Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
International Journal of Forecasting 20 (2004) 375–387
international journal of forecasting
www.elsevier.com/locate/ijforecast

A new approach to forecasting intermittent demand for service parts inventories☆

Thomas R. Willemain*, Charles N. Smart, Henry F. Schwarz

Smart Software, Inc., Belmont, MA 02478, USA

Abstract

A fundamental aspect of supply chain management is accurate demand forecasting. We address the problem of forecasting intermittent (or irregular) demand, i.e. random demand with a large proportion of zero values. This pattern is characteristic of demand for service parts inventories and capital goods and is difficult to predict. We forecast the cumulative distribution of demand over a fixed lead time using a new type of time series bootstrap. To assess accuracy in forecasting an entire distribution, we adapt the probability integral transformation to intermittent demand. Using nine large industrial datasets, we show that the bootstrapping method produces more accurate forecasts of the distribution of demand over a fixed lead time than do exponential smoothing and Croston's method.

Keywords: Accuracy; Bootstrapping; Croston's method; Exponential smoothing; Intermittent demand; Inventory; Spare parts; Service parts

1. Introduction

A fundamental aspect of supply chain management is accurate demand forecasting. We address the problem of forecasting intermittent (or irregular) demand. Intermittent demand is random demand with a large proportion of zero values (Silver, 1981). Items with intermittent demand include service (spare) parts and high-priced capital goods, such as heavy machinery. Such items are often described as 'slow moving'. Demand that is intermittent is often also 'lumpy', meaning that there is great variability among the nonzero values.

Accurate forecasting of demand is important in inventory control (Buffa & Miller, 1979; Hax & Candea, 1984; Silver, Pyke, & Peterson, 1998), but the intermittent nature of demand makes forecasting especially difficult for service parts (Swain & Switzer, 1980; Tavares & Almeida, 1983; Watson, 1987). Similar problems arise when an organization manufactures slow-moving items and requires sales forecasts for planning purposes.

Motivated by these common business problems, this paper develops a patented algorithm for forecasting the cumulative distribution of intermittent demand over a fixed lead time and a new method of assessing the accuracy of those forecasts. Both these tasks are made difficult by a high proportion of zero values in the demand history. We make two original contribu-

☆ PII of linked article S0169-2070(03)00034-7.
* Corresponding author. Department of Decision Sciences and Engineering Systems, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180-3590, USA. Tel.: +1-518-276-6622; fax: +1-518-276-8227.
E-mail address: willet@rpi.edu (T.R. Willemain).

0169-2070/$ - see front matter
doi:10.1016/S0169-2070(03)00013-X

tions: adapting the bootstrap to forecast intermittent data, and adapting the probability integral transformation to produce a measure of forecast accuracy appropriate for intermittent data. Using historical demand data on over 28,000 inventory items from nine industrial companies, we show that the new forecasting method estimates the cumulative distribution of lead time demand (LTD) over a fixed lead time more accurately than exponential smoothing and Croston's variant of exponential smoothing.

2. Related research

Much inventory management is done subjectively. However, among objective approaches, the most popular appears to be the theory of economic order quantities (EOQ), which determines two quantities for each item: a reorder point and an order quantity. When on-hand inventory reaches the reorder point, one orders an amount equal to the order quantity to replenish the stock. Calculating the order quantity normally requires forecasts only of the average demand per period. In contrast, calculating the correct reorder point requires estimates of the entire distribution of demand over the interval, known as the lead time, between the generation of a replenishment order and its arrival in inventory. (For example, a specific percentile of the predicted demand distribution, e.g. 99%, indicates the level at which the inventory reorder point should be set to insure a corresponding likelihood of not stocking out of that item during the lead time.) Textbooks simplify this problem by assuming that demands in each time period are independent and normally distributed; neither assumption is valid for intermittent demand. The bootstrap forecasting method developed here requires neither assumption.

The relevant research literature can be divided into studies of:

- assessing or relaxing the standard assumptions for non-intermittent demand (Lau & Wang, 1987; Tyworth & O'Neill, 1997);
- non-extrapolative approaches, such as reliability theory and expert systems (Petrovic & Petrovic, 1992);
- variants of the Poisson model (Ward, 1978; Williams, 1982, 1984; Mitchell, Rappold, & Faulkner, 1983; Van Ness & Stevenson, 1983; Bagchi, 1987; Schultz, 1987; Watson, 1987; Dunsmuir & Snyder, 1989);
- simple statistical smoothing methods (Bier, 1984; Sani & Kingsman, 1997);
- Croston's variant of exponential smoothing (Croston, 1972; Rao, 1973; Segerstedt, 1994; Willemain, Smart, Shockor, & Desautels, 1994; Johnston & Boylan, 1996);
- bootstrap methods (Bookbinder & Lordahl, 1989; Künsch, 1989; Wang & Rao, 1992; Efron & Tibshirani, 1993; Kim, Haddock, & Willemain, 1993; Politis & Romano, 1994; Bühlmann & Künsch, 1995; Park & Willemain, 1999).

To date, there have been two published applications of the bootstrap to inventory management for products with smooth demand. Bookbinder and Lordahl (1989) found the bootstrap superior to the normal approximation for estimating high percentiles of LTD distributions for independent data. Our approach improves on theirs by dealing with autocorrelation between successive demands and by incorporating a variant of the smoothing that they speculated would be helpful. Wang and Rao (1992) also found the bootstrap to be effective when dealing with smooth demand. They extended the analysis to autocorrelated data but did not build up LTD values from the constituent process, as we do. Neither paper considered the special problems of managing intermittent demand.

3. Industrial datasets

We assessed the relative accuracy of the various forecasting methods using industrial data from nine companies ranging in size from local to multinational. In total, we used over 28,000 inventory items. More items were available, but we excluded any with only zero values, since these items lacked any basis for a statistical forecast. The nine datasets spanned a range of types of inventory items. The items provided to us were deliberately biased toward series considered most difficult to forecast. In all cases, the items were 'live' rather than at the very end of their life cycles (as indicated, for example, by end-of-life items having 99% zero values). Table 1 summarizes the companies and datasets.

Table 1
Description of datasets and companies

Dataset	Company	Description of data	Comments
A	Major common carrier managing a fleet of aircraft	Aircraft service parts	
B	Fortune 500 manufacturer of heavy equipment	Low usage service parts	
C	British manufacturer of hardware service parts	Hardware service parts	Demand patterns often consisted of repeated values, suggesting that orders were broken into several consecutive and equal monthly installments
D	Electronics distributor	Electronics components	Previously analyzed by Flores et al. (1993); wide range of demand values
E	Manufacturer of microwave communications equipment	Service parts	
F	Small supplier of industrial refrigeration equipment	Service parts	Nonzero monthly demands were constant for many items until the last 4 months
G	Fortune 500 manufacturer of computers and instruments	Service parts with low maximum monthly demand	
H	Fortune 500 manufacturer of aircraft	Service parts	
I	Manufacturer of marine engines and equipment	Service parts	Almost all 0s, 1s and 2s; frequent negative values (returns)

All observations were monthly shipment totals, which we, like many companies, used as a proxy for monthly demand. Occasionally, negative values appeared in the datasets, indicating returned items. Although the bootstrap can accommodate negative demands, we replaced these values with zeros, in part because there was no way to attribute the returns to particular earlier shipments. Furthermore, replacing with zeros is conservative, since the alternative would be to look upon returns as a kind of random replenishment. However, one should not rely on returns as an integral part of inventory management: continuous improvement in customer service and product quality should drive returns to zero.

Table 2 summarizes the statistical characteristics of these intermittent demands. The total number of observations available per item, including observations we held out to assess accuracy, ranged from 24 to 60. Typically, well over half of the data values for an item were zeros. Nonzero demands were generally in the single-digit range and quite variable, as evidenced by the large values of the coefficient of variation (CV).

4. Forecasting methods

In this section, we explain the details of the three competing forecasting methods: exponential smoothing, Croston's method, and the new variant of the bootstrap. Despite their prominence in the literature, we did not evaluate methods based on the Poisson model due to the inappropriateness of the assumptions of some of these models (e.g. demands of only 0 or 1) and because they appeared to perform poorly on many of our industrial datasets (but see our comments below on doubly-stochastic Poisson processes). While some Poisson-based models may perform well for some items in some companies, our goal was to develop forecasting methods robust across many situations, not necessarily optimal for a limited number of specific situations. We did not

Table 2
Summary statistics for monthly intermittent demand data

Dataset	A	B	C	D	E	F	G	H	I
Data features									
No. series	8363	1000	4654	967	611	237	1464	1058	10184
No. obs./series	29	54	60	24	62	27	36	60	36
% Zero values									
Mean	49	75	84	67	93	47	85	65	93
S.D.	29	17	20	17	8	36	13	25	5
Maximum	100	98	98	88	100	100	100	98	100
75%ile	76	89	97	79	98	85	94	85	97
50%ile	48	80	92	71	95	41	89	75	94
25%ile	21	63	82	58	92	11	78	48	89
Minimum	3	9	0	8	37	4	22	2	56
Average of nonzero demands									
Mean	54	5	2777	815	19	10	2	32	1
S.D.	305	46	43202	1680	29	15	2	170	1
Maximum	5827	1393	930000	25665	250	106	35	3969	9
75%ile	13	3	12	789	23	10	2	12	2
50%ile	4	2	4	306	9	4	1	5	1
25%ile	2	1	2	117	4	2	1	2	1
Minimum	1	1	1	1	1	1	1	1	1
CV of nonzero demands									
Mean	72	58	40	96	66	65	45	83	21
S.D.	40	40	46	41	34	45	36	38	25
Maximum	437	294	364	343	163	312	261	315	111
75%ile	91	72	71	117	89	83	64	102	40
50%ile	67	53	34	92	67	61	43	78	0
25%ile	47	36	0	69	44	31	18	60	0
Minimum	0	0	0	0	0	0	0	0	0

Results rounded to zero decimals.

evaluate moving average forecasts because they are not much used in practice and we do not expect them to have any advantage over exponential smoothing in the task of forecasting the entire distribution of LTD.

Let $X(t)$ be the observed demand in period $t, t = 1 \ldots T$. This integer demand is often zero, and when it is nonzero it tends to be highly variable. Let L be the fixed lead time over which forecasts are desired. Our goal is to estimate the entire distribution of the sum of the demands over the lead time, called the lead time demand (LTD),

$$\mathrm{LTD} = \sum X(t), t = T + 1 \ldots T + L. \quad (1)$$

In our analysis of industrial datasets, we held out the last L observations, fitting our models on the first $T - L$ values. Since our datasets were recorded monthly, we used lead times of 1, 3 and 6 to reflect a realistic range of lead times.

4.1. Exponential smoothing

Exponential smoothing has proven to be a robust forecasting method and is probably the most used of the statistical approaches to forecasting intermittent demand. Our exponential smoothing forecasts assume that LTD is a normally distributed sum of L iid random variables and use the smoothing process to estimate the mean and variance of the normal distribution. Of course, the normality assumption is implausible for intermittent demand, especially for short lead times that do not permit Central Limit Theorem effects for the sum. Nevertheless, following

Croston (1972), we do use a normal distribution, which requires specification of the mean and standard deviation.

We computed the mean and variance of the assumed normal distribution of LTD as follows. We estimated the mean level of demand at time $t, M(t)$, using

$$M(t) = \alpha X(t) + (1 - \alpha)M(t-1), t = 1 \ldots T, \quad (2)$$

where α is a smoothing constant between 0 and 1. For each inventory item, we selected the value of α that minimized the sum of squared residuals $\sum(X(t) - M(t))^2, t = 1 \ldots T$. We initialized the smoothing using the average of the first two demands

$$M(0) = [X(1) + X(2)]/2. \quad (3)$$

We estimated the mean of the L demands over the lead time as $L \cdot M(T)$. We estimated the common variance, V, from the one-step ahead forecast errors using

$$V = (1/T) \cdot \sum [X(t) - M(t-1)]^2, t = 1 \ldots T. \quad (4)$$

If an item had no variation (i.e. the only nonzero demands were in the holdout period), we used $V = 0.001$ as a default value. Finally, assuming iid demands, we estimated the variance of the LTD distribution as $L \cdot V$.

4.2. Croston's method

Croston's method was developed to provide a more accurate estimate of the mean demand per period. Like exponential smoothing, Croston's method assumes that LTD has a normal distribution. Previous work (Willemain et al., 1994; Johnston & Boylan, 1996) established that Croston's method provides more accurate forecasts of the mean demand per period than exponential smoothing.

Croston's method estimates the mean demand per period by applying exponential smoothing separately to the intervals between nonzero demands and their sizes. Let $I(t)$ be the smoothed estimate of the mean interval between nonzero demands, and let $S(t)$ be the smoothed estimate of the mean size of a nonzero demand. Let q be the time interval since the last nonzero demand. Croston's method works as follows: if $X(t) = 0$ then

$$S(t) = S(t-1) \quad (5)$$

$$I(t) = I(t-1)$$

$$q = q + 1$$

else

$$S(t) = \alpha X(t) + (1 - \alpha)S(t-1)$$

$$I(t) = \alpha q + (1 - \alpha)I(t-1)$$

$$q = 1.$$

Combining the estimates of size and interval provides the estimate of mean demand per period

$$M(t) = S(t)/I(t). \quad (6)$$

These estimates are only updated when demand occurs. When demand occurs every review interval, Croston's method is identical to conventional exponential smoothing, with $S(t) = M(t)$. We initialized Croston's method using the time until the first event and the size of the first event. As with exponential smoothing, Croston's method regards LTD as normal with mean $L \cdot M(T)$ and variance $L \cdot V$.

4.3. Bootstrap method

Efron's (1979) bootstrap creates pseudo-data by sampling with replacement from the individual observations. Applied naively to the problem of forecasting LTD, this simple approach would ignore autocorrelation in the demand sequence and produce as forecast values only the same numbers that had already appeared in the demand history. We developed a modified bootstrap responsive to three difficult features of intermittent demand: autocorrelation, frequent repeated values, and relatively short series.

From extensive observation of the industrial data, we realized that demand often runs in streaks, with longer sequences of zero or nonzero values than one would expect from a simple Bernoulli process. Such a pattern evidences positive autocorrelation. Occasionally, the demand shows more frequent alternation between runs of zero and nonzero values than one would expect, evidencing negative autocorrelation. In either case, short term forecasts should capture and exploit this autocorrelation.

We model autocorrelation using a two state, first order Markov process. We begin with a forecast of the sequence of zero and nonzero values over the L periods of the lead time. These forecasts are conditional on whether the last demand, $X(T)$, is zero or nonzero. We estimate the state transition probabilities from the historical demand series using started counts (Mosteller & Tukey, 1977).

Once we have a forecast of the sequence of zero and nonzero values over the lead time from the two-state Markov process, we need to give specific numerical values to the nonzero forecasts. The most direct way to do this is to simply sample from the nonzero values that have appeared in the past. However, this would imply that no different values would ever appear in the future. The result would be unrealistic bootstrap samples, especially poor estimates of the tails of the LTD distribution made from short data series. This issue has long been recognized as a potential weakness of bootstrapping when the goal is to produce life-like samples (Taylor & Thompson, 1992). An analogy is available from order statistics from a continuous distribution. There, a well-known result (Mosteller & Rourke, 1973) is that the largest observed value among n independent samples estimates, on average, only the $n/(n+1)$st fractile of the distribution. Instead, after we select one of the nonzero demand values at random, we jitter it (i.e. add some random variation) to allow some chance for a nearby value to be used instead. For instance, if the randomly chosen nonzero demand is 7, we might use, say, 7 or 6 or 10 as the forecast. The jittering process is designed to allow greater variation around larger demands. This comports with the common empirical phenomenon, noted in the industrial datasets, that variances often increase when means increase.

Let X^* be one of the historical demand values selected at random, and let Z be a standard normal random deviate. The jittering process works as follows:

$$\begin{aligned}&\text{JITTERED} = 1 + \text{INT}\{X^* + Z\sqrt{X^*}\}\\&\text{IF JITTERED} \leq 0, \text{THEN JITTERED} = X^*.\end{aligned} \tag{7}$$

Experiments not reported here revealed that this form of jittering generally improves accuracy, especially for short lead times. Using the measure of accuracy defined in the next section, jittering improved accuracy by 48±19% (±S.E.) for lead times of 1 month, 41±11% for 3 months, and 11±10% for 6 months.

Summing the forecasts over each period of the lead time gives one forecast of LTD. We repeat the process until we have 1000 bootstrap forecasts estimating the entire distribution of LTD. (Commercial software allows for many more than 1000 replications and therefore produces even better performance than that reported below.)

Here is a concise summary of the steps in the bootstrap approach.

- Step 0 Obtain historical demand data in chosen time buckets (e.g. days, weeks, months).
- Step 1 Estimate transition probabilities for two-state (zero vs. nonzero) Markov model.
- Step 2 Conditional on last observed demand, use Markov model to generate a sequence of zero/nonzero values over forecast horizon.
- Step 3 Replace every nonzero state marker with a numerical value sampled at random with replacement from the set of observed nonzero demands.
- Step 4 Jitter the nonzero demand values.
- Step 5 Sum the forecast values over the horizon to get one predicted value of LTD.
- Step 6 Repeat steps 2–5 many times.
- Step 7 Sort and use the resulting distribution of LTD values.

5. Assessing forecast accuracy

Assessing the performance of forecasting methods for intermittent demand required some invention. When assessing the accuracy of forecasts of many series of smooth demand, it is conventional to use the mean absolute percentage error (MAPE). However, when one or more of the observed demands is zero, as is often true with intermittent demand, the MAPE is undefined.

A more fundamental problem is that we need to assess the quality not of a point forecast of the mean but of a forecast of an entire distribution. Calculations of reorder points in (s, Q) and (R, S) inventory models require an integral over the distribution of LTD. More informal approaches also require knowledge of the distribution rather than just the mean value. Inventory managers concerned with worst case parts availability scenarios focus on the upper tail, i.e. high percentiles, of the distribution. Treasurers concerned with worst case revenue scenarios focus on the lower tail.

Estimating the entire distribution of LTD from sparse data is challenging. Given historical data on item demand, one can hold out the final L periods, sum their demands, and thereby obtain a single observed value for the LTD for that item. The problem is how to use this single holdout value to assess the quality of an entire predicted distribution of LTD for that item. Consistent with the literature (Gardner, 1990; Flores, Olson, & Pearce, 1993), the industrial datasets we studied were 'short and wide', e.g. there were only a few dozen monthly observations on each of hundreds or thousands of inventory items. This meant that we could afford to create only one holdout value of LTD for each item, making it impossible to compare the three forecast distributions on an item-specific basis. However, there were data for many items, so we pooled across items. To respond to these two problems, we adapted the well-known probability integral transformation method.

Our solution involves pooling percentile estimates across items. The bootstrap method estimates the entire distribution of LTD. Any single observed LTD value corresponds to some percentile of that distribution. If the bootstrap method accurately predicts the distribution, then the observed percentiles will be uniformly distributed between 0 and 100.

To understand this approach, first consider the easier problem of assessing the predicted distribution of a continuous random variable. Among the methods used to solve such problems are the chi-square test and tests based on the probability–integral transformation; our solution uses both.

The probability–integral transformation converts an observed value of a random variable X into its corresponding fractile using the CDF, $F(X)$. For example, if the median value is 2.35, then the transformation would convert an observed value of 2.35 into 0.5. If $F(X)$ is known exactly, the transformation converts the observed distribution into a uniform distribution. In the more realistic case of having to use an estimate $F^\wedge(X)$, sampling variability leads to departures from uniformity (Kendall & Stuart, 1979). Nevertheless, we can take the degree of conformance of the estimated fractiles to the uniform ideal as an indicator of the relative quality of the estimated CDF.

However, we are not dealing with a continuous random variable, since LTD is discrete. Because the CDF of LTD is defined only at integer values in this case, only a discrete set of fractiles are observed. To convert this problem back into one in which we assess departures from uniformity, we first divide the possible fractiles into 20 bins of width 0.05. When an LTD count of X is observed, we calculate $F^\wedge(X) - F^\wedge(X-1)$ and distribute that probability proportionally (i.e. using linear interpolation) over the relevant bins (taking $F^\wedge(-1) = 0$). Then we compare the distribution of counts in the 20 bins to a uniform distribution, summarizing the departure from uniformity using the chi-square statistic. The choice of 20 bins is an appropriate compromise between resolution and statistical stability, given the large numbers of items in our industrial datasets. Other numbers of bins would be appropriate if there were more or fewer items.

To illustrate this method of fractional allocation of counts, consider a distribution of LTD that is estimated to be Poisson with mean 1.3. In this case, $F^\wedge(0) = 0.273$ and $F^\wedge(1) = 0.627$. Now suppose that the observed LTD is 1. Our method distributes $F^\wedge(1) - F^\wedge(0) = 0.354$ uniformly over the range from 0.273 to 0.627. Thus we credit 0.076 = (0.30−0.273)/0.354 of a count to the bin spanning 0.25 to 0.30, 0.141 = 0.05/0.354 of a count to each of the six bins spanning the range from 0.30 to 0.60, and 0.076 = (0.627−0.60)/0.354 of a count to the bin spanning 0.60 to 0.65. We accumulate these fractional counts across items, then assess the uniformity of the total counts across the 20 bins using the chi-square statistic. The resulting chi-square statistic is a meaningful figure of merit whose logarithm is suitably distributed for confidence interval calculations based on Student's *t* distribution. Forecasting methods with smaller mean values of the chi-square statistic are more accurate.

6. Results

We compared exponential smoothing and Croston's method with the bootstrap on the basis of the uniformity of observed LTD percentiles. For each of the nine industrial datasets, we compared the three methods using lead times of 1, 3 and 6 months. Table 3 shows the results of the accuracy comparisons. We analyzed the logarithms of the chi-square values in Table 3 using fixed-effects ANOVA, with forecasting method, lead time, and company as factors. Table 4 shows the ANOVA results. All main effects were highly significant, as were the interactions with forecast method. The main effects and interaction of method and lead time are clearly visible in Fig. 1, which plots 95% confidence intervals for the mean values of log chi-square.

These results, supplemented by Newman–Keuls multiple-comparison tests, support the following statements.

1. The bootstrap method was the most accurate forecasting method.
2. Despite its ability to provide more accurate estimates of mean demand per period, Croston's method had no statistically significant advantage over exponential smoothing at forecasting the entire LTD distribution; in fact, Croston's method was slightly less accurate at every lead time.

Table 3
Chi-square values by lead time and forecasting method

Dataset	Lead time	No. of series	Expo smooth	Croston's method	Bootstrap
A	1	8335	2481	2853	43
B	1	999	602	648	9
C	1	4633	3747	4896	50
D	1	967	559	561	32
E	1	598	493	586	10
F	1	219	143	131	18
G	1	1446	1068	1210	4
H	1	1058	778	760	100
I	1	9758	7758	10706	6
	3	8333	1239	1935	130
A					
B	3	999	495	537	65
C	3	4571	3504	6130	175
D	3	967	569	555	193
E	3	570	416	526	29
F	3	214	157	203	61
G	3	1378	628	770	44
H	3	1057	739	706	259
I	3	9608	4244	9277	228
	6	8332	2768	4646	3030
A					
B	6	998	339	385	168
C	6	4484	3333	7987	388
D	6	951	924	1109	790
E	6	539	360	559	67
F	6	211	236	118	110
G	6	1278	422	563	176
H	6	1055	778	839	528
I	6	9347	2838	10239	441

Table 4
Analysis of variance of logarithm of chi-square values

Source	DF	Sum of squares	Mean square	F-ratio	Prob. level	Power ($\alpha = 0.05$)
Method	2	110.30	55.15	299.38	0.00	1.00
Lead	2	10.98	5.49	29.8	0.00	0.99
Method×lead	4	27.09	6.77	36.76	0.00	1.00
Company	8	67.94	8.49	46.11	0.00	1.00
Method×company	16	20.73	1.30	7.03	0.00	1.00
Lead×company	16	3.81	0.24	1.29	0.26	0.39
Error	32	5.89	0.18			
Total (adjusted)	80	246.74				
Total	81					

3. The accuracy of the bootstrap method decreased with lead time, which is to be expected; however, the bootstrap remained the most accurate of the three methods even at a lead time of 6 months.
4. Lead time had little effect on the accuracy of exponential smoothing and Croston's method. This surprising result might be caused by offsetting factors: accuracy should decrease with lead time, but the normality assumption should improve as more demands are summed.
5. There were large differences in the average 'forecastability' of items from different companies. Compared to items from the rest of the companies, those from company F were significantly easier to forecast, whereas items from companies A, C and I were significantly more difficult.

The results above aggregate across the full range of possible LTD values. As a referee pointed out, it is also useful to disaggregate the results so that performance at the tails of the distribution is visible. Fig. 2 compares the three methods in terms of the accuracy of their calibration. In a perfectly calibrated forecast of the distribution of LTD, each of the 20 percentile bins (0–5%, 5–10%, etc.) would contain exactly 5% of the held-out LTD values. It is clear from Fig. 2 that all the methods were least well calibrated in the tails, but the

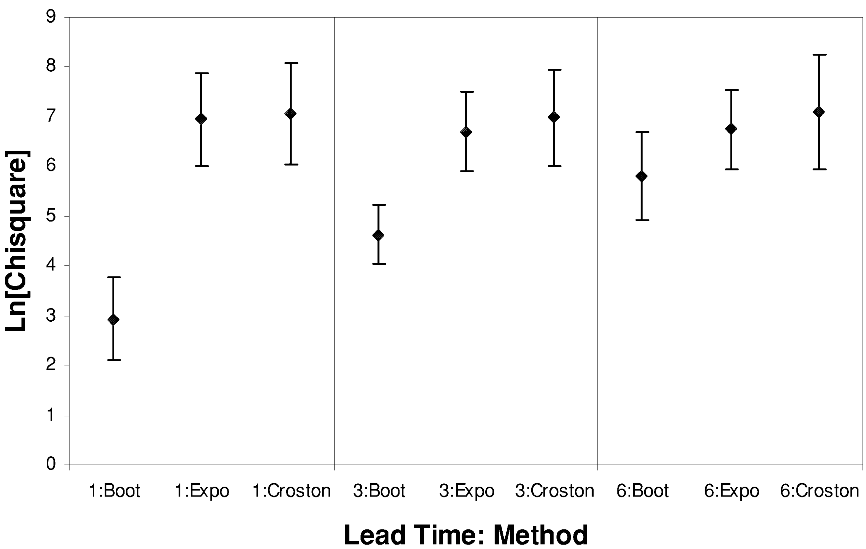

Fig. 1. Accuracy based on uniform distribution of percentiles. Note: bars show 95% confidence intervals for mean of logarithm of chi-square values across companies. Lower values are better.

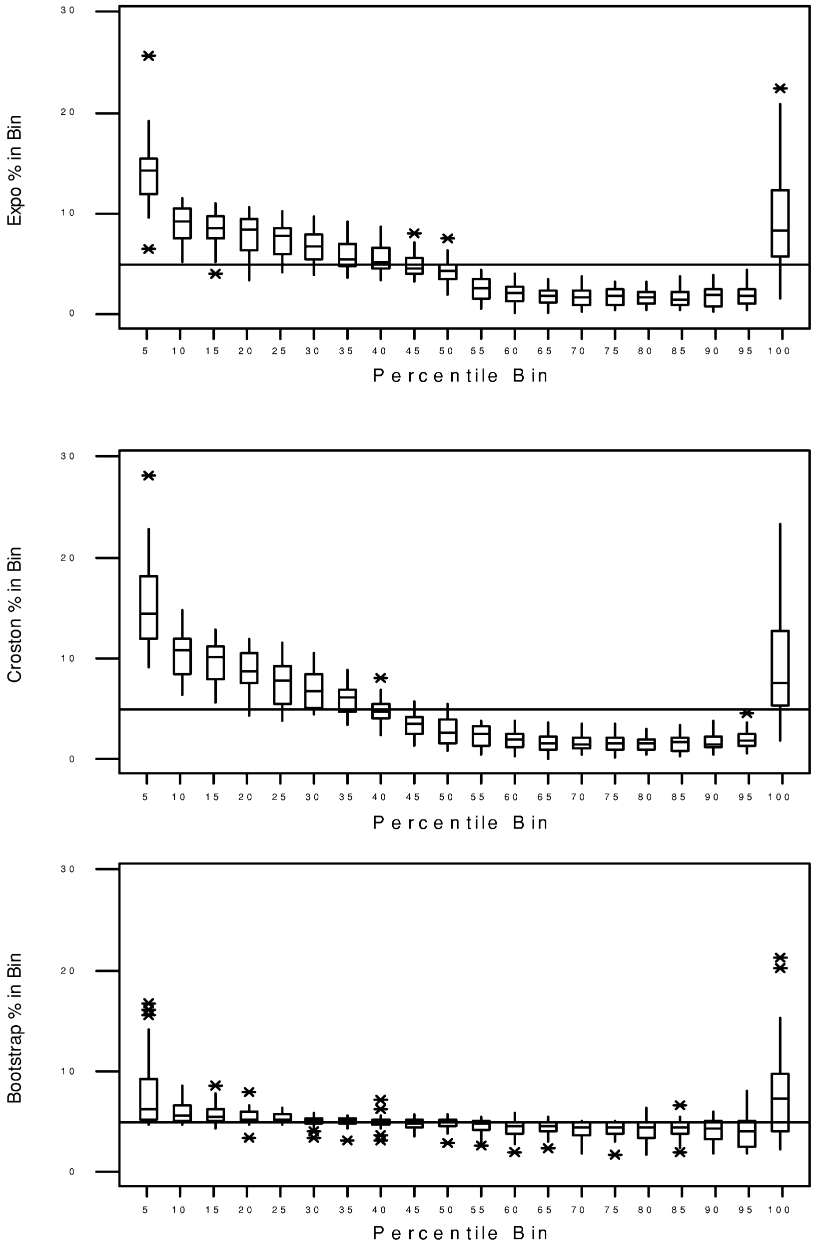

Fig. 2. Percentage of held-out LTD values falling into each 5-percentile bin.

bootstrap forecasts performed best across all the percentile bins. For example, consider the 95–100% bin. The mean percentages of held-out LTDs falling into that bin were 9.6% for exponential smoothing, 9.7% for Croston's method, and 7.8% for the bootstrap.

7. Conclusions

Intermittent demand for service parts and high-priced capital goods is difficult to forecast. This difficulty complicates the problem of managing inventory efficiently, which requires an acceptable balance between costly excess inventory on the one hand and disruptive stockouts on the other. To strike this balance, inventory management models require accurate forecasts of the entire distribution of an aggregated quantity: cumulative demand over the fixed lead time required to receive replenishment orders.

When managing thousands of items whose demand is intermittent, it is valuable to have a fast, automatic mechanism for calculating the parameters of inventory models. Current practice, where it uses statistical forecasting, favors exponential smoothing. An alternative recommended in inventory control textbooks is Croston's method.

In this paper, we have developed a new bootstrapping approach to forecasting the distribution of the sum of intermittent demands over a fixed lead time. To compare the relative accuracy of bootstrapping against exponential smoothing and Croston's method, we modified the probability integral transformation to work with intermittent data. We pooled one holdout value from each of many inventory items into an assessment of the accuracy of the forecast of the entire distribution of lead time demand across all items.

We compared the accuracy of the forecasting methods by applying them to over 28,000 items provided by nine industrial companies. Croston's method, though it can provide more accurate estimates of the mean level of demand at moments when demand occurs, did not provide an overall improvement on exponential smoothing when the task was to forecast the entire distribution of LTD. However, the bootstrap clearly did outperform exponential smoothing, especially for short lead times.

Besides improving forecast accuracy, the bootstrap approach has other advantages. One is that it can easily be modified to accommodate variable lead times simply by sampling from empirical distributions of item-specific lead times. Another is its attractiveness to practitioners. In conversations, they indicate an appreciation for the bootstrap's reliance on the detailed empirical history of each inventory item rather than on tenuous mathematical assumptions, and they find the skewness of the bootstrap distributions more realistic than a conventional normal distribution. Finally, users intuitively grasp the simple procedural explanation of how the bootstrap works. Their comfort with the bootstrap approach may derive from the concrete, algorithmic nature of computational inference, in contrast to the more abstract character of traditional mathematical approaches to statistical inference.

While we believe the bootstrapping approach to forecasting intermittent demand is a powerful new option, we recognize that there remain several difficulties and unsolved problems.

(a) Problems with jittering. As a reviewer pointed out, there are some circumstances in which jittering will create forecasts that are not physically meaningful. For example, suppose that the product in question is a beer that is sold only in packages containing either 6, 12, or 24 cans. Clearly, every LTD must be divisible by 6, yet jittering could produce a forecast LTD of 13. In such cases, one may wish to skip the jittering step.
(b) Problems with nonstationarity. Every bootstrapping algorithm developed to date assumes that the data series is stationary or can be easily transformed to stationarity. Complicated trends and seasonality would therefore be a concern. The industrial datasets analyzed here did not have this problem, but we have seen retail demand data for gardening products that showed clear seasonality. Bootstrapping separately each phase of the seasonal cycle could cope with this problem, yet it might fragment the data to the point that the resulting forecasts become very unstable. More research is needed here.
(c) Alternatives based on Poisson processes. We began our research assuming that some modification of a Poisson process would be a good way to forecast intermittent demand. We were not able to find a robust, flexible and estimable model, so we switched our attention to the bootstrap approach. We discovered three problems with Poisson-based models. Firstly, they were not as flexible as nonparametric models at fitting arbitrary marginal distributions. Secondly, it was difficult to recreate the serial correlation evident in industrial data. Thirdly, it seemed that the model would inevitably be a parametric doubly-stochastic Poisson process of some kind, and we were unable to obtain good parameter estimates for such models using the short datasets typically seen in practice. Nevertheless, we regard doubly-

stochastic Poisson process models as major potential competitors to the bootstrap approach and consider them an interesting area for further research.

Acknowledgements

This research was supported by the US National Science Foundation under Small Business Innovation Research Grant DMI-9204101 to Smart Software, Inc. The authors thank the following individuals for helpful conversations: Nelson Hartunian, Eric Ratti, So Young Sohn, and Pasquale Sullo. We also acknowledge the assistance of Benito Flores and the other individuals who provided data for our study. Smart Software has received US patent #6,205,431 B1 on the intermittent demand forecasting methodology described in this article.

References

Bagchi, U. (1987). Modeling lead-time demand for lumpy demand and variable lead time. *Naval Research Logistics*, *34*, 687–704.

Bier, I. J. (1984). *Boeing commercial airplane group spares department: simulation of spare parts operations*. Detroit, MI: ORSA/TIMS Joint National Meeting.

Bookbinder, J. H., & Lordahl, A. E. (1989). Estimation of inventory re-order levels using the bootstrap statistical procedure. *IIE Transactions*, *21*, 302–312.

Buffa, E. S., & Miller, J. G. (1979). In 3rd ed, *Production–inventory systems: Planning and control*. Homewood, IL: Irwin.

Bühlmann, P., & Künsch, H. R. (1995). The blockwise bootstrap for general parameters of a stationary time series. *Scandinavian Journal of Statistics*, *22*, 35–54.

Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, *23*, 289–303.

Dunsmuir, W. T. M., & Snyder, R. D. (1989). Control of inventories with intermittent demand. *European Journal of Operational Research*, *40*, 16–21.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, *7*, 1–26.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall.

Flores, B., 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*, 139–160.

Gardner, E. S. (1990). Evaluating forecast performance in an inventory control system. *Management Science*, *36*, 490–499.

Hax, A. C., & Candea, D. (1984). *Production and inventory management*. Englewood Cliffs, NJ: Prentice-Hall.

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

Kendall, M., & Stuart, A. (1979). In 4th ed, *The advanced theory of statistics* vol. 2. London: Charles Griffin.

Kim, Y. B., Haddock, J., & Willemain, T. R. (1993). The binary bootstrap: Inference with autocorrelated binary data. *Communications in Statistics—Simulation and Computation*, *22*, 205–216.

Künsch, H. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, *17*, 1217–1241.

Lau, H. -S., & Wang, M. -C. (1987). Estimating the lead-time demand distribution when the daily demand is non-normal and autocorrelated. *European Journal of Operational Research*, *29, 60–69.*

Mitchell, C. R., Rappold, R. A., & Faulkner, W. B. (1983). An analysis of Air Force EOQ data with an application to reorder point calculation. *Management Science*, *29*, 440–446.

Mosteller, F., & Rourke, R. E. K. (1973). *Sturdy statistics*. Reading, MA: Addison-Wesley.

Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression: A second course in statistics*. Reading, MA: Addison-Wesley.

Park, D., & Willemain, T. R. (1999). The threshold bootstrap and threshold jackknife. *Computational Statistics and Data Analysis*, *31*, 187–202.

Petrovic, D., & Petrovic, R. (1992). SPARTA II: Further development in an expert system for advising on stocks of spare parts. *International Journal of Production Economics*, *24*, 291–300.

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, *89*, 1303–1313.

Rao, A.V. (1973). A comment on 'Forecasting and stock control for intermittent demands'. *Operational Research Quarterly*, *24*, 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*, 700–713.

Schultz, C. R. (1987). Forecasting and inventory control for sporadic demand under periodic review. *Journal of the Operational Research Society*, *37*, 303–308.

Segerstedt, A. (1994). Inventory control with variation in lead times, especially when demand is intermittent. *International Journal of Production Economics*, *35*, 365–372.

Silver, E. A. (1981). Operations research in inventory management: A review and critique. *Operations Research*, *29*, 628–645.

Silver, E. A., Pyke, D. F., & Peterson, R. (1998). In 3rd ed, *Inventory management and production planning and scheduling. New York: Wiley.*

Swain, W., & Switzer, B. (1980). Data analysis and the design of automatic forecasting systems. *In Proceedings of the business and economic statistics section*. American Statistical Association, pp. 219–223.

Tavares, L. V., & Almeida, L. T. (1983). A binary decision model for the stock control of very slow moving items. *Journal of the Operational Research Society*, *34*, 249–252.

Taylor, M. S., & Thompson, J. R. (1992). A nonparametric density estimation based resampling algorithm. In LePage, R., & Billard, L. (Eds.), *Exploring the limits of bootstrap*. New York: Wiley, pp. 397–403.

Tyworth, J. E., & O'Neill, L. (1997). Robustness of the normal approximation of lead-time demand in a distribution setting. *Naval Research Logistics*, *44*, 165–186.

Van Ness, P. D., & Stevenson, W. J. (1983). Reorder-point models with discrete probability distributions. *Decision Sciences*, *14*, 363–369.

Wang, M., & Rao, S. S. (1992). Estimating reorder points and other management science applications by bootstrap procedure. *European Journal of Operations Research*, *56*, 332–342.

Ward, J. B. (1978). Determining reorder points when demand is lumpy. *Management Science*, *24*, 623–632.

Watson, R. B. (1987). The effects of demand-forecast fluctuations on customer service and inventory cost when demand is lumpy. *Journal of the Operational Research Society*, *38*, 75–82.

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.

Williams, T. M. (1982). Reorder levels for lumpy demand. *Journal of the Operational Research Society*, *33*, 185–189.

Williams, T. M. (1984). Stock control with sporadic and slow-moving demand. *Journal of the Operational Research Society*, *35*, 939–948.

Biographies: Thomas R. WILLEMAIN is Vice President of Smart Software, Inc. of Belmont, MA and Professor in the Department of Decision Sciences and Engineering Systems, Rensselaer Polytechnic Institute. He holds the BSE degree from Princeton University and the SM and Ph.D. degrees from Massachusetts Institute of Technology.

Charles N. SMART is President and CEO of Smart Software, Inc. He holds the BA and MA degrees from Harvard University and the MBA degree from Massachusetts Institute of Technology.

Henry F. SCHWARZ holds the BA degree from the University of Pennsylvania and the MS degree from Rensselaer Polytechnic Institute.