sensitivity analysis. Note that separate smoothing constants are used for the sizes and intervals in the Croston-type approaches; therefore we provide in this paper the statistical properties of the estimators in this general case which has never been presented before in the literature dealing with intermittent demand.

The remainder of our paper is structured as follows. In the next section, we present the forecasting methods considered for the purposes of our work, followed by a summary of arguments on the smoothing constants that should be employed for forecasting purposes in Section 3. In Section 4 the statistical properties of the methods are discussed and linked to the issue of smoothing constants. Section 5 provides details related to the empirical investigation and the results of our study. The conclusions of our work are discussed in Section 6 along with suggestions for further research.

2. Forecasting methods

First we provide the notation used in our work followed by the description of the various estimation procedures considered.

2.1 Notation

The following notation is used in the remainder of this paper:

Y_t: Demand for an item in period t

Y_t': Estimate of mean demand per period made in period t for period $t+1$

z_t: Demand size in period t, when demand occurs, with a mean μ and variance σ^2

z_t': Estimate of mean demand size in period t

p_t: Indicator of demand occurrence in period t,

$$p_t = \begin{cases} 1 & \text{if demand occurrs at time } t \text{ (i.e. } Y_t > 0) \\ 0 & \text{otherwise} \end{cases}$$

$p_t^{'}$: Estimate of the probability of demand occurrence in period t

T_t: Actual demand interval in period t

$T_t^{'}$: Estimate of mean demand interval in period t

α , β: smoothing parameters ($0 \le \alpha, \beta \le 1$)

2.2 Estimators

Four Croston's type estimators are considered in our work, namely: Croston's method (Croston 1972), the SBA method (Syntetos and Boylan 2005), the SY method (after Syntetos, 2001) and the TSB method (after Teunter, Syntetos and Babai, 2011). We also consider the Simple Exponential Smoothing (SES) method, the zero forecast method and the Naïve method, the latter two serving as benchmarks. We remark that these methods can be applied for any (intermittent) demand process, both stationary and non-stationary, as will be discussed in Section 3.

- **Croston's method**

Croston's estimator relies upon separate exponentially smoothed estimates of the interval between consecutive demands and the size of demands. The estimates are updated only in periods with positive demand. We remark that Croston originally suggested using the same value for both smoothing constants, but others (e.g. Schultz, 1987) later argued that using different values can be beneficial. We consider the more general variant.

If $Y_t = 0: T_t^{'} = T_{t-1}^{'}$, $z_t^{'} = z_{t-1}^{'}$, $Y_t^{'} = Y_{t-1}^{'}$.

If $Y_t > 0: T_t^{'} = T_{t-1}^{'} + \beta(T_t - T_{t-1}^{'})$, $z_t^{'} = z_{t-1}^{'} + \alpha(Y_t - z_{t-1}^{'})$, $Y_t^{'} = z_t^{'} / T_t^{'}$.

- **The SBA method**

SBA relies upon the employment of a correction factor to the estimates produced by Croston's method to reduce the bias associated with those estimates.

If $Y_t = 0: T_t^{'} = T_{t-1}^{'}$, $z_t^{'} = z_{t-1}^{'}$, $Y_t^{'} = Y_{t-1}^{'}$.

If $Y_t > 0: T_t^{'} = T_{t-1}^{'} + \beta(T_t - T_{t-1}^{'})$, $z_t^{'} = z_{t-1}^{'} + \alpha(D_t - z_{t-1}^{'})$, $Y_t^{'} = (1 - \beta/2) z_t^{'} / T_t^{'}$.

- **The SY method**

The SY estimator also relies upon the employment of a correction factor to the estimates produced by Croston's method.

If $Y_t = 0: T_t^{'} = T_{t-1}^{'}$, $z_t^{'} = z_{t-1}^{'}$, $Y_t^{'} = Y_{t-1}^{'}$.

If $Y_t > 0: T_t^{'} = T_{t-1}^{'} + \beta(T_t - T_{t-1}^{'})$, $z_t^{'} = z_{t-1}^{'} + \alpha(D_t - z_{t-1}^{'})$, $Y_t^{'} = (1 - \beta/2) z_t^{'} / (T_t^{'} - \beta/2)$.

- **The TSB method**

The TSB estimator uses separate exponentially smoothed estimates of the demand probability and the demand size. The estimate of the probability of occurrence is updated every time period. The estimate of the demand size is updated only at the end of periods with positive demand.

If $Y_t = 0: p_t^{'} = p_{t-1}^{'} + \beta(0 - p_{t-1}^{'}), \quad z_t^{'} = z_{t-1}^{'}, \quad Y_t^{'} = p_t^{'} z_t^{'}$

If $Y_t > 0: p_t^{'} = p_{t-1}^{'} + \beta(1 - p_{t-1}^{'}), \quad z_t^{'} = z_{t-1}^{'} + \alpha(z_t - z_{t-1}^{'}), \quad Y_t^{'} = p_t^{'} z_t^{'}$

Please note that, although the TSB method is unbiased when all points in time are considered, it does suffer from a statistical bias for issue points only.

- **SES method**

$$Y_t^{'} = Y_{t-1}^{'} + \alpha(Y_t - Y_{t-1}^{'}) \text{ for all } t$$

- **Zero forecast**

$$Y_t^{'} = 0 \text{ for all } t$$

- **Naïve Forecast**

$$Y_t^{'} = Y_{t-1} \text{ for all } t$$

Please note that the Naïve method may be seen as a special case of the TSB method and of the SES method with the smoothing constant(s) set to 1.

3. Smoothing constants

Croston's method and its variants were originally proposed for a model that assumes stationarity, both for the demand sizes and demand intervals. However, forecasts are produced through the employment of SES procedures, implying a non-stationary demand process. This renders the model and method inconsistent. Shenstone and Hyndman (2005) examined the models that may be underlying Croston's estimator. In their paper they commented on the wide prediction intervals that arise for non-stationary models and recommended that stationary models should be reconsidered. However, they concluded, "*...the possible models underlying Croston's and related methods must be non-stationary and defined on a continuous sample space. For Croston's original method, the sample space for the underlying model included negative values. This is inconsistent with reality that demand is always non-negative*, Shenstone and Hyndman (op. cit), pp. 389-390".

This does not mean that Croston's method and its variants are not useful. Such methods do constitute the current state of the art in intermittent demand parametric forecasting. In addition, three arguments may be put forward to demonstrate that the inconsistency between Croston's model and the employment of SES estimates is not very restrictive (at least within the context of our research): a practical, a statistical and a methodological argument.

- **A practical argument**

By using the SES estimator the forecast is always alert for any changes in the situation, which would reveal them-selves through the forecast error (Johnston and Boylan 1994). This is exactly the feature required if we are to forecast a constant mean model which is valid locally rather than globally. Unless we can be absolutely sure that the Stationary Mean Model (SMM) has applied in the past and will continue to apply in the future, it is 'safer' to use the weighted mean (SES) rather than the ordinary mean.

- **A statistical argument**

By using SES we account for any small undetected changes in level. The changes cannot be detected because of the very nature of the data (few demand occurrences). To strengthen our point: unless we have long demand histories of, say, 10 years, we cannot detect a Steady State Model (SSM – that implies SES). That is, we cannot tell that the mean changes over time because we can hardly observe that mean. Use of SES ensures that any real changes in the underlying demand level that cannot be practically detected (and theoretically be accounted for) because of the scarcity of data, will be reflected in the demand estimates.

- **A methodological argument**

Theoretical results have previously been generated (Syntetos, 2001), assuming that SES methods are used in conjunction with the stationary mean model assumption. The results were tested on stationary theoretically generated data showing a very robust behaviour in the sense that SES performs well for a variety of models, not just the one for which it is theoretically optimal. Teunter *et al*. (2011) also tested the Croston method and its variants on theoretically generated non-stationary data, and their results suggest that the TSB method is fairly robust against non-stationarity, whereas the performance of the other methods is more affected, especially for decreasing demand. In this paper, the various methods are tested on real intermittent demand data that may or may not be stationary. The empirical results will allow us to gain insight as to what extent the model-estimator inconsistency is reflected in real world situations.

Next we discuss the (ranges of) values for the smoothing constants that have been suggested in the literature, where we remark that most authors set $\beta = \alpha$. Burgin and Wild (1967) found that $\alpha = 0.2$ is suitable for most weekly data but they implicitly recommended lower α values for slow movers. Croston (1972), and later Gutierrez *et al*. (2007) and Petropoulos *et al.* (2008), recommended the use of α values in the range 0.05 - 0.2, when demand is intermittent. Croston also demonstrated, numerically, that for deterministic demands of size μ, occurring every p (constant) review intervals, the SES forecast error reduces with the smoothing constant value. Finally, he suggested that higher values of α, in the range of 0.2 - 0.3, may be found necessary only if there is a high proportion of items that is known to be non-stationary. Sani (1995) and Johnston and Boylan (1996) reported results on a comparative studies for $\alpha = 0.15$, whereas Willemain *et al.* (1994) set $\alpha = 0.1$. Since one of our goals is to assess the effect of the smoothing constants on the performance of forecasting methods by using data that is thought to be non-stationary (see sub-section 5.1), we consider the 'wide' range of 0.05 - 0.3 in this study.

4. Linking performance to the smoothing constants for stationary demand

4.1 Forecasting methods' properties

In Table 1, the statistical properties (expected estimate and variance of estimates) of the estimators for all points in time discussed in Section 2 are presented in detail, under the assumption of independent stationary demand processes for the demand size (following any theoretical distribution) and the demand probability (following a Bernoulli distribution, resulting in geometrically distributed demand intervals). See also Syntetos, (2001), Syntetos and Boylan (2010) and Teunter *et al.* (2011).

INSERT TABLE 1 ABOUT HERE

4.2 Effect of the smoothing constants

Considering the results presented in Table 1, it is easy to see that the Croston's method is positively biased and the percentage bias increases with the smoothing constant, whereas, the SY and the TSB estimators are exactly unbiased (the latter for all points in time only). The SBA estimate is slightly negatively biased and the absolute value of the percentage bias increases with the smoothing constant. The absolute bias can easily be shown to be generally lower than that of the Croston method.

Moreover, a comparison of the Croston's and SBA variances shows that the variance of the SBA estimates is lower than that related to Croston since the former is obtained by multiplying the latter by the square of a coefficient smaller than 1. SBA is also associated with a variance smaller than that characterising the SY method. The variance of the Naïve method is equal to the variance of the demand, so if the demand is lumpy, the Naïve is expected to provide a poor performance as a benchmark, but if demand is associated with a low variability of the sizes, the variance of the Naïve estimates should be low as well.

Let us denote the β of the Croston and TSB methods by β_C and β_T, respectively, for just this paragraph. A comparison of the TSB and Croston variances for the same value of α shows that they are equal if $\beta_T/(2-\beta_T) = p\beta_C/(2-\beta_C)$. Since smoothing constants are usually not set larger than 0.3 (see Section 3) this condition is almost equivalent to $\beta_T = p\beta_C$, which has the following intuitive explanation. The TSB method updates the demand probability $1/p$ times as often as the Croston method (that updates its inverse, the demand interval). By setting $\beta_T = p\beta_C$, the total effect of $1/p$ updates is approximately $\beta_T / p = \beta_C$. In other words, the TSB method then updates the demand probability at approximately the same speed as the Croston method. For this reason, we expect the optimized β_T to generally be smaller than the optimized β_C, especially for SKUs with very infrequent demand.

Please bear in mind that the above discussion on optimizing the values for the smoothing constants is theoretical as, in practice, we are unsure whether the demand process is stationary. Indeed, this is the rationale for constantly updating forecasts using exponential smoothing, as discussed before. In the next section, we therefore continue with an empirical investigation.

5. Empirical investigation

5.1 Empirical data

Two datasets are used. Dataset 1 consists of the individual demand histories of 3,000 SKUs covering 24 consecutive months from the automotive industry. Dataset 2 consists of the individual monthly demand histories of 5,000 SKUs over 7 years (84 monthly demand observations, from Jan. 1996 to Dec. 2002 inclusive) from the Royal Air Force (RAF). Descriptive statistics (across all SKUs) for both datasets are given in Tables 2 and 3.

INSERT TABLES 2 AND 3 ABOUT HERE

Note from Tables 2 and 3 that the level of intermittency and lumpiness are quite different among the two datasets. The automotive industry dataset consists of low demand items with low degree of intermittence. The maximum of the mean demand intervals is equal to 2 with a 75th percentile equal to 1.4 and the maximum mean demand size is equal to 193. Dataset 2 consists of items with a very high degree of lumpiness. The maximum of the mean demand intervals is equal to 24 with a 75th percentile equal to 11. The minimum mean demand size is 1 and the maximum mean demand size is equal to 668.

We have also identified (and separately analysed) those SKUs in both datasets that are associated with a decreasing demand pattern. This was achieved as follows: in the case of dataset 1, the entire demand history was divided in 3 equal non-overlapping blocks of 8 periods and the mean demand was recorded over every such block. The SKUs related to a strict decrease of the mean from the older time block to the most recent were selected. The same procedure was used for dataset 2, in which case, 4 blocks were used (each consisting of 21 periods). Teunter et al. (2011) argued and showed using simulation that the TSB performs particularly well for such SKUs (with a high inventory obsolescence risk), an assertion that will be tested empirically using the resulting 469 SKUs for dataset 1 and 190 SKUs for dataset 2.

5.2 Accuracy measuring and simulation details

We assess the forecasting performance of the different estimators through three accuracy measures, namely: the Mean Error (ME), the Mean Square Error (MSE) and the Mean Absolute Scaled Error (MASE). These accuracy measures are defined respectively as follows:

$$ME = \frac{1}{n}\sum_{t=1}^{n}(Y_t^{'} - Y_t), \qquad MSE = \frac{1}{n}\sum_{t=1}^{n}(Y_t^{'} - Y_t)^2, \qquad MASE = \frac{\frac{1}{n}\sum_{t=1}^{n}\left|Y_t^{'} - Y_t\right|}{\frac{1}{n_1 - 1}\sum_{i=2}^{n_1}\left|Y_i - Y_{i+1}\right|},$$

where n_1 is the length of the in-sample demand history and n is the length of the out-of-sample demand history. Following their calculation on a specific series all accuracy measures are then arithmetically averaged across series (SKUs).

The ME provides a way to determine whether specific methods consistently underestimate or overestimate the level of demand, i.e. to indicate whether the estimation procedures under concern are biased, and if so in which direction. The MSE calculation is similar to the statistical measure of the variance of forecast errors but not quite the same since theoretically bias is also taken into account (should it exist). The MASE was proposed by Hyndman and Koehler (2006) as a generally applicable scale-free measurement of forecast accuracy where the scaling of the errors is based on the in-sample Mean Absolute Error from the naïve forecast method.

To evaluate the performance of the methods, we split the demand history available for each SKU into two parts. The 1st part (i.e. within-sample) is used in order to initialize the estimates of the level and variance of demand. (For that purpose, 13 and 36 periods are considered for datasets 1 and 2 respectively). The 2nd part (out-of-sample) is used for the out-of-sample generation of results and evaluation of performance. The Croston, SBA and SY initial estimates are calculated at the end of the initialisation part by considering the average demand sizes and intervals. If no demand occurs in the initialisation part, the inter-demand interval estimate is set to 13 or 36 (for dataset 1 or dataset 2, respectively), whereas the demand size is assigned an initial estimate of 1. The TSB estimate is initialised by using the same procedure described above for the other Croston type approaches, with the only difference that the demand occurrence probability is initialised with its average value over the initialisation part. The SES estimate is initialised at the end of the initialisation part by considering the average demand per period over that part. For all methods, the first MSE estimate at the end of the initialisation part is computed as the square error between the initial estimate of the demand level and the actual demand.

In order to generate results, we vary the smoothing constants from 0.05 to 0.30 in steps of 0.05. For the demand probability smoothing constant of TSB, we additionally consider values from 0.01 to 0.04 in steps of 0.01. This is in line with what has been suggested in the literature, as discussed in Sections 3 and 4.

5.3 Empirical results on forecast accuracy

We first present results for the complete datasets. In Tables 4 and 5, we show the average performance for all considered performance measures, when the smoothing constants are selected on the basis of providing the smallest MSE averaged across all SKUs. Later, we will discuss optimization of smoothing constants at the individual SKU level. We remark that selecting values for smoothing constants based on MASE instead of MSE produces similar results, which are therefore not presented separately.

INSERT <u>TABLES 4 - 5</u> ABOUT HERE

The results for dataset 1 are generally in line with what has been suggested in the literature and discussed in Sections 3 and 4. The Croston method is considerably positively biased, SBA has a smaller negative bias, and the theoretically unbiased estimators (SES, SY, TSB, Naïve) have even lower absolute bias. The Zero Forecast obviously has the largest (negative) bias. All four Croston-type approaches show similar performance with respect to MSE and MASE. As expected, the Naïve and Zero Forecast methods perform much worse for both these measures.

The results for dataset 2 are more surprising. The Naïve method is the one with the bias closest to zero, although it still performs worst with respect to both MSE and MASE. The Zero Forecast has the best MASE performance and a reasonable MSE performance, but it is of course very heavily biased.

The optimal values of the smoothing constants are also quite stable over the Croston, SY and SBA methods for each dataset. As expected, TSB has a relatively small optimal smoothing constant for the demand probability, since it is updated in every period. Overall, smoothing constants are larger for dataset 2, indicating that the demand is less stationary.

Before discussing results for the selection of SKUs with decreasing mean demand, we first consider optimization of smoothing constants at the SKU level. As it turns out, doing so leads to much worse performance than optimization across SKUs. This becomes clear from comparing the results for dataset 2 in Table 6 to those in Table 5. (The results for dataset 1 also worsen considerably.)

INSERT TABLE 6 ABOUT HERE

Apparently, there simply is not enough data available at the SKU level for these intermittent demand items to optimize the smoothing constants on an individual SKU basis. In what remains, we will therefore restrict to optimization across SKUs. Tables 7 and 8 present those results for the SKUs with decreasing demand. (See Section 5.1 for the selection procedure.)

INSERT TABLES 7 AND 8 ABOUT HERE

The results are generally in line with those reported for the complete datasets. Somewhat surprisingly, the TSB method does not perform considerably better than Croston or other variants, despite the fact that it can adjust quicker to the decreasing demand by lowering the demand period in periods without demand. Indeed, SBA is the clear winner for dataset 1 when it comes to the MSE.

Please note that, the difference between the empirical findings of this work and those by Teunter et al. (2011) may be partly attributed to the empirical nature of the data that do not necessarily reflect all the assumptions (i.e. demand stationarity, distributions of the demand sizes and demand occurrence, etc.) upon which the TSB method was developed and evaluated in the study by Teunter et al. (2011).

5.4 Empirical results on inventory control

As discussed in Section 1, most of the comparative forecasting studies for intermittent demand focus solely on forecast accuracy rather than inventory control implications. In particular, the TSB has never before been evaluated in an inventory control context, although its proposal was partially motivated by the ability to adapt to changing and in particular decreasing demand, thereby possibly preventing obsolescence.

Our empirical setup is as follows. We consider a base stock inventory control strategy which is suitable for slow moving, intermittent demand items (Axsäter 2006). Let us denote the lead time by L periods. The order of events in a period is as follows: observe demand, receive an order (placed L periods ago), place a new order. It is well known from the inventory control literature (Axsäter 2006) that the base stock level should include the expected demand and safety stock for L+1 periods (given this order of events). Therefore, we calculate the base stock level at the end of each period as: $(L+1)Y_t^{'} + k\sqrt{(L+1)MSE_t^{'}}$, where $MSE_t^{'}$ is the updated mean square error (where we use a smoothing constant of 0.25) and k can be interpreted as the safety stock factor. By varying the safety factor, we obtain solutions that differ in holding and backordering volumes. It also allows us to draw efficiency curves of holding versus backordering, which can then be compared across the different forecasting methods. This enables us to make fair comparisons despite the fact that the different forecasting methods do not result in exactly the same service performance (for the same value of the safety factor). The efficiency curves obtained for the complete datasets 1 and 2 are reported in Figures 1 and 2, respectively. The results obtained after the selection of SKUs with decreasing mean demand are given in Figure 3 and Figure 4. Please note that Naive and the Zero forecast stock control results are much less good, and excluded from Figures 1-4 to allow for a clearer graphical comparison of the other methods.

INSERT <u>FIGURES 1, 2, 3 AND 4</u> ABOUT HERE

Figure 1 shows that the SES and all Croston-type approaches provide similar inventory performance as their efficiency curves almost overlap. This is in line with the results of Section 5.3, showing not much difference between the ME and MSE results for dataset 1. It is hence difficult to spot the outperformance of any forecasting method.

However, Figure 2 shows that Croston's method clearly outperforms all other methods. Apparently, despite the considerable bias of this method that exists both theoretically (see Section 4) and empirically (Section 5.3), the Croston method achieves a better inventory holding-backlog efficiency.

Figure 4 shows that, even for the SKUs with decreasing demand from dataset 2, Croston's method still outperforms the other methods. This is again in line with the results from Section 5.3, but it remains a surprising outcome that may be attributed to the very high degree of lumpiness associated with dataset 2.

Returning to dataset 1, Figure 3 shows that for the SKUs with decreasing demand from this dataset, the TSB method provides the best inventory performance. So, these results *are* in line with those from the simulation experiment on generated data conducted by Teunter et al. (2011), whereas dataset 2 leads to different empirical results.

6. Conclusions and extensions

Most of the work on intermittent demand forecasting in the last fifteen years has been based on Croston's method. Many modifications have been proposed in the academic literature to improve its performance and especially to reduce the bias of this method. The original method and variants all use smoothing constants to update the estimates. A number of authors have recommended (ranges of) values for the smoothing constants (under certain conditions). However, empirical justification has been limited.

In this paper, we analysed the performance of intermittent demand estimation methods by considering the bias (ME) of the mean, the variance of the estimates – as measured through the MSE, and a scale-free measure provided by the MASE. First, for fixed smoothing constants and under the assumption of stationary demand, exact expressions for the first two moments of all estimators were given and compared, leading to a theoretical discussion of the methods' properties. Then, an empirical investigation on two real-life datasets was performed including a sensitivity analysis of the error measures with respect to the smoothing constants.

The two datasets that we used lead to different results. Those for dataset 1 (with relatively fast moving SKUs) are most in line with our theoretical expectations and simulation findings as reported in Teunter *et al*. (2011). Considering all SKUs of that dataset, Croston has the largest positive bias, SBA a smaller negative bias, and TSB and SY have even smaller biases. The different Croston-type methods show very similar MASE, MSE and therefore also inventory performances. For SKUs from dataset 1 with a negative trend in demand, TSB does show a superior inventory performance, although the other methods do not lag far behind.

The results for dataset 2 paint a very different picture. Croston is still the most biased, but has the best inventory performance. Even for only SKUs with decreasing demand from that dataset, Croston outperforms TSB and all other methods. This could relate to a high degree of non-stationarity of that dataset, rendering theoretical results based on the assumption of stationary demand less indicative.

The relatively high optimal (across SKUs) values of the smoothing constants for dataset 2 do seem to indicate a higher degree of non-stationarity. Croston and all considered variants have an optimal value for α of 0.1 and 0.15 for datasets 1 and 2, respectively. The optimal value for β increases from at most 0.1 to at least 0.2 for Croston, SBA and SY. The TSB method does show a small decrease from 0.05 to 0.02. We remark that small values for the demand probability smoothing constant of TSB are expected for dataset 2 due the very infrequent demand for most SKUs. Another interesting finding on optimizing smoothing constants is that doing so at the individual SKU level, rather than across all SKUs, considerably deteriorates performance for both datasets.

There are many opportunities for further research. If nothing else, this research has shown that there is a need for more empirical testing of forecasting methods for intermittent demand, as the two datasets that we considered lead to different and sometimes opposite findings. In such further empirical studies, inventory performance should be considered alongside forecasting accuracy, as our results have shown that the two do not always go hand-in-hand.

One particularly interesting aspect to be studied in relation to comparative forecasting performance is that of the levels of intermittency, lumpiness and non-stationarity. This should offer insights into what method to select when, and what values to use for the smoothing constants.

Another important research avenue is to consider demand (dis)aggregation. For both datasets that we considered, optimizing smoothing constants at the individual SKU level performs much worse than across SKU. However, there may be a way in between by aggregating sets of SKUs based on e.g. their description and/ or demand process characteristics. Future research can address the amount of data (i.e. number of observations per SKU or set of SKUs) needed for optimization of smoothing constants, again in relation to the intermittency, lumpiness and non-stationarity of the data.

References

Axsäter, S., 2006. Inventory Control, 2nd edition, Springer.

Boylan, J.E. and Syntetos, A.A., 2010. Spare parts management: a review of forecasting research and extensions. *IMA Journal of Management Mathematics*, 21, 227-237.

Burgin, T.A. and Wild, A.R., 1967. Stock control experience and usable theory. *Operational Research Quarterly*, 18, 35-52.

Chatfield, D.C. and Hayya, J.C., 2007. All-zero forecasts for lumpy demand: a factorial study. *International Journal of Production Research*, 45, 935-950.

Croston, J.D., 1972. Forecasting and stock control for intermittent demand. *Operational Research Quarterly*, 23, 289-303.

Eaves, A. and Kingsman, B.G., 2004. Forecasting for ordering and stock holding of spare parts. *Journal of the Operational Research Society*, 55, 431-437.

Fildes, R., Nikolopoulos., K., Crone, S. and Syntetos, A.A., 2008. Forecasting and operational research: a review. *Journal of the Operational Research Society*, 59, 1150-1172..

Gutierrez, R.S., Solis, A.O. and Mukhopadhyay, S., 2007. Lumpy demand forecasting using neural networks. *International Journal of Production Economics*, 111, 409-420.

Hyndman, R.J. and Koehler, A.B., 2006. Another look at measures of forecast accuracy, *International Journal of Forecasting*, 22, 679-688.

Johnston, F.R. and Boylan, J.E., 1994. How far ahead can an EWMA model be extrapolated?. *Journal of the Operational Research Society*, 45, 710-713.

Johnston, F.R. and Boylan, J.E., 1996. Forecasting for items with intermittent demand. *Journal of the Operational Research Society,* 47, 113-121.

Johnston, F.R., Boylan, J.E. and Shale, E.A., 2003. An examination of the size of orders from customers, their characterization and the implications for inventory control of slow moving items. *Journal of the Operational Research Society*, 54, 833-837.

Petropoulos, F., Nikolopoulos, K. and Assimakopoulos, V., 2008. An evaluation of the accuracy of intermittent demand forecasting approaches. *Proceedings of the 28th International Symposium on Forecasting*, Nice (France) - available at: http://www.forecasters.org/isf/past-isfs.html.

Quintana, R. and Leung, M.T., 2007. Adaptive exponential smoothing versus conventional approaches for lumpy demand forecasting: case of production planning for a manufacturing line. *International Journal of Production Research*, 45, 4937-4957.

Sani, B., 1995. Periodic inventory control systems and demand forecasting methods for low demand items. Unpublished PhD thesis, Lancaster University, UK.

Sani, B. and 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, 700-713.

Schultz, C.R., 1987. Forecasting and inventory control for sporadic demand under periodic review. *Journal of the Operational Research Society*, 38, 453-458.

Shenstone, L. and Hyndman, R.J., 2005. Stochastic models underlying Croston's method for intermittent demand forecasting. *Journal of Forecasting*, 24, 389-402.

Syntetos, A.A., 2001. Forecasting of intermittent demand. Unpublished PhD thesis, Brunel University – Buckinghamshire New University, UK.

Syntetos, A.A., Babai, M.Z. and Altay, N. 2012. On the demand distributions of spare parts. *International Journal of Production Research*,.50, 2101-2117.

Syntetos, A.A. and Boylan, J.E., 2001. On the bias of intermittent demand estimates. *International Journal of Production Economics,* 71, 457-466.

Syntetos, A.A. and Boylan, J.E., 2005. The accuracy of intermittent demand estimates. *International Journal of Forecasting*, 21, 303-314.

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

Syntetos, A.A. and Boylan, J.E., 2010. On the variance of intermittent demand estimates. *International Journal of Production Economics*, 128, 546-555.

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

Teunter, R.H. and Duncan, L., 2009. Forecasting intermittent demand: a comparative study. *Journal of the Operational Research Society*, 60, 321-329.

Teunter, R. and Sani, B., 2009. On the bias of Croston's forecasting method. *European Journal of Operational Research*, 194, 177-183.

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

Willemain, T.R., Smart, C.N., Shockor, J.H. and DeSautels, P.A., 1994. Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method. *International Journal of Forecasting*, 10, 529-538.

List of Tables

Estimator	Moments	
Croston	$\mathrm{E}(Y'_t)$	$\approx p\mu + \frac{\beta}{2-\beta}\mu\, p(1-p)$
	$Var(Y'_t)$	$\approx (\frac{\alpha}{2-\alpha})(\frac{\beta}{2-\beta})\sigma^2 p^2(1-p) + \frac{\alpha}{2-\alpha}\sigma^2 p^2 + \frac{\beta}{2-\beta}\mu^2 p^2(1-p)$
SBA	$\mathrm{E}(Y'_t)$	$\approx p\mu - \frac{\beta}{2} p^2 \mu$
	$Var(Y'_t)$	$= (1-\beta/2)^2 Var(Croston)$
SY	$\mathrm{E}(Y'_t)$	$= p\mu$
	$Var(Y'_t)$	$\approx (1-\frac{\beta}{2})^2 \frac{\left[(\frac{1}{p}-\frac{\beta}{2})^2 \frac{\alpha}{2-\alpha}\sigma^2 + (\frac{\beta}{2-\beta})\frac{1}{p^2}(1-p)\mu^2 + (\frac{\alpha}{2-\alpha})(\frac{\beta}{2-\beta})\frac{1}{p^2}(1-p)\sigma^2\right]}{(\frac{1}{p}-\frac{\beta}{2})^4}$
TSB estimator	$\mathrm{E}(Y'_t)$	$= p\mu$
	$Var(Y'_t)$	$= (\frac{\alpha}{2-\alpha})(\frac{\beta}{2-\beta})\sigma^2 p(1-p) + \frac{\alpha}{2-\alpha}\sigma^2 p^2 + \frac{\beta}{2-\beta}\mu^2 p(1-p)$
SES	$\mathrm{E}(Y'_t)$	$= p\mu$
	$Var(Y'_t)$	$= \frac{\alpha}{2-\alpha}\left[p(1-p)\mu^2 + p\sigma^2\right]$
Zero forecast	$\mathrm{E}(Y'_t)$	$= 0$
	$Var(Y'_t)$	$= 0$
Naïve method	$\mathrm{E}(Y'_t)$	$= p\mu$
	$Var(Y'_t)$	$= p(1-p)\mu^2 + p\sigma^2$

Table 1. Statistical properties of intermittent demand estimators under stationary demand

3,000 SKUs	Demand Intervals		Demand Sizes		Demand per period	
	Mean	St. Deviation	Mean	St. Deviation	Mean	St. Deviation
Min	1.043	0.209	1.000	0.000	0.542	0.504
25%ile	1.095	0.301	2.050	1.137	1.458	1.319
Median	1.263	0.523	2.886	1.761	2.333	1.922
75%ile	1.412	0.733	5.000	3.357	4.167	3.502
Max	2.000	1.595	193.750	101.415	129.167	122.746

Table 2. Descriptive statistics – dataset 1

5,000 SKUs	Demand Intervals		Demand Sizes		Demand per period	
	Mean	St Deviation	Mean	St Deviation	Mean	St Deviation
Min	3.824	0.000	1	0.408	0.036	0.187
25%ile	7.273	5.431	1.556	2.964	0.155	0.538
Median	9.000	6.930	3.833	7.203	0.369	1.452
75%ile	11.571	8.630	11.333	14.328	1.155	4.434
Max	24	16.460	668.000	1138.408	65.083	275.706

Table 3. Descriptive statistics – dataset 2

Estimator	α	β	ME	MSE	MASE
SES	0.05		-0.066	74.970	0.891
SY	0.1	0.05	0.099	75.144	0.892
SBA	0.1	0.1	-0.107	74.923	0.883
Croston	0.1	0.05	0.116	75.137	0.893
TSB	0.1	0.05	-0.064	75.033	0.889
Naïve			0.054	136	1.127
Zero			-4.422	146.219	1.229

Table 4. Minimum (across SKUs) MSE results – dataset 1

Estimator	α	β	ME	MSE	MASE
SES	0.05		-0.118	233.804	1.510
SY	0.15	0.3	0.190	231.267	1.424
SBA	0.15	0.3	0.156	230.961	1.413
Croston	0.15	0.2	0.414	232.906	1.492
TSB	0.15	0.02	-0.178	232.202	1.402
Naïve			0.001	449.615	1.642
Zero			-1.344	238.172	0.872

Table 5. Minimum (across SKUs) MSE results – dataset 2

Estimator	α	β	ME	MSE	MASE
SES	0.061		0.024	265.975	2.075
SY	0.181	0.162	0.178	265.059	1.959
SBA	0.181	0.163	0.153	264.333	1.949
Croston	0.177	0.152	0.318	266.924	2.012
TSB	0.181	0.041	-0.063	262.057	1.841

Table 6. Minimum (per SKU) MSE results – dataset 2

Estimator	α	β	ME	MSE	MASE
SES	0.25		-0.903	35.151	0.669
SY	0.3	0.15	1.031	34.980	0.679
SBA	0.2	0.3	0.541	31.271	0.637
Croston	0.3	0.15	1.074	35.113	0.683
TSB	0.05	0.15	-0.951	34.825	0.902

Table 7. Minimum (across SKUs) MSE results for SKUs with decreasing demand – dataset 1

Estimator	α	β	ME	MSE	MASE
SES	0.05		-1.269	60.566	0.457
SY	0.3	0.3	1.556	62.033	0.491
SBA	0.3	0.3	1.500	61.323	0.484
Croston	0.3	0.3	1.931	67.248	0.539
TSB	0.3	0.05	-1.345	61.310	1.479

Table 8. Minimum (across SKUs) MSE results for SKUs with decreasing demand – dataset 2

List of Figures

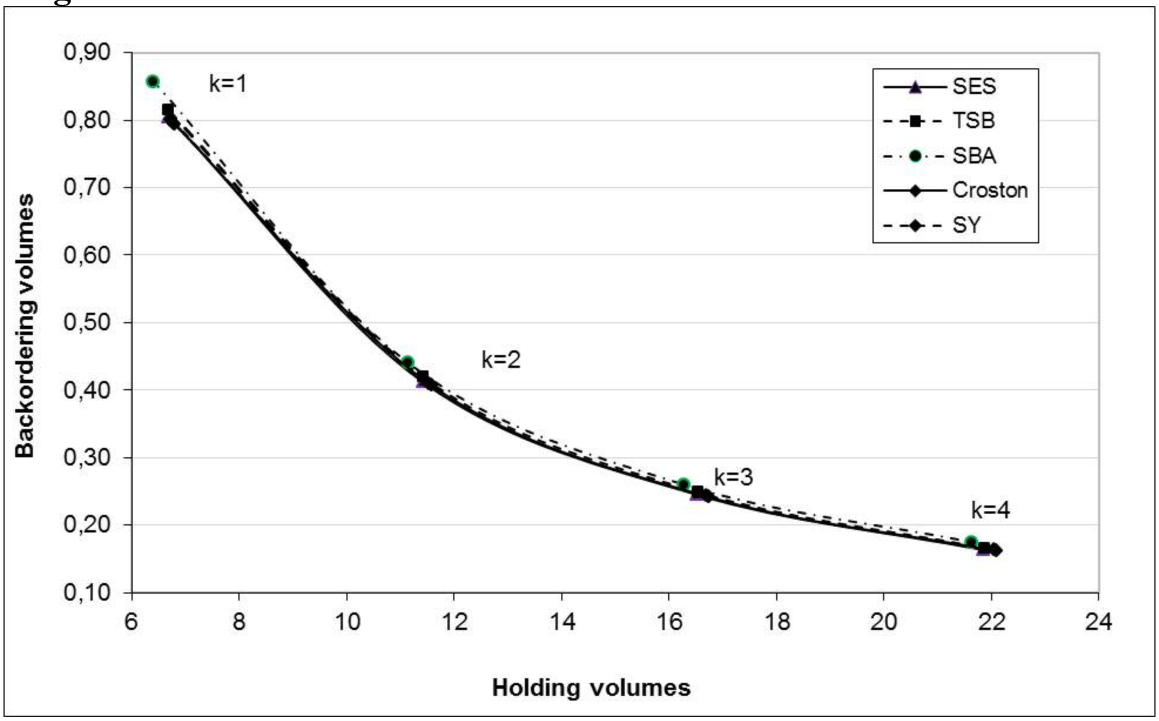

Figure 1. Stock control efficiency curves – dataset 1

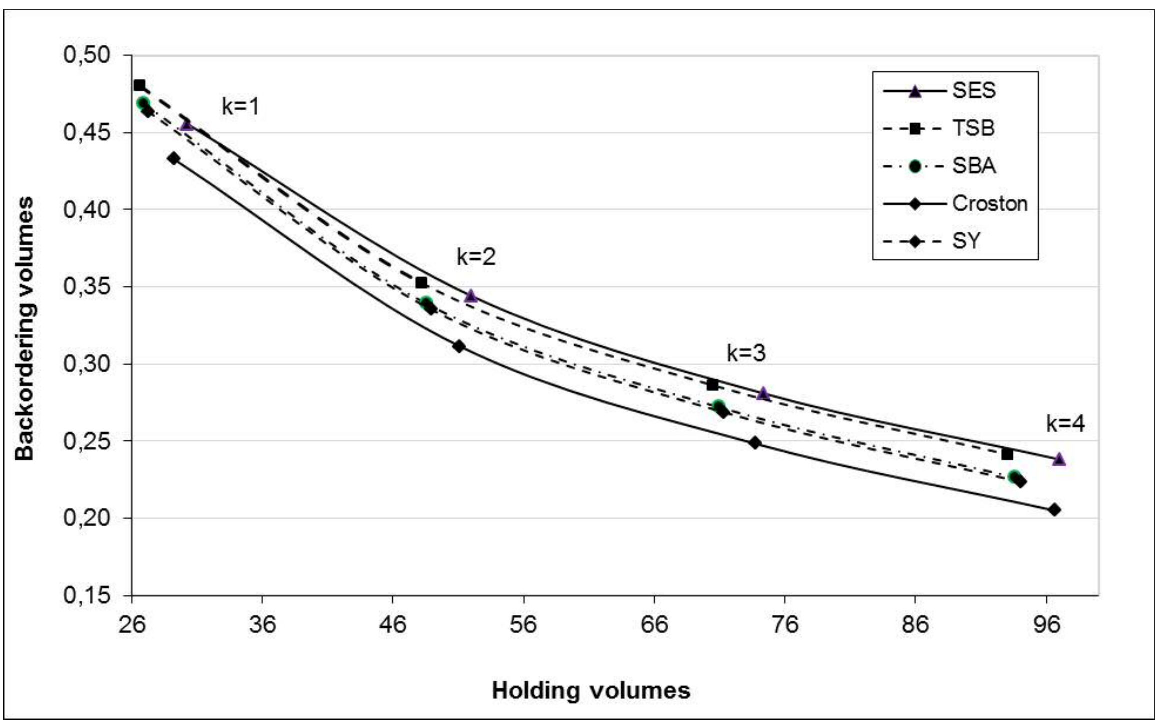

Figure 2. Stock control efficiency curves – dataset 2

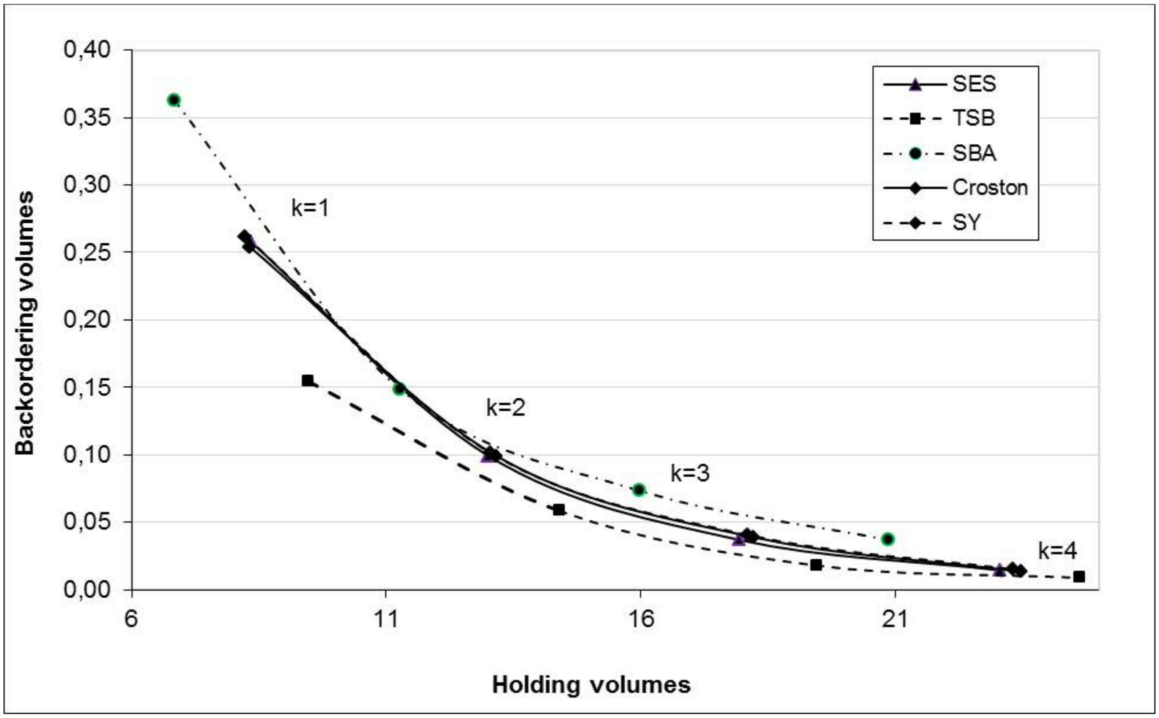

Figure 3. Stock control efficiency curves for SKUs with decreasing demand – dataset 1

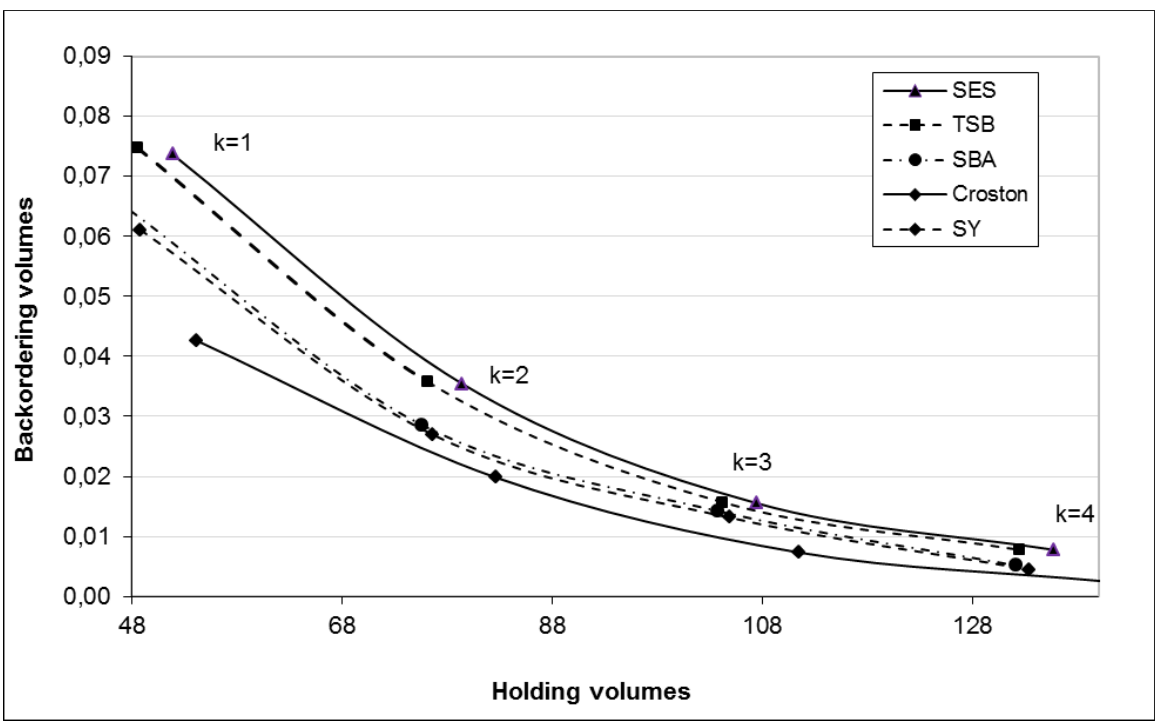

Figure 4. Stock control efficiency curves for SKUs with decreasing demand – dataset 2