Int. J. Production Economics 265 (2023) 109013

Contents lists available at ScienceDirect

International Journal of Production Economics

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

iETS: State space model for intermittent demand forecasting

Ivan Svetunkov [*], John E. Boylan

Department of Management Science, Lancaster University Management School, Lancaster, Lancashire, LA1 4YX, UK

ARTICLE INFO

Keywords:
Forecasting
State space models
Exponential smoothing
Croston
Intermittent demand
Supply chain
Inventory management

ABSTRACT

Inventory decisions relating to items that are demanded intermittently are particularly challenging. Decisions relating to termination of sales of product often rely on point estimates of the mean demand, whereas replenishment decisions depend on quantiles from interval estimates. It is in this context that modelling intermittent demand becomes an important task. In previous research, this has been addressed by generalised linear models or integer-valued ARMA models, while the development of models in state space framework has had mixed success. In this paper, we propose a general state space model that takes intermittence of data into account, extending the taxonomy of single source of error state space models. We show that this model has a connection with conventional non-intermittent state space models used in inventory planning. Certain forms of it may be estimated by 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 a simulation experiment, empirically evaluating the inventory implications.

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 are commonly observed in the demand for service parts, in retail and in many other supply chain contexts (e.g. as a result of lot-sizing rules in the upstream stages of the supply chain or due to customer purchasing behaviour in the downstream stages). In this case, inventory decisions must be made when demand is highly uncertain, and rely not only on point estimates but also on interval estimates. Intermittent series arise in other, non-inventory related contexts, but these are not the main subject of interest of this paper.

There are several forecasting methods that are typically used in the case of intermittent demand (Croston, 1972; Syntetos and Boylan, 2005; Teunter et al., 2011) and there have been some attempts to propose forecasting (statistical) models that would underlie these methods (Snyder, 2002; Shenstone and Hyndman, 2005; Snyder et al., 2012). For what follows we introduce two definitions related to the topic of demand forecasting. A **forecasting model** is a mathematical representation of a real phenomenon with a complete specification of distribution and parameters. A **forecasting method** is a mathematical procedure that generates point and/or interval forecasts, with or without a forecasting model. The classical example of the latter is the Simple Exponential Smoothing (SES) method (Brown, 1956), which has an underlying ETS(A,N,N) model (Hyndman et al., 2002).

In this paper, we propose a general state space forecasting framework for intermittent data and demonstrate its application for intermittent demand. The models that we propose allow producing conditional moments and selecting appropriate time series components using information criteria. We also show the connection between the conventional and the proposed forecasting models. We then conduct an experiment on a dataset, demonstrating how the proposed approach works both in terms of forecasting accuracy and inventory performance. Thus we contribute towards filling a gap of modelling intermittent time series using state space models, opening up new research directions in this area.

2. Literature review

The most popular intermittent demand forecasting method was proposed by Croston (1972). His method has been implemented in widely adopted supply chain software packages (e.g. SAS, SAP APO and others). Croston was the first to note that, when demand is intermittent, SES 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:

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

* Corresponding author.
E-mail addresses: i.svetunkov@lancaster.ac.uk (I. Svetunkov), j.boylan@lancaster.ac.uk (J.E. Boylan).

https://doi.org/10.1016/j.ijpe.2023.109013
Received 24 October 2022; Received in revised form 11 May 2023; Accepted 11 August 2023
Available online 16 August 2023

where y_t is the actual observation, o_t is a binary Bernoulli distributed variable taking a value of one when demand occurs and zero otherwise, z_t is the potential demand size, having some conditional distribution, becoming the realised demand when $o_t = 1$ and, finally, t is the time of the observation. 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. This also means that instead of having the series $t = 1, \dots, T$, we have two time series, namely the demand intervals q_{j_t} and demand sizes z_{j_t}, where $j_t = 1, \dots, N$ reflects the sequential numbers of demand intervals and demand sizes and N is the number of non-zero demands. If q_{j_t} is the time elapsed since the last non-zero observation, then it represents the demand interval when the next non-zero observation occurs. Croston (1972) assumed that the probability of occurrence is constant between the non-zero demands, while the mean demand sizes are considered to be the same during the zero demands. Both demand sizes z_{j_t} and demand intervals q_{j_t} are forecasted in this method using SES, which leads to the following system:

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

where $\hat{y}_{j_t}$ is the predicted mean demand, $\hat{z}_{j_t}$ is the predicted demand size, $\hat{q}_{j_t}$ is the predicted demand interval, α_q and α_z are the smoothing parameters for intervals and sizes respectively. The formulation (2) demonstrates how each observation at time t translates to respective j_t element of demand sizes and demand intervals. 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).

Analysing Croston's method, 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 an approximation (known as the Syntetos-Boylan Approximation, SBA). 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).

Trying to find a statistical model underlying Croston's method, Snyder (2002) examined the following form:

$$y_t = o_t \mu_{t|t-1} + \epsilon_{z,t}, \tag{3}$$

where $\mu_{t|t-1}$ is the conditional expectation of demand sizes. Snyder (2002) showed that the model (3) contradicts some basic assumptions about intermittent demand. The main reason for this is because the error term $\epsilon_{z,t}$ is assumed to be normally distributed, but this means that demand can be negative. So, Snyder (2002) proposed the following modified intermittent demand model:

$$y_t^+ = o_t \exp(\mu_{t|t-1} + \epsilon_{z,t}), \tag{4}$$

where y_t^+ represents the demand at time t. This model does not underlie Croston's method, but rather introduces a new forecasting method.

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:

1. be non-stationary,
2. be defined for continuous demand sizes,
3. assume that demand can be negative.

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. While we agree that non-stationarity is an important property, it seems that Shenstone and Hyndman (2005) assume that intermittent and count (or integer-valued) demands are the same thing. However, we argue that the two terms are not equivalent. Count data is not the same as intermittent data. The former may have zeroes (but not necessarily), while the latter must have at least one zero. Furthermore, intermittent data can have continuous demand sizes (e.g. electricity or fuel consumption), while count data can have only discrete variables (e.g. number of boxes of milk sold). From the more fundamental modelling point of view, zero is just one of the observations in the count data process (people buy zero units of product), being at the same time a result of a different process in case of the intermittent data: people either buy or do not buy a product; if they buy they will purchase a non-zero amount. In some situations, count demand can be considered as a special case of the intermittent demand, but this is not a universal rule. As for the negativity, we show later in the paper that it is possible to formulate a model that would work for positive demand sizes only and would underlie Croston's method. We argue that pure multiplicative models make more sense for the intermittent demand statistical model than the additive or mixed ones, because they restrict the space of demand sizes to positive numbers.

Hyndman et al. (2008), pp. 281–283 proposed a basis for Croston's method using a Poisson distribution of demand sizes:

$$\begin{aligned} y_t &= o_t z_{j_t} \\ z_{j_t} &\sim \text{Poisson}(\hat{z}_{j_t-1} - 1) + 1 \\ \hat{z}_{j_t} &= \alpha_z z_{j_t-1} + (1-\alpha_z)\hat{z}_{j_t-1}, \\ o_t &\sim \text{Bernoulli}\left(\frac{1}{\hat{q}_{j_t}}\right) \\ \hat{q}_{j_t} &= \alpha_p q_{j_t-1} + (1-\alpha_p)\hat{q}_{j_t-1} \end{aligned} \tag{5}$$

where $\hat{\mu}_{z,j_t}$ is the estimate of the average number of events per trial and j_t is defined in (2). The authors point out that the proposed filter "gives one-step-ahead forecasts equivalent to Croston's method". Eq. (5) defines a stochastic process and retains useful statistical properties, allowing the generation of both point forecasts and prediction intervals, which, as the authors propose, should be done using simulations.

The fundamental issue with this approach is that the authors claim that intermittent demand is equivalent to count data. However, as discussed earlier, this is not necessarily the case. Besides, Eq. (5) is not a complete statistical model, because it does not have a complete specification of distributions: both $\hat{\mu}_{z,j_t}$ and $\hat{q}_{j_t}$ employ the SES method, which means that $\hat{\mu}_{z,j_t}$ is the weighted average of the previous observed demand sizes, while $\hat{q}_{j_t}$ is the weighted average of the previous observed demand intervals. And neither of them can be defined in terms of an underlying distribution and a fixed parameter. As a result of this, (5) should be considered a filter rather than a model. It is a restrictive approach, which cannot be easily generalised to other exponential smoothing models and does not support exogenous variables.

Snyder et al. (2012) proposed several count data approaches. The filter (5) was called the "Hurdle shifted Poisson". The authors also suggested using the Negative Binomial distribution with time varying mean value to intermittent data, and found that it performs better than the other approaches. The critique applied to the previous approaches in terms of the assumed equivalence of intermittent and count data applies to this filter as well. Furthermore, because the authors did not model the occurrence variable separately, the proposed Negative Binomial filter is more restrictive than the filter (5), updating its values on every observation, in a manner similar to SES, and thus potentially reintroducing the decision-point bias.

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

directly 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:

$$\begin{aligned} \hat{y}_t &= \hat{p}_t \hat{z}_{j_t} \\ \hat{z}_{j_t} &= \alpha_z z_{j_t-1} + (1-\alpha_z)\hat{z}_{j_t-1}, \\ \hat{p}_t &= \alpha_p o_{t-1} + (1-\alpha_p)\hat{p}_{t-1} \end{aligned} \tag{6}$$

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 on each observation, while demand sizes are updated only when $o_t = 1$, thus index j_t instead of t for the variable. An advantage of this method is that the conditional h-steps ahead expectation is straight forward and the update mechanism is simple. 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, variance calculation and prediction intervals construction.

Both TSB and Croston can be applied to fast moving products, where they become equivalent to simple exponential smoothing. They both perform well on real-world datasets (Kourentzes, 2014); however they are not easily extendable and are disconnected from other exponential smoothing methods and are considered to be a different group, because no common model base has yet been identified.

Assessing the performance of intermittent demand methods, Kourentzes (2014) showed that the difference between several of them is insubstantial in terms of point forecasts. The main benefits appear, when distributions are considered and these methods are used for inventory decisions.

Finally, when we switch the discussion to forecasting models, there has been a lot of progress in the direction of count data models (see, for example, Davis et al., 2003; Fokianos et al., 2009; Fokianos and Tjøstheim, 2011) and in the direction of integer ARMA models (Kachour and Yao, 2009; Kachour, 2014). However, the only state space based approach for count data that we are aware of is the one proposed by Harvey and Fernandes (1989). The authors used a Multiple Source of Error framework and assumed a Gamma distribution for the level of demand and Poisson or Negative Binomial for the demand itself. While this is a viable approach for count data, in our opinion it has issues similar to the ones in the filters of Snyder (2002). We suggest using the mixture distribution approach instead, where the demand is split to demand sizes and demand occurrence. This gives additional flexibility in modelling and forecasting intermittent data.

3. 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.). This way the model that we end up with can be considered as a mixture distribution model, consisting of two parts:

1. the one underlying z_t corresponds to potential demand for a product,
2. the other one, underlying o_t, corresponds to demand realisation, when there is occurrence of demand for 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; Ord et al., 1997; Hyndman et al., 2002) whilst acknowledging that other model forms are possible (e.g. multiple source of error or ARIMA). We use the SSOE model for z_t, which in a very general way has the following form, based on (1):

$$\begin{aligned} y_t &= o_t z_t \\ z_t &= w(\mathbf{v}_{t-\boldsymbol{l}}) + r(\mathbf{v}_{t-\boldsymbol{l}})\epsilon_{z,t}, \\ \mathbf{v}_t &= f(\mathbf{v}_{t-\boldsymbol{l}}) + g(\mathbf{v}_{t-\boldsymbol{l}})\epsilon_{z,t} \end{aligned} \tag{7}$$

where o_t is a Bernoulli distributed random variable, $\mathbf{v}_t$ is the state vector, $\epsilon_{z,t}$ is the error term, $f(\cdot)$ is the transition function, $w(\cdot)$ is the measurement function, $g(\cdot)$ is the persistence function, $r(\cdot)$ is the error term function and $\boldsymbol{l}$ is the vector of lags of components, which denotes that the components of the vector $\mathbf{v}_t$ can have different lags (similar to how the SSOE models are formulated in Svetunkov, 2023; Svetunkov and Boylan, 2022; Svetunkov et al., 2023). The functions in (7) correspond to the functions in Hyndman et al. (2008, p.54) and support 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 (7) from an intermittent into a non-intermittent conventional model. This modification expands the Hyndman et al. (2008) taxonomy and allows the introduction of 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.

In order for the model (7) 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 in its value. This assumption might appear to be restrictive, but we show later in this paper that such a model performs well even in the case of discrete demand sizes;
2. Potential demand size may change in time even if we do not observe it. This reflects the idea that the quantities customers intend to purchase might change over time even when they do not purchase the product;
3. Demand size z_t is independent of its occurrence o_t. Relaxing this assumption will lead to a different statistical model;
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).

Having introduced the general framework (7), in this paper, we focus on one special case, the ETS(M,N,N) model for demand sizes, 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 (Chatfield et al., 2001), 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.

In this paper we discuss the main model and several variants of it, stemming from several possible restrictions. In order to distinguish intermittent state space models from the conventional ones we use the letter 'i' in front of the name of the model.

3.1. $iETS_G$ – the general model

The general continuous intermittent state space model (7) based on ETS(M,N,N) reduces to the special case, called iETS(M,N,N). It can be written as:

$$\begin{aligned} y_t &= o_t z_t \\ z_t &= l_{z,t-1}\left(1+\epsilon_{z,t}\right) \quad , \\ l_{z,t} &= l_{z,t-1}(1+\alpha_z \epsilon_{z,t}) \end{aligned} \tag{8}$$

where $l_{z,t}$ is the level of the potential demand sizes. An important distinction between the multiplicative error ETS of Hyndman et al. (2008) and our implementation is that, while they assume that the error term follows a normal distribution, we lift this assumption, following the work of Svetunkov and Boylan (2022), who proposed using the Inverse Gaussian ($\mathcal{IG}$), Gamma (Γ) or Log-Normal ($\log \mathcal{N}$) distributions. Given that we deal with data with potentially low values, the normal distribution becomes unsuitable, because it implies that the error term can be either positive, or negative, or zero and that the distribution itself is symmetric. These are the assumptions that do not hold in the case of intermittent demand. The distributions discussed in Svetunkov and Boylan (2022), on the other hand, can only be positive and have a skewness and kurtosis that are regulated by the parameters of the distributions, giving the necessary flexibility for real time series (for example, see Tadikamalla, 1981, who used $\mathcal{IG}$ for lead time demand).

Whatever the distribution used, the important condition that should be satisfied in order for the model to make sense is $\mathrm{E}(1+\epsilon_{z,t}) = 1$ (see Svetunkov and Boylan, 2022, for explanation). Based on that, the conditional mean and variance for h-steps ahead values can be obtained via the formulae (Svetunkov and Boylan, 2022):

$$\mu_{z,t+h|t} = l_{z,t}, \tag{9}$$

$$\sigma^2_{z,t+h|t} = l^2_{z,t} \left(\left(1+\alpha_z^2\sigma_\epsilon^2\right)^{h-1}(1+\sigma_\epsilon^2) - 1 \right), \tag{10}$$

where σ^2_ϵ is the variance of the error term.

As for the demand occurrence part, it is natural to assume in (8) that o_t has a Bernoulli distribution:

$$o_t \sim \mathrm{Bernoulli}(p_t), \tag{11}$$

where the probability p_t is assumed in a general case to vary over time. It can be modelled in different ways. We propose to consider the probability as a variable driven by two latent variables: the occurrence of demand $\mu_{a,t}$ and the absence of demand $\mu_{b,t}$ and to calculate it based on their conditional means:

$$p_t = \frac{\mu_{a,t}}{\mu_{a,t}+\mu_{b,t}}. \tag{12}$$

According to (12) when $\mu_{b,t}$ increases, the probability of occurrence decreases, while the effect of $\mu_{a,t}$ on p_t is the opposite. The natural restriction for both $\mu_{a,t}$ and $\mu_{b,t}$ is for them to be positive only.

Summarising, the general iETS$_\mathrm{G}$(M,N,N) model can be written as:

$$\begin{aligned} & y_t = o_t l_{z,t-1}\left(1+\epsilon_{z,t}\right) \\ & l_{z,t} = l_{z,t-1}(1+\alpha\epsilon_{z,t}) \\ & o_t \sim \mathrm{Bernoulli}\left(p_t\right) \\ & p_t = \frac{\mu_{a,t}}{\mu_{a,t}+\mu_{b,t}} \\ & \mu_{a,t} = w(\mathbf{v}_{a,t-\boldsymbol{l}_a}) \\ & \mathbf{v}_{a,t} = f(\mathbf{v}_{a,t-\boldsymbol{l}_a}) + g(\mathbf{v}_{a,t-\boldsymbol{l}_a})\epsilon_{a,t} \\ & \mu_{b,t} = w(\mathbf{v}_{b,t-\boldsymbol{l}_b}) \\ & \mathbf{v}_{b,t} = f(\mathbf{v}_{b,t-\boldsymbol{l}_b}) + g(\mathbf{v}_{b,t-\boldsymbol{l}_b})\epsilon_{b,t} \end{aligned}, \tag{13}$$

where the variables with index a denote the part of the model for demand occurrence, the variables with index b denote the part for the non-occurrence of demand, $1+\epsilon_{a,t}$ and $1+\epsilon_{b,t}$ are the mutually and serially independent error terms and α_a and α_b are the smoothing parameters. Note that we do not make any distributional assumptions about $1+\epsilon_{a,t}$ and $1+\epsilon_{b,t}$, because they are not directly observable. The only two natural restrictions on these variables would be for them to be positive and for their means to be equal to one so that the underlying ETS models would produce adequate point forecasts (Svetunkov and Boylan, 2022). Given that the model (13) has two time varying parameters $\mu_{a,t}$ and $\mu_{b,t}$, it covers all the possible cases of probability change over time, including building up demand, demand obsolescence and stable demand.

Note that any multiplicative ETS model could potentially be used in both the demand occurrence and demand sizes parts of (13), 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. Instead we will focus our analysis on the ETS(M,N,N) model, which leads to the following:

$$\begin{aligned} & y_t = o_t l_{z,t-1}\left(1+\epsilon_{z,t}\right) \\ & l_{z,t} = l_{z,t-1}(1+\alpha\epsilon_{z,t}) \\ & o_t \sim \mathrm{Bernoulli}\left(p_t\right) \\ & p_t = \frac{\mu_{a,t}}{\mu_{a,t}+\mu_{b,t}} \\ & \mu_{a,t} = l_{a,t-1} \\ & l_{a,t} = l_{a,t-1}(1+\alpha_a\epsilon_{a,t}) \\ & \mu_{b,t} = l_{b,t-1} \\ & l_{b,t} = l_{b,t-1}(1+\alpha_b\epsilon_{b,t}) \end{aligned}, \tag{14}$$

where $l_{a,t}$ and $l_{b,t}$ are the levels for each of the variables. For simplicity, further in the paper, we will drop the "(M,N,N)" part when we refer to the model (14). There are several special cases arising from the iETS$_\mathrm{G}$ model (14), appearing when some of the parameters of the model are restricted. These models are useful because they allow modelling several real life scenarios, such as "demand building up" and "demand becoming obsolescent". They are discussed in the next subsections.

3.2. iETS$_\mathrm{F}$

The first model appears if we assume that both $\mu_{a,t}$ and $\mu_{b,t}$ are constant over time, which leads to the simplified iETS$_\mathrm{F}$ model (dropping the demand sizes equations):

$$\begin{aligned} & o_t \sim \mathrm{Bernoulli}\left(p\right) \\ & p = \frac{\mu_a}{\mu_a+\mu_b} \end{aligned}. \tag{15}$$

In this model we assume that the probability of demand occurrence does not change over time, so that the probability of having sales is the same for every time period.

The advantage of this model is that it does not contain as many parameters as the iETS$_\mathrm{G}$ does. However, it does not have the same flexibility, so cannot be used for the purposes of modelling and forecasting of demand with time varying occurrence. Still, this model can be used as one of the benchmarks for purposes of model selection in real life.

3.3. iETS$_\mathrm{O}$

Another model arising from (14) is the model with the restriction $\mu_{b,t}=1$, resulting in the iETS$_\mathrm{O}$, the odds ratio model:

$$\begin{aligned} & o_t \sim \mathrm{Bernoulli}\left(p_t\right) \\ & p_t = \frac{\mu_{a,t}}{\mu_{a,t}+1} \\ & \mu_{a,t} = l_{a,t-1} \\ & l_{a,t} = l_{a,t-1}(1+\alpha_a\epsilon_{a,t}) \end{aligned}. \tag{16}$$

This model is called "odds ratio", because the probability of occurrence in (16) is calculated using the classical logistic transform. This also means that $\mu_{a,t}$ is equal to:

$$\mu_{a,t} = \frac{p_t}{1-p_t}. \tag{17}$$

When $\mu_{a,t}$ decreases in the iETS$_\mathrm{O}$ model, the odds ratio decreases as well, meaning that the probability of occurrence goes down. Given the stochastic convergence of the sample path of ETS(M,N,N) to zero (Akram et al., 2009), the model should be appropriate for the cases of demand obsolescence.

3.4. iETS$_I$

Another special case of the model (14) is obtained by restricting $\mu_{a,t} = 1$ instead of $\mu_{b,t}$. In this case we impose specific dynamics towards the increase of probability (based on the convergence of sample path to zero, as discussed in Akram et al., 2009). The general model reduces then to the iETS$_\mathrm{I}$ model:

$$\begin{aligned} & o_t \sim \text{Bernoulli}\left(p_t\right) \\ & p_t = \frac{1}{1+\mu_{b,t}} \\ & \mu_{b,t} = l_{b,t-1} \\ & l_{b,t} = l_{b,t-1}(1+\alpha_b \epsilon_{b,t}) \end{aligned} . \tag{18}$$

The variable b_t can be represented in terms of the probability as:

$$\mu_{b,t} = \frac{1-p_t}{p_t}. \tag{19}$$

So the occurrence part of the (18) models the inverse of the odds ratio (19), thus the name of the model. This means that the decrease of $\mu_{b,t}$ will increase the odds ratio, increasing the probability of occurrence.

3.5. iETS$_D$

The last special case of (14) that we discuss is the model with the restrictions:

$$\mu_{a,t} + \mu_{b,t} = 1, \mu_{a,t} \leq 1, \tag{20}$$

an iETS$_\mathrm{D}$ model (the letter "D" stands for the "direct" probability). After inserting (20) in the formula of the probability (14), it becomes apparent that $p_t = \mu_{a,t}$. But in order to make sure that the probability is kept in the region [0, 1], a condition on the value of $\mu_{a,t}$ needs to be introduced:

$$\begin{aligned} & o_t \sim \text{Bernoulli}\left(\mu_{a,t}\right) \\ & \mu_{a,t} = \min(l_{a,t-1}, 1) \\ & l_{a,t} = l_{a,t-1}(1+\alpha_a \epsilon_{a,t}) \end{aligned} . \tag{21}$$

The model (21) has similar dynamics of the probability of occurrence as iETS$_\mathrm{O}$ has, but it should be more reactive, with the probability changing more rapidly over time.

3.6. iETS processes

Finally, we summarise the properties of the processes according to iETS models. Based on how the models are formulated and the stochastic convergence of ETS(M,N,N) (Akram et al., 2009), we can conclude that:

- iETS$_\mathrm{F}$ should be appropriate in cases, when the probability of occurrence does not change over time (i.e. average demand interval is fixed);
- iETS$_\mathrm{O}$ will work well on data exhibiting obsolescence, when demand slowly disappears and the customers stop purchasing the product, i.e. probability of occurrence eventually converges to zero;
- iETS$_\mathrm{I}$ should capture the dynamics correctly on the data where demand is building up, i.e. the probability of occurrence eventually reaching one;
- iETS$_\mathrm{D}$ should work well on the data with obsolescence, capturing the change in probability of occurrence more rapidly than iETS$_\mathrm{O}$;
- iETS$_\mathrm{G}$ can capture the cases of fixed demand, demand building up and demand obsolescence, depending on the values of its parameters, together with a situation, when demand evolves from one state to the other and does not converge to either zero or one.

Fig. 1 shows several examples of probability of occurrence generated using the models discussed above. In each of these examples, the 10,000 observations were generated to show the long term dynamics of models. These examples summarise visually the main properties of the models discussed above.

4. Estimation of iETS

In order to estimate one of the models discussed in the previous section, we should first consider the question of how to construct them. When the occurrence of demand happens ($o_t = 1$) the construction is straightforward: the forecast error $e_{z,t} = \frac{z_t - \hat{l}_{z,t-1}}{\hat{l}_{z,t-1}}$ is used as an estimate of the error term $\epsilon_{z,t}$, so that the state is updated on that observation. The main difficulty arises, with the estimation of the level of demand sizes $l_{z,t}$, when $o_t = 0$. Given that the demand sizes are not observable in this case, the expectation conditional on the last observed level $\hat{l}_{z,t}$ can be used instead, which for ETS(M,N,N) is:

$$\hat{l}_{z,t+h|t} = \hat{l}_{z,t}. \tag{22}$$

It follows directly from the properties of the ETS(M,N,N) model (Svetunkov and Boylan, 2022) that the several steps ahead conditional expectation of the demand sizes $\hat{z}_{t+h|t}$ is also equal to (22).

As for the demand occurrence, the h-steps ahead point forecast of probability $\hat{p}_{t+h|t}$ follows directly from the model and is equal to:

$$\hat{p}_{t+h|t} = \frac{\hat{l}_{a,t+h|t}}{\hat{l}_{a,t+h|t} + \hat{l}_{b,t+h|t}}, \tag{23}$$

where $\hat{l}_{a,t+h|t} = \hat{l}_{a,t}$ and $\hat{l}_{b,t+h|t} = \hat{l}_{a,t}$ are the corresponding parts of the model. The special cases of the iETS$_\mathrm{G}$ model have the following predicted probabilities of occurrence:

- iETS$_\mathrm{F}$: $\hat{p}_{t+h|t} = \hat{p} = \frac{T_1}{T}$, where T_1 is the number of non-zero observations;
- iETS$_\mathrm{O}$: $\hat{p}_{t+h|t} = \frac{\hat{l}_{a,t}}{\hat{l}_{a,t}+1}$;
- iETS$_\mathrm{I}$: $\hat{p}_{t+h|t} = \frac{1}{1+\hat{l}_{b,t}}$;
- iETS$_\mathrm{D}$: $\hat{p}_{t+h|t} = \min(\hat{l}_{a,t-1}, 1)$.

Finally, in order to construct the occurrence part of the iETS models, we need to have the estimates of the one-step-ahead forecast errors for the demand occurrence. However, they are not observable, so we propose estimating them using the following proxies (see Appendix A for derivations):

$$e_{a,t} = \frac{u_t}{1-u_t} - 1 \tag{24}$$

and

$$e_{b,t} = \frac{1-u_t}{u_t} - 1, \tag{25}$$

where $u_t = \frac{1+o_t-\hat{p}_{t|t-1}}{2}$ and $\hat{p}_{t|t-1}$ is the expected one-step-ahead probability of occurrence. The idea of the proxies is that they transform the distance between the outcome and the probability into the scale of the $l_{a,t}$ and $l_{b,t}$.

Finally, given the assumption of independence between demand occurrence and demand sizes, the conditional one-step-ahead mean of the demand y_t is calculated as:

$$\hat{y}_{t|t-1} = \hat{p}_{t|t-1}\hat{z}_{t|t-1}. \tag{26}$$

Similarly, the h steps ahead conditional expectation of the demand y_{t+h} can be calculated as:

$$\hat{y}_{t+h|t} = \hat{p}_{t+h|t}\hat{z}_{t+h|t}. \tag{27}$$

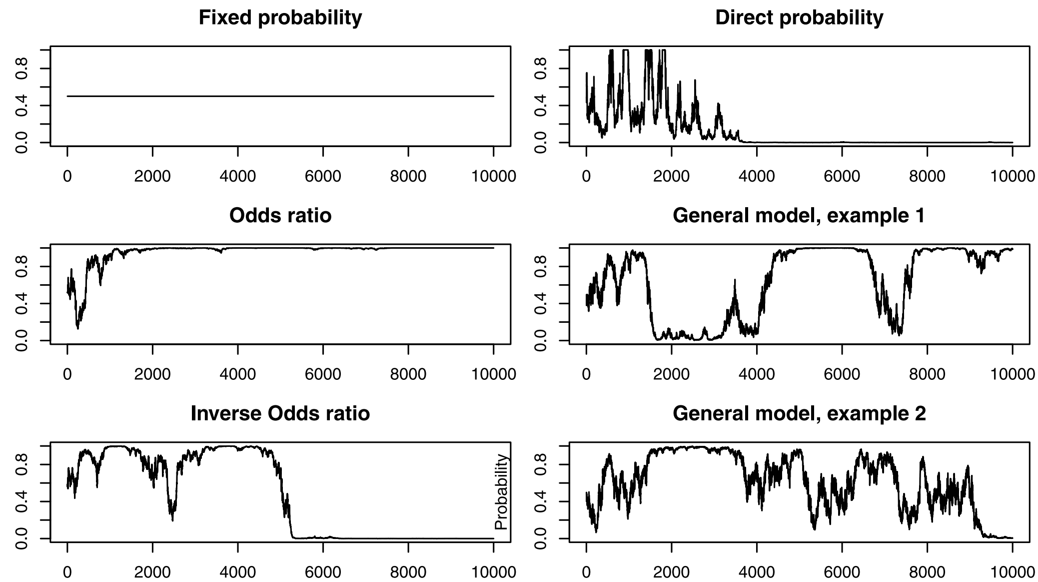

Fig. 1. Several examples of probability of occurrence generated using the iETS models.

Summarising all of the above, the estimation of the iETS_{G} model can be represented as a set of the following equations:

$$
\begin{aligned}
&\hat{y}_{t|t-1} = \hat{p}_{t|t-1}\hat{z}_{t|t-1} \\
&e_{z,t} = o_t \frac{y_t - \hat{z}_{t|t-1}}{\hat{z}_{t|t-1}} \\
&\hat{z}_t = \hat{l}_{z,t-1} \\
&\hat{l}_{z,t} = \hat{l}_{z,t-1}(1 + \hat{\alpha}_z e_{z,t}) \\
&\hat{p}_{t|t-1} = \frac{\hat{l}_{a,t}}{\hat{l}_{a,t} + \hat{l}_{b,t}} \\
&u_t = \frac{1 + o_t - \hat{p}_{t|t-1}}{2} \\
&e_{a,t} = \frac{u_t}{1 - u_t} - 1 \\
&\hat{l}_{a,t} = \hat{l}_{a,t-1}(1 + \hat{\alpha}_a e_{a,t}) \\
&e_{b,t} = \frac{1 - u_t}{u_t} - 1 \\
&\hat{l}_{b,t} = \hat{l}_{b,t-1}(1 + \hat{\alpha}_b e_{b,t})
\end{aligned} . \tag{28}
$$

Note that in order to construct the iETS_{G} model, we need to initialise it with some values of $\hat{l}_{z,0}$, $\hat{l}_{a,0}$ and $\hat{l}_{b,0}$. The conventional ETS approach to this problem is to estimate them together with the smoothing parameters $\hat{\alpha}_z$, $\hat{\alpha}_a$ and $\hat{\alpha}_b$ via the maximisation of the likelihood function (Hyndman et al., 2002). In our case, the log-likelihood function for the model (14) can be written as (see Appendix B for the derivation):

$$
\begin{aligned}
\ell(\boldsymbol{\theta}, \sigma_\epsilon^2 | \mathbf{Y}) = & \sum_{o_t=1} \log f_z\left(z_t | l_{z,t-1}\right) + \sum_{o_t=0} \mathcal{H}_z(z_t) \\
& + \sum_{o_t=1} \log(\hat{p}_t) + \sum_{o_t=0} \log(1 - \hat{p}_t)
\end{aligned} , \tag{29}
$$

where $\mathbf{Y}$ is the vector of all the in-sample observations, $\boldsymbol{\theta}$ is the vector of parameters to estimate (initial values and smoothing parameters), $f_z\left(z_t | l_{z,t-1}\right)$ is the density function of the assumed distribution for the demand sizes and $\mathcal{H}_z(z_t)$ is the differential entropy of that distribution, $\hat{p}_t$ is the probability of occurrence produced by the selected iETS model.

We should note at this stage that all the iETS models estimated in the proposed way will produce consistent, efficient, but positively biased estimates of the smoothing parameter α_z. The first two are inherited by the MLE, while the discussion of the latter phenomenon is provided in Appendix C. The bias appears due to the assumption of change of level between the non-zero demands and increases with the decrease of the probability of occurrence. If the probability of occurrence p_t converges to one, the bias disappears (see Appendix C for details).

The other important part of the model application to the data is the number of estimated parameters k, which differs from model to model:

- iETS_{G} – seven parameters: the initial values of $\hat{l}_{z,0}$, $\hat{l}_{a,0}$ and $\hat{l}_{b,0}$, the smoothing parameters $\hat{\alpha}_z$, $\hat{\alpha}_a$ and $\hat{\alpha}_b$ and the scale parameter for the error term of the demand sizes $\hat{\lambda}_\epsilon$;
- iETS_{F} – four parameters: the initial value of $\hat{l}_{z,0}$, the smoothing parameter $\hat{\alpha}_z$, the probability of occurrence $\hat{p}$ and the scale parameter $\hat{\lambda}_\epsilon$;
- iETS_{O} – five parameters: the initial values of $\hat{l}_{z,0}$ and $\hat{l}_{a,0}$, the smoothing parameters $\hat{\alpha}_z$, $\hat{\alpha}_a$ and the scale parameter $\hat{\lambda}_\epsilon$;
- iETS_{I} – five parameters, similar to iETS_{O}, but with $\hat{l}_{b,0}$ instead of $\hat{l}_{a,0}$ and $\hat{\alpha}_b$ instead of $\hat{\alpha}_a$;
- iETS_{D} – also five parameters, similar to iETS_{O}.

Note that the parameters of the demand occurrence part are estimated via the maximisation of the last two terms in (29) and given the assumption of independence, this can be done independently of the estimation of the parameters of the demand sizes part.

Having discussed how to estimate the iETS models, we now move to the discussion of two important special cases of the general model.

4.1. Estimation of iETS_I

There are two ways of estimating the probability of occurrence in the iETS_{I} model. The first one is based on (28), as discussed above, while the second one appears as soon as we observe that the probability of occurrence p_t in this model is inversely proportional to the value of $\mu_{b,t}$. This property implies that the model can also be estimated, assuming that $\mu_{b,t}$ is rounded down, which becomes equal to the observed demand intervals. We can then make a substitution of $\hat{q}_{j_t} = \lfloor 1 + \hat{\mu}_{b,t} \rfloor$ in order to arrive to the mechanism of probability update in Croston's method (2):

$$
\hat{p}_{j_t} = \frac{1}{\lfloor 1 + \hat{\mu}_{b,t} \rfloor} = \frac{1}{\hat{q}_{j_t}}. \tag{30}
$$

This demonstrates that Croston's method is just one of the ways of estimating the model iETS_{I}. It is important to note that although $\hat{\mu}_{b,t}$ may vary in time on each observation, influencing the corresponding probability $\hat{p}_t$, iETS_{I} model estimated with the probability (30) cannot be estimated when demand is zero. So during the estimation of the model based on (30), it should be assumed that the states of $\mu_{b,t}$ do not change between demand occurrences, which corresponds to the original Croston's method assumption. This demonstrates that the iETS$_\mathrm{I}$ is a model underlying Croston's method.

4.2. Estimation of iETS$_D$

If all the values of $\mu_{a,t} \leq 1$, then all the conditional values of iETS$_\mathrm{D}$ are equivalent to those from the ETS(M,N,N) model. This situation may occur, when level is low and converges to zero, meaning that demand for the product is sparse and becomes obsolete. In this case, the iETS$_\mathrm{D}$ model can be estimated using the TSB method (6).

When it comes to the construction of the iETS$_\mathrm{D}$ model (21), it is implied that the error term is equal to:

$$1 + e_{a,t} = \frac{o_t}{\hat{\mu}_{a,t}}. \tag{31}$$

However, this is unrealistic, because in case of $o_t = 0$ the error (31) becomes equal to zero, thus making the model inestimable. In order to estimate the iETS$_\mathrm{D}$, we have to introduce the following approximation for the error term $e_{a,t}$, which guarantees non-zero forecast errors for the boundary cases:

$$e_{a,t} = \frac{o_t(1-2\kappa) + \kappa - \hat{\mu}_{a,t}}{\hat{\mu}_{a,t}}, \tag{32}$$

where κ is a very small number (for example, $\kappa = 10^{-10}$). This modification is artificial but it helps in the estimation of the model. It is worth stressing that the only purpose of κ is to make model estimable.

An alternative way of estimating iETS$_\mathrm{D}$ is using TSB method, which has a direct connection with the model because ETS(M,N,N) underlies SES method, implying that the set of Eqs. (6) can be used to estimate the model (21). This means that iETS$_\mathrm{D}$ is a model underlying TSB.

4.3. Prediction interval for iETS models

In order to calculate prediction interval for the iETS model, the cumulative distribution function (CDF) can be used:

$$F_y(y_{t+h} \leq Q) = \hat{p}_{t+h|t} F_z(z_{t+h} \leq Q) + (1 - \hat{p}_{t+h|t}), \tag{33}$$

where $F_z(z_{t+h})$ is the h-steps ahead CDF of demand sizes z_{t+h}, $F_y(y_{t+h})$ is the final CDF of the variable y_{t+h} and Q is the value of the desired quantile of distribution. $F_y(y_{t+h})$ should correspond to the desired probability (for example, 0.95), and the only unknown element in (33) is $F_z(z_{t+h})$, which can be calculated as:

$$F_z(z_{t+h} \leq Q) = \frac{F_y(y_{t+h} \leq Q) - (1 - \hat{p}_{t+h|t})}{\hat{p}_{t+h|t}}. \tag{34}$$

So, in the construction of a prediction interval, the formula (34) can be used for the calculation of the probability for the demand sizes part of the model, which will give the necessary quantile Q. Unfortunately, the conditional h-steps ahead distribution function of demand sizes is not known for any of the known distributions because that value involves complex permutations of error terms (see Svetunkov and Boylan, 2022). The relatively simple way of obtaining the values is via simulations of possible demand trajectories based on the applied model.

Furthermore, given that in many cases the data might contain a preponderance of zeroes, the lower bound of prediction interval would correspond to zero and might not be useful. In this case, the one-sided prediction interval for the upper bound (which is important, for example, for safety stock calculation) can be calculated based on (34). In order to do so, the upper quantile is calculated for $1-\alpha$ rather than $1-\frac{\alpha}{2}$ (where α is the significance level).

Finally, in many contexts the intermittent demand approaches are applied to count data, thus it makes more sense to have integer numbers in quantiles rather than the fractional ones. One of the simplest and most practical solutions in this case is to round up the resulting quantile. From the theoretical point of view, this will correspond to the quantiles from the model:

$$y_t = o_t \lceil z_t \rceil. \tag{35}$$

The proof of this is provided in Appendix D. We argue that rounding up makes more sense for the iETS model than rounding down, because in the latter case, zero values might be obtained, which would correspond to an exotic situation of "having a zero demand".

4.4. Model with constant demand sizes

One of the other special cases of the iETS model that needs to be discussed separately is the case when $z_t = z$ is constant over time. This means that the demand y_t takes at random only two values: 0 or z. The demand model becomes:

$$\begin{aligned} & y_t = o_t z \\ & o_t \sim \text{Bernoulli}\left(p_t\right)' \end{aligned} \tag{36}$$

where p_t can be modelled using one of the types of the occurrence parts of iETS models discussed above. The log-likelihood for this model is then simplified to:

$$\ell(\boldsymbol{\theta}|\mathbf{Y}) = \sum_{o_t=1} \log(\hat{p}_t) + \sum_{o_t=0} \log(1-\hat{p}_t). \tag{37}$$

As for the prediction interval, given that o_t is the only source of randomness in this model, the upper bound of the prediction interval will typically be equal to z, while the lower one will be 0.

4.5. Model selection in the iETS framework

Having the likelihood function for all the discussed intermittent state space models: iETS$_\mathrm{G}$ (14), iETS$_\mathrm{F}$ (15), iETS$_\mathrm{O}$ (16), iETS$_\mathrm{I}$ (18) and iETS$_\mathrm{D}$ (21) – and knowing the number of parameters to estimate, we can calculate any information criterion and use it for model selection between the models, or for the selection of components inside each of the parts of each model, or the selection of the most appropriate distributions. For example, the Akaike Information Criterion (Akaike, 1974) can be calculated as:

$$\text{AIC} = 2k - 2\ell(\boldsymbol{\theta}, \hat{\sigma}_\epsilon^2|\mathbf{Y}), \tag{38}$$

where for the intermittent models k was discussed in Section 4 and, for example, for the conventional ETS(A,N,N), $k = 3$. Note that the only difference between iETS$_\mathrm{O}$, iETS$_\mathrm{I}$ and iETS$_\mathrm{D}$ is in the probability modelling mechanism — they have exactly the same number of degrees of freedom. In order to distinguish the automatic model selected via an information criterion from the individual cases, we use the notation iETS$_\mathrm{A}$.

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 ETS(M,N,N). For example, using a model with fixed level (setting smoothing parameter α_z to zero) allows preserving one degree of freedom without substantial loss in generality and fitting the model to the data with at least four non-zero observations.

5. Real time series experiment

5.1. The setting

In order to examine the performance of the proposed intermittent state space models, we conduct an experiment on M5 competition dataset (Makridakis et al., 2021). The dataset contains 30,490 daily time series, 1969 observations each. The distribution of non-zero observations in the dataset is shown in Fig. 2. As it can be seen from the

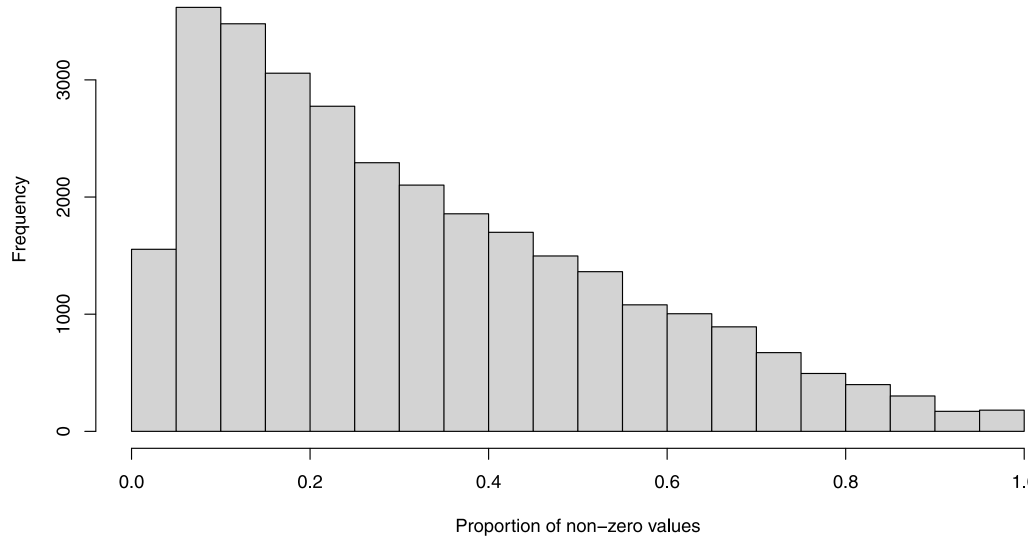

Fig. 2. The distributions of proportions of the non-zero demands in the dataset.

figure, there are more series with low value of non-zero observations, implying that the intermittent demand models should be appropriate.

We have used the following set of models from the `smooth` package v3.2.0 (Svetunkov, 2022) for R (R Core Team, 2022) in this experiment using the `adam()` function:

1. ETS(A,N,N), which is needed as a benchmark;
2. ETS(X,X,N) – automatically selected pure additive non-seasonal ETS model, which was used in the original M5 competition;
3. iETS_F – the model with fixed probability;
4. iETS_O – the odds ratio model;
5. iETS_I – the inverse odds ratio model;
6. iETS_D – the direct probability model;
7. iETS_G – the general iETS model;
8. iETS_A – the model with automatic selection between the first five models using corrected AIC (Sugiura, 1978);

All the models except for iETS_A assumed that the demand sizes follow the Gamma distribution. The iETS_A model implemented an automatic distribution selection. We have also added the following filters implemented in the `counter`, v0.2.0 package for R (Svetunkov, 2019):

1. Hurdle shifted Poisson filter (denoted "HSP") discussed in Snyder et al. (2012), implemented in the `hsp()` function;
2. Negative Binomial filter (denoted "NegBin") from Snyder et al. (2012), implemented in the function `negbin()`.

Finally, we have included the two benchmark methods, discussed earlier in this paper, implemented in `tsintermittent` v1.9 package for R (Kourentzes and Petropoulos, 2016):

1. TSB method implemented in `tsb()` function;
2. Croston's method implemented in `crost()` function;
3. SBA method implemented in `crost()` function.

For each of the time series in the dataset we produced 1 to 14 steps ahead point forecasts and prediction intervals, withholding the respective 14 observations for the model evaluation. Some time series contain zeroes in the beginning, implying that the product is introduced at a later stage. These zeroes were removed before applying forecasting approaches. As a result the in-sample size would differ from one series to another depending on when the first non-zero demand happened. We measure the accuracy of point forecasts of all the competing methods and models using the following error metrics, discussed in Kourentzes (2014) and Petropoulos and Kourentzes (2015):

- sCE — scaled Cumulative Error, measuring the potential bias in forecasts;
- sAPIS — scaled Absolute Periods in Stock, measuring the bias over the lead time;
- sRMSE — scaled Root Mean Squared Error, measuring the accuracy of point forecasts;

Given that iETS models and the filters allow producing prediction intervals, we used Scaled Mean Interval Score (based on Gneiting and Raftery, 2007):

$$\text{sMIS} = \frac{\text{MIS}}{\bar{y}}, \tag{39}$$

where MIS is the Mean Interval Score for the model/method under consideration and $\bar{y}$ is the in-sample mean. MIS is calculated as the arithmetic mean of the interval score (IS by Gneiting and Raftery, 2007) over the forecasting horizon. The width of the prediction interval is set to 0.9, 0.95, 0.99, and we produced the upper bound only, setting the lower one to zero for all the models under consideration (as discussed in Section 4.3).

However, MIS gives only an aggregate information in terms of performance of the methods. In order to better understand how specifically the methods perform, we also provide values for coverage (the closer it is to the nominal value, the better it is) and pinball for the levels of 0.9, 0.95 and 0.99.

When comparing the performance in terms of quantiles, we also report the values for the rounded-up quantiles, as discussed in Section 4.3.

Finally, we measure the computational time (in seconds) of the approaches to better understand how efficient they are in terms of used resources.

5.2. Point forecasts

Table 1 summarises performance of the models and methods under consideration in terms of point forecasts.

From Table 1, we can conclude that iETS models perform very well — almost all of them outperform HSP and NegBin benchmarks. The only specific iETS model that does not do well is the one with the fixed probability. This means that the data exhibits changes in probability of occurrence, which should be taken into account. Note also that in terms of sAPIS and sRMSE, the ETS(A,N,N) model performs better than NegBin, HSP, Croston, TSB and iETS_D. This could be due to the nature of the data — these models were developed for highly intermittent data, especially for spare parts problem, while the retail

Table 1
Performance of models in terms of point forecasts.

Approaches	Mean values				Median values			
	sCE	sAPIS	sRMSE	Time	sCE	sAPIS	sRMSE	Time
$iETS_F$	−2.112	35.163	0.487	0.114	−2.272	27.981	0.423	0.116
$iETS_O$	−1.196	19.084	0.347	0.553	−1.168	11.502	0.169	0.481
$iETS_I$	**0.150**	**18.753**	**0.331**	0.506	**−0.578**	**8.074**	**0.119**	0.457
$iETS_D$	−1.211	24.422	0.393	0.380	−1.628	18.404	0.288	0.407
$iETS_G$	−0.773	20.742	0.360	1.000	−1.295	14.615	0.227	1.111
$iETS_A$	−0.795	21.035	0.362	2.082	−1.296	14.704	0.231	2.156
HSP	−2.450	38.002	0.512	0.647	−2.513	29.936	0.447	0.654
NegBin	−1.259	23.927	0.390	0.262	−1.591	17.932	0.287	0.255
TSB	−1.346	25.663	0.403	**0.017**	−1.695	19.643	0.318	**0.013**
Croston	−2.976	39.830	0.533	0.024	−2.692	30.704	0.475	0.018
SBA	−2.818	39.252	0.526	0.024	−2.594	30.356	0.467	0.018
ETS(A,N,N)	−1.464	23.824	0.393	0.089	−1.613	18.527	0.314	0.090
ETS(X,X,N)	−1.461	23.832	0.393	0.225	−1.605	18.520	0.314	0.238

Table 2
Performance of models in terms of quantile forecasts.

Approaches	Mean values			Median values	
	sMIS	Pinball	Coverage	sMIS	Pinball
$iETS_F$	2.658	4.556	0.891	1.386	0.651
$iETS_O$	**1.880**	2.745	0.937	0.937	0.300
$iETS_I$	2.231	4.413	0.886	**0.659**	**0.152**
$iETS_D$	2.305	3.805	0.907	1.141	0.443
$iETS_G$	2.211	3.716	0.908	0.989	0.337
$iETS_A$	2.221	3.740	0.908	1.019	0.355
HSP	2.654	3.163	**0.946**	2.041	1.375
NegBin	2.306	**2.332**	0.970	2.046	1.300
ETS(A,N,N)	2.749	2.941	0.982	2.456	1.926
ETS(X,X,N)	2.749	2.941	0.982	2.456	1.926
Rounded up quantiles					
$iETS_F$	2.574	3.165	**0.950**	1.855	1.137
$iETS_O$	**2.126**	**2.236**	0.971	1.837	1.027
$iETS_I$	2.150	2.935	0.944	**1.631**	**0.844**
$iETS_D$	2.312	2.679	0.959	1.814	1.024
$iETS_G$	2.266	2.617	0.960	1.799	1.008
$iETS_A$	2.272	2.642	0.959	1.803	1.016

data has different characteristics, with more rapidly changing patterns. The best performing model in our case is $iETS_I$. Given the theoretical implications of this model, this also demonstrates that often time series in the dataset start as highly intermittent but over time exhibit more frequent demand occurrences.

Also note that $iETS_O$ and $iETS_G$ performed very well as well, outperforming the conventional intermittent demand forecasting approaches both in terms of mean and median error measures. Still, $iETS_I$ was the least biased and most accurate model.

While the automatic model selection in iETS did not outperform the other approaches in terms of point forecasts, it did not do poorly, making it a plausible alternative for the cases when it is not obvious which of the iETS models to use.

Finally, in terms of computational time, all iETS models take more time than the simple approaches, such as TSB, Croston and SBA. The slowest approach (as expected) is the iETS with model selection, which takes on average 2 s per time series.

5.3. Quantile forecasts

Table 2 summarises performance of models in terms of the upper bound of the prediction interval for the 95% confidence level. The values for 90% and 99% are provided in Appendix E, in Tables E.3 and E.4 respectively. Given that Croston, SBA and TSB are forecasting methods, we cannot easily produce quantile forecasts from them, which is why we do not include them in the further evaluation.

We can see from Table 2 that now the $iETS_O$ model performs better than the others in terms of several error measures: it is the best in terms of both mean and median sMIS and it is the second best model in terms of mean and median pinball. The reason why iETS models performed well in terms of sMIS could be because they produced narrower intervals, with actual values lying not far from the upper bound. In terms of coverage, the HSP is the closest to the nominal 95%, while all iETS model fall short, producing intervals that include much smaller percentage of observations than the nominal one. Overall, if the prediction interval is generated without rounding the values, the iETS models perform well in terms of some measures, falling behind the NegBin and HSP filters in terms of the others.

However, the situation changes when we round up the quantiles of iETS models. In that situation, the coverage for all the models increases and gets much closer to the nominal 0.95. The mean pinball values improve as well, making iETS models attractive alternatives to the count distribution filters. Interestingly, both median sMIS and pinball of iETS models increase when the values are rounded up. In order to understand this phenomenon, consider the case when the initially predicted quantile of 1.5 is rounded up to 2, while the actual value was 1. In this situation, 1.5 is closer to 1 than 2, thus we would expect in many cases an increase in sMIS after rounding up the values because of the sampling of the continuous values. At the same time, we do not see such a clear effect in terms of mean values: some of them increase slightly, some of them decrease. This is because rounding up allows avoiding more extreme situations, when the prediction interval does not cover the actual values. This is reflected in the higher coverage for the iETS with rounded up quantiles.

It is also worth noting that ETS(A,N,N) performed very similar to the ETS(X,X,N), which implies that the local level model was selected in the majority of cases in the experiment. Because of that, we drop the latter from further analysis.

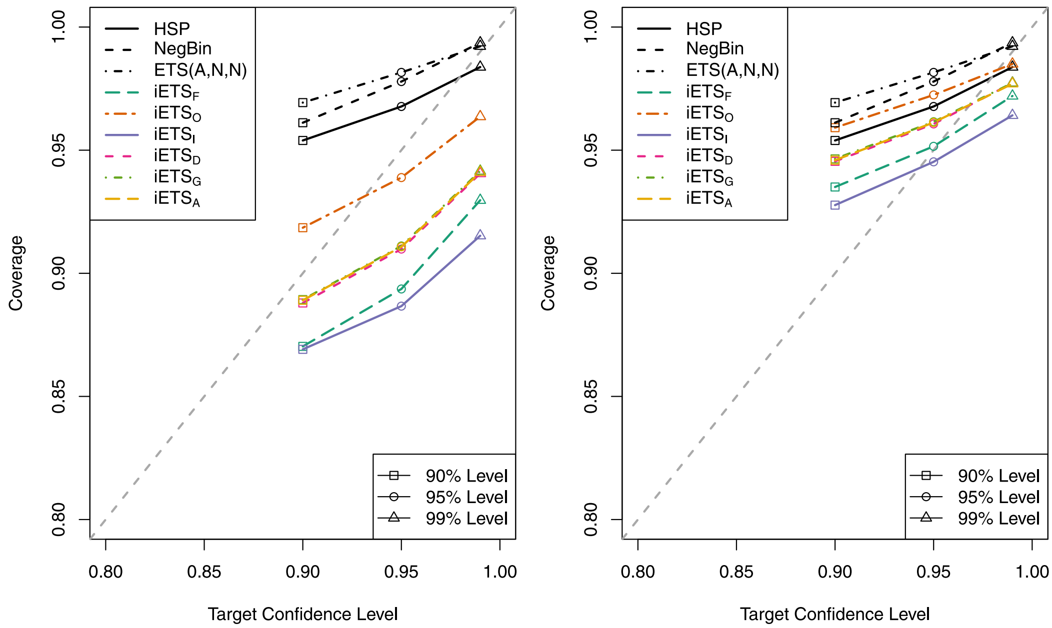

Fig. 3. Performance of models in terms of coverage. Left plot shows performance of conventional iETS models, right plot shows iETS models with rounding up.

The results for the 90% and 99% confidence levels are similar, so we do not discuss them here in detail (see Appendix E). The only point worth noting is that all the models tend to produce higher coverage for the lower confidence level and lower coverage for the higher level (see Fig. 3), implying that the models under consideration are not well calibrated for the data. NegBin, HSP and ETS(A,N,N) tend to produce higher coverage than needed, while the conventional iETS models tend to produce lower coverage. However, the performance of the latter improves as soon as the rounding is introduced.

5.4. Summary

As can be seen from the results of this experiment, iETS models have the flexibility necessary for intermittent demand forecasting, outperforming the existing benchmarks in terms of point forecasts. iETS$_O$ and iETS$_I$ models perform better than many other models in our experiment, making them attractive for application to a wide variety of intermittent time series. In terms of prediction intervals, we find that the iETS models without rounding up include fewer observations than needed, but after rounding the quantiles up, the iETS model becomes a viable alternative to the conventional approaches. Once again, iETS$_O$ and iETS$_I$ look attractive in terms of their performance.

6. Inventory implications

In order to see how the proposed models behave in terms of inventory, we conduct a smaller experiment on a subset of M5 data. Given that we do not have information about the inventory system in Walmart, we will use the order-up-to policy, assuming that the cycle service level is of the main interest for the company. To simplify the setup, we also assume that the inventory system does weekly reviews and that the lead time is equal to 7 days. We will focus on one category of products, "FOODS" (which has 14,370 series), to have an inventory simulation of a wide ranging yet coherent set of products. We have chosen this category because of the importance of food products in retail: producing accurate forecasts in this category is crucial for reducing both costs and waste. We will also assume that the products that are on hand at the end of the week are discarded (i.e. they are perishable and last only one week). In terms of data split and approaches used, we maintain the same setting as in Section 5, i.e. we keep the last 14 observations as the holdout set to test the performance of models and filters in terms of inventory costs. We do not include Croston and TSB, because being forecasting methods rather than filters or models, they do not support predictive distributions and thus it is not possible to generate quantiles from them. For each time series, we measure the performance of models in two periods: for the first week and for the second one.

To evaluate the performance of models we measure the following:

- Achieved Cycle Service Level, with targets of 90%, 95% and 99%,
- Scaled lost sales,
- Scaled inventory on hand.

The scaling of lost sales and inventory on hand is needed in order to calculate the aggregate measures. We use the same scaling as Kourentzes et al. (2020), dividing the values by the in-sample mean demand.

Fig. 4 summarises the performance of models in terms of lost sales and inventory on hand. The closer the curve is to the origin, the better the performance of the model is. As we see, the best performing model in this respect is iETS$_O$ - it has lowest lost sales and lowest inventory on hand for each of the service levels. The HSP filter produces forecasts with the highest excessive inventory, but at the same time with the lowest lost sales. It can be considered as one of the least balanced approaches in our experiment. NegBin is doing better than ETS(A,N,N), but similar to some iETS models, such as iETS$_I$, iETS$_A$ and iETS$_G$, not being able to outperform iETS$_O$. Finally, we note that iETS$_F$ is performing worse than all the other models, which indicates that the mechanism for the probability update is important and contributes to models in terms of inventory performance.

When we consider rounding up of quantiles (second plot in Fig. 4), the performance of all iETS models improves and all of them, except for iETS$_F$, outperform the benchmarks ETS(A,N,N), HSP and NegBin.

In terms of achieved service level, we can see from Fig. 5 that there are some iETS models that perform better than others. Arguably, iETS$_O$ is more balanced than the others, because while it over-performs for lower service levels (90% and 95%), it gets close to the nominal in the case of 99%. iETS$_I$, iETS$_F$ and iETS$_D$ tend to under-perform, achieving a service level lower than the target. At the same time, HSP and ETS(A,N,N) tend to overshoot the nominal level due to higher inventory on hand, as we saw from Fig. 4. Finally, NegBin performs

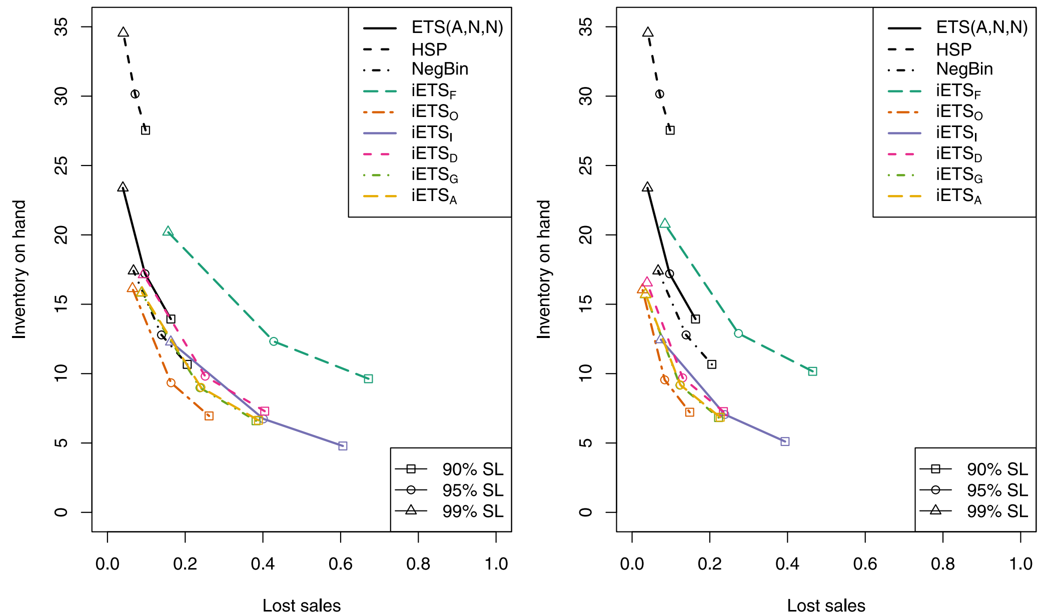

Fig. 4. Performance of models on FOODS category from M5 data in terms of lost sales and inventory on hand. Left plot shows performance of conventional iETS models, right plot shows iETS models with rounding up.

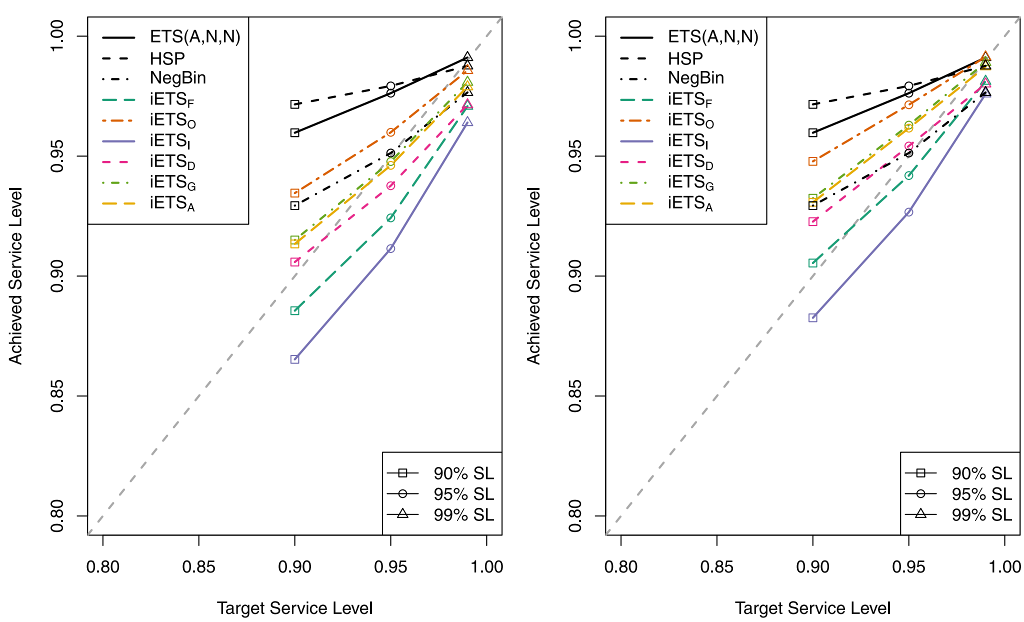

Fig. 5. Performance of models on FOODS category from M5 data in terms of service level. Left plot shows performance of conventional iETS models, right plot shows iETS models with rounding up.

on average well in terms of service level, overshooting slightly on the lower levels and undershooting on the higher ones.

The situation changes slightly when rounding up is introduced (right plot in Fig. 5). We can see that in that situation all iETS models achieve a higher service level with only iETS$_I$ and iETS$_F$ lagging behind. This agrees with the findings related to the coverage, discussed at the end of Section 5.3.

Given the discussion in Section 3.3, we can conclude that the iETS$_O$ is more appropriate for this data, because in many cases the demand is building up, which is then efficiently captured by this model.

Overall, as we see, the proposed iETS models perform very well in terms of inventory measures. While there is a clear winner (iETS$_O$), other models also perform well. Note also that iETS$_A$ can be considered as a robust alternative, performing well in many cases, although not being able to outperform iETS$_O$. We argue that in case of a doubt, researchers should use the model with automatic selection, and if quantile forecasts are required, then rounding up should be preferred.

7. Conclusions

In this paper, we have proposed a new statistical model for intermittent demand. This model expands the Hyndman et al. (2008) taxonomy by the inclusion of intermittent models and unites the worlds of continuous and intermittent data. 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). We have discussed both the demand sizes and the demand occurrence parts of the model, demonstrating the potential for the extendability of the model in both directions. We also proposed mechanisms for model construction, estimation and selection.

This paper was focused on the ETS(M,N,N) model and the intermittent equivalent of this model was called iETS(M,N,N). The most general model, "iETS$_G$", is the most complicated one but, at the same time, is the most flexible. We then discussed several special cases of this model, appearing when specific restrictions on the parameters of the original model are imposed. The simplest state space model arising from this is the one with the fixed probability (denoted as "iETS$_F$"), which is very easy to estimate and use. We then discussed two more complicated ones, namely odds ratio "iETS$_O$", and the inverse odds ratio "iETS$_I$" models, showing that the "iETS$_I$" model can be estimated by Croston's method. Finally, we proposed "iETS$_D$" model, which can be estimated using the TSB method. 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. Now a demand planner does not need to make a preliminary time series classification, trying to understand whether they are dealing with intermittent demand or not — this can be done automatically inside the iETS model.

Lastly, the experiment on the M5 data shows that the proposed approach is applicable to real life problems and that the proposed models can perform very well on real-world datasets, outperforming the conventional intermittent demand approaches in terms of point forecasts and prediction intervals. iETS$_I$ and iETS$_O$ were generally the most robust forecasting models, and it can be concluded that the continuous iETS models performed very well overall. We also found that rounding up quantiles from the iETS models improves their performance in terms of prediction intervals. This does not contradict the idea of the model and can be considered as a simple and logical step that could be used in practice. Finally, the inventory simulations showed that iETS models outperform the standard benchmarks, leading to lower inventory on hand and lost sales, getting closer to the target service level. The iETS$_O$ model performed particularly well, beating other approaches in terms of inventory measures. We also found, across the experiments, that iETS with automatic model selection was a robust option, performing well according to a variety of measures. We argue that if it is not clear, which of the models to use, the iETS with automatic selection is a viable alternative to other iETS models.

We should remark that the focus of this paper was on a specific iETS(M,N,N) model with several special cases. 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 general model, discussed in the paper can be denoted as iETS(M,N,N)$_G$(M,N,N)(M,N,N), where the letters in the first brackets indicate the type of ETS model for demand sizes and the letters after the subscript "G" refer to the models for the variables $\mu_{a,t}$ and $\mu_{b,t}$ respectively in the demand occurrence part of the model. Using this notation, new types of models can be studied in future research. For example, a model with additive trend in demand sizes and multiplicative trend in the demand occurrence with the odds ratio mechanism can be denoted as iETS(M,A,N)$_O$(M,M,N) (note that the third bracket is omitted because there is only one variable in the demand occurrence part of the model). This allows the extension of the Hyndman et al. (2008) taxonomy and opens new avenues for research.

While this paper focused on inventory application of intermittent demand models, we would like to point out that the context of application is much wider than just inventory and includes other areas, such as healthcare (e.g. for staff rostering), energy (e.g. for electricity load planning), and other.

Looking at the more technical side of the model, although iETS works well in real life applications, it has also been shown that the proposed construction and estimation procedures inevitably lead to positive bias in parameter estimation, leading to higher than required smoothing parameters. The same phenomenon should be applicable to the conventional intermittent demand approaches. How to fix this bias and whether to fix it at all, is one of the topics for future 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, diffusion and other models, which could be applied to a wide range of time series. In fact, any statistical model can be applied either to the demand sizes, or to the $\mu_{a,t}$, or $\mu_{b,t}$ parts of the model, as long as the necessary transformations are made, ensuring that the model produces positive values. For example, working with the pure additive ETS models on the log-transformed demand sizes, would simplify the derivations and could give potential benefits in terms of the ease of use of the model. Studying properties of these models would be another large area of research.

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. While this assumption is practical, lifting it would imply derivation of another model, with different properties than the iETS model.

Data availability

The openly available M5 dataset was used in the paper

Acknowledgements

We would like to thank Stephan Kolassa and Nikolaos Kourentzes for their comments to this paper, which helped to improve it substantially.

Appendix A. Proxies for the error terms of iETS model

The probability of occurrence in the model (14) at time t is defined as:

$$p_t = \frac{\mu_{a,t}}{\mu_{a,t} + \mu_{b,t}}. \tag{A.1}$$

Based on this formula, either $\mu_{a,t}$ or $\mu_{b,t}$ can be calculated if the probability is known:

$$\mu_{a,t} = \mu_{b,t} \frac{p_t}{1 - p_t} \tag{A.2}$$

and

$$\mu_{b,t} = \mu_{a,t} \frac{1 - p_t}{p_t}. \tag{A.3}$$

While the probability is never known, it can be estimated on observation t, and the value $\hat{p}_t$ can be used instead of p_t for the calculation of the error, assuming that when $o_t = 1$, the probability should be as close to one as possible, and when $o_t = 0$, the probability should be as close to zero as possible. Based on this idea the following error can be calculated:

$$v_t = o_t - \hat{p}_t. \tag{A.4}$$

This error lies in the interval $(-1, 1)$, depending on the values of o_t and $\hat{p}_t$. We then transform the variable in order to make it lie in the interval $(0, 1)$:

$$u_t = \frac{1 + v_t}{2}. \tag{A.5}$$

In this way the value of $u_t = 0.5$ corresponds to the ideal situation, when the predicted and the actual outcomes are equal (e.g. demand has occurred and the probability is equal to one). In the boundary case when $\hat{p}_t = 1$ and $o_t = 0$, the error proxy (A.5) is equal to 0, while in the opposite case it is equal to 1. In all the other cases it lies between zero and one.

Now inserting this new variable u_t in (A.2) and (A.3) instead of p_t and setting the unobserved $\mu_{a,t}$ and $\mu_{b,t}$ to one (which implies independence of the two models), we obtain the following two proxies of the error terms:

$$1 + e_{a,t} = \frac{u_t}{1 - u_t} \tag{A.6}$$

and

$$1 + e_{b,t} = \frac{1 - u_t}{u_t} \tag{A.7}$$

respectively.

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

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 depends on the previous observations of y_t and leads to the likelihood (based on the likelihood in Hyndman et al., 2008, p.35):

$$L(\boldsymbol{\theta}, \lambda_\epsilon | y_t, o_t = 1) = p_t f_z(z_t | l_{z,t-1}), \tag{B.1}$$

where $f_z(\cdot)$ is the probability density function of the log normal distribution and $\boldsymbol{\theta}$ is the vector of all the parameters of the model. In the latter case it is similarly equal to:

$$L(\boldsymbol{\theta}, \lambda_\epsilon | y_t, o_t = 0) = (1 - p_t) f_z(z_t | l_{z,t-1}). \tag{B.2}$$

The likelihood function for the statistical model (14) for all the T observations, is then:

$$L(\boldsymbol{\theta}, \lambda_\epsilon | \mathbf{Y}) = \prod_{o_t=1} p_t \prod_{o_t=0} (1 - p_t) \prod_{t=1}^{T} f_z(z_t | l_{z,t-1}), \tag{B.3}$$

where $\mathbf{Y}$ is the set of all the available variables y_t. However, the likelihood (B.3) cannot be used for the estimation of the model, because z_t is not observable when $o_t = 0$, thus making the estimation of the probability density function in those points not tractable. This means that the cases of z_t when $o_t = 0$ should be treated as missing values, and the likelihood should take into account the uncertainty about the distribution of z_t, conditional on $l_{z,t-k}$, where z_{t-k} is the last observed demand size. There are two options of calculating the likelihood in this case:

1. Calculate the marginal probability density function of $f_z\left(z_t | l_{z,t-k}\right)$ (similar to the derivation by Barnea et al., 2006):

$$f_z\left(z_t | l_{z,t-k}\right) = \int_0^\infty \dots \int_0^\infty f_z\left(z_t | l_{z,t-1}\right) \prod_{j=1}^{k-1} f_l\left(l_{z,t-j} | l_{z,t-j-1}\right) dl_{z,t-1} \dots dl_{z,t-k+1}, \tag{B.4}$$

   where $f_z\left(z_t | l_{z,t-1}\right)$ is the probability density function of the demand sizes (based on the log normal distribution) and $f_l\left(l_{z,t-j} | l_{z,t-j-1}\right)$ is the probability density function of the level (based on the three parameter log normal distribution).
2. Calculate the likelihood using Expectation Maximisation (EM) algorithm. In this case the expectation of the logarithm of the (B.3) is taken, which is then maximised.

We will be using the latter approach, because the former does not have analytical solutions and involves numerical optimisation, for example, based on Monte Carlo simulations, which is more computationally expensive than the second option. The EM algorithm, applied to our problem, leads to the following expectation of the log-likelihood:

$$\begin{aligned} \ell(\boldsymbol{\theta}, \lambda_\epsilon | \mathbf{Y}) = & \sum_{o_t=1} \log f_z\left(z_t | l_{z,t-1}\right) + \sum_{o_t=0} \mathrm{E}\left(\log f_z\left(z_t | l_{z,t-1}\right)\right) \\ & + \sum_{o_t=1} \log(p_t) + \sum_{o_t=0} \log(1 - p_t) \end{aligned}. \tag{B.5}$$

The log-likelihood (B.5) can be simplified further, given that $\mathcal{H}(z_t) = -\mathrm{E}\left(\log f_z\left(z_t | l_{z,t-1}\right)\right)$ is the differential entropy (Lazo and Rathie, 1978). For each distribution, the entropy will be different. We do not discuss them here as they are well-known for all the distributions discussed in this paper.

For each specific distribution, a concentrated log-likelihood can be obtained based on (B.5) by inserting the PDF and the differential entropy and then inserting the MLE of scale of each distribution. We do not provide these details here. Finally, the probability p_t is not known and its predicted value can be used instead to get the concentrated log-likelihood:

$$\begin{aligned} \ell(\boldsymbol{\theta}, \lambda_\epsilon | \mathbf{Y}) = & \sum_{o_t=1} \log f_z\left(z_t | l_{z,t-1}\right) + \sum_{o_t=0} \mathcal{H}_z(z_t) \\ & + \sum_{o_t=1} \log(\hat{p}_t) + \sum_{o_t=0} \log(1 - \hat{p}_t) \end{aligned}. \tag{B.6}$$

Appendix C. Bias in the estimation of iETS models based on the example of iETS$_\mathrm{F}$

In order to see the bias in the estimation of the smoothing parameter in iETS models, consider a case, when the level needs to be updated on some observation t after having $j_t - 1$ zeroes, where j_t is in fact the demand interval for the observation t. Given that the update is done using the conditional expectation of $l_{z,t|t-j_t}$, the formula will be:

$$\hat{l}_{z,t} = \hat{l}_{z,t-j_t}(1 + \hat{\alpha}_z e_{z,t}), \tag{C.1}$$

where $e_{z,t} = \frac{z_t}{\hat{l}_{z,t-j_t}} - 1$. At the same time the true level $l_{z,t}$ can be represented as a product of the previous error terms:

$$\begin{aligned} l_{z,t} = & l_{z,t-1}(1 + \alpha_z \epsilon_{z,t}) = l_{z,t-2}(1 + \alpha_z \epsilon_{z,t-1})(1 + \alpha_z \epsilon_{z,t}) = \\ & l_{z,t-j_t} \prod_{i=0}^{j_t-1} (1 + \alpha_z \epsilon_{z,t-i}), \end{aligned} \tag{C.2}$$

and the respective actual value z_t can be expressed in terms of the error term and the true value of level (C.2):

$$z_t = l_{z,t-1}(1 + \epsilon_{z,t}) = l_{z,t-j_t} \prod_{i=1}^{j_t-1} (1 + \alpha_z \epsilon_{z,t-i})(1 + \epsilon_{z,t}).$$

The forecast error, used in the update of states of the constructed model is:

$$e_{z,t} = \frac{z_t}{\hat{l}_{z,t-j_t}} - 1 = \frac{l_{z,t-j_t}}{\hat{l}_{z,t-j_t}} \prod_{i=1}^{j_t-1} (1 + \alpha_z \epsilon_{z,t-i})(1 + \epsilon_{z,z,t}) - 1, \tag{C.3}$$

and the updated state (C.1) in the constructed model using (C.3) is:

$$\hat{l}_{z,t} = \hat{l}_{z,t-j_t} \left(1 + \hat{\alpha_z} \left(\frac{l_{z,t-j_t}}{\hat{l}_{z,t-j_t}} \prod_{i=1}^{j_t-1} (1 + \alpha_z \epsilon_{z,t-i})(1 + \epsilon_{z,t}) - 1 \right) \right). \tag{C.4}$$

The asymptotic expectation of the forecast error (C.3) is equal to zero, because asymptotically the estimated level will converge to the true one, meaning that the ratio $\frac{l_{z,t-j_t}}{\hat{l}_{z,t-j_t}} = 1$. Furthermore, the error term $1 + \epsilon_{z,t}$ has a unit mean by definition. This implies that the conditional expectation of the level $\hat{l}_{z,t}$ is equal to $\hat{l}_{z,t-j_t}$, which is similar to the conditional expectation of the true level (C.2): $l_{z,t} = l_{z,t-j_t}$. However, the variance of (C.4) conditional on the values available at time $t - j_t$ will differ between the true and the estimated models (based on the formula (10)):

$$\begin{aligned} \mathrm{V}\left(\hat{l}_{z,t} | \hat{l}_{z,t-j_t}\right) = & \hat{l}^2_{z,t-j_t} \hat{\alpha_z}^2 \mathrm{V}\left(\prod_{i=1}^{j_t-1} (1 + \alpha_z \epsilon_{z,t-i})(1 + \epsilon_{z,t}) \right) = \\ & \hat{l}^2_{z,t-j_t} \hat{\alpha_z}^2 \left(\left(1 + \alpha_z^2 s_\epsilon^2\right)^{j_t-1} (1 + s_\epsilon^2) - 1 \right), \end{aligned} \tag{C.5}$$

where s_ϵ^2 is the estimate of σ_ϵ^2, the variance of the error term. At the same time, the true variance of the level based on (C.2) is:

$$\begin{aligned} \mathrm{V}\left(l_{z,t} | l_{z,t-j_t}\right) = & \mathrm{V}\left(l_{z,t-j_t} \prod_{i=0}^{j_t-1} (1 + \alpha_z \epsilon_{z,t-i}) \middle| l_{z,t-j_t} \right) = \\ & l^2_{z,t-j_t} ((1 + \alpha_z^2 \sigma_\epsilon^2)^{j_t} - 1). \end{aligned} \tag{C.6}$$

Table E.3
Performance of models in terms of quantile forecasts for the 90% level.

Approaches	Mean values			Median values	
	sMIS	Pinball	Coverage	sMIS	Pinball
iETS_F	2.061	4.876	0.870	1.238	1.038
iETS_O	**1.452**	**3.110**	0.919	0.742	0.380
iETS_I	1.644	4.544	0.869	**0.504**	**0.177**
iETS_D	1.773	4.077	0.888	0.991	0.672
iETS_G	1.680	3.969	**0.889**	0.853	0.503
iETS_A	1.689	3.991	**0.889**	0.878	0.530
HSP	2.307	4.046	0.954	1.994	2.580
NegBin	1.880	3.189	0.961	1.728	1.834
ETS(A,N,N)	2.486	4.670	0.969	2.251	3.173
ETS(X,X,N)	2.486	4.670	0.969	2.251	3.173
		Rounded up quantiles			
iETS_F	2.104	3.970	0.935	1.680	1.833
iETS_O	1.592	**2.862**	0.959	1.541	1.416
iETS_I	**1.511**	3.403	**0.928**	**1.176**	**0.820**
iETS_D	1.874	3.382	0.945	1.604	1.581
iETS_G	1.773	3.222	0.946	1.545	1.466
iETS_A	1.787	3.269	0.946	1.563	1.496

Table E.4
Performance of models in terms of quantile forecasts for the 99% level.

Approaches	Mean values			Median values	
	sMIS	Pinball	Coverage	sMIS	Pinball
iETS_F	4.112	2.798	0.930	1.696	0.193
iETS_O	**2.787**	1.556	0.964	1.439	0.137
iETS_I	3.884	3.035	0.915	**1.174**	**0.090**
iETS_D	3.565	2.292	0.941	1.503	0.149
iETS_G	3.455	2.252	0.942	1.372	0.125
iETS_A	3.474	2.275	0.941	1.412	0.132
HSP	3.149	0.965	0.984	2.559	0.431
NegBin	2.867	**0.744**	0.994	2.535	0.406
ETS(A,N,N)	3.235	0.885	**0.992**	2.773	0.503
ETS(X,X,N)	3.235	0.885	0.992	2.772	0.503
		Rounded up quantiles			
iETS_F	3.393	1.528	0.972	2.009	0.270
iETS_O	**2.606**	**0.888**	**0.985**	1.987	0.241
iETS_I	3.255	1.578	0.964	1.906	0.227
iETS_D	3.004	1.234	0.977	1.936	0.232
iETS_G	2.932	1.209	0.977	**1.905**	**0.220**
iETS_A	2.945	1.221	0.977	1.915	0.224

If the estimator used for the model is consistent then the estimate of variance of the error term will converge to its true value, the estimated level will coincide with the true one, and in the perfect situation the conditional variances (C.5) and (C.6) will be equal, implying that:

$$(1+\alpha_z^2\sigma_\epsilon^2)^{j_t} - 1 = \hat{\alpha_z}^2\left(\left(1+\alpha_z^2\sigma_\epsilon^2\right)^{j_t-1}(1+\sigma_\epsilon^2) - 1\right),$$

which in turn means that the estimated smoothing parameter $\hat{\alpha_z}$ is equal to:

$$\hat{\alpha_z} = \sqrt{\frac{(1+\alpha_z^2\sigma_\epsilon^2)^{j_t} - 1}{\left(1+\alpha_z^2\sigma_\epsilon^2\right)^{j_t-1}(1+\sigma_\epsilon^2) - 1}}, \tag{C.7}$$

where the positive root is taken because the smoothing parameter is always restricted to the (0, 1) range (see, for example, Svetunkov and Boylan, 2022). In the iETS model with fixed probability it can be taken that $j_t = \frac{1}{p}$, so the formula (C.7) transforms into:

$$\hat{\alpha_z} = \sqrt{\frac{(1+\alpha_z^2\sigma_\epsilon^2)^{\frac{1}{p}} - 1}{\left(1+\alpha_z^2\sigma_\epsilon^2\right)^{\frac{1-p}{p}}(1+\sigma_\epsilon^2) - 1}}, \tag{C.8}$$

This implies that, even under perfect conditions, the estimated smoothing parameter will converge asymptotically to (C.8) instead of the true α_z. Thus it will always be positively biased: the smaller the probability p is, the larger the bias becomes. However, the bias will disappear in the case of the $p = 1$, because according to (C.8) the estimated smoothing parameter becomes equal to the true α_z.

Appendix D. 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:

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

and

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

where n is the quantile of the distribution of rounded up values (the smallest integer number that satisfies the inequality (D.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:

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

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 (D.3), the probability (D.2) can be rewritten as:

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

Note also that the following is true:

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

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

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

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

$$\lceil Q_\alpha(z_t) \rceil = Q_\alpha\left(\lceil z_t \rceil\right). \tag{D.7}$$

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:

$$\lfloor Q_\alpha(z_t) \rfloor = Q_\alpha\left(\lfloor z_t \rfloor\right). \tag{D.8}$$

Appendix E. Quantile forecasts for M5

See Tables E.3 and E.4.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Control 19 (6), 716–723.

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

Barnea, O., Solow, A.R., Stone, L., 2006. On fitting a model to a population time series with missing values. Israel J. Ecol. Evol. 52, 1–10.

Brown, R.G., 1956. Exponential Smoothing for Predicting Demand. Arthur D. Little, Inc, Cambridge 42, Massachusetts, URL http://legacy.library.ucsf.edu/tid/dae94e00.

Chatfield, C., Koehler, A.B., Ord, J.K., Snyder, R.D., 2001. A new look at models for exponential smoothing. J. R. Stat. Soc. Ser. D (the Statistician) 50 (2), 147–159.

Croston, J.D., 1972. Forecasting and stock control for intermittent demands. Oper. Res. Q. 23 (3), 289.

Davis, R.A., Dunsmuir, W.T.M., Streett, S.B., 2003. Observation-driven models for Poisson counts. Biometrika 90 (4), 777–790.

Fokianos, K., Rahbek, A., Tjøstheim, D., 2009. Poisson autoregression. J. Amer. Statist. Assoc. 104 (488), 1430–1439.

Fokianos, K., Tjøstheim, D., 2011. Log-linear Poisson autoregression. J. Multivariate Anal. 102 (3), 563–578.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. J. Amer. Statist. Assoc. 102 (477), 359–378.

Harvey, A.C., Fernandes, C., 1989. Time series models for count or qualitative observations. J. Bus. Econom. Statist. 7 (4), 407–417.

Hyndman, R.J., Koehler, A.B., Ord, J.K., Snyder, R.D., 2008. Forecasting with Exponential Smoothing. In: 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. Int. J. Forecast. 13 (8), 439–454.

Kachour, M., 2014. On the rounded integer-valued autoregressive process. Comm. Statist. Theory Methods 43 (2), 355–376.

Kachour, M., Yao, J.F., 2009. First-order rounded integer-valued autoregressive (RINAR(1)) process. J. Time Series Anal. 30 (4), 417–448.

Kourentzes, N., 2014. On intermittent demand model optimisation and selection. Int. J. Prod. Econ. 156, 180–190.

Kourentzes, N., Petropoulos, F., 2016. Tsintermittent: Intermittent time series forecasting. URL https://CRAN.R-project.org/package=tsintermittent, R package version 1.9.

Kourentzes, N., Trapero, J.R., Barrow, D.K., 2020. Optimising forecasting models for inventory planning. Int. J. Prod. Econ. 225 (November 2019), 107597.

Lazo, A., Rathie, P., 1978. On the entropy of continuous probability distributions (Corresp.). IEEE Trans. Inform. Theory 24 (1), 120–122.

Makridakis, S., Spiliotis, E., Assimakopoulos, V., 2021. The M5 competition: Background, organization, and implementation. Int. J. Forecast. (xxxx).

Ord, K.J., Koehler, A.B., Snyder, R.D., 1997. Estimation and prediction for a class of dynamic nonlinear statistical models. J. Amer. Statist. Assoc. 92 (March 2013), 1621–1629.

Petropoulos, F., Kourentzes, N., 2015. Forecast combinations for intermittent demand. J. Oper. Res. Soc. 66 (6), 914–924.

R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.

Schultz, C.R., 1987. Forecasting and inventory control for sporadic demand under periodic review. J. Oper. Res. Soc. 38 (5), 453–458.

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

Snyder, R.D., 1985. Recursive estimation of dynamic linear models. J. R. Stat. Soc. Ser. B Stat. Methodol. 47 (2), 272–276.

Snyder, R.D., 2002. Forecasting sales of slow and fast moving inventories. European J. Oper. Res. 140 (3), 684–699.

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

Sugiura, N., 1978. Further analysis of the data by Akaike' s information criterion and the finite corrections. Comm. Statist. Theory Methods 7 (1), 13–26.

Svetunkov, I., 2019. Counter: Modelling and forecasting of count data. URL https://github.com/config-i1/counter, R package version 0.2.0.

Svetunkov, I., 2022. Smooth: Forecasting using state space models. R package version 3.2.0, URL https://github.com/config-i1/smooth.

Svetunkov, I., 2023. Forecasting and analytics with the Augmented Dynamic Adaptive Model (ADAM). Chapman and Hall/CRC, https://openforecast.org/adam/.

Svetunkov, I., Boylan, J.E., 2022. Dealing with Positive Data Using Pure Multiplicative ETS Models. 3, Department of Management Science, Lancaster University, pp. 1–29.

Svetunkov, I., Chen, H., Boylan, J.E., 2023. A new taxonomy for vector exponential smoothing and its application to seasonal time series. European J. Oper. Res. 304 (3), 964–980.

Syntetos, A.A., Boylan, J.E., 2001. On the bias of intermittent demand estimates. Int. J. Prod. Econ. 71 (1–3), 457–466.

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

Tadikamalla, P.R., 1981. The inverse Gaussian approximation to the lead time demand. Int. J. Prod. Res. 19 (2), 213–219.

Teunter, R.H., Syntetos, A.A., Babai, M.Z., 2011. Intermittent demand: Linking forecasting to inventory obsolescence. European J. Oper. Res. 214 (3), 606–615.