Please cite this paper as:

Svetunkov Ivan and Boylan John E. (2017). Multiplicative State-Space Models for Intermittent Time Series. Working Paper of Department of Management Science, Lancaster University, 2017:4 , 1–43.

Lancaster University Management School

Managmeent Science

Working Paper 2017:4

Multiplicative State-Space Models for Intermittent Time Series

Ivan Svetunkov and John E. Boylan

The Department of Management Science
Lancaster University Management School
Lancaster LA1 4YX
UK

Multiplicative State-Space Models for Intermittent Time Series

Ivan Svetunkov[a,*], John E. Boylan[a]

[a] *Centre for Marketing Analytics and Forecasting*
Lancaster University Management School, Lancaster, LA1 4YX, UK

Abstract

Intermittent demand forecasting is an important supply chain task, which is commonly done using methods based on exponential smoothing. These methods however do not have underlying statistical models, which limits their generalisation. In this paper we propose a general state-space model that takes intermittence of data into account, extending the taxonomy of exponential smoothing models. We show that this model has a connection with conventional non-intermittent state space models and underlies Croston's and Teunter-Syntetos-Babai (TSB) forecasting methods. We discuss properties of the proposed models and show how a selection can be made between them in the proposed framework. We then conduct experiments on simulated data and on two real life datasets, demonstrating advantages of the proposed approach.

Keywords: Inventory forecasting, state space models, exponential smoothing, intermittent demand, Croston, count data

1. Introduction

An intermittent time series is a series that has non-zero values occurring at irregular frequency. The data is usually, but not necessarily, discrete and often takes low integer values. Intermittent series occur in many application areas where there are rare events. Examples include security breaches, natural disasters and the occurrence of demand for slow-moving products. In

*Correspondence: Ivan Svetunkov, Department of Management Science, Lancaster University Management School, Lancaster, Lancashire, LA1 4YX, UK.

Email address: `i.svetunkov@lancaster.ac.uk` (Ivan Svetunkov)

the last case we usually refer to "intermittent demand" which, in addition to irregularity of occurrence, contains only zeroes and positive values. The final application area is important in a supply-chain setting, where decisions need to be made about the quantity to order and the discontinuation of ordering.

In this paper, we establish a general modelling framework for intermittent series, which does not depend on any particular application area. However, its development has been motivated by the requirements of inventory management. Two forecasting methods that have been described in the supply-chain literature to inform how much to order and when to discontinue a product, are Croston's method (Croston, 1972) and the TSB method (Teunter et al., 2011). Neither of these methods has, so far, been furnished with a satisfactory statistical model.

From a practical supply-chain perspective, there are a number of issues that need to be resolved. We need to decide, in a systematic way, which intermittent demand forecasting method to use, rather than allowing these choices to be arbitrary. Having chosen the forecasting method, it needs to be parametrised appropriately. Finally, replenishment decisions should be informed by reliable estimates. For inventory systems based on the probability of stock-out, good estimates of upper percentiles of demand are required. For systems based on fill-rates (percentage of demand filled immediately from stock), good estimates of probabilities of demand are required. In the latter case, these may be calculated from an estimated Cumulative Distribution Function. The modelling approach recommended in this paper supports the estimation of both percentiles and Cumulative Distribution Functions. Indeed, we argue that a sound statistical model can support method choice, parametrisation and, ultimately, replenishment decisions in supply chains.

In this paper we propose a statistical model for intermittent data, discuss its properties, then derive reasonable and concise models underlying Croston's method and TSB and demonstrate their advantages. We also show the connection between conventional forecasting models and the intermittent demand model. We then demonstrate how the proposed intermittent demand model works on several examples. Thus we contribute towards filling a gap of modelling intermittent time series, which opens new research directions in the area.

2. Literature review

The most popular intermittent demand forecasting method was proposed by Croston (1972). His method has been researched extensively in recent years and has been implemented in widely adopted supply chain software packages (e.g. SAP APO). Croston was the first to note that, when demand is intermittent, simple exponential smoothing produces biased forecasts immediately after demand occurrences (known as 'decision-point bias'). So he proposed splitting the observed data into two parts: demand sizes and demand occurrences. The proposed model in Croston (1972) has the following simple form:

$$y_t = o_t z_t, \tag{1}$$

where o_t is a binary Bernoulli distributed variable taking a value of one when demand occurs and zero otherwise and z_t is demand size, having some conditional distribution. Proposing the model (1), Croston suggested to work with each of these two parts separately, showing that the probability of occurrence can be estimated using intervals between demands. If q_t is the time elapsed since the last non-zero observation, then it represents the demand interval when the next non-zero observation occurs. Both demand sizes z_t and demand intervals q_t are forecasted in this method using simple exponential smoothing, which leads to the following system:

$$\begin{aligned} \hat{y}_t &= \frac{1}{\hat{q}_t}\hat{z}_t \\ \hat{z}_t &= \alpha_z z_{t-1} + (1-\alpha_z)\hat{z}_{t-1}, \\ \hat{q}_t &= \alpha_q q_{t-1} + (1-\alpha_q)\hat{q}_{t-1} \end{aligned} \tag{2}$$

where $\hat{y}_t$ is the predicted mean demand, $\hat{z}_t$ is the predicted demand size and $\hat{q}_t$ is the predicted demand interval and α_q and α_z are smoothing parameters for intervals and sizes respectively. In Croston's initial formulation it was assumed that $\alpha_q = \alpha_z$, but separate smoothing parameters were later suggested by Schultz (1987), and this additional flexibility has been supported by other researchers (e.g. Snyder, 2002; Kourentzes, 2014). Note that the update of variables $\hat{q}_t$ and $\hat{z}_t$ happens in the method (2) only when $o_t = 1$. When $o_t = 0$, then there is no updating, $\hat{z}_t = \hat{z}_{t-1}$ and $\hat{q}_t = \hat{q}_{t-1}$.

Syntetos and Boylan (2001, 2005) showed that estimating the mean demand using the first equation in (2) leads to 'inversion bias' and in order to correct it, they proposed the following approximation (known as the

Syntetos-Boylan Approximation, SBA):

$$\hat{y}_t = \left(1 - \frac{\alpha_q}{2}\right) \frac{1}{\hat{q}_t} \hat{z}_t. \tag{3}$$

They conducted an experiment on 3000 real time series and showed that forecasting accuracy of SBA is higher than Croston's method (Syntetos and Boylan, 2005).

Although various models have been proposed, none have so far been identified which would be appropriate for non-negative integer series and would underlie Croston's method. This means that heuristic methods of initialisation and parameter estimation are used instead of statistically rigorous ones. Several authors over the years have looked into this problem.

Snyder (2002) discussed possible statistical models underlying Croston's method. He examined the following form:

$$y_t = o_t \mu_{t|t-1} + \epsilon_t, \tag{4}$$

where $\mu_{t|t-1}$ is the conditional expectation of demand sizes. Snyder (2002) showed that the model (4) contradicts some basic assumptions about intermittent demand. The main reason for this is because the error term ϵ_t is assumed to be normally distributed, but this means that demand can be negative. So, Snyder (2002) proposed the following modified intermittent demand model:

$$y_t^+ = o_t \exp(\mu_{t|t-1} + \epsilon_t), \tag{5}$$

where y_t^+ represents the demand at time t.

Shenstone and Hyndman (2005) studied several possible statistical models with additive errors, including those of Snyder (2002), to identify a model for which Croston's method is optimal. They argued that any model underlying Croston's method must be non-stationary and defined on continuous space including negative values. They concluded that such a model has unrealistic properties.

However, one of the main conclusions of Shenstone and Hyndman (2005) is open to misinterpretation. One should not conclude that intermittent demand methods do not have and cannot have any reasonable underlying statistical model. This conclusion depends on the important assumption of an additive error term. In this paper, we shall propose statistical models with multiplicative error terms as an alternative to the models discussed by Shenstone and Hyndman (2005).

(Hyndman et al., 2008, pp. 281 - 283) proposed a model underlying Croston's method based on a Poisson distribution of demand sizes with time varying parameter λ_t:

$$\begin{aligned} & y_t = o_t z_t \\ & z_t \sim \text{Poisson}(\lambda_{t-1} - 1) + 1 \\ & \lambda_t = \alpha_z z_{t-1} + (1 - \alpha_z)\lambda_{t-1}, \\ & o_t \sim \text{Bernoulli}\left(\frac{1}{q_t}\right) \\ & q_t = \alpha_p \tau_{t-1} + (1 - \alpha_p) q_{t-1} \end{aligned} \tag{6}$$

where λ_t is the average number of events per trial and τ_t is the observed demand intervals. The authors point out that the proposed model "gives one-step-ahead forecasts equivalent to Croston's method". This is because the conditional expectation of z_t for one-step-ahead in (6) is equal to $\lambda_{t-1}-1+1 = \lambda_{t-1}$, which corresponds to Croston's $\hat{z}_t$ in (2). However, the proposed model has two problems.

First, (6) cannot be considered as an appropriate statistical model, because it uses the SES method for λ_t, without giving a rationale for the generating process. The system of equations (6) should be considered as a filter instead. It still retains useful statistical properties, but it is not as powerful as appropriate statistical models, and sidesteps the ETS taxonomy. Furthermore although the Poisson distribution becomes closer to the normal distribution with an increase of λ, the connection between the model (6) and the conventional ETS models is not apparent, making two separate cases. The authors also do not propose any statistical model underlying the demand intervals q_t and once again use SES. This limits the properties of the occurrence part of the model to the specific forecasting method.

Second, using the filter described in (6) restricts its generalisation, because introduction of new components or exogenous variables is not straightforward in this framework.

Overall, while the filter (6) solves some problems for intermittent demand, and has a connection with Croston's method, it cannot be considered as a complete solution to the problem.

In Snyder et al. (2012) several intermittent demand models were proposed. The model (6) is called in that paper the "Hurdle shifted Poisson" model. The authors suggested applying the Negative Binomial distribution with time varying mean value to the intermittent data, and found that it performs better than the other filters. However this filter has the same problems as the Poisson one, discussed above. In addition, although it is well-known that with the increase of trials the Binomial distribution asymptotically becomes equivalent to the normal distribution, it cannot be explicitly used when the probability of success is equal to one, because it is not supported in this case. Furthermore because the authors did not model the occurrence variable separately, the proposed Negative Binomial filter is more restrictive than the filter (6). Finally, the authors did not make a comparison with the ETS(A,N,N) model in their paper, so it is not possible to assess the accuracy advantage of the proposed filters in comparison with the simpler non-intermittent models.

Another intermittent demand method was proposed by Teunter et al. (2011), which has been known in the literature as TSB. It was derived for obsolescence of inventory, but can be used for other cases as well. The authors proposed using the same principle as in (1), but estimating the time varying probability of demand occurrence p_t using simple exponential smoothing based on the variable o_t rather than switching to intervals between demands. Their method can be represented by the following system of equations:

$$\begin{aligned} \hat{y}_t &= \hat{p}_t \hat{z}_t \\ \hat{z}_t &= \alpha_z z_{t-1} + (1-\alpha_z)\hat{z}_{t-1} \ , \\ \hat{p}_t &= \alpha_p o_{t-1} + (1-\alpha_p)\hat{p}_{t-1}. \end{aligned} \tag{7}$$

where $\hat{p}_t$ is the predicted probability of demand occurrence and α_p is the smoothing parameter for this probability estimate. The update of probability in TSB is done after each observation, while demand sizes are updated only when $o_t = 1$. In cases when $o_t = 0$ there is no updating, $\hat{z}_t = \hat{z}_{t-1}$. An advantage of this method is that the conditional expectation does not need any corrections similar to (3). However the authors did not propose a statistical model for their method, which leads to issues similar to the ones for Croston's method. These include problems with the correct estimation of the model parameters, conditional mean and variance.

Both TSB and Croston can be applied to fast moving products, where they become equivalent to simple exponential smoothing. They both perform well on several datasets (Kourentzes, 2014); however they are disconnected from other exponential smoothing methods and are considered to be a different group.

3. Statistical model

3.1. Filter or statistical model?

Before proceeding to the main part of the paper, it is important to discuss two forecasting approaches, that are often met in the literature, namely using filters and using statistical models.

The filtering approach assumes that the data is used as an input of some equation (or set of equations) and is transformed into a value of the same scale as the original data. The classical example of a filter is Simple Exponential Smoothing (SES), which has the form:

$$\hat{y}_t = \alpha y_{t-1} + (1-\alpha)\hat{y}_{t-1}, \tag{8}$$

where α is the smoothing constant. The advantage of filters is in their simplicity and the small number of assumptions. For example, SES is very easy to interpret and can be used without assuming normality of the residuals. The main disadvantage of filters is in the lack of statistical rationale. This leads to ambiguity in estimation of the parameters of the filter and problems in the construction of prediction intervals. For example, different initialisation procedures and different estimators can be applied to (8) and there is no way to say which of them should be preferred without a rigorous analysis of its predictive performance. This was one of the main arguments against using exponential smoothing in the statistical literature and one of the reasons that statisticians used to prefer ARIMA models to exponential smoothing (Box and Jenkins, 1976). Finally, the selection of a filter appropriate to the data is one of the problems that does not have a straightforward solution. Both Croston's and TSB methods are filters.

As for the statistical models, they have their own advantages and disadvantages as well. For example, statistical models usually have strict assumptions on the error term, which means that in order to select the correct model and correctly estimate its parameters, those assumptions need to hold. The upside of having a model is in a simplified model selection procedure (which nowadays is based on information criteria), statistically rigorous estimation of parameters and in a simplified derivation of prediction intervals. Overall, models allow working with the distributions of values, while filters are focused on point values. Finally, it is well known that there is no such thing as a "true model". But even if the model is wrong it can still be useful.

Overall, both approaches are used in practice and both of them have advantages and disadvantages. In this paper we employ the modelling approach, keeping in mind its upsides and downsides mentioned above.

3.2. General Intermittent State-Space Model

We start from Croston's original formulation (1) and split intermittent demand into two parts in a similar way, but assuming that z_t is generated using a statistical model on its own. We argue that the assumption that the error term interacts with the final demand y_t rather than demand sizes z_t is the main flaw in the logic of derivation of statistical models underlying intermittent demand forecasting methods. Moving the error term into z_t allows using any statistical model that a researcher prefers (e.g. ARIMA, ETS, regression, diffusion model etc). The model underlying z_t corresponds to potential demand for a product, while the other model, underlying o_t, corresponds to demand realisations, when a customer makes a purchase of a product.

Taking into account that both Croston's method and TSB use exponential smoothing methods, we propose to use a model form that underlies this forecasting approach. We adopt the single source of error (SSOE) state-space model, as this has been well-established (Snyder, 1985; Hyndman et al., 2002) whilst acknowledging that other model forms are possible (e.g. multiple source of error, MSOE). We use the SSOE model for z_t, which in a very general way has the following form, based on (1):

$$\begin{aligned} y_t &= o_t z_t \\ z_t &= w(\mathbf{v}_{t-1}) + r(\mathbf{v}_{t-1})\epsilon_t , \\ \mathbf{v_t} &= f(\mathbf{v}_{t-1}) + g(\mathbf{v}_{t-1})\epsilon_t \end{aligned} \tag{9}$$

where o_t is a Bernoulli distributed random variable, $\mathbf{v}_t$ is the state vector, ϵ_t is the error term, $f(\cdot)$ is the transition function, $w(\cdot)$ is the measurement function, $g(\cdot)$ is the persistence function and $r(\cdot)$ is the error term function. These correspond to the functions in (Hyndman et al., 2008, p.54) and allow both additive and multiplicative state-space models. One advantage of this approach is that in cases of fast moving demand o_t becomes equal to one for all t, which transforms the model (9) from an intermittent into a non-intermittent conventional model. This modification expands the Hyndman et al. (2008) taxonomy and allows introducing simple modifications of the model by inclusion of time series components and exogenous variables.

In our new model, the first equation corresponds to Croston's original formulation (1), while the second equation, called the measurement equation, reflects the potential demand size evolution over time. The third equation is the standard transition equation for an SSOE model, describing the change of components of the model over time.

An interpretation of the new intermittent demand model (9) is that a potential demand size may change in time even when an actual demand is not observed. In these cases, $o_t = 0$, leading to $y_t = 0$ in the first equation of (9). However the measurement and the transition equations in (9) are not affected by o_t, leading to potential evolution of z_t regardless of whether there is an actual demand occurrence or not.

One thing to note about this model is that it can be applied to intermittent data with continuous non-zero observations. Such series arise in the context of natural disasters and other natural phenomena. They are less common in a supply chain context, but time series with such characteristics do exist. For example, daily sales of an expensive coffee sold per ounce can exhibit such behaviour with zeroes in some days and then fractional quantities in the others.

However, while the model (9) solves the problem identified by Shenstone and Hyndman (2005) of negative values (because now a multiplicative model can be used for z_t), there is still a need for an integer-valued model. In order to solve this problem, we propose a simple modification of the first equation in (9):

$$y_t = o_t \lceil z_t \rceil, \tag{10}$$

where $\lceil z_t \rceil$ is the rounded up value of z_t. This way the statistical model we propose becomes integer-valued, and it does not contradict any reasonable assumptions about intermittent demand. Furthermore, any statistical model can be used for z_t. It is worth noting that the rounding is an important issue, which will be explored further in this paper, in Section 3.7. But before looking into model (10) we need to study the properties of the basic model (9), keeping in mind that it is an approximation of the more realistic model (10).

The model with rounded up values will be called in this paper "integer" model, while the simpler model (9) will be referred to as "continuous".

In order for the model (9) to work we make the following assumptions, some of which can be relaxed and would lead to different models:

1. Demand size z_t is continuous. This assumption is relaxed in (10) and

discussed later in this paper, in Section 3.7.

2. Demand size z_t is independent of its occurrence o_t. Relaxing this assumption will lead to a different statistical model;
3. Potential demand size may change in time even if we do not observe it. Relaxing this assumption means that we need to impose additional restrictions on the transition equation;
4. o_t has a Bernoulli distribution with some probability p_t that in the most general case varies in time. This is a natural assumption, following the idea of Croston (1972). Making some other assumption in its place will also lead to a different statistical model.

With these assumptions the proposed intermittent state-space model allows calculating conditional expectation and variance for several steps ahead using the following formulae:

$$\begin{aligned} &\mu_{y,t+h|t} = \mu_{o,t+h|t}\mu_{z,t+h|t} \\ &\sigma^2_{y,t+h|t} = \sigma^2_{o,t+h|t}\sigma^2_{z,t+h|t} + \sigma^2_{o,t+h|t}\mu^2_{z,t+h|t} + \mu^2_{o,t+h|t}\sigma^2_{z,t+h|t} \end{aligned}, \tag{11}$$

where $\mu_{y,t+h|t}$ and $\sigma^2_{y,t+h|t}$ are respectively conditional expectation and conditional variance of y_t; $\mu_{o,t+h|t}$ and $\sigma^2_{o,t+h|t}$ are conditional expectation and variance of occurrence variable o_t and finally $\mu_{z,t+h|t}$ and $\sigma^2_{z,t+h|t}$ are the respective values for the demand sizes z_t.

The important point is that, taking into account intermittent demand, pure multiplicative models make more sense for the measurement equation in (9) than additive or mixed ones, because they restrict the space of demand sizes to positive numbers. In this paper we discuss the multiplicative error ETS model for demand sizes, namely ETS(M,N,N), which denotes multiplicative error, no trend and no seasonality. The reason for this choice is because ETS(M,N,N) is a simple well-known model underlying simple exponential smoothing (Hyndman et al., 2008, p.54), which is a core method in both Croston and TSB. However more complicated models can also be used instead of ETS(M,N,N), but they are not of the main interest in this paper.

The general continuous intermittent state-space model (9) reduces to the special case, called iETS(M,N,N), and can be written as:

$$\begin{aligned} &y_t = o_t l_{z,t-1}\left(1+\epsilon_t\right) \\ &l_{z,t} = l_{z,t-1}(1+\alpha_z\epsilon_t) \end{aligned}, \tag{12}$$

where $l_{z,t}$ is the level of the series of non-zero observations and $l_{z,t-1}\left(1+\epsilon_t\right) = z_t$. A natural assumption about $(1+\epsilon_t)$ is that it is i.i.d. and log-normal with location parameters μ_ϵ and σ^2_ϵ: $(1+\epsilon_t) \sim \log\mathcal{N}(\mu_\epsilon, \sigma^2_\epsilon)$. The general properties of the ETS(M,N,N) model with the proposed assumptions are discussed briefly in Appendix A.

The statistical model (12) is useful because it allows estimating all the parameters via likelihood maximisation. The concentrated log-likelihood function for the model (12) in case of log-normal distribution of error has the following form (see Appendix B for the derivation):

$$\begin{aligned} \ell(\theta, \hat{\sigma_\epsilon}^2 | Y) = & -\frac{T_1}{2}\left(\log(2\pi e) + \log(\hat{\sigma_\epsilon}^2)\right) - \sum_{o_t=1} \log(z_t) \\ & + \sum_{o_t=1} \log(\hat{p}_t) + \sum_{o_t=0} \log(1-\hat{p}_t) \end{aligned} , \quad (13)$$

where θ is the vector of parameters to estimate (initial values and smoothing parameters), T_1 is the number of non-zero observations, $\hat{\sigma_\epsilon}^2 = \frac{1}{T_1}\sum_{o_t=1} \log^2(1+\epsilon_t)$ is the variance of the one-step-ahead forecast error for the demand sizes and $\hat{p}_t$ is the estimated probability of a non-zero demand at time t. This likelihood function allows estimation of the parameters of iETS models and implementation of model selection even between different intermittent and conventional ETS models.

The only variable that still needs to be estimated is the probability p_t, which can be modelled in different ways. In the simplest case it can be assumed that it is fixed, meaning that:

$$o_t \sim \text{Bernoulli}(p). \quad (14)$$

In more complicated cases it may vary in time, leading to a Croston's style approach:

$$o_t \sim \text{Bernoulli}\left(\frac{1}{1+q_t}\right) \quad (15)$$

or an approach in the TSB style:

$$o_t \sim \text{Bernoulli}(p_t). \quad (16)$$

All these cases and their properties are discussed in the following sections.

In order to distinguish intermittent state-space model from the conventional one we use the letter 'i'. We use a subscript in order to distinguish the three cases of demand occurrence variable discussed above. We denote the model (12) with the case (14) as iETS$_F$, the model with the case (15) as iETS$_I$ (acknowledging that the occurrence part is modelled via demand intervals) and the model with the case (16) as iETS$_P$ (pointing out that the probability is modelled directly). We drop the part denoting the type of ETS model used for demand sizes, implying that the ETS(M,N,N) is the standard model for demand sizes.

In this paper we discuss four types of models: with fixed, Croston's, TSB probability and the one selected automatically between the three. The model without any specified probability does not have a subscript.

3.3. iETS$_F$ – the model with fixed probability

This is the simplest model. It is formulated in the following system of equations:

$$\begin{aligned} & y_t = o_t l_{z,t-1} (1 + \epsilon_t) \\ & l_{z,t} = l_{z,t-1}(1 + \alpha_z \epsilon_t) \\ & o_t \sim \text{Bernoulli}(p) \\ & (1 + \epsilon_t) \sim \log\mathcal{N}(\mu_\epsilon, \sigma_\epsilon^2) \end{aligned} . \tag{17}$$

In this model we assume that the probability of demand occurrence is fixed. The conditional expectation of the occurrence variable o_t can be then calculated as:

$$\mu_{o,t+h|t} = p. \tag{18}$$

The conditional variance of demand occurrence does not change over time as well and is equal to:

$$\sigma^2_{o,t+h|t} = p(1-p). \tag{19}$$

Finally the conditional expectation and variance for intermittent demand can be calculated using (see derivations in Appendix C):

$$\begin{aligned} & \mu_{y,t+h|t} = p\mu_{z,t+h|t} \\ & \sigma^2_{y,t+h|t} = p\sigma^2_{z,t+h|t} + p(1-p)\mu^2_{z,t+h|t} \end{aligned} . \tag{20}$$

In order to estimate this model we use the log-likelihood function (13), which with the assumption of fixed probability becomes simpler. The probability p can then be estimated via the maximisation of the likelihood:

$$\hat{p} = \frac{T_1}{T}, \tag{21}$$

where T is the number of all the observations. The probability estimate (21) can be inserted in the concentrated log-likelihood (13) and then used in information criteria calculation. The number of parameters of the model (17) (which is needed for information criteria) is equal to four: initial value $l_{z,0}$, smoothing parameter α_z, variance σ^2_ϵ and the probability p. The location parameter μ_ϵ is usually assumed to be equal to zero, so that the underlying normal distribution has zero mean. However if the parameter is not equal to zero, then it should be taken into account, increasing the number of parameters by one.

3.4. iETS$_I$ – the model with Croston's probability

It is assumed in Croston's method that the probability of occurrence p_t does not change between demand sizes that we observe, and that the intervals q_t between demands are inversely proportional to the probability. Taking this into account, the occurence part of iETS$_I$ in Croston style can be formulated as:

$$\begin{matrix} o_t \sim \text{Bernoulli}\left(\frac{1}{1+q_t}\right) \\ q_t = l_{q,t-1}(1+\epsilon_{q,t}) \\ l_{q,t} = l_{q,t-1}(1+\alpha_q \epsilon_{q,t}) \\ (1+\epsilon_{q,t}) \sim \log\mathcal{N}(\mu_q, \sigma^2_q) \end{matrix}, \tag{22}$$

where $l_{q,t}$ is the level component for intervals between demands, α_q is the smoothing parameter and $\epsilon_{q,t}$ is the error term of demand intervals. The error term implies that the demand intervals have their own stochastic nature, so even if the level of intervals is fixed, the actual observed intervals may have their own variability. Note that the interval variable q_t is continuous in the model (22), distributed log-normally and defined on $(0, \infty)$. But in many applications it is often the case that we can only observe integer-valued intervals $\tilde{q}_t$. So we deal with discretisation of a continuous unobservable variable due to the process of measurement of data. This imposes some restrictions on the model (22), because the observed $\tilde{q}_t$ will have a discrete log-normal distribution and will be represented by integer numbers greater than zero with the smallest value of $\tilde{q}_t = 1$ in contrast with $(1 + q_t)$ always being greater than one. So the observed intervals $\tilde{q}_t$ can be set equal to $\lceil q_t \rceil$. This means that the first equation in (22) should be amended in order to make the model more realistic:

$$o_t \sim \text{Bernoulli}\left(\frac{1}{\lceil q_t \rceil}\right), \tag{23}$$

At the same time we argue that the continuous log-normal distribution can be used as a good approximation of discrete log-normal in this situation. So finally the model that underlies Croston's method and can be estimated on data collected in discrete time has the following form:

$$
\begin{array}{l}
y_t = o_t l_{z,t-1} (1 + \epsilon_t) \\
l_{z,t} = l_{z,t-1}(1 + \alpha_z \epsilon_t) \\
(1 + \epsilon_t) \sim \log\mathcal{N}(\mu_\epsilon, \sigma_\epsilon^2) \\
o_t \sim \text{Bernoulli}\left(\frac{1}{\lceil q_t \rceil}\right) \\
q_t = l_{q,t-1}(1 + \epsilon_{q,t}) \\
l_{q,t} = l_{q,t-1}(1 + \alpha_q \epsilon_{q,t}) \\
(1 + \epsilon_{q,t}) \sim \log\mathcal{N}(\mu_q, \sigma_q^2)
\end{array} . \tag{24}
$$

It is also important to note that although the intervals q_t may vary in time on each observation, influencing the corresponding probability p_t, iETS$_I$ model cannot be estimated when demand is zero. So during the estimation of the model it is assumed that the states of q_t do not change between demand occurrences. This may be an artificial assumption but it follows logically from the original Croston's method.

Finally it follows straight from the connection between simple exponential smoothing and ETS(M,N,N) that Croston's method (2) has (24) as an underlying statistical model. Note that this implies that the method has an underlying intermittent model with continuous demand sizes.

Overall, the usage of continuous distribution for demand intervals in the model is motivated by the following:

1. We assume that the demand intervals are continuous in their nature, and only their measurement makes them integer;
2. Using the continuous ETS(M,N,N) model for demand intervals shows the connection of the iETS$_I$ model with Croston's method;
3. The model for demand intervals becomes extendable. This means that in future research it is easy to modify the demand intervals part of the model, introducing different time series components and exogenous variables.

Taking into account the discussion of sample paths of ETS(M,N,N) in Appendix A based on (A.1), it can be noted that iETS$_I$ model implies one of the two cases:

1. If $\exp\left(\mu_q + \frac{\sigma_q^2}{2}\right) \leq 1$, then the sample path of q_t will converge to zero (either asymptotically or almost surely), meaning that the probability

of occurrence will converge to one. So in this case the demand changes from slow moving to fast moving.

2. If $\exp\left(\mu_q + \frac{\sigma_q^2}{2}\right) > 1$, then q_t will diverge, meaning that the probability of occurrence becomes close to zero, which characterises the obsolescence of a product.

Although the latter is implied by the model, it cannot be correctly estimated because of the aforementioned updating properties of Croston's method. This means that iETS$_I$ model should be more efficient for the former cases.

We can also calculate conditional (on the information available on the observation t) expectation and variance of the model, which depend on value of q_t. The former is straightforward and is equal to:

$$\mu_{o,t+h|t} = E\left(p_{t+h|t}\right) = E\left(\frac{1}{q_{t+h|t}}\right). \tag{25}$$

Syntetos and Boylan (2001, 2005) showed that the conditional expectation in Croston's method is biased and proposed a correction, but we do not incorporate it here for reasons discussed later in Section 3.6. Knowing the value (25) and using the Bernoulli distribution assumption, the conditional variance of the occurrence part can be calculated as:

$$\sigma^2_{o,t+h|t} = \mu_{o,t+h|t}(1 - \mu_{o,t+h|t}). \tag{26}$$

Because the model underlying the occurrence part of Croston's method is ETS(M,N,N), its parameters can be estimated by maximising the likelihood function of the log-normal distribution. In order to estimate the likelihood of the final intermittent demand model (13), we would need to derive the density function for p_t. However it can be shown (see Appendix D) that for the estimation of the occurrence part of the model (24), a likelihood, derived from the assumption of log-normal distributions of $(1 + \epsilon_{q,t})$, can be used directly:

$$\ell(\theta, \hat{\sigma}_\epsilon{}^2|Y) = -\frac{T_q}{2}\left(\log(2\pi e) + \log(\hat{\sigma}_{\epsilon_q}{}^2)\right) - \sum_{o_t=1} \log(q_t), \tag{27}$$

where $\hat{\sigma}^2_q = \frac{1}{T_q}\sum_{t=1}^{T_q}(1 + \epsilon_{q,t})$ and T_q is the number of intervals between nonzero demands. After that the conditional one-step-ahead expectation

(25) for each observation t can be inserted in (13) instead of $\hat{p}_t$ in order to obtain the following concentrated likelihood:

$$\begin{aligned} \ell(\theta, \hat{\sigma}^2|Y) = -\frac{T_1}{2}\left(\log(2\pi e) + \log(\hat{\sigma}^2)\right) - \sum_{o_t=1} \log(z_t) \\ +\sum_{o_t=1} \log\left(\hat{p}_t\right) + \sum_{o_t=0} \log\left(1-\hat{p}_t\right) \end{aligned} . \tag{28}$$

This simplifies the estimation process of the iETS$_I$ model as it can now be done in two steps:

1. Estimation of parameters of the demand occurrence part of the model using (27);
2. Estimation of parameters of the demand size part of the model using (28).

In order to calculate information criteria we need to know the number of parameters used in the model. While the demand sizes part of the model is still the same, the occurrence part now has a smoothing parameter and its own variance along with the initial value. So the number of parameters in (24) is equal to six: initial value $l_{z,0}$, smoothing parameter α_z, variance ${\hat{\sigma}_\epsilon}^2$, initial value $l_{q,0}$, the smoothing parameter α_q and the variance $\hat{\sigma}_q^2$.

Note that any multiplicative ETS model could potentially be used instead of ETS(M,N,N) in the demand occurrence part of (24), including models with exogenous variables. This enlarges the spectrum of potential intermittent demand models. However we do not aim to study all these models in this paper.

3.5. iETS$_P$ – the model with TSB probability

Similarly to Croston's method, it is assumed in TSB that the probability of demand occurrence may vary in time, which means that it has its own conditional expectation and variance. In order to model this behaviour we use a Beta distribution for the probability of occurrence. This means that we need to deal with a compound Beta-Bernoulli distribution:

$$o_t \sim \text{Beta-Bernoulli}\left(a_t, b_t\right), \tag{29}$$

where a_t and b_t are Beta shape parameters for the left and right sides of the distribution respectively. The Beta-Bernoulli distribution is a special case of the Beta-Binomial distribution, which has been shown to perform well in intermittent demand forecasting (Dolgui and Pashkevich, 2008a,b). However, in previous papers it was assumed that the parameters of the Beta distribution do not vary in time, while our modification assumes that both parameters of the Beta distribution may change in time. In order to model this we propose using two ETS(M,N,N) models, which leads to the following large state-space model:

$$\begin{array}{l} y_t = o_t l_{z,t-1} \left(1 + \epsilon_t\right) \\ l_{z,t} = l_{z,t-1}(1 + \alpha_z \epsilon_t) \\ (1 + \epsilon_t) \sim \log\mathcal{N}(\mu_\epsilon, \sigma_\epsilon^2) \\ o_t \sim \text{Beta-Bernoulli}\left(a_t, b_t\right) \\ a_t = l_{a,t-1} \left(1 + \epsilon_{a,t}\right) \\ l_{a,t} = l_{a,t-1}(1 + \alpha_a \epsilon_{a,t}) \\ (1 + \epsilon_{a,t}) \sim \log\mathcal{N}(\mu_a, \sigma_a^2) \\ b_t = l_{b,t-1} \left(1 + \epsilon_{b,t}\right) \\ l_{b,t} = l_{b,t-1}(1 + \alpha_b \epsilon_{b,t}) \\ (1 + \epsilon_{b,t}) \sim \log\mathcal{N}(\mu_b, \sigma_b^2) \end{array}, \tag{30}$$

where $l_{a,t}$ and $l_{b,t}$ are levels for each of the shape parameters, $\epsilon_{a,t}$ and $\epsilon_{b,t}$ are mutually independent error terms and α_a and α_b are the smoothing parameters. Similarly to previous models, error terms are distributed log-normally with location parameters μ_a, μ_b, σ_a^2 and σ_b^2. This guarantees that both shape parameters are positive.

Using the properties of sample paths discussed in Appendix A, we can show that (30) allows modelling the following four situations:

1. If $\exp\left(\mu_a + \frac{\sigma_a^2}{2}\right) \leq 1$ and $\exp\left(\mu_b + \frac{\sigma_b^2}{2}\right) \leq 1$, then the sample paths of both a_t and b_t will converge to zero (either asymptotically or almost surely), in which case the Beta distribution becomes equivalent to Bernoulli with $p = 0.5$.
2. If both $\exp\left(\mu_a + \frac{\sigma_a^2}{2}\right) > 1$ and $\exp\left(\mu_b + \frac{\sigma_b^2}{2}\right) > 1$, then a_t and b_t will diverge, meaning that distribution of probability p_t is concentrated around 0.5. This means once again that asymptotically $p = 0.5$.
3. If $\exp\left(\mu_a + \frac{\sigma_a^2}{2}\right) \leq 1$ while $\exp\left(\mu_b + \frac{\sigma_b^2}{2}\right) > 1$, then the sample path of a_t will converge to zero, while the sample path of b_t will diverge. This corresponds to a situation of product obsolescence, because the Beta distribution becomes degenerate with all the values concentrated around zero.

4. If $\exp\left(\mu_a + \frac{\sigma_a^2}{2}\right) > 1$ while $\exp\left(\mu_b + \frac{\sigma_b^2}{2}\right) \leq 1$, then the sample path of a_t will diverge, while the sample path of b_t will converge to zero. In this case the Beta distribution becomes degenerate with values concentrated around one, which means that a product switches from slow moving to fast moving.

So, in general, model (30) underlies a wide variety of real life processes. However, the TSB method is underpinned by a more specific model, because it does not have the part corresponding to b_t. In order to derive the connection between the model (30) and the TSB method we need to impose the following restriction:

$$a_t + b_t = 1, a_t \in (0, 1) \tag{31}$$

This means that instead of using two models for shape parameters we can use only one:

$$\begin{aligned} a_t &= l_{a,t-1} (1 + \epsilon_{a,t}) \\ l_{a,t} &= l_{a,t-1}(1 + \alpha_a \epsilon_{a,t}) \end{aligned}. \tag{32}$$

In turn, this means that the expectation of the Beta-distributed probability p_t can be simplified to:

$$\mathrm{E}(p_t) = \mathrm{E}\left(\frac{a_t}{a_t + b_t}\right) = \mathrm{E}(a_t) = l_{a,t-1}. \tag{33}$$

The very same expectation is calculated in the TSB method (7), which implies that the model (32) underlies the occurrence part of the TSB method, because ETS(M,N,N) underlies the simple exponential smoothing method. However there is one important element – this connection implies that $a_t = o_t$ and $b_t = 1 - o_t$ – which is not realistic, because the density function of the Beta distribution is equal to zero for cases of $a_t = 0$ or $b_t = 0$. This means that the model (30) will underlie TSB only when the difference between a_t and o_t is infinitesimal. In order to estimate this model, we can introduce the following approximation for a_t:

$$a_t = o_t(1 - 2\kappa) + \kappa, \tag{34}$$

where κ is a very small number (for example, $\kappa = 10^{-10}$). This modification is artificial but it helps estimation of the model. It is worth stressing that κ is not a natural element of the model and does not affect the simulated time series paths. Its only purpose is to make model estimable.

So, summarising all of the derivations above, the following model underlies TSB:

$$\begin{aligned}
& y_t = o_t l_{z,t-1} (1 + \epsilon_t) \\
& l_{z,t} = l_{z,t-1}(1 + \alpha\epsilon_t) \\
& (1 + \epsilon_t) \sim \log\mathcal{N}(\mu_\epsilon, \sigma_\epsilon^2) \\
& o_t \sim \text{Beta-Bernoulli}\,(a_t, 1 - a_t), \\
& a_t = l_{a,t-1} (1 + \epsilon_{a,t}) \\
& l_{a,t} = l_{a,t-1}(1 + \alpha_a \epsilon_{a,t}) \\
& (1 + \epsilon_{a,t}) \sim \log\mathcal{N}(\mu_a, \sigma_a^2)
\end{aligned} \tag{35}$$

where (34) is used purely for model estimation purposes. The model (35) will have only two sample paths cases out of four, corresponding to items (3) and (4) in the list above. So this model is suitable for modelling demand for products that either become obsolete or fast moving. However there is a potential problem with the model (35) in the case (4), because a_t can become greater than one, implying that $b_t = 1 - a_t < 0$. In this situation the model does not make sense. So iETS$_P$ model (35) implies that there is only one sensible situation, corresponding to item (3) of the list: when a_t decreases and converges almost surely or asymptotically to zero, while b_t converges to one.

Once again we may use a likelihood function for estimation of the model (35). It can be derived using the same likelihood as in (13) (see Appendix E for the details):

$$\begin{aligned}
\ell(\theta, \hat{\sigma_\epsilon}^2 | Y) = \; & -\frac{T_1}{2} \left(\log(2\pi e) + \log(\hat{\sigma_\epsilon}^2)\right) - \sum_{o_t=1} \log(z_t) \\
& + \sum_{o_t=1} \log(l_{a,t-1}) + \sum_{o_t=0} \log(1 - l_{a,t-1})
\end{aligned} . \tag{36}$$

Taking into account that demand sizes are independent of demand occurrences in the proposed model, the latter can be optimised separately by minimising the following cost function (instead of maximising the compound Beta-Bernoulli likelihood directly):

$$CF = -\sum_{o_t=1} \log\left(l_{a,t-1}\right) - \sum_{o_t=0} \log\left(1 - l_{a,t-1}\right). \tag{37}$$

So the iETS$_P$ model should be estimated using cost function (37) rather than any other.

The conditional mean and variance of the demand occurrence part of the $iETS_P$ model are much simpler to calculate than in $iETS_I$ model. They are:

$$\begin{aligned} \mu_{o,t+h|t} &= l_{a,t} \\ \sigma^2_{o,t+h|t} &= l_{a,t}(1-l_{a,t}) \end{aligned}. \quad (38)$$

Finally, due to (37), there is no need to estimate the variance of the occurrence part of the model (35). This means that the $iETS_P$ model has only five parameters: initial value $l_{z,0}$, smoothing parameter α_z, variance $\hat{\sigma_\epsilon}^2$, initial value $l_{a,0}$ and the smoothing parameter α_a.

Once again any other multiplicative ETS model can be used instead of ETS(M,N,N) in (35), which leads to completely new types of models. But this is yet another potential research direction.

3.6. Conditional values and prediction intervals for iETS models

One of the advantages of statistical models, as discussed in section 3.1 is the ability to work with distributions of variables rather than point values. For intermittent state-space models, there are some peculiarities that need to be taken into account, which are mainly caused by the assumption of log-normal distribution of residuals.

It has already been discussed that the mean of the log-normal distribution can be calculated based on location parameters μ_ϵ and σ^2_ϵ using (A.1):

$$E(1+\epsilon_t) = \exp\left(\mu_\epsilon + \frac{\sigma^2_\epsilon}{2}\right).$$

However in cases of skewed distributions, the median value is usually considered to be more robust and useful than the mean. It can be shown that the conditional median of $z_{t+h|t}$ can be calculated using (see Appendix F):

$$\text{Md}(z_{t+h|t}) = l_{z,t}. \quad (39)$$

This is exactly the same forecast as in the conventional ETS(M,N,N) model. However because of the different assumption about the error term, the value (39) in Hyndman et al. (2008) corresponds to the conditional mean. The difference between the conditional mean and median will be negligible if σ^2_ϵ in (A.1) is close to zero. Recalling that ETS(M,N,N) is also used for demand occurrences modelling, $\mu_{o,t+h|t}$ will also correspond to the conditional median value rather than the mean for o_t. If a researcher needs the conditional mean, it can be calculated using (A.1). However this means that the models will produce different point forecasts than Croston's method and TSB. Furthermore, if (A.1) is used for the final calculation, the forecast trajectory may demonstrate increase over time, because the variance in (A.1) typically increases with the increase of the forecast horizon. That is why this will not be discussed further in this paper.

The formula (39) also implies that Croston's method in our formulation does not need any bias correction, because $\text{Md}\left(\frac{1}{x}\right) = \frac{1}{\text{Md}(x)}$.

All of the above also means that the final forecasts produced by all the models discussed in this paper are not mean forecasts, but a multiplication of median forecasts for demand sizes and the median of demand occurrence parts. Although this is not what is usually used in forecasting, we argue that this is not a critical issue, because for the typical task of inventory management it is more important to have a distribution of values rather than a single value.

In order to calculate prediction intervals for intermittent state-space models, the cumulative distribution function (CDF) can be used:

$$F(y_{t+h} < x) = \mu_{o,t+h|t} F_h(z_{t+h} < x) + (1 - \mu_{o,t+h|t}), \tag{40}$$

where $F_h(z_{t+h})$ is the h-steps ahead CDF for z_{t+h} and x is the value of the desired quantile of the distribution. In order to find the appropriate x for intervals of a desired width, an optimisation procedure can be used for each horizon h. In this procedure the parameters of the distribution for each step h and x are inserted in (40) and the probability $F(y_{t+h} < x)$ is then compared with the desired level $1 - \alpha$. The procedure continues until the difference between $F(y_{t+h} < x)$ and $1 - \alpha$ becomes infinitesimal.

These prediction intervals will in general be asymmetric, because of the log-normality assumption of residuals. One of the nice properties of the proposed iETS model is that it allows producing meaningful one-sided intervals, which is important for safety stock calculation. In order to do that the upper quantile is calculated for $1 - \alpha$ rather than $1 - \alpha/2$.

3.7. Integer state-space model

The integer iETS model is more complicated than the continuous model (9) and it has two important aspects that distinguish it from its counterpart.

First, conditional expectation and variance cannot be analytically derived for this model. However, they can be both calculated via simulations. In order for the model to be consistent with the other models discussed in this paper, the median demand sizes need to be taken during the simulation instead of mean for the final value of point forecasts. Simulations can also be used for the calculation of quantiles of distribution for the prediction intervals construction. However, we can use a simplification, which still allows using analytical derivations instead of simulations for both point forecasts and prediction intervals. This simplification is based on the following equality for any quantiles of any distribution (see Appendix G):

$$q_\alpha\left(\lceil z_t \rceil\right) = \lceil q_\alpha(z_t) \rceil , \tag{41}$$

where $q_\alpha(\cdot)$ is α quantile of a random variable. The equality (41) implies that the quantiles of the log-normal distribution imposed by the continuous model underlying demand sizes can be used and then rounded up. As a result there is no need to work directly with the integer model and to produce values via the simulations. Furthermore, the following equality holds for all z_t as well (Appendix G):

$$q_\alpha\left(\lfloor z_t \rfloor\right) = \lfloor q_\alpha(z_t) \rfloor . \tag{42}$$

This means that the decision of whether to round up or round down values can be made by a forecaster depending on their preferences after producing quantiles of the continuous model. The result will be equivalent to using the model with the respective rounding mechanism directly.

Second, the likelihood function for the integer model is more complicated than for the continuous one, because in the former case a multitude of values of z_t correspond to one rounded up value $\lceil z_t \rceil$: all the values in the region $(\lceil z_t \rceil - 1, \lceil z_t \rceil]$ need to be taken into account. This means that the density function cannot be used for the estimation of the likelihood for demand sizes, but the CDF for the interval should be used instead:

$$F_z(\lceil z_t \rceil - 1 < z_t \leq \lceil z_t \rceil) = F_z(z_t \leq \lceil z_t \rceil) - F_z(z_t > \lceil z_t \rceil - 1), \tag{43}$$

where F_z is the CDF of demand sizes (log-normal distribution assumed throughout this paper). Note that with the increase of the level and variance of time series, the distance between $F_z(z_t \leq \lceil z_t \rceil)$ and $F_z(z_t < \lceil z_t \rceil - 1)$ will decrease, and the (43) will asymptotically be equal to the PDF of the same distribution.

The log-likelihood function for the model (10) based on (43) is:

$$\begin{aligned} \ell(\theta, \hat{\sigma_\epsilon}^2|Y) = & \sum_{o_t=1} \log\left(F_z(\lceil z_t \rceil - 1 < z_t \leq \lceil z_t \rceil)\right) \\ & + \sum_{o_t=1} \log(\hat{p}_t) + \sum_{o_t=0} \log(1-\hat{p}_t) \end{aligned} . \qquad (44)$$

The parameters of the model (10) can be estimated directly via maximisation of the concentrated likelihood function (44), which is a more computationally intensive task than the maximisation of the likelihood function for the continuous model (9). However, after the estimation, the likelihood function (44) can be used in model selection. In order to simplify the process we propose to use two-stage optimisation, where in the first stage the parameters of the continuous model are estimated, and in the second the likelihood (44) is used for the correction of the estimated parameters.

3.8. Model selection in the iETS framework

Having the likelihood functions for all three intermittent state-space models (17), (24) and (35) and knowing the number of parameters to estimate, we can calculate any information criterion and use it for model selection. For example, the Akaike Information Criterion can be calculated as:

$$AIC = 2k - 2\ell(\theta, \hat{\sigma_\epsilon}^2|Y), \qquad (45)$$

where for intermittent models k is equal to 4, 5 or 6 (depending on the model underlying the occurrence part) and, for example, for a basic ETS(A,N,N) $k = 3$. So the only difference between intermittent models is in probability modelling. If demand occurs and the probability of occurrence is high, then the likelihood value will be high as well, meaning that the model is more appropriate for the data.

Note that we can also compare conventional non-intermittent ETS models (with trend and seasonality) with the intermittent ones using information criteria. However we do not aim to cover all the possible models in this paper and focus on the level models only.

It is also important to note at this point that having at least four parameters to estimate, iETS models need at least five non-zero demand observations. If for some reason the sample is smaller, then simpler models for demand sizes should be used instead of iETS(M,N,N). For example, using a model with fixed level (setting smoothing parameter α to zero) allows preserving one degree of freedom without substantial loss in generality and fitting the model to data with at least four non-zero observations.

4. Experiments

4.1. Forecasting performance on simulated data

In order to see how the iETS models work and under which conditions, we have conducted a simulation experiment. We have generated data using three integer valued iETS models with:

- fixed probability, where p is chosen randomly from the interval (0, 1),
- Croston's probability with $l_{q,0} = 10$, $\alpha_q = 0.1$,
- TSB probability with $l_{a,0} = 0.5$, $\alpha_a = 0.1$.

The distribution used in all parts of models is $\log\mathcal{N}\left(-\frac{0.16}{2}, 0.16\right)$. These parameters of the distribution lead to almost sure decline of sample paths, discussed in Section 3.2, but also give some variability of data. In case of TSB we had to ensure that a_t converges to zero and that condition (31) is satisfied. That is why we have set a different location parameter $\mu_a = -\frac{0.16}{2} - 0.05$ (which leads to asymptotic convergence to zero) and have selected only those series, where $a_t \in (0, 1)$. We have generated 10,000 time series with 1,000 observations each and made sure that all of them are intermittent and have at least 5 non-zero observations. This was done using the `sim.es()` function from the `smooth` package v2.0.0 for R.

We have then applied several iETS models to randomly selected consequent parts of data containing 24, 48, 96 and 1000 observations, withholding the last 12 observations in order to measure forecasting accuracy. This gives us in-sample sizes of 12, 36, 84 and 988. The first two sizes are typical for supply chain data, the third one is usually considered in practice as a large sample and finally the last is needed in order to see the asymptotic properties of models. The models we used were:

1. ETS(A,N,N), which is needed as a benchmark;
2. iETS_F – the model with fixed probability;
3. iETS_I – the model with demand intervals (Croston probability);
4. iETS_P – the model with varying probability (TSB probability);
5. iETS_A – the model with model selection between the first four models using AIC corrected;
6. int iETS_F – integer counterpart of iETS_F;
7. int iETS_I – integer counterpart of iETS_I;

8. int iETS_P – integer counterpart of iETS_P;
9. int iETS_A – similar to (5), but with rounded up values.

We use the `es()` function from the `smooth` package v2.0.0 for R for the estimation of all these models.

Taking into account the ability of all the models to produce distributions of values, we decided to assess their performance using distributions rather than looking at mean, median or quantile values. In order to do that we use the prediction likelihood score (PLS) discussed in Snyder et al. (2012) and shortly in Kolassa (2016). Other methods are also available, but we think that PLS is easier to interpret and work with. We use the log-likelihood functions (13) and (44) for each of the models and insert the withheld values of generated series instead of z_t and o_t, estimating the joint distribution of 1 to h steps ahead forecasts. Then we can measure if the distributions are estimated correctly by each of the models and decide which of them performs well in each case.

In order to summarise PLS across all the series, we use the arithmetic mean. Higher values of PLS indicate better estimation of the distribution. The results of the simulation are shown in Tables 1 and 2. The first column in both tables shows the number of observations in the sample, and the second column shows the data generating processes used. The other columns show iETS models applied to the data. The continuous iETS models are shown in Table 1, while their integer analogues are summarised in Table 2. The highest values for each sample size and data generating process are shown in bold.

It can be seen from Table 1 that the iETS_F outperformed all the other models on smaller samples (12, 36 and 84 in-sample observations) with iETS_P performing similarly but slightly worse. This can be explained by negligible difference in data characteristics for the three DGPs on small samples – even if probability decreases over time, it does not change substantially on such a small amount of data. It is worth noting that the iETS_P model outperformed all the other models on the large sample of 988 observations. iETS_I performed well only on large samples, still not being able to beat the best model for each of the DGPs, even for the data generated using iETS_I model. This is probably because of the mechanism of updating in the model, when probability is updated only with non-zero demand occurrences. Finally, the selection mechanism does not perform very well on small samples, but it outperforms most of the models on the large sample of 988 observations,

In-sample	DGP	ETS(ANN)	$iETS_F$	$iETS_I$	$iETS_P$	$iETS_A$
12	$iETS_F$	-33.63	**-18.67**	-19.44	-18.72	-26.53
	$iETS_I$	-32.34	**-13.95**	-15.80	-14.00	-29.60
	$iETS_P$	-37.06	**-18.13**	-19.94	-18.15	-34.25
36	$iETS_F$	-33.27	**-19.19**	-19.76	-19.30	-19.97
	$iETS_I$	-30.44	**-11.65**	-12.47	-11.72	-16.18
	$iETS_P$	-33.23	**-12.20**	-13.16	-12.28	-16.22
84	$iETS_F$	-32.79	**-19.61**	-19.90	-19.65	-19.75
	$iETS_I$	-29.56	**-11.20**	-11.64	-11.24	-12.14
	$iETS_P$	-29.38	**-8.04**	-8.61	-8.06	-10.36
988	$iETS_F$	-34.20	-19.01	-19.05	**-18.75**	**-18.75**
	$iETS_I$	-30.76	-13.66	-13.26	**-13.11**	-13.12
	$iETS_P$	-22.62	-1.14	-0.67	**-0.31**	**-0.31**

Table 1: PLS values for continuous models applied to simulated data.

where the difference between data produced by different models becomes more substantial.

Table 2 shows similar results concerning the performance of intger $iETS_F$, $iETS_I$ and $iETS_P$. However the model selection mechanism for integer models does not seem to work for small samples. The important thing to note is that all the integer valued models perform worse than their continuous counterparts on all the samples and all the DGPs. For example, continuous $iETS_F$ applied to data generated from fixed probability model with sample size of 84 has PLS of -19.61 (Table 1), while its integer counterpart produced PLS=-21.69 (Table 2). This shows that although the integer valued models are designed to work on this data, they produce less accurate distributions than their continuous analogues. We will investigate this effect further in the next section on real data.

In addition we analysed the percentage of selected models out of all the cases for the Auto for each of the DGPs. $iETS_F$ and ETS(A,N,N) were selected for the majority of cases. This is the expected result, taking into account that intermittent models have more parameters than ETS(A,N,N). $iETS_I$ model has the largest number of parameters and it has not been selected at all, probably because $iETS_P$ is able to perform better, but with fewer parameters to estimate.

In-sample	DGP	iETS_F	iETS_I	iETS_P	iETS_A
12	iETS_F	**-22.71**	-23.49	-22.77	-30.52
	iETS_I	**-16.03**	-17.88	-16.07	-31.89
	iETS_P	**-18.12**	-19.92	-18.14	-34.32
36	iETS_F	**-22.11**	-22.68	-22.22	-29.95
	iETS_I	**-12.95**	-13.78	-13.02	-28.57
	iETS_P	**-12.29**	-13.25	-12.37	-31.60
84	iETS_F	**-21.69**	-21.97	-21.73	-31.10
	iETS_I	**-12.12**	-12.56	-12.16	-29.28
	iETS_P	**-8.13**	-8.70	-8.14	-29.70
988	iETS_F	-21.79	-21.83	**-21.53**	-34.41
	iETS_I	-14.74	-14.34	**-14.19**	-31.77
	iETS_P	-1.15	-0.68	**-0.32**	-25.33

Table 2: PLS values for integer models applied to simulated data.

Overall we have expected that each of the applied models would perform better on the data generated from the respective processes, but our experiment shows that this is not true. The model with fixed probability performs better than all the others on small samples and may be preferred to more complicated ones. However iETS_P performed very well in many cases and especially well on large samples. Taking its overall good performance and flexibility, we would recommend it as a basic model for intermittent demand forecasting. The iETS_I model did not perform well in our experiment. But this does not mean that it is not applicable at all. It may perform better on time series with slowly increasing probability (for example, with lower variance σ_q^2). Finally, we found that integer valued iETS models perform worse than their continuous counterparts.

4.2. Real time series experiment

In order to examine the performance of the proposed intermittent state-space models, we conduct experiment on two datasets.

The first is 3000 real time series of automotive spare parts. This dataset originates from Syntetos and Boylan (2005) and was also used in Kourentzes (2014). This is monthly time series, containing 24 observations. We withheld 5 observations from each time series for measuring forecasting accuracy.

The second dataset is Royal Air Force data, which contains 5000 real time series (Eaves and Kingsman, 2004). Each of the time series in this dataset has 84 observations. We withheld 12 observations and use them in order to measure forecasting accuracy of tested models.

We have used the same set of models as in Section 4.1 in this experiment and added the following benchmark methods and filters implemented in the `tsintermittent`, v2.0 package for R:

1. Hurdle shifted Poisson filter (denoted "HSP") discussed in Snyder et al. (2012) implemented in `hsp()` function;
2. Negative Binomial filter (denoted "NegBin") from Snyder et al. (2012), implemented in the function `negbin()`.
3. TSB method implemented in `tsb()` function;
4. Croston's method implemented in `crost()` function;
5. SBA method implemented in `crost()` function.

We can calculate PLS only for iETS model, HSP and NegBin filters. The last three methods produce expected values only and the distribution of values cannot be estimated correctly for them. So we measure accuracy of point forecasts of all the competing methods and models as well. In order to measure performance of all of them we use the following error metrics, discussed in Kourentzes (2014) and Petropoulos and Kourentzes (2015):

- sME – scaled Mean Error;
- sMSE – scaled Mean Squared Error;
- sAPIS – scaled Absolute Periods in Stock.

We have calculated mean and median values of these errors across all the series and summarised them in two tables. In cases when data had no variability in-sample and when models produced unrealistic forecasts, PLS returned infinite values. So we excluded those cases, when calculating PLS, which left us with 2785 series instead of 3000 in Automotive data and 4365 series instead of 5000 in RAF data.

The results of the experiment on Automotive data are given in the Table 3, which shows that all the methods performed very similarly. Continuous iETS$_F$ appears to be more accurate than the other models in median values of sMSE and sAPIS and in mean values of sAPIS, while iETS$_P$ was very close to it, producing almost identical forecasts. The continuous iETS$_I$ is the least

Methods	Mean values				Median Values			
	sME	sMSE	sAPIS	PLS	sME	sMSE	sAPIS	PLS
ETS(A,N,N)	-0.04	1.07	6.84	-11.99	-0.05	0.72	5.59	-10.15
iETS_F	0.10	1.03	**6.48**	**-6.63**	0.09	**0.66**	**5.19**	**-7.26**
iETS_I	**0.02**	1.04	6.55	-6.97	**0.01**	0.69	5.21	-7.62
iETS_P	0.10	1.03	6.49	-6.68	0.08	**0.66**	**5.19**	-7.29
iETS_A	0.06	1.04	6.57	-8.68	0.04	**0.66**	5.34	-8.38
int iETS_F	0.11	1.07	6.80	-6.95	0.09	0.69	5.40	-7.99
int iETS_I	0.03	1.08	6.90	-7.29	0.02	0.71	5.47	-8.37
int iETS_P	0.10	1.07	6.80	-7.00	0.09	0.69	5.38	-8.02
int iETS_A	-0.05	1.09	7.00	-11.01	-0.05	0.72	5.76	-10.53
HSP	-0.07	1.07	6.85	-14.49	-0.09	0.73	5.51	-10.11
NegBin	-0.03	1.02	6.57	-11.30	-0.05	0.71	5.48	-9.96
TSB	-0.03	**1.01**	6.49	NA	-0.06	0.70	5.41	NA
Croston	-0.04	1.02	6.54	NA	-0.08	0.70	5.40	NA
SBA	-0.04	1.02	6.52	NA	-0.06	0.70	5.33	NA

Table 3: Automotive data results.

biased model judging by both mean and median values of sME. The TSB method was more accurate on mean value of sMSE. The differences between the methods and models do not look substantial. As for the PLS, it is worth noting that although the data we deal with is count, the continuous models outperform consistently integer models both in mean and median values. It is also worth pointing out that almost all the iETS models outperform both Hurdle shifted Poisson and Negative Binomial filters of Snyder et al. (2012) in the majority of measures.

In order to determine if the differences between the models are statistically significant, we have conducted a Nemenyi test (Demšar, 2006) on PLS values. The results of this test are shown in Figure 1. The ranking was done so that the model with the highest PLS would have the score of 1 and the model with the lowest PLS would have the score of 10. The Y-axis in Figure 1 shows average ranks for each of the models. The vertical lines in the figure show the groups of models, in which the difference between the ranks is statistically insignificant. The significance level used in this experiment is 5%.

As we see from the graph, the continuous iETS_F performed significantly better than all the other models. The second best model is continuous iETS_P.

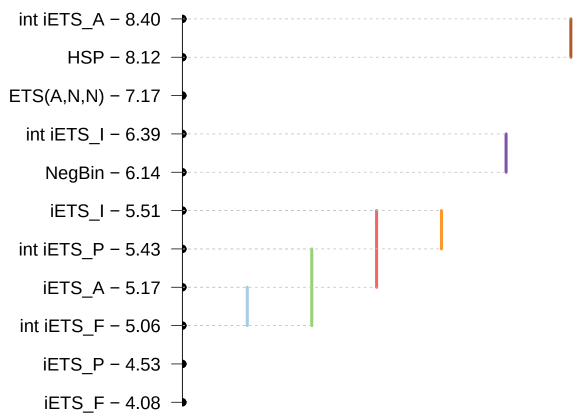

Figure 1: Nemenyi test for models applied to automotive data.

The difference between integer iETS$_F$, continuous iETS$_A$ with model selection and integer iETS$_P$ is statistically insignificant. The HSP filter performed significantly worse than all the models outperforming only the integer iETS$_A$ with model selection. At the same time, the Negative Binomial filter performs significantly better than the Posson filter (which agrees with the finding of Snyder et al., 2012) and some of the iETS models. It is at least as good as the integer iETS$_I$ model, but significantly worse than the integer iETS with fixed and TSB probabilities. It also performed worse than all the other continuous iETS models. It is also worth pointing out that the model selection in case of integer model does not perform well. So we would advise either using continuous iETS with model selection instead or to use iETS$_F$ or iETS$_P$ for this data.

The results of the experiment for the Royal Air Force data are shown in Table 4.

The differences between the applied models and methods are more substantial on this dataset. The leading model in both sMSE and sAPIS is the continuous iETS$_P$. It outperforms the TSB method (which is optimised and initialised differently), but performs very closely to the model with fixed probability. In terms of PLS, iETS$_F$ and iETS$_P$ perform very similarly, sharing the first place in this competition. Once again the continuous models in general outperform integer ones with the one exception of integer iETS$_A$,

Methods	Mean values				Median Values			
	sME	sMSE	sAPIS	PLS	sME	sMSE	sAPIS	PLS
ETS(A,N,N)	-0.37	11.06	76.17	-24.04	-0.62	**2.25**	81.22	-18.83
iETS$_F$	-0.16	**10.54**	66.63	**-4.53**	-0.39	2.27	59.76	**-4.71**
iETS$_I$	-0.65	11.47	89.27	-4.96	-0.81	3.62	80.20	-5.05
iETS$_P$	-0.16	**10.54**	**66.60**	-4.54	-0.39	**2.25**	59.69	**-4.71**
iETS$_A$	-0.20	10.68	68.24	-7.08	-0.41	2.28	61.24	-5.05
int iETS$_F$	-0.15	10.59	66.94	-5.95	-0.32	2.63	56.88	-4.90
int iETS$_I$	-0.63	11.69	89.22	-6.38	-0.71	3.86	74.68	-5.35
int iETS$_P$	-0.14	10.59	66.88	-5.95	-0.32	2.61	**56.87**	-4.90
int iETS$_A$	**0.02**	11.50	77.66	-23.72	**0.00**	3.48	65.33	-21.46
HSP	-0.50	10.98	81.25	-10.28	-0.76	2.65	87.68	-5.04
NegBin	-0.34	10.63	74.21	-6.16	-0.64	2.30	79.76	-5.03
TSB	-0.31	10.61	73.21	NA	-0.62	2.26	76.55	NA
Croston	-0.37	10.63	74.73	NA	-0.68	2.30	78.16	NA
SBA	-0.36	10.63	74.48	NA	-0.67	2.26	77.36	NA

Table 4: Royal Airforce data results.

which has the least biased forecasts measured by both mean and median sME. The worst performing model (in terms of sME, sMSE and sAPIS) on this dataset is iETS$_I$ (both continuous and integer versions). It even performs worse than Croston's method, which points to the differences at the estimation of parameters. This may be explained by the maximum likelihood estimation of parameters of the model leading to less accurate point forecasts on this dataset than in case with the simple estimation methods (implemented in the `tsintermittent` package). The HSP filter performed similarly to how it performed on a previous dataset, this time slightly outperforming ETS(A,N,N) and outperforming iETS$_I$ model. As for the Negative Binomial filter, it performed better than HSP in all the measures (once again agreeing with Snyder et al., 2012). Finally, the model selection mechanism in case of integer iETS model does not seem to work well.

Following the same procedure as with automotive data, we have conducted a Nemenyi test with the results shown in Figure 2.

The test shows that continuous iETS$_P$ is significantly more accurate than the other models. The integer version of this model is the second best. The worst performing model is the integer iETS(M,N,N) with model selection. As

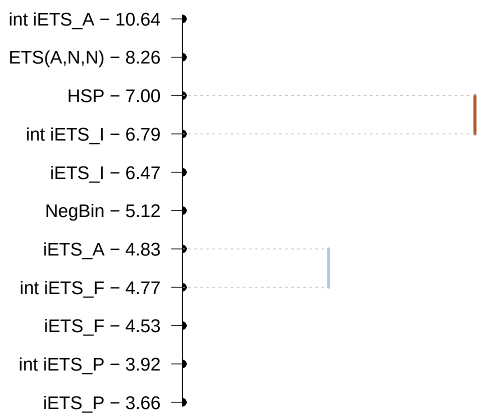

Figure 2: Nemenyi test for models applied to Royal Air Force data.

for the HSP filter, it was the third worst model in the comparison, performing similar to the integer iETS$_I$ model. Finally, Negative Binomial filter outperformed both iETS$_I$ models and HSP, but it could not produce as accurate forecasts as iETS$_F$ and iETS$_P$.

As can be seen from this experiment the proposed intermittent state-space models perform very well and can be applied to real life problems. The iETS$_P$ model seems to be the most robust of all of them. Although we know that the data we deal with is count and that the continuous model is wrong in this case, we found that it is still useful.

5. Conclusions

In this paper, we have proposed a statistical model that underpins both Croston's and the TSB method. These methods are important in a supply chain context, as they are used to inform replenishment and discontinuation decisions. The model also unites intermittent and non-intermittent processes, expanding the Hyndman et al. (2008) taxonomy by the inclusion of intermittent models. This is vital for the forecasting of a wide range of stock keeping units, which may evolve from slow moving to fast moving products (or vice-versa). The model allows for a systematic approach to method selection, rigorous parametrisation, and estimation of upper percentiles of demand.

This paper was focused on the ETS(M,N,N) model and the intermittent equivalent of this model was called iETS(M,N,N). Firstly we have proposed a simple state-space model with fixed probability (denoted as "iETS$_F$"), which is very easy to estimate and use. We have shown that Croston's method has an underlying statistical model (denoted as "iETS$_I$"), which allows the calculation of conditional expectation and variance. After that we have shown that the TSB method also has an underlying statistical model (denoted as "iETS$_P$"), which allows estimation of the model parameters. We have also derived the likelihood functions for all the iETS models, which allow not only obtaining efficient and consistent estimates of parameters, but also selecting between several state-space models. This also includes selecting between intermittent and non-intermittent models, thereby simplifying the forecasting process. We have shown that the forecasts produced by iETS$_I$ and iETS$_P$ correspond to the conditional median of demand sizes rather than the mean, which in case of intermittent data is a useful property. We have also proposed an algorithm of parametric prediction intervals construction using the proposed intermittent state-space model. Finally, we developed integer counterparts of iETS models which address the issue of count data modelling.

We conducted several experiments on simulated and real data. The simulation that we have conducted shows several interesting results. First, it seems that integer iETS models do not perform as accurately as their continuous counterparts. Second, iETS$_F$ and iETS$_P$ work very well on small samples of data generated from different iETS models. Third, iETS$_P$ works better than the other models on large samples, being able to produce the most accurate forecasts for all the DGPs. Fourth, iETS with model selection improves its performance with an increasing sample size. We argue that iETS$_P$ should be preferred to other models on small samples as a more robust and more flexible model. It is able to produce accurate forecasts on a wide variety of time series from different data generating processes.

Finally, the experiment on automotive data and on data from the Royal Air Force shows that the proposed approach is applicable to real life supply chain problems and that the proposed models perform very well on different datasets. They outperformed the existing forecasting methods and several filters previously proposed in the literature. iETS$_P$ generally was one of the best forecasting models on both data sets. We would advise it as one of the most robust models, applicable to wide variety of series.

We should remark that the focus of this paper was on a specific iETS(M,N,N) model, which underlies the two intermittent demand forecasting methods studied in the paper (Croston's and TSB). We simplified the notation for this model in the paper. However, we propose a more detailed one, which acknowledges the flexibility of the proposed approach and the fact that both demand sizes and demand occurrence parts may have their own ETS models (potentially with exogenous variables). So, the model with demand intervals, discussed in the paper can be denoted as iETS(M,N,N)$(M,N,N)_I$, where the letters in the first brackets indicate the type of ETS model for demand sizes and the letters in the second ones indicate the type of ETS model used for demand intervals. Using this notation, new types of models can be studied in future research. For example, model with additive trend in demand sizes and multiplicative trend in time varying probability can be denoted as iETS(M,A,N)$(M,M,N)_P$. This allows extending the Hyndman et al. (2008) taxonomy and opens new avenues for the research.

It is also worth mentioning that the approach of intermittent state-space modelling allows using (for both demand sizes and demand occurrence parts of the model) ETS, ARIMA, regression models or diffusion models, which could be applied to a wide range of time series (not limited with intermittent demand). Studying properties of such models would be another large area of research. The other possible direction of research is the development of a new model for demand occurrence, as both Croston's and TSB mechanisms have their own flaws. Finally, in order to show the connections between the methods and the models, we assumed throughout this paper that demand occurrence and demand size parts are independent. This could be modified in a new model using the state-space approach discussed in the paper.

Appendix A. Properties of ETS(M,N,N) model

The main properties of ETS(M,N,N) are well studied in Akram et al. (2009) and are not discussed here. The important thing to note is that the authors use Kakutani's theorem, showing that if the mean value of $(1+\alpha_z\epsilon_t)$ is equal to one and the distribution is non-degenerate, then the sample path of ETS(M,N,N) tends to converge almost surely to zero. This is based on the assumption of normal distribution with zero mean of ϵ_t, which leads to $\mathrm{E}(1+\epsilon_t) = 1$. However in cases of log-normal distribution of error term, the mean value of $(1+\epsilon_t)$ is in general not equal to one, because of the following connection between the mean of the log-normal distribution with the parameters of the normal distribution:

$$\mathrm{E}(1+\epsilon_t) = \exp\left(\mu_\epsilon + \frac{\sigma_\epsilon^2}{2}\right). \tag{A.1}$$

Even if $\mu_\epsilon = 0$ in the model (12), σ_ϵ^2 is not equal to zero. If μ_ϵ is close to $-\frac{\sigma_\epsilon^2}{2}$, then the sample path of ETS(M,N,N) will converge almost surely to zero as discussed in Akram et al. (2009), because the expected value of $(1+\epsilon_t)$ will be close to one in this case, meaning that $\mathrm{E}(\epsilon_t)$ tends to zero and $\mathrm{E}(1+\alpha\epsilon_t)$ tends to one. In the other case, when $\mu_\epsilon < -\frac{\sigma_\epsilon^2}{2}$, the sample path will converge to zero asymptotically, because $\mathrm{E}(1+\epsilon_t)$ becomes less than one, making $\mathrm{E}(1+\alpha\epsilon_t) < 1$ as well, leading to diminishing values of the level of time series l_t. Finally, when $\mu_\epsilon > -\frac{\sigma_\epsilon^2}{2}$ the sample path will asymptotically diverge from zero, because $\mathrm{E}(1+\alpha\epsilon_t) > 1$, which causes growth of level. This is an important property, because it implies that with different values of μ_ϵ and σ_ϵ^2 the model will behave differently.

The properties of the log-normal distribution and the multiplicative model also restrict the smoothing parameter with the interval [0, 1]. Assuming that the smoothing parameter is always positive, the inequality $(1+\epsilon_t) > 0$ implies that:

$$\begin{aligned} \epsilon_t &> -1 \\ \alpha_z\epsilon_t &> -\alpha_z \\ 1+\alpha_z\epsilon_t &> 1-\alpha_z \end{aligned} \tag{A.2}$$

The ETS(M,N,N) model makes sense only when $1+\alpha_z\epsilon_t > 0$. So, if $\alpha_z > 1$, then $1+\alpha_z\epsilon_t$ may become negative, which breaks the model, because the level may become negative. The model however still makes sense for boundary values of α_z: when $\alpha_z = 0$, the level is not updated, while in the case of $\alpha_z = 1$, the level has the dynamics of a random walk process. The condition $\alpha_z \in [0,1]$ is rather restrictive, because there may be some cases when even with $\alpha_z > 1$ the value of $(1+\alpha_z\epsilon_t)$ will be greater than zero. However it guarantees that the level of time series is always positive whatever the error value is.

Appendix B. Likelihood function for iETS(M,N,N)

The likelihood function for the log-normal distribution $L(\theta, \sigma_\epsilon^2|z_t)$ is well-known and is not presented here. It is worth noting that θ is a vector of all the parameters of the model, $\mu_{z,t|t-1}$ is the conditional mean and σ^2_ϵ is the variance of one-step-ahead forecast error for demand sizes.

There are two cases for the intermittent demand model: when demand occurs and when it does not. In the former case, the probability of obtaining the value y_t is equal to:

$$P(y_t|\theta, \sigma^2_\epsilon, o_t = 1) = L(\theta, \sigma^2_\epsilon|y_t, o_t = 1) = p_t L(\theta, \sigma^2_\epsilon|z_t). \tag{B.1}$$

In the latter case it is just equal to the probability of non-occurrence:

$$L(\theta, \sigma^2_\epsilon|y_t, o_t = 0) = (1 - p_t). \tag{B.2}$$

The likelihood function for all the T observations, which include T_0 cases of non-occurrence and T_1 cases of demand occurrence, is then:

$$L(\theta, \sigma^2_\epsilon|Y) = \prod_{o_t=1} p_t L(\theta, \sigma^2_\epsilon|z_t) \prod_{o_t=0} (1 - p_t), \tag{B.3}$$

where Y is the set of all the variables y_t. Taking the logarithm of (B.3), we obtain:

$$\begin{aligned} \ell(\theta, \sigma^2_\epsilon|Y) = & -\sum_{o_t=1} \left(\log(z_t) + \frac{1}{2}\log(2\pi\sigma^2_\epsilon) + \frac{1}{2}\frac{\left(\log z_t - \log \mu_{z,t|t-1}\right)^2}{\sigma^2_\epsilon} \right) \\ & + \sum_{o_t=1} \log(p_t) + \sum_{o_t=0} \log(1 - p_t) \end{aligned}. \tag{B.4}$$

The variance σ^2_ϵ can be estimated using this likelihood (by taking the derivative of (B.4) with respect to σ^2_ϵ and equating it to zero) and is equal to:

$$\hat{\sigma}^2_\epsilon = \frac{1}{T_1} \sum_{o_t=1} \log^2 (1 + \epsilon_t), \tag{B.5}$$

where $\log(1 + \epsilon_t) = \log z_t - \log \mu_{z,t|t-1}$. In addition, the probability p_t is also not known and needs to be substituted by the estimated value $\hat{p}_t$. All of this leads to the following concentrated log-likelihood:

$$\begin{aligned} \ell(\theta, \hat{\sigma}^2_\epsilon|Y) = & -\frac{T_1}{2}\left(\log(2\pi e) + \log(\hat{\sigma}^2_\epsilon)\right) - \sum_{o_t=1} \log(z_t) \\ & + \sum_{o_t=1} \log(\hat{p}_t) + \sum_{o_t=0} \log(1 - \hat{p}_t) \end{aligned}. \tag{B.6}$$

Appendix C. Conditional variance for iETS$_F$

Knowing that $\sigma^2_{o,t+h|t} = p(1-p)$ for a Bernoulli process and inserting it in (11) leads to:

$$\sigma^2_{y,t+h|t} = p(1-p)\sigma^2_{z,t+h|t} + p(1-p)\mu^2_{z,t+h|t} + p^2\sigma^2_{z,t+h|t}, \tag{C.1}$$

which then can be simplified to:

$$\sigma^2_{y,t+h|t} = p\sigma^2_{z,t+h|t} + p(1-p)\mu^2_{z,t+h|t}. \tag{C.2}$$

Appendix D. Likelihood function for iETS$_I$

In order to derive the likelihood for iETS$_I$ model, the probability density function for $p_t = \frac{1}{1+q_t}$ needs to be derived. This can be done, taking into account that q_t has a log-normal distribution, using the formula:

$$f_y(y) = \left| \frac{d}{dy} g^{-1}(y) \right| f_x(g^{-1}(y)), \tag{D.1}$$

where $y = g(x)$, $x = g^{-1}(y)$ is the inverse function of y and $f_x(\cdot)$ is the density function of x. q_t can be reformulated as $q_t = \frac{1-p_t}{p_t}$, the differential of which is equal to $-\frac{1}{p_t^2}$. Inserting this in (D.1), the density function of p_t is:

$$f(p_t|\theta_q) = \frac{1}{p_t^2} \frac{1}{\frac{1-p_t}{p_t}\sqrt{2\pi\sigma_q^2}} e^{-\frac{\left(\log\left(\frac{1-p_t}{p_t}\right) - \log\mu_{q,t|t-1}\right)^2}{2\sigma_q^2}}, \tag{D.2}$$

which becomes:

$$f(p_t|\theta_q) = \frac{1}{p_t(1-p_t)} \frac{1}{\sqrt{2\pi\sigma_q^2}} e^{-\frac{\left(\log(1-p_t) - \log(p_t) - \log\mu_{q,t|t-1}\right)^2}{2\sigma_q^2}}, \tag{D.3}$$

where θ_q is the vector of parameters relating to the occurrence part of the model. The probability of having an occurrence is now a compound with the following density function:

$$f(o_t = k|\theta_q) = \int_0^1 p_t^k (1-p_t)^{1-k} f(p_t|\theta_q) dp_t, \tag{D.4}$$

where $k = 1$, when demand occurs and $k = 0$ otherwise. The density functions (D.4) and (D.3) can now be inserted in the general likelihood function (B.3), discussed in Appendix B. However there is no point in doing so, and it is sufficient to note that maximisation of the likelihood (B.3) means automatically the maximisation of (D.3). And if we make the substitution $p_t = \frac{1}{q_t}$ in (D.3) we will have the likelihood function for q_t based on the log-normal distribution. This means that the estimation of iETS$_I$ model can be done in two steps: first the occurrences part should be estimated via maximisation of the likelihood function for q_t, then the general model can be estimated using (B.3) and the expected values of probabilities from iETS$_I$ model. This also demonstrates the connection between the optimisation procedure on demand intervals level and on the level of the model as a whole.

Appendix E. Likelihood function for iETS$_P$

The likelihood function for iETS$_P$ resembles the likelihood of the general iETS model. The only difference is in the probability of occurrences. So the concentrated log-likelihood can be written as:

$$\begin{aligned}\ell(\theta, \sigma_\epsilon^2 | Y) = & -\frac{T_1}{2} \left(\log(2\pi e) + \log(\hat{\sigma}_\epsilon^2)\right) - \sum_{o_t=1} \log(z_t) \\ & + \sum_{o_t=1} \log \left(\frac{\mathrm{B}(o_t + a_t, 1 - o_t + b_t)}{\mathrm{B}(a_t, b_t)} \right) + \sum_{o_t=0} \log \left(\frac{\mathrm{B}(o_t + a_t, 1 - o_t + b_t)}{\mathrm{B}(a_t, b_t)} \right),\end{aligned} \tag{E.1}$$

where B(a,b) is the Beta function with parameters a and b. The likelihood (E.1) can be simplified to (taking TSB restriction of $a_t + b_t = 1$):

$$\begin{aligned}\ell(\theta, \sigma_\epsilon^2 | Y) = & -\frac{T_1}{2} \left(\log(2\pi e) + \log(\hat{\sigma}_\epsilon^2)\right) - \sum_{o_t=1} \log(z_t) \\ & + \sum_{o_t=1} \log \left(\frac{\mathrm{B}(1 + a_t, 1 - a_t)}{\mathrm{B}(a_t, 1 - a_t)} \right) + \sum_{o_t=0} \log \left(\frac{\mathrm{B}(a_t, 1 + 1 - a_t)}{\mathrm{B}(a_t, 1 - a_t)} \right).\end{aligned} \tag{E.2}$$

Now we should recall two important properties of the Beta function. They are:

$$\begin{aligned}\mathrm{B}(1 + a, b) &= a \frac{\mathrm{B}(a, b)}{a + b} \\ \mathrm{B}(a, 1 - a) &= \frac{\pi}{\sin(\pi a)}.\end{aligned} \tag{E.3}$$

Using these properties, it can be shown that:

$$\frac{\mathrm{B}(1+a_t, 1-a_t)}{\mathrm{B}(a_t, 1-a_t)} = a_t \frac{\mathrm{B}(a_t, 1-a_t)}{a_t + 1 - a_t} \frac{\sin(\pi a_t)}{\pi} = a_t \frac{\pi}{\sin(\pi a_t)} \frac{\sin(\pi a_t)}{\pi} = a_t. \tag{E.4}$$

Similarly it can be shown that:

$$\frac{\mathrm{B}(a_t, 1+1-a_t)}{\mathrm{B}(a_t, 1-a_t)} = 1 - a_t. \tag{E.5}$$

This means that the log-likelihood function for this model is:

$$\begin{aligned} \ell(\theta, \sigma_\epsilon^2 | Y) = & -\frac{T_1}{2} \left(\log(2\pi e) + \log(\hat{\sigma}_\epsilon^2)\right) - \sum_{o_t=1} \log(z_t) \\ & + \sum_{o_t=1} \log(a_t) + \sum_{o_t=0} \log(1-a_t) \end{aligned} . \tag{E.6}$$

However a_t is unknown and thus should be estimated using ETS(M,N,N). This gives us the following final concentrated log-likelihood:

$$\begin{aligned} \ell(\theta, \hat{\sigma}_\epsilon^2 | Y) = & -\frac{T_1}{2} \left(\log(2\pi e) + \log(\hat{\sigma}_\epsilon^2)\right) - \sum_{o_t=1} \log(z_t) \\ & + \sum_{o_t=1} \log(l_{a,t-1}) + \sum_{o_t=0} \log(1-l_{a,t-1}) \end{aligned} , \tag{E.7}$$

Appendix F. Conditional values of iETS(M,N,N)

If we rewrite the demand size part of the general intermittent state-space model (12) in logarithms:

$$\begin{aligned} & y_t = o_t \left(\log l_{z,t-1} + \log\left(1+\epsilon_t\right)\right) \\ & \log l_{z,t} = \log l_{z,t-1} + \log(1+\alpha_z \epsilon_t) \end{aligned} \tag{F.1}$$

then the measurement equation for y_{t+h} can be written as:

$$y_{t+h} = o_{t+h} \exp\left(\log l_{z,t} + \sum_{j=1}^{h-1} \log(1+\alpha_z \epsilon_{t+h-j}) + \log(1+\epsilon_{t+h})\right). \tag{F.2}$$

The part inside the exponent in (F.2) will have a normal distribution if $(1+\epsilon_t)$ has a log-normal distribution. The conditional expectation and variance of that part will then be:

$$\begin{aligned} & \tilde{\mu}_{z,t+h|t} = \log l_{z,t} \\ & \tilde{\sigma}^2_{z,t+h|t} = \sigma^2_\epsilon \left(1 + \textstyle\sum_{j=1}^{h-1} \sigma^2_\alpha\right), \end{aligned} \tag{F.3}$$

where $\sigma^2_\alpha = \frac{1}{T_1}\sum_{t=1}^{T_1}\log(1+\alpha\epsilon_t)^2$. Knowing these values allows us to calculate the conditional mean, median and variance of z_t:

$$\begin{aligned}
&\mu_{z,t+h|t} = \exp\left(\tilde{\mu}_{z,t+h|t} + \frac{\tilde{\sigma}^2_{z,t+h|t}}{2}\right) \\
&\mathrm{Md}(z_{t+h|t}) = \exp\left(\tilde{\mu}_{z,t+h|t}\right) \\
&\sigma^2_{z,t+h|t} = \left(\exp\left(\tilde{\sigma}^2_{z,t+h|t}\right) - 1\right)\exp\left(2\tilde{\mu}_{z,t+h|t} + \tilde{\sigma}^2_{z,t+h|t}\right)
\end{aligned}, \tag{F.4}$$

where $\mu_{z,t+h|t}$ is the conditional mean, $\mathrm{Md}(z_{t+h|t})$ is the conditional median and $\sigma^2_{z,t+h|t}$ is the conditional variance of z_{t+h}. Note that the smaller the conditional variance $\tilde{\sigma}^2_{z,t+h|t}$ is, the closer the final conditional mean and median are to each other.

It is also worth mentioning that in these derivations we only look at the dynamics of demand sizes, which may change over time between $t+1$ and $t+h$. Thus we ignore the possibility of demand occurrence, which can be done because of the assumption of independence of demand sizes and occurrences.

Appendix G. Quantiles of rounded up random variables

Before proceeding with the proof we need to give the definition of the quantiles of the continuous and rounded up random variables:

$$P(z_t < k) = 1 - \alpha, \tag{G.1}$$

and

$$P(\lceil z_t \rceil \leq n) \geq 1 - \alpha, \tag{G.2}$$

where n is the quantile of the distribution of rounded up values (the smallest integer number that satisfies the inequality (G.2)) and k is the quantile of the continuous distribution of the variable.

In order to prove that $n = \lceil k \rceil$, we need to use the following basic property:

$$\lceil z_t \rceil \leq n \iff z_t \leq n, \tag{G.3}$$

which means that the rounded up value will always be less than or equal to n if and only if the original value is less than or equal to n.

Taking into account (G.3), the probability (G.2) can be rewritten as:

$$P(z_t \leq n) \geq 1 - \alpha. \tag{G.4}$$

Note also that the following is true:

$$P\left(\lceil z_t \rceil \leq n-1\right) = P\left(z_t \leq n-1\right) < 1-\alpha. \tag{G.5}$$

Taking the inequalities into account (G.1), (G.2), (G.4) and (G.5), the following can be summarised:

$$P\left(z_t \leq n-1\right) < P\left(z_t < k\right) \leq P\left(z_t \leq n\right), \tag{G.6}$$

which is possible only when $k \in (n-1, n]$, which means that $\lceil k \rceil = n$. So the rounded up quantile of continuous random variable z_t will always be equal to the quantile of the descritised value of z_t.

It is also worth noting that the same results can be obtained with the floor function instead of ceiling, following the same logic. So the following equation will hold for all z_t as well:

$$q_\alpha\left(\lfloor z_t \rfloor\right) = \lfloor q_\alpha(z_t) \rfloor, \tag{G.7}$$

References

Akram, M., Hyndman, R. J., Ord, J. K., 2009. Exponential smoothing and non-negative data. Australian & New Zealand Journal of Statistics 51 (4), 415–432.

Box, G., Jenkins, G., 1976. Time series analysis: forecasting and control. Holden-day, Oakland, California.

Croston, J. D., 1972. Forecasting and Stock Control for Intermittent Demands. Operational Research Quarterly 23 (3), 289.

Demšar, J., 2006. Statistical Comparisons of Classifiers over Multiple Data Sets. Journal of Machine Learning Research 7, 1–30.

Dolgui, A., Pashkevich, M., 2008a. Extended beta-binomial model for demand forecasting of multiple slow-moving inventory items. International Journal of Systems Science 39 (7), 713–726.

Dolgui, A., Pashkevich, M., 2008b. On the performance of binomial and beta-binomial models of demand forecasting for multiple slow-moving inventory items. Computers & Operations Research 35 (3), 893–905.

Eaves, A., Kingsman, B., 2004. Forecasting for the ordering and stock-holding of spare parts. Journal of the Operational Research Society 55 (4), 431–437.

Hyndman, R. J., Koehler, A. B., Ord, J. K., Snyder, R. D., 2008. Forecasting with Exponential Smoothing. Springer Series in Statistics. Springer Berlin Heidelberg.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., Grose, S., 2002. A state space framework for automatic forecasting using exponential smoothing methods. International Journal of Forecasting 13 (8), 439–454.

Kolassa, S., 2016. Evaluating predictive count data distributions in retail sales forecasting. International Journal of Forecasting 32 (3), 788–803.

Kourentzes, N., 2014. On intermittent demand model optimisation and selection. International Journal of Production Economics 156, 180–190.

Petropoulos, F., Kourentzes, N., 2015. Forecast combinations for intermittent demand. Journal of the Operational Research Society 66 (6), 914–924.

Schultz, C. R., 1987. Forecasting and Inventory Control for Sporadic Demand under Periodic Review. Journal of the Operational Research Society 38 (5), 453–458.

Shenstone, L., Hyndman, R. J., 2005. Stochastic models underlying Croston's method for intermittent demand forecasting. Journal of Forecasting 24 (6), 389–402.

Snyder, R. D., 1985. Recursive Estimation of Dynamic Linear Models. Journal of the Royal Statistical Society, Series B (Methodological) 47 (2), 272–276.

Snyder, R. D., 2002. Forecasting sales of slow and fast moving inventories. European Journal of Operational Research 140 (3), 684–699.

Snyder, R. D., Ord, J. K., Beaumont, A., 2012. Forecasting the intermittent demand for slow-moving inventories: A modelling approach. International Journal of Forecasting 28 (2), 485–496.

Syntetos, A. A., Boylan, J. E., 2001. On the bias of intermittent demand estimates. International Journal of Production Economics 71 (1-3), 457–466.

Syntetos, A. A., Boylan, J. E., 2005. The accuracy of intermittent demand estimates. International Journal of Forecasting 21 (2), 303–314.

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