International Journal of Forecasting 28 (2012) 485–496

Contents lists available at SciVerse ScienceDirect

International Journal of Forecasting

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

Forecasting the intermittent demand for slow-moving inventories: A modelling approach

Ralph D. Snyder[a,*], J. Keith Ord[b], Adrian Beaumont[a]

[a] *Department of Econometrics and Business Statistics, Monash University, Clayton, VIC 3800, Australia*

[b] *McDonough School of Business, Georgetown University, Washington, DC 20057, USA*

A R T I C L E   I N F O

Keywords:
Croston's method
Exponential smoothing
Hurdle shifted Poisson distribution
Intermittent demand
Inventory control
Prediction likelihood
State space models

A B S T R A C T

Organizations with large-scale inventory systems typically have a large proportion of items for which demand is intermittent and low volume. We examine various different approaches to demand forecasting for such products, paying particular attention to the need for inventory planning over a multi-period lead-time when the underlying process may be non-stationary. This emphasis leads to the consideration of prediction distributions for processes with time-dependent parameters. A wide range of possible distributions could be considered, but we focus upon the Poisson (as a widely used benchmark), the negative binomial (as a popular extension of the Poisson), and a hurdle shifted Poisson (which retains Croston's notion of a Bernoulli process for the occurrence of *active* demand periods). We also develop performance measures which are related to the entire prediction distribution, rather than focusing exclusively upon point predictions. The three models are compared using data on the monthly demand for 1046 automobile parts, provided by a US automobile manufacturer. We conclude that inventory planning should be based upon dynamic models using distributions that are more flexible than the traditional Poisson scheme.

1. Introduction

Modern inventory control systems may involve thousands of items, many of which show very low levels of demand. Furthermore, such items may be requested only on an occasional basis. When events corresponding to positive demands occur only sporadically, we refer to the demand as *intermittent*. When the average size of a customer order is large, a continuous distribution is a suitable description, but when it is small, a discrete distribution is more appropriate.

In this paper, our interest focuses upon intermittent demand with low volume. On occasion, such stock keeping units (SKUs) may be of very high value, such as, for example, spare aircraft engines. However, even when the individual units are of low value, it is not unusual for them to represent a large percentage of the number of SKUs, so that they collectively represent an important element in the planning process. Johnston and Boylan (1996a, p. 121) cite an example where the average number of purchases of an item by a customer was 1.32 occasions per year, and "For the slower movers, the average number of purchases was only 1.06 per item [per] customer". Similarly, in the study of car parts discussed in Section 6, out of 2509 series with complete records for 51 months, only 1046 had (a) ten or more months with positive demands, *and* (b) at least some positive demands in the first 15 and the last 15 months.

Demand forecasting for high volume products can be handled successfully using exponential smoothing methods, for which a voluminous body of literature exists; see for example Hyndman, Koehler, Ord, and Snyder (2008)

* Corresponding author. Tel.: +61 3 9905 2366; fax: +61 3 9905 5474.
E-mail addresses: ralph.snyder@monash.edu (R.D. Snyder), ordk@georgetown.edu (J.K. Ord), adrian.beaumont@monash.edu (A. Beaumont).

0169-2070/$ – see front matter © 2011 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2011.03.009

and Ord, Koehler, and Snyder (1997). When volumes are low, the exponential smoothing framework must be based upon a distribution that describes count data, rather than the normal distribution. Further, as was recently emphasized by Syntetos, Nikolopoulos, and Boylan (2010), it is not sufficient to look at point forecasts when making inventory decisions. Instead, they recommend the use of stock control metrics. We accept their viewpoint completely, but since such metrics depend upon the underlying prediction distribution, we have opted to work with such distributions directly. This choice is reinforced by the observation that prediction distributions are applicable to count problems beyond inventory control. Moreover, the information on costs and lead times required when using inventory criteria was not available for the data considered in Section 6.

The remainder of the paper is structured as follows. It begins in Section 2 with a review of the literature on forecasting intermittent demand. The focus here is on models that allow for both non-stationary and stationary features. For example, the demand for spare parts may increase over time as the machines age and then decline as they fail completely or are withdrawn from service. In Section 3, we summarize the different models which will be considered in the empirical analysis and examine how they might be estimated and how they might be used to simulate various prediction distributions. Since our particular focus is on the ability of a model to furnish the entire prediction distribution, not just point forecasts, we examine suitable performance criteria in Section 4. Issues relating to model selection are examined briefly in Section 5. In Section 6 we present an empirical study using data on the monthly demand for 1046 automobile parts. Then, in Section 7, we examine the links between forecasting and management decision making, with an illustration of the use of prediction distributions in inventory management. Finally, various conclusions from our research are summarized briefly in Section 8.

2. Review of the literature on intermittent demand

The classic paper on this topic is that of Croston (1972), with corrections by Rao (1973). Croston's key insight was that:

> When a system is being used for stock replenishment, or batch size ordering, the replenishment will almost certainly be triggered by a demand which has occurred in the most recent interval (Croston, 1972, p. 294).

The net effect of this phenomenon when forecasting the demand for a product which is required only intermittently is that the mean demand is over-estimated and the variance is under-estimated. Thus, an inventory decision based on the application of the usual exponential smoothing formulae will typically produce inappropriate stock levels. Croston then proceeded to develop an alternative approach based upon:

- an exponential smoothing scheme for updating the expected time gap between successive active demand periods;
- an exponential smoothing scheme for updating the expected demand in active periods; and
- an assumption that the time gaps and the demand in active periods are statistically independent.

Since the original paper by Croston, a number of extensions and improvements to the method have been made, notably by Johnston and Boylan (1996a) and Syntetos and Boylan (2005). Syntetos and Boylan (2001) first showed that the original Croston estimators were biased; they then (see Syntetos & Boylan, 2005) developed a new method, which we refer to as the bias-adjusted Croston method, and evaluated its performance in an extensive empirical study. Their out-of-sample comparisons indicate that the new method provides superior point forecasts for "faster intermittent" items; that is, those with relatively short mean times between active demand periods.

Snyder (2002) identified some logical inconsistencies in the original Croston method and examined the use of a time-dependent Bernoulli process. Unlike with Croston's method, distinct smoothing parameters were used for the positive demands and the time gaps. Snyder went on to develop a simulation procedure which provides a numerical determination of the predictive distribution for the lead-time demand. Shenstone and Hyndman (2005) showed that there is no possible model which will lead to the Croston forecast function unless we allow a sample space for active period demands that can take on either negative or positive values.

2.1. Low volume, intermittent demand

There is an extensive body of literature on low count time series models which is potentially applicable to forecasting the demand for slow moving items. Most expositions rely on a Poisson distribution to represent the counts but introduce serial correlation through a changing mean (and variance). Models based on lagged values of the count variable essentially have a single source of randomness (Davis, Dunsmuir, & Wang, 1999; Heinen, 2003; Jung, Kukuk, & Liesenfeld, 2006; Shephard, 1995). By contrast, models which are based upon unobservable components have an additional source of randomness driving the evolution of the mean (Davis, Dunsmuir, & Wang, 2000; Durbin & Koopman, 2001; Harvey & Fernandes, 1989; West & Harrison, 1997; West, Harrison, & Migon, 1985; Zeger, 1988). In addition, there are also several multiple source of error approaches based upon integer-valued autoregressive (INAR) models (Al-Osh & Alzaid, 1987; McCabe & Martin, 2005; McKenzie, 1988).

Single- and dual-source of error models for count data were compared by Feigin, Gould, Martin, and Snyder (2008), who found the dual source of error model to be more flexible, which is not true for Gaussian measurements (Hyndman et al., 2008). However, their analysis was conducted under a stationarity assumption, while, as was noted earlier, demand series are typically non-stationary. The results in Section 6 suggest that non-stationary single source of error models are competitive with other approaches for count time series.

2.2. Evaluation of the Croston method

Willemain, Smart, Shockor, and DeSautels (1994) conducted an extensive simulation study which violated some of the original assumptions (such as cross-correlations between active demand periods and the positive demand quantities) and found that substantial improvements were possible. When they tested the method on real data, the benefits for one-step-ahead forecasts were modest. However, as Johnston and Boylan (1996b) pointed out in a comment, improvements are only to be expected when the average time between active demand periods is appreciably longer than the periodic review time. Willemain, Smart, and Schwartz (2004) described a bootstrap-based approach which allows for a Markov chain development of the probability of a non-zero demand, and indicated that their method produces better inventory decisions than either exponential smoothing or the Croston method. However, Gardner and Koehler (2005) point out that Willemain et al. (2004) did not use the correct lead-time distributions for either of these benchmark methods, nor did they examine extensions of the Croston method. Finally, from the perspective of prediction intervals, Willemain et al. provided an incorrect variance expression.

Sani and Kingsman (1997) also conducted a sizeable simulation study which compared various methods. They used multiple criteria, including overall cost criteria and the service level; they too found that the Croston method performed well, although a simple moving average provided the best overall performance. In an empirical study, Eaves and Kingsman (2004) found little difference between exponential smoothing and the bias-adjusted Croston method when using traditional point measures (mean absolute deviation, root mean squared error and mean absolute percentage error). They went on to argue that a better measure would be to examine the average stock holdings for a given safety stock policy. Their simulation results suggest that the bias-adjusted Croston method works significantly better than the original method in this context.

Syntetos and Boylan (2005) provided a new method, in the spirit of the Croston approach, which they found to be more accurate at issue points, although the results were inconclusive at other time points.

Teunter and Duncan (2009) provided a comparative study of a number of methods. They also concluded that their modified Croston method is to be preferred, based upon a comparison of target and realized service levels.

2.3. Point predictions versus prediction distributions

One interesting aspect of the empirical work done thus far is the strong emphasis on point forecasts. Given that the main purpose behind forecasting intermittent demands is to plan inventory levels, a more compelling analysis should examine service levels, or, more generally, prediction distributions. Indeed, as Chatfield (1992) has pointed out, prediction intervals, which can be derived from prediction distributions if required, deserve a much greater prominence in forecasting applications. Furthermore, the limited empirical evidence available, as cited above, is consistent with the notion that Croston-type methods may provide more accurate prediction distributions, even if they offer little or no advantage for point forecasts.

When we consider processes with low counts, the discrete nature of the distributions can lead to prediction intervals whose actual levels may be higher than the nominal level. Accordingly, we focus upon complete prediction distributions rather than intervals in this paper.

3. Models for intermittent demand and low volume

The literature contains relatively little discussion of this case, although, interestingly, at the end of their paper Johnston and Boylan (1996a) indicate that a simple Poisson process might suffice for slow movers. We move in a different direction in two respects: first, we will retain the idea that the demand is measured at the ends of regular periods of time. Second, we wish to allow for lumpy demand such that the measured demand may exceed one.

3.1. The basic models

Our framework encompasses two possible ways of viewing time. The first assumes that time is continuous and that transactions occur sporadically. At the same time, demands are observed periodically, so that the number of transactions in each period has a Poisson distribution. In cases where the transaction size always equals 1, the quantity demanded in each period also has the same Poisson distribution. To allow for the possibility that transaction sizes are random, we also consider the possibility that they are governed by a logarithmic series distribution. Then, the combined demand Y in a time period has a negative binomial distribution (Quenouille, 1949; Stuart & Ord, 1994, pp. 179–187). It belongs to the family of compound Poisson distributions and is also known as a *randomly stopped sum* distribution, which describes the mechanism just outlined.

The second approach assumes that time is discrete and is divided into equal (or approximately equal) periods of time such as a month. When transactions take place in a particular period, their combined size Y is assumed to follow a shifted Poisson distribution defined by $Y = Z + 1$, where Z is Poisson-distributed. This shift to the right leaves the probability of a zero undefined, so it is assigned a finite probability q. The shifted Poisson probabilities are then weighted by $p = 1-q$ in order to ensure that new probabilities sum to one. We call the result a hurdle shifted Poisson process (HSP). It will be seen that one advantage of the HSP is that it provides a link with the modified Croston method outlined in Section 3.1.2.

Various other possibilities have been proposed over the years, including other compound Poisson forms such as the stuttering Poisson distribution. However, we view the Poisson, negative binomial and shifted hurdle Poisson distributions as reasonable representations of the set of possible distributions. The three distributions are summarized in Table 1 and are used in the empirical comparisons in Section 6.

In Table 1, y designates the values that can be taken by a discrete random variable Y. Its potential probability distributions are all defined over the domain $y = 0, 1, 2, \ldots$.

Table 1
Count distributions used in the empirical study.

Distribution	Mass function	Parameter restrictions	Mean (μ)
Poisson	$\frac{\lambda^y}{y!}\exp(-\lambda)$	$\lambda > 0$	λ
Negative binomial	$\frac{\Gamma(a+y)}{\Gamma(a)y!}\left(\frac{b}{1+b}\right)^a\left(\frac{1}{1+b}\right)^y$	$a > 0, b > 0$	$\frac{a}{b}$
Hurdle shifted Poisson	$\begin{cases} q & y = 0 \\ p\lambda^{y-1}\exp(-\lambda)/(y-1)! & y = 1, 2, \ldots \end{cases}$	$p \geq 0,\ q > 0, \lambda > 0, p + q = 1$	$p(\lambda + 1)$

Notes: When the iterative estimation procedure generated an estimated value of b that exceeded 99, the negative binomial was replaced by the Poisson distribution. This condition was typically the result of under-dispersion, which would lead to having the Poisson as a limiting case. The negative binomial form was chosen rather than the more orthodox version with $p = b/(1 + b)$, for the sake of consistency with the Harvey-Fernandes version.

Table 2
Recurrence relationships for the mean.

Relationship	Recurrence relationship	Restrictions
Static	$\mu_t = \mu_{t-1}$	
Damped dynamic	$\mu_t = (1 - \phi - \alpha)\mu + \phi\mu_{t-1} + \alpha y_{t-1}$	$\mu > 0, \phi > 0, \alpha > 0$ $\phi + \alpha < 1$
Undamped dynamic	$\mu_t = \delta\mu_{t-1} + \alpha y_{t-1}$	$\delta > 0, \alpha > 0$ $\delta + \alpha = 1$

Table 3
Conversion of latent factors to distribution parameters.

Distribution	Formula
Poisson	$\lambda_t = \mu_t$
Negative binomial	$a_t = b\mu_t$
Hurdle shifted Poisson	$\lambda_t = \frac{\mu_t}{p_t} - 1$

For each distribution, we allow for the possibility that the mean of a demand distribution may change randomly over time to reflect the effects of possible structural change. Although the random variable is discrete, we assume that the mean is continuous. Three possibilities are summarized in Table 2, corresponding to a static or constant mean model, a damped dynamic mean (which may be thought of as a stationary autoregressive model for the mean), and an undamped dynamic model (which corresponds to an integrated moving average model for the mean), respectively. The dynamic cases involve a local or short-run mean μ_t. The damped case, being stationary, also includes a long-run mean μ. Additional lags are conceivable but are likely to be of dubious value relative to the gains achieved by allowing the mean to evolve over time. The undamped dynamic relationship has no long-run mean because the process is non-stationary; the associated updating relationship corresponds to that for simple exponential smoothing.

The distribution parameters λ and a are determined from the mean using the conversion formulae in Table 3.

There are two versions of the dynamic negative binomial model: the so-called unrestricted case, which uses the full parameterization as presented in these tables, and the restricted case, where the restriction $\alpha = 1/(1 + b)$ is applied to reduce the number of parameters. This restriction is included so that we can explore its links with the Harvey-Fernandes method (outlined in Section 3.1.1).

There is one disadvantage of using undamped models. The simulation of prediction distributions from such models is hampered by a general problem that applies to all non-stationary count models which are defined on non-negative integers: the simulated series values stochastically converge to zero (Grunwald, Hazma, & Hyndman, 1997), where they get trapped over moderate to long prediction horizons, a behavior which is examined further by Akram, Hyndman, and Ord (2009). This problem does not occur with the damped stationary models.

It will be noted that in the case of the hurdle shifted Poisson distribution, we allow the probability of a non-zero demand in a period to change over time. Designated by p_t, it may be viewed as a local or short-run probability. The damped case is governed by the recurrence relationship

$$p_t = (1 - \phi - \alpha)p + \phi p_{t-1} + \alpha x_{t-1}, \tag{1}$$

where $x_t = 0$ if there is no demand and $x_t = 1$ if there is demand in period t. It involves a long-run probability designated by p. The parameters ϕ and α are the same as those used in the corresponding recurrence relationship for the mean demand, consistent with the parameterization originally used by Croston. The effect is to contain the growth in the number of parameters and to force p_t to be less than μ_t for all t, so that $\lambda_t = \dfrac{\mu_t}{p_t} - 1$ is never negative. The long-run parameters must satisfy the condition $p < \mu$ for a similar reason. Both the seed probability p_1 and the long-run probability p are restricted to the unit interval.

The undamped analogue of Eq. (1) with $\alpha + \delta = 1$ is

$$p_t = \delta p_{t-1} + \alpha x_{t-1}. \tag{2}$$

This corresponds to a simple exponential smoothing recurrence relationship for the probability. Again, α and δ are the same as the parameters used in the corresponding recurrence relationship for the mean.

Several other models have been considered in the literature on count time series, as noted earlier, and those which are included in the simulation study are now summarized briefly.

3.1.1. The Harvey-Fernandes method

Harvey and Fernandes (1989) describe a method based on a local level state space model with Poisson measurements. Their method does not allow for intermittent demands but is a Poisson analogue of the Kalman filter based upon a negative binomial distribution, defined as a mixture of Poisson distributions with a gamma mixture distribution. The negative binomial distribution has a time dependent mean given by the finite exponentially weighted average

$$\mu_t = \frac{\delta y_{t-1} + \delta^2 y_{t-2} + \delta^3 y_{t-3} \cdots + \delta^{t-1} y_1}{\delta + \delta^2 + \cdots + \delta^{t-1}}, \quad (3)$$

where δ is a parameter called the *discount factor* which satisfies the condition $0 \leq \delta \leq 1$. The numerator and denominator of this expression are designated by a_t and b_t respectively. Used as the parameters a and b of the negative binomial distribution (see Table 1) in typical period t, they are calculated recursively using the expressions

$$a_{t+1} = \delta(a_t + y_t) \quad \text{and} \quad b_{t+1} = \delta(b_t + 1),$$

where $a_1 = b_1 = 0$.

As t increases in Eq. (3), b_t converges to a constant value $b = \delta/(1 - \delta)$, and the mean $\mu_t = b_t/a_t$ is then governed by the simple exponential smoothing update relationship $\mu_{t+1} = \delta\mu_t + \alpha y_t$. The negative binomial probability parameter is $p = b/(1+b) = \delta$, and consequently $q = 1 - p = \alpha$. The smoothing and discount parameters used for the update of the mean also become an integral part of the negative binomial distribution formula. This corresponds to what we earlier called the restricted undamped negative binomial model. Given that this is effectively the asymptotic form of the Harvey-Fernandes method, the two approaches should give similar predictions.

3.1.2. The modified Croston model

The original Croston method examined a series of the time gaps between periods with demand occurrences and a series of the non-zero demand quantities. Simple exponential smoothing is then applied to the two derived series to obtain estimates of their means, and a point prediction is established from the two results. The same parameters δ and α are used in both of the simple exponential smoothing recurrence relationships.

Building on earlier ideas from the work of Snyder (2002) and Shenstone and Hyndman (2005), the Croston method was modified by Hyndman et al. (2008, pp. 281–283) to incorporate probabilistic assumptions. It was envisaged that:

1. The time gaps between active demand periods will be governed by a shifted geometric distribution

   $$\Pr\{\tau = t\} = q^{t-1}p \quad t = 1, 2, \ldots, \quad (4)$$

   where p is the probability of a positive demand in a given period and $q = 1 - p$.
2. The positive demands Y^+ will be governed by a shifted Poisson distribution

   $$\Pr\{Y^+ = y|\mu^+\} = \frac{(\lambda)^{y-1}}{(y-1)!}\exp(-\lambda), \quad y = 1, 2, \ldots, \quad (5)$$

   where $\mu^+ = \lambda + 1$ is the mean of the positive demands.

Moreover, it was envisaged that the parameters of these distributions would change over time, with their values being derived from the mean time gaps and mean non-negative demands obtained from the application of simple exponential smoothing.

The main advantage of the stochastic assumptions is that it expands Croston's method to enable it to produce entire prediction distributions by simulation, rather than point predictions alone. It also enables the derivation and use of maximum likelihood estimates of the parameters δ and α.

3.2. Estimation

All of the unknown model parameters were estimated using the maximum likelihood method. The likelihood function is based on the joint distribution $p(y_1, \ldots, y_n|\mu_1, \theta)$, where θ represents all unknown parameters other than the first mean μ_1. Using induction, in conjunction with the conditional probability law $\Pr\{A, B\} = \Pr\{B|A\}\Pr\{A\}$, it can be established, for every model considered, that

$$p(y_1, \ldots, y_n|I_{n-1}) = \prod_{t=1}^{n} p(y_t|I_{t-1}), \quad (6)$$

where $I_{t-1} = \{\mu_t, \theta\},\ t = 1, 2, \ldots, n$. The univariate distributions in this decomposition are a succession of one-step-ahead prediction distributions. In the case of the static models, the maximum likelihood estimate of the common mean is just a simple average. In the other cases, the appropriate dynamic relationship is applied to obtain the means of the univariate distributions.

The same basic approach is used for the Harvey-Fernandes (HF) model. In this case, the initial mean μ_1 is not needed and the likelihood is calculated from the period with the first demand. Successive means are calculated using Eq. (3), and the term corresponding to $t = 1$ in Eq. (6) is dropped. In the case of the modified Croston method, estimates of the parameters were also obtained using the prediction decomposition of the likelihood. The details are provided by Hyndman et al. (2008, pp. 281–283).

3.3. Prediction distributions

A simulation approach is used to obtain all of the prediction distributions, although analytical methods could be developed for static models — see Ord, Snyder, and Beaumont (2010). Given that the models involve first-order recurrence relationships, the joint prediction distribution can be decomposed into a product of univariate one-step-ahead prediction distributions, as follows:

$$p(y_{n+1}, \ldots, y_{n+h}|I_n) = \prod_{t=n+1}^{n+h} p(y_t|I_{t-1}). \quad (7)$$

In the simulation approach, successive future series values are then generated from each future one-step-ahead distribution. This process is repeated 100,000 times, to give a sample from the joint distribution. Marginal and lead time relative frequency distributions are then used as approximations for the prediction distributions.

4. Prediction performance measures

Many different measures can be used to evaluate predictive performance, but we focused primarily on three: the mean absolute scaled error, the prediction likelihood score and the discrete ranked probability score. Each of these will be defined in the following sub-sections. For the moment, we consider these measures from a general perspective.

Let $\mathcal{M}$ denote a measure of the prediction performance that can be calculated for all models under consideration. If this measure is defined so that an increase means an improved prediction, we write $\mathcal{S} = 1$; otherwise, when a decrease signifies an improvement, we let $\mathcal{S} = -1$.

It is convenient to benchmark all models against the static Poisson model. We use $\mathcal{M}_p$ and $\mathcal{M}_i$ to represent the measure for a static Poisson distribution and for another model i, respectively. Moreover, it makes sense to use a scale-free summary statistic to facilitate comparisons. We therefore recommend the use of statistics of the form

$$\mathcal{R}_{ip} = 100\mathcal{S}(\log \mathcal{M}_i - \log \mathcal{M}_p). \tag{8}$$

$\mathcal{R}_{ip}$ may be interpreted as the percentage change in the measure for model i relative to the static Poisson model. $\mathcal{R}_{ip} > 0$ indicates that model i is a better predictor of $y_{n+1}, \dots, y_{n+h}$ than the static Poisson model.

Once the statistic in Eq. (8) is in place, it is straightforward to compare any two models. $\mathcal{R}_{12} = \mathcal{R}_{1p} - \mathcal{R}_{2p}$ measures the percentage difference between any two models 1 and 2. Hence, $\mathcal{R}_{12} > 0$, or equivalently $\mathcal{R}_{1p} > \mathcal{R}_{2p}$, indicates that model 1 is a better predictor than model 2.

4.1. Mean absolute scaled error

One performance measure in common use is the *mean absolute percentage error* (MAPE), defined as

$$\text{MAPE} = \frac{100}{h} \sum_{j=1}^{h} \left| \frac{\hat{y}_n(j) - y_{n+j}}{y_{n+j}} \right|, \tag{9}$$

where $\hat{y}_n(j)$ designates the prediction of the series value y_{n+j} made at origin n, and h is the prediction horizon. However, it fails for low count data whenever the value of zero is encountered in the series. We therefore used the *mean absolute scaled error* (Hyndman & Koehler, 2006) instead:

$$\text{MASE} = \frac{1}{h} \sum_{j=1}^{h} |y_{n+j} - \hat{y}_n(j)| \Big/ \frac{1}{n-1} \sum_{t=2}^{n} |y_t - y_{t-1}|. \tag{10}$$

In Section 6, the empirical study is based upon $h = 6$; that is, lead times of 1–6 are employed in the calculations. Other measures such as the geometric mean absolute error (GMAE) or geometric root mean square error (GRMSE) could be employed, but would be expected to give similar results. In all cases, extreme values may result if the number of active demand periods in the estimation sample is very small (or even zero), which is why we required a minimum number of active demand periods in the empirical work reported in Section 6.

4.2. Distribution based scores

Although the MASE and similar measures are useful in determining the performances of point forecasting methods, they do not provide any information on other characteristics of the predictive distributions. In the following sections we describe two criteria that can be used to measure the forecasting performance relative to the predictive distribution. Typically, such measures could be used with fairly small numbers of hold-out observations for a single series, but would be averaged over a number of series to determine the overall performance for a group of series, as in Section 6.

4.2.1. Prediction likelihood score (PLS)

The joint prediction distribution $p(y_{n+1}, \dots, y_{n+h}|I_n)$ summarizes *all* of the characteristics of a future series, including the central tendency, variability, autocorrelation, skewness and kurtosis. Since we are interested in prediction distributions rather than point forecasts, this joint distribution is a natural criterion to use. Here, I_n consists of all quantities that inform the calculation of these probabilities, including the estimation sample $y_1, \dots, y_n$, the parameters, and the states of the process at the end of period n. Assuming that we withhold the series values $y_{n+1}, \dots, y_{n+h}$ for evaluation purposes, $p(y_{n+1}, \dots, y_{n+h}|I_n)$ is the *likelihood* that these values come from the model under consideration. We call this the *prediction likelihood score* (PLS), although its logarithm is more commonly known as the *logarithmic score* (Gneiting & Raftery, 2007). Note that Czado, Gneiting, and Held (2009) use this measure in a study of cross-sectional Poisson and negative binomial regression models, but with constant coefficients.

The PLS could be evaluated in several ways. We consider the joint distribution of $\{y_{n+1}, \dots, y_{n+h}\}$ given the information up to and including period n, namely I_n. Applying the same logic as in the derivation of Eq. (7), it can be established that

$$p(y_{n+1}, \dots, y_{n+h}|I_n) = \prod_{t=n+1}^{n+h} p(y_t|I^*_{t-1}), \tag{11}$$

where $I^*_t = \{I_n, y_{n+1}, \dots, y_t\}$; the change in notation serves to indicate that the parameters are estimated using only the first n observations, whereas the means are updated each time. Each univariate distribution describes the uncertainty in the typical 'future' period t, as seen from the *beginning* of this period, using the 'past' information contained in I^*_{t-1}. Thus, the joint prediction mass function can be found from the product of h one-step-ahead univariate prediction distributions. This is the forecasting analogue of the prediction error decomposition of the likelihood function used in estimation; see Hyndman et al. (2008). The means of these one-step prediction distributions are calculated using the various forms of exponential smoothing implied by the damped and undamped transition relationships given in Table 2. Since interest may focus upon forecasts for the demand over the lead-time as well as one-step-ahead, we also examine the PLS for the sum over the lead-time $S_n(h) = y_{n+1} + \cdots + y_{n+h}$

with prediction distribution $p(S_n(h)|I_n)$. In Section 6, the PLS is averaged over all of the series we consider in order to provide an overall assessment of each model, rather than a selection criterion for individual series.

The following simple example illustrates why it is important to use a measure such as the PLS rather than one which focuses exclusively on point forecasts.

Example 1: *Model choice using PLS*

Consider two competing static Gaussian forecasting models, indexed as 1 and 2, with a common mean μ and variances, from the estimation sample, estimated as $V_1 < V_2$. (For example, the two models might correspond to estimation with and without the removal of outliers.) From Eq. (11), the comparative form of the PLS reduces to

$$\mathcal{R}_{12} = 100h\left(\log V_2 - \log V_1 + \frac{V}{V_2} - \frac{V}{V_1}\right), \qquad (12)$$

where V denotes the one-step-ahead forecast mean squared error evaluated over the hold-out sample for periods $n+1, \ldots, n+h$.

If an in-sample criterion, based upon forecast variances, were used to select a model, it is clear that model 1 would be selected. However, if we are interested in the validity of the prediction distribution (or, more specifically, in making a safety stock decision), we need to ensure that the model selected represents the uncertainty in the forecasts properly. In this simple case, the two procedures would give rise to the same value of V. It can be shown that expression (12) is positive at $V = V_1$ and negative at $V = V_2$, so that the choice of model depends upon the prediction distribution. Thus, expression (12) indicates that we should only choose model 1 if V is sufficiently close to V_1; that is, if $\mathcal{R}_{12} > 0$. In other words, the method with the variance closest to that of the mean squared prediction error is chosen.

4.2.2. Discrete ranked probability score (DRPS)

Another possible measure is the ranked probability score (Epstein, 1969; Murphy, 1971). It uses the L_2-norm to measure the distance between two distributions:

$$L_2(x, F) = \sum_{y=0}^{\infty}(\hat{F}(y) - F(y))^2,$$

where $\hat{F}(y)$ is the sample-based approximation of the distribution function F. When F is discrete and there is only a single observation x,

$$\hat{F}(y) = \begin{cases} 0 & \text{if } y \le x \\ 1 & \text{if } y > x. \end{cases}$$

In our calculations, the infinite sum was truncated at $y = 100$; the numerical error caused by this truncation is negligible.

In Section 6, we calculate the DRPS for each one-step-ahead forecast relating to the sample withheld, and then average over the $h(=6)$ values; the procedure follows the same logic as for the PLS. The DRPS for the total lead-time demand is also considered.

5. Model selection

There are two principal approaches to model selection. The first uses an information criterion such as the AIC or BIC (see, for example, Hyndman et al., 2008, pp. 105–108), and relies upon the fit of the data to the estimation sample, with suitable penalties for extra parameters. The second method, known as prediction validation, uses an estimation sample to specify the parameter values, and then selects a procedure based upon the out-of-sample forecasting performances of the competing models. Despite the popularity of prediction validation (e.g. Makridakis & Hibon, 2000), Billah, King, Snyder, and Koehler (2006) found it to be generally inferior to other methods for point forecasting, particularly to those based upon information criteria. This conclusion is not affected by the particular out-of-sample point forecasting criterion selected, such as those described in the previous section. We ran several small simulation experiments for the distributions currently under consideration, which confirmed the conclusions of Billah et al. (2006). This conclusion is especially true in the present case, when the hold-out sample for a single series is based upon only six observations.

Although the prediction validation methods are not recommended for model selection for individual series, they are useful for assessing the overall performance across multiple series when we use criteria such as PLS and DRPS to evaluate the prediction distributions. Such comparisons are common in forecasting competitions, and are useful when a decision as to a general approach to forecasting a group of series such as a set of SKUs must be made. Accordingly, the comparisons in the next section are made using the criteria discussed in Section 4 for comparing the overall model performances.

6. An empirical study of auto parts demand

The study used data on slow-moving parts for a US automobile company; these data were also discussed by Hyndman et al. (2008, pp. 283–286). The data set consists of 2674 monthly series, of which 2509 had complete records. The data cover a period of 51 months; 45 observations were used for estimation and 6 were withheld for comparing the forecasting performances one to six steps ahead. Considering only those series with at least two active periods, the average time lapse or gap between positive demands is 4.8 months. The average positive demand is 2.1, with an average variance-to-mean ratio of 2.3, meaning that most of the series are over-dispersed. As was noted earlier, only 1046 of these series had (a) ten or more months with positive demands, *and* (b) at least some positive demands in the first 15 and the last 15 months; our forecasting study was restricted to these 1046 series, for which the average time lapse was 2.5, with an average variance-to-mean ratio of 1.9. The purpose of these restrictions was to ensure that each part had at least some inventory activity during the estimation periods.

We examined the performance of the models defined in Tables 1–3 for one-step, multiple-step and lead time demand predictions. For the purpose of comparison, we

Table 4
Comparison of the forecasting performances of different methods for 1046 US automobile parts series (best cases in bold).

Predictions	One-step			Multiple-step		Lead time demand		
Criterion	PLS	DRPS	MASE[b]	DRPS	MASE	PLS[a]	DRPS	MASE
Static models								
Poisson	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Hurdle shifted Poisson	12.0	9.5	0.0	9.5	0.0	6.6	1.7	0.0
Negative binomial	14.5	13.7	0.0	13.7	0.0	9.2	11.1	0.0
Damped dynamic models								
Poisson	10.9	16.7	15.4	15.0	13.0	10.8	33.0	24.4
Hurdle shifted Poisson	16.9	21.8	12.8	20.1	10.9	13.4	34.1	22.2
Negative binomial – unrestricted	**20.5**	25.7	15.9	24.3	13.5	**15.3**	41.6	25.3
Negative binomial – restricted	18.3	21.4	11.1	10.7	−9.2	14.4	22.0	−13.9
Undamped dynamic models								
Poisson	10.2	18.4	19.4	18.2	18.8	10.9	38.2	32.4
Hurdle shifted Poisson	17.2	22.7	15.8	22.2	15.1	14.1	37.1	26.8
Negative binomial – unrestricted	20.1	**26.9**	18.9	**26.3**	18.2	15.2	**44.2**	31.7
Negative binomial – restricted	15.1	23.2	23.3	22.0	21.6	13.4	41.8	34.5
Others								
Croston	15.8	18.5	8.9	18.5	8.8	12.2	28.2	17.4
Harvey and Fernandes	16.0	24.9	25.7	22.9	23.0	14.3	43.9	**35.7**
Zeros	− inf	10.0	**68.4**	10.0	**68.4**	− inf	−2.8	26.8
Info criteria based on the best choice of 9 models								
AIC	18.6	23.9	15.0	23.3	14.3	13.5	37.0	25.6
BIC	16.4	20.7	11.7	20.8	11.7	11.7	30.0	20.2

[a] The PLS values are averaged using a 2% trimmed mean in order to avoid a small number of extreme values.
[b] The MASEs for the static models are the same, as the best predictor for each of them is the mean.

also included the Harvey-Fernandes and modified Croston models described in Sections 3.1.1 and 3.1.2, as well as the simplistic all-zeros forecast which ignores the data and simply forecasts a zero demand for all periods. The required prediction distributions were obtained using the simulation approach outlined in Section 3.3.

The measures PLS, DRPS and MASE were calculated using withheld data from periods 46 to 51. These measures were computed for the 1046 series and benchmarked against the static Poisson case using Eq. (8), but the computational process was different for each. In the case of the PLS, Eq. (8) was applied to each series and then averaged. Given the form of the log-likelihood ratio, this is equivalent to averaging the log-likelihood across all series and then taking the ratios for different models. We would have liked to have applied a series-by-series approach to the DRPS and MASE, but the single series measures could take on zero values. Therefore, the DRPS and MASE were averaged across series before Eq. (8) was applied to the resulting averages.

A summary of the results, in terms of *percentage improvements*, is reported in Table 4. In all cases, larger values are better. A summary using medians instead of averages was also produced, but they were found to lead to similar conclusions and so are not reported here. The summary for the PLS of lead-time demand was based on a trimmed mean, in order to avoid distortions from a few extreme cases. The zeros prediction method has $\Pr(Y = 0) = 1$ and $\Pr(Y \neq 0) = 0$, so its PLS results are reported as minus infinity.

A number of interesting observations can be made. First, the results in Table 4 are reasonably consistent across the different types of forecasts. Second, the PLS and DRPS results suggest that the traditional static Poisson distribution can be too restrictive for intermittent inventories. They confirm that better predictions may be obtained from distributions which allow for over-dispersion, with the negative binomial distribution being the best option. This outcome was to be expected, because the negative binomial distribution has been used widely in inventory control for slow moving items, presumably because it has been found to work well in practice. Interestingly, the MASE, and similar measures based on point predictions, fail to reflect this important conclusion. Instead, they imply that the zero prediction is best for both one-step and multi-step predictions, which would result in zero stocks for all items!

The main aim of this study was to assess methods that allow for serial correlation through the use of dynamic specifications. The performance measures all indicate that there is a significant improvement, and the PLS and DRPS are reasonably consistent in their rankings. Each measure indicates that the unrestricted negative binomial distribution is best, although there is no clear indication of whether damped or undamped dynamics should be used. The difference between the two forms of the dynamics is small according to both criteria. Undamped dynamics gets our vote here, because it has strong links with the widely used simple exponential smoothing forecasting method, and because it has been found that most business and economic time series are not stationary. Furthermore, fewer parameters need to be estimated, which is important when the series are short. However, as is indicated in Section 3.1, this choice has one inconvenient aspect: simulated series values eventually converge to a fixed point of zero.

Table 5
Percentage breakdown of AIC model selections.

Distribution	Dynamics			
	Static	Damped	Undamped	Total
Poisson	15	18	1	34
Negative binomial	21	11	4	36
Hurdle Poisson	24	5	0	30
Total	60	34	6	100

Table 6
Percentage breakdown of BIC model selections.

Distribution	Dynamics			
	Static	Damped	Undamped	Total
Poisson	31	10	5	46
Negative binomial	24	2	5	31
Hurdle Poisson	22	0	0	23
Total	77	13	10	100

One interesting finding is that the Harvey-Fernandes method out-performed the widely-used Croston method of forecasting, or at least our adaptation of it, which allows for maximum likelihood estimation of its parameters. It was shown in Section 3.1.1 that the Harvey-Fernandes method has a limiting form corresponding to our restricted version of the negative binomial model with undamped dynamics. Not surprisingly, the two models had similar forecasting performances, with the Harvey-Fernandes method having a slight edge. However, the unrestricted version of the negative binomial model was markedly better than either of these approaches. What clearly emerges from this study is the fact that, based on the PLS and DRPS criteria, our undamped negative binomial model significantly out-performs both the adapted Croston method and the Harvey-Fernandes method.

The assumptions underlying the hurdle shifted Poisson distribution, as outlined in Section 3.1, appear to make it a strong candidate for intermittent demand forecasting. However, the results from Table 4 indicate that its performance lags behind that of the damped negative binomial model, according to all of the measures. In earlier numerical studies using the same data set, a similar outcome (not reported here) was obtained for the zero-inflated Poisson distribution.

In our framework, the undamped hurdle shifted Poisson model is the closest to the modified Croston model. Instead of smoothing the time gaps, we smooth the demand occurrence indicator variable using Eq. (2). Interestingly, the undamped hurdle shifted Poisson model does better than the adapted Croston method.

Multi-model approaches using information criteria for model selection were also explored in the study. Both the methods designated as *Others* and the restricted versions of our negative binomial models in Table 4 were excluded from the set of possible models for this part of the project. Both the Akaike (AIC) and Bayesian (BIC) information criteria were considered. As might be expected, the AIC approach yielded the better forecasts on average. Nevertheless, the unrestricted damped negative binomial case employed as an encompassing model out-performed the multi-model approaches.

The percentage breakdowns of the models selected by the AIC and the BIC are shown in Tables 5 and 6. Static models proved to be quite adequate for about 60% of the series according to the AIC and 77% according to the more stringent BIC. We need to keep in mind the fact that the asymptotic justification for the AIC is based upon the forecasting performance, whereas the BIC is a consistent criterion for selecting the true model. Given the forecasting focus of this study, the AIC seems more appropriate. Moreover, no particular distribution dominated. It is not clear how much emphasis should be placed on these particular outcomes when the multi-model approach did not work as well as the encompassing approach. Nevertheless, at first sight, the prominence of static and Poisson selections may be surprising. For those series, it is likely that all of the models (except 'zeros') would perform well, and the AIC is simply guiding the modeler to the simpler schemes. In the remaining cases, the static and Poisson versions are inadequate and more complex models are needed. With a different mix of series, the AIC may well outperform the encompassing approach, but it needs to be kept in mind that non-stationary models are better able to adapt to structural changes in a series, and thus may be preferable as an 'insurance policy'.

There may be concern that our initial culling of the series has had a distorting effect on the results. If all of the series had been included, the results for the MASE and the DRPS would have shifted in the direction of the zeros method, with the latter still having the value minus infinity for the PLS. In practice, items with such low levels of demand would often be covered, not by carrying stock, but simply by placing special replenishment orders as needed.

7. Use of simulated demands in inventory control

Once forecasts of the demand have been obtained, they can be fed into the inventory control decision process. In the Gaussian case (Harrison, 1967; Snyder, Koehler, & Ord, 1999), this can be done analytically, but in non-Gaussian cases it is often necessary to resort to simulation. We will provide an example in this section, but a more comprehensive exposition is given by Hyndman et al. (2008, Chapter 18).

Our focus is on a part, the demand for which is governed by a Poisson distribution, the mean of which changes according to the undamped dynamic equation underlying simple exponential smoothing, as defined in Tables 1 and 2. We assume that the maximum likelihood estimate of the smoothing parameter α has been found using 55 months of data and is 0.1. It is now the beginning of month 56; a simple exponential smoothing forecasting routine has yielded a point prediction for this month of 0.75, and a replenishment order must be placed with a supplier. There is a delivery lead time of 2 months, meaning that a replenishment order placed now is not delivered until the beginning of month 58. The size of the replenishment order is found by comparing the current *total* supply (stock *minus* backlog *plus* outstanding replenishment orders) with a pre-determined order-up-to level (OUL). The size of the OUL determines the size of the replenishment order,

Table 7
Simulation of future demands from the undamped dynamic Poisson model.

Period	Point prediction (mean)	Simulated demand
56	0.75	0
57	0.675	2
58	0.8075	1

which in turn determines the level of service provided to customers in month 58. The problem is to find a value for the OUL which ensures that the expected fill rate in month 58 is at least 90%. Demands occurring during shortages are backlogged.

The analysis begins by assigning a trial value to the OUL and determining the consequent expected fill rate in month 58. We then adjust the OUL until the expected fill rate equals the target value of 90%. Initially, the analysis ignores the integer property of the stock and treats the OUL as a continuous quantity. This search procedure is typically automated by the use of a solver such as Goal-Seeker in Microsoft Excel.

Since the predicted demand is 0.75 for each future month, as seen at the beginning of month 56, and the *extended* lead time of interest is 2+1 = 3 months, we set the initial trial value of the OUL equal to the mean extended lead time demand (three times the monthly demand, or 2.25). There is no safety stock for this initial situation.

We can then simulate the future demands for months 56, 57 and 58, as shown in Table 7. The first value of 0 was simulated from a Poisson distribution with a mean of 0.75. The mean was then revised in the light of this new simulated demand using simple exponential smoothing (with $\alpha = 0.1$), to give a new mean of 0.675, with this change being a reflection of presumed permanent changes in the market for the inventory. The second value of 2 was simulated from a Poisson distribution with a mean of 0.675. This procedure was then run again (new mean $= 0.8075$), resulting in a simulated demand of 1 for month 58.

Given these three simulated future demands, it is then possible to find the corresponding sales in month 58, as shown in Table 8. The OUL represents the total stock which is available to meet demand over the extended lead time of 3 months. After the OUL of 2.25 is used to meet demands of 0 and 2 in months 56 and 57, respectively, month 58 begins with a stock of 0.25. This is not sufficient to meet the simulated demand of 1 in month 58. Simulated sales immediately from stock can only be 0.25. The unsatisfied demand must be deferred until the next delivery.

This experiment can be repeated to create, say, 100 hypothetical future ensembles of the demand and sales in month 58. The ratio of the ensemble average of sales to the ensemble average demand in month 58 can be viewed as an estimate of the fill-rate. This is unlikely to exactly equal the target fill-rate with the initial trial value for the OUL, and therefore, keeping the demand scenarios unchanged, the OUL is adjusted by a solver until the estimated fill rate equals the target value of 90%. The deviation of OUL from its initial value of the mean extended lead time demand represents the safety stock. For practical reasons, any resulting non-integer values of the OUL are rounded to the next highest integer, in which case the target fill rate is usually exceeded.

Regardless of which dynamic demand model is used, a simulation approach is needed because the analytical form of future distributions of demands, as seen from the forecast origin, are unknown, with the exception of the distribution for the first future period. Because of the associated computational loads, an approach like this would have been completely impractical in the era when most of the basic ideas underpinning inventory control theory were originally developed, ideas which still form the mainstay of most texts on the subject. Analytical approaches were once necessary for tractable computations, but now the raw computational power provided by modern computers enables us to find ordering parameters such as the OUL using approaches like the one described here in a fraction of a second. Approaches like this are therefore now feasible even for large numbers of parts.

8. Conclusions

In this paper we have introduced some new models for forecasting intermittent demand time series based on a variety of count probability distributions, coupled with a variety of dynamic specifications to account for potential serial correlation. These models were then compared with established forecasting procedures using a database of car parts demands. Particular emphasis was placed on prediction distributions from these models rather than point forecasts, because the latter ignore features such as variability and skewness, which can be important for safety stock determination.

The empirical results suggest that although many series may be able to be modeled adequately using traditional static schemes, substantial gains may be achievable by using dynamic versions for many of the others. A similar argument favors the use of richer models than the Poisson. Thus, an effective forecasting framework for SKUs that have low volume intermittent demands must look beyond the traditional static Poisson format.

Table 8
Sales calculation from simulated demands.

Period	Variable	Value	
56	Order-up-to level	2.25	Trial value specified by user (OUL)
56	Demand	0	Simulated demand (D56)
57	Demand	2	Simulated demand (D57)
58	Open stock	0.25	max (OUL–D56–D57, 0)
58	Demand	1	Simulated demand (D58)
58	Sales	0.25	min (Open stock, D58)

Our study has indicated that simple exponential smoothing can work well in conjunction with an unrestricted negative binomial distribution. It has also indicated that there is little advantage in using a multi-model approach with information criteria for model selection. The usual caveat that such results are potentially data dependent must be made. Nevertheless, we expect that similar results would emerge for other datasets. This being the case, it appears that the Croston method, even in an adapted form, should possibly be replaced by exponential smoothing, coupled with a negative binomial distribution.

There are a number of series, most of which were excluded from the sample of 1046 SKUs used here, for which the demand is very low, perhaps in the order of one or two units per year. In such cases, a static model might be preferable, although from the stock control perspective the decision will often lie between holding one unit of stock and holding zero stock and submitting replenishment orders as needed.

There still remains the important issue of what to do with new parts which have no or only limited demand data. In such cases, maximum likelihood methods applied to single series are going to be ineffective. This question is the subject of an on-going investigation into the forecasting of demand for slow moving inventories, which explores the possibility of extending the maximum likelihood principle to multiple time series in a quest to overcome the paucity of data.

Acknowledgement

This research was supported by Australian Research Council Discovery Grant DP0877424, and by the Sebes Fellowship funded by Mr Alvin J Delaire.

References

Akram, M., Hyndman, R. J., & Ord, J. K. (2009). Exponential smoothing and non-negative data. *Australian and New Zealand Journal of Statistics*, *51*, 415–432.

Al-Osh, M. A., & Alzaid, A. A. (1987). First-order integer valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis*, *8*, 261–275.

Billah, B., King, M. L., Snyder, R. D., & Koehler, A. B. (2006). Exponential smoothing model selection for forecasting. *International Journal of Forecasting*, *22*, 239–247.

Chatfield, C. (1992). Calculating interval forecasts. *Journal of Business and Economic Statistics*, *11*, 121–144.

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

Czado, C., Gneiting, T., & Held, L. (2009). Predictive model assessment for count data. *Biometrics*, *65*, 1254–1261.

Davis, R. A., Dunsmuir, W. T., & Wang, Y. (1999). Modelling time series of count data. In S. Ghosh (Ed.), *Asymptotics, nonparametrics and time series* (pp. 63–114). New York: Marcel Dekker.

Davis, R. A., Dunsmuir, W. T. M., & Wang, Y. (2000). On autocorrelation in a Poisson regression model. *Biometrika*, *87*, 491–505.

Durbin, J., & Koopman, S. J. (2001). *Time series analysis by state space methods*. Oxford: Oxford University Press.

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.

Epstein, E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, *8*, 985–987.

Feigin, P. D., Gould, P., Martin, G. M., & Snyder, R. D. (2008). Feasible parameter regions for alternative discrete state space models. *Statistics and Probability Letters*, *78*, 2963–2970.

Gardner, E. S., & Koehler, A. B. (2005). Comment on a patented bootstrapping method for forecasting intermittent demand. *International Journal of Forecasting*, *21*, 617–618.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction and estimation. *Journal of the American Statistical Association*, *102*, 359–378.

Grunwald, G. K., Hazma, K., & Hyndman, R. J. (1997). Some properties and generalizations of Bayesian time series models. *Journal of the Royal Statistical Society, Series B*, *59*, 615–626.

Harrison, P. J. (1967). Exponential smoothing and short-term sales forecasting. *Management Science*, *13*, 821–842.

Harvey, A. C., & Fernandes, C. (1989). Time series models for count or qualitative observations. *Journal of Business and Economic Statistics*, *7*, 407–422.

Heinen, A. (2003). *Modelling time series count data: an autoregressive conditional Poisson model*. Working paper. University of California. San Diego.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*, 679–688.

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Berlin and Heidelberg: Springer-Verlag.

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

Johnston, F. R., & Boylan, J. E. (1996b). Forecasting intermittent demand: a comparative evaluation of Croston's method. Comment. *International Journal of Forecasting*, *12*, 297–298.

Jung, R. C., Kukuk, M., & Liesenfeld, R. (2006). Time series of count data: modeling, estimation and diagnostics. *Computational Statistics and Data Analysis*, *51*, 2350–2364.

Makridakis, S., & Hibon, M. (2000). The M3-competition: results, conclusions and implications. *International Journal of Forecasting*, *16*, 451–476.

McCabe, B. P. M., & Martin, G. M. (2005). Bayesian predictions of low count time series. *International Journal of Forecasting*, *21*, 315–330.

McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability*, *20*, 822–835.

Murphy, A. H. (1971). A note on the ranked probability score. *Journal of Applied Meteorology*, *10*, 155–156.

Ord, J. K., Koehler, A. B., & Snyder, R. D. (1997). Estimation and prediction of a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association*, *92*, 1621–1629.

Ord, J. K., Snyder, R. D., & Beaumont, A. (2010). *Forecasting for intermittent demand for slow-moving items*. Department working paper 12/10.

Quenouille, M. H. (1949). A relation between the logarithmic, Poisson, and negative binomial series. *Biometrics*, *5*, 162–164.

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.

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

Shephard, N. (1995). *Generalized linear autoregressions*. Working paper. Nuffield College. Oxford.

Snyder, R. D. (2002). Forecasting sales of slow and fast moving inventories. *European Journal of Operational Research*, *140*, 684–699.

Snyder, R. D., Koehler, A. B., & Ord, J. K. (1999). Lead time demand for simple exponential smoothing: an adjustment factor for the standard deviation. *Journal of the Operational Research Society*, *50*, 1079–1082.

Stuart, A., & Ord, J. K. (1994). *Kendall's advanced theory of statistics* (6th ed.) London: Edward Arnold.

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., Nikolopoulos, K., & Boylan, J. E. (2010). Judging the judges through accuracy implication metrics: The case of inventory forecasting. *International Journal of Forecasting*, *26*, 134–143.

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

West, M., & Harrison, J. (1997). *Bayesian forecasting and dynamic models* (2nd ed.) New York: Springer-Verlag.

West, M., Harrison, J., & Migon, H. S. (1985). Dynamic generalized models and Bayesian forecasting (with discussion). *Journal of the American Statistical Association*, *80*, 73–97.

Willemain, T. R., Smart, C. N., & Schwartz, H. F. (2004). A new approach to intermittent forecasting demand for service parts inventories. *International Journal of Forecasting*, *20*, 375–387.

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

Zeger, S. L. (1988). A regression model for time series counts. *Biometrika*, *75*, 621–629.

Ralph D. Snyder is an associate professor in the Department of Econometrics and Business Statistics at Monash University, Australia. Having published extensively on business forecasting and inventory management, he has played a leading role in the establishment of the class of innovations state space models for exponential smoothing.

J. Keith Ord is a professor in the McDonough School of Business, Georgetown University, Washington, DC. He has authored over 100 research papers on statistics and its applications, and ten books, including Kendall's *Advanced Theory of Statistics*. He is a fellow of the American Statistical Association and the International Institute of Forecasters.

Adrian Beaumont, a statistician, is a research assistant at Monash University.