IMA Journal of Management Mathematics (2023) **00**, 1–20
https://doi.org/10.1093/imaman/dpad028

Staying Positive: Challenges and Solutions in Using Pure Multiplicative ETS Models

IVAN SVETUNKOV*
Centre for Marketing Analytics and Forecasting, Department of Management Science, Lancaster University, Lancaster University Management School, LA1 4YX, UK

AND

JOHN E. BOYLAN
Centre for Marketing Analytics and Forecasting, Department of Management Science, Lancaster University, Lancaster University Management School, LA1 4YX, UK
*Corresponding author: i.svetunkov@lancaster.ac.uk

[Received on 21 May 2023; revised on 2 October 2023; accepted on Date Month Year]

Exponential smoothing in state space form (ETS) is a popular forecasting technique, widely used in research and practice. While the additive error ETS models have been well studied, the multiplicative error ones have received much less attention in forecasting literature. Still, these models can be useful in cases, when one deals with positive data, because they are supposed to work in such situations. Unfortunately, the classical assumption of normality for the error term might break this property and lead to non-positive forecasts on positive data. In order to address this issue we propose using Log-Normal, Gamma and Inverse Gaussian distributions, which are defined for positive values only. We demonstrate what happens with ETS(M,*,*) models in this case, discuss conditional moments of ETS with these distribution and show that they are more natural for the models than the Normal one. We conduct the simulation experiments in order to study the bias introduced by point forecasts in these models and then compare the models with different distributions. We finish the paper with an example of application, showing how pure multiplicative ETS with a positive distribution works.

Keywords: exponential smoothing; forecasting; state space model.

1. Introduction

Exponential smoothing is one of the most popular models used for demand forecasting in practice [32]. It was originally proposed by [5] and [7], developed over the years by [20] and [16] as a Single Source of Error model, and finally presented as a holistic framework by [8]. This framework includes Error, Trend and Seasonal components, thus being called ‘ETS’. Different components of the model can be ‘A’ - additive, ‘M’ - multiplicative, or ‘N’ - none (with exception of the error term, which can only be either ‘A’ or ‘M’, and the trend component, which can also be ‘Ad’ - damped additive and ‘Md’ - damped multiplicative). The developed taxonomy presents 30 models based on different types of components, making the framework flexible and applicable to a wide variety of time series including positive, negative and mixed data. Since then, the ETS model has become popular in forecasting community, performing well in different competitions [3, 14] and being used as the base model for development of new approaches [see, for example, 11, 17, 25, 26, 27] and as one of the common benchmarks [for example, 4, 18, 31].

However, one of the main limitations of the approach of [9] is that it is built on the assumption of normality of the error term for both additive and multiplicative models. As a result, the latter form is applicable only to the data with high positive values – in this case, the variance of the error term becomes small enough for the forecast values in the multiplicative error models not to become negative. Unfortunately, in case of the lower volume data (e.g. not thousands of units, but tens), the variance of the error term might become large, which might lead to issues with the model, making it produce unrealistic forecasts in some cases and completely inapplicable in the others. For example, this impacts the prediction intervals and might lead to negative point forecasts even if the data is strictly positive.

The pure multiplicative error models have been studied by [2], who mentioned that there were several possibilities for the distributions of the error term, including Normal, Truncated Normal, Log-Normal and Gamma. The authors never investigated the idea further and instead of developing the ETS(M,*,*) model with these distributions, proposed a modified ETS (a model in logarithms) and focused on investigating its properties. While this is a solution to the problem, it sidesteps the original ETS taxonomy, introducing the new group of models. This means that the mixed models in the original ETS framework do not benefit from the fix introduced by [2]. Furthermore, the approach expands the number of possible ETS models that can be used, making the model selection even more challenging. We argue that modifying the original ETS instead of introducing new models is a more sustainable approach, as it simplifies the model extension and the process of the model selection. There are no other papers on the subject, and the issue with normality has never been fully investigated in the literature. ETS(M,*,*) have been overlooked in terms of theoretical problems and practical implementations. These models are particularly important, when we observe skewness, especially with lower volume data in such contexts as supply chain and retail, where the normality assumption becomes unreasonable.

In this paper we propose using several positive distributions in the original ETS(M,*,*) models, showing how the models work with them. We study their properties and discuss how to estimate ETS model with them and then use in forecasting. But before moving to the distributions in ETS(M,*,*), we discuss the properties of pure multiplicative ETS models to understand what their main limitations are and how they might impact the applicability of the proposed assumption. We finish the paper with an example of application and a discussion of the managerial implications of the pure multiplicative ETS models.

2. Properties of multiplicative error ETS models

While the main properties of ETS models are well studies in [9], several important elements have been neglected over the years. These mainly apply to the conditional expectation of the models.

Discussing all the 30 ETS models, [9] mention on page 23 that, when either the trend, or the seasonal components are multiplicative, then the point forecasts for the horizon $h > 1$, produced by models might differ from the conditional expectations [9, pages 21 and 22]. In order to investigate this issue further, we start by studying the properties of pure multiplicative models, reformulating them using logarithms via the following state space equations [24, Chapter 6]:

$$\begin{aligned} \log y_t &= \mathbf{w}' \log(\mathbf{v}_{t-l}) + \log(1+\varepsilon_t) \\ \log \mathbf{v}_t &= \mathbf{F} \log \mathbf{v}_{t-l} + \log(\mathbf{1}_k + \mathbf{g}\varepsilon_t) \end{aligned}, \tag{2.1}$$

where y_t is the actual value on observation t, $\mathbf{w}$ is the measurement vector, $\mathbf{v}_t$ is the state vector, $\mathbf{F}$ is the transition matrix, $\mathbf{g}$ is the persistence vector, ε_t is the error term, l is the lag vector, $\mathbf{1}_k$ is the vector of ones of the size k and log is the natural logarithm. Note that the logarithm and exponent in the formulae (2.1) are taken element-wise. This formulation differs from the original [9] one, because we use the lags of components l, while the original one uses a bigger transition matrix, tracking how the seasonality changes from one observation to another. The reason for this formulation is its compactness, because the transition matrix in the model (2.1) contains only up to 3^2 elements versus the $(2+m)^2$ in the original formulation of the ETS. At the same time, this compactness does not impact specific models, which do not differ between the approach (2.1) and the conventional one. An example of a model that can be written in the state space form based on (2.1) is ETS(M,Md,M) [originally proposed by 28]:

$$\begin{aligned} \log y_t &= \log l_{t-1} + \phi \log b_{t-1} + \log s_{t-m} + \log(1+\varepsilon_t) \\ \log l_t &= \log l_{t-1} + \phi \log b_{t-1} + \log(1+\alpha\varepsilon_t) \\ \log b_t &= \phi \log b_{t-1} + \log(1+\beta\varepsilon_t) \\ \log s_t &= \log s_{t-m} + \log(1+\gamma\varepsilon_t) \end{aligned} , \tag{2.2}$$

for which:

$$\mathbf{w} = \begin{pmatrix} 1 \\ \phi \\ 1 \end{pmatrix}, \mathbf{v}_t = \begin{pmatrix} l_t \\ b_t \\ s_t \end{pmatrix}, l = \begin{pmatrix} 1 \\ 1 \\ m \end{pmatrix}, \mathbf{v}_{t-l} = \begin{pmatrix} l_{t-1} \\ b_{t-1} \\ s_{t-m} \end{pmatrix}, \mathbf{F} = \begin{pmatrix} 1 & \phi & 0 \\ 0 & \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \mathbf{g} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}, \mathbf{1}_k = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

In this case l_t is the level, b_t is the trend, s_t is the seasonal components, ϕ is the dampening parameter, α, β and γ are the smoothing parameters and m is the seasonal lag. The model (2.2) also shows that there is a connection between the pure multiplicative ETS models and the pure additive ones applied to the data in logarithms, because in case of low values of smoothing parameters and the variance of the error term, the following approximation holds:

$$\log(\mathbf{1}_k + \mathbf{g}\varepsilon_t) \approx \mathbf{g}\varepsilon_t. \tag{2.3}$$

Note that by taking exponent in the model (2.2), we obtain the original model as formulated by [9]:

$$\begin{aligned} y_t &= l_{t-1} b_{t-1}^{\phi} s_{t-m}(1+\varepsilon_t) \\ l_t &= l_{t-1} b_{t-1}^{\phi}(1+\alpha\varepsilon_t) \\ b_t &= b_{t-1}^{\phi}(1+\beta\varepsilon_t) \\ s_t &= s_{t-m}(1+\gamma\varepsilon_t) \end{aligned} . \tag{2.4}$$

In order to obtain the conditional h steps ahead expectation (given the information on the observation t), we need to have the following recursive formula, obtained directly from (2.1):

$$\begin{aligned} \log y_{t+h} =& \mathbf{w}_1' \mathbf{F}_1^{h-1} \log \mathbf{v}_t + \sum_{j=1}^{h-1} \mathbf{w}_1' \mathbf{F}_1^{j-1} \log(\mathbf{1}_k + \mathbf{g}_1 \varepsilon_{t+h-j}) + \\ & \mathbf{w}_m' \mathbf{F}_m^{\lceil \frac{h}{m} \rceil - 1} \log \mathbf{v}_t + \sum_{j=1}^{\lceil \frac{h}{m} \rceil - 1} \mathbf{w}_m' \mathbf{F}_m^{j-1} \log(\mathbf{1}_k + \mathbf{g}_m \varepsilon_{t+m\lceil \frac{h}{m} \rceil - j}) + \log(1+\varepsilon_{t+h}), \end{aligned} \tag{2.5}$$

where $\lceil \cdot \rceil$ is the ceiling function, the subscript 1 corresponds to the non-seasonal part of the model, and m corresponds to the seasonal ones (e.g. the vector $\mathbf{g}_m$ contains zeroes for α and β and non-zero γ if the seasonal model is under consideration). An example for the same ETS(M,Md,M) model is:

$$\mathbf{w}_1 = \begin{pmatrix}1\\1\\0\end{pmatrix}, \mathbf{w}_m = \begin{pmatrix}0\\0\\1\end{pmatrix}, \mathbf{F}_1 = \begin{pmatrix}1 & 1 & 0\\0 & 1 & 0\\0 & 0 & 0\end{pmatrix}, \mathbf{F}_m = \begin{pmatrix}0 & 0 & 0\\0 & 0 & 0\\0 & 0 & 1\end{pmatrix}, \mathbf{g}_1 = \begin{pmatrix}\alpha\\\beta\\0\end{pmatrix}, \mathbf{g}_m = \begin{pmatrix}0\\0\\\gamma\end{pmatrix},$$

Assuming that the error term is homoscedastic, not serially correlated and that $\log(1+\varepsilon_t) = 0$, the conditional expectation of the logarithm of the actual value h steps ahead based on (2.5) is:

$$\begin{aligned} \mathrm{E}(\log y_{t+h}|t) =& \mathbf{w}_1'\mathbf{F}_1^{h-1}\log\mathbf{v}_t + \sum_{j=1}^{h-1}\mathbf{w}_1'\mathbf{F}_1^{j-1}\mathrm{E}\left(\log\left(\mathbf{1}_k + \mathbf{g}_1\varepsilon_{t+h-j}\right)\right) + \\ & \mathbf{w}_m'\mathbf{F}_m^{\lceil\frac{h}{m}\rceil-1}\log\mathbf{v}_t + \sum_{j=1}^{\lceil\frac{h}{m}\rceil-1}\mathbf{w}_m'\mathbf{F}_m^{j-1}\mathrm{E}\left(\log\left(\mathbf{1}_k + \mathbf{g}_m\varepsilon_{t+m\lceil\frac{h}{m}\rceil-j}\right)\right). \end{aligned} \tag{2.6}$$

Its exponent is the geometric mean conditional on the information on the observation t. The main issue with the expectation (2.6) is that in general, in case, when $\mathbf{g}$ contains non-zero elements, **we cannot assume** that:

$$\mathrm{E}(\log(\mathbf{1}_k + \mathbf{g}\varepsilon_{t+h-j})) = \mathbf{0}_k, \tag{2.7}$$

where $\mathbf{0}_k$ is the vector of zeroes. So, the expectation (2.6) cannot be simplified any further. Even if we assume that all elements of $\mathbf{g}_i$ are zero in (2.7), making it hold, the conditional expectation of y_{t+h} will be greater than the exponent of the expectation (2.6) due to Jensen's inequality. At the same time, according to [9], the point forecast from the pure multiplicative models can be calculated as (based on our formulation):

$$\hat{y}_{t+h} = \exp\left(\left(\mathbf{w}_1'\mathbf{F}_1^{h-1} + \mathbf{w}_m'\mathbf{F}_m^{\lceil\frac{h}{m}\rceil-1}\right)\log\mathbf{v}_t\right), \tag{2.8}$$

which does not correspond to any known statistics of a distribution. In a very special case, when $\mathbf{g} = \mathbf{0}_k$, the point forecast (2.8) coincides with the conditional geometric expectation rather than the arithmetic one. Another thing to note is that the geometric expectation (2.6) is in general less than or equal to the point conditional expectation. On the other hand, it will coincide with the point forecast only, when condition (2.7) is satisfied. In all the other cases the geometric mean will be larger because the elements of $\mathbf{g}$ will be greater than zero. This means that the following condition should hold:

$$\hat{y}_{t+h} \leq \check{y}_{t+h} \leq \mu_{y,t+h}, \tag{2.9}$$

where $\mu_{y,t+h}$ is the conditional h steps ahead expectation and $\check{y}_{t+h}$ is the conditional h steps ahead geometric expectation. The are only three special cases, when the point forecast (2.8) will coincide with the geometric and the arithmetic expectations:

1. When $h = 1$ for all models;
2. When $h \leq m$ for the non-trended models with multiplicative seasonality;
3. For a special case, the model ETS(M,N,N).

(1) is important for the construction of the model, where the fitted values coincide with the one-step-ahead forecast. (3) becomes apparent, when studying the following recursion for the ETS(M,N,N)

model (based on the exponent of (2.5)):

$$y_{t+h} = l_t \prod_{j=1}^{h-1} (1+\alpha\varepsilon_{t+j})(1+\varepsilon_{t+h}), \tag{2.10}$$

the expectation of which is equal to l_t, as long as

$$\mathrm{E}(1+\varepsilon_t) = 1, \tag{2.11}$$

and the error term is i.i.d. The assumption about the expectation of the error term becomes important in ETS, because it guarantees that the one-step-ahead point forecasts correspond to the expectations of models. This is necessary for the construction of the model, where the fitted value is equal to one-step-ahead point forecast and is required to start the recursion. Furthermore, if it does not hold, then the transition equation in the holdout will differ from the typically assumed one, implying that the expectation of products of random variables needs to be calculated.

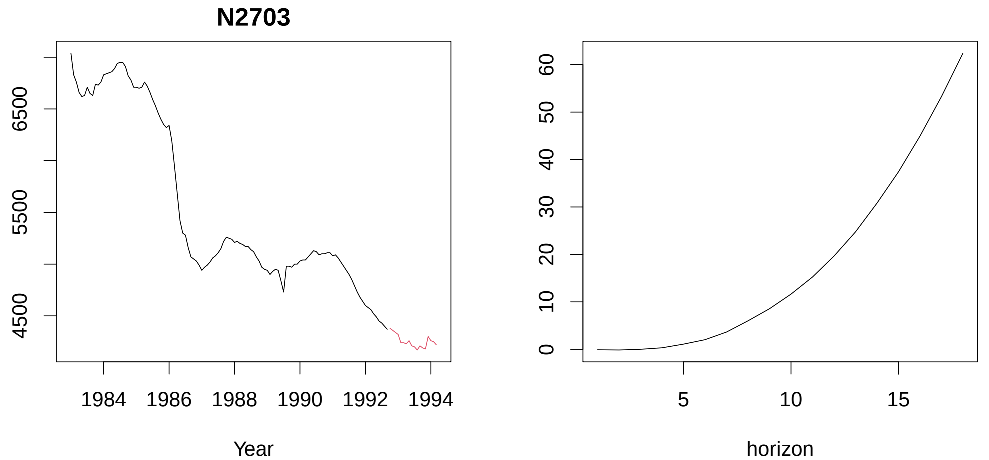
While [9] acknowledge the issue with the multiplicative error models, they claim that the difference between the point forecast and the expectation will in general be negligible. In order to counter this claim we propose to consider the following example of time series N2703 from M3 competition data [13]. The series itself is shown in Figure 1a and exhibits a negative trend, which can be modelled using ETS(M,M,N). This model is then applied to the data via `adam()` function from `smooth` package v3.2.0 [23] for R [30], producing one to 18 steps ahead forecasts. Figure 1b demonstrates the difference between the point forecast (using equation 2.9) and the simulated conditional expectation of the model on this series. It becomes apparent that on shorter horizons, the difference between the two values is negligible, but with the increase of the horizon, the difference between them grows exponentially. This example shows that in general the conventional formula (2.8) should not be used for point forecasts generation – the simulations should be done instead.

Analysing the formulae (2.6), (2.8) and (2.9), we argue that there are several elements that impact the conditional expectation, influencing the bias of the point forecasts of the model:

- With the increase of the values of smoothing parameters α, β and γ, the bias of point forecasts should increase;
- The distance between the conditional expectation and the point forecast will be larger for the larger values of the variance of the error term σ^2;
- Finally, the difference between the two will increase with the increase of the forecast horizon h.

As a side note, some of mixed ETS(M,*,*) models have been studied by [9] more thoroughly than the others. For example, ETS(M,A,M) and ETS(M,Ad,M) are discussed in Chapter 6 of their monograph. The authors suggest approximations for them, based on the assumption that the variance of the error term σ^2 is small. However, this assumption might be violated if one deals with low volume data, so the approximations proposed in the textbook might not be appropriate as well. As for the other mixed ETS(M,*,*) models, the conditional expectations for them do not have closed forms.

As a conclusion from this section, we argue that conditional expectation from the multiplicative error models should be produced using simulations, not relying on the conventional point forecasts formulae, because the latter might be misleading.



(a) Series N2703 from M3 data.

(b) Difference between the expectation and point forecast on the series.

FIG. 1. An example of ETS(M,M,N) on the Series 2703 from M3 data.

3. Simulation experiment 1: bias of point forecasts

In order to better understand what impacts the difference between the point forecast and the conditional expectation, we conduct a simulation experiment, investigating the influence of scale parameter, forecast horizon and values of smoothing parameters on the difference. We use ETS(M,Md,N) model, as it potentially exhibits the highest deviation of point forecasts from the conditional expectation. We consider the following options:

- Scale $\sigma = \{0.01, 0.05, 0.1, 0.5, 1\}$;
- Smoothing parameters: $\alpha, \beta = \{0.1, 0.2, 0.5\}$;
- Forecast horizon $h = \{10, 50, 100\}$;
- Dampening parameter for trend $\phi = \{0.5, 0.9, 0.95, 1\}$.

We then generated 100 observations from ETS(M,Md,N) model with an Inverse Gaussian distribution (see discussion of distributions in Section 4) using the `sim.es()` function from the `smooth` package v3.2.0 for R [23], and applied the same model with the known parameters on the data, producing point forecasts and conditional expectations. We then calculate the percentage difference between the two:

$$\text{PD} = \frac{\mu_{t+h} - \hat{y}_{t+h}}{\mu_{t+h}}, \tag{3.1}$$

thus measuring the deviation of the point forecasts from the conditional expectation. The value (3.1) will always lie between zero and one due to the inequality (2.9) and has a simple interpretation, i.e. the conditional expectation is PD$\times$100% higher than the point forecast. We then aggregate it over the forecast horizon (i.e. from $h = 1$ to the selected value) and repeat this experiment for 10,000 iterations in order to get a better understanding of the performance of the ETS(M,Md,N) model.

The distribution of the variable (3.1) is shown on Figure 2.

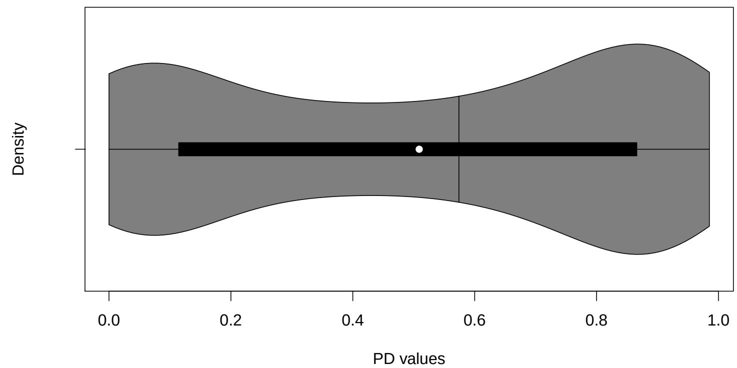


FIG. 2. Violinplot for the distribution of the percentage difference. The black vertical line corresponds to the median of the distribution, while the white dot is the mean value.

We can see that the percentage difference has two modes: one corresponds to the lower values of σ, α, β and horizon, while the upper corresponds to the higher values of these variables. In order to summarise the results and see the impact of these variables on the bias of point forecasts, we fit the regression model on percentage differences, taking the scale, smoothing parameters and horizon (scaled in order to lie between 0 and 1, similar to the other variables) as independent variables [using `alm` function from greybox package v1.0.3 for R, 22]. Before fitting the model, we transform the response variable into odds via:

$$\text{odds} = \frac{\text{PD}}{1-\text{PD}},$$

in order to change its domain from (0, 1) to (0, ∞) and then do Box-Cox transformation, selecting λ that maximises the likelihood function (this is done automatically in `alm` function with the parameters `distribution="dbcnorm"`). All these transformations were needed in order to ensure that the model fits the data adequately and captures well the relations, which are non-linear in their nature in the original scale. The parameters of this model are summarised in Table 1, together with their standard errors and 99.9% confidence intervals. We need to note that the aim of this exercise is not to create the most suitable model for purposes of prediction or interpretation, but rather to capture complex relations between the variables and see, whether the bias is affected positively or negatively by them and what is the impact of each separate variable on it.

The analysis of Table 1 shows that with the increase of any of these variables, the percentage difference between the conditional expectation and the point forecasts increases: all parameters are positive and their confidence intervals do not include zero values, so we can be sure about the values of the parameters on 99.9% confidence level. We have selected such a high confidence level to make sure that there is a small margin for the error. This meets our expectations and supports the earlier discussions in Section 2. Given that all the variables in the model lie between 0 and 1, we can also conclude that the increase of the parameter ϕ leads to the highest increase in bias among all variables, followed by β and then forecast horizon, while the impact of α on bias is the smallest. The specific values of parameters are difficult to interpret, and they do not provide any useful information.

Variables	Estimate	Std. Error	Lower 0.05%	Upper 99.95%	
(Intercept)	-9.222	0.006	-9.242	-9.202	*
σ	2.755	0.003	2.745	2.765	*
horizon	3.656	0.003	3.646	3.667	*
α	1.578	0.007	1.555	1.600	*
β	3.739	0.007	3.716	3.762	*
ϕ	6.167	0.006	6.148	6.187	*
$R^2 = 0.986$	Number of degrees of freedom: 534,884				

TABLE 1 *Regression output for the bias of point forecasts data, experiment 1. Box-Cox parameter* $\lambda = 0.146$.

4. Distributional assumptions in ETS framework

After discussing some general properties of multiplicative error models, we can now move to the distributional assumptions of the model.

4.1. *Potential distributions for ETS(M,*,*) model*

We argue that in order for the distribution to be efficiently used in ETS(M,**), it needs to have the following properties:

1. Ease of use. Ideally, the distribution should have a simple location and scale parameters, which should have direct connection to the expectation and variance. This would simplify the inference with the distribution;
2. Flexible skewness and kurtosis. In order to cover different real life situation, the distribution should have skewness and kurtosis regulated using simple parameters (preferably location and scale in order not to introduce additional parameters);
3. Relation to the Normal distribution. It should have some relation to Normal distribution, preferably converging to it under some circumstances, i.e. increase of level of series. In this case, the multiplicative error models would behave similar to the additive error ones on high volume data;
4. Convexity of the likelihood function. This is a desired property, which guarantees consistency of the estimator [15]. If the function is not convex then this introduces additional complications in the estimation and makes the distribution less desirable.

The candidate distributions for the error term $(1+\varepsilon_t)$ in ETS(M,*,*) models, as mentioned earlier [and in 2], are Normal ($\mathscr{N}$), Log-Normal ($\log\mathscr{N}$), Gamma (Γ) and truncated Normal. We also point out that Inverse Gaussian ($\mathscr{IG}$) can be considered as a viable candidate distribution. We drop truncated Normal from the consideration, because its statistics are too complicated to summarise, which makes it very difficult to use in ETS framework (thus violating the property (1)). In order to understand the potential limitations of the other distributions, we summarise their main properties in Table 2 based on [10].

In order for these distributions to work in ETS model, we need to make sure that the expectation of the error term $(1+\varepsilon_t)$ is equal to 1 (as mentioned in Section 2), which is easily achievable in $\mathscr{IG}$ by setting $\mu = 1$. In case of Log$\mathscr{N}$ distribution, the condition $\mathrm{E}(x) = 1$ implies that $\mu = -\frac{\sigma^2}{2}$, while in case of Γ it is: $\xi = \frac{1}{\sigma^2}$. All these restrictions will affect the log-likelihood functions for the ETS(M,*,*)

	$\mathcal{N}(\mu, \sigma^2)$	$\mathcal{IG}(\mu, \sigma^2)$	$\log\mathcal{N}(\mu, \sigma^2)$	$\Gamma(\xi, \sigma^2)$
Support	$(-\infty, \infty)$	$(0, \infty)$	$(0, \infty)$	$(0, \infty)$
Mean, $\mathrm{E}(1+\varepsilon_t)$	μ	μ	$e^{\mu+\frac{\sigma^2}{2}}$	$\xi\sigma^2$
Parameters restrictions in ETS	$\mu = 1$	$\mu = 1$	$\mu = -\frac{\sigma^2}{2}$	$\xi = \frac{1}{\sigma^2}$
Variance, $\mathrm{V}(1+\varepsilon_t) = \varsigma^2$	σ^2	σ^2	$e^{\sigma^2} - 1$	σ^2
Skewness	0	3ς	$\varsigma^3 + 3\varsigma$	2ς
Kurtosis	3	$15\varsigma^2$	$\left((\varsigma^2+3)(\varsigma^2+1)+3\right)(\varsigma^2+1)^2$	$6\varsigma^2$

TABLE 2 *Comparison of main statistics of several distributions. μ is the location, σ^2 is the scale, ξ is the shape parameter and ς is the standard deviation of the error term.*

models with different assumptions, which are presented in Table 3 and are obtained directly from the probability density functions of respective distributions with the mentioned restrictions [the case of the Normal distribution was derived in 9, p.68].

Distribution of $1+\varepsilon_t$	log-likelihood
$\mathcal{N}(1, \sigma^2)$	$\ell\left(\theta, \sigma^2 \vert \mathbf{y}\right) = -\frac{T}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{t=1}^T \varepsilon_t^2 - \sum_{t=1}^T \log \mu_{y,t}$
$\mathcal{IG}(1, \sigma^2)$	$\ell\left(\theta, \sigma^2 \vert \mathbf{y}\right) = -\frac{T}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{t=1}^T \frac{\varepsilon_t^2}{1+\varepsilon_t} - \frac{3}{2}\sum_{t=1}^T \log y_t$
$\log\mathcal{N}\left(-\frac{\sigma^2}{2}, \sigma^2\right)$	$\ell\left(\theta, \sigma^2 \vert \mathbf{y}\right) = -\frac{T}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{t=1}^T \left(\log(1+\varepsilon_t) + \frac{\sigma^2}{2}\right)^2 - \sum_{t=1}^T \log y_t$
$\Gamma\left(\sigma^{-2}, \sigma^2\right)$	$\ell\left(\theta, \sigma^2 \vert \mathbf{y}\right) = -T\log\Gamma\left(\sigma^{-2}\right) - \frac{T}{\sigma^2}\log\sigma^2 + \frac{1}{\sigma^2}\sum_{t=1}^T \log\left(\frac{1+\varepsilon_t}{\exp(1+\varepsilon_t)}\right) - \sum_{t=1}^T \log y_t$

TABLE 3 *The log-likelihoods for ETS(M,*,*) models, where θ is the vector of all estimated parameters, except for the scale σ^2; $\mathbf{y}$ is the vector of actual values, $\mu_{y,t}$ is the one step ahead conditional expectation of the model and T is the sample size.*

The MLE of scale parameter are straightforward for $\mathcal{N}$ and $\mathcal{IG}$ and are obtained directly from the log-likelihoods in Table 3 by taking derivatives with respect to the σ^2:

- $\mathcal{N}$: $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^T \varepsilon_t^2$;
- $\mathcal{IG}$: $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^T \frac{\varepsilon_t^2}{1+\varepsilon_t}$;

The case with $\log\mathcal{N}$ is more complicated because of the restriction on the location parameter, but still tractable:

- $\log\mathcal{N}$: $\hat{\sigma}^2 = 2\left(1 \pm \sqrt{1 - \frac{1}{T}\sum_{t=1}^T \log^2(1+\varepsilon_t)}\right)$;

The main issue with this estimate is that it can only be used in the case of $\frac{1}{T}\sum_{t=1}^T \log^2(1+\varepsilon_t) \leq 1$ - assuming that the scale can only be a positive real number. In the other cases it would need to be estimated directly using numerical optimisation. Also note that the formula will produce two roots, only one of which (the smaller one) will corresponds to the local maximum of the likelihood. However, given that the derivative of the log-likelihood has two roots, the likelihood itself is not convex, which might imply that identification and compactness assumptions for the likelihood are violated, making the estimates of parameters potentially inconsistent [15]. Having said that, the analysis of the likelihood function shows that the first root (for the case of subtraction in the formula above) should correspond to the local maximum, while the second one (addition in the formula) should correspond to the local minimum of the likelihood function. This means that the subtraction should be used in order to get the local maximum of the likelihood function. All of this means that, while $\log\mathcal{N}$ can be considered as a possible option for the ETS(M,*,*) models, it has issues in the estimation of the scale, and should be used with care.

Finally, in case of Γ distribution, the maximum likelihood estimate of scale does not have a closed form, because of the imposed restrictions in Table 2. A numerical optimisation would need to be used in order to estimate it, which makes it less desirable than other distributions due to the property (1) discussed in the beginning of this section. However, the scale can be estimated using method of moments, based on how the mean and variance can be calculated in Γ distribution (Table 2) – given the parameters restriction the estimate of scale is:

- Γ: $\hat{\sigma}^2 = \mathrm{V}(1+\varepsilon_t) = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2$.

This does not guarantee that the likelihood is maximised, but at least allows estimating the model. Furthermore, this estimate can be used as a starting point in the estimation of the scale of Γ distribution via maximisation of likelihood.

Another important property of all the distributions from Table 2 is scalability, meaning that:

$$\begin{aligned}
&\text{if } 1+\varepsilon_t \sim \mathcal{N}(1,\sigma^2), \text{ then } k(1+\varepsilon_t) \sim \mathcal{N}\left(k, k^2\sigma^2\right) \\
&\text{if } 1+\varepsilon_t \sim \mathcal{IG}(1,\sigma^2), \text{ then } k(1+\varepsilon_t) \sim \mathcal{IG}\left(k, \frac{\sigma^2}{k}\right) \\
&\text{if } 1+\varepsilon_t \sim \log\mathcal{N}\left(-\frac{\sigma^2}{2},\sigma^2\right), \text{ then } k(1+\varepsilon_t) \sim \log\mathcal{N}\left(\log(k)-\frac{\sigma^2}{2},\sigma^2\right) \\
&\text{if } 1+\varepsilon_t \sim \Gamma(\sigma^{-2},\sigma^2), \text{ then } k(1+\varepsilon_t) \sim \Gamma(\sigma^{-2},k\sigma^2).
\end{aligned} \tag{4.1}$$

This property is useful, when we need to consider the distribution of the variable y_t in the ETS(M,*,*) model.

When it comes to the convolution, it is not well defined for $\log\mathcal{N}$, but is well defined for $\mathcal{N}$, i.i.d. $\mathcal{IG}$ and i.i.d. Γ random variables. The situation with the product distribution is the opposite: it is well defined for $\log\mathcal{N}$, but does not have a closed form for either $\mathcal{N}$, $\mathcal{IG}$, or Γ. These two properties are necessary for the derivation of the conditional h steps ahead distribution of a variable from the ETS(M,*,*) model. Unfortunately, none of the discussed distributions supports both of them. So, the conditional h-steps-ahead distribution does not have closed form for any of them.

If only one distribution between the four needs to be selected, in addition to the discussion above, some findings from literature can be used. [6] argued after investigating a set of data that they can be equally well described by Log-Normal, Gamma and Inverse Gaussian distributions, so the choice of the distribution might be based on convenience of working with it. We argue that the Inverse Gaussian is more convenient than its competitors. However, it is still possible to have all the discussed distributions in the ETS(M,*,*) model. Furthermore, it is also possible to select the most appropriate distribution automatically using information criteria (this will be discussed in Subsection 4.3).

4.2. *ETS(M,N,N) model with different distributions*

The model with multiplicative error, no trend and no seasonality, ETS(M,N,N), can be considered as a model of a special interest, because, as discussed in Section 2, it has closed forms for the conditional moment, so we briefly discuss it in this section. It is formulated in the following way:

$$\begin{aligned} y_t &= l_{t-1}(1+\varepsilon_t) \\ l_t &= l_{t-1}(1+\alpha\varepsilon_t) \end{aligned}. \tag{4.2}$$

The classical assumption about this model [16] is that $\varepsilon_t \sim \mathcal{N}(0,\sigma^2)$. [2] proposed rewriting this model in the following way:

$$\begin{aligned} y_t &= l_{t-1}\epsilon_t \\ l_t &= l_{t-1}(1-\alpha+\alpha\epsilon_t) \end{aligned}, \tag{4.3}$$

where $\epsilon_t = 1+\varepsilon_t \sim \mathcal{N}(1,\sigma^2)$. The form (4.3) demonstrates why the distributional assumptions should be imposed on $1+\varepsilon_t$ rather than ε_t, although the models (4.2) and (4.3) are identical.

The main limitation of the model (4.2) is that, when the level of the data is low, the assumption of normality might not be reasonable anymore, because the model may produce negative values even if the data is strictly positive. As discussed in Subsection 4.1, a more natural distribution for the model is either the Inverse Gaussian, or the Log-Normal, or Gamma, implying one of the following:

$$\begin{aligned} &1+\varepsilon_t \sim \mathcal{IG}(1,\sigma^2) \\ &1+\varepsilon_t \sim \log\mathcal{N}\left(-\frac{\sigma^2}{2},\sigma^2\right). \\ &1+\varepsilon_t \sim \Gamma(\sigma^{-2},\sigma^2) \end{aligned} \tag{4.4}$$

Note that due to the restrictions on the expectation in (4.4), the conditional h steps ahead mean and variance of the ETS(M,N,N) model can be calculated using the same formulae as in the conventional ETS(M,N,N) relying on the normality assumption [9]:

$$\begin{aligned} &\mathrm{E}(l_{t+h}|t) = \mathrm{E}\left(l_t \prod_{j=1}^{h}(1+\alpha\varepsilon_{t+j}) \middle| l_t \right) = l_t, \\ &\mathrm{V}(l_{t+h}|t) = l_t^2\left(\left(1+\alpha^2\sigma^2\right)^h - 1\right). \end{aligned} \tag{4.5}$$

In addition, the one-step-ahead actual value from this model can be obtained using the scalability property (4.1):

$$\begin{aligned} &\mathcal{IG}: y_{t+1} \sim \mathcal{IG}\left(l_t, \frac{\sigma^2}{l_t}\right) \\ &\log\mathcal{N}: y_{t+1} \sim \log\mathcal{N}\left(\log(l_t) - \frac{\sigma^2}{2}, \sigma^2\right). \\ &\Gamma: y_{t+1} \sim \Gamma(\sigma^{-2}, l_t\sigma^2) \end{aligned} \tag{4.6}$$

Based on the discussion from the Subsection 4.1, related to convolution and product distribution, we can conclude that ETS(M,N,N) does not have a closed form for the h-steps ahead conditional distribution and thus cannot be easily characterised using the parameters of distributions. This means that in order to obtain the h-steps ahead prediction distribution from ETS(M,N,N) with any of the distributions above, the simulations need to be used. This aligns with what we have discussed for the general case of ETS(M,*,*) in Section 2.

Furthermore, it can be shown that the properties of the distributions and the multiplicative model also restrict the smoothing parameter with the interval [0, 1]. Assuming that the smoothing parameter is always positive, the inequality $(1+\varepsilon_t) > 0$ implies that:

$$\begin{aligned} \varepsilon_t &> -1 \\ \alpha\varepsilon_t &> -\alpha \\ 1+\alpha\varepsilon_t &> 1-\alpha \end{aligned} \tag{4.7}$$

The ETS(M,N,N) model makes sense only when $1+\alpha\varepsilon_t > 0$. So, if $\alpha > 1$, then $1+\alpha\varepsilon_t$ may become negative, which breaks the model, because the level may become negative. The model however still makes sense for boundary values of α: when $\alpha = 0$, the level is not updated, while in the case of $\alpha = 1$, the level has the dynamics of a random walk process. While there may be some cases when even with $\alpha > 1$ the value of $(1+\alpha\varepsilon_t)$ will be greater than zero, this cannot be guaranteed universally. So, the condition for the smoothing parameter in ETS(M,N,N) is $\alpha \in [0,1]$.

As we see, the case of ETS(M,N,N) is interesting, because this is the only pure multiplicative ETS model that has closed forms for the expectation and variance for any horizon h. The more complicated pure multiplicative ETS models will not have the analytical solutions for the moments of distribution (as discussed in Section 2), but they will share the main properties with the ETS(M,N,N). For example, all the smoothing parameters should have restrictions similar to (4.7) for the reasons discussed above.

4.3. *Distribution selection*

Based on the discussions in previous sections, we can conclude that there are several possible options for the ETS(M,*,*) models in terms of distributions. Given, that the models are formulated similarly, we can calculate the values of the likelihood functions in their maxima and use them in the model selection via information criteria, for example, using Akaike's Information Criterion [1]:

$$\mathrm{AIC} = 2k - 2\ell\left(\theta, \sigma^2 | \mathbf{y}\right), \tag{4.8}$$

where k is the number of all estimated parameters and $\ell\left(\theta, \sigma^2 | \mathbf{y}\right)$ is the value of the log-likelihood function from Table 3. The procedure in this case is straightforward: select the most appropriate ETS model for each distribution, then compare them with each other and select the one with the lowest information criterion. Other criteria can be used in the place of AIC to select the appropriate models. For example, AICc by [21] is typically recommended instead of AIC on small samples.

5. Simulation experiment 2: distributional assumptions

In order to assess performance of models in terms of their predictive distributions, we conduct another simulation experiment, generating 10,000 time series from ETS(M,Md,M) model with the distributions discussed in Subsection 4.1:

1. Normal,
2. Inverse Gaussian,
3. Gamma,

4. Log-Normal.

We have also included the DGP ETS(A,Ad,A) with the Normal distribution, the data of which was then exponentiated to get the model similar to the one in [2]. We called this DGP "Akram" (by the surname of the first author of the original paper).

Each series contained 500 observations, in every case the standard deviation of the error term was set to 0.1, while the smoothing parameters were set to $\alpha = 0.1$, $\beta = 0.05$, $\gamma = 0.2$, while $\phi = 0.95$, $l_0 = 100$, $b_0 = 1$ and the initial seasonal indices were randomly generated and normalised. The data generation was done using the `sim.es()` function from the `smooth` package for R [23]. We then applied ETS(M,Md,M) model with each of the distributions (and one with AICc-based selected distribution), using the `adam()` and `auto.adam()` functions from the `smooth` package, and produced quantile forecasts for the next 48 observations. We also applied the ETS(A,Ad,A) to the logarithms of the generated data to get the values from the model proposed in [2] (method called "Akram" in the results below).

After that we assessed the generated quantiles via scaled pinball score, calculated using the formula:

$$\mathrm{sPS} = \frac{1}{\bar{y}} \left((1-\tau) \sum_{y_{t+j} < q_{t+j}, j=1,\dots,h} |y_{t+j} - q_{t+j}| + \tau \sum_{y_{t+j} \geq q_{t+j}, j=1,\dots,h} |y_{t+j} - q_{t+j}| \right),$$

where $\bar{y}$ is the in-sample mean of actual values and τ is the level of probability for the specific quantile. The scaling in pinball score is needed in order to make the values comparable across different generated time series. In addition, we calculated the quantile coverage to measure how good the quantiles are estimated by the models. This is based on the calculation of the percentage of cases below each of quantiles. Quantiles were produced for $\tau = \{0.01, 0.02, \dots, 0.99\}$ levels.

Applied Model	DGP $\mathcal{N}$	$\mathcal{IG}$	Γ	$\log\mathcal{N}$	Akram
$\mathcal{N}$	5.255	5.204	5.193	9.417	5.192
$\mathcal{IG}$	5.300	5.221	5.224	9.464	5.229
Γ	5.287	5.192	5.231	9.388	5.214
$\log\mathcal{N}$	5.295	5.215	5.224	9.423	5.220
Akram	**5.110**	**5.052**	**5.074**	**9.180**	**5.089**
Auto	5.258	5.175	5.208	9.332	5.178

TABLE 4 *Mean scaled pinball scores for the ETS(M,Md,M) model with various distributions.*

Table 4 summarises the overall performance of the models under consideration. The lowest values are marked in boldface. The pinball scores are very close to each other across the different applied models, although Akram's approach seems to be doing consistently better on average than the others. This is probably due to computational complexities arising in estimation of the pure multiplicative models. Interestingly enough, specific distributions applied to the data of the same type (e.g. model with $\mathcal{IG}$ distribution on $\mathcal{IG}$ DGP) do not necessarily perform the best, although the difference between different models is typically small and not statistically significant (on 5% significance level), with the exception of the Akram's approach.

Applied Model	DGP $\mathcal{N}$	$\mathcal{IG}$	Γ	$\log\mathcal{N}$	Akram
		Quantile bias			
$\mathcal{N}$	-0.002	0.011	0.003	0.007	-0.010
$\mathcal{IG}$	-0.004	0.009	0.001	**0.001**	-0.013
Γ	-0.004	**0.008**	0.001	0.003	-0.012
$\log\mathcal{N}$	-0.005	**0.008**	**0.000**	**0.001**	-0.013
Akram	**0.000**	0.010	0.004	0.006	**-0.007**
Auto	-0.003	**0.008**	**0.000**	0.002	-0.012
		Quantile precision			
$\mathcal{N}$	0.007	0.011	0.005	0.009	0.010
$\mathcal{IG}$	0.009	0.009	0.006	0.009	0.015
Γ	0.011	0.009	0.005	0.008	0.014
$\log\mathcal{N}$	0.008	**0.008**	**0.004**	**0.004**	0.014
Akram	**0.004**	0.010	0.005	0.008	**0.007**
Auto	0.008	0.009	**0.004**	0.007	0.013

TABLE 5 *Quantile bias and precision values for the ETS(M,Md,M) model with various distributions.*

We also calculated the average quantile bias and quantile precision to assess the predictive distributions of the models from a different angle. Table 5 summarises performance of models in terms of the quantile bias (mean of all quantile coverages minus the nominal values), and the quantile precision (mean absolute value of quantile coverages minus the nominal values). The values of quantile bias can be interpreted as average percentage points below (in case of negative) or above (in case of positive values) the nominal values of quantiles. So, for example, the value of 0.008 for the Gamma distribution on the $\mathcal{IG}$ data in quantile bias tells us that it tends to produce quantiles that cover 0.8% more observations than expected on average (across all the confidence levels). The quantile precision shows the average deviation from the nominal value for each of the approaches.

Judging by the results in Table 5, we can conclude that all the approaches perform very similar and that there is no obvious winner: all values are small and close to zero, showing that the approaches are well calibrated, irrespective of the data they are applied to.

Finally, we produce the pinball and quantile coverage plots to see how the approaches perform specifically for each confidence level. Figures 3, A.1, A.2, A.3 and A.4 summarise performance of the approaches for each of the DGPs in the form of reliability diagrams. Given that the performance of models for different DGPs is very similar, we focus on the case of Normal distribution, Figure 3, leaving the rest in Appendix.

Judging by the pinball values in Figure 3, we can conclude that all approaches perform very similar for the lower and higher quantiles (below 0.2 and above 0.8), while Akram's model does slightly better in the middle of levels (i.e. between 0.2 and 0.8). Arguably, the tails are more important in practice than the middle because they correspond to prediction intervals and the decisions are typically made based on the higher confidence levels (e.g. 90%, 95%, or 99%). Furthermore, we can see that the performance of approaches is indeed very similar for different levels in terms of quantile coverage: the fluctuations

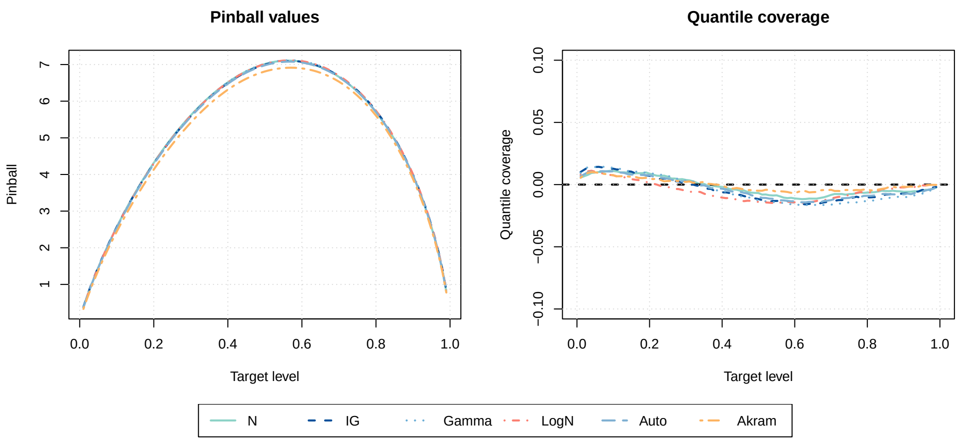


FIG. 3. Pinball and quantile coverage values for DGP with the Normal distribution.

around the horizontal line are random, and a new run of the experiment produced different shapes, with lines having different curvatures. This means that all the models under consideration are well calibrated, and that the decision, which one to use comes to personal preferences of the analyst.

As a conclusion from this simulation experiment, we see that ETS can be used efficiently with all discussed distributions, meaning that there is no reason to always use the Normal one, especially given its restrictions, discussed in Section 4.1. Furthermore, the AIC-based selection between the distributions looks as a robust option for the usage of ETS with the discussed distributions. Finally, Akram's approach does better in terms of mean pinball values, but this comes mainly because of the good performance for the target levels in the middle, which are not practically useful. At the same time, it does very similar to other approaches in the tails of distributions, meaning that it does not have any obvious practical benefit in comparison with the ETS with the positive distributions. Given that this approach is detached from the ETS framework, and makes it impossible to introduce mixed ETS models, we argue that the approaches proposed in this paper are competitive and can be used efficiently in practice.

6. Real life demonstration

Finally, to demonstrate how the pure multiplicative ETS works in the case of non-normal distributions, we apply the model to the data of electric vehicles (EV) charging from a company that tracks the amount of electricity used by such vehicles. This is a daily data on a high aggregation level (EV users in a geographical area) containing 511 observations. The data is strictly positive, but starts with a very low volume, demonstrating the situation when the demand builds up over time. It exhibits a multiplicative seasonal pattern (repeating every 7 days) and probably should be modelled with an ETS with multiplicative error term.

We keep the last 14 observations (two weeks) for the model evaluation and use the rest of the sample for the model fitting and selection. We conduct the selection based on AICc in two dimensions:

1. Selection of components of ETS models from the pool of pure multiplicative ones;
2. Selection of the appropriate distribution.

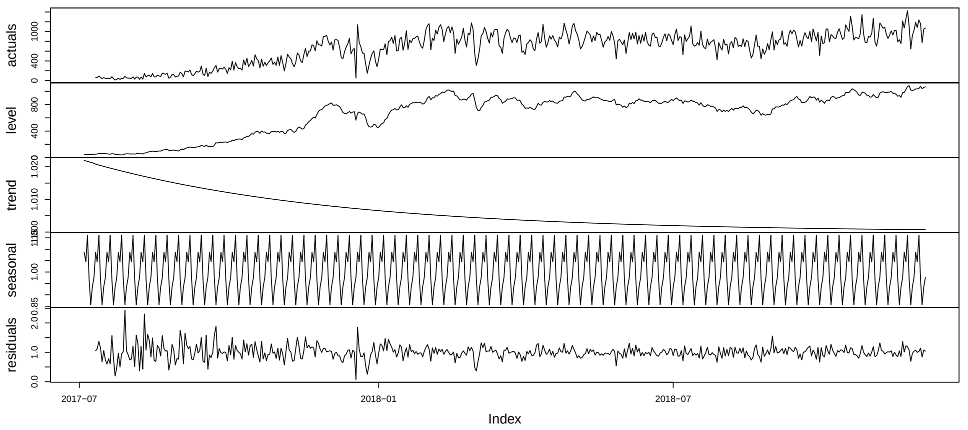


FIG. 4. Time series decomposition according to ETS(M,Md,M) with the Gamma distribution.

This was done using `auto.adam()` function from the `smooth` package for R.

As a result, the ETS(M,Md,M) with Gamma distribution had the lowest AICc of 6,398.570. The model had the following parameters: $\hat{\alpha} = 0.1683$, $\hat{\beta} = 0.0000$, $\hat{\gamma} = 0.0013$, $\hat{\phi} = 0.9925$, $\hat{\sigma} = 0.276$, implying that the seasonality does not change abruptly and that the trend component is not updated over time. This is reflected in dynamics of components shown in Figure 4, where the trend component declines over time due to the dampening and seasonality changes slowly.

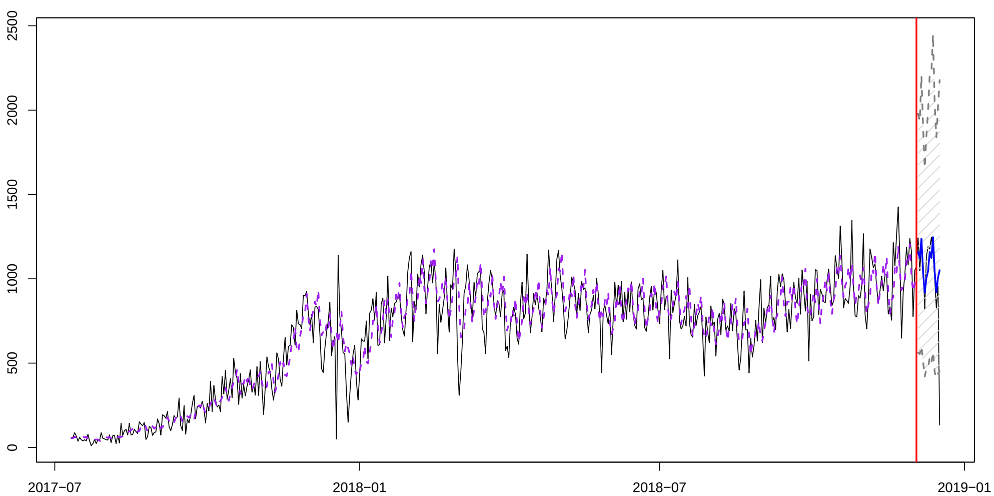


FIG. 5. An example of application of ETS model to the electric vehicle charging data. ETS(M,Md,M) with the Gamma distribution.

Figure 5 shows the model fit, the conditional expectation and the 95% prediction interval from the model. This example demonstrates how pure multiplicative ETS model with non-normal distribution can be efficiently used for time series analysis and forecasting in practice.

To compare the pure multiplicative ETS model with the pure additive one, we have also fitted ETS(A,Ad,A) model to the data, assuming Normal distribution of the residuals. The resulting model had AICc of 6,418.296 versus 6,398.570 of the pure multiplicative ETS with the Gamma distribution. This indicates that the pure multiplicative model is more appropriate for this type of data than the additive one.

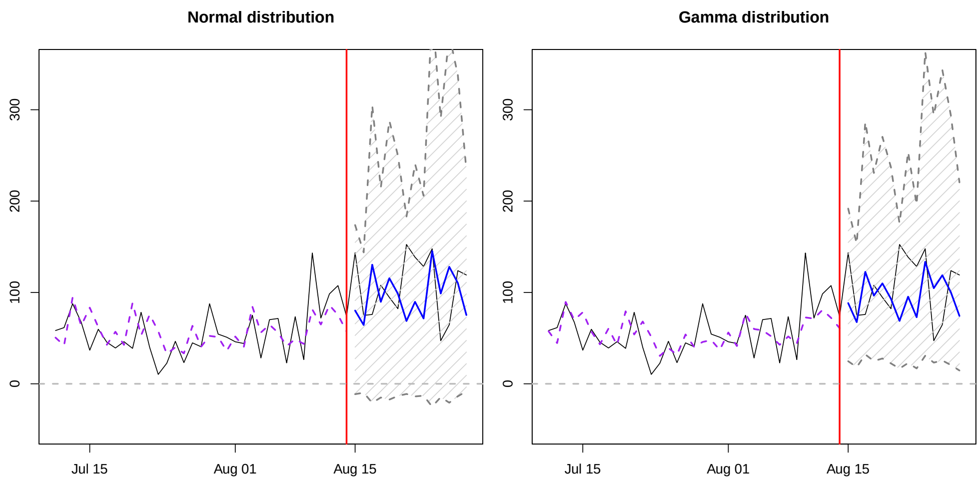


FIG. 6. An example of application of pure multiplicative ETS to a shorter sample of data. The left plot shows forecasts generated by the model with the Normal distribution (with $\hat{\sigma} = 0.471$), while the right one shows the one with Gamma (with $\hat{\sigma} = 0.493$).

Finally, we fit the ETS(M,Md,M) model to the first 49 observations only, keeping the last 14 of them for the holdout evaluation to demonstrate what happens on low volume data when the pure multiplicative model with normality is used against the model with a positive distribution (Gamma in this example). Figure 6 shows the point forecasts and the 95% prediction intervals for the two cases. It becomes apparent that although the data is strictly positive in this example, the model with the Normal distribution generates an unrealistic prediction interval, covering the negative values. In contrast, the ETS with the Gamma distribution generates a reasonable prediction interval. This demonstrates the advantage of the proposed approach in comparison with the conventional ETS.

7. Managerial implications

Pure multiplicative ETS models discussed in this paper not only have theoretical interest, but also have important practical implications. Given that now they can work with the positive distributions discussed in this paper (Log-Normal, Gamma and Inverse Gaussian), they become useful in such contexts as supply chain, retail, healthcare and energy. In these cases, the values typically exhibit asymmetry, making the conventional normality assumption inappropriate. Given that ETS(M,*,*) is a proper statistical model, it is possible to generate predictive distributions for several steps ahead from it. `adam()` function from the `smooth` package in R allows doing that, generating different types of forecasts, including the cumulative over the lead time ones.

In supply chain and retail, the demand is typically non-negative [12], and the decisions are typically made on weekly or daily basis, making the multiplicative ETS especially useful. In this context, the stocking decisions are usually made based on the cumulative demand over the lead time, forecasts for which can be done using the ETS(M,*,*) models. The generated quantile forecasts can then be used directly to define the safety stock levels.

The pure mutliplicative ETS models can even be used on intermittent data as shown in [25]. The authors have amended the conventional ETS model to deal with zeroes and used multiplicative ETS to capture the non-zero demands. The approach works well both in terms of forecasting accuracy and in terms of reducing inventory costs.

Another example of application would be electricity demand forecasting [29], where the demand is always non-negative, making the multiplicative models especially useful. In that context, ETS needs to be extended to multiple seasonalities and possibly be modified to deal with potential zeroes in the data [as done by 25]. But fundamentally, a pure multiplicative ETS would allow producing predictive distributions for several steps ahead, which then can be used to determine how much electricity needs to be generated and/or sent to specific substations.

Finally, the pure multiplicative ETS models can be used in healthcare, for example, for forecasting Emergency Department attendance. That demand is typically count and non-negative. While it might seem that continuous models, such as ETS, would be not as appropriate as count models (e.g. Poisson regression) in this situation, it has been shown that the continuous distributions, such as Truncated Normal and Gamma perform well in that context [19]. This makes the multiplicative ETS models with distributions discussed in this paper especially attractive.

All in all, the pure multiplicative ETS models with positive distributions can be useful in a variety of contexts, making them practically appealing.

8. Conclusion

In this paper, we had a new look at the pure multiplicative ETS models, discussed the statistical properties of these models and showed that the point forecasts from them typically do not coincide with the conditional expectations. While this is not a new result in the literature, it has been neglected by many researchers and is worth mentioning. We also showed the relation between the point forecasts, geometric and arithmetic conditional means in such models. In general, due to the multiplicative nature of the model, the three statistics do not coincide. Aligned to this, we conducted a simulation experiment, showing what impacts the difference between the point forecast and the conditional expectation. We found that the dampening parameter ϕ has the highest impact on it, which is then followed by the smoothing parameter of the trend, β and the variance of the error term of the model. We also demonstrated that the difference increases with the increase of the forecasting horizon. This has serious practical implications: if the analyst needs conditional expectations from a pure multiplicative ETS model, they should in general use simulations, because point forecast will not coincide with the expectation.

After that we proposed using several positively-defined distributions in pure multiplicative ETS models, showing how they should be used, and what restrictions should be imposed on their parameters. We derived the concentrated log-likelihoods for each distribution and showed how their scales can be calculated. We then conducted a simulation experiment, showing how ETS models with different distributions can be applied to the data generated from similar ETS models but with other distributions.

We found that all considered models perform very similar, with automatic selection based on AICc performing well in the majority of cases. This means that in practice one can use such procedure to select the most appropriate distribution for their ETS model and that in general there is no need to use the Normal distribution.

Finally, we showed how the appropriate pure multiplicative ETS model with the most fitting distribution can be selected on an example of electric vehicles charging data. The proposed approach is already implemented in the `auto.adam()` function from the `smooth` package [23] for R.

We should acknowledge at this stage that the simplest benchmark we used in this paper, was the Akram et al. model [2]. We did not include Simple Exponential Smoothing, Naïve or any other simple forecasting method like that intentionally. This is because we focused our investigation on predictive distributions, and models with the additive error term would not be appropriate for the low volume purely positive data.

As for the future works, given the discussion of multiplicative trend models in the paper, we feel that the development of a new model which would have properties of a pure multiplicative model but would not exhibit explosive trends, is an important direction of future research. It is not clear how such model can be formulated, but having it instead of ETS(M,M,*) models would make the pure multiplicative models more attractive.

Data Availability Statements

The data from Sections 4 and 5 was generated using the `sim.es()` function from the `smooth` package in R. The source code will be shared on reasonable request to the corresponding author. The data from Section 6 cannot be shared publicly due to non-disclosure agreement.

References

1. H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19 (6):716–723, 1974. doi: 10.1109/TAC.1974.1100705.
2. M. Akram, R. J. Hyndman, and J. K. Ord. Exponential Smoothing and Non-negative Data. *Australian & New Zealand Journal of Statistics*, 51(4):415–432, 2009.
3. G. Athanasopoulos, R. J. Hyndman, H. Song, and D. C. Wu. The tourism forecasting competition. *International Journal of Forecasting*, 27(3):822–844, 2011. doi: 10.1016/j.ijforecast.2010.04.009.
4. M. Z. Babai, A. Tsadiras, and C. Papadopoulos. On the empirical performance of some new neural network methods for forecasting intermittent demand. *IMA Journal of Management Mathematics*, 31(3):281–305, jun 2020. ISSN 1471-678X. doi: 10.1093/imaman/dpaa003.
5. R. G. Brown. Exponential Smoothing for predicting demand. 1956.
6. J. L. Folks and R. S. Chhikara. The Inverse Gaussian Distribution and Its Statistical Application–A Review. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40(3):263–289, 1978.
7. C. C. Holt. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1):5–10, 2004.
8. R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3):439–454, 2002.
9. R. J. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Snyder. *Forecasting with Exponential Smoothing*. Springer Berlin Heidelberg, 2008.
10. S. Kotz, N. Johnson, and C. Read. *Encyclopedia of statistical sciences*. Number v. 2 in Encyclopedia of Statistical Sciences. Wiley, 1988.
11. N. Kourentzes, F. Petropoulos, and J. R. Trapero. Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting*, 30(2):291–302, 2014.

12. C. Li and A. Lim. A greedy aggregation–decomposition method for intermittent demand forecasting in fashion retailing. *European Journal of Operational Research*, 269(3):860–869, 2018. doi: 10.1016/j.ejor.2018.02.029.
13. S. Makridakis and M. Hibon. The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, 16:451–476, 2000.
14. S. Makridakis, E. Spiliotis, and V. Assimakopoulos. The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1):54–74, 2020.
15. W. K. Newey and D. McFadden. Chapter 36 large sample estimation and hypothesis testing. volume 4 of *Handbook of Econometrics*, pages 2111 – 2245. Elsevier, 1994.
16. J. K. Ord, A. B. Koehler, and R. D. Snyder. Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association*, 92(440):1621–1629, 1997.
17. A. F. Osman and M. L. King. A new approach to forecasting based on exponential smoothing with independent regressors. 2015.
18. P. Ramos, N. Santos, and R. Rebelo. Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and Computer-Integrated Manufacturing*, 34:151–163, 2015. ISSN 07365845. doi: 10.1016/j.rcim.2014.12.015.
19. B. Rostami-Tabar, J. Browell, and I. Svetunkov. Probabilistic forecasting of hourly emergency department arrivals. *Health Systems*, 00(00):1–17, 2023. doi: 10.1080/20476965.2023.2200526.
20. R. D. Snyder. Recursive Estimation of Dynamic Linear Models. *Journal of the Royal Statistical Society, Series B (Methodological)*, 47(2):272–276, 1985.
21. N. Sugiura. Further analysis of the data by akaike' s information criterion and the finite corrections. *Communications in Statistics - Theory and Methods*, 7(1):13–26, 1978.
22. I. Svetunkov. *greybox: Toolbox for Model Building and Forecasting*, 2021. URL https://github.com/config-i1/greybox. R package version 1.0.3.
23. I. Svetunkov. *smooth: Forecasting Using State Space Models*, 2023. URL https://github.com/config-i1/smooth. R package version 3.2.0.
24. I. Svetunkov. *Forecasting and Analytics with the Augmented Dynamic Adaptive Model (ADAM)*. Chapman and Hall/CRC, 2023. ISBN 978-1-003-45265-2. doi: 10.1201/9781003452652. URL https://openforecast.org/adam/.
25. I. Svetunkov and J. E. Boylan. iETS: State space model for intermittent demand forecasting. *International Journal of Production Economics*, 265(October 2022):109013, 2023. ISSN 09255273. doi: 10.1016/j.ijpe.2023.109013.
26. I. Svetunkov, N. Kourentzes, and J. K. Ord. Complex exponential smoothing. *Naval Research Logistics (NRL)*, page 31, 8 2022. doi: 10.1002/nav.22074.
27. I. Svetunkov, N. Kourentzes, and R. Killick. Multi-step estimators and shrinkage effect in time series models. *Computational Statistics*, pages 1–43, jun 2023. ISSN 0943-4062. doi: 10.1007/s00180-023-01377-x.
28. J. W. Taylor. Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting*, 19(4):715–725, 2003.
29. J. W. Taylor. Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, 204(1):139–152, 2010. doi: 10.1016/j.ejor.2009.10.003.
30. R. C. Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL https://www.R-project.org/.
31. D. Thomakos and K. Nikolopoulos. Fathoming the theta method for a unit root process. *IMA Journal of Management Mathematics*, 25(1):105–124, 2014. ISSN 1471-678X. doi: 10.1093/imaman/dps030.
32. M. Weller and S. Crone. Supply Chain Forecasting: Best Practices and Benchmarking Study. Technical report, Lancaster Centre for Forecasting, 2012.

A. Pinball and quantile coverage values

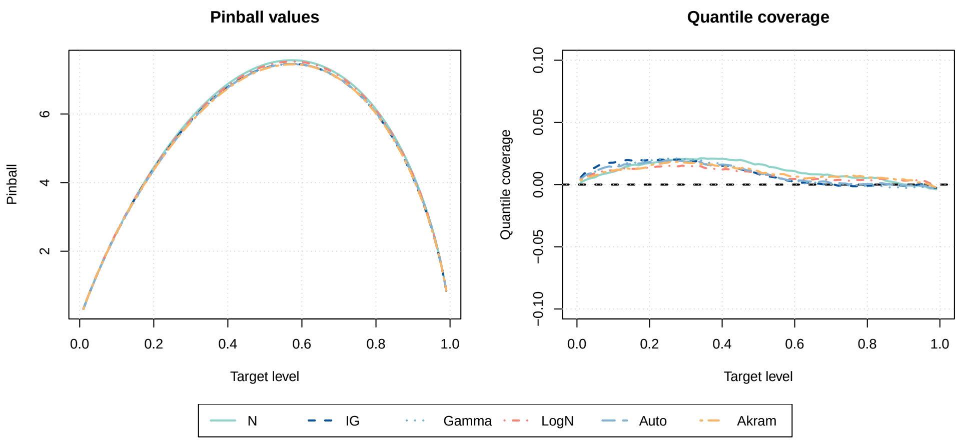


FIG. A.1. Pinball and quantile coverage values for DGP with the Inverse Gaussian distribution.

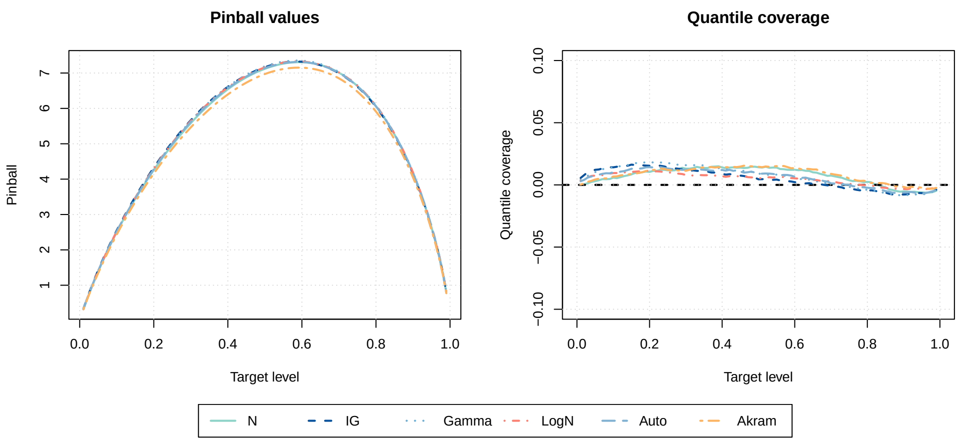


FIG. A.2. Pinball and quantile coverage values for DGP with the Gamma distribution.

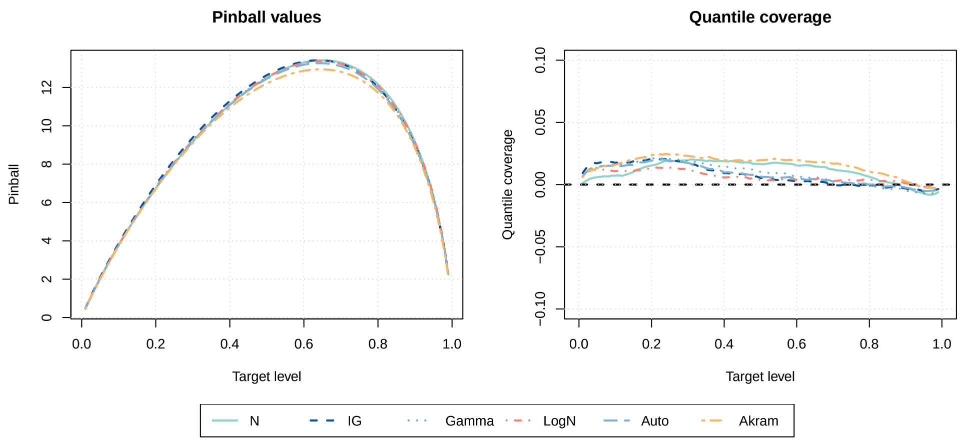


FIG. A.3. Pinball and quantile coverage values for DGP with the Log-Normal distribution.

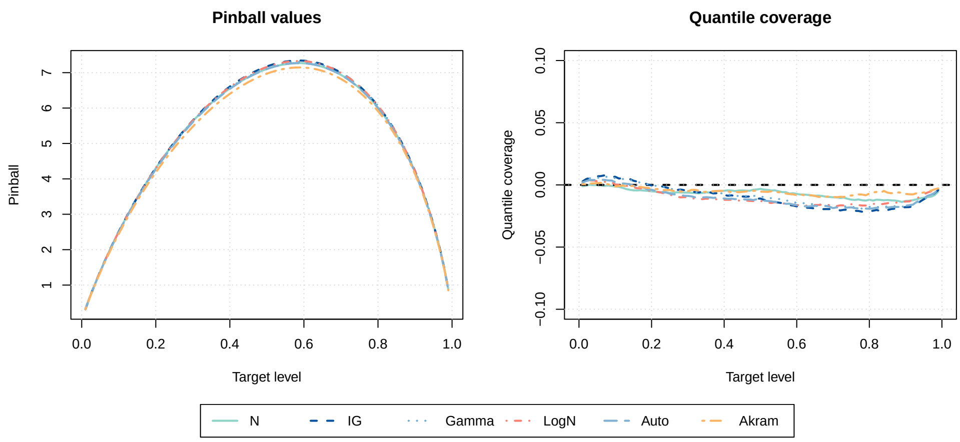


FIG. A.4. Pinball and quantile coverage values for DGP with the model from [2].