How to Make Multiplicative ETS Work for You

Ivan Svetunkov, John E. Boylan

ISF2021

30th June 2021

Marketing Analytics and Forecasting
Lancaster University Management School

Introduction

Exponential Smoothing methods are very popular in forecasting

They performed very well in many competitions:

- M-Competitions in 1982, 2000, 2019, 2020, (Makridakis et al., 1982; Makridakis and Hibon, 2000; Makridakis et al., 2020a,b)
- Competition on telecommunication data in 1998 and 2008 (Fildes et al., 1998),
- Tourism forecasting competition in 2011, (Athanasopoulos et al., 2011)

Introduction

Hyndman et al. (2008) proposed a taxonomy that includes:

- 2 types of error terms (additive and multiplicative);
- 5 types of trend components (none, additive, multiplicative, damped additive and damped multiplicative);
- 3 types of seasonality (none, additive, multiplicative).

In theory it leads to 30 types of ETS models

Model selection procedure based on IC is widely used.

Introduction

All the models in that taxonomy rely on Normal distribution.

This is an issue, when you deal with low level data.

On low level data the variance might become large

This will lead to issues with the model and its forecasts

Introduction

Akram et al. (2009) studied properties of multiplicative models

They mentioned several possibilities for the distributions of the error term:

1. normal, $\mathcal{N}$,
2. truncated normal,
3. log normal, $\log\mathcal{N}$,
4. Gamma, Γ.

They never investigated the idea further

Introduction

We feel that multiplicative ETS models have been overlooked

They are useful but they have not been developed properly

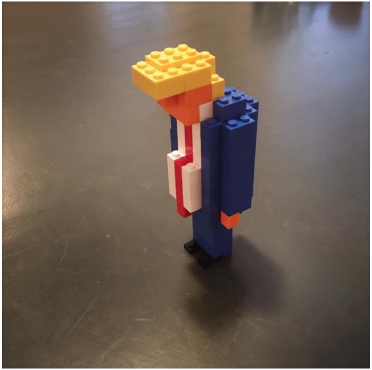
Their properties are not very well known to wide audience

We want to make them work

The truth that was concealed from us

Properties of ETS(M,*,*)

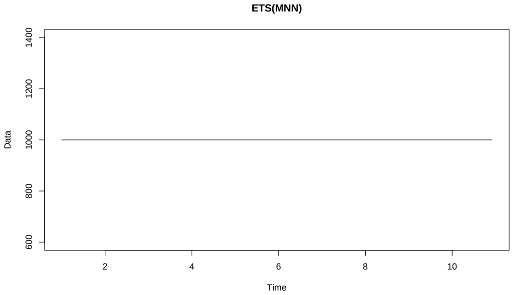
ETS can be considered as a LEGO kit (analogy by Kandrika Fadhlan Pritularga).



Properties of ETS(M,*,*)

We can start with a level of series:

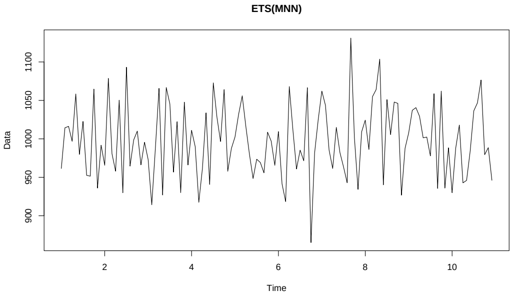
$$y_t = l_{t-1}$$



Properties of ETS(M,*,*)

Add multiplicative error term:

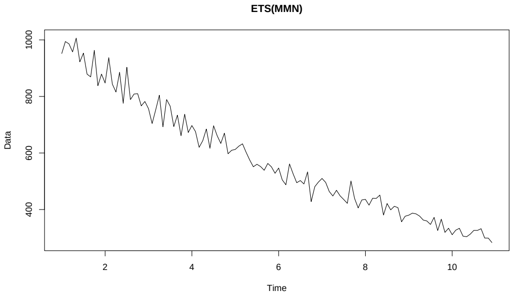
$$y_t = l_{t-1}(1 + \epsilon_t)$$



Properties of ETS(M,*,*)

Add multiplicative trend:

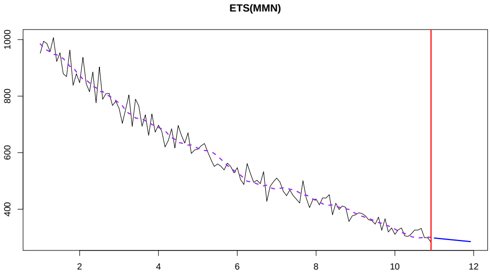
$$y_t = l_{t-1} b_{t-1} (1 + \epsilon_t)$$



Properties of ETS(M,*,*)

Let's take ETS(M,M,N) as an example.

It produces the point forecasts via: $\hat{y}_{t+h|t} = l_t b_t^h$



Properties of ETS(M,*,*)

Did you know that this is not a conditional expectation?

Hyndman et al. (2008) mention this on page 23.

They mention that the bias between point forecast and expectation might increase with the increase of h.

But they claim that it will be typically small.

Properties of ETS(M,*,*)

To investigate this further, we formulate pure multiplicative models as:

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

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, $\mathbf{l}$ is the lag vector, $\mathbf{1}_k$ is the vector of ones of the size k.

Properties of ETS(M,*,*)

This gives exactly the same models as in Hyndman et al. (2008). For example, ETS(M,Md,N):

$$\begin{aligned} \log y_t &= \log l_{t-1} + \phi \log b_{t-1} + \log\left(1 + \epsilon_t\right) \\ \log l_t &= \log l_{t-1} + \phi \log b_{t-1} + \log(1 + \alpha \epsilon_t) \ , \\ \log b_t &= \phi \log b_{t-1} + \log(1 + \beta \epsilon_t) \end{aligned} \tag{2}$$

is equivalent to (after taking exponent):

$$\begin{aligned} y_t &= l_{t-1} b_{t-1}^{\phi} \left(1 + \epsilon_t\right) \\ l_t &= l_{t-1} b_{t-1}^{\phi} (1 + \alpha \epsilon_t). \\ b_t &= b_{t-1}^{\phi} (1 + \beta \epsilon_t) \end{aligned} \tag{3}$$

Properties of ETS(M,*,*)

The conditional expectation in logarithms is:

$$\mathrm{E}(\log y_{t+h}|t) = \log l_t + h \log b_t +$$

$$\sum_{j=1}^{h-1} \mathbf{w}'\mathbf{F}^{j-1}\mathrm{E}\left(\log\left(\mathbf{1}_k + \mathbf{g}\epsilon_{t+h-j}\right)\right) + \mathrm{E}\left(\log\left(1+\epsilon_{t+h}\right)\right). \quad (4)$$

The point forecast from this model is:

$$\hat{y}_{t+h|t} = \exp(\log l_t + h \log b_t) = l_t b_t^h \quad (5)$$

In general $\mathrm{E}(\log(\mathbf{1}_k + \mathbf{g}_i\epsilon_{t+h-j})) \geq \mathbf{0}_k$

Properties of ETS(M,*,*)

And according to Jensen's inequality $\mathsf{E}(y_{t+h}|t) \geq \exp(\mathsf{E}(\log y_{t+h}|t))$

So, the expectation will be typically greater than point forecasts.

So for ETS(M,*,*) models we will have:

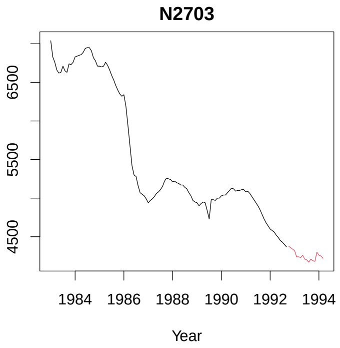
$$\hat{y}_{t+h} \leq \check{y}_{t+h} \leq \mu_{y,t+h} \quad (6)$$

(point forecast $\leq$ geometric mean $\leq$ expectation).

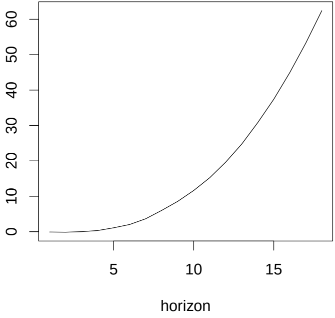
They will be equal only when either:

- $h = 1$,
- or $\alpha = \beta = \gamma = 0$.

An example of ETS(M,M,N) on the N2703 from M3



(a) Series N2703 from M3 data.



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

Properties of ETS(M,*,*)

Several take-aways from a side experiment:

- the increase of ϕ leads to the highest increase in bias,
- this is followed by β,
- and then forecast horizon,
- the impact of α on bias is the smallest,
- if variance is low (high value data), this is not important...
- ...but on low level data, it's not!

Use simulations for expectations!

Already done in `adam()` from `smooth`.

Making distributions cool again

Distributions for ETS(M,*,*)

Now, what was that about Normal distribution?

Hyndman et al. (2008) assume that: $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$

This implies that: $1 + \epsilon_t \sim \mathcal{N}(1, \sigma^2)$

This works fine on high level data.

But breaks on the low values.

Note! ETS models rely on $\mathrm{E}(1 + \epsilon_t) = 1$

If this is violated, the transition equation might break.

Distributions for ETS(M,*,*)

We can have the following options for $1 + \epsilon_t$:

	$\mathcal{N}(\mu, \sigma^2)$	$\mathcal{IG}(\mu, \sigma^2)$	$\log\mathcal{N}(\mu, \sigma^2)$	$\Gamma(\xi, \sigma^2)$
Support	$(-\infty, \infty)$	$(0, \infty)$	$(0, \infty)$	$(0, \infty)$
Mean	μ	μ	$e^{\mu+\frac{\sigma^2}{2}}$	$\xi\sigma^2$
Variance	σ^2	$\sigma^2\mu^3$	$(e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$	$\xi\sigma^4$
Skewness	0	$3\sqrt{\mu}\sigma$	$(e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1}$	$\frac{2}{\sqrt{\sigma^2}}$
Kurtosis	3	$15\mu\sigma^2$	$e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 3$	$\frac{6}{\xi}$
Parameters restrictions in ETS	$\mu = 1$	$\mu = 1$	$\mu = -\frac{\sigma^2}{2}$	$\xi = \frac{1}{\sigma^2}$

Table: Comparison of several distributions. μ is the location, σ^2 is the scale and ξ is the shape parameters.

Distributions for ETS(M,*,*)

log$\mathcal{N}$, $\mathcal{IG}$ and Γ are more flexible than $\mathcal{N}$.

They have flexible skewness and kurtosis.

They converge to $\mathcal{N}$ with the increase of level of data.

$\mathcal{IG}$ has fatter tail than Γ and log$\mathcal{N}$.

Distributions for ETS(M,*,*)

Given the restrictions, the log-likelihoods will be:

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

Table: $\boldsymbol{\theta}$ is the vector of all estimated parameters $\mathbf{y}$ 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.

Distributions for ETS(M,*,*)

The estimates of scale for each distribution:

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

Distributions for ETS(M,*,*)

ETS works perfectly with all these distributions.

What is different:

- Estimates of parameters,
- Fit,
- Prediction intervals.

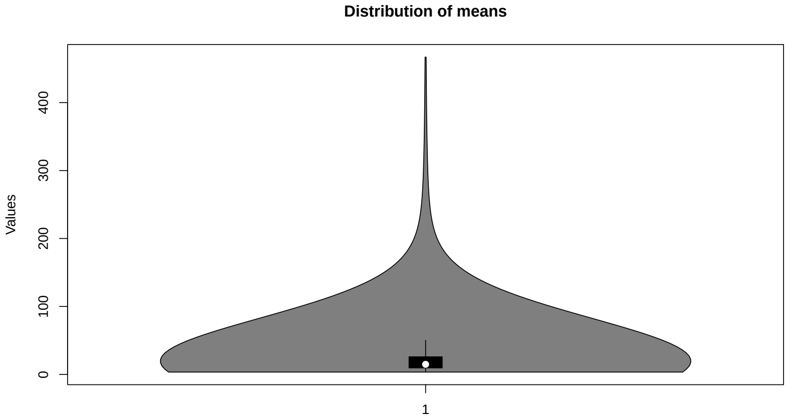
Having likelihoods, we can select between the distributions and choose the most suitable.

This is already implemented in `adam()` from `smooth`.

An experiment

An experiment. The data

Weekly data of a retailer – 1606 series, 111 observations



An experiment. The models

ETS with distributions discussed above

+ Selection between pure multiplicative ETS models with:

- log$\mathcal{N}$,
- $\mathcal{IG}$,
- Γ.

+ Point forecasts

Done via `adam()` and `auto.adam()` from `smooth`.

With horizon of 6 weeks ahead

An experiment. The error measures

1. MASE from Hyndman and Koehler (2006),
2. RMSSE from M5 (Makridakis et al., 2020b),
3. sMIS from M4 (Makridakis et al., 2020a),
4. Coverage of prediction intervals.

(1) is needed for point forecasts, just in case they are geometric mean

(2) is needed to see arithmetic means performance

An experiment. Mean values

	MASE	RMSSE	sMIS	Coverage
ETS Normal	0.890	0.793	2.482	0.931
ETS Point Normal	0.887	0.791		
ETS LogNormal	0.991	0.874	3.156	0.942
ETS Point LogNormal	1.036	0.918		
ETS Gamma	0.992	0.874	2.759	0.927
ETS Point Gamma	1.022	0.901		
ETS IG	1.021	0.902	3.248	0.916
ETS Point IG	1.109	0.977		
ETS Auto	0.994	0.879	2.974	0.926
ETS Point Auto	1.041	0.921		

An experiment. Mean values

	MASE	RMSSE	sMIS	Coverage
ETS Normal	0.890	0.793	**2.482**	0.931
ETS Point Normal	**0.887**	**0.791**		
ETS LogNormal	0.991	0.874	3.156	**0.942**
ETS Point LogNormal	1.036	0.918		
ETS Gamma	0.992	0.874	2.759	0.927
ETS Point Gamma	1.022	0.901		
ETS IG	1.021	0.902	3.248	0.916
ETS Point IG	1.109	0.977		
ETS Auto	0.994	0.879	2.974	0.926
ETS Point Auto	1.041	0.921		

An experiment. Mean values

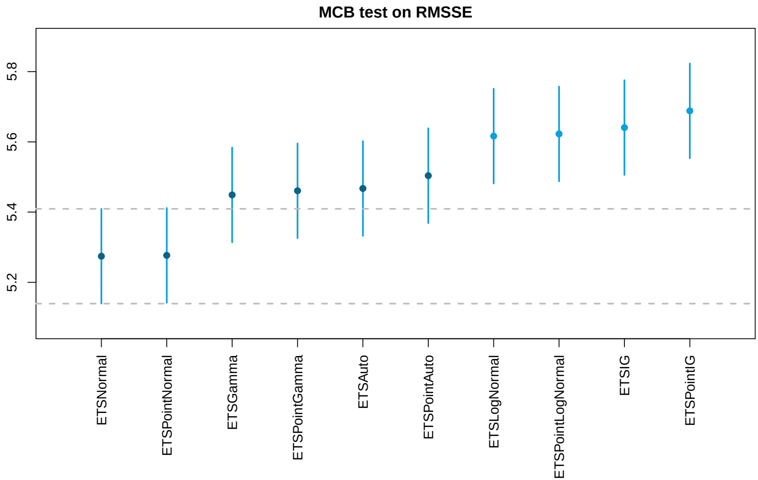
	MASE	RMSSE	sMIS	Coverage
ETS Normal	0.890	0.793		
ETS Point Normal	0.887	0.791		
ETS LogNormal	0.991	0.874		
ETS Point LogNormal	1.036	0.918		
ETS Gamma	0.992	0.874		
ETS Point Gamma	1.022	0.901		
ETS IG	1.021	0.902		
ETS Point IG	1.109	0.977		
ETS Auto	0.994	0.879		
ETS Point Auto	1.041	0.921		

An experiment. Median values

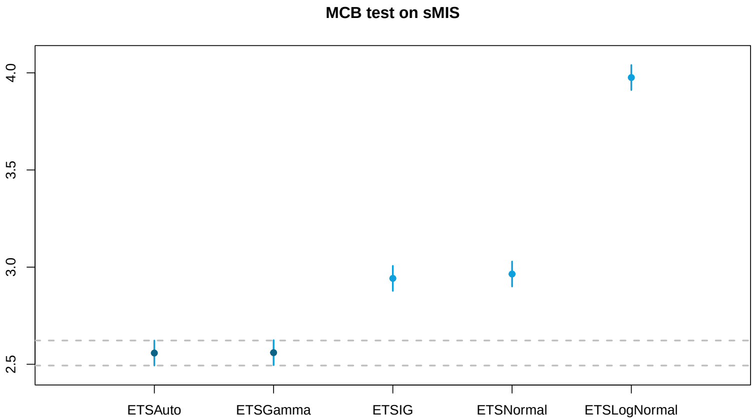
	MASE	RMSSE	sMIS
ETS Normal	0.741	0.683	**1.689**
ETS Point Normal	0.741	0.685	
ETS LogNormal	0.753	0.686	1.970
ETS Point LogNormal	0.754	0.687	
ETS Gamma	0.743	0.683	1.712
ETS Point Gamma	0.744	0.683	
ETS IG	0.749	0.687	1.765
ETS Point IG	0.749	0.690	
ETS Auto	0.746	0.685	1.706
ETS Point Auto	0.746	0.687	

Not clear, which is better...

An experiment. MCB on RMSSE



An experiment. MCB on sMIS



Conclusions

Conclusions

- Point forecasts are neither arithmetic, not geometric means in ETS(M,*,*) models;
- The difference is bigger with the increase of h, variance of error and smoothing parameters;
- We can use log Normal, Inverse Gaussian and Gamma distributions for the error term;
- We can have an automatic selection between them;

Conclusions

- Point forecasts are better than expectations for Normal...
- ...but they don't make sense!
- For all the other distributions, expectations are better;
- Differences in terms of point forecasts are not significant;
- As for sMIS, ETS Auto and ETS Gamma are better than the others;
- All of this is already available in `adam()` from `smooth`;
- Also check out the textbook (work in progress) on ADAM: https://openforecast.org/adam.

Thank you for your attention!

Ivan Svetunkov

i.svetunkov@lancaster.ac.uk

https://forecasting.svetunkov.ru

https://openforecast.org/adam

twitter: @iSvetunkov

Marketing Analytics and Forecasting

Lancaster University Management School

References I

Akram, M., Hyndman, R. J., Ord, J. K., dec 2009. Exponential Smoothing and Non-negative Data. Australian & New Zealand Journal of Statistics 51 (4), 415–432.
URL `http://doi.wiley.com/10.1111/j.1467-842X.2009.00555.x`

Athanasopoulos, G., Hyndman, R. J., Song, H., Wu, D. C., 2011. The tourism forecasting competition. International Journal of Forecasting 27 (3), 822–844.
URL `http://dx.doi.org/10.1016/j.ijforecast.2010.04.009`

References II

Fildes, R., Hibon, M., Makridakis, S., Meade, N., 1998. Generalising about univariate forecasting methods: further empirical evidence. International Journal of Forecasting 14 (3), 339–358.

Hyndman, R. J., Koehler, A. B., 2006. Another look at measures of forecast accuracy. International Journal of Forecasting 22 (4), 679–688.

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

References III

Makridakis, S., Andersen, A. P., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., Winkler, R. L., 1982. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. Journal of Forecasting 1 (2), 111–153.
URL http://doi.wiley.com/10.1002/for.3980010202

Makridakis, S., Hibon, M., oct 2000. The M3-Competition: results, conclusions and implications. International Journal of Forecasting 16 (4), 451–476.
URL https://linkinghub.elsevier.com/retrieve/pii/S0169207000000571

References IV

Makridakis, S., Spiliotis, E., Assimakopoulos, V., 2020a. The M4 Competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting 36 (1), 54–74.
URL
https://doi.org/10.1016/j.ijforecast.2019.04.014

Makridakis, S., Spiliotis, E., Assimakopoulos, V., 2020b. The M5 Accuracy Competition: Results, Findings and Conclusions.
URL
https://www.researchgate.net/publication/344487258