2008

Adapting Wright's modification of Holt's method to forecasting intermittent demand

Nezh Altay, *University of Richmond*

Frank Rudisill, *University of South Carolina - Upstate*

Lewis A Litteral, *University of Richmond*

Adapting Wright's modification of Holt's method to forecasting intermittent demand

Nezih Altay^{a,*}, Frank Rudisill^b, Lewis A. Litteral^a

^a*Robins School of Business, University of Richmond, Richmond, VA 23173, USA*

^b*School of Business Administration and Economics, University of South Carolina Upstate, USA*

Received 1 January 2006; accepted 1 January 2007

Available online 1 March 2007

Abstract

The challenge of accurately forecasting demand for spare parts is due to the intermittent nature of their demand. This study compares a recent modification of Croston's method and a method based on Holt's double exponential smoothing taking the firm's competitive priorities into consideration. Wright's modification of Holt's method is presented as a viable alternative when forecasting demand for spares especially when trend is present. Results indicate that firms focusing on minimizing inventory levels as a priority should consider forecasting using the Modified Croston's method. If their priority is high customer service, then the modified Holt's method is superior.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Spare parts; Intermittent demand; Forecasting; Trend

1. Introduction

Intermittent demand is common in the service parts businesses of the aerospace, automotive, high-tech/electronics, utilities, and industrial machinery industries. The airline industry alone spends in excess of \$10 billion annually on spare parts, with about 10% originating from pre-owned stock units of airlines and overhaul agencies. AMR Research estimates that \$700 billion was spent on service parts in USA in 2001 (Smart, 2002). It is not only the variability of the demand size, but also the variability of the demand timing that makes intermittent demand so difficult to forecast. The

fact that equipment does not fail every demand period creates a lumpy demand structure. Hence such demand cannot be accurately predicted by standard forecasting techniques such as exponentially weighted moving averages (EWMA). EWMA places more weight on the most recent data resulting—due to many zero demand periods—in a series of estimates that are highest just after a demand occurrence and lowest just before demand occurs again.

Although a significant amount of research has been done on the general problem, very little of it has considered the effect of the presence of trend in the data. The studies that have recognized this aspect of the problem to be important include Lindsey and Pavur (2005), Ghobbar and Friend (2003), and Snyder (2002). In the first and last of these cases the presence of trend is explained by the

*Corresponding author. Tel.: +1 804 289 8259;
fax: +1 804 289 8878.

E-mail address: naltay@richmond.edu (N. Altay).

life cycle stage of the item. A significant percentage of the real data series used in this study also exhibits trend.

An approach developed by Croston (1972) has been considered the benchmark in intermittent demand forecasting. Later, Wright (1986) developed a method for missing or irregularly reported data that could potentially be used to forecast intermittent demand. Wright's method is a modification of Holt's method in which a trend line is calculated and updated using exponential smoothing. This characteristic of the method makes it a good candidate for forecasting intermittent demand with trend.

The purpose of this study is to adapt Wright's approach to forecast intermittent demand, compare its performance to Croston's method, and investigate the environmental conditions favorable to each of these approaches. The methodologies are applied to simulated scenarios as well as a large set of demand data obtained from an aircraft parts supply center. Methods are evaluated using statistical measures of forecasting accuracy as well as accuracy implication metrics based on a common stock replenishment rule. The results indicate that, when service levels are a top priority, i.e. stock-out costs are relatively high relative to holding costs, Wright's method is preferred.

The rest of the paper is organized into four sections. Section 2 presents the two methods compared here, and discussion of the bootstrapping approach, and other related literature. Section 3 describes the methodology of this study including the experimental design and the assumptions. Subsequently, Section 4 presents the results of the computational tests and their statistical analysis. The conclusion is found in Section 5 where the managerial implications are summarized and future research directions are suggested.

2. Methods and literature review

To design and/or compare techniques to forecast demand where data contains sporadic zeros it is important to understand the nature of this sporadicity. Williams (1984) dissected demand variance into variance of order sizes, transaction variability and variance of the lead times. Expanding on Williams' idea Syntetos developed a categorization scheme based on the average demand interval (ADI) and the square coefficient of variation (CV^2) calculated using non-zero demand values only

(Syntetos, 2001; Syntetos et al., 2005). Fig. 1 displays this framework. In this paper we focus our attention on the lumpy and intermittent demand categories.

2.1. Croston's method

For demand streams with high CV^2 and high ADI, EWMA fails to generate forecasts with acceptable accuracy. The seminal work to address such situations was done by Croston (1972) who developed a new method based on exponential smoothing to forecast demand size and demand interval separately. Croston used a common smoothing constant to generate both forecasts and recommended using small values, between 0.1 and 0.2 for example. Consider the following notation to explain Croston's procedure:

z_t	demand at period t
$\hat{z}_t$	forecast of size of next demand beyond period t
z^*	average demand per period
p_t	time between two positive demands
$\hat{p}_t$	forecast of demand interval
q	time interval since the last positive demand
α	smoothing constant

The method updates the forecasts for demand size and interval only after positive demand occurs. That means, if in a period t demand is zero the method only increments the count of time periods since last positive demand. The procedure for Croston's

Fig. 1. Categories of demand patterns (adopted from Syntetos et al., 2005).

method is described below:

$$\text{IF } z_t = 0 \text{ THEN } \hat{z}_t = \hat{z}_{t-1}; \hat{p}_t = \hat{p}_{t-1}; \\ \text{and } q = q + 1 \quad (1)$$

$$\text{ELSE } \hat{z}_t = \hat{z}_{t-1} + \alpha(z_t - \hat{z}_{t-1}); \\ \hat{p}_t = \hat{p}_{t-1} + \alpha(q - \hat{p}_{t-1}); \text{ and } q = 1 \quad (2)$$

$$\text{Combining these forecasts provides } z^* = \frac{\hat{z}_t}{\hat{p}_t}. \quad (3)$$

Croston's method, though widely accepted in literature has frequently been challenged. Rao (1973) corrected algebraic errors in Croston's error estimators. Johnston and Boylan (1996a) criticized not incorporating the effect of uncertainty in the time intervals between positive demand periods. Snyder (2002) suggested using different smoothing constants for demand size and demand interval along with smoothing in log-space to avoid negative demands.

Syntetos and Boylan showed that Croston's method is positively biased due to a mistake in the mathematical derivation of the estimate of average demand per period (Syntetos and Boylan, 2001). Referring to the results of Sani and Kingsman (1997) they stated that this mistake contributes towards the modest performance of the method on practical problems. Their modification corrects the expected mean demand per interval by replacing $z^* = \hat{z}_t/\hat{p}_t$ with $z^* = \hat{z}_t/(\hat{p}_t c^{\hat{p}_t-1})$, where $c > 100$ and eliminates the bias associated with Croston's method. The modified algorithm was then tested using a simulated experiment with 200 factor combinations: 10 different smoothing constants, five different demand intervals, and four different combinations of means and standard deviations of demand size were generated using only the normal distribution. Each simulation run consisted of 20,000 demand periods of which the first 100 were considered the warm-up period. They concluded that Croston's method is superior only for low values of the smoothing constant and the bias becomes pronounced for α values above 0.15.

Another simulation experiment intentionally created scenarios violating Croston's assumptions and compared Croston's method to EWMA (Willemain et al., 1994). These experiments used the lognormal distribution for demand size, and assumed cross-correlation between demand size and intervals, and autocorrelated demand sizes and intervals. The two

methods were also compared using real industrial data and their performances were reported using mean absolute percent error (MAPE) as the measure of accuracy. In all scenarios Croston's method performed better than EWMA. Willemain et al. also suggested that there may be degrees of intermittency where too many zero demand periods make it essentially impossible to forecast well using any statistical method, while too few zero demand periods make it unnecessary to switch from EWMA to another method.

Inspired by this suggestion, Johnston and Boylan investigated the level of ADI when Croston's method becomes favorable to EWMA (Johnston and Boylan, 1996b). They reported that Croston's method is superior provided that the ADI is greater than 1.25. This essentially translates to a data stream with 20% zeros. Syntetos (2001) moves this threshold to ADI = 1.32 in his framework, which corresponds to 24.24% zeros.

Syntetos and Boylan (2005) compare a newly developed forecasting method based on Croston's assumptions with three other forecasting methods, simple moving average with 13 periods, single exponential smoothing, and Croston's method. The methods are compared on 3000 real intermittent demand data series from the automotive industry and a useful discussion of error measures is presented. The new method presented performs better than the others compared in the study on most of the error measures considered. In particular, superior performance of their method was observed on scaled mean error and relative geometric root mean square error. Syntetos and Boylan (2006) also used the same data series to simulate and compare the stock control performance of these methods. Their results indicate overall that the estimator developed in their 2005 paper demonstrates superior stock control performance.

Levén and Segerstedt (2004) suggest a procedure for a modern ERP system that is a periodic review system built around a modified Croston's method while fitting the Erlang distribution to the data. Their simulation studies show that the Croston based procedure presents fewer shortages than a system based on exponential smoothing and the normal distribution. Of particular interest is their conclusion that the reduced shortages are not due to increased inventory levels and that improved performance is a result of the use of the modified Croston's method and not particularly dependent

on their choice of the Erlang distribution. Boylan and Syntetos (2006a), however, showed that Levén and Segerstedt's modified version is based on invalid measure of forecast accuracy and produced biased forecasts. They found Croston's method to be generally more accurate than its modification.

While attempting to identify stochastic models that underly Croston's method, Shenstone and Hyndman (2005) show that such models are inconsistent with the properties of intermittent demand data. They are quick to point out, however, that this result does not preclude the usefulness of Croston's method in real world applications.

To summarize the literature to date, Croston's method is well accepted as a benchmark. Its assumptions and formulation have been well scrutinized. It has been shown to be superior to EWMA when demand is intermittent (i.e. at least 20% of data contains zeros).

2.2. Wright's modified Holt's method

Wright (1986) presented a modification for Holt's double exponential smoothing for data with changing reporting frequency and data with irregularity in time spacing. Holt's method consists of fitting a trend line to the data where estimates of intercept and slope are obtained from single exponential smoothing equations. Since Croston's method is based on the assumption of stationary mean model, this modification of Holt's method may provide some flexibility and improvement in performance. Wright listed his reasons for choosing Holt's method rather than Brown's double exponential smoothing as its greater computational efficiency, flexibility in terms of having more smoothing constants, and better performance with empirical data as shown in the literature.

Wright applied his modified algorithm to six series of published data in Canada ranging from electric power availability to calcium concentration measurements in rivers. Although these data streams contain no zero input in any given period, they contain periods with no input at all, creating irregular and erratic data. Therefore, our research takes a heuristic approach in applying this method to intermittent demand series. The forecast obtained F_{n+1} is the one step ahead forecast of demand size. The average demand per period is then calculated by dividing the forecasted demand size by the ADI up to that period. The notation and derivation of

Wright's extension to Holt's method is given below:

L_n	intercept of the trend line at demand period n
M_n	slope of the trend line at demand period n
X_n	demand size in demand period n
t_n	demand period n (ignoring zero demand periods)
q	average time spacing of data (calculated from historical figures and used to initialize)
α	smoothing constant for the intercept
β	smoothing constant for the slope
A	intercept of trend line used to initialize
B	slope of trend line used to initialize
F_{n+1}	one period ahead forecast of the next demand value

$$F_{n+1} = L_n + M_n(t_n - t_{n-1}), \quad (4)$$

$$L_n = (1 - V_n)[L_{n-1} + M_{n-1}(t_n - t_{n-1})] + V_n X_n, \quad (5)$$

$$M_n = (1 - U_n)M_{n-1} + \frac{U_n(L_n - L_{n-1})}{(t_n - t_{n-1})}, \quad (6)$$

$$U_n = \frac{U_{n-1}}{(d_n + U_{n-1})}, \quad (7)$$

$$V_n = \frac{V_{n-1}}{(b_n + V_{n-1})}, \quad (8)$$

$$d_n = (1 - \beta)^{(t_n - t_{n-1})}, \quad (9)$$

$$b_n = (1 - \alpha)^{(t_n - t_{n-1})}, \quad (10)$$

$$U_0 = 1 - (1 - \beta)^q, \quad (11)$$

$$V_0 = 1 - (1 - \alpha)^q, \quad (12)$$

$$L_0 = A, \quad (13)$$

$$M_0 = B. \quad (14)$$

2.3. Bootstrapping methods

Although beyond the scope of the comparisons made in this paper, there is another promising class of methods known as bootstrapping that is being applied to forecast intermittent demand. The definitive work in this area appears in Willemain et al. (2004). In that paper the authors develop a patented algorithm to address the problem of

forecasting the cumulative distribution of intermittent demand over a fixed lead time as well as a new method of assessing the accuracy of those forecasts. They used demand data from nine industrial companies and showed that the bootstrapping method estimated the cumulative distribution of lead time demand over a fixed lead time more accurately than exponential smoothing and Croston's method. Earlier work on bootstrapping was reported in Smart and Willemain (2000) and Smart (2002). The fact that the algorithm is patented and the patent's potential effects on the conduct of academic research, produced a lively discussion in Lawrence (2004), Gardner and Koehler (2005) and Willemain et al. (2005).

In addition to the stream of bootstrapping research described above, Snyder (2002) compared adaptations of Croston's method to a parametric bootstrap approach that integrates demand forecasting with inventory control. Illustrations are made on real demand data for car parts. In a very recent study, Hua et al. (2007) introduce a method to estimate the cumulative demand of spare parts in process industries over a fixed lead time while forecasting the occurrence of non-zero demand. They used data sets from a petrochemical enterprise to compare this method with exponential smoothing, Croston's method and a slightly modified version of the bootstrapping method of Willemain et al. (2004). Their statistical results show that the proposed method provides superior performance with measures of average error ratio of non-zero demand judgments over lead time and absolute percentage error of lead time demand.

3. Experimental design

We have originally compared three intermittent demand forecasting methods using simulated scenarios in Excel: Croston's method (CM), Syntetos and Boylan's modification to Croston's method (MCM), and Wright's modified Holt's algorithm (WMH). However, since the results of CM and MCM were virtually the same under the experimental conditions created, we only report the results of MCM in the remainder of the paper. In order to evaluate the two methods we generated, via simulation, intermittent demand data with varying underlying distributions and characteristics. The data were generated based on a $4 \times 2 \times 2$ design, with four demand distributions (Erlang, exponential, uniform and normal), two mean demand levels (10 and 500) and two levels of

intermittency (30% and 70% zeros). The two forecast methods (MCM and WMH) were applied to the same data set of the 16 generated conditions.

Each of the simulated demand data sets is 1000 periods in length. In each case, the first 200 observations are held out and used to initialize the forecasts. The remaining 800 observations are used to collect data for the experiment. This long warm-up period is essentially unrealistic since most real service parts series we have encountered do not contain more than 72 observations. However, our intention with producing long strings of demand data is to strip the output of the study from short term effects such as the choice of initial values. In order to make meaningful statistical comparisons, 30 data sets are generated for each demand distribution, mean level and intermittency level combination. The same sets of demand data are used with both forecasting approaches. Consequently, there is a total of 480 data streams (four distributions, two mean levels, two intermittency levels, and 30 separate runs) tested for each criterion examined. The study is then repeated on a real data set, the characteristics of which are described in Section 3.3.

The objective of the simulation study is to investigate the forecast accuracy and service performance of the two forecasting methods across the different demand behaviors generated. Analysis of variance (ANOVA) is performed to evaluate the statistical significance of the experimental factors in the simulation. Although all the experimental factors (distribution, demand mean, intermittency and forecast method) and their interactions are included in the ANOVA models, we are specifically interested in the statistical significance of the forecast methods and the interactions in which they appear. The response variables for forecast accuracy were MASE, GMAE, MAPE and the response variables for service were AVGINV, SERV0/1, SERV% and TC which are explained in Sections 3.4.1 and 3.4.2. All effects in the model are considered to be fixed except for simulation run (1 through 30 for each combination of distribution, mean demand level, and level of intermittency) which is a random effect. The model residuals are analyzed to test compliance with the ANOVA assumptions. The residuals should be normal and display a constant variance across the range of the experimental factors. Some of the response variables do not satisfy the ANOVA assumptions so transformations are made and statistical significance is determined using the ANOVA results for

the transformed data according to Hicks (1982). Residual analyses are performed and evaluated using the transformed responses. In most cases, the transformations satisfy the ANOVA requirements.

3.1. Demand distributions and intermittency

Johnston and Boylan (1996a) simulated the demand sizes using the exponential, Erlang, and uniform distributions. Croston (1972) assumed demand size is normally distributed. In this study, we also use these four distributions to generate demand sizes. These distributions are selected because they are utilized in previous studies and more importantly cover a range of CV^2 values. Exponential distribution falls undoubtedly in the upper right quadrant in Fig. 1. The Erlang distribution we generated falls on the CV^2 threshold. Uniform distribution comes close to the threshold while the normal distribution is clearly in the lower right quadrant of Fig. 1. Data generated were rounded to the nearest integer to create discrete data. We recognize that rounding provides an approximation to the actual distribution of demand. Parameters used to generate the two levels of mean demand with these four probability distributions are listed in Table 1. For details about the probability distributions and their parameters, the reader is advised to consult Law and Kelton (2000).

We considered two levels of intermittency: 30% and 70% zero demand periods. These two intermittency levels translate to ADIs of 1.43 and 3.33, respectively. This is important for two reasons. First, Johnston and Boylan (1996b) suggested that CM is superior to simple exponential smoothing when 20% of data contains zero values. And

secondly, the ADI threshold between intermittent demand and smooth demand in Fig. 1 is 1.32, which translates to 24.24% of zeros in the data (Syntetos et al., 2005). Therefore at 30% and 70% the data are considered intermittent by any measure in the literature, and we are able to observe forecasting method behavior even in extremely intermittent scenarios.

In this study, intermittency was introduced by modeling the number of zero demand periods using the Bernoulli distribution as suggested by Croston (1972). Janssen (1998) tested the Bernoulli demand generation process on real data obtained from a Dutch wholesaler of fasteners and concluded that the Bernoulli process is a reasonable approximation for intermittent demand processes.

3.2. Selection of smoothing constants

The choice of the smoothing constants will surely affect the accuracy of forecasts. However, it is impractical to go over every single data set in order to select the best constant especially when a firm is dealing with thousands of stock keeping units in their inventory. Johnston and Boylan (1996a) suggested choosing a low value of the smoothing constant for low average demand and designed their simulations based on a smoothing constant $\alpha = 0.15$. Croston (1972) advised that α values between 0.1 and 0.2 worked well in practice. Willemain et al. (1994) reported test results for $\alpha = 0.1$ even though they ran experiments using 0.01 and 0.5 as alternatives for the smoothing constant. For the modified Holt's method Wright (1986) reported that the smoothing constant for the intercept, α , had to be more precisely chosen than the slope parameter β . He further added that for erratic demand, the value of the intercept

Table 1
Demand distributions and mean levels of demand used in the study

	Parameters		Means		Standard deviations		Theoretical CV^2 s	
Exponential	$\lambda = 0.1$	$\lambda = 0.002$	10	500	10	500	1	1
Erlang	$r = 2$ $\lambda = 0.2$	$r = 2$ $\lambda = 0.004$	10	500	7	353.5	0.49	0.499
Uniform	$a = 1$ $b = 19$	$a = 1$ $b = 999$	10	500	5.2	288.1	0.27	0.33
Normal	$\mu = 10$ $\sigma = 3$	$\mu = 500$ $\sigma = 166$	10	500	3	166	0.09	0.11

parameter was critical but the slope parameter could vary over a wide range with little effect on forecast accuracy.

This study utilizes a single level of smoothing constants to compare forecasting performance. To find the optimal smoothing constant for each methodology the full experiment was conducted for a range of smoothing constants (480 runs per constant). MCM was tested using α values starting from 0.05 and going to 0.45 with increments of 0.05 and two levels of parameter c in the modification (100 and 1000). WMH method was tested using $0.1 \leq \alpha \leq 0.9$ and $0.1 \leq \beta \leq 0.9$ with increments of 0.1. Only statistical error measures MAD and MAPE (calculated at issue points only) were used as evaluation criteria.

Parallel to the conclusions in literature, for MCM the smoothing constant $\alpha = 0.1$ was found to provide the best results for both, MAD and MAPE. The modification constant c was taken as 1000, although the values of MAD and MAPE did not change significantly between $c = 100$ and 1000. The preliminary study showed that $\alpha = 0.3$ and $\beta = 0.1$ provide the best combination for WMH.

3.3. Description of the real data set

In addition to the simulated environment, a large data set of demand for aircraft service parts was utilized. Items included in the data set range from nuts and bolts to circuits and engine parts. Sixty periods of demand history are available for each item. Average positive demand ranges from one unit to 1326.8 units, with the maximum demand in the whole data set being 4568 units. Items in this data set cover a wide range of ADI (from 1.33 to 12.0) and CV²s (from zero to 12.25). A total of 4225 items are included in the data set where 3126 items have a CV² > 0.49, while the CV² of the remaining 1099 is less than or equal to the threshold value. Furthermore, within this data set, 594 parts indicate a significant trend ($p < 0.05$) for non-zero demand.

The small number of periods in this data set makes the choice of initialization critical. WMH requires good initial intercept and slope estimates. Wright used the first six data points to find the initial estimates. In the case of highly intermittent demand, however (e.g. 92% zeros), the series might not have any positive input for more than six periods at the beginning of the data stream, or even have less than six positive data points throughout

the whole data stream. In this study, methods were initialized using the first 20 periods of data and performance measures were calculated using the last 40 periods. In cases where no non-zero demand was observed within the first 20 periods, the slope was taken to be zero and the first non-zero demand in the series was used to initialize the intercept. Initial estimates for demand size for the MCM were found assuming a perfect forecast (equal to the first period's demand). Initial demand interval was always equal to one.

3.4. Performance evaluation

3.4.1. Statistical measures of forecast accuracy

Standard forecast error measures frequently appearing in the literature do not provide fair evaluations for intermittent demand cases due to the large number of zero demand periods. There is an ongoing discussion about which error measures are appropriate for this type of series. The June 2006 issue of *Foresight—the International Journal of Applied Forecasting*—has a special feature on this topic in which Hoover (2006), Boylan and Syntetos (2006b), Willemain (2006), and Hyndman (2006) suggest various metrics and discuss their strengths and weaknesses. Although it seems that no single metric comes out clean, in our opinion the mean absolute scaled error (MASE) developed by Hyndman and Koehler (2006) is the least problematic one.

MASE is a scale-free error measure which uses naïve forecasts as a benchmark. Let e_t indicate forecast error, $e_t = X_t - F_t$. The scaled error, q at time t is then calculated using (15) and MASE is the average of absolute values of q_t

$$q_t = \frac{e_t}{(1/(n-1)) \sum_{i=2}^n |X_i - X_{i-1}|}. \quad (15)$$

The denominator of Eq. (15) is simply the MAD for naïve forecasts. A scaled error less than one suggests better forecasts than the benchmark method.

Syntetos and Boylan (2005) proposed using the geometric root mean squared error (GRMSE) for measuring forecast accuracy when demand is intermittent. Later, Hyndman (2006) proved that the geometric mean absolute error (GMAE) and GRMSE are essentially the same for intermittent demand, but GMAE is relatively simpler to calculate. The GMAE, calculated using (16), however, is problematic if for any observation error

equals to zero

$$\text{GMAE} = \left(\prod_{t=1}^n |X_t - F_t| \right)^{1/n}. \quad (16)$$

This may happen in two ways: (1) non-zero demand and identical non-zero forecast, and (2) zero demand and zero forecast. The first case is a rare occasion if demand forecasts are not rounded to the nearest integers. Since Croston's method updates demand size forecasts only at non-zero demand period the second case is also unlikely to happen. However, WMH uses a trend line to forecast demand and a small quantity demanded following a large quantity will result in a steep negative slope and consequently in a zero forecast. Hence, the second case is very likely to happen with items that demonstrate a negative slope and/or high intermittency.

One rough approach has been to calculate errors only at demand points (Eaves and Kingsman, 2004). Hoover (2006) mentions that he found some software packages reporting MAPE for intermittent series in this fashion. Calculating errors only on non-zero demand periods, however, does not give a complete picture since this approach is essentially treating zero demand as missing data. Therefore, we used MAPE cautiously in this paper. On the other hand, MAD is not a good choice for our study due to the wide variability of demand sizes in the aircraft parts data set.

Let J be the set of demand periods with non-zero demand and $j \in J$. If X_j denotes actual demand in demand period j , F_j the forecast accumulated until period j where a non-zero demand occurs and m the total number of non-zero demand periods, then

$$\text{MAPE} = \frac{\sum_{j=1}^m (|X_j - F_j|/X_j) * 100}{m}. \quad (17)$$

Since MCM forecasts average demand per period we used this average to find a cumulative forecast for demand and calculate error. We did, however, use the demand size forecast as an upper bound for the cumulative forecast. For instance, if MCM forecasts in period i that 10 units of demand is expected to occur in five periods, the cumulative forecast F_{i+1} , F_{i+2} , and F_{i+3} would be 2, 4, and 6, respectively. If demand does not occur by period $i+5$, the cumulative forecast would remain as 10 units until demand occurs. The same approach was also applied to WMH.

3.4.2. Accuracy implication metrics

Flores et al. (1993) introduced economic measures of forecast accuracy based on the cost burden of inventories. Boylan and Syntetos (2006b) suggest that if we use the same inventory replenishment rule on different forecasting methods the outcome implies the accuracy of the forecasts. In our study, we used an (s, S) policy where an order was placed when on-hand inventory plus in-transit shipments drops to or below reorder point s . The quantity ordered is then the difference between on-hand (including in-transit shipments) and S . The calculations necessary to find the optimal values of s and S can be very demanding. In an extensive study, searching for the best periodic inventory control and forecasting method for lumpy demand items, Sani and Kingsman (1997) concluded that Naddor's heuristic (Naddor, 1975) along with the Power Approximation is very effective for finding s and S . Naddor's heuristic only requires estimates of the mean and variance of demand.

To operationalize this replenishment policy we assumed that lead time is one period and holding cost per unit per period is \$1. The choice of the shortage cost leads to the desired cycle service level in Naddor's heuristic. We have run our simulation for four different shortage cost values (4, 9, 24, and 99 dollars per unit short) which result in 80%, 90%, 96%, and 99% desired cycle service levels. The cost of ordering was assumed to be \$32 and \$64 per order. Each simulation run started with an initial on-hand inventory of S units calculated from the values in the warm-up sample.

To have a sense regarding the forecast performance we monitored average inventory (AVGINV), proportion of orders immediately fulfilled from stock (SERV0/1), average percentage of orders satisfied (SERV%), and total cost (TC). We measured service in two different ways because as also pointed by Sani and Kingsman (1997) orders that cannot be completely satisfied from stock may lead to loss of future sales from that customer or even to the loss of that particular order.

4. Results and discussion

The two demand forecast methods (MCM and WMH) were separately applied to the same 30 sets of simulated demand from each of the 16 combinations of distribution, mean and intermittency. For each set of the 480 sets of data, forecast accuracy measures (GMAE, MASE and MAPE)

and accuracy implication measures (AVGINV, SERV0/1, SERV%, and TC) were calculated for the two forecast methods. Ninety-five percent confidence intervals were generated for each combination of forecast method and demand characteristic. An ANOVA was also conducted to evaluate the statistical significance of the forecast methods across the 16 demand conditions. The p -values from these ANOVAs for each measure appear in Table 2. Since, the factors of primary interest are those involving forecast methods, Table 2 only reports results related to Method and its interactions. This technique permits the evaluation of the main effect of Method and each interaction involving Method. Factors and interactions with p -values less than 0.05 are considered to be statistically significant.

Table 2 also displays the p -values for ANOVAs performed on transformations of some of the response measures. The transformations were applied to the data in some cases to ensure that the statistical assumptions associated with the ANOVA model were satisfied. There is more discussion of the particular transformations and the associated residual analysis in the sections that follow.

The aircraft data were analyzed in three ways; as a complete set, by partitioning the data set into

those data streams that exhibited a statistically significant trend and those that did not, and by partitioning the data set into data streams that were characterized by low CV² and those with high CV².

4.1. Results of the analysis of the simulated data

4.1.1. Forecast accuracy metrics

In the case of GMAE, WMH performs better or there is no statistically significant difference in every case. This conclusion is easily verified by Fig. 2 which presents 95% confidence intervals for the difference in GMAE when WMH is subtracted from MCM for each combination of experimental conditions.

The authors note that the ANOVA requirements of normality of residuals and homogeneity of variances are not met for this error measure but transformations that were applied to some of the other measures (log 10) were not possible in this case because of the occurrence of zero values for this measure. As noted in Section 3.4.1 the GMAE measure for a series turns out to be zero anytime any one of the forecast errors in the series is zero. When demand is very intermittent and WMH finds a negative trend, the possibility of a zero forecast

Table 2
 p -Values from ANOVAs for simulated data (Method and its interactions only)

Source	DF	GMAE	MASE	MAPE	AVGINV	SERV0/1	SERV%	TC
Method	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Dist*Method	3	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Mean*Method	1	0.000	0.024	0.251	0.000	0.000	0.000	0.000
Zeros*Method	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Dist*Mean*Method	3	0.000	0.632	0.664	0.000	0.000	0.000	0.000
Dist*Zeros*Method	3	0.000	0.026	0.000	0.000	0.000	0.000	0.000
Mean*Zeros*Method	1	0.000	0.286	0.092	0.000	0.000	0.000	0.000
Dist*Mean*Zeros*Method	3	0.000	0.441	0.057	0.000	0.004	0.000	0.000
Error	464							
Total	959							
Source	DF		Log MASE	Log MAPE	Log AVGINV	Arcsin SERV0/1	Arcsin SERV%	Log TC
Method	1		0.000	0.000	0.000	0.000	0.000	0.000
Dist*Method	3		0.000	0.000	0.000	0.000	0.000	0.000
Mean*Method	1		0.023	0.000	0.000	0.000	0.000	0.000
Zeros*Method	1		0.000	0.000	0.000	0.000	0.000	0.000
Dist*Mean*Method	3		0.630	0.023	0.000	0.000	0.000	0.000
Dist*Zeros*Method	3		0.003	0.000	0.000	0.000	0.000	0.000
Mean*Zeros*Method	1		0.342	0.558	0.000	0.000	0.000	0.000
Dist*Mean*Zeros*Method	3		0.425	0.662	0.965	0.000	0.761	0.000
Error	464							
Total	959							

Note: The degrees of freedom (DF) will not sum to 959 since some factors are intentionally omitted from this table.

Fig. 2. Confidence limits for the difference of $GMAE_{MCM} - GMAE_{WMH}$.

Fig. 3. Confidence limits for MASE in the case of simulated data.

coinciding with a zero demand in a given period is highly probable. Therefore WMH results in more zero GMAE values than MCM which rarely does so. This makes a fair comparison also very difficult. However, Fig. 2 clearly shows that WMH is less than or equal to MCM in each case.

With regard to the MASE and MAPE error measures, MCM performs better or ties in every case. It can be seen from Fig. 3 that the case of 70% zeros is harder for both methods.

The log transformation was applied to the MASE data to satisfy the statistical assumptions underlying

the ANOVA model. Fig. 4 gives the residual plots associated with the log MASE data and is an example of the type of analysis conducted regarding the ANOVA models for each of the error measures. Thus the p -values in Table 2 accurately reflect the statistical significance. Note that Method and several interactions involving Method are statistically significant.

Fig. 5 shows the results for MAPE where MCM performs better or ties in every case. It can be seen that there is little difference within distributions especially when the mean is 500 but that the

Fig. 4. Residual plots associated with the log transformation of MASE.

Fig. 5. Confidence limits for MAPE in the case of simulated data.

exponential case is much more difficult. Such a result is expected since the data variability is much greater in the exponential distribution as measured by its CV^2 value of 1. The ANOVA results for the log MAPE show that Method and interactions

involving Method are statistically significant for this error measure. The ANOVA assumptions were satisfied using the log transformation.

The mean error was calculated for each of the 32 cases and it was not significantly different from zero

in any case except exponential with mean equal to 500 and 70% zeros. This indicates that neither method exhibited any bias in its forecasts. Since 32 95% confidence intervals were constructed, it would be expected that, considering the type I error value of .05, at least one of the cases would result in an indicated difference from zero even if such a difference did not exist. However, indications are that the exponential case of mean 500 and 70% zeros is the most difficult which might also explain this result.

4.1.2. Accuracy implication metrics

The simulation results indicate mixed results for the two methods regarding the accuracy implication metrics. The general conclusion is that MCM tends to have lower average inventory levels than WMH at the expense of service level. This tradeoff is consistently observed and intuitively appealing. The TC results are mixed. With regard to AVGINV, MCM performs better (carries a lower average value) or performs the same as WMH in all cases where the mean demand is 500. Fig. 6 shows no significant difference in performance when the mean demand is 10. Table 2 also shows that, for log AVGINV, there are statistically significant two-way and three-way interactions involving Method. Again, the ANOVA assumptions were satisfied by the log transformation. Table 3 summarizes the results of

the accuracy implication metrics for the simulated data.

With regard to both measures of service level WMH performs better or ties MCM in every case. As can be seen in Fig. 7, the difference in performance between the two methods on the two service level measures is greater when the mean demand is 500.

The differences range from about 0% to about 23% in the case of SERV0/1 and from about 0% to 5.5% in the case for SERV%. In each of these cases, statistical significance is determined by performing an ANOVA after applying the arcsine transformation (which satisfies the ANOVA requirements). The TC results are mixed with no difference at mean demand equal to 10 as can be seen in Fig. 8.

WMH performs better on TC when there are 30% zeros and mean 500 while MCM performs better with 70% zeros and mean 500. The ANOVA based on the logarithm of TC (log TC) also confirms statistically significant interactions involving Method. Again the ANOVA assumptions were satisfied using the log transformation.

4.2. Results of the analysis of the aircraft data

4.2.1. Forecast accuracy metrics

A graph of 95% confidence intervals for the differences in means between MCM and WMH for each accuracy metric (MAPE, GMAE, and MASE)

Fig. 6. Confidence limits for the difference of $AVGINV_{MCM} - AVGINV_{WMH}$.

Table 3
Means of accuracy implication metrics for simulated data

Distribution	Means	Zeros (%)	Method	Means			
				AVGINV	SERV0/1	SERV%	TC
Normal	10	30	MCM	22.3	99.88	99.99	25,049
			WMH	22.8	99.92	100.00	25,409
		70	MCM	25.8	99.78	99.99	23,762
			WMH	27.6	99.95	100.00	25,087
	500	30	MCM	375.9*	90.91*	99.17*	416,281
			WMH	454.0*	96.60*	99.76*	413,315
		70	MCM	593.6*	92.65*	99.36*	520,394*
			WMH	772.6*	99.32*	99.97*	633,697*
Uniform	10	30	MCM	21.6	98.79	99.92	24,115
			WMH	23.1	99.49	99.98	25,048
		70	MCM	25.0	98.35*	99.88	23,021
			WMH	29.6	99.96*	100.00	26,371
	500	30	MCM	429.2*	75.32*	96.74*	766,385*
			WMH	588.8*	93.95*	99.77*	521,103*
		70	MCM	624.2*	76.74*	97.00*	671,491*
			WMH	987.7*	99.97*	100.00*	803,560*
Erlang	10	30	MCM	22.9	96.27*	99.31*	27,760
			WMH	25.5	97.46*	99.56*	28,692
		70	MCM	26.7	95.82*	99.20*	25,680
			WMH	33.8	98.52*	99.77*	30,219
	500	30	MCM	468.3*	81.95*	95.51*	1,170,221*
			WMH	686.8*	91.25*	98.26*	926,454*
		70	MCM	638.7*	81.09*	95.24*	872,565*
			WMH	1106.7*	95.66*	99.23*	972,031*
Exponential	10	30	MCM	24.4	93.89*	98.50*	34,304
			WMH	29.0	95.73*	99.08*	34,867
		70	MCM	27.3	93.65*	98.47*	28,060
			WMH	39.9	97.97*	99.59*	35,487
	500	30	MCM	547.2*	81.25*	94.11*	1,754,816*
			WMH	868.9*	90.53*	97.61*	1,368,247*
		70	MCM	709.5*	80.22*	93.51*	1,194,424*
			WMH	1434.4*	94.88*	98.86*	1,311,264*

*Significant at the 0.05 level.

is presented in Fig. 9. The x -axis designates the trend value. (All = all data, No = data with no trend component, Yes = data with trend.)

Table 4 gives results of the analysis of the forecast accuracy measures for the aircraft data. Means, medians and their associated 95% confidence intervals are given for each of the error measures for each of the three ways that the data were considered.

WMH yields statistically significantly better values for the mean and median of GMAE in all cases except when considering only the data with trend where the mean difference is not statistically significant.

No significant difference in the mean MASE is observed between MCM and WMH. However, the

median MASE is statistically lower for the WMH method in each of the three data cases. For the MAPE measure: MCM performs better for items with no trend; WMH performs better, on average, for items that exhibit a trend; and MCM yields a statistically smaller median value based on all the data.

4.2.2. Accuracy implication metrics

Table 5 gives results of the analysis of the accuracy implication metrics for the aircraft data. Means, medians and their associated 95% confidence intervals are given for each of the error measures for each of the three ways that the data were considered.

Fig. 7. Confidence limits for the difference of SERV0/1 and SERV%.

These results are remarkably consistent and give clear guidance regarding the choice of methods when these accuracy implication metrics are important. There were statistically significant differences in almost all pairs of means and medians except for mean TC. MCM gave lower mean and median values for AVGINV in every case. WMH produced much higher mean and median service levels in all cases and for both definitions of service level. With regard to mean TC, there was no

significant difference between the methods in any case. However, MCM yielded lower median TC in each case.

4.2.3. Relationship of the criteria to degree of intermittency

For each of the criteria considered, the relationship of the criteria to the proportion of zeros is examined. This is done by partitioning the real data into two sets: one where $0.25 \leq CV^2 \leq 0.49$ and the

Fig. 8. Confidence limits for the difference of $TC_{MCM} - TC_{WMH}$.

Fig. 9. Confidence limits for the difference in MAPE, GMAE and MASE.

other where $0.50 \leq CV^2 \leq 0.75$, labeled low CV^2 and high CV^2 , respectively. The proportion of zeros in the data series ranges from 30% to 90%.

For the low CV^2 case, when the 95% confidence intervals for mean values of GMAE for MCM and WMH are plotted as a function of the % zeros the intervals always overlap for the two methods as depicted in Fig. 10. The same holds true for the high CV^2 case although this graph is omitted.

When the criterion is MASE the forecasting problem is much more difficult at extremely high levels of intermittency as shown with the lower chart in Fig. 10. The confidence intervals for the mean performance of the two methods overlap in every case except at 80% zeros in this low CV^2 case although the point estimate for the mean value for MCM is always equal to or higher than that for WMH. A similar result appears in the high CV^2 case although that graph is omitted.

Although graphs such as shown in Fig. 10 were constructed for every forecast accuracy metric and every accuracy implication metric, little of note was observed except in the case of service level where the measures performed similarly for SERV0/1 and SERV% for high and low CV^2 . The most interesting plot is shown in Fig. 11 which plots AVGINV, TC and SERV0/1 as a function of % zeros for the low CV^2 case. In the case of

Table 4

Confidence intervals for forecast accuracy metrics for aircraft data

Data	Measure	Method	Confidence intervals for mean			Confidence intervals for median		
			LCL	Mean	UCL	LCL	Median	UCL
Without trend $n = 3631$	GMAE	MCM	3.93	4.38*	4.82	1.52	1.60*	1.69
		WMH	1.81	2.13*	2.45	0.37	0.42*	0.47
	MASE	MCM	1.47	1.79	2.10	1.04	1.05*	1.07
		WMH	1.14	1.36	1.57	0.94	0.96*	0.97
	MAPE	MCM	119.35	128.45*	137.55	74.26	75.61*	77.27
		WMH	142.25	178.80*	215.35	93.30	95.67*	97.98
With trend $n = 594$	GMAE	MCM	4.87	6.65	8.42	2.10	2.37*	2.77
		WMH	1.72	3.56	5.40	0.27	0.49*	0.63
	MASE	MCM	1.36	2.11	2.86	1.06	1.10*	1.15
		WMH	1.04	1.42	1.81	0.91	0.94*	0.97
	MAPE	MCM	151.41	171.66*	191.91	87.25	95.61	107.20
		WMH	114.49	127.05*	139.61	83.09	87.04	90.51
All data $n = 4225$	GMAE	MCM	4.24	4.67*	5.15	1.60	1.69*	1.76
		WMH	1.95	2.33*	2.71	0.38	0.42*	0.48
	MASE	MCM	1.54	1.83	2.12	1.05	1.06*	1.07
		WMH	1.17	1.37	1.56	0.94	0.95*	0.96
	MAPE	MCM	126.21	134.56	142.91	75.87	77.48*	79.31
		WMH	139.97	171.53	203.09	91.77	93.88*	96.05

*Significant at the 0.05 level.

Fig. 10. Behavior of GMAE and MASE with respect to intermittency.

Fig. 11. Behavior of service measures with respect to intermittency.

SERV0/1 the performances of the two methods begin to diverge significantly at 70% zeros. At 90% zeros WMH satisfies demand in almost 100% of the cases while the performance of MCM falls to under 40. This improved performance is achieved with statistically significantly more average inventory in the WMH case. The compelling case in favor of WMH is that the superior performance in service level is achieved but there is no statistically significant difference between the two methods in TC.

5. Conclusions and future research directions

Spare parts inventories frequently display intermittent demand streams, which are difficult to forecast. Croston's method introduced in 1972 is the seminal work and is often used as benchmark when testing new methods. In this study we investigated an alternative method developed by Wright (1986) for data sets with missing values and compared it to a modified version of Croston's method in a simulated environment. The two

Table 5
Confidence intervals for accuracy implication metrics for aircraft data

Data	Measure	Method	Confidence intervals for mean			Confidence intervals for median		
			LCL	Mean	UCL	LCL	Median	UCL
Without trend $n = 3631$	AVGINV	MCM	14.74	15.76*	16.78	8.40	8.68*	8.99
		WMH	38.02	42.18*	46.34	18.66	19.32*	19.98
	SERV0/1	MCM	72.55	73.53*	74.51	80.00	80.95*	81.82
		WMH	94.52	94.81*	95.10	100.00	100.00*	100.00
	SERV%	MCM	81.86	82.74*	83.62	91.96	92.55*	93.07
		WMH	97.93	98.06*	98.19	100.00	100.00*	100.00
	TC	MCM	2051.20	2337.74	2624.28	752.10	785.46*	824.08
		WMH	2143.59	2405.41	2667.23	929.31	963.49*	999.90
With trend $n = 594$	AVGINV	MCM	21.55	24.87*	28.19	11.85	13.40*	15.32
		WMH	49.61	57.71*	65.81	26.28	28.87*	32.33
	SERV0/1	MCM	77.35	79.64*	81.93	85.71	87.75*	90.00
		WMH	94.27	95.18*	96.09	100.00	100.00*	100.00
	SERV%	MCM	86.41	88.21*	90.02	95.09	96.07*	97.33
		WMH	97.66	98.14*	98.62	100.00	100.00*	100.00
	TC	MCM	1927.74	3900.99	5874.25	1000.24	1108.29*	1215.22
		WMH	2464.11	3364.89	4265.67	1291.98	1404.63*	1575.93
All data $n = 4225$	AVGINV	MCM	16.05	17.05*	18.05	8.79	9.11*	9.46
		WMH	40.62	44.38*	48.14	19.56	20.24*	20.95
	SERV0/1	MCM	73.49	74.39*	75.29	81.25	81.82*	83.33
		WMH	94.58	94.86*	95.14	100.00	100.00*	100.00
	SERV%	MCM	82.71	83.51*	84.31	92.59	93.12*	93.49
		WMH	97.94	98.07*	98.20	100.00	100.00*	100.00
	TC	MCM	2185.96	2556.94	2927.92	788.03	823.02*	859.71
		WMH	2281.37	2539.75	2798.13	979.74	1019.22*	1058.35

*Significant at the 0.05 level.

methods were also tested using a real data set of aircraft parts demand. In each case a common (s, S) inventory policy, Naddor's heuristic, was employed in order to obtain accuracy implication metrics.

Some of the data streams in the real data set exhibit a statistically significant trend and this effect is considered. The performance of each of the methods is examined using three forecast accuracy measures as well as four accuracy implication metrics. The analysis confirms that, regardless of how the performance of the methods is measured, the mean level of demand and the level of intermittency are significant factors as expected. When the forecast accuracy measures are considered the results are mixed for the simulated data as well as the real data set. The most practically significant finding of the study appears when the accuracy implication metrics are analyzed for the real data set. It is consistently true that MCM always carries less inventory than WMH at the expense of service level. WMH provides superior

service levels in every case and it does so with no significant difference in total cost. These results hold for the complete data set as well as when the data is partitioned into that with and without trend. If a firm's competitive priorities lie in customer satisfaction the results of this study strongly suggest that WMH should be chosen over MCM.

A significant portion of the items in the rather large real data set considered in this study exhibited significant trend. This suggests that identifying key characteristics of an item with intermittent demand may help in developing alternative forecasting and ordering policies that can relate to specific characteristics of stock keeping units. For example, a cross-correlation between demand interval and size may lead to better forecasting techniques that exploit this characteristic. Furthermore, we need to have a better understanding of which theoretical distributions may represent intermittent demand when designing simulation studies. In this study, we used four continuous distributions all of which have appeared

in intermittent demand literature. We do not, however, make a case for any of these distributions being an excellent fit.

This study focused on one-step-ahead forecasts. Another direction of future research is to investigate the effects of long lead times. Bootstrapping is already being successfully applied to the estimation of lead time distribution of demand. The non-parametric approach of bootstrapping along with its relaxation of the Bernoulli process assumption regarding the distribution of zeros in demand streams should have wide application.

Finally, in this study we utilized forecast accuracy metrics as well as accuracy implication metrics. There is an ongoing discussion on appropriate accuracy metrics for intermittent demand forecasting. Our study showed that GMAE is not a good measure to be used with Wright's method because this method using a trend line does actually forecast zero demand, frequently resulting in zero errors when intermittency is high. This result suggests that further research in designing performance metrics for forecasting intermittent demand is needed.

Acknowledgments

This research was partially supported by a Robins School of Business research grant. Their support is gratefully acknowledged. The author's would also like to thank the three referees for their insightful review and valuable comments.

References

- Boylan, J.E., Syntetos, A., 2006a. The accuracy of a modified Croston procedure. *International Journal of Production Economics*, available online, doi:10.1016/j.ijpe.2006.10.005.
- Boylan, J.E., Syntetos, A., 2006b. Accuracy and accuracy-implication metrics for intermittent demand. *Foresight: The International Journal of Applied Forecasting* 4, 39–42.
- Croston, J.D., 1972. Forecasting and stock control for intermittent demands. *Operational Research Quarterly* 23 (3), 289–303.
- Eaves, A.H.C., Kingsman, B.G., 2004. Forecasting for the ordering and stock-holding of spare parts. *Journal of the Operational Research Society* 55, 431–437.
- Flores, B.E., Olson, D.L., Pearce, S.L., 1993. Use of cost and accuracy measures in forecasting method selection: a physical distribution example. *International Journal of Production Research* 31 (1), 139–160.
- Gardner, E.S., Koehler, A.B., 2005. Comments on a patented bootstrapping method for forecasting intermittent demand. *International Journal of Forecasting* 21, 617–618.
- Ghobbar, A.A., Friend, C.H., 2003. Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model. *Computers & Operations Research* 30, 2097–2114.
- Hicks, C.R., 1982. *Fundamental Concepts in the Design of Experiments*, third ed. Saunders College Publishing, Orlando, FL.
- Hoover, J., 2006. Measuring forecast accuracy: omissions in today's forecasting engines and demand planning software. *Foresight: The International Journal of Applied Forecasting* 4, 32–35.
- Hua, Z.S., Zhang, B., Yang, J., Tan, D.S., 2007. A new approach of forecasting intermittent demand for spare parts inventories in the process industries. *Journal of the Operational Research Society* 58, 52–61.
- Hyndman, R.J., 2006. Another look at forecast-accuracy metrics for intermittent demand. *Foresight: The International Journal of Applied Forecasting* 4, 43–46.
- Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting* 22, 679–688.
- Janssen, F., 1998. *Inventory management systems: control and information issues*. Published Ph.D. Thesis, Centre for Economic Research, Tilburg University, The Netherlands.
- Johnston, F.R., Boylan, J.E., 1996a. Forecasting for items with intermittent demand. *Journal of the Operational Research Society* 47 (1), 113–121.
- Johnston, F.R., Boylan, J.E., 1996b. Forecasting intermittent demand: a comparative evaluation of Croston's method. *International Journal of Forecasting* 12, 297–298.
- Law, A.M., Kelton, W.D., 2000. *Simulation Modeling and Analysis*, third ed. McGraw-Hill, New York, NY.
- Lawrence, M., 2004. Commentary on: a new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting* 20, 389–390.
- Leven, E., Segerstedt, A., 2004. Inventory control with a modified Croston procedure and Erlang distribution. *International Journal of Production Economics* 90, 361–367.
- Lindsey, M., Pavur, R., 2005. A comparison of methods for forecasting intermittent demand with increasing or decreasing demand occurrences. In: *Proceedings of the Annual Decision Sciences Institute Conference*, November 19–22, pp. 16921–16926.
- Naddor, E., 1975. Optimal and heuristic decisions in single- and multi-item inventory systems. *Management Science* 21 (11), 1234–1249.
- Rao, A.V., 1973. A comment on: forecasting and stock control for intermittent demands. *Operational Research Quarterly* 24 (4), 639–640.
- 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.
- Shenstone, L., Hyndman, R.J., 2005. Stochastic models underlying Croston's method for intermittent demand forecasting. *Journal of Forecasting* 24, 389–402.
- Smart, C.N., 2002. Accurate intermittent demand/inventory forecasting: new technologies and dramatic results. In: *Proceedings of APICS International Conference*, p. C-08.
- Smart, C.N., Willemain, T.R., 2000. A new way to forecast intermittent demand. *APICS The Performance Advantage* 64–68.

- Snyder, R., 2002. Forecasting sales of slow and fast moving inventories. *European Journal of Operational Research* 140, 684–699.
- Syntetos, A.A., 2001. Forecasting of intermittent demand. Unpublished Ph.D. Thesis, Buckinghamshire Business School, Brunel University.
- Syntetos, A.A., Boylan, J.E., 2001. On the bias of intermittent demand estimates. *International Journal of Production Economics* 71, 457–466.
- Syntetos, A.A., Boylan, J.E., 2005. The accuracy of intermittent demand estimates. *International Journal of Forecasting* 21, 303–314.
- Syntetos, A.A., Boylan, J.E., 2006. On the stock control performance of intermittent demand estimators. *International Journal of Production Economics* 103, 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, 495–503.
- Willemain, T.R., 2006. Forecast-accuracy metrics for intermittent demands: look at the entire distribution of demand. *Forecasting: The International Journal of Applied Forecasting* 4, 36–38.
- 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, 529–538.
- 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, 375–387.
- Willemain, T.R., Smart, C.N., Schwarz, H.F., 2005. Author's response to Koehler and Gardner. *International Journal of Forecasting* 21, 619–620.
- Williams, T.M., 1984. Stock control with sporadic and slow-moving demand. *Journal of the Operational Research Society* 35 (10), 939–948.
- Wright, D.J., 1986. Forecasting data published at irregular time intervals using an extension of Holt's method. *Management Science* 32 (4), 499–510.