ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

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

Forecasting the Intermittent Demand for Slow-Moving Items

Keith Ord, Ralph Snyder and Adrian Beaumont

May 2010

Working Paper 12/10

Forecasting the Intermittent Demand for Slow-Moving Items

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

Abstract

Organizations with large-scale inventory systems typically have a large proportion of items for which demand is intermittent and low volume. We examine different approaches to 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. We develop a forecasting framework based upon the zero-inflated Poisson distribution (ZIP), which enables the explicit evaluation of the multi-period lead-time demand distribution in special cases and an effective simulation scheme more generally. We also develop performance measures related to the entire predictive distribution, rather than focusing exclusively upon point predictions. The ZIP model is compared to a number of existing methods using data on the monthly demand for 1,046 automobile parts, provided by a US automobile manufacturer. We conclude that the ZIP scheme compares favorably to other approaches, including variations of Croston's method as well as providing a straightforward basis for inventory planning.

Author Keywords: Croston's method; Exponential smoothing; Intermittent demand; Inventory control; Prediction likelihood; State space models; Zero-inflated Poisson distribution

JEL classification: C22

---

[1] McDonough School of Business, Georgetown University, Washington, DC 20057, USA ordk@georgetown.edu

[2] Department of Econometrics and Business Statistics, Monash University, Clayton, VIC 3800, Australia ralph.snyder@buseco.monash.edu.au and adrian.beaumont@buseco.monash.edu

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 demand as *intermittent.* When the average size of a customer order is large, a continuous distribution is a suitable description, but when orders are placed for a relatively small number of items a discrete distribution is more appropriate. In this paper the term “order” will refer only to orders from customers.

In this paper our interest focuses upon intermittent demand with low volume. On occasion, such stock keeping units (SKUs) may be very high value as, for example, spare aircraft engines. But even when individual units are of low value, it is not unusual for such components to represent a large percentage of the number of SKUs, so that collectively they represent an important element in the planning process. For example, Johnston and Boylan (1996a) cite an example where the average number of purchases by a customer for an item was 1.32 occasions per year and that “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 2,509 series with complete records for 51 months, only 1,046 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 is successfully handled using exponential smoothing methods, for which a voluminous literature exists; see, for example Ord et al. (1997) and Hyndman et al. (2008). When volumes are lower, the exponential smoothing framework remains appropriate but a distribution that describes count data must be used in place of the normal distribution. Further, as recently emphasized by Syntetos, Nikolopoulos and Boylan (2010), it is not sufficient to look at point forecasts when making inventory decisions. Those authors recommend the use of stock control metrics. We accept their viewpoint completely, but have opted to work with measures that evaluate the performance of the overall prediction distribution.

In the remainder of this section we review the literature on forecasting intermittent demand. In particular, the demand for spare parts may either increase over time as equipment ages, or decline as units fail completely or are withdrawn from service. Thus, our focus is upon models that allow for non-stationary processes so we do not consider exclusively stationary schemes.

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

- an exponential smoothing scheme to update expected order size
- an exponential smoothing scheme to update the time gap to the next order
- an assumption that timing and order size are independent.

Since that time a number of extensions and improvements have been made to Croston's method, notably by Johnston and Boylan (1996a) and Syntetos and Boylan (2001). The latter authors develop a method, which we refer to as the bias-adjusted Croston method, for removing bias in the expectation. Snyder (2002) identifies some logical inconsistencies in the original Croston method and examines the use of a time-dependent Bernoulli process. Unlike Croston, distinct smoothing parameters were used for the positive demands and the time gaps. In the empirical testing phase, Snyder found that the maximum likelihood estimate for the smoothing parameter for the Bernoulli component was typically zero, an indication that little is lost by using a constant probability model. Snyder went on to develop a simulation procedure that provides a numerical determination of the predictive distribution for lead-time demand. Shenstone and Hyndman (2005) show that there is no possible model leading to the Croston forecast function unless we allow a sample space for order size that can take on negative as well as positive values. They assume a constant parameter Bernoulli process for the number of orders in a given lead-time and combine this with a normal distribution for order size and then derive the resulting predictive distribution for lead-time demand. Given these results, we focus upon models with a fixed Bernoulli parameter, which has the advantage that it enables us to develop explicit results in some circumstances for the predictive distributions. Nevertheless, when we refer to the Croston method later in the paper, we shall mean the single smoothing parameter version.

1.2 Low volume, intermittent demand

There is an extensive literature on low count time series models that are 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). They have been categorized by Cox (1981) as 'observation-driven', when the mean depends on lagged values of the count variable, and 'parameter driven', when the mean is a function of latent factors. The observation-driven models (Shephard, 1995; Davis et. al., 1999; Heinen, 2003; Jung et al., 2006) essentially have a single source of randomness. The parameter driven models (West et al, 1985; Zeger, 1988; Harvey and Fernandes, 1989; West and Harrison, 1997; Davis et al, 2000; Durbin and Koopman, 2001) have an additional source of randomness driving the evolution of the mean. In addition there are multiple source of error approaches (Al-Osh and Alzaid, 1987; McKenzie, 1988; McCabe and Martin, 2005) involving adaptations of the ARMA framework to a count data setting using binomial thinning (INAR models).

Models belonging to both categories have been compared in Feigin et al (2008). A theoretical analysis indicated that a dual source of error model was more flexible than its single source of error analogue, something that is not true for Gaussian measurements (Hyndman et al., 2008). However, their analysis was conducted under a stationarity assumption. As noted earlier, demand series are typically non-stationary, and we believe that in this context a single source of error approach still provides sufficient flexibility for modeling intermittent demand series.

1.3 The evidence

Notwithstanding the theoretical objections just raised to Croston's method and its extensions, we might still expect that it would out-perform exponential smoothing methods that completely fail to account for the intermittent nature of demand. Yet, the results have been less than compelling.

Willemain et al. (1994) conducted an extensive simulation study that violated some of the original assumptions (such as cross-correlations between order size and the time between orders) and found that substantial improvements were possible. When they tested the method on real data the benefits were modest for one-step-ahead forecasts. However, as Johnston and Boylan (1996b) point out in a comment, improvements are only to be expected when the average time between orders is appreciably longer than the periodic review time. Willemain et al. (2004) describe a bootstrap –based approach that allows for a Markov chain development of the probability of an order and indicate that their method produces better inventory decisions than either exponential smoothing or Croston's 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.

Sani and Kingsman (1997) also conducted a sizeable simulation study that compared various methods. They used multiple criteria including overall cost criteria and service level; they too found that Croston's method performed well. In an empirical study, Eaves and Kingman (2004) found little difference between exponential smoothing and the bias-adjusted Croston method when using traditional point measures (MAD, RMSE and MAPE). They go on to argue that a better measure is to examine average stock holdings for a given safety stock policy. Their simulation results suggest the bias-adjusted Croston method works slightly better in this context.

Syntetos and Boylan (2005) provide a new method, in the spirit of Croston's approach, which they find to be more accurate at issue points, although the results are inconclusive at other time points.

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

1.4 A comment

An interesting aspect of the empirical work done thus far is the heavy emphasis on point forecasts; the paper by Sani and Kingsman (1997) being a notable exception. 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, one-sided prediction intervals. Indeed, as Chatfield (1992) has pointed out, prediction intervals deserve much greater prominence in forecasting applications. Furthermore, the empirical evidence is consistent with the notion that Croston-type methods may provide more accurate prediction intervals yet 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 level may be in excess of those stated; see Table 1 in section 2.6. Accordingly, we focus upon complete prediction distributions rather than intervals in this paper.

1.5 Remainder of the paper

In section 2 we develop an approach based upon zero-inflated distributions, with particular emphasis upon the zero-inflated Poisson distribution (ZIP), which enables explicit evaluation of the lead-time demand distribution. Then, in section 3, we summarize the different models that will be considered in the empirical analysis. Since our particular focus is on the ability of a model to predict the entire predictive distribution, and not just to provide point forecasts, we examine suitable performance criteria in section 4. Issues relating to model selection are briefly examined in section 5. Finally, in section 6 we present an empirical study using data on monthly demand for 1,046 automobile parts.

2 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 follow a different direction in two respects: first, we will retain the idea that inventory is subject to periodic review, implying a discrete time series. Second, we wish to allow for lumpy demand so that the number of units demanded will be relatively small, but may exceed one.

We assume that the event of orders being placed in any time period follows a Bernoulli process with $P(\text{orders}) = q$ and $P(\text{no order}) = p,\ p+q=1$. Although we believe that the value of p is likely to decrease initially and then increase later in the lifetime of a product, the empirical evidence suggests that a constant value for p is often a convenient approximation (Snyder, 2002). Further, we assume that the total size of the orders in a given time period follows some discrete distribution with probabilities defined as $\{P(\text{order size } = y) = Q_y,\ y = 0,1,\ldots\}$. Note that the probability of an order of size zero is included, which may seem redundant. The probability of a zero order may be set to zero by left truncation of the distribution if needed, but for theoretical developments it is somewhat more convenient to maintain the present formulation, as illustrated by the examples in section 2.3.

2.1 The basic model

Let Y denote the size of order placed in a given time so that:

$$P\{Y=0\} = p + qQ_0$$
$$P\{Y=y\} = qQ_y,\ y \ge 1$$

Such distributions are known as *zero-inflated.* Define the probability generating function (PGF) for Y as:

$$G(z)=\sum_{y=0}^{\infty} z^{y} P(Y=y)=p+qQ(z)$$

where $Q(z)=\sum_{y=0}^{\infty} z^{y} Q_{y}$. It follows directly that when the mean and variance of the order size distribution are $\{\mu, \omega\}$, the mean and variance of Y are given by:

$$E(Y)=q\mu$$
$$V(Y)=q\omega+pq\mu^{2}$$

Thus, the variance to mean ratio for Y is:

$$\frac{V(Y)}{E(Y)}=\frac{\omega}{\mu}+p\mu>\frac{\omega}{\mu}$$

so that this ratio is greater than that for the order size distribution, but may be less than or greater than one depending on the original choice of distribution.

2.2 Lead-time demand

A periodic review system needs to accommodate lead times for delivery that may be much longer that the review period itself. The PGF for total demand during a lead time *h*, denoted by $H_{h}(z)$ is:

$$H_{h}(z)=[G(z)]^{h}=[p+qQ(z)]^{h} \tag{1}$$

However, equation (1) is valid only if the order size distribution does not change. Typically, the parameters will evolve over time, as in section 3, and the uncertainty will increase as the number of periods ahead is increased. We use the subscript *j* to denote the jth period during the lead time. Thus, the PGFs for period *j* become $G_{j}(z)$ and $Q_{j}(z)$, and the PGF for the total demand during lead time may be written as:

$$H_h(z) = \prod_{j=1}^{h} G_j(z) = \prod_{j=1}^{h} [p + qQ_j(z)] \tag{2}$$

The effect of the increased uncertainty will be to make distribution (2) somewhat heavier in the tails than (1). Nevertheless, when the order size distribution changes relatively slowly, expression (1) will serve as an adequate approximation to (2). In section 6, we provide empirical results based upon these approximations and contrast them with solutions based upon a simulation approach, which is described in section 3.3.

2.3 Probabilities for lead-time demand

Expression (1) may be represented by the series expansion:

$$H_h(z) = \sum_{j=0}^{h} \binom{h}{j} p^{h-j} q^j [Q(z)]^j \tag{3}$$

If we define $[Q(z)]^j = \sum_{k=0}^{\infty} w_{j,k} z^k$ it follows that the probabilities are:

$$P_k = \sum_{j=0}^{h} \binom{h}{j} p^{h-j} q^j w_{j,k} \tag{4}$$

This expression is not pretty and evaluation of the $\{w\}$ terms becomes tedious. Fortunately, simpler expressions are available in realistic special cases.

Example 1: Poisson demand

If the mean is λ we have the PGF $Q(z) = \exp\{\lambda(z-1)\}$ so that $[Q(z)]^j = \exp\{\lambda j(z-1)\}$ and equation (4) becomes:

$$P_k = \sum_{j=1}^{h} \binom{h}{j} p^{h-j} q^j e^{-\lambda j} \frac{(\lambda j)^k}{k!} \tag{5}$$

When $k=0$ this expression reduces to

$$P_0 = (p + qe^{-\lambda})^h$$

as may be verified directly by putting *z=0* in equation (3).

Example 2: Hurdle Poisson demand with minimum order size equal to one

The notion of making allowances for orders of size zero may seem artificial and an alternative use of the Poisson is to assume that the mean is $\lambda + 1$so that the PGF for order size becomes $Q(z) = z\exp\{\lambda(z-1)\}$; the distribution is known as the Hurdle Poisson (Winklemann, 2008, chapter 6). Going through the algebra in the same way as before, we arrive at the probabilities, setting $u = \min(k,h)$:

$$\begin{aligned} P_k &= \sum_{j=1}^{u} \binom{h}{j} p^{h-j} q^j e^{-\lambda j} \frac{(\lambda j)^{k-j}}{(k-j)!}, \quad k \geq 1 \\ P_0 &= p^h \end{aligned} \tag{6}$$

Example 3: Geometric demand with minimum order size equal to one

The PGF for the geometric distribution with minimum order size equal to one is $Q(z) = \dfrac{z(1-\theta)}{(1-\theta z)}$. The resulting expression for the probabilities becomes:

$$\begin{aligned} P_k &= \sum_{j=1}^{u} \binom{h}{j} p^{h-j} q^j \binom{k-1}{j-1} (1-\theta)^j \theta^{k-j}, \quad k \geq 1 \\ P_0 &= p^h \end{aligned} \tag{7}$$

If greater flexibility is required, the negative binomial distribution may be used in place of the geometric. For the low volume demands we are considering, this extension would not seem to be needed as the modal size of order will almost always be one.

2.4 Limiting distributions

A continuous review system corresponds to letting $h \to \infty$ with $hq = \gamma$ fixed. That is, the binomial distribution in equation (1) is replaced by its Poisson limit. The resulting compound

Poisson distributions are well-known. For example, a Poisson assumption for order size yields the Neyman Type A distribution and the geometric version produces the Polyà-Aeppli distribution (Johnson, Kotz and Kemp, 1992, pp. 368-382). If the logarithmic series distribution (LSD) is used for order size, the limiting form is the negative binomial distribution, but the LSD is not a convenient form to use for periodic review since it does not yield the simplifications illustrated in Examples 1 – 3.

2.5 Variable lead-times

The models we have discussed so far all assume a fixed lead-time for planning purposes. However, lead-time itself may be stochastic. We do not pursue this topic in depth but it is worth noting that if lead-time, h is a Poisson random variable with parameter β, the compound distribution derived from equation (1) has PGF

$$H_*(z) = \exp[\beta\{p + qQ(z)\} - 1]$$

That is, we again have a compound Poisson distribution. The variety of mechanisms that lead to these distributional forms encourages the view that they can provide a successful planning framework for intermittent demand even though we know we cannot hope to capture all aspects of a complex process in a simple model.

2.6 Numerical examples

How do the various factors, such as probability of not receiving an order in a unit time period (p), mean size of order (λ) and duration of lead-time (h) affect the overall probability distribution for Lead Time Demand (LTD)? We carried out a small numerical study for the set-up described in Example 1 and determined the minimum level of inventory required to ensure that the system is in-stock just prior to restocking at least 95 percent of the time; that is we constructed one-sided prediction intervals. The results are summarized in Table 1. It is evident that the level of inventory required for longer lead-times is proportionally much less than that required for a single period.

Table 1: Minimum inventory levels required to achieve at least a 95 percent service

Mean size of order	P(No order)	Horizon h = 1		Horizon h = 4	
		Inventory	Service	Inventory	Service
λ =1.5	0.5	3	.967	7	.958
	0.8	2	.962	4	.951
λ =3.0	0.5	5	.958	13	.960
	0.8	4	.963	8	.960

3. Probability distributions

Applications of this approach require that we focus upon a particular distribution and we now consider the Zero-Inflated Poisson (ZIP) distribution. This distribution and others considered in the empirical study in section 6 are summarized in Table 2. In the table 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 2: Count distributions used in the empirical study

Distribution	**Mass Function**	**Parameter Restrictions**	**Mean**
Poisson	$\frac{\lambda^y}{y!}\exp(-\lambda)$	$\lambda > 0$	λ
Negative Binomial	$\frac{\Gamma(a+y)}{\Gamma(a)y!}\left(\frac{b}{1+b}\right)^a\left(\frac{1}{1+b}\right)^y$	$a > 0, b > 0$	$\frac{a}{b}$
Zero Inflated Poisson	$\begin{cases} p + q\exp(-\lambda) & y = 0 \\ q\lambda^y \exp(-\lambda)/y! & y = 1, 2, \ldots \end{cases}$	$p \geq 0, q > 0, q\lambda > 0$ $p + q = 1$	$q\lambda$

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

In each case, we wish to allow for the possibility that the mean μ of a distribution may change randomly over time. Three possibilities are summarized in Table 3.

Table 3: Recurrence relationships for the mean

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

Note that the undamped dynamic relationship corresponds to the updating relationship used in simple exponential smoothing. The distribution parameter λ (or a) is then determined from the mean using the following conversion formulae.

Distribution	**Conversion Formula**
Poisson	$\lambda_t = \mu_t$
Negative Binomial	$a_t = b\mu_t$
Zero Inflated Poisson	$\lambda_t = \mu_t / q$

We considered two versions of the dynamic negative binomial model; the first case uses the full parametrization as presented in these tables (Nbin unres) whereas in the second (Nbin res) we applied the restriction $\alpha = 1/(1+b)$ to reduce the number of parameters. The reason for the restriction is to explore the effect of making the undamped dynamic version of the negative binomial correspond to the equilibrium version of the Harvey-Fernandes model.

There is one disadvantage in 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 defined on the non-negative integers: the simulated series values stochastically converge to zero (Grunwald et al., 1997) and get trapped there over moderate to long prediction horizons; a behavior which is further examined by Akram et al. (2009). This problem does not occur with the damped stationary models.

3.1 Other models

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

The Harvey-Fernandes model

The Harvey and Fernandes (1989) method is based on a local level state space model with Poisson measurements. It is a Poisson analogue of the Kalman filter. The details are eschewed here, because in the end it reduces to the use of negative binomial distributions for counts with time dependent means given by the finite exponentially weighted average

$$\mu_{t+1} = \frac{y_t + \delta y_{t-1} + \delta^2 y_{t-2} \ldots + \delta^{t-1} y_1}{1 + \delta + \delta^2 + \ldots \delta^{t-1}} \qquad (8)$$

where δ is a parameter called the *discount factor* satisfying the condition $0 \le \delta \le 1$. 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 2) in typical period t, they are calculated recursively with the expressions:

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

where $a_1 = b_1 = 0$.

The modified Croston model

Based on theory from Snyder (2002) and Shenstone and Hyndman (2005), this approach models the time τ between successive periods with transactions as a shifted geometric distribution with a mean that is governed by an undamped local level function. The shifted geometric distribution has probability mass function

$$\Pr\{\tau = t\} = p^{t-1}q \quad t = 1, 2, \ldots \tag{9}$$

Here q is the probability of a transaction in a given period. The mean of this distribution is $\mu^{\tau} = q^{-1}$ so that the sample version of $q = 1/\mu^{\tau}$ provides an estimate of the probability.

It is assumed that in those periods with transactions, the demand Y is governed by a hurdle Poisson distribution

$$\Pr\{Y = y \mid \mu^{+}\} = \frac{(\lambda)^{y-1}}{(y-1)!}\exp(-\lambda) \qquad \mathrm{y} = 1, 2, \ldots \tag{10}$$

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

3.2 Estimation

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

$$p(y_1, \ldots, y_n \mid I_{n-1}) = \prod_{t=1}^{n} p(y_t \mid I_{t-1}) \tag{11}$$

where $I_0 = \{\mu_1, \theta\}$ and $I_{t-1} = \{\mu_t, \theta\}, t = 2, 3, \ldots, n.$ The univariate distributions in this decomposition are a succession of one-step-ahead prediction distributions. In the case of the global level 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 these univariate distributions.

The Harvey-Fernandes model (HF)

In this case the initial mean μ_1 is not needed. Successive means are calculated with equation (8) and the term corresponding to $t=1$ in (11) is dropped.

The modified Croston model

Given that the processes governing transaction sizes and the time gaps between transactions are independent it might be thought that the estimation of their parameters should be divided into two separate maximum likelihood problems. However, suppose that there are m transactions over the n periods. The effective sample sizes in this 'dual' approach are m and $m-1$ respectively. Since m is usually much smaller than the sample size n, it transpires that the estimates that result from this 'dual' approach are often quite poor. An integrated maximum likelihood approach is needed to obtain results based on the full sample of size n.

Suppose that at period t, τ_t designates the number of periods since the previous transaction. Let p_j and q_j designate probability parameters satisfy the condition $p_j+q_j=1$ and let α be a parameter lying in the unit interval $\left[0,1\right]$. The component of the likelihood function for typical period t is determined as follows:

Case 1 ($y_t=0$)	Case 2 ($y_t>0$)
$\mathscr{L}_t = p_t$	$\mathscr{L}_t = q_t \Pr\left\{y_t \mid \mu_t^+\right\}$
$\mu_{t+1}^+ = \mu_t^+$	$\mu_{t+1}^+ = \left(1-\alpha\right)\mu_t^+ + \alpha y_t$
$\mu_{t+1}^\tau = \mu_t^\tau$	$\mu_{t+1}^\tau = \left(1-\alpha\right)\mu_t^\tau + \alpha \tau_t$
$\tau_{t+1} = \tau_t + 1$	$\tau_{t+1} = 1$

In both cases $q_{t+1} = 1/\mu_{t+1}^\tau$ and $p_{t+1} = 1 - q_{t+1}$.

The likelihood function is formed from the product of the $\mathscr{L}_t$. It depends on μ_1^+, μ_1^τ and α

3.3 Prediction and the prediction likelihood

A mix of analytical and simulation methods are used to obtain the prediction distributions. For static models, the analytical methods of section 2.3 provide exact results for the total lead time demand but they underestimate the variability for dynamic models. Given that the models involve first-order recurrence relationships, the joint distribution may be decomposed into a product of univariate one-step-ahead prediction distributions. In the simulation approach future series values are then generated from each future one-step-ahead distribution in succession.

Each model is fitted using the estimation sample. The future one-step-ahead prediction distributions are then estimated using 100,000 simulations. The one-step-ahead prediction distributions are used in conjunction with the actual test series values to calculate the associated probabilities. The product of these probabilities is the prediction likelihood required for the calculation of the CPA (defined in section 4.2).

4. Prediction Performance measures

Most measures for evaluating the performance of various prediction methods examine only point forecasts. One criterion in common use is the *mean absolute percentage error*, defined as

$$\text{MAPE} = \frac{100}{h}\sum_{j=1}^{h}\left|\frac{\hat{y}_n(j) - y_{n+j}}{y_{n+j}}\right| \tag{12}$$

where $\hat{y}_n(j)$ designates the prediction made at origin n of the series value y_{n+j} and h is the prediction horizon. It fails for low count data whenever the value of zero is encountered in the series. One way to circumvent this issue is to rely on the *mean absolute scaled error* (Hyndman and Koehler, 2006) defined as

$$\text{MASE} = \frac{1}{h}\sum_{j=1}^{h}\left|y_{n+j} - \hat{y}_n(j)\right| \Big/ \frac{1}{n}\sum_{t=1}^{n}\left|y_t - y_{t-1}\right|. \tag{13}$$

This measure is used in our study, but like the MAPE, it evaluates only point forecasts. In section 6, the empirical study is based upon $h = 6$; that is, lead times 1 – 6 are employed in the calculations.

Although the MASE and similar measures are useful in determining the performance of point forecasting methods, it does not provide any information regarding predictive distributions. In the remainder of this section we describe two criteria that can be used to measure forecasting performance relative to the predictive distribution. Typically such measures might 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 overall performance for a group of series, as in section 6.

4.1 Comparative Prediction Advantage (CPA)

The joint prediction distribution $p(y_{n+1},\ldots,y_{n+h} \mid I_n)$ summarizes *all* the characteristics of a future series including central tendency, variability, autocorrelation, skewness and kurtosis. Here I_n consists of all quantities that inform the calculation of these probabilities including the sample $y_1,\ldots,y_n$, the parameters, and the states of the process at the end of period n. Assuming that we withhold the series values $y_{n+1},\ldots y_{n+h}$ for evaluation purposes, $p(y_{n+1},\ldots,y_{n+h} \mid I_n)$ is a measure of the *likelihood* that these values come from the model under consideration. The mode of this joint distribution may be considered to be the point predictions. Series values $y_{n+1},\ldots y_{n+h}$ that are more remote from the mode have lower probabilities. Thus, we suggest that this probability can be interpreted as the *likelihood of the model under consideration*.

The log-likelihood is a standard measure of information. Thus, when there are two competing models M_1 and M_2, and their prediction likelihoods are designated by $\mathscr{L}_1$ and $\mathscr{L}_2$ respectively, the quantity

$$\mathscr{R}_{1,2} = 100(\log \mathscr{L}_1 - \log \mathscr{L}_2) \tag{14}$$

measures the percentage difference in their prediction likelihoods. Thus, we suggest that M_1 can be considered a better predictor of $y_{n+1},\ldots y_{n+h}$ than M_2 if $\mathscr{R}_{1,2}$ is positive. We call this the *comparative prediction advantage* (CPA) of M_1 over M_2.

It is convenient to use one model, designated by M_0, as a benchmark. The prediction performance of all other models can be calculated relative to this benchmark. Thus, $\mathscr{R}_{i,0}$ measures the comparative prediction advantage of a model M_i over the benchmark. This quantity can be calculated for all the models under consideration. Then M_i can be compared with M_j using the percentage difference in their prediction likelihoods derived with the relationship

$$\mathscr{R}_{i,j} = \mathscr{R}_{i,0} - \mathscr{R}_{j,0} \,. \tag{15}$$

In our study we will use the Poisson static model as the benchmark.

The CPA could be evaluated in several ways. We consider the joint distribution of $\{y_{n+1},\ldots,y_{n+h}\}$ given the information up to and including period n, namely I_n. Applying the same logic as in the derivation of equation (11), it can be established that

$$p\left(y_{n+1},\ldots,y_{n+h} \mid I_n\right) = \prod_{t=n+1}^{n+h} p\left(y_t \mid I_{t-1}^{*}\right). \tag{16}$$

where $I_t^{*} = \{I_n, y_n, \ldots, y_{t-1}\}$; 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 typical 'future' period t as seen from the *beginning* of this period with 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 3. Since interest may focus upon forecasts for the demand over lead-time as well as one-step-ahead, we also examine the CPA for the sum over the lead-time $S_n(h) = y_{n+1} + \ldots + y_{n+h}$ with prediction distribution $p(S_n(h) | I_n)$. The following simple example illustrates why it is important to use a measure such as CPA rather than one that focuses exclusively on point forecasts.

Example 4: Model choice using CPA

Consider two competing static Gaussian forecasting models (M_1 and M_2) with common mean μ and their variances, from the estimation sample, are estimated as $V_1 < V_2$. (For example, the two models might correspond to estimation with and without removal of outliers). The first reaction is to assume that the forecasting model with the smaller sample variance is better. However, a comparison of the prediction likelihoods using the CPA reveals that we should choose M_1 only if the hold-out sample mean square error is sufficiently close to V_1. In this simple case, both procedures would give rise to the same sum of squared errors in the hold-out sample of size *h*, which we denote by $S = hV$. From equation(14) the CPA becomes:

$$\mathscr{R}_{1,2} = 100h\left(\log V_2 - \log V_1 + \frac{V}{V_2} - \frac{V}{V_1} \right)$$

It may be shown that this expression is positive at $V = V_1$ and negative at $V = V_2$ so that the choice of model depends upon the prediction distribution.

4.2 Rank Probability Score (RPS)

The rank probability score was introduced in Epstein (1969) and Murphy (1971). It uses the L_2 -norm to measure the distance between two distributions. It is simplest to develop the theory in terms of continuous random variables, but the results obtained also apply with obvious modifications to count random variables. The details are provided in the Appendix.

When F is discrete and we consider a single observation x, the RPS distance function becomes

$$L_2(x,F) = \sum_{y=0}^{\infty} \left(\delta(y \ge x) - F(y) \right)^2$$

where $\delta(y \ge x) = 1$ if $y \ge x$ and 0 otherwise. In our calculations the infinite sum was truncated at $y = 100$.

In section 6, we calculate the RPS for each one-step-ahead forecast relating to the withheld sample and then average over the $h(=6)$ values; the procedure follows the same logic as for CPA. The RPS 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 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 performance of the competing model. Despite the popularity of prediction validation (e.g. Makridakis and Hibon, 2000), Billah et al. (2006) found that method to be generally inferior to other methods for point forecasting, particularly those based upon information criteria. This conclusion is unaffected by the particular choice of 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 overall performance across multiple series when we use criteria such as CPA and RPS to evaluate the prediction distributions. Such comparisons are common in forecasting competitions and are useful when a decision must be made on a general approach too forecasting a group of series such as a set of SKUs. Accordingly, the comparisons in the next section are made using the criteria discussed in section 4 to compare overall model performance.

6. An empirical study of auto parts demand

The empirical study had two components: the comparison of the different models proposed for intermittent demand forecasting, and the evaluation of the analytical approach developed in section 2. The analytical approach provides exact results for the one-step-ahead static case for the Poisson, negative binomial and zero-inflated Poisson distributions (assuming the model truly represents the data generating process). However, it is only an approximation for the total lead-time-demand since equation (1) ignores the increased variability in forecasts made multiple steps ahead[3]. Thus, all other cases were evaluated using simulation. The performance of the different approaches was evaluated using each of MASE, CPA and RPS.

We examined the performance of the Poisson, negative binomial [NBin, both unrestricted and restricted parameter versions] and zero-inflated Poisson [ZIP] distributions, using each of the parameter specifications described in Table 3. For purposes of comparison, we also examined the Harvey-Fernandes [HF] and modified Croston [Croston] models described in section 3.1. Finally, as a benchmark we included the simplistic all-zeros [Zero] forecast, which ignores the data and simply forecasts zero demand for all periods. On encountering a non-zero observation in a withheld series the associated CPA is assumed to be minus infinity because such an event is deemed to be impossible in the zero-forecast model. The dynamic forms of these different approaches are evaluated using the simulation method described in section 3.3.

The study used data on slow-moving parts for a US automobile company; these data were previously discussed in Hyndman et al. (2008, pp. 283-286). The data set consists of 2,674 monthly series of which 2,509 had complete records. Restricting attention to those series with

[3] A variance correction could be applied using the results in Snyder et al. (1999) but the resulting distribution would no longer be based upon the ZIP.

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. As noted earlier, 1,046 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 series for which the average time lapse was 2.5, with an average variance-to-mean ratio of 1.9. The data cover a period of 51 months; 45 observations were used for estimation and 6 were withheld for comparing forecasting performance one to six steps ahead.

6.1 Comparison of the different models

The results from the analytical calculations and the simulations are summarized in Table 4. It should be kept in mind that smaller values of RPS and MASE are preferable, whereas larger values of CPA are better. The results in the table represent the means for these measures, taken across the 1,046 series; the figures in the table represent averages across series (and across realizations for the multiple-step case). The results for the CPA of lead-time demand use a trimmed mean to avoid distortions from a few extreme cases. The results for the medians are very similar and are not reported here. Three cases are presented:

- one-step-ahead forecasts
- multiple-step-ahead forecasts based upon the recursions implicit in equation (11)
- forecasts for lead-time demand, based upon $h = 6$.

The two versions of the negative binomial (Nbin unres and Nbin res) refer to the full and restricted parameter sets summarized in section 3. The analytical results are exact for one-step-ahead, so the simulations do not provide any new insights. These simulation results differ from the analytical forms only very slightly and are omitted.

Table 4: Comparison of the forecasting performance of different methods for 1,046 US automobile parts series, based upon both the analytical approach and the simulation study [best cases in bold]

Predictions	one-step			multiple-step		lead time demand		
Criterion	CPA	RPS	MASE	RPS	MASE	CPA*	RPS	MASE
Static Models (Analytical Approach)								
Poisson	0.00	0.46	0.82	0.46	0.82	0.00	0.40	0.54
ZIP	13.29	0.41	0.82	0.41	0.82	8.37	0.37	0.54
Nbin	13.80	0.40	0.82	0.40	0.82	8.79	0.36	0.54
Damped Dynamic Models								
Poisson	10.93	0.39	0.70	0.39	0.72	10.82	0.29	0.42
ZIP	18.81	0.36	0.71	0.37	0.76	14.51	0.28	0.46
Nbin unres	**20.48**	**0.35**	0.70	0.36	0.72	**15.30**	0.27	0.42
Nbin res	18.25	0.37	0.74	0.41	0.90	14.37	0.32	0.62
Undamped Dynamic Models								
Poisson	10.16	0.38	0.68	0.38	0.68	10.93	0.27	0.39
ZIP	17.96	0.36	0.68	0.36	0.69	14.93	0.27	0.40
Nbin unres	20.11	**0.35**	0.68	**0.35**	0.69	15.19	**0.26**	0.40
Nbin res	15.11	0.36	0.65	0.37	0.66	13.37	**0.26**	**0.38**
Others								
Croston	15.76	0.38	0.75	0.38	0.76	12.22	0.30	0.46
Harvey-F	15.97	0.36	0.64	0.36	0.65	14.32	**0.26**	**0.38**
Zeros	-inf	0.41	**0.41**	0.41	**0.41**	-inf	0.41	0.41

* *The CPA values are averaged using a 2% trimmed mean to avoid a small number of extreme values.*

The following comments are in order:

1. The prediction distributions for the one-step-ahead forecasts based on the dynamic unrestricted negative binomial (Nbin unres) are consistently better than those generated by other methods, followed by the zero-inflated Poisson (ZIP).This result is not surprising, given the high proportion of periods with zero demands, but rather serves to reflect the inadequacy of criteria based only upon point forecasts.

2. The dynamic models outperform the static models according to both CPA and RPS. Thus the traditional use of the static Poisson distribution for slow-moving items should be superseded by a dynamic version. However, even the static Nbin and ZIP represent a considerable improvement over the Poisson.

3. The results for lead-time demand are less pronounced, reflecting the smaller proportions of zeros encountered after the aggregation of several time periods.

4. The study demonstrates the limitations of measures based upon point forecasts when we are interested in the prediction distribution.

    a. The zero prediction method has the lowest MASE simply because such a large proportion of the observations are zero.

    b. MASE for the static Poisson and negative binomial distributions are the same since they simply reflect the sample mean.

    c. The Harvey- Fernandes method and its close cousin, undamped dynamic Nbin res, have the next lowest MASE, in part because they generate more forecasts that are closer to zero. However, these methods also perform well on other criteria, as seen from Table 4.

5. The Croston method did not perform as well as many other methods. We attribute this to the fact that it uses the same smoothing parameter for smoothing both the positive demands and the time gaps. This acts as an artificial constraint because, as shown in Snyder (2002), the maximum likelihood estimate of the time gap smoothing parameter in the two-parameter adaptation of the Croston method was often zero, suggesting that something akin to the zero-inflated Poisson distribution is more appropriate.

6. Given that the undamped models involve fewer parameters, which would be an advantage for shorter series, we are inclined to conclude that the undamped models should be chosen. This is done reluctantly, because we feel that the CPA has a somewhat more compelling logic than the other criteria and the undamped schemes are subject to the stochastic convergence problem noted in section 3.

7. There are a number of series, mostly excluded for the sample of 1,046 SKUs used here, for which demand is very low, perhaps 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 or holding zero inventory and submitting special orders as needed.

8. The analytical results provide reasonable approximations when the lead time is relatively short or the demand process is relatively stable (low value of α in the updating expressions in Table 3). When dynamic models provide an improved description and the lead time is relatively long, the simulation approach should be adopted.

Summary

We developed a forecasting framework for SKUs that have low volume, intermittent demands. The models considered included the negative binomial and the zero-inflated Poisson distributions. These were evaluated for their capacity to predict demand in individual future periods and aggregate demand over a specified lead time. The static versions provided an improvement over the traditional static Poisson distribution, but the dynamic versions offered further improvements. Using a data set on automobile spare parts, we conducted an empirical comparison of forecasting performance with other extant methods. The results indicated that these schemes perform well and that they are likely to provide an effective framework for inventory planning in other settings.

References

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

Al-Osh M.A. and 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. and 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.

Cox, D.R. (1981) Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics*, 8, 93-115.

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

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

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

Durbin, J. and Koopman, S.J. (2001) *Time Series Analysis by State Space Methods*. Oxford University Press: Oxford.

Eaves, A.H.C. and 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. and Snyder, R.D. (2008) Feasible parameter regions for alternative discrete state space models. *Statistics and Probability Letters* 78 2963-2970.

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

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

Harvey, A.C. and 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. and 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 and Snyder, R.D. (2008) *Forecasting with Exponential Smoothing: The State Space Approach*. Springer-Verlag: Berlin and Heidelberg.

Johnson, N.L., Kotz, S. and Kemp, A.W. (1992) *Univariate Discrete Distributions.* Wiley: New York. Second edition.

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

Johnston, F.R. and 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. and Leisenfeld, R. (2006) Time series of count data: modeling, estimation and diagnostics. *Computational Statistics and Data Analysis*, 51, 2350-2364.

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

McCabe, B.P.M. and 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. and Snyder, R.D. (1997) Estimation and prediction of a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association*, 92, 1621-1629.

Rao, A.V. (1973) A comment on: Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, 24, 639-640.

Sani, B. and Kingsman, B.G. (1997) Selecting the best periodic inventory control and demand forecasting methods for low demand items. *Journal of the Operational Research Society*, 48, 700-713.

Shenstone, L. and 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. and 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.

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

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

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

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

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.

West, M. and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models*, second edition, Springer-Verlag, New York.

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

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

Winklemann, R. (2008) *Econometric Analysis of Count Data*, 5th edition, Springer-Verlag, New York.

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

Appendix: Properties of the Rank Probability Score

The rank probability score (RPS) is the distance between distribution functions F and G, defined by

$$L_2(G,F)=\int (G(\xi)-F(\xi))^2\,d\xi\,. \tag{17}$$

Suppose F_n represents an empirical distribution formed from a sample of size n governed by the distribution F. The distance between the two distributions is given by

$$L_2(F_n,F)=\int (F_n(\xi)-F(\xi))^2\,d\xi \tag{18}$$

In the special case where $n=1$,

$$F_1(\xi)=\begin{cases}0 & \xi<x\\ 1 & \xi\geq x\end{cases}$$

where x is the observed value of a random variable X. The RPS becomes

$$L_2(x,F)=\int_{-\infty}^{x} F^2(\xi)d\xi+\int_{x}^{\infty}(1-F(\xi))^2\,d\xi$$

This L_2 norm is a function of the observed value x. This function is convex with a minimum corresponding to the median of the distribution. More specifically,

$$\frac{dL_2}{dx}=F^2(x)-(1-F(x))^2=2F(x)-1.$$

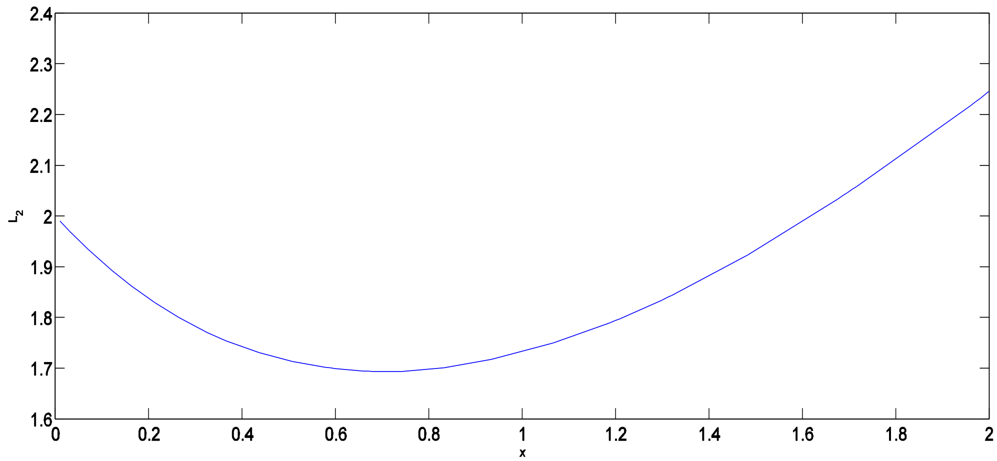
A turning point occurs where $F(x)=\frac{1}{2}$. Moreover, letting f designate the density function, the second derivative is

$$\frac{dL_2^2}{d^2x} = 2f(x).$$

This quantity is non-negative so the turning point is a minimum.

Consider, for example the case of an exponential distribution with a mean of 1. The distribution function is $F(\xi) = 1 - \exp(-\xi)$. It may be shown that $L_2(x, F) = x + 2\exp(-x)$, a relationship which is plotted in Figure 1. The median is found by solving the equation $\exp(-x) = 1/2$, or equivalently $x = -\log(1/2) = 0.693$.

Figure 1: Plot of the RPS for the exponential distribution



If F corresponds to a prediction distribution for a particular period, its median can be interpreted as the point prediction a which minimizes the mean absolute prediction error $\int |x - a| f(x)dx$. In a sense the RPS transforms a prediction error (the deviation of the observed value from the predicted value) so that the shape of the prediction distribution is taken into account. When the distribution is asymmetric, two points on either side of, but equidistant from, the median have different RPSs, the size of which reflects the skewness of the distribution.

In general the interest may be in a succession of future periods, in which case we simply find the average of the RPSs associated with each period. The following example serves to illustrate the computational procedure with a fitted static Poisson model which has $\hat{\lambda} = 0.5$ on 6 withheld observations (0,0,0,1,0,2). For each observation x in the future horizon, the formula for the RPS is

$$\text{RPS}(x) = \sum_{y=0}^{5} \left(\delta(y \geq x) - F(y)\right)^2$$

where 5 is the truncation level for the sum. The calculations are shown in Table 5 where $p(y)$ designates the probability mass function.

Table 5 Calculation of the RPS for $x = 2$.

y	0	1	2	3	4	5
$p(y)$	0.6065	0.3033	0.0758	0.0126	0.0016	0.0002
$F(y)$	0.6065	0.9098	0.9856	0.9982	0.9998	1.0000
$\delta(y \geq 2)$	0	0	1	1	1	1
$\delta - F(y)$	-0.6065	-0.9098	0.0144	0.0018	0.0002	0.0000
$(\delta - F(y))^2$	0.3679	0.8277	0.0002	0.0000	0.0000	0.0000

The RPS at $x = 2$ is the sum of the entries in the last row of the table and is 1.1958. Similarly, it can be shown that $\text{RPS}(0) = 0.1632$ and $\text{RPS}(1) = 0.3762$. The RPSs are then averaged Recognizing that there are 4 zeros, one 1 and one 2 the calculation is

$$\begin{aligned}\text{avg RPS} &= \frac{4 \times \text{RPS}(0) + \text{RPS}(1) + \text{RPS}(2)}{6} \\ &= \frac{4 \times 0.1632 + 0.3762 + 1.1958}{6} \\ &= 0.371\end{aligned}$$