ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Intermittent demand forecasting for inventory control: A multi-series approach

Ralph Snyder, Adrian Beaumont and J. Keith Ord

July 2012

Working Paper 15/12

Intermittent demand forecasting for inventory control: A multi-series approach

Ralph Snyder, Adrian Beaumont[1] and J. Keith Ord[2]

Abstract

This paper is concerned with identifying an effective method for forecasting the lead time demand of slow-moving inventories. Particular emphasis is placed on prediction distributions instead of point predictions alone. It is also placed on methods which work with small samples as well as large samples in recognition of the fact that the typical range of items has a mix of vintages due to different commissioning and decommissioning dates over time. Various forecasting methods are compared using monthly demand data for more than one thousand car parts. It is found that a multi-series version of exponential smoothing coupled with a Pólya (negative binomial) distribution works better than the other twenty-four methods considered, including the Croston method.

Author Keywords: demand forecasting; inventory control; shifted Poisson distribution

JEL classification: C22

[1] Department of Statistics, The University of Melbourne, Parkville, Victoria, Australia.
a.beaumont@student.unimelb.edu.au
[2] McDonough School of Business, Georgetown University, Washington, DC 20057, USA
ordk@georgetown.edu

1 Introduction

Plausible state space models for count time series are developed and benchmarked against standard methods of forecasting (Croston 1972; Harvey and Fernandes 1989). These models permit a rational approach to estimation based on the principle of maximum likelihood and they provide a systematic framework for the derivation of discrete prediction distributions. The latter are particularly important in the context of inventory control where features other than central tendency and variability, such as skewness, may affect safety stock levels.

The models considered are based upon three discrete probability distributions: the Poisson, the Pólya (negative binomial distribution with continuous index parameter) and the hurdle Poisson. Examination of the estimated prediction distributions enables consideration of the competing risks of stock-outs and excess inventory. As unanticipated changes in market structure may affect demand, corresponding changes to the values of the parameters of these distributions are accommodated by consideration of three forms of dynamic adjustment: *static*, where the parameters of the distributions remain unchanged from one period to the next; *undamped*, where the means and similar quantities are governed by the recurrence relationship associated with simple exponential smoothing (Brown, 1959); and *damped*, where shocks have only a transitory effect.

Hyndman, Koehler, Ord & Snyder (2008) and Snyder, Ord, & Beaumont (2012) examined forecasts for monthly demand time series for 1046 slow moving car parts. These studies indicated that the majority of series are adequately handled by static models although a minority of them are better managed with simple exponential smoothing dynamics. In addition it was found that the commonly used Croston method is not particularly effective and that simple exponential smoothing dynamics coupled with the Pólya distribution provide the most robust of the approaches considered.

From a practical perspective, these studies have a serious limitation: the sample size considered is the same for all items. In reality, an inventory system consists of a mix of items introduced at different points of time. Every year, tens of thousands of new items must be accommodated, for example, with the release of each new model of car. Sample sizes, as a consequence, can be quite variable and may often be quite small. Demand forecasts are required for all items, regardless of sample size, but standard maximum likelihood estimates are often not reliable when based upon short series.

One objective of this paper is to consider an extension of the earlier studies to account for variable sample sizes. In particular, we examine the effects of pooling data across time series in an attempt to resolve the small sample size problem, a method we refer to as the *multi-series approach*.

The paper is structured as follows: the basic time series models are summarised in Section 2; methods for estimation are considered in Section 3; a method for deriving lead time demand distributions is presented in Section 4; and a criterion used for scoring predictions from

different models is defined in Section 5. The theory from these sections is then applied in a comparison of the various forecasting methods on the demand data for car parts in Section 6. A survey of the current literature on intermittent demand forecasting may be found in Snyder, Ord, and Beaumont (2012).

2 Models

2.1 Basic Framework

The original study explored the possible use of the Poisson, Pólya and hurdle Poisson distributions to represent possible forms of the primary source of randomness in the state space models. Their mass functions, parameter restrictions, and other associated formulae are shown in Table 1. A conventional Poisson distribution in typical period *t* has a mean and variance equal to its parameter λ_t and is said to be equi-dispersed. The shifted Poisson distribution, in contrast, has a shifted mean $\lambda_t + 1$ but with the same variance λ_t. This distribution consequently only allows for under-dispersion of *positive* demands, yet in practice over-dispersed *positive* demands may be found.

Table 1: Discrete probability distributions where *y* is the value of a random variable and *λ, a, b, p* and *q* are parameters.

Distribution	Mass Function	Parameter Restrictions	Mean	Variance	Dispersion ratio
Poisson	$\frac{\lambda^y}{y!}\mathrm{e}^{-\lambda}$	$\lambda > 0$	λ	λ	1
Pólya	$\frac{\Gamma(r+y)}{\Gamma(r)\,y!}p^r q^y$	$r > 0$ $0 < p < 1$	rq/p	rq/p^2	$1/p$
Hurdle Poisson (Cragg, 1971)	$\begin{cases} q & \text{if } y = 0 \\ p\frac{\lambda^{y-1}}{(y-1)!}\mathrm{e}^{-\lambda} & \text{if } y = 1,2,\ldots \end{cases}$	$p \geq 0$ $q > 0$ $\lambda > 0$ $p + q = 1$	$p(\lambda+1)$	$pq(1+\lambda)^2 + p\lambda$	$q(1+\lambda) + \frac{\lambda}{1+\lambda}$

The mean of a probability distribution may vary over time to reflect changes in market structure. Denoted by μ_t, its evolution is potentially governed by any one of the recurrence relationships shown in Table 2. In the case where the means are the same from one period to the next, the presumption is that the market structure remains unchanged over time. A model with this property is said to be *static*. In contrast, a model with a mean that changes over time is said to be *dynamic*. The relationship in the second row of Table 2 is one important case corresponding to the updating relationship from simple exponential smoothing; the mean follows a random walk and so the associated time series is non-stationary. The complementary parameters δ and α determine the balance between the history of the

process, reflected by the preceding mean μ_{t-1}, and information contained in the most recent observation y_t. Another case, shown in the third row of Table 2, depicts means that follow a stationary first-order auto-regression. It is referred to here as the *damped* dynamic recurrence relationship. The parameter μ in this recurrence relationship represents the mean of the associated steady state distribution.

The hurdle Poisson distribution involves a parameter p which represents the probability of demand in a month. It is constant in the static case, but may change over time as depicted by the relationships in the third column of Table 2. The updating relationship involves a binary random variable x_t which equals 1 when some demand occurs in period *t*, but equals 0 otherwise. In general $x_t \le y_t$, so that $p \le \mu$ in the static case and $p_t \le \mu_t$ in the dynamic case. The recent Teunter, Syntetos, & Babai (2011) method relies on a similar recurrence relationship for the probability, but is not based on a modelling approach.

Table 2: State recurrence relationships where α, δ and μ designate parameters.

Relationship	Mean	Demand Occurrence Probability	Restrictions
Static	$\mu_t = \mu_{t-1}$	$p_t = p_{t-1}$	
Undamped dynamic	$\mu_t = \delta\mu_{t-1} + \alpha y_{t-1}$	$p_t = \delta p_{t-1} + \alpha x_{t-1}$	$\alpha > 0, \delta > 0$ $\alpha + \delta = 1$
Damped dynamic	$\mu_t = \delta\mu_{t-1} + \alpha y_{t-1}$ $+ (1-\delta-\alpha)\mu$	$p_t = \delta p_{t-1} + \alpha x_{t-1}$ $+ (1-\delta-\alpha)p$	$\alpha > 0, \delta > 0$ $\alpha + \delta < 1$

Given the value for the mean obtained from a recurrence relationship, the corresponding parameter value can be calculated. The required relationships are shown in Table 3. In the case of the Pólya distribution, its second parameter p is assumed to be time invariant. For the hurdle Poisson distribution, the condition $p_t \le \mu_t$ ensures that $\lambda_t \ge 0$, a condition which might be violated if the recurrence relationships in Table 2 for the mean and demand occurrence probability rely on different parameter values.

Table 3: Links between distribution parameters and means.

Distribution	Parameter
Poisson	$\lambda_t = \mu_t$

Pólya	$r_t = p\mu_t / q$
Hurdle-Poisson	$\lambda_t = \mu_t / p_t - 1$

2.2 Adapted Croston Model

In the original Croston approach, the mean time gaps between successive *active* periods and the mean non-zero demand quantities are related by suitably adapted simple exponential smoothing recursions. To be more specific, let α_t be a random variable that equals α in *active* periods and 0 in *inactive* periods. Let $\delta_t = 1 - \alpha_t$. The mean demands in active periods, designated by μ_t^+, are governed by the recurrence relationship

$$\mu_t^+ = \delta_t \mu_{t-1}^+ + \alpha_t y_{t-1}. \tag{2.1}$$

The mean times between gaps $\overline{\tau}_t$ evolve according to a similar formula

$$\overline{\tau}_t = \delta_t \overline{\tau}_{t-1} + \alpha_t \tau_{t-1} \tag{2.2}$$

where τ_{t-1} is the number of periods since the previous active period.

Hyndman et al., (2008) effectively augment the Croston recurrence relationships with a hurdle Poisson distribution. The time dependent probability p_t of the geometric distribution is linked to the mean gap $\overline{\tau}_t$ by the relationship:

$$p_t = 1/\overline{\tau}_t \,. \tag{2.3}$$

The other parameter of the hurdle Poisson distribution is obtained with

$$\lambda_t = \mu_t^+ - 1. \tag{2.4}$$

When equation (2.3) is used in conjunction with the recurrence relationship (2.2), τ_t and hence p_t, are fixed quantities at the start of period *t*. The mean μ_t^+ computed with the relationship (2.1) is also fixed. The effect of these calculations is to condition these quantities on the known and fixed demand data from periods 1 to $t-1$. Strictly speaking we should introduce notation to represent this conditioning, but to reduce potential complexity the required notation is suppressed. For fixed quantities in period *t*, there is an additional requirement: the parameters and the seed values μ_0^+ and $\overline{\tau}_0$ must also be fixed. These quantities are unknown, but for the moment it is assumed that they have been assigned known trial values. A method for estimating these unknown quantities is outlined in Section 3.4.1.

2.3 Harvey and Fernandes Model

The Harvey and Fernandes model is a discrete analogue of a linear Gaussian state space model. The series value in a typical period, conditional on the mean, is generated by a Poisson distribution instead of Gaussian distribution. The transition equation which describes

the evolution of the mean over time is non-linear with a beta distributed disturbance. The details can be found in Harvey & Fernandes (1989). The filter associated with their model, an analogue of the Kalman filter, turns out to be relatively simple: it is described in Section 3.4.2.

3 Estimation

Estimates of model parameters are required before prediction distributions can be derived. They may, however, be quite unreliable for relatively new items because of the paucity of data. A practical tactic is to assume that the demand for new items is likely to be similar to the demand for established or even obsolete items, and to use the latter's data to inform the estimation process for active items. In this section two basic strategies for estimation are examined, both of which are informed by statistics from a *training* group of obsolete items.

The first approach assumes that model parameters are potentially different for each item. This leads to adaptations of the single series approaches from Snyder, Ord, & Beaumont (2012) . The second approach, in contrast, assumes that one or more parameters are the same for all items. A multi-series approach to estimation is proposed as a means for reducing the effect of small sample sizes. Both approaches are considered in turn.

3.1 Single Series Approach to Estimation

The single series approach to estimation is divided into three stages: an initialisation stage which occurs when a part is completely new and no demand data is yet available; a run-in stage which encompasses the first M months (we use months 2 to 12 in our application) where there is some data but it is insufficient to estimate anything but the simplest model; and, a normal stage following the first M months where sample sizes are sufficient to contemplate the prospect of estimating more complex models. For convenience and for consistency with the application we now refer to the run-in period as the first year, and normal operations as the second year and beyond. Each of these stages is now considered in turn.

3.1.1 Initialisation

When an item is completely new and no demand data are available to inform the forecasting process, there is little choice but to make a crude approximation. One possibility is the following:

1. Monthly demand is governed by a single parameter distribution such as the Poisson distribution;
2. New items have a *common* mean estimated by a simple average $\hat{\mu}_1$ of the demands in the first month of items in the *training* collection.
3. The Poisson parameter is set equal to this estimate of the common mean.

3.1.2 Run-In (First Year)

There is a risk that the common mean used for the initial prediction is atypical of the demands for some items and so it is sensible to seek early predictions which adjust quickly to

the actual levels of early demands. Sample sizes are small in the first year, so we use a static Poisson model. The emphasis must remain on simple averages of demands. Let y_{it} designate the demand for car part *i* in month *t*. Then the simple average associated with part *i* in month *n* is

$$\hat{\mu}_{in} = \left(\hat{\mu}_1 + \sum_{t=1}^{n-1} y_{it} \right) \Big/ n, \tag{3.1}$$

which may be computed recursively with the weighted average formula

$$\hat{\mu}_{i,n+1} = \delta_n \hat{\mu}_{i,n} + \alpha_n y_{i,n} \tag{3.2}$$

where $\alpha_n = 1/(n+1)$ and $\delta_n = 1 - \alpha_n$. Equation (3.2) is reminiscent of the weighted average formula used for simple exponential smoothing, but the weights change over time. In particular, the weight given to the potentially atypical seed prediction $\hat{\mu}_1$ declines quickly as the first year unfolds.

3.1.3 Normal Prediction (Year 2 and onwards)

It is debatable when the simple model used in the run-in stage should be replaced by something more flexible. We elect to do so after the first year. The model considered may be based on any of the probability distributions in Table 1, any of the dynamics specified in Table 2, and may possess more than one parameter. The time invariant parameters are collected together into a vector θ; and the states are collected into a vector s_n. The time dependent parameters θ_n are computed from the state vector s_n using the appropriate formulae from Table 3.

Maximum likelihood methods are used to estimate both the unknown invariant parameter vector θ and the unknown seed state vector s_1. The states are then revised every month to ensure that forecasts track persistent changes in the demand patterns of individual items. In our application the invariant parameter vector estimates are only updated at the beginning of each year, but the frequency of updating is clearly at choice. It is ultimately an empirical question whether or not it makes sense to use an approach like this with small samples. The issue is implicitly examined in the empirical study reported in Section 6.

3.1.4 Maximum likelihood estimation

The likelihood function used in estimation is based on the joint density function $p(y_1, \ldots, y_n \mid \theta, s_1)$, which depends directly on the time invariant parameters θ and indirectly on the seed state s_1. The states are usually random quantities so this approach conditions on a fixed but unknown value of s_1. The estimates $\hat{\theta}$ and $\hat{s}_1$ obtained by maximising this likelihood function are referred to as *conditional* maximum likelihood estimates.

The process of likelihood evaluation for trial values of the parameter and seed state vectors is simplified by use of the decomposition

$$p(y_1,\ldots,y_n \mid \theta, s_1) = p(y_1 \mid \theta, s_1)\prod_{t=2}^{n} p(y_t \mid \theta, s_t) \tag{3.3}$$

where successive states s_t are generated by the relevant recurrence relationships in Table 2. In the case of the static models the recurrence relationship (3.2) is used to recursively compute the estimates of mean demand and the active periods' probability. The components of this decomposition consist of successive one-step-ahead prediction distributions. Although originally based on a multivariate distribution, the evaluation of the likelihood is thus simplified to finding the product of a sequence of univariate mass functions.

The Harvey-Fernandes and Croston methods are also adapted to handle the small sample issue for new items. The Harvey-Fernandes method depends on only one parameter: the discount factor δ. This parameter is optimised annually except for the first year. For the Croston method, the adapted framework from Hyndman et al. (2008) is used. The likelihood function used to obtain maximum likelihood estimates is outlined therein.

3.2 Multi-Series Approach to Estimation

The multi-series approach presumes that certain features of demand are common to all items in an inventory and that cross-series information can be exploited to improve the forecasts of individual series, particularly those of relatively new items constrained by short demand histories. Accordingly, the time invariant parameter vector θ is assumed to be the same for all items. Without prior knowledge, the seed state vector s_1 is also assumed to be the same for all items. Unlike θ, however, state vectors associated with subsequent periods are not forced to be the same. Idiosyncratic features of a series are captured by applying the appropriate state recurrence relationships in isolation from other demand series.

The common parameters θ and the common seed state s_1 are estimated from the demand data for a *training collection* of old or obsolete items. An advantage of this tactic is that we may select (relatively homogenous) demand series that are relatively long and are likely to reflect the variety of features observed in demands for all items under consideration. The common parameter vector and common seed state are again estimated using maximum likelihood. This time, however, they are selected to maximize *total* log-likelihood $\log \mathsf{L} = \sum_{i=1}^{m} \log \mathsf{L}_i(\theta, s_1)$ where $\mathsf{L}_i(\theta, s_1)$ designates the likelihood function for item i and m is the number of items. The training collection is kept relatively small to avoid unwarranted computational loads. We found in the application reported in Section 6 that $m = 20$ produces reliable estimates.

Once this estimation exercise is completed, the training collection is no longer needed. The focus changes to current items which form what we call the *prediction* collection. The estimates of the common parameters from the training collection are used in the relationships and distributions employed to generate successive prediction distributions for all items in the prediction collection. The recurrence relationships are seeded with the estimate of the common seed state for all items in the prediction collection. The first prediction distributions

are the same for all items in the prediction collection. When the prediction origin rolls forward through successive periods, the states are revised on an individual series basis, and successive prediction distributions soon diverge between items. Unlike the single series approach where the time invariant parameters are re-estimated periodically (e.g. at the beginning of each year), in the multi-series approach the parameter estimates remain unchanged. When the series for a particular item is sufficiently long, we may switch to a single series approach. Again, this is an empirical issue and not one that is pursued in this paper.

3.3 Filters

The common seed state in the multi-series approach will be atypical for some items. Its impact may be quickly reduced by accelerating the rate of adjustment of earlier predictions to the demand levels of individual items. The adjustment is normally too slow when states are calculated with the dynamic relationships in Table 2. S*imple exponential smoothing*, for example, relies on the undamped recurrence relationship

$$\hat{\mu}_{n+1} = \delta\hat{\mu}_n + \alpha y_n \tag{3.4}$$

for $n = 1, 2, \ldots$. to calculate the one-step ahead predictions $\hat{\mu}_{n+1}$. When the parameter α adopts values typically used in practice such as 0.1 or 0.2, it takes some time for the effects of the seed mean to effectively wash out of the calculations. Filters which adapt more quickly in their early stages of application are needed.

Quicker adaptation can be achieved by using larger values for α. As forecasts adjust to the actual levels of a series, however, the need for rapid adjustment declines and then smaller values of α should be used. Otherwise, the predictions remain too responsive to randomness in the data. A time dependent parameter, represented by α_n, thought of as a *short-run smoothing parameter* can be used, but it should possess the property that it eventually converges to a constant long-run value α. Implementation of a filter with the short-run smoothing constant must rely on the adaptation of the simple exponential smoothing recursion to the form:

$$\hat{\mu}_{n+1} = \delta_n\hat{\mu}_n + \alpha_n y_n . \tag{3.5}$$

We have already seen an example of this in Section 3.1.2 where the short-run smoothing parameter was governed by the formula $\alpha_n = 1/(n+1)$ for $n = 1, 2, \ldots$. That is, α_n begins with a large value of $\alpha_1 = 1/2$ but then it changes to successively smaller values, converging eventually to a constant value of zero. This makes sense if a static model applies so that the least squares estimate is a simple average. In dynamic models, the need for forecasts to adjust to structural change remains, even in the long-run. The short-run smoothing parameter α_n should converge to an appropriate constant non-zero value α.

A systematic approach to choosing the short-run smoothing parameter can be based on a variation of the theory of recursive least-squares estimation outlined in the Appendix. It involves the geometric series $S_n = \sum_{t=1}^{n} \left(\delta^2\right)^{t-1}$, something which can be computed recursively with the formula

$$S_{n+1} = \delta^2 S_n + 1 \tag{3.6}$$

after being seeded with the value $S_1 = 1$. It also involves a short-run smoothing parameter

$$\alpha_n = \frac{\delta\left(S_{n+1} - S_n\right)}{S_{n+1}} + \alpha \tag{3.7}$$

and the complementary parameter $\delta_n = 1 - \alpha_n$ which we call the *short-run discount factor*. $\delta\left(S_{n+1} - S_n\right)/S_n$ S_{n+1} S_n The effect is that the short-run smoothing parameter α_n converges as required to a constant long-run value α.

The case $\alpha = 0$ is not normally considered with exponential smoothing but it corresponds to the important static case where $\mu_t = \mu_{t-1}$. The formula (3.7) yields $\alpha_n = 1/\left(n+1\right)$ as used in Section 3.1.2 for the recursive calculation of a simple average. The other extreme is represented by the random walk case where $\alpha = 1$. Then $\delta_n = 0$ for all *n*, which ensures that the history of the time series is always ignored. The naïve prediction $\hat{\mu}_{n+1} = y_n$, emerges with an entire focus on the most recent observation.

The recursive least squares algorithm as outlined in the Appendix is designed to ensure that as the months pass, the weight given to the seed mean declines.

A similar filter applies for damped dynamics. The formula $\mu_t = \delta\mu_{t-1} + \alpha y_{t-1} + \left(1 - \delta - \alpha\right)\mu$ can be rewritten as $m_t = \delta m_{t-1} + \alpha \tilde{y}_{t-1}$ where $m_t = \mu_t - \mu$ and $\tilde{y}_t = y_t - \mu$. The transformed recurrence relationship has the same form as the recurrence relationship for an undamped model, the only difference being that $\delta + \alpha < 1$ instead of $\delta + \alpha = 1$. The least-squares predictions of the mean of the deviations of the short-run means from the long run mean are updated with

$$\hat{m}_{n+1} = \delta_n \hat{m}_n + \alpha_n \tilde{y}_n . \tag{3.8}$$

Remarkably, α_n is still calculated using the formula (3.7), the only difference being that $\delta_n = \alpha + \delta - \alpha_n$ instead of $\delta_n = 1 - \alpha_n$.

3.4 Other Approaches

3.4.1 Adapted Croston Method

The adapted Croston method is based on the model in Section 2.1.2. The evolution of the mean positive demand and the mean time gap is governed by the simple exponential smoothing recurrence relationships (2.1) and (2.2). The estimates of these means are also

computed using these relationships. Quicker adjustment, however, is needed in the early stages of the algorithm so the rules for determining S_t and α_t are altered to the following to reflect the theory in the previous section:

$$S_{n+1} = \begin{cases} S_n & \text{if } y_n = 0 \\ \delta^2 S_n + 1 & \text{if } y_n > 0 \end{cases} \tag{3.9}$$

$$\alpha_n = \begin{cases} 0 & \text{if } y_n = 0 \\ \dfrac{\delta\left(S_{n+1} - S_n\right)}{S_{n+1}} + \alpha & \text{if } y_n > 0 \end{cases}. \tag{3.10}$$

3.4.2 Harvey and Fernandes Filter

The filter for the Harvey-Fernandes model described in Section 2.3 is the Poisson analogue of the Kalman filter, but is very much simpler. The one-step-ahead prediction distributions needed for evaluating the likelihood function turn out to be Pólya distributions with a time dependent mean given by the finite exponentially weighted average

$$\hat{\mu}_{n+1} = \frac{\delta y_n + \delta^2 y_{n-1} + \delta^3 y_{n-2} \ldots + \delta^n y_1}{\delta + \delta^2 + \ldots + \delta^n}, \tag{3.11}$$

where δ satisfies the condition $0 \le \delta \le 1$. The numerator and denominator of this expression are designated by a_t and b_t respectively. They are calculated recursively with the expressions:

$$a_{n+1} = \delta(a_n + y_n) \text{ and } b_{n+1} = \delta(b_n + 1)$$

where $a_1 = b_1 = 0$. The parameters of the Pólya distribution in Table 1 are given by $r = a_{n+1}$ and $p = b_{n+1} / \left(1 + b_{n+1}\right)$.

4 Derivation of Prediction Distributions

Our focus is on prediction distributions of lead time demand. The quantity of interest from a forecast origin *n* is total demand over a lead time *L*. Designated by Y_n, it is defined as

$$Y_n = \sum_{t=n+1}^{n+L} y_t \,. \tag{4.1}$$

Y_n depends on *L* but this is not shown explicitly in the notation.

No formula for the probability distribution of Y_n is available for most of the models, but numerical approximations can be obtained using a parametric bootstrap The procedure is seeded with $t = n$, the maximum likelihood estimates of the parameter vector θ and the state vector s_n. The steps are:

1. Increment t by one month.

2. Obtain the new state s_t from the old state s_{t-1} and the old series value y_{t-1} using the appropriate least-squares filter[3].
3. Calculate the time dependent distribution parameters from the new state s_t using the appropriate formulae from Table 3.
4. Simulate a series value y_t from the appropriate distribution in Table 1 using the parameter values from Step 3.

These steps are repeated until $t > n + L$. The L simulated values are then used to calculate a value $Y_n = \sum_{t=n+1}^{n+L} y_t$ of lead time demand.

A sample of *R* lead time demands is generated by repeating this procedure *R* times The estimated lead time demand distribution may then be used in a computerised inventory control system to determine key control parameters such as order-up-to levels and reorder levels. An insight is provided by an example where the 90*th* percentile of the simulated distribution is used in an order-up-to level inventory system. This would ensure that the shortage probability *immediately before a replenishment delivery* is limited to about 10 per cent.

Care must be taken when this procedure is used with models which possess undamped dynamics. A problem common to all non-stationary random processes involving non-negative numbers may occur: simulated series values converge stochastically to a fixed point of zero in larger samples (Grunwald, Hamza, & Hyndman, 1997). The problem also exists for probabilities governed by non-stationary dynamics where an additional fixed point of one exists: simulated probabilities eventually get trapped at values of zero or one. This general problem has never been resolved. It does not appear to create difficulties in practice, however, provided the lead-time is of short to moderate duration (Akram, Hyndman and Ord, 2009). One way of guaranteeing that it never occurs is to use only models with damped dynamics, but this choice might result in a reduced prediction performance if the true underlying dynamics are undamped.

4.1 Prediction Scoring

Demand data that are withheld from the estimation process may be used to evaluate ex-ante predictions distributions. An observed lead time demand Y_t^{obs} may be used in measures such as the absolute prediction error, the rank probability score, and the logarithmic score (Good I J, 1952), defined as the negative of the log-likelihood.

The absolute prediction error is given by $\left|Y_t^{\text{obs}} - \hat{Y}_t\right|$ where $\hat{Y}_t$ is the *point* prediction. In situations where a forecasting method is applied repetitively to the same series using a rolling origin, the many resulting absolute prediction errors can be averaged to provide the *mean*

[3] Often the series would be simulated from the model underpinning the method. The advantage of the least squares filter is that it also provides an approximation for the estimation error associated with states such as the local mean. This is important given the small estimation samples that we consider later in the paper.

absolute error. Alternatively, the prediction errors may be squared and averaged to give the *mean squared error*. Such measures focus on point predictions, but for slow moving items other features such as the skewness of the distribution may be important when establishing safety stocks. This suggests the need for a focus on entire prediction distributions rather than point predictions alone.

We use the logarithmic score to evaluate prediction performance. It has a focus on the entire distribution and is readily extended if required to joint probability distributions of future time series values. It may be defined in the context of two models M_1 and M_2 which have been estimated from a time series $y_1, y_2, \ldots, y_n$ and which have been used to derive (normally by parametric bootstrap as described in Section 4) two distinct prediction distributions of lead time demand Y_n over periods $n+1, n+2, \ldots, n+L$. Let p_1 and p_2 denote the probabilities of the observed lead time demand Y_n^{obs} according to the two lead time distributions. When $p_1 > p_2$, the model M_1 is considered to be more likely than the model M_2 to have generated the observed lead time demand Y_n^{obs} An equivalent way of comparing the models is to rely on the log-ratio $\ln(p_1/p_2)$. In cases where a rolling prediction is used to generate many lead time distributions from the same model, the mean logarithmic score may be used to measure prediction performance and to rank prediction models.

5 Comparison of Approaches on Car Parts Data

An extension of a study of Snyder et al. (2012) of demand series for car parts from a US automobile company is now reported. The primary aim is to gauge the benefits, if any, of changing from the single series approach to the corresponding multi-series approach. Other aims are to evaluate the relative performance of various discrete probability distributions and of the various forms of dynamics. A final aim is to contrast these with established approaches for forecasting intermittent demand series (Croston, 1972; Harvey & Fernandes, 1989).

The data from the study relate almost entirely to SKUs with intermittent demands, the gap between active periods averaging about 2.5 months. Demands, when they occur, average about two. About 96 per cent of the monthly demand series are over-dispersed with a variance-to-mean ratio in excess of one. Even when the series are stripped of their zero values, about 22 per cent have a variance-to-mean ratio in excess of one.

As no complete data series for obsolete parts are available for the multi-series approach, 20 series were chosen at random from the 1046 series to use for training purposes. The various methods were then evaluated on the remaining 1026 series.

Twenty-four methods are compared in this study: the 18 methods obtained from the various combinations of measurement distributions (Poisson, Pólya and Hurdle Poisson), with various forms of dynamics (static, undamped and damped); another method that uses the best of these 18 approaches according to the traditional Akaike information criterion; and our adaptations of the traditional Croston method and the Harvey-Fernandes method. Both single series and multiple series versions of these approaches are considered.

Unlike in the earlier study, where models are fitted to series with a fixed length of 45 months and another six months of data are withheld for ex-ante prediction evaluation, the series lengths in this study are varied. In the single series approaches parameters are estimated at the beginning of each year but the state variables are revised each month. The sample sizes for parameter estimation are 0, 12, 24, 36 and 48. As it is not possible to fit the models to samples of size of 0, a Poisson distribution is used with a mean estimated by a simple average of demand in the first month for the 20 parts of the training sample. The simple average is adapted with each new observation in the first year. The models are only fitted to samples of size twelve and larger.

Lead time demand prediction distributions are obtained using the parametric bootstrap method (Section 4) using R=10,000 for lead times of one, three and six months (designated in the tables below by LT1, LT3 and LT6). The distributions for LT1 are of particular interest because they correspond to the one-step ahead prediction distributions. Unlike other lead time distributions, they can be derived using analytical methods. Because analytical methods give the most reliable results, analytical methods were used for LT1, while the parametric bootstrap method was used for LT3 and LT6, where no analytical method was available.

The lead time demand prediction distributions are gauged against withheld data using the logarithmic score. A calendar year is divided into 12, 4 or 2 contiguous, *non-overlapping* lead times for the lead times LT1, LT3 and LT6 respectively. The start of each lead time is used as a forecast origin. For each part, the logarithmic score is calculated for each lead time and averaged over all the lead times in a calendar year to give a statistic we call the annual logarithmic score (ALS). Being incomplete with only three months of data, the fifth year has to be treated a little differently, it not being possible to obtain the ALS for LT6. Of particular interest is the fact that the ALS for the case LT1 corresponds to the logarithmic score for the *joint* prediction distribution of monthly demands over a calendar year (see Section 3.1.3 on the prediction decomposition of the likelihood function).

One complication with the numerical approach to prediction evaluation is that sometimes actual lead time demand lies in the extreme tail of a prediction distribution. In such cases, where the prediction distribution is highly unlikely to have generated the observed lead time demand, the probability mass estimate produced by parametric bootstrap may be zero so that the corresponding logarithmic score is effectively infinite. Indeed, in this empirical study some ALSs turn out to be infinite.

A number of approaches were explored for presenting the results of the study and it transpired that they all lead to similar conclusions. In the end we opted to focus on ranks of the various methods found by comparing the ALSs on each of the 1026 evaluation series. There are a number of advantages in presenting the results in terms of ranks. First, any method with an ALS of minus infinity is automatically ranked last. There is no need to take any special measures to handle such aberrant cases. Second, average ranks emphasise *robustness,* a feature that is often seen as desirable in practice. When only two methods are compared, this approach is equivalent to looking at percentage wins.

5.1 Relative performance of multi-series models

Our primary aim is to determine whether the potential data shortages that accompany single series methods can be offset by pooling demand series. Pooling of series brings more information to bear and possibly improves estimation accuracy. However, the need to assume that parameters are the same across all series may offset this advantage. We had no preconceived notions as to which of these two conflicting forces prevail. We formulate this question as our first hypothesis.

H_1: Multi-series methods will perform as well or better than single series methods.

An answer for the car parts data is provided by Table 5. It has pairwise comparisons of average ranks for single and multiple series implementations of each method, including the Croston and Harvey-Fernandes methods. The lowest average ranks are highlighted in bold. Friedman rank score tests (not shown here) applied to each row of the table have most *p*-values close to zero, indicating that most of the depicted differences are statistically significant (not that this should be confused with practical significance). The main conclusion to be derived from this table is that most methods improve with data pooling. Accordingly, most of the remaining analysis will be restricted to the multi-series approaches.

It had been found in earlier studies (Billah, King, Snyder, & Koehler, 2006) that there were benefits to be realized from using a multi-model approach to forecasting with the Akaike information criterion (AIC) (Akaike, 1970) for model selection. These studies were conducted on the M3 dataset which consists mainly of non-intermittent time series. Our study indicates that those benefits may not be achievable when applied to low count time series with non-Gaussian distributions; see the last two rows of Table 5.

Table 5. Pairwise comparisons of single and multiple series approaches with ranks averaged across 1026 series using the ALS criterion. The best ranks are shown in bold font. S=single series approach; M=multiple series approach; HPoiss=hurdle Poisson. Avg 2-4 are the ranks averaged over years 2 to 4. The grand averages are the averages of the annual averages across the three lead times. Each average rank would be 1.5 if there were no difference between the approaches.

	Lead time		LT1							LT3							LT6						Grand	
		Year	1	2	3	4	5	Avg 1-5	Avg 2-5	1	2	3	4	5	Avg 1-5	Avg 2-5	1	2	3	4	Avg 1-4	Avg 2-4	Avg 1-5	Avg 2-5
Static	Poisson	S	1.54	1.64	**1.41**	**1.29**	**1.21**	**1.42**	**1.39**	1.69	1.70	1.57	**1.50**	1.39	1.57	1.54	1.64	1.69	1.64	1.58	1.64	1.64	1.54	1.52
		M	**1.46**	**1.36**	1.59	1.71	1.79	1.58	1.61	**1.31**	**1.30**	**1.43**	**1.50**	**1.61**	**1.43**	**1.46**	**1.36**	**1.31**	**1.36**	**1.42**	**1.36**	**1.36**	**1.46**	**1.48**
	Polya	S	1.76	1.62	1.53	1.51	1.52	1.59	1.54	1.71	1.62	1.57	1.56	1.51	1.60	1.57	1.62	1.62	1.59	1.57	1.60	1.59	1.59	1.57
		M	**1.24**	**1.38**	**1.47**	**1.49**	**1.48**	**1.41**	**1.46**	**1.29**	**1.38**	**1.43**	**1.44**	**1.49**	**1.40**	**1.43**	**1.38**	**1.38**	**1.41**	**1.43**	**1.40**	**1.41**	**1.41**	**1.43**
	HPoiss	S	**1.51**	1.68	**1.49**	**1.42**	**1.35**	1.49	1.49	**1.33**	1.64	1.66	1.66	1.64	1.59	1.65	**1.33**	1.62	1.64	1.64	1.56	1.63	1.55	1.59
		M	1.49	**1.32**	1.51	1.58	1.65	**1.51**	**1.51**	1.67	**1.36**	**1.34**	**1.34**	**1.36**	**1.41**	**1.35**	1.67	**1.38**	**1.36**	**1.36**	**1.44**	**1.37**	**1.45**	**1.41**
Undamped	Poisson	S	**1.48**	1.61	1.58	1.58	1.54	1.56	1.58	1.69	1.63	1.59	1.59	1.55	1.61	1.59	1.64	1.63	1.58	1.59	1.61	1.60	1.59	1.59
		M	1.52	**1.39**	**1.42**	**1.42**	**1.46**	**1.44**	**1.42**	**1.31**	**1.37**	**1.41**	**1.41**	**1.45**	**1.39**	**1.41**	**1.36**	**1.37**	**1.42**	**1.41**	**1.39**	**1.40**	**1.41**	**1.41**
	Polya	S	1.77	1.67	1.59	1.57	**1.49**	1.62	1.58	1.72	1.64	1.57	1.58	1.51	1.60	1.58	1.63	1.63	1.58	1.57	1.60	1.59	1.61	1.58
		M	**1.23**	**1.33**	**1.41**	**1.43**	1.51	**1.38**	**1.42**	**1.28**	**1.36**	**1.43**	**1.42**	**1.49**	**1.40**	**1.42**	**1.37**	**1.37**	**1.42**	**1.43**	**1.40**	**1.41**	**1.39**	**1.42**
	HPoiss	S	1.51	1.67	**1.50**	**1.49**	**1.50**	1.53	1.54	**1.34**	1.62	1.54	1.54	1.51	1.51	1.56	**1.34**	1.58	1.53	1.53	**1.49**	1.55	1.51	1.55
		M	**1.49**	**1.33**	**1.50**	1.51	**1.50**	**1.47**	**1.46**	1.66	**1.38**	**1.46**	**1.46**	**1.49**	**1.49**	**1.44**	1.66	**1.42**	**1.47**	**1.47**	1.51	**1.45**	**1.49**	**1.45**
Damped	Poisson	S	**1.42**	1.61	1.52	**1.47**	**1.42**	**1.49**	1.51	**1.34**	1.58	**1.50**	**1.43**	**1.48**	**1.47**	**1.50**	**1.33**	1.56	**1.50**	**1.42**	**1.45**	**1.49**	**1.47**	**1.50**
		M	1.58	**1.39**	**1.48**	1.53	1.58	1.51	**1.49**	1.66	**1.42**	**1.50**	1.57	1.52	1.53	**1.50**	1.67	**1.44**	**1.50**	1.58	1.55	1.51	1.53	**1.50**
	Polya	S	1.75	1.66	1.53	**1.47**	**1.42**	1.57	1.52	1.72	1.63	1.51	**1.46**	**1.42**	1.55	1.51	**1.68**	1.61	1.53	**1.45**	1.57	1.53	1.56	1.52
		M	**1.25**	**1.34**	**1.47**	1.53	1.58	**1.43**	**1.48**	**1.28**	**1.37**	**1.49**	1.54	1.58	**1.45**	**1.49**	**1.32**	**1.39**	**1.47**	1.55	**1.43**	**1.47**	**1.44**	**1.48**
	HPoiss	S	1.89	1.56	**1.39**	**1.31**	**1.27**	**1.48**	**1.38**	1.54	1.58	**1.42**	**1.33**	**1.29**	**1.43**	**1.40**	1.55	1.56	**1.39**	**1.32**	**1.46**	**1.42**	**1.46**	**1.40**
		M	**1.11**	**1.44**	1.61	1.69	1.73	1.52	1.62	**1.46**	**1.42**	1.58	1.67	1.71	1.57	1.60	**1.45**	**1.44**	1.61	1.68	1.54	1.58	1.54	1.60
Other	Harvey	S	1.73	1.64	1.60	1.55	**1.50**	1.60	1.57	1.66	1.65	1.62	1.56	**1.49**	1.60	1.58	1.62	1.65	1.62	1.56	1.61	1.61	1.60	1.59
		M	**1.27**	**1.36**	**1.40**	**1.45**	**1.50**	**1.40**	**1.43**	**1.34**	**1.35**	**1.38**	**1.44**	1.51	**1.40**	**1.42**	**1.38**	**1.35**	**1.38**	**1.44**	**1.39**	**1.39**	**1.40**	**1.41**
	Croston	S	**1.50**	1.67	1.57	1.54	**1.42**	1.54	1.55	1.67	1.66	1.56	1.55	**1.43**	1.57	1.55	1.66	1.64	1.54	1.54	1.59	1.57	1.57	1.56
		M	**1.50**	**1.33**	**1.43**	**1.46**	1.58	**1.46**	**1.45**	**1.33**	**1.34**	**1.44**	**1.45**	1.57	**1.43**	**1.45**	**1.34**	**1.36**	**1.46**	**1.46**	**1.41**	**1.43**	**1.43**	**1.44**
	AIC	S	1.89	**1.47**	**1.34**	**1.30**	**1.27**	**1.45**	**1.35**	1.54	1.51	**1.40**	**1.33**	**1.32**	**1.42**	**1.39**	1.55	**1.50**	**1.40**	**1.32**	**1.44**	**1.41**	**1.44**	**1.38**
		M	**1.11**	1.53	1.66	1.70	1.73	1.55	1.65	**1.46**	**1.49**	1.60	1.67	1.68	1.58	1.61	**1.45**	**1.50**	1.60	1.68	1.56	1.59	1.56	1.62

5.2 Choice of distribution and dynamic structure

Of particular interest is the choice of probability distribution. Our initial thoughts before undertaking the study may be formulated as a series of hypotheses:

H_2: The Pólya models will perform better than the corresponding Poisson versions.

The Pólya distribution assumes that demands occur sporadically but, unlike the Poisson distribution, it allows for an uncertain quantity governed by a logarithmic distribution rather than a fixed quantity of one. The variance-mean ratio is greater than one, so we anticipated that its static model would out-perform the static Poisson distribution. We were unsure how the dynamic Pólya model would compare with dynamic Poisson models, but thought the flexibility provided by its additional parameter might lead to better predictions.

H_3: The hurdle Poisson models will perform better than the others.

The hurdle Poisson distribution assumes that demands occur in sporadic *periods* as distinct from sporadic *points of time*. The quantities demanded in active periods are also uncertain, following a shifted Poisson distribution. The logarithmic distribution underpinning the Pólya distribution always has a mode of one, whereas the shifted Poisson distribution can have a mode of any positive integer value. Thus, our initial thoughts were that the hurdle Poisson distribution would prove to be more flexible than the Pólya distribution and that as a consequence, its models would outperform all others. However, we also anticipated that this

effect might be mitigated by the fact that the hurdle Poisson distribution cannot properly accommodate the 20 per cent of series with over-dispersed *positive* demands.

Table 6 has three way comparisons of the three probability distributions for static, undamped and damped dynamics. Interestingly, our expectations only proved to be partially correct. Again, using a bold font to highlight cases with the lowest average rank, the Pólya distribution is a clear winner, validating H_2 but not H_3. The reasons why the Pólya is better than the hurdle Poisson distribution are unclear but probably relate to the heavier tails of the predictive distribution.

Table 6. Comparisons of probability distributions for each dynamic form using ranks of average annual logarithmic scores averaged across 1026 series for multiple series approaches only. Best ranks are shown in bold font. HPoiss=hurdle Poisson; Undamp=Undamped; Avg 2-4 are the ranks averaged over years 2 to 4. The grand averages are the averages of the annual averages across the three lead times. Each average rank would be 2 if there were no difference between the distributions.

	Lead time	LT1							LT3							LT6						Grand	
	Year	1	2	3	4	5	Avg 1-5	Avg 2-5	1	2	3	4	5	Avg 1-5	Avg 2-5	1	2	3	4	Avg 1-4	Avg 2-4	Avg 1-5	Avg 2-5
Static	Poisson	2.28	2.18	2.26	2.41	2.41	2.31	2.31	1.89	2.22	2.41	2.51	2.25	2.26	2.35	1.66	2.18	2.35	2.43	2.16	2.32	2.24	2.33
	Polya	**1.63**	**1.62**	**1.60**	**1.58**	**1.71**	**1.63**	**1.63**	**1.44**	**1.73**	**1.69**	**1.65**	**1.83**	**1.67**	**1.73**	**1.58**	**1.78**	**1.75**	**1.65**	**1.69**	**1.73**	**1.66**	**1.69**
	HPoiss	2.09	2.20	2.14	2.00	1.88	2.06	2.06	2.67	2.05	1.89	1.84	1.92	2.07	1.92	2.76	2.04	1.90	1.91	2.15	1.95	2.10	1.98
Undamp	Poisson	2.30	2.13	2.05	2.06	2.08	2.12	2.08	1.89	2.06	2.05	2.00	2.00	2.00	2.03	1.66	1.97	**1.92**	**1.88**	1.86	1.92	1.99	2.01
	Polya	**1.62**	**1.64**	**1.68**	**1.72**	**1.70**	**1.67**	**1.69**	**1.44**	**1.75**	**1.78**	**1.85**	**1.83**	**1.73**	**1.80**	**1.58**	**1.85**	**1.92**	1.94	**1.82**	**1.90**	**1.74**	**1.80**
	HPoiss	2.08	2.24	2.27	2.22	2.22	2.21	2.24	2.66	2.19	2.16	2.16	2.17	2.27	2.17	2.76	2.19	2.16	2.18	2.32	2.18	2.27	2.19
Damped	Poisson	2.67	2.11	2.11	2.19	2.23	2.26	2.16	2.34	2.18	2.18	2.23	2.11	2.21	2.18	2.34	2.18	2.15	2.21	2.22	2.18	2.23	2.17
	Polya	2.03	**1.63**	**1.57**	**1.56**	**1.61**	**1.68**	**1.59**	**1.56**	**1.65**	**1.62**	**1.58**	**1.74**	**1.63**	**1.65**	**1.68**	**1.71**	**1.65**	**1.62**	**1.66**	**1.66**	**1.66**	**1.63**
	HPoiss	**1.30**	2.26	2.32	2.25	2.15	2.06	2.25	2.10	2.16	2.20	2.18	2.15	2.16	2.17	1.97	2.11	2.20	2.18	2.12	2.16	2.11	2.19

Another issue of interest is the choice of dynamics. Our initial view was that given the nature of the data, it was unlikely that the undamped or damped models would do substantially better than their static analogues.

H_4: The dynamic versions will out-perform their static counterparts.

Table 7 shows three way comparisons of the dynamics for each probability distribution. It clearly shows that undamped models have the lowest average ranks. Again, although the calculations are not shown here, the differences between the approaches are statistically significant. Whether this translates into practical significance in terms of more effective control of inventory costs is unclear. The additional economic data needed to examine this important issue was not available.

The statistical analysis summarised by Tables 5, 6 and 7 strongly suggests that if one seeks a single approach to forecasting lead time demands, one can do little better than to use simple exponential smoothing. However, in recognition of the intermittent nature of demand, this traditional forecasting method must be coupled with a Pólya probability distribution to describe the predictive distribution of lead time demand.

Table 7. Comparisons of dynamics for each for each probability distribution using ranks of average annual logarithmic scores averaged across 1026 series for multiple series approaches only. Best ranks are shown in bold font. HPoiss=hurdle Poisson; Avg 2-4 are the ranks averaged over years 2 to 4. The grand averages are the averages of the annual averages across the three lead times. Each average rank would be 2 if there were no difference between the distributions.

	Lead time	LT1							LT3							LT6						Grand	
	Year	1	2	3	4	5	Avg 1-5	Avg 2-5	1	2	3	4	5	Avg 1-5	Avg 2-5	1	2	3	4	Avg 1-4	Avg 2-4	Avg 1-5	Avg 2-5
Poisson	Static	**1.88**	2.07	2.20	2.28	2.38	2.16	2.24	**1.70**	2.11	2.24	2.31	2.41	2.15	2.27	**1.72**	2.05	2.24	2.28	2.07	2.19	2.13	2.23
	Undamped	2.03	**1.87**	**1.77**	**1.69**	**1.60**	**1.79**	**1.73**	1.75	**1.72**	**1.65**	**1.61**	**1.66**	**1.68**	**1.66**	1.85	**1.75**	**1.70**	**1.61**	**1.73**	**1.68**	**1.73**	**1.69**
	Damped	2.08	2.06	2.03	2.03	2.01	2.04	2.03	2.55	2.17	2.10	2.08	1.94	2.17	2.07	2.43	2.20	2.06	2.11	2.20	2.13	2.14	2.08
Polya	Static	**1.90**	2.18	2.20	2.24	2.31	2.17	2.23	**1.91**	2.11	2.17	2.24	2.35	2.16	2.22	**1.95**	2.04	2.18	2.24	2.10	2.15	2.14	2.20
	Undamped	1.98	**1.80**	**1.74**	**1.73**	**1.68**	**1.79**	**1.74**	1.93	**1.94**	**1.83**	**1.77**	**1.77**	**1.85**	**1.83**	2.01	2.01	**1.90**	**1.81**	**1.93**	**1.91**	**1.86**	**1.82**
	Damped	2.12	2.02	2.06	2.03	2.01	2.05	2.03	2.16	**1.94**	1.99	1.99	1.88	1.99	1.95	2.04	**1.95**	1.92	1.95	1.97	1.94	2.00	1.97
HPoiss	Static	2.36	1.89	1.78	1.84	1.89	1.96	1.85	2.19	1.99	1.86	1.89	1.88	1.96	1.91	2.40	1.97	1.88	1.89	2.03	1.91	1.98	1.89
	Undamped	2.58	**1.78**	**1.77**	**1.68**	**1.56**	**1.87**	**1.70**	2.27	**1.86**	**1.72**	**1.61**	**1.55**	**1.80**	**1.69**	2.43	**1.90**	**1.71**	**1.61**	**1.91**	**1.74**	**1.86**	**1.71**
	Damped	**1.06**	2.33	2.45	2.48	2.54	2.17	2.45	**1.54**	2.16	2.41	2.49	2.56	2.23	2.41	**1.17**	2.13	2.41	2.49	2.05	2.34	2.15	2.40

The above conclusion ignores the traditional Croston method and the Harvey-Fernandes method. Our basic hypotheses for these distributions were:

H_5: The dynamic Croston model will out-perform its static counterpart (the static hurdle Poisson).

H_6: The dynamic version of the Pólya *model will out-perform the Croston and Harvey-Fernandes models.*

The Harvey-Fernandes method is restrictive in the sense that it sets a probability parameter in the Pólya distribution equal to the smoothing parameter implied by its discount factor and there seems to be no ostensible reason for this. The undamped Pólya model has a different parameter for the Pólya distribution so it involves one additional parameter, which should lead to an improved fit. The results are summarised in Table 8. They indicate that the undamped Pólya model is better than both the Croston and Harvey-Fernandes methods for this car parts data set. Another curious point which we cannot explain is that the adapted Croston method is not as good as its static analogue: the method based on the static hurdle Poisson model.

One interesting aspect to these results is that they appear to be largely independent of the length of the lead times. Moreover, they are consistent with our earlier study based on single series approaches where there the undamped Pólya model was best and where the Croston method did not shine forth.

Table 8. Comparisons of common forecasting methods and some of our methods using ranks of average annual logarithmic scores averaged across 1026 series for multiple series approaches only. Best ranks are shown in bold font. HPoiss=hurdle Poisson; Undamp=Undamped; Avg 2-4 are the ranks averaged over years 2 to 4. The grand averages are the averages of the annual averages across the three lead times. The first three pairwise comparisons would have an average rank 1.5 if there were no differences; the final 3-way comparisons would have a rank of 2 if there were no differences.

Lead time	LT1							LT3							LT6						Grand	
Year	1	2	3	4	5	Avg 1-5	Avg 2-5	1	2	3	4	5	Avg 1-5	Avg 2-5	1	2	3	4	Avg 1-4	Avg 2-4	Avg 1-5	Avg 2-5
Static HPoiss	1.97	**1.36**	**1.24**	**1.21**	**1.20**	**1.39**	**1.25**	1.73	**1.42**	**1.27**	**1.22**	**1.17**	**1.36**	**1.27**	1.91	**1.42**	**1.27**	**1.22**	**1.46**	**1.31**	**1.41**	**1.28**
Croston	**1.03**	1.64	1.76	1.79	1.80	1.61	1.75	**1.27**	1.58	1.73	1.78	1.83	1.64	1.73	**1.09**	1.58	1.73	1.78	1.54	1.69	1.59	1.72
Undamped Polya	**1.37**	**1.37**	**1.31**	**1.38**	**1.37**	**1.36**	**1.36**	**1.50**	**1.46**	**1.39**	**1.40**	**1.39**	**1.43**	**1.41**	1.60	**1.50**	**1.44**	**1.42**	**1.49**	**1.45**	**1.43**	**1.41**
Harvey	1.63	1.63	1.69	1.62	1.63	1.64	1.64	**1.50**	1.54	1.61	1.60	1.61	1.57	1.59	**1.40**	**1.50**	1.56	1.58	1.51	1.55	1.57	1.59
Undamped Polya	1.80	**1.33**	**1.29**	**1.32**	**1.32**	**1.41**	**1.31**	**1.35**	**1.37**	**1.37**	**1.35**	**1.34**	**1.36**	**1.36**	**1.38**	**1.42**	**1.39**	**1.36**	**1.39**	**1.39**	**1.39**	**1.35**
Croston	**1.20**	1.67	1.71	1.68	1.68	1.59	1.69	1.65	1.63	1.63	1.65	1.66	1.64	1.64	1.62	1.58	1.61	1.64	1.61	1.61	1.61	1.65
Undamped Polya	2.18	**1.70**	**1.61**	**1.70**	**1.69**	**1.77**	**1.67**	1.85	**1.82**	**1.76**	**1.75**	**1.74**	**1.78**	**1.77**	1.97	1.92	**1.83**	**1.79**	1.88	**1.85**	**1.81**	**1.76**
Harvey	2.61	2.03	2.04	1.92	1.99	2.12	2.00	**1.83**	1.90	1.98	1.92	1.98	1.92	1.94	**1.79**	**1.84**	1.88	1.87	**1.85**	1.86	1.96	1.93
Croston	**1.21**	2.27	2.35	2.39	2.32	2.11	2.33	2.31	2.27	2.27	2.33	2.28	2.29	2.29	2.24	2.24	2.29	2.34	2.28	2.29	2.23	2.30

5.3 Performance during run-in period

We also expected that the forecasts in the first year based on a Poisson distribution, although informed by data from the training sample, would be rather poor. However, we also expected that when richer models cut-in from the second year onwards, the predictions would improve with increases in sample size. Table 9 reports a representative example which refers to the multiple series version of the undamped Pólya model with a focus on the one-month prediction distributions. It shows the median ALS calculated across the 1026 series. The median scores decrease with increases in the sample size as anticipated.

Table 9. Median ALSs for the LT1 distribution from the undamped Pólya model

Year	1	2	3	4	5
Score	1.49	1.33	1.11	0.94	0.82

5.4 Specification of parameters

We had fewer expectations about parameter values. Table 10 shows the values obtained by maximising the log-likelihood across the 20 training series. The undamped multi-series approaches based on the Poisson and Pólya distributions, together with the Croston method, all had optimized values of the smoothing parameter α quite close to 0.1 (obtained by subtracting the long-run δ from one). This is consistent with the common viewpoint (Brown, 1959) that 0.1 should normally be used when demand is reasonably stable. This viewpoint was originally formed in an age when computers were primitive and not widely available in business and industry. It was not possible to countenance the use of maximum likelihood methods on large numbers of demand series. The now dramatically different situation makes it practicable to optimise the value of α on an individual series basis. Indeed, Fildes, Hibon, Makridakis, & Meade (1998) found that when applied to telecommunications series, there

were distinct advantages from optimisation. Such benefits are not apparent for the slow moving car parts demand data.

Table 10. Estimated parameter values from training series.

Distribution	Dynamic	μ_1	μ	δ	α	p_1	p	τ_1
Poisson	static	1.547						
Poisson	undamped	1.617		0.892	0.108			
Poisson	damped	1.541	2.470	0.334	0.159			
Polya	static	1.473					0.531	
Polya	undamped	1.488		0.904	0.096		0.487	
Polya	damped	1.238	2.597	0.296	0.146		0.521	
HPoisson	static	0.156				0.056		
HPoisson	undamped	0.152		0.938	0.062	0.055		
HPoisson	damped	0.048	2.295	0.010	0.010	0.017	0.001	
Harvey	Other			0.820				
Croston	Other	2.518		0.880	0.120			3.197

6 Concluding Remarks

The main aim of this paper was to adapt the maximum likelihood approach to exponential smoothing to a form that may be applied to the realities of forecasting demand for slow moving inventories when historical data are sparse. We explored a multi-series approach where critical parameters such as the smoothing constant were assumed to be the same across all series. This implies that all series are subject to the same levels of structural change, an assumption that may seem to be rather dubious at first sight. Yet, the advantage of this approach is that the small sample size problem is mitigated by being able to pool data across series. Our study on car parts data showed that the statistical efficiency gains from this approach offset any losses that may result by not tailoring the parameter values to the characteristics of individual series. Indeed we confirmed that the common practice of using a blanket value of 0.1 for the smoothing parameter can be most effective for most versions of exponential smoothing, so obviating the need for optimisation in these cases.

A surprise in our earlier study was that the Croston method seemed not to work nearly as well as other methods such as the one based on a Pólya distribution with undamped dynamics for the mean. This finding is confirmed in the current study where the multi-series version of the Croston method proposed in this paper is still outperformed by most other methods. We suspect that this is due to the fact that estimates are only revised when there are positive demands so that effective samples sizes are much smaller.

7 References

Akaike, H. (1970). Statistical predictor identification, *Annals of the Institute of Statistical Mathematics, 22,* 203-217.

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.

Brown, R. G. (1959). *Statistical Forecasting for Inventory Control*. New York: McGraw-Hill.

Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica, 39,* 829-389.

Croston, J. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly, 23*(3), 289-303.

Fildes, R., Hibon, M., Makridakis, S., & Meade, N. (1998). Generalising about univariate forecasting methods: further empirical evidence. *International Journal of Forecasting, 14*(3), 339-358. doi:10.1016/S0169-2070(98)00009-0

Good I J. (1952). Rational decisions. *Journal of the Royal Statistical Society, Series B, 14,* 107-114.

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

Harvey, A. C., & Fernandes, C. (1989). Time series models for count or qualitative observations. *Journal of Business and Economic Statistics, 7*(4), 407-417.

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Berlin & New York: Springer.

Snyder, R D. (1985). Recursive estimation of dynamic linear models. *Journal of the Royal Statistical Society, Series, B., 47*, 272-276.

Snyder, R.D., Ord, J. K., & Beaumont, A. (2011). Forecasting the intermittent demand for slow-moving inventories: A modelling approach. *International Journal of Forecasting*. Elsevier. Retrieved from http://www.sciencedirect.com/science/article/pii/S0169207011000781

Teunter, R., Syntetos, A. A., & Babai, M. Z. (2011). Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research, 214*(3), 606-615.

Appendix

The least-squares filter is derived in this Appendix. It is based on the assumption that there is no prior information about demands. It forms the basis of a filter described in Section 3.3 where its equations have been adapted to incorporate prior information from the series in the tuning collection. The least-squares filter is a special case of the Kalman filter for a linear

innovations state space model (Snyder, 1985; Hyndman et al., 2008) . The derivation presented here is justified because it is more transparently related to simple exponential smoothing than the Kalman filter; and, because it provides a new and more revealing formula for the Kalman gain (short-run smoothing parameter).

It is assumed that the trajectory for the mean is described in a model by a recurrence relationship

$$\mu_t = \delta\mu_{t-1} + \alpha y_{t-1}. \tag{A.1}$$

Its solution has the general form

$$\mu_t = a_t + \delta^{t-1}\mu_1 \tag{A.2}$$

where the quantities a_t can calculated with the recurrence relationship

$$a_t = \delta a_{t-1} + \alpha y_{t-1} \tag{A.3}$$

where $a_1 = 0$. This can be established by substituting (A.2) into (A.1) and equating like terms. Note that (A.3) has the same form as (A.1), but the quantities μ_t and a_t differ because their common recurrence relationship is seeded with different values.

When μ_t is used as a predictor of y_t, the associated prediction error is $e_t = y_t - \mu_t$. The latter can be rewritten in terms of the seed mean, using (A.2) to give $e_t = y_t - a_t - \delta^{t-1}\mu_1$. The seed mean μ_1 is unknown, but an approximation $\hat{\mu}_1$ may be found by minimising the sum of squared errors $\sum_{t=1}^{n}\left(y_t - a_t - \delta^{t-1}\hat{\mu}_1\right)^2$. The least-squares estimator of the seed mean, based on a sample of size *n* and designated by $\hat{\mu}_{1|n}$, is given by

$$\hat{\mu}_{1|n} = \sum_{t=1}^{n}\delta^{t-1}\left(y_t - a_t\right) \Big/ \sum_{t=1}^{n}\left(\delta^2\right)^{t-1}. \tag{A.4}$$

The sample size expands by one each month and the estimate of the seed mean must be revised. Unnecessary duplication of calculations involved in the re-use of equation (A.4) can be avoided by employing a recursive scheme where $\hat{\mu}_{1|n}$ is calculated from the previous estimate $\hat{\mu}_{1|n-1}$. By comparing the formula (A.4) for $\hat{\mu}_{1|n}$ and $\hat{\mu}_{1|n-1}$ it can be established that

$$\hat{\mu}_{1|n} = \hat{\mu}_{1|n-1} + \frac{\delta^{n-1}}{S_n}\left(y_n - a_n - \delta^{n-1}\hat{\mu}_{1|n-1}\right). \tag{A.5}$$

where

$$S_n = \sum_{t=1}^{n}\left(\delta^2\right)^{t-1} = \frac{1-\delta^{2n}}{1-\delta^2}. \tag{A.6}$$

The means in subsequent periods are linearly related to the seed mean and so is the estimate $\hat{\mu}_{n|n}$ of μ_n based on a sample of size *n*. Multiply both sides of (A.5) by δ^{n-1}, add a_n and use (A.2) to get the result.

$$\hat{\mu}_{n|n} = \frac{S_{n-1}}{S_n}\hat{\mu}_{n|n-1} + \frac{\left(\delta^2\right)^{n-1}}{S_n} y_n . \tag{A.7}$$

How do we obtain the one-step ahead prediction $\hat{\mu}_{n+1|n}$? Using (A.1) seeded with $\hat{\mu}_{1|n}$ we get

$$\hat{\mu}_{n+1|n} = \delta\hat{\mu}_{n|n} + \alpha y_n . \tag{A.8}$$

Multiply (A.7) by δ and add αy_n to give

$$\hat{\mu}_{n+1|n} = \frac{\delta S_{n-1}}{S_n}\hat{\mu}_{n|n-1} + \left(\frac{\delta\left(S_n - S_{n-1}\right)}{S_n} + \alpha\right) y_n . \tag{A.9}$$

This has the form

$$\hat{\mu}_{n+1|n} = \delta_n\hat{\mu}_{n|n-1} + \alpha_n y_n \tag{A.10}$$

where

$$\delta_n = \delta S_{n-1}/S_n \tag{A.11}$$

and

$$\alpha_n = \frac{\delta\left(S_n - S_{n-1}\right)}{S_n} + \alpha \tag{A.12}$$

Note that $\delta_n + \alpha_n = 1.$

As an example of the rate of convergence of these weights to their limiting values, when $\alpha = 0.1, \delta = 0.9$ and it follows that for periods 2 and 12 $\delta_2 = 0.50$ and $\delta_{12} = 0.88.$

A similar result emerges when the mean follows the damped process

$$\mu_t = \left(1-\delta-\alpha\right)\mu + \delta\mu_{t-1} + \alpha y_{t-1} . \tag{A.13}$$

The prediction error becomes $e_t = \left(y_t - \left(1-\delta-\alpha\right)\mu - \delta\mu_{t-1} - \alpha y_{t-1}\right)$ which can be rewritten as $e_t = \tilde{y}_t - \delta\tilde{\mu}_{t-1} - \alpha\tilde{y}_{t-1}$ where $\tilde{y}_t = y_t - \mu$ and $\tilde{\mu}_t = \mu_t - \mu$. The arguments then follow as above except that now $\delta + \alpha < 1$ for stationarity.