Stochastic models underlying Croston's method for intermittent demand forecasting

Lydia Shenstone[1], Rob J. Hyndman[1] 3 February 2003

Abstract: Croston's method is a widely used to predict inventory demand when it is intermittent. However, it is an ad hoc method with no properly formulated underlying stochastic model. In this paper, we explore possible models underlying Croston's method and three related methods, and we show that any underlying model will be inconsistent with the properties of intermittent demand data. However, we find that the point forecasts and prediction intervals based on such underlying models may still be useful. [JEL: C53, C22, C51]

Keywords: Croston's method, exponential smoothing, forecasting, intermittent demand.

1 Introduction

Inventories with intermittent demands are quite widespread in practice. Data for such items consist of time series of non-negative integer values where some values are zero. We shall denote the historical demand series $Y_1, Y_2, \ldots, Y_n$ and assume these take non-negative integer values.

Croston's (1972) method is the most widely used approach for intermittent demand forecasting (IDF), and involves separate simple exponential smoothing (SES) forecasts on the size of a demand and the time period between demands. Other authors, including Johnston & Boylan (1996) and Syntetos & Boylan (2001), have suggested a few modifications to Croston's method that can provide improved forecast accuracy. One such modification is to apply Croston's method to the logarithms of the demand data and to the logarithms of the inter-demand time.

However, all of these methods provide only point forecasts and are not based on a stochastic model. In fact, no underlying model for Croston's method has ever been properly formulated. Consequently, there are no forecast distributions and prediction intervals associated with forecasts obtained using these methods.

This paper aims to identify stochastic models that underly Croston's method and related meth-

[1]Department of Econometrics and Business Statistics, Monash University, Clayton, VIC 3800, Australia.
Correspondence to: Lydia.Shenstone@buseco.monash.edu.au

ods, and hence obtain forecast distributions and other properties. However, we end up showing that the models that might be considered as underlying Croston's and related methods are inconsistent with the properties of intermittent demand data. In particular, the possible models underlying Croston's and related methods must be non-stationary and defined on a continuous sample space. For Croston's original method, the sample space for the underlying model includes negative values. This is inconsistent with reality that demand is always non-negative.

This does not mean that Croston's method itself is not useful. Its long history of use shows that many people find the point forecasts obtained in this way to be satisfactory. Indeed, several studies have demonstrated that it gives superior point forecasts to some competing methods (e.g., Willemain et al. 1994). Furthermore, we show how the underlying stochastic models can be used to construct prediction intervals which are helpful in calculating appropriate levels of safety stock for inventory demand.

In Section 2, we discuss Croston's method and its potential underlying models. Section 3 describes three additional models that are related to modifications of Croston's method. We present in Section 4 the model properties such as forecast means and variances, forecast distributions and lead-time demand distributions. Prediction intervals are discussed in Section 5 and we conclude in Section 6 by comparing the results and considering alternative approaches. Proofs of the main results are provided in the Appendix.

2 Croston's Method

Let Y_t be the demand occurring during the time period t and X_t be the indicator variable for non-zero demand periods; i.e., $X_t = 1$ when demand occurs at time period t and $X_t = 0$ when no demand occurs. Furthermore, let j_t be number of periods with nonzero demand during interval $[0,\ t]$ such that $j_t = \sum_{i=1}^{t} X_i$, i.e., j_t is the index of the the non-zero demand. For ease of notation, we will usually drop the subscript t on j. Then we let Y_j^* represent the size of the jth non-zero demand and Q_j the inter-arrival time between Y_{j-1}^* and Y_j^*. Using this notation, we can write $Y_t = X_t Y_j^*$.

Croston's (1972) method separately forecasts the non-zero demand size and the inter-arrival time between successive demands using simple exponential smoothing (SES), with forecasts being updated only after demand occurrences. Let Z_j and P_j be the forecasts of the $(j+1)$th demand size and inter-arrival time respectively, based on data up to demand j. Then Croston's method gives

$$Z_j = (1-\alpha)Z_{j-1} + \alpha Y_j^*, \tag{2.1}$$

$$P_j = (1-\alpha)P_{j-1} + \alpha Q_j. \tag{2.2}$$

The smoothing parameter α takes values between 0 and 1 and is assumed to be the same for both Y_j^* and Q_j. Let $\ell = j_n$ denote the last period of demand. Then the mean demand rate, which is used as the h-step ahead forecast for the demand at time $n + h$, is estimated by the ratio

$$\hat{Y}_{n+h} = Z_\ell / P_\ell. \tag{2.3}$$

Several variations on this procedure have been proposed including Johnston & Boylan (1996) and Syntetos & Boylan (2001).

Croston (1972) stated that the assumptions of this method were (1) the distribution of non-zero demand sizes Y_j^* is iid normal; (2) the distribution of inter-arrival times Q_j is iid Geometric; and (3) demand sizes Y_j^* and inter-arrival times Q_j are mutually independent. These assumptions are clearly incorrect, as the assumption of iid data would result in using the simple mean as the forecast, rather than SES, for both processes. Nevertheless, much of the published empirical analyses of Croston's method have been based on the same assumptions (e.g., Willemain et al. 1994, Syntetos & Boylan 2001).

One goal of this paper is to discuss what assumptions could lead to Croston's method of forecasting. Specifically, is there a model that would lead to forecasts Z_j and P_j as specified in (2.1) and (2.2), and what would the properties of such a model be? We have already seen that such models must be autocorrelated; are there other properties that can be determined?

Note that (2.1) can be rewritten as an exponentially weighted average of past values:

$$Z_j = \sum_{k=0}^{j-1} \alpha(1-\alpha)^k Y_{j-k}^* + (1-\alpha)^j Z_0. \tag{2.4}$$

A similar equation can be obtained for P_j. This immediately means that the underlying models must be non-stationary (e.g., Abraham & Ledolter 1983, Section 3.3).

Now, if the sample space of a model defined as an exponentially weighted moving average is bounded to any subset of $[0, \infty]$, (e.g., taking only positive values or integers), Grunwald et al. (1997) show that the original process will converge almost surely to a constant. Thus, the models underlying Croston's forecasts must assume continuous data with sample space including negative values. This is clearly problematic when we have intermittent demand data that is always integer-valued and non-negative.

Thus, any models underlying Croston's method should be based on assumptions that the process is autocorrelated, non-stationary and has a continuous sample space including negative values. However, we are unable to identify a *unique* underlying model. Chatfield et al. (2001) summarizes a general class of state space models for which SES provides optimal forecasts.

Nevertheless, this general class of models does not cover all the possible models leading to SES forecasts.

The ARIMA(0,1,1) process is a special case of this class and is often used as the underlying model for SES (Box et al. 1994). Because it is the most studied model underlying SES forecasting, we will use the Gaussian ARIMA(0,1,1) model in our empirical comparisons involving Croston's method. That is, we will assume $Y_j^* \sim \text{ARIMA}(0,1,1)$ and $Q_j \sim \text{ARIMA}(0,1,1)$ where Y_j^* and Q_j are independent. The state-space representation of this model is

$$\begin{aligned} Y_j^* &= Z_{j-1} + e_j, \\ Z_j &= Z_{j-1} + \alpha e_j, \\ Q_j &= P_{j-1} + \varepsilon_j, \\ \text{and} \qquad P_j &= P_{j-1} + \alpha \varepsilon_j, \end{aligned} \tag{2.5}$$

where $e_j \overset{\text{iid}}{\sim} \text{N}(0, \sigma_e^2)$ and $\varepsilon_j \overset{\text{iid}}{\sim} \text{N}(0, \sigma_\varepsilon^2)$. We shall refer to this as the "Croston model". In practice, when we use this model for simulating data (as in Section 5), we round the interarrival times to the next highest positive integer. In our analytical results for this model, we ignore this complication.

From (2.5) we note that

$$\text{E}(Y_{j+h}^* \mid Y_1^*, \dots, Y_j^*, Z_0) = Z_j \qquad \text{and} \qquad \text{E}(Q_{j+h} \mid Q_1, \dots, Q_j, P_0) = P_j$$

and so this model does give the required forecasts (2.1) and (2.2), although the mean of the forecast distribution $Y_{n+h} \mid [Y_1, \dots, Y_n, Z_0, P_0]$ is not given by (2.3). Note that the initial states Z_0 and P_0 have negligible effect provided $0 < \alpha < 2$.

3 Modifications of Croston's Method

One simple modification to Croston's method is to use log transformations of both demands and interarrival times to restrict the sample space of the underlying model to be positive. Of course, the underlying models are still defined on a continuous sample space, but they may provide a reasonable approximation to the data. The assumed underlying model involves two independent ARIMA(0,1,1) processes:

$$\begin{aligned} \log(Y_j^*) &\sim \text{ARIMA}(0,1,1), \\ \log(Q_j) &\sim \text{ARIMA}(0,1,1). \end{aligned} \tag{3.1}$$

Again, in simulations with this model we need to round the interarrival times to the next highest positive integer.

Both the Croston model and the log-Croston model (3.1) assume nonstationary interarrival times whereas we find that in practice the interarrival times often appear stationary and uncorrelated. In addition, we will see in the next section that the nonstationary interarrival time also makes it more difficult to explore the properties of the model. Thus, it is reasonable and necessary to assume independent interarrival times.

Snyder (2002) proposed two methods in which the interarrival times are assumed to have an iid Geometric distribution. (Note that this is consistent with Croston's stated assumptions.) The modified Croston model is specified as

$$\begin{aligned} Y_j^* &\sim \text{ARIMA}(0,1,1), \\ Q_j &\overset{\text{iid}}{\sim} \text{Geometric}(p), \end{aligned} \tag{3.2}$$

where p is the mean interarrival time of the demand series. The iid Geometric distribution of Q_j indicates that the probability of demand occurring at each time period is $1/p$, i.e.,

$$X_t \overset{\text{iid}}{\sim} \text{Binomial}(1, 1/p). \tag{3.3}$$

This is useful in calculating the forecast distribution and the corresponding means and variances. The other model, the modified log-Croston model, is similar but uses logarithms of the demands:

$$\begin{aligned} \log(Y_j^*) &\sim \text{ARIMA}(0,1,1), \\ Q_j &\overset{\text{iid}}{\sim} \text{Geometric}(p). \end{aligned} \tag{3.4}$$

The distribution of X_t in (3.3) also holds for the modified log-Croston model.

4 Model Properties

Before discussing the model properties, we shall introduce some distributions that will be useful. First, the Binomial-Normal distribution is constructed from a binomial distribution and several normal distributions. If $X \sim \text{Binomial}(n, q)$ and $Y \mid X \sim \text{N}(m_X, \sigma_X^2)$, then Y has the Binomial-Normal distribution denoted by BN$(n, q, \boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ where $\boldsymbol{\mu} = (m_0, \ldots, m_n)$ and $\boldsymbol{\sigma}^2 = (\sigma_0^2, \ldots, \sigma_n^2)$. Its distribution function is given by

$$\Pr(Y \le y) = \sum_{i=0}^{n} \binom{n}{i} q^i (1-q)^{n-i} \Phi[(y - m_i)/\sigma_i]$$

where $\Phi(\cdot)$ is the standard normal distribution function. Similarly, the Binomial-LogNormal distribution is denoted by BLN$(n, q, \boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ with distribution function

$$\Pr(Y \le y) = \sum_{i=0}^{n} \binom{n}{i} q^i (1-q)^{n-i} \Phi[(e^y - m_i)/\sigma_i].$$

We now consider the properties of the four IDF models introduced in the previous sections. The forecasting methods corresponding to these four models, except that for the log-Croston model, have been used and discussed by various authors. However, none of these underlying models have been studied and explored for their theoretical properties.

For each model, we are interested in exploring properties of the h-step ahead future demand, Y_{n+h}, for any positive integer h, given the historical demand series $Y_1, Y_2, \dots, Y_n$ and the initial states P_0 and Z_0. For each model, provided $0 < \alpha < 2$ these initial states will have asymptotically zero effect.

We shall denote the forecast mean by $m_h = \mathrm{E}(Y_{n+h} \mid \mathcal{I})$ and the forecast variance by $v_h = \mathrm{Var}(Y_{n+h} \mid \mathcal{I})$ where $\mathcal{I} = (Y_1, \dots, Y_n, Z_0, P_0)$. As above, we will assume the model parameters are such that the effect of the initial states Z_0 and P_0 is negligible. Furthermore, we use $Y_n(h)$ to denote the h-step ahead lead-time demand

$$Y_n(h) = Y_{n+1} + \cdots + Y_{n+h}$$

with forecast mean $M_h = \mathrm{E}(Y_n(h) \mid \mathcal{I})$ and forecast variance $V_h = \mathrm{Var}(Y_n(h) \mid \mathcal{I})$.

Due to the nonstationary interarrival time assumed by the model, it is difficult to derive the properties of the Croston model and the log-Croston model. Only the one-step ahead forecast properties for these two models are obtained (they are not in a simple form, however) and presented here. We start with the two simpler models, the modified Croston and log-Croston models, since their properties are easier to analysis.

For each model, we use Z_j and P_j to represent respectively the underlying states corresponding to the (log) demand size and the (log) interarrival time. We also let $\ell = j_n$ denote the last period of demand. Proofs of the following results are given in the Appendix.

Theorem 4.1 *The h-step forecast distribution for model (3.2) is*

$$Y_{n+h} \mid \mathcal{I} \sim \begin{cases} BN(h-1, 1/p, Z_\ell \mathbf{1}, \boldsymbol{\sigma}^2) & w.p.\ 1/p \\ 0 & w.p.\ 1-1/p. \end{cases} \tag{4.1}$$

where $\mathbf{1}$ is a vector of ones and $\sigma_i^2 = \sigma_e^2\left[1 + i\alpha^2\right]$ is the typical element of $\boldsymbol{\sigma}^2$. The forecast mean and variance are given by

$$m_h = Z_\ell/p \tag{4.2}$$

and

$$v_h = \Big\{(p-1)Z_\ell^2 + \sigma_e^2\left[p + \alpha^2(h-1)\right]\Big\}/p^2. \tag{4.3}$$

The lead-time forecast mean and variance are given by

$$M_h = hZ_\ell/p \tag{4.4}$$

and

$$V_h = \frac{h}{p^3}\Big\{p(p-1)Z_\ell^2 + \sigma_e^2\big[p^2 + p\alpha(1+\alpha/2)(h-1) + \alpha^2(h-1)(h-2)/3\big]\Big\}. \tag{4.5}$$

Theorem 4.2 *The h-step forecast distribution for model (3.4) is*

$$Y_{n+h} \mid \mathcal{I} \sim \begin{cases} LBN(h-1, 1/p, Z_\ell\mathbf{1}, \boldsymbol{\sigma}^2) & w.p.\ 1/p \\ 0 & w.p.\ 1-1/p. \end{cases} \tag{4.6}$$

where $\mathbf{1}$ is a vector of ones and $\sigma_i^2 = \sigma_e^2\big[1 + i\alpha^2\big]$ is the typical element of $\boldsymbol{\sigma}^2$. The forecast mean and variance are given by

$$m_h = \exp\Big\{Z_\ell + \tfrac{1}{2}\sigma_e^2\left[1 + \alpha^2(h-1)/p\right]\Big\}/p \tag{4.7}$$

and

$$v_h = m_h^2\Big(p\ \exp\left\{\sigma_e^2[1+\alpha^2(h-1)/p]\right\} - 1\Big). \tag{4.8}$$

The lead-time forecast mean and variance are given by

$$M_h = \theta_1 \exp\left\{Z_\ell + \sigma_e^2/2\right\} \tag{4.9}$$

and

$$V_h = \frac{M_h^2}{\theta_1^2}\left[\theta_4\exp\{\sigma_e^2\} - \theta_1^2 + \frac{2(\theta_2-\theta_1)}{r-1} + \frac{2(\theta_2 - h/p)\exp\{\alpha\sigma_e^2\}}{r^2-1} - \frac{h(h-1)}{p^2}\right], \tag{4.10}$$

where $\theta_i = \{[1 + (r^i-1)/p]^h - 1\}/(r^i-1)$ and $r = \exp\{\alpha^2\sigma_e^2/2\}$.

It is assumed in both the Croston model and the log-Croston model that interarrival times are continuous and nonstationary. This makes it difficult to calculate the probability of demand occurring at each future time period, and consequently we are unable to derive other properties of the forecast distributions. Instead, we approximate the conditional probability given the previous demand and only look at the one-step forecast distributions.

Theorem 4.3 *For the Croston model (2.5), the one-step forecast distribution is*

$$Y_{n+1} \mid \mathcal{I} \sim \begin{cases} N(Z_\ell, \sigma_e^2) & w.p.\ \rho \\ 0 & w.p.\ 1-\rho \end{cases} \tag{4.11}$$

with mean and variance given by $m_1 = \rho Z_\ell$ *and* $v_1 = \rho(1-\rho)Z_\ell^2 + \rho\sigma_e^2$*, where* $\rho = \Phi\left[(1-P_\ell)/\sigma_\varepsilon\right]$.

For the log-Croston model (3.1), the one-step forecast distribution is given by

$$Y_{n+1} \mid \mathcal{I} \sim \begin{cases} LogN(Z_\ell, \sigma_e^2) & w.p.\ \psi \\ 0 & w.p.\ 1-\psi, \end{cases} \tag{4.12}$$

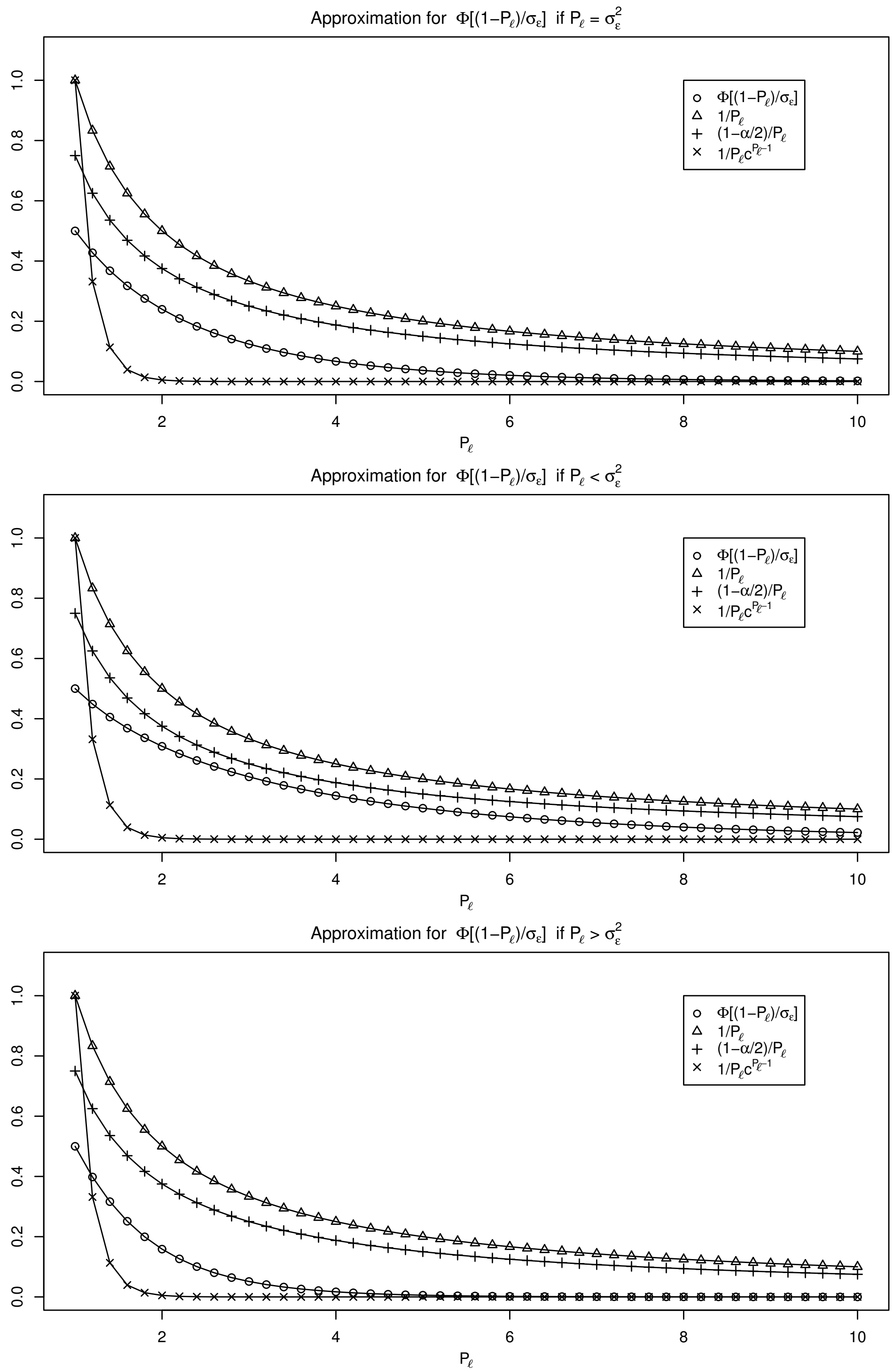
with mean and variance given by $m_1 = \psi \exp\{Z_\ell + \sigma_e^2/2\}$ *and* $v_1 = m_1^2(\exp\{\sigma_e^2\}/\psi - 1)$ *where* $\psi = \Phi\left(-P_\ell/\sigma_\varepsilon\right)$.

Properties for further future steps (i.e., $h \geq 2$) of these two models are not easy to work out or even to approximate.

For Croston's method, several point forecasts have previously been suggested including Croston's (1972) original forecast and the two revised forecasts proposed by Syntetos & Boylan (2000) and Syntetos & Boylan (2001). We shall denote these by F_{cr}, F_{sb0} and F_{sb1} respectively. They are given by

$$F_{cr} = Z_\ell/P_\ell, \qquad F_{sb0} = (1-\alpha/2)Z_\ell/P_\ell \qquad \text{and} \qquad F_{sb1} = Z_\ell/(P_\ell c^{P_\ell - 1})$$

where α and c are constants to be selected. The authors suggest using $c = 100$ and we use this value in the analysis described below. We use $\alpha = 0.5$; other values of α made only small differences to the results. These three one-step ahead point forecasts can be considered as approximations to $m_1 = \Phi[(1-P_\ell)/\sigma_\varepsilon]Z_\ell$. Figure 1 shows the approximation of $\Phi[(1-P_\ell)/\sigma_\varepsilon]$ implied by these three methods for different values of P_ℓ and σ_ε^2. This shows that none of these methods are particularly good. Croston's method F_{cr} is a poor approximation to m_1 for all values of P_ℓ and σ_ε^2. The Syntetos & Boylan (2001) method F_{sb1} works well only when P_ℓ is large. The Syntetos & Boylan (2000) method F_{sb0} provides a better approximation when the value of σ_ε^2 is greater than or equal to P_ℓ. However, when σ_ε^2 is small, it also does poorly.



Sfrag replacements

Figure 1: *Approximations for $\Phi[(1-P_\ell)/\sigma_\varepsilon]$ based on the point forecasts of Croston (1972), Syntetos & Boylan (2000) and Syntetos & Boylan (2001). Here, $c = 100$ and $\alpha = 0.5$. In the middle and the bottom graphs, $P_\ell = \sigma_\varepsilon^2/2$ and $P_\ell = 2\sigma_\varepsilon^2$, respectively.*

5 Prediction Intervals

One of the most important uses of stochastic models in forecasting is the construction of prediction intervals. These can be obtained for any of the models described in the preceding sections by simply simulating many sample paths from the model, and calculating the relevant empirical percentiles from the sample at each forecast horizon. In all simulations, we round the inter-arrival times to the next highest positive integer.

For the modified Croston and modified log-Croston models, we can also obtain analytical prediction intervals which will be simpler to compute. The proof of the following Theorem is given in the Appendix.

Theorem 5.1 *Let* $\delta(h) = \sigma_e[1+\alpha^2(h-1)/p]^{1/2}$*, and let* $k_1 = \Phi^{-1}(\alpha p/2)$ *and* $k_2 = \Phi^{-1}(1-p+\alpha p/2)$*. Also, let* $I(x) = x$ *when* $p < 2/(2-\alpha)$ *and* $I(x) = -\infty$ *otherwise. Then, for the modified Croston model (3.2), a* $(1-\alpha)100\%$ *prediction interval for* h*-step ahead demand* Y_{n+h} *is*

$$\Big(\max\big\{\min[0, Z_\ell + k_1\delta(h)], I[Z_\ell + k_2\delta(h)]\big\}\ ,\ \ Z_\ell - k_1\delta(h)\Big), \tag{5.1}$$

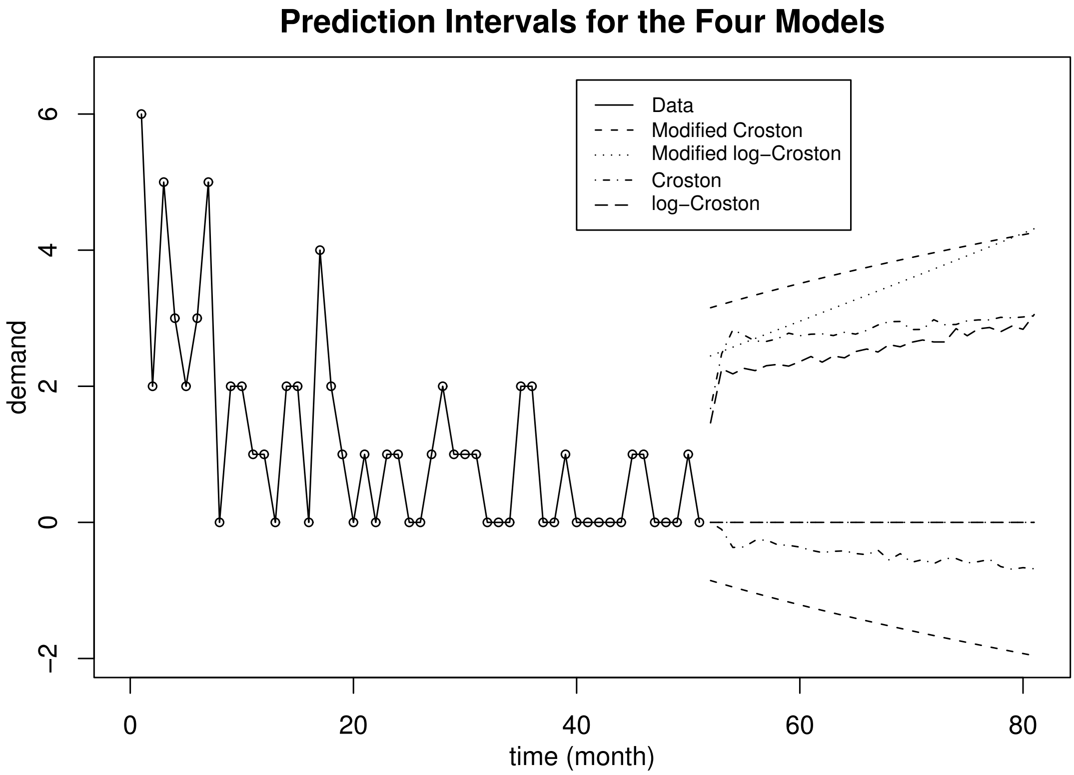
and for the modified log-Croston model (3.4), a $(1-\alpha)100\%$ *prediction interval for* h*-step ahead demand* Y_{n+h} *is*

$$\Big(\max\big\{\min[0, \exp\{Z_\ell + k_1\delta(h)\}], I[\exp\{Z_\ell + k_2\delta(h)\}]\big\}\ ,\ \ \exp\{Z_\ell - k_1\delta(h)\}\Big). \tag{5.2}$$

We demonstrate the use of these prediction intervals by computing them for a real data set. These data consist of monthly demand for a service part for Saturn motor vehicles, over the period January 1998 to March 2002.

For the Croston and log-Croston models, 10000 sample paths were simulated using the fitted parameters. Then the 2.5% and 97.5% percentiles of the sample paths were calculated to give 95% prediction intervals. For the modified models, we use Theorem 5.1 to obtain 95% prediction intervals. These are all shown in Figure 2. Also shown are the point forecasts for each model. Again, the means for the Croston and log-Croston models are computed from the simulated sample paths, while the others are obtained using Theorems 4.1 and 4.2.

The point forecasts are very similar for all models, while the prediction intervals are very different, reflecting the different assumptions made in formulating the models. Obviously, those intervals which include negative values (the Croston and modified Croston models) would normally be truncated in practice, although this will inevitably alter the probability coverage.



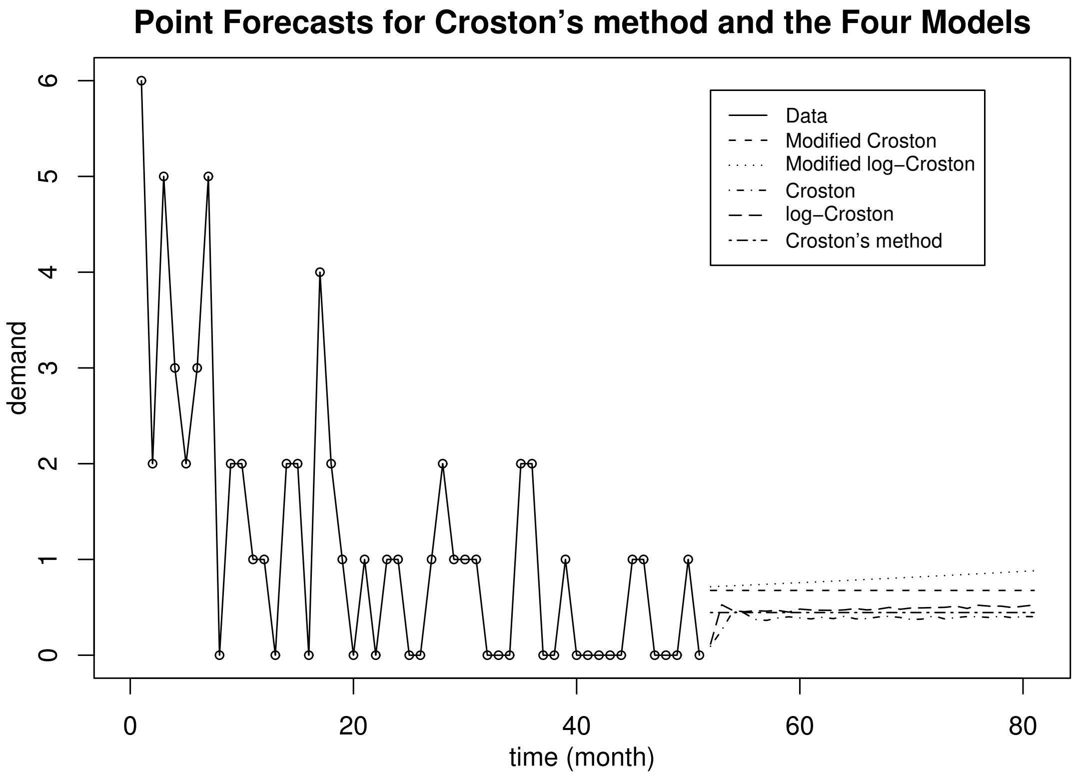


Figure 2: *Prediction intervals and point forecasts for the four models considered here applied to monthly demand for a spare part for Saturn motor vehicles (January 1998 – March 2002). Results for the Croston and log-Croston models were obtained from simulation. In the bottom graph, point forecasts for the original Croston's method are also presented.*

Model name	Forecast mean: $m_h = \mathrm{E}[Y_{n+h} \mid \mathcal{I}]$	Forecast variance: $v_h = \mathrm{Var}[Y_{n+h} \mid \mathcal{I}]$
Modified Croston (Snyder 2002)	$m_h = Z_\ell / p$	$v_h = \frac{1}{p^2}\left\{(p-1)Z_\ell^2 + \sigma_e^2\left[p + \alpha^2(h-1)\right]\right\}$
Modified log-Croston (Snyder 2002)	$m_h = \exp\left\{Z_\ell + \frac{\sigma_e}{2p^2}\left[p + \alpha^2(h-1)\right]\right\}$	$v_h = m_h^2\left(p \exp\left\{\frac{\sigma_e^2}{p}[p + \alpha^2(h-1)]\right\} - 1\right)$
Croston (Croston 1972)	$m_1 = \rho Z_\ell$	$v_1 = \rho(1-\rho)Z_\ell^2 + \rho\sigma_e^2$
log-Croston	$m_1 = \psi \exp\{Z_\ell + \sigma_e^2/2\}$	$v_1 = m_1^2(\exp\{\sigma_e^2\}/\psi - 1)$

Table 1: *Forecast means and variances for the four IDF models. Here* $\rho = \Phi\left[(1-P_\ell)/\sigma_\varepsilon\right]$, $\psi = \Phi\left(-P_\ell/\sigma_\varepsilon\right)$ *and* $\Phi(\cdot)$ *is the standard normal distribution function.*

6 Conclusion

We have shown that any model assumed to be underlying Croston's method must be non-stationary and defined on a continuous sample space including negative values. Hence, the implied model has properties that don't match the demand data being modelled.

We have also studied models underlying some of the suggested modifications to Croston's method. Table 1 summarizes the forecast means and variances for the four models discussed in this paper. Of these, only the modified log-Croston model is defined on the positive real line (which may be considered an approximation to the non-negative integers) and has tractable expressions for the forecast mean and variance. This makes it a more attractive candidate for IDF modelling than either Croston's original proposal or the other suggested modifications.

However, notably all four models discussed here are nonstationary. Consequently, the forecast variances are all increasing over time which can result in wide prediction intervals, especially at long forecast horizons. It would be useful to also consider stationary models for IDF rather than restrict attention to models based on SES. Potential stationary models are Poisson autoregressive models (see Grunwald et al. 2000) and other time series models for counts (e.g. Cameron & Trivedi 1998, Winkelmann 2000). To our knowledge, these have not be used for intermittent demand data before, and are worthy of investigation in this context.

Acknowledgments

We thank Donald Parent of Saturn Service Parts, Spring Hill, Tennessee USA, for providing the data discussed in Section 5. The authors are also grateful to the Monash University Business and Economic Forecasting Unit for supporting this research.

References

Abraham, B. & Ledolter, J. (1983), *Statistical methods for forecasting*, John Wiley & Sons, New York.

Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (1994), *Time series analysis: forecasting and control*, 3rd edn, Holden-Day, San Francisco.

Cameron, A. C. & Trivedi, P. K. (1998), *Regression Analysis of Count Data*, Econometric Society Monograph No.30, Cambridge University Press.

Chatfield, C., Koehler, A. B., Ord, J. K. & Snyder, R. D. (2001), 'A new look at models for exponential smoothing', *The Statistician* **50**(2), 147–159.

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

Grunwald, G. K., Hamza, K. & Hyndman, R. J. (1997), 'Some properties of generalizations of non-negative Bayesian time series models', *Journal of Royal Statistical Society, Series B* **59**(3), 615–626.

Grunwald, G. K., Hyndman, R. J., Tedesco, L. & Tweedie, R. L. (2000), 'Non-Gaussian conditional linear AR(1) models', *Australian & New Zealand Journal of Statistics* **42**(4), 475–479.

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

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

Syntetos, A. A. & Boylan, J. E. (2000), 'Developments in forecasting intermittent demand'. Paper presented at 20th International Symposium on Forecasting, June 21–24, 2000, Lisbon, Portugal.

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

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

Winkelmann, R. (2000), *Econometric analysis of count data*, 3rd edn, Springer-Verlag, Berlin.

Appendix: Proofs

Proof of Theorem 4.1

Assume there are s arrivals of non-zero demands within $(n, n+h]$. We first derive results conditional on s. It is useful to note that $Y_{n+h} = X_{n+h}Y^*_{\ell+s}$ and, by the properties of SES,

$$Y^*_{\ell+s} = Z_\ell + e_{\ell+s} + \alpha(e_{\ell+s-1} + \cdots + e_{\ell+1}).$$

So $Y^*_{\ell+s}$ has a normal distribution and the forecast distribution conditional on s is

$$Y_{n+h} \mid \mathcal{I}, s \sim \begin{cases} \mathrm{N}\left(Z_\ell\,,\ \sigma_e^2\big[1+\alpha^2(s-1)\big]\right) & \text{w.p. } 1/p \\ 0 & \text{w.p. } 1-1/p. \end{cases} \tag{A.1}$$

Since Y_{n+h} takes values from the normal distribution in (A.1) only if demand occurs in time $n+h$ (i.e., $X_{n+h}=1$), then, in this case, $s \geq 1$ and $(s-1) \sim \text{Binomial}(h-1, 1/p)$. Thus we obtain (4.1), from which we obtain

$$m_h = \mathrm{E}[X_{n+h} \mid \mathcal{I}]\mathrm{E}[Y^*_{\ell+s} \mid \mathcal{I}] = Z_\ell/p$$

and

$$\begin{aligned} v_h &= \mathrm{Var}[X_{n+h} \mid \mathcal{I}](\mathrm{E}[Y^*_{\ell+s} \mid \mathcal{I}])^2 + \mathrm{E}[X^2_{n+h} \mid \mathcal{I}]\mathrm{Var}[Y^*_{\ell+s} \mid \mathcal{I}] \\ &= (1-1/p)Z_\ell^2/p + \sigma_e^2(1+\alpha^2\mathrm{E}[s-1])/p \\ &= \Big\{(p-1)Z_\ell^2 + \sigma_e^2\big[p+\alpha^2(h-1)\big]\Big\}/p^2. \end{aligned}$$

We can rewrite the h-step ahead lead-time demand as

$$Y_n(h) = \sum_{k=1}^{s} Y^*_{\ell+k} = Y^*_\ell(s) = sZ_\ell + \sum_{k=1}^{s}\big[1+\alpha(s-k)\big]e_{\ell+k}$$

so that $\mathrm{E}[Y^*_\ell(s) \mid \mathcal{I}, s] = sZ_\ell$ and $\mathrm{Var}[Y^*_\ell(s) \mid \mathcal{I}, s] = s\{1+\alpha(s-1)+\frac{\alpha^2}{6}(s-1)(2s-1)$. Hence, the conditional distribution of h-step ahead lead-time demand will be

$$Y_n(h) \mid [\mathcal{I}, s] \sim \mathrm{N}\Big(sZ_\ell\,,\ \sigma_e^2 s\big[1+\alpha(s-1)+\tfrac{\alpha^2}{6}(s-1)(2s-1)\big]\Big), \tag{A.2}$$

where $s \geq 0$ and therefore $s \sim \text{Binomial}(h, 1/p)$. Note the distribution of s in (A.2) differs from that in (A.1). Thus we obtain $M_h = \mathrm{E}[sZ_\ell] = hZ_\ell/p$ and

$$\begin{aligned} V_h &= \mathrm{Var}[\mathrm{E}(Y_n(h) \mid \mathcal{I}, s)] + \mathrm{E}[\mathrm{Var}(Y_n(h) \mid \mathcal{I}, s)] \\ &= \mathrm{Var}(sZ_\ell \mid \mathcal{I}) + \mathrm{E}[\sigma_e^2 s(1 + \alpha(s-1) + \tfrac{\alpha^2}{6}(s-1)(2s-1)) \mid \mathcal{I}] \\ &= h(p-1)Z_\ell^2/p^2 + \frac{h\sigma_e^2}{p}\left\{1 + \alpha(1+\alpha/2)(h-1)/p + \tfrac{\alpha^2}{3p^2}(h-1)(h-2)\right\}. \end{aligned}$$

Here, we have used the fact that, if $s \sim \mathrm{Binomial}(h, 1/p)$, then $\mathrm{E}[s(s-1)] = h(h-1)/p^2$ and $\mathrm{E}[s(s-1)(s-2)] = h(h-1)(h-2)/p^3$.

Proof of Theorem 4.2

The derivations are similar to that for Theorem 4.1, so we omit some of the details and use similar notation. We can write $Y_{n+h} = X_{n+h}Y^*_{\ell+s} = X_{n+h}\exp\{W_{\ell+s}\}$, where $\{W_j\}$ is the ARIMA(0,1,1) process underlying the logarithm of the demand size. Thus the h-step ahead forecast distribution of Y_{n+h} conditional on s for this model is

$$Y_{n+h} \mid \mathcal{I}, s \sim \begin{cases} \mathrm{LogN}\Big(Z_\ell\ ,\ \sigma_1^2\left[1 + \alpha^2(s-1)\right]\Big) & \text{w.p. } 1/p \\ 0 & \text{w.p. } 1 - 1/p, \end{cases} \tag{A.3}$$

where $(s-1) \sim \mathrm{Binomial}(h-1, 1/p)$ given $X_{n+h} = 1$. This gives (4.6) and we find the h-step ahead forecast mean is given by (4.7) and the variance is given by

$$\begin{aligned} v_h &= \mathrm{Var}[X_{n+h} \mid \mathcal{I}]\left(\mathrm{E}[e^{W_{\ell+s}} \mid \mathcal{I}]\right)^2 + \mathrm{E}[X_{n+h}^2 \mid \mathcal{I}]\mathrm{Var}[e^{W_{\ell+s}} \mid \mathcal{I}] \\ &= \frac{1}{p}\left(\exp\left\{\sigma_e^2\left[1 + \alpha^2(h-1)/p\right]\right\} - \frac{1}{p}\right)\exp\Big\{2Z_\ell + \sigma_e^2\left[1 + \alpha^2(h-1)/p\right]\Big\} \end{aligned}$$

from which (4.8) follows.

The h-step ahead lead time demand $Y_n(h) = Y^*_\ell(s)$ is the sum of correlated lognormal random variables, for which the distribution is unknown. However, we can derive the mean by first noting that, if $s \sim \mathrm{Binomial}(h, 1/p)$, then $\mathrm{E}[x^s] = (1 + (x-1)/p)^h$. Thus, the mean lead time demand for this model is given by

$$\begin{aligned} M_h = \mathrm{E}[Y^*_\ell(s) \mid \mathcal{I}] &= \mathrm{E}\left[\sum_{k=1}^{s}\exp\{W_{\ell+k}\} \mid \mathcal{I}\right] \\ &= \mathrm{E}\left[\sum_{k=1}^{s}\exp\left\{Z_\ell + e_{\ell+k} + \alpha(e_{\ell+k-1} + \cdots + e_{\ell+1})\right\}\right] \\ &= \mathrm{E}\left[\frac{r^s - 1}{r - 1}\exp\{Z_\ell + \sigma_e^2/2\}\right] \end{aligned}$$

from which we obtain (4.9). The variance of the h-step ahead lead-time demand can be derived similarly as follows.

$$
\begin{aligned}
V_h &= \mathrm{E}\left[\mathrm{Var}\left(\sum_{k=1}^{s}\exp\{W_{\ell+k}\}\mid \mathcal{I}, s\right)\right] + \mathrm{Var}\left[\mathrm{E}\left(\sum_{k=1}^{s}\exp\{W_{\ell+k}\}\mid \mathcal{I}, s\right)\right] \\
&= \mathrm{E}\left[\sum_{k=1}^{s}\mathrm{Var}\left(\exp\{W_{\ell+k}\}\mid \mathcal{I}, s\right) + 2\sum_{1\le i\le k\le s}\mathrm{Cov}\left(\exp\{W_{\ell+i}\}, \exp\{W_{\ell+k}\}\mid \mathcal{I}, s\right)\right] \\
&\quad + \mathrm{Var}\left[\mathrm{E}\left(\sum_{k=1}^{s}\exp\{W_{\ell+k}\}\mid \mathcal{I}, s\right)\right] \\
&= \mathrm{E}\left[\left\{\tfrac{(r^{4s}-1)\exp\{\sigma_e^2\}}{r^4-1} - \tfrac{r^{2s}-1}{r^2-1} + \tfrac{2[r^{2s}-1-s(r^2-1)]\exp\{\alpha\sigma_e^2\}}{(r^2-1)^2} - s(s-1)\right\}\exp\{2Z_\ell + \sigma_e^2\}\right] \\
&\quad + \mathrm{Var}\left[\frac{r^s-1}{r-1}\exp\{Z_\ell + \sigma_e^2/2\}\right].
\end{aligned}
$$

Then simple algebra leads to (4.10).

Proof of Theorem 4.3

For simplicity, we assume that the last non-zero demand Y_ℓ^* occurs at time n, i.e., $Y_n = Y_\ell^*$. Then, $Q_{\ell+1}$ is the number of time periods we wait from Y_ℓ^* until the next non-zero demand $Y_{\ell+1}^*$. Then, using the properties of SES, for the Croston model the conditional distribution of $Q_{\ell+1}$ is $Q_{\ell+1} \mid Q_1, \dots, Q_\ell \sim \mathrm{N}(P_\ell, \sigma_\varepsilon^2)$, and for the log-Croston model $Q_{\ell+1} \mid Q_1, \dots, Q_\ell \sim \mathrm{LogN}(P_\ell, \sigma_\varepsilon^2)$.

In the Croston model, the conditional probability of demand occurring probability at time $n+1$ can be approximated as

$$\Pr(X_{n+1} = 1 \mid \mathcal{I}) = \Pr(Q_{\ell+1} \le 1 \mid Q_1, \dots, Q_\ell) = \Phi[(1 - P_\ell)/\sigma_\varepsilon] = \rho.$$

Then the one-step ahead forecast distribution is given by (4.11) from which we can obtain the one-step forecast mean and variance.

A similar argument for the log-Croston model leads to the one-step ahead forecast distribution in the form (4.12) from which we can obtain the one-step ahead forecast mean and variance.

Proof of Theorem 5.1

For each of the two models, we obtain a prediction interval with its lower and upper limits, a and b, being respectively the $(\alpha/2)$th and the $(1-\alpha/2)$th quantiles of the forecast distribution such that $\Pr(Y_{n+h} \le a) = \Pr(Y_{n+h} \ge b) = \alpha/2$.

For the modified Croston model, the forecast distribution is given by (4.1). Let s be the number of non-zero demands within $(n, n+h]$ and let $D(s) = Y_{n+h} \mid (s, X_{n+h} = 1)$. Then

$$\begin{aligned}\Pr(Y_{n+h} \le y) &= (1 - \tfrac{1}{p})1_{\{y\ge 0\}} + \tfrac{1}{p}\sum_s \Pr(D(s) \le y)\Pr(X = s) \\ &= (1 - \tfrac{1}{p})1_{\{y\ge 0\}} + \tfrac{1}{p}\mathrm{E}[\Pr(D(s) \le y)]\end{aligned}$$

where $s - 1 \sim$ Binomial$(h-1, 1/p)$ provided $s \ge 1$, and $1_{\{y\ge 0\}}$ is an indicator function taking value 1 if $y \ge 0$ and 0 otherwise.

Now if $s \ge 1$ then $D(s) \sim \mathrm{N}\left(Z_\ell\,,\ \delta^2(s)\right)$ where $\delta^2(s) = \sigma_e^2\left[1 + \alpha^2(s-1)\right]$. Because $Z_\ell > 0$ for demand data, it is always true that the upper limit of the prediction interval $b > 0$. So we have $\Pr(D(s) \le b) = \Phi\{[b - Z_\ell]/\delta(s)\}$ and therefore $\alpha/2 = \mathrm{E}[\Pr(D(s) > b)]/p$. Thus $b = Z_\ell - k_1\mathrm{E}[\delta(s)] = Z_\ell - k_1\delta(h)$.

Similarly, for the lower limit a, if $\mathrm{E}[\Pr(D(s) < 0)] \ge \alpha p/2$, then $a \le 0$ and we have $\alpha/2 = \mathrm{E}[\Pr(D(s) < a)]/p$ which gives $a = Z_\ell + k_1\delta(h)$. Now let $d = \alpha/2 - \mathrm{E}[\Pr(D(s) < 0)]/p$. Then when $\mathrm{E}[\Pr(D(s) < 0)] < \alpha p/2$, we have $d > 0$. If $p \ge 2/(2-\alpha)$ then $a = 0$ since $d \le 1 - 1/p$. When $p < 2/(2-\alpha)$, if $\Pr[D(s) < 0] \ge \alpha p/2 - p + 1$, then $d \le 1 - 1/p$ and therefore $a = 0$; otherwise $d > 1 - 1/p$ and $a = Z_\ell + k_2\delta(h)$.

A similar argument for forecast distribution (4.6) leads to the prediction interval in (5.2) for the modified log-Croston model.