

# Forecasting using exponential smoothing: the past, the present, the future

Ivan Svetunkov, Nikos Kourentzes

OR60

13th September 2018

Marketing Analytics  
and Forecasting



Lancaster University  
Management School

# Introduction

Exponential smoothing (ES) is one of the most popular forecasting methods in practice:

(Weller and Crone, 2012) report that ES is used the most frequently by practitioners, in  $\sim 32\%$  of cases.

Any other sources?

# Introduction

ES has performed very well in different competitions over the years (Makridakis et al., 1982; Fildes et al., 1998; Makridakis and Hibon, 2000; Athanasopoulos et al., 2011).

It is now considered as one of the basic benchmarks and base forecasting models (see, for example, Maia and de Carvalho, 2011; Wang et al., 2012; Ramos et al., 2015; Kolassa, 2016).

But has this always been like that?

# Introduction

Before we proceed, we need define two terms (Svetunkov and Boylan, 2017):

**A statistical model** is as 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 statistical model.

# Introduction

We can distinguish three streams of application of ES methods and the related models:

1. Demand on fast moving products;
2. Intermittent demand (see presentation by John Boylan today);
3. Demand with multiple seasonalities (partially covered by Devon Barrow yesterday).

We will mainly focus on (1).

Now we can continue...

## The past



The past of exponential smoothing

# The past

Good overview of the past of ES is done by Gardner (1985) and Gardner (2006).

Here we discuss it briefly.

Simple Exponential Smoothing (SES) method was proposed in Brown (1956) and Holt (2004) in the form:

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha) \hat{y}_t, \quad (1)$$

## Simple exponential smoothing

SES in the error correction form:

$$\hat{y}_{t+1} = \hat{y}_t + \alpha e_t, \quad (2)$$

Or as a system of two equations:

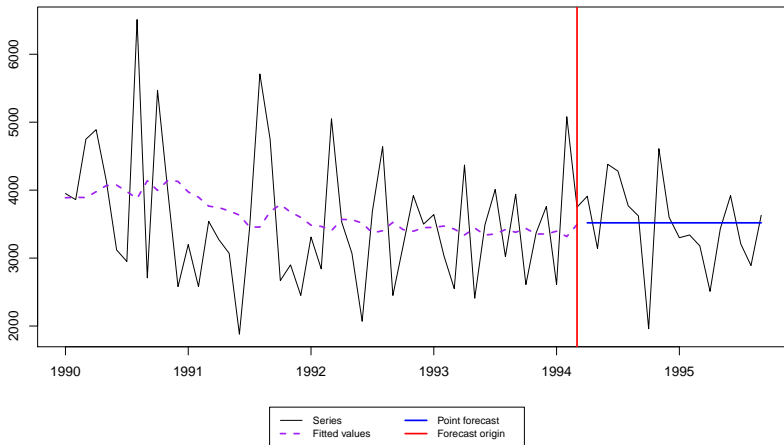
$$\begin{aligned} \hat{y}_{t+1} &= l_{t-1} \\ l_t &= l_{t-1} + \alpha e_t' \end{aligned} \quad (3)$$

where  $l_t$  is the level of series,  $\alpha$  is the smoothing parameter and  $e_t$  is the one-step-ahead forecast error.

Muth (1960) demonstrated that SES has underlying ARIMA(0,1,1) model with  $\theta = 1 - \alpha$ .



# Simple Exponential Smoothing



## Holt's method

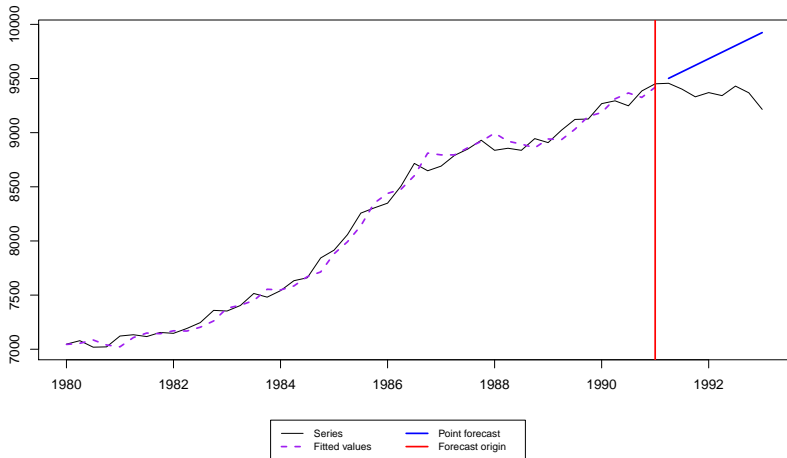
Holt (2004) proposed the trend model in 1957 (presented here in error correction form):

$$\begin{aligned}\hat{y}_{t+1} &= l_t + b_t \\ l_t &= l_{t-1} + b_{t-1} + \alpha e_t, \\ b_t &= b_{t-1} + \beta e_t\end{aligned}\tag{4}$$

where  $b_t$  is the trend component.

Nerlove and Wage (1964) demonstrated that this has an underlying ARIMA(0,2,2) model with  $\theta_1 = 2 - \alpha - \beta$  and  $\theta_2 = \alpha - 1$

# Holt's method



## Holt-Winters method

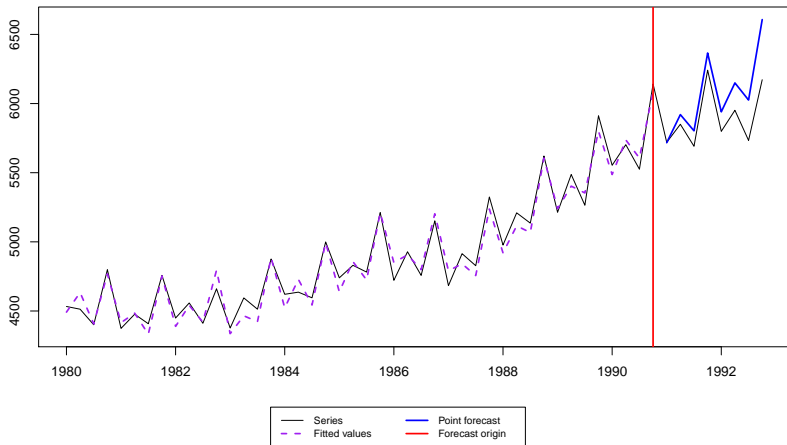
Winters (1960) developed a seasonal version of Holt's model that contains seasonal indices:

$$\begin{aligned}\hat{y}_{t+1} &= (l_t + b_t)s_{t-m+1} \\ l_t &= l_{t-1} + b_{t-1} + \alpha \frac{e_t}{s_{t-m}} \\ b_t &= b_{t-1} + \beta \frac{e_t}{s_{t-m}}, \\ s_t &= s_{t-m} + \gamma \frac{e_t}{s_{t-m}}\end{aligned}\tag{5}$$

where  $s_t$  is the seasonal component.

Chatfield (1977) shows that there is no underlying ARIMA for the multiplicative seasonal Holt-Winters method (5).

# Holt-Winters method



# Pegels taxonomy

There were other modifications proposed over the time.

Pegels (1969) suggested classifying exponential smoothing by types of trend and seasonal components.

According to him trend and seasonal components can be:

1. None,
2. Additive,
3. Multiplicative

This gives 9 possible exponential smoothing methods.

## Damped trend

Exponential smoothing performed well in M-Competition  
Makridakis et al. (1982).

Observing the performance of Holt's method in M-Competition,  
Gardner and McKenzie (1985) proposed a damped trend model,  
expanding the set of ES methods:

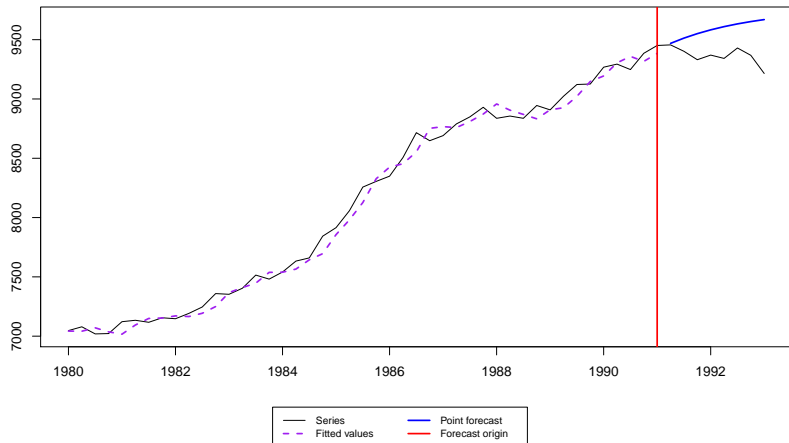
$$\begin{aligned}\hat{y}_{t+1} &= l_t + \phi b_t \\ l_t &= l_{t-1} + \phi b_{t-1} + \alpha e_t, \\ b_t &= \phi b_{t-1} + \beta e_t\end{aligned}\tag{6}$$

where  $\phi \in (0, 1)$  is a damping parameter.

This method has an underlying ARIMA(1,1,2) model.

Gardner and McKenzie (1989) discuss seasonal version of this.

# Damped trend (Gardner's method)





# State space models, MSOE

In parallel, Kalman (1960) proposed a model that will be later called “Multiple Source of Errors State Space model”.

It emerged from the field of engineering.

It consists of two parts: measurement (or prediction) equation and transition (or updating) equation.

The original model is quite complicated but then it was simplified by Harvey (1984).

## State space models, MSOE

In general, state space model with MSOE can be represented with:

$$\begin{aligned} y_t &= \mathbf{w}'_t \mathbf{v}_t + \epsilon_t \leftarrow \text{Measurement equation} \\ \mathbf{v}_t &= \mathbf{F} \mathbf{v}_{t-1} + \boldsymbol{\eta}_t \leftarrow \text{Transition equation} \end{aligned} \quad (7)$$

where  $\mathbf{v}_t$  is a vector of some components (for example, level, trend, seasonality),

$\epsilon_t \sim \mathcal{N}(0, \sigma^2)$  is an observation error and  $\boldsymbol{\eta}_t \sim \mathcal{N}(0, \Sigma)$  is the vector of error terms for the components

$\mathbf{w}_t$  is a measurement vector,  $\mathbf{F}$  is a transition matrix,

Matrix  $\mathbf{F}$  and vector  $\mathbf{w}_t$  are usually defined by a researcher.

# MSOE state space model

Muth (1960) and Theil and Wage (1964) showed that several exponential smoothing methods produce forecasts equivalent to the ones produced by MSOE.

Thus, additive exponential smoothing methods have underlying MSOE models (with some restrictions on parameters).

## The problem of ES in 20th century

The main concern of statisticians: if ARIMA underlies ES, why bother with the “special cases”?

The main concern of engineers: if MSOE underlies ES, why bother with the forecasting method?

It appears that ES as a method performs very well in practice (Makridakis et al., 1982).

It appears that ES is much easier to understand and work with than ARIMA and MSOE.

Moreover, MSOE is hard to estimate.

## SSOE state space model

Snyder (1985) modified MSOE and proposed a “Single Source of Error” state space model:

$$\begin{aligned}y_t &= \mathbf{w}'\mathbf{v}_{t-1} + \epsilon_t \\ \mathbf{v}_t &= \mathbf{F}\mathbf{v}_{t-1} + \mathbf{g}\epsilon_t\end{aligned}\quad (8)$$

where  $\mathbf{g}$  is a persistence vector, containing smoothing parameters.

Now all the components are influenced by the same error.

This is much easier to understand and estimate than MSOE

## SSOE state space model

An example is a local level model, where  $\mathbf{w} = 1$ ,  $\mathbf{F} = 1$  and  $\mathbf{g} = \alpha$ :

$$\begin{aligned}y_t &= l_{t-1} + \epsilon_t \\l_t &= l_{t-1} + \alpha\epsilon_t\end{aligned}\quad (9)$$

Reminds us SES in the expanded error correction form:

$$\begin{aligned}\hat{y}_{t+1} &= l_t \\l_t &= l_{t-1} + \alpha e_t\end{aligned}\quad (10)$$

Thus the main critique of SSOE was that this is not a model, but just an ES with an error term.

## SSOE state space model

Still Ord et al. (1997) expanded SSOE for the cases of multiplicative components and multiplicative error:

$$\begin{aligned}y_t &= w(\mathbf{v}_{t-1}) + r(\mathbf{v}_{t-1})\epsilon_t \\ \mathbf{v}_t &= f(\mathbf{v}_{t-1}) + g(\mathbf{v}_{t-1})\epsilon_t'\end{aligned}\tag{11}$$

so that SSOE now underlied all the possible ES methods.

Hyndman et al. (2002) expanded the Pegels' taxonomy.

Model selection in SSOE framework based on an information criterion.

Given 2 types of errors, 4 types of trends and 3 types of seasonal components, ES in SSOE contained 24 models.

## SSOE state space model

To make things even more exciting, Hyndman et al. (2008) added Taylor (2003) ES with damped multiplicative trend and presented a framework uniting 30 models.

They called the framework ETS - Error, Trend, Seasonality.




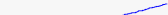





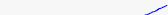
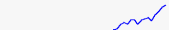
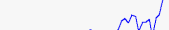
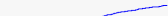


ETS is relatively easy to use and estimate, it provides appropriate prediction intervals and supports model selection (with information criteria).

It underlies 15 ES methods.

It can be expanded by inclusion of exogenous variables or additional seasonal components.



# ETS framework

Trend	Seasonality		
	None	Additive	Multiplicative
None			
Additive			
Additive Damped			
Multiplicative			
Multiplicative Damped			

## The present



The present of exponential smoothing

# The present

We define “present” as a period starting from Hyndman et al. (2008)

ETS is a standard in forecasting nowadays.

But there are still new papers discussing ES method instead of SSOE / MSOE frameworks.



## Demand on fast moving products

Kourentzes et al. (2014) develop a Multi Aggregation Prediction Algorithm (MAPA) based on ETS model.

MAPA performs very well in different competitions.

Kourentzes and Petropoulos (2016) extended MAPA to include exogenous variables.

## Demand on fast moving products

Svetunkov and Kourentzes (2015) propose Complex Exponential Smoothing (CES), that sidesteps ETS taxonomy.

CES produces exponential trends in a style of multiplicative trends models, but performs more accurately.

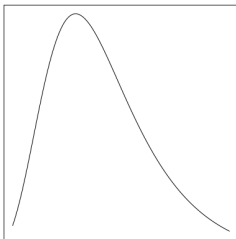
Osman and King (2015) discuss a time varying parameter model with exogenous variables in SSOE framework.

It is built upon ETS and allows updating the parameters in the exponential smoothing style.

## Demand on fast moving products

Guo et al. (2016) discuss new type of trend - life-cycle trend.

They use pure multiplicative ETS(M,M,N) model, and add a turn-down parameter.



## Demand on fast moving products

de Silva et al. (2010) develops Vector Exponential Smoothing - the multivariate counterpart of ETS.

Athanasopoulos and de Silva (2012) demonstrate the advantages of VES on an example of Australian tourism data.

Not to mention progress in intermittent demand and multiple seasonalities demand...



## The future



And now for something completely different...

# The future (missing bits)

New models with new types of trends.

New developments in multivariate domain.

Intermittent models with multiple seasonalities.

Models connecting the three worlds.

## Generalised Univariate Model

All the state-space models with SSOE have similar characteristics and can be summarised with the formula:

$$\begin{aligned}y_t &= \mathbf{w}'\mathbf{v}_{t-\mathbf{L}} + \epsilon_t \\ \mathbf{v}_t &= \mathbf{F}\mathbf{v}_{t-\mathbf{L}} + \mathbf{g}\epsilon_t\end{aligned}\quad (12)$$

where  $\mathbf{L}$  is a vector of lags of each component.

For example, ETS(A,N,A) can be summarised as with  $\mathbf{v}_t = \begin{pmatrix} l_t \\ s_t \end{pmatrix}$

and  $\mathbf{v}_{t-\mathbf{L}} = \begin{pmatrix} l_{t-1} \\ s_{t-m} \end{pmatrix}$ , where  $m$  is the lag of seasonality.

Thus  $\mathbf{L} = \begin{pmatrix} 1 \\ m \end{pmatrix}$  in this example.

# Generalised Univariate Model

Why not define the general model, with **fuzzy components**, where the elements of  $\mathbf{w}$ ,  $\mathbf{F}$  and  $\mathbf{g}$  would be estimated?

Each component can have a specific lag (e.g. 1, 2,  $m_1$ ,  $m_{41}$ , ...).

Number of components can be arbitrary.

We specify GUM the following way:

$$\text{GUM}(p_1^{m_1}, p_2^{m_2}, \dots, p_k^{m_k}),$$

where  $p_j$  is number of components of type  $j$ ,  $m_j$  is the lag of components  $j$ .

# Generalised Univariate Model

Example: GUM( $2^1$ ) underlies Holt's method.

It is formulated as:

$$\begin{aligned}y_t &= w_1 v_{1,t-1} + w_2 v_{2,t-1} + \epsilon_t \\v_{1,t} &= f_{1,1} v_{1,t-1} + f_{1,2} v_{2,t-1} + g_1 \epsilon_t, \\v_{2,t} &= f_{2,1} v_{1,t-1} + f_{2,2} v_{2,t-1} + g_2 \epsilon_t\end{aligned}\tag{13}$$

# Generalised Univariate Model

GUM( $2^1$ ):

$$\begin{aligned}y_t &= w_1 v_{1,t-1} + w_2 v_{2,t-1} + \epsilon_t \\v_{1,t} &= f_{1,1} v_{1,t-1} + f_{1,2} v_{2,t-1} + g_1 \epsilon_t, \\v_{2,t} &= f_{2,1} v_{1,t-1} + f_{2,2} v_{2,t-1} + g_2 \epsilon_t\end{aligned}$$

ETS(A,Ad,N):

$$\begin{aligned}y_t &= 1 \cdot l_{t-1} + \phi \cdot b_{t-1} + \epsilon_t \\l_t &= 1 \cdot l_{t-1} + \phi \cdot b_{t-1} + \alpha \epsilon_t, \\b_t &= 0 \cdot l_{t-1} + \phi \cdot b_{t-1} + \beta \epsilon_t\end{aligned} \tag{14}$$

# Generalised Univariate Model

$GUM(2^1)$  has  $ETS(A,A,N)$  and  $ETS(A,Ad,N)$  as special cases.

$GUM(2^1)$ : also underlies:  $ARIMA(0,2,2)$ ,  $ARIMA(1,1,2)$ ,  
 $ARIMA(2,0,2)$ ,  $ARIMA(1,0,2)$ ,  $ARIMA(0,1,2)$ ,  $ARIMA(0,0,2)$ ,  
 $ARIMA(0,2,1)$ ,  $ARIMA(2,0,1)$ ,  $ARIMA(0,2,0)$ ,  $ARIMA(2,0,0)$

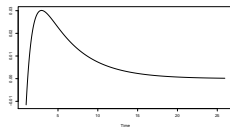
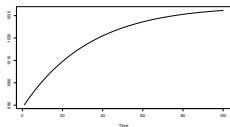
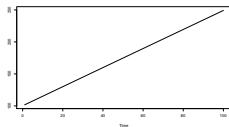
In fact, GUM underlies all the existing additive ETS and ARIMA models.

# Generalised Univariate Model

Why bother with GUM?

It can model and extrapolate any types of trends.

For example,  $GUM(2^1)$ :



GUM is a perfect approximator.



# Generalised Univariate Model

`gum()` function in `smooth` package is on CRAN  
Limitations of GUM:

- High number of parameters to estimate,
- Unrestricted model has a multitude of global minima (how to select the best?).



Potential solution – restrict parameters of the model.

e.g. assume that  $f_j, w_j \in (0, 1)$

# Conclusions



# Conclusions

- ES has gone through a lot of changes over the years,
- It started as an arbitrary forecasting method,
- It now has different underlying statistical models,
- ETS is one of the most popular models in forecasting,
- The field is still developing,
- There is a lot of exciting things in the future...
- IMHO, GUM is one of them.



# Thank you for your attention!

Ivan Svetunkov

[i.svetunkov@lancaster.ac.uk](mailto:i.svetunkov@lancaster.ac.uk)

<https://forecasting.svetunkov.ru>

twitter: @iSvetunkov

Marketing Analytics  
and Forecasting



Lancaster University

Management School

Athanasopoulos, G., de Silva, A., 2012. Multivariate Exponential Smoothing for Forecasting Tourist Arrivals. *Journal of Travel Research* 51 (5), 640–652.

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>

Brown, R. G., 1956. Exponential Smoothing for predicting demand.

Chatfield, C., 1977. Some Recent Developments in Time-Series Analysis. *Journal of the Royal Statistical Society. Series A (General)* 140 (4), 492.

URL [https:](https://www.jstor.org/stable/2345281?origin=crossref)

[//www.jstor.org/stable/2345281?origin=crossref](https://www.jstor.org/stable/2345281?origin=crossref)

de Silva, A., Hyndman, R. J., Snyder, R., dec 2010. The vector innovations structural time series framework. *Statistical*



Modelling: An International Journal 10 (4), 353–374.

URL <http://journals.sagepub.com/doi/10.1177/1471082X0901000401>

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.

Gardner, E. S., 1985. Exponential smoothing: The state of the art. Journal of Forecasting 4 (1), 1–28.

URL <http://doi.wiley.com/10.1002/for.3980040103>

Gardner, E. S., 2006. Exponential smoothing: The state of the art-Part II. International Journal of Forecasting 22 (4), 637–666.

Gardner, E. S., McKenzie, E., 1985. Forecasting trends in time series. Management Science 31 (10), 1237–1246.

Gardner, E. S., McKenzie, E., 1989. Seasonal Exponential



Smoothing with Damped Trends. *Management Science* 35 (3), 372–376.

Guo, X., Lichtendahl, K. C., Grushka-Cockayne, Y., 2016.

Forecasting Life Cycles with Exponential Smoothing.

URL <http://www.ssrn.com/abstract=2805244>

Harvey, A. C., jul 1984. A Unified View of Statistical Forecasting Procedures. *Journal of Forecasting* 3 (3), 245–275.

URL <http://doi.wiley.com/10.1002/for.3980030302>

Holt, C. C., 2004. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting* 20 (1), 5–10.

Hyndman, R. J., Koehler, A. B., Ord, J. K., Snyder, R. D., 2008.

Forecasting with Exponential Smoothing. Springer Berlin Heidelberg.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., Grose, S., jul 2002.

A state space framework for automatic forecasting using



exponential smoothing methods. *International Journal of Forecasting* 18 (3), 439–454.

URL <http://linkinghub.elsevier.com/retrieve/pii/S0169207001001108>

Kalman, R. E., 1960. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering* 82 (1), 35.  
URL <http://fluidsengineering.asmedigitalcollection.asme.org/article.aspx?articleid=1430402>

Koehler, A. B., Snyder, R. D., Ord, J. K., Beaumont, A., 2012. A study of outliers in the exponential smoothing approach to forecasting. *International Journal of Forecasting* 28 (2), 477–484.  
URL

<http://dx.doi.org/10.1016/j.ijforecast.2011.05.001>

Kolassa, S., 2011. Combining exponential smoothing forecasts using Akaike weights. *International Journal of Forecasting* 27 (2), 238–251.





## URL

<http://dx.doi.org/10.1016/j.ijforecast.2010.04.006>

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

## URL

<http://dx.doi.org/10.1016/j.ijforecast.2015.12.004>

Kourentzes, N., Petropoulos, F., nov 2016. Forecasting with multivariate temporal aggregation: The case of promotional modelling. *International Journal of Production Economics* 181, 145–153.

Kourentzes, N., Petropoulos, F., Trapero, J. R., 2014. Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting* 30 (2), 291–302.



## URL

<http://dx.doi.org/10.1016/j.ijforecast.2013.09.006>

Maia, A. L. S., de Carvalho, F. D. a. T., 2011. Holt's exponential smoothing and neural network models for forecasting interval-valued time series. *International Journal of Forecasting* 27 (3), 740–759.

Makridakis, S., Andersen, A. P., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., Winkler, R. L., apr 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., 2000. The M3-Competition: results, conclusions and implications. *International Journal of Forecasting* 16, 451–476.



Muth, J. F., 1960. Optimal Properties of Exponentially Weighted Forecasts. *Journal of the American Statistical Association* 55 (1), 299–306.

Nerlove, M., Wage, S., 1964. On the Optimality of Adaptive Forecasting. *Management Science* 10 (2), 207–224.

Ord, J. K., Koehler, A. B., Snyder, R. D., 1997. Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association* 92 (March 2013), 1621–1629.

Osman, A. F., King, M. L., 2015. A new approach to forecasting based on exponential smoothing with independent regressors. URL <http://econpapers.repec.org/paper/mshebswps/2015-2.htm>

Pegels, C. C., 1969. Exponential Forecasting : Some New Variations. *Management Science* 15 (5), 311–315.



- Ramos, P., Santos, N., Rebelo, R., 2015. Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and Computer-Integrated Manufacturing* 34, 151–163. URL <http://dx.doi.org/10.1016/j.rcim.2014.12.015>
- Snyder, R. D., 1985. Recursive Estimation of Dynamic Linear Models. *Journal of the Royal Statistical Society, Series B (Methodological)* 47 (2), 272–276.
- Svetunkov, I., Boylan, J. E., 2017. Multiplicative State-Space Models for Intermittent Time Series.
- Svetunkov, I., Kourentzes, N., 2015. Complex Exponential Smoothing. Tech. rep., Department of Management Science, Lancaster University. URL <https://ideas.repec.org/p/prs/mprapa/69394.html>
- Taylor, J. W., 2003. Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting* 19 (4), 715–725.

- Theil, H., Wage, S., 1964. Some Observations on Adaptive Forecasting. *Management Science* 10 (2), 198–206.  
URL <http://pubsonline.informs.org/doi/abs/10.1287/mnsc.10.2.198>
- Wang, J.-J., Wang, J.-Z., Zhang, Z.-G., Guo, S.-P., 2012. Stock index forecasting based on a hybrid model. *Omega* 40 (6), 758–766.  
URL <http://dx.doi.org/10.1016/j.omega.2011.07.008>
- Weller, M., Crone, S., 2012. *Supply Chain Forecasting: Best Practices and Benchmarking Study*. Tech. rep., Lancaster Centre for Forecasting.
- Winters, P. R., 1960. Forecasting Sales by Exponentially Weighted Moving Averages. *Management Science* 6 (3), 324–342.

