Modified simple exponential smoothing

Guckan Yapar*†

Abstract

In this study, we propose a new exponential smoothing method, modified simple exponential smoothing (MSES) as an alternative to simple exponential smoothing (SES). Despite its success and widespread use in many areas, SES has some shortcomings that negatively affect the accuracy of forecasts made using this method. For example, there is no agreed upon concensus on choosing an initial value and determining an optimum smoothing parameter and these decisions greatly affect the forecasting accuracy of SES. The proposed method will help cope with these shortcomings. It is compared to SES on popular metrics that are commonly used for evaluating performance of forecasting techniques and is shown to have better performance. The two models are applied to the 1001 time series of the M-competition data simultaneously and their prediction accuracies are compared under various settings.

1. Introduction

Time series data arise in many different contexts including finance and industry, whenever something is observed over time. The main purpose in these cases involves using a sequence of observations on some variable to predict a future value of it. This is achieved by using some aggregation of the past observations to predict the future values. There are many studies in the literature dealing with this problem utilizing forecasting and smoothing techniques. Let the observed values of a random variable over time be denoted by x_t, $t = 1, \ldots, n$. The aim is then to obtain an estimate for x_{n+1}. For simplicity, it is assumed that the data do not display any clear trending behavior or any seasonality, although the mean of the data may be changing slowly over time. The method proposed later on can be easily adapted to handle data that involve such components. For now, assume x_t can be modeled using only a random error component as below:

$$(1.1) \qquad x_t = a + e_t,$$

---

* Department of Statistics, Faculty of Science, Dokuz Eylul University, Tinaztepe, Izmir, Turkey, Email: `guckan.yapar@deu.edu.tr`
†Corresponding Author.

where e_t is some random noise with mean zero and variance σ^2. Under the model in (1.1), the aim is then reduced to finding a good estimator for the constant a so that it can be used to forecast future values. The general form of this estimator should involve some sort of an average of the observed values. It can be notated as:

$$(1.2) \qquad \hat{a} = F(x_1, \ldots, x_n) = \sum_{t=1}^{n} w_t x_t,$$

where w_t are a collection of weights called weighting vector such that $w_t \in [0,1]$ for $t = 1, \ldots, n$ and $\sum_{t=1}^{n} w_t = 1$. The estimators of form (1.2) will be unbiased. In order to deal with sequential updating, the term a_n is sometimes used to indicate the smoothed value at time n, therefore a_n and $\hat{a}$ are synonyms.

Since there are a lot of ways to obtain estimators of form (1.2) and there is not an estimator that will be universally satisfactory, researchers need a way to choose among all these potential smoothing methods. When making a choice, some criteria to judge the relative merits of each alternative that are important to the researchers are needed. Most importantly, researchers have a preference for fresh data and therefore in practice weighting vectors that assign more weight to recent observations are preferred. In other words, weighting vectors with $w_j \geq w_i$ for $j > i$ are preferred. One popular metric that is used for measuring a smoothing method's ability to utilize fresh data is the average age (AA) of the data used:

$$(1.3) \qquad AA = n - \sum_{t=1}^{n} t w_t.$$

Another important metric to consider is the variance of the estimator at hand as usual. Since the estimator in (1.2) is unbiased, its variance can be written as:

$$(1.4) \qquad Var(\hat{a}) = E\left[\left(\sum_{t=1}^{n} w_t x_t - a \right)^2 \right] = \sum_{t=1}^{n} w_t^2 \sigma^2 = V\sigma^2.$$

Even though it is desirable to keep both of the metrics in (1.3) and (1.4) minimal simultaneously, it is not an achievable goal. Consider two extreme weighting schemes which will result in boundary values of these metrics. The first scheme is the average method which assigns equal weights to all observations over time, i.e. $w_t = \frac{1}{n}$, for $t = 1, \ldots, n$. Here the estimator $\hat{a}$ is simply the simple average and it is well known that for the conditions $w_t \in [0,1]$ and $\sum_{t=1}^{n} w_t = 1$ the variance in (1.4) is minimized since $V = \frac{1}{n}$. On the contrary, AA attains its largest value under this weighting scheme which is equal to $\frac{n-1}{2}$. Under the naive weighting scheme where all the observations other than the latest one are discarded, i.e. $w_n = 1$ and $w_t = 0$ for $t = 1, \ldots, n-1$, the estimator is simply equal to the latest observation. AA under this scenario will be equal to zero thus minimized but this time the variance of the estimator will be maximized since V now reaches its maximum value which is equal to 1.

Even though average and naive weighting schemes are simple methods that work remarkably well for many economic and financial time series, it is more realistic to use weighting schemes that assign more weight on current observations without having to give up all the remaining observations. One such parameterized method is the classic moving average (MA) where for the parameter, the size of the window p, the model can be written as:

$$(1.5) \qquad \hat{a} = \frac{x_n + x_{n-1} + \cdots + x_{n-p+1}}{p} = \frac{1}{p} \sum_{j=0}^{p-1} x_{n-j},$$

for $p \leq n$. For this model the weights are $w_j = 0$ for $j \leq n - p$ and $w_j = \frac{1}{p}$ for $n - p + 1 \leq j \leq n$. This model has $AA = \frac{p-1}{2}$ and $V = \frac{1}{p}$ (Brown, 1962).

Another classic and well known approach that allows the researchers to utilize more data is the SES method where for the smoothing constant $\alpha \in [0, 1]$, the model can be written as:

$$(1.6) \qquad a_n = \alpha x_n + (1 - \alpha)a_{n-1},$$

where a_n is the smoothed value at time n which is as mentioned earlier in the paper a synonym for $\hat{a}$. Substituting the model in (1.6) into itself successively, the model can be re-written as:

$$(1.7) \qquad a_n = \alpha \sum_{k=0}^{n-1} (1 - \alpha)^k x_{n-k} + (1 - \alpha)^n a_0,$$

so a_n represents a weighted moving average of all past observations with weights decreasing exponentially, a_0 is the initial value. It can be seen that for large α recent observations get more weight. For large sample sizes the ES estimator in (1.6) is unbiased and has $AA = \frac{1-\alpha}{\alpha}$ and $V = \frac{\alpha}{2-\alpha}$ (Brown, 1962).

Exponential smoothing (ES) methods are the most widely used techniques in forecasting due to their simplicity, robustness and accuracy as an automatic forecasting procedure. ES was proposed in the late 1950s (Holt, 1957; Brown, 1959; Winters, 1960). However, their popularity in time series analysis is not just as a result of simplicity but also their proven superiority against more sophisticated approaches (Makridakis et al., 1984; Makridakis and Hibon, 2000). ES models assume that the time series have up to three underlying data components: level, trend and seasonality. Estimates for the final values of these components are used to construct the forecast. An ES model can include of one of five types of trend (none, additive, damped additive, multiplicative, or damped multiplicative) and one of three types of seasonality (none, additive, or multiplicative). Pegels (1969) proposed taxonomy of ES methods, which was extended and modified later by Gardner Jr and McKenzie (1985), Hyndman et al. (2002), Taylor (2003) and Hyndman and Athanasopoulos (2014). Thus, there are 15 different ES models, the best known of which are simple exponential smoothing (SES) (no trend, no seasonality), Holt's linear model (additive trend, no seasonality) and Holt-Winters' additive model (additive trend, additive seasonality)(Goodwin et al., 2010). For this reason we can say that ES is not a simple model but rather a family of models.

Appropriate choice of smoothing constants and initial values in any ES model play key roles in successful forecasting. An extensive review and discussion of ES models and initial value and smoothing constant selection for the various ES models is given by Gardner (2006). Despite their success and popularity and large body of research on this topic, there has never been a consensus among forecasters and there are no consistent guidelines in the forecasting literature on how smoothing constants and initial values should be selected. In this study, a new smoothing framework will be introduced as an alternative to traditional SES method to cope with this shortcoming.

Even though we focus on providing and studying in detail only MSES, the alternative to the SES, in this paper, it should be kept in mind that the proposed framework can easily be adapted to higher order ES models.

In Section 2, we introduce the proposed method MSES and provide explicit formulas for its average age and variance. In Section 3, we compare MSES and SES in great detail. In Section 4, we illustrate the method by applying it to a data set from the M-competition data and then compare the overall forecasting performances of various MSES and SES models under two specific settings by applying both models to all 1001

series of the M-competition. Finally, in Section 5, we conclude by suggesting some future research.

2. Proposed method: Modified simple exponential smoothing (MSES)

Our purpose was to develop a parametric procedure for obtaining weights that provide the best forecast for x_{n+1} which also satisfy $\sum_{t=1}^{n} w_t = 1$ and $w_i \leq w_j$ for $i < j$. In order to do this, the shortcomings of ES models were examined in detail and a new model that can help cope with these was proposed. Let the smoothed value at time n be written as:

$$(2.1) \qquad a_n = \left(\frac{m}{n}\right) x_n + \left(\frac{n-m}{n}\right) a_{n-1},$$

where $m = 0, 1, \ldots, n$. This model has very similar form to traditional SES model so it will be called modified simple exponential smoothing (MSES) henceforth. To see that the weights sum to unity and therefore a_n can be interpreted as a weighted average of past observations, the model in (2.1) is applied recursively to all successive observations in the series as below:

$$a_{n-1} = \left(\frac{m}{n-1}\right) x_{n-1} + \left(\frac{n-m-1}{n-1}\right) a_{n-2},$$

$$a_n = \left(\frac{m}{n}\right) x_n + \left(\frac{m}{n}\right)\left(\frac{n-m}{n-1}\right) x_{n-1} + \left(\frac{n-m}{n}\right)\left(\frac{n-m-1}{n-1}\right) a_{n-2},$$

$$a_{n-2} = \left(\frac{m}{n-2}\right) x_{n-2} + \left(\frac{n-m-2}{n-2}\right) a_{n-3},$$

$$a_n = \left(\frac{m}{n}\right) x_n + \left(\frac{m}{n}\right)\left(\frac{n-m}{n-1}\right) x_{n-1} + \left(\frac{m}{n}\right)\left(\frac{n-m}{n-1}\right)\left(\frac{n-m-1}{n-2}\right) x_{n-2}$$
$$+ \left(\frac{n-m}{n}\right)\left(\frac{n-m-1}{n-1}\right)\left(\frac{n-m-2}{n-2}\right) a_{n-3},$$

$$\vdots$$

$$a_n = \left(\frac{m}{n}\right) x_n + \left(\frac{m}{n}\right)\left(\frac{n-m}{n-1}\right) x_{n-1} + \ldots$$
$$+ \frac{\binom{m}{m-1}}{\binom{n}{m}} x_{m+1} + \left(\frac{n-m}{n}\right)\left(\frac{n-m-1}{n-1}\right) \ldots \left(\frac{1}{m+1}\right) a_m.$$

Therefore, the smoothed value at time n obtained by MES can be re-written as:

$$(2.2) \qquad a_n = \sum_{k=0}^{n-(m+1)} \frac{\binom{n-k-1}{m-1}}{\binom{n}{m}} x_{n-k} + \frac{1}{\binom{n}{m}} a_m,$$

where a_m is the starting or initial value for MSES which can be simply the mth observation or the average of the oldest m observations. It can now easily be seen that the smoothed value at time n is a weighted average of past observations and the initial value a_m. The weights of MSES as given in (2.2) can be thought of as the probabilities from a Negative Hyper-Geometric distribution with parameters $(n, m, 1)$ where if a random variable X follows this distribution then

$$P(X = x) = \frac{\binom{n-x-1}{m-1}}{\binom{n}{m}},$$

for $x = 0, 1, 2, \ldots, n - m$ (Johnson and Kotz, 1977). Utilizing the expected value of this distribution, the average age of MSES can then be easily found as:

$$(2.3) \qquad AA_{MSES} = \frac{n - m}{m + 1}.$$

In order to calculate the variance of the MSES estimator, it is needed to calculate the sum of squared weights, V, once again. With a slight re-arrangement of the numerators in the weights, the sum can be written as:

$$
\begin{aligned}
V_{MSES} &= \sum_{t=1}^{n} w_t^2 \\
&= \left(\frac{m}{n}\right)^2 + \left(\frac{m}{n}\right)^2 \left(\frac{n-m}{n-1}\right)^2 + \left(\frac{m}{n}\right)^2 \left(\frac{n-m}{n-1}\right)^2 \left(\frac{n-m-1}{n-2}\right)^2 + \cdots \\
&\quad + \left(\frac{m}{n}\right)^2 \left(\frac{n-m}{n-1}\right)^2 \left(\frac{n-m-1}{n-2}\right)^2 \cdots \left(\frac{1}{m}\right)^2 \\
&= \left(\frac{m}{n}\right)^2 \left[1 + \sum_{i=0}^{n-m-1} \prod_{j=0}^{i} \left(\frac{n-m-j}{n-1-j}\right)^2\right] \\
(2.4) \qquad &= \left(\frac{m}{n}\right)^2 {}_3F_2\left((1, m-n, m-n), (1-n, 1-n), 1\right)
\end{aligned}
$$

From equation (2.4) it can be seen that the variance of the MSES estimator involves the Generalized Hyper-Geometric series:

$$ {}_3F_2\left((1, m-n, m-n), (1-n, 1-n), 1\right) $$

(Bailey, 1935). This framework can be easily adapted to incorporate higher order components when needed.

3. Comparison of SES and MSES

There is no doubt that the two methods SES and MSES are closely related and MSES lies somewhere between SES and MA. MSES attaches weights to only the most recent m observations like MA and the weights decrease exponentially like SES for some m ($m \geq 3$). Both SES and MSES methods need smoothing constants and initial values. The smoothing constants for SES are commonly estimated by minimizing the mean squared error (MSE), although the mean absolute error (MAE) and the mean absolute percentage error (MAPE) are also used. Gardner (1985) discusses various theoretical and empirical arguments for selecting an appropriate smoothing constant and concludes that it is best to estimate an optimum α from the data.

The main idea in SES is that the recent history is more representative of the near future and therefore more emphasis should be given to recent observations. So, intuitively, the starting point for grid search should be weighting all past observations equally (average method) and then giving greater emphasis to recent observations gradually until ending up by weighting the last observation by 1 (naive method). This would guarantee that the weight of the initial value stays less than or equal to the weight of the most current observation. This can easily be achieved with an MSES model with $m = 1$ for any n as below:

$$
\begin{aligned}
a_n &= \left(\frac{1}{n}\right) x_n + \left(\frac{n-1}{n}\right) a_{n-1} \\
&= \left(\frac{1}{n}\right) x_n + \left(\frac{1}{n}\right) x_{n-1} + \cdots + \left(\frac{1}{n}\right) x_2 + \left(\frac{1}{n}\right) a_1,
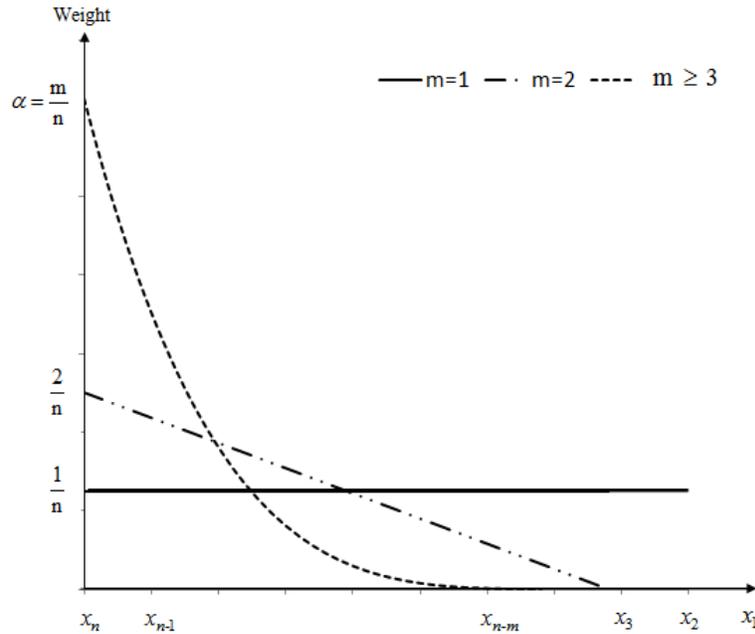\end{aligned}
$$

Figure 1. Weights attached to the observations by MSES for different m values.

where $a_1 = x_1$ and $a_n = \bar{x}$. This can not be achieved by a SES model for any parameter value of α. When $m = 2$, for any n the MSES model produces weights that decrease linearly with slope $2/(n(n-1))$ which again can never be achieved by SES since it always assigns exponentially decreasing weights to observations no matter the parameter choice. For $m = 2$ the MSES smoothed value at time n can be written as:

$$
\begin{aligned}
a_n &= \left(\frac{2}{n}\right) x_n + \left(\frac{n-2}{n}\right) a_{n-1} \\
&= \frac{2}{n} x_n + \frac{2\,(n-2)}{n\,(n-1)} x_{n-1} + \frac{2\,(n-3)}{n\,(n-1)} x_{n-2} + \cdots + \frac{2}{n}\frac{2}{(n-1)} x_3 + \frac{2}{n(n-1)} a_2,
\end{aligned}
$$

where $a_2 = (x_1 + x_2)/2$ or simply $a_2 = x_2$. For $m > 2$ the weights start to decrease exponentially as the observations get older as in SES but not exactly at the same rate. The weights of MSES can be visualized as in Figure 1. As a result, the MSES model is more flexible and intuitive compared to SES since it allows for more meaningful weighting schemes when searching for an optimal parameter while potentially reducing the number of iterations needed for reaching that optimal smoothing parameter.

After a smoothing parameter is obtained, it is also needed to find an initial value for both of the aforementioned models. For SES, most practitioners work with α values between 0.01 and 0.3. However, also known as the "initialization problem", when either n or α is small SES attaches more weight to initial value than even the most current observation. The choice of starting value then becomes particularly important for SES. It is conceptually wrong to think of the smoothing constant α alone without paying any attention to the sample size n. The main idea of ES is to assign more weight to recent observations and therefore an ES model should assign the most recent observation at
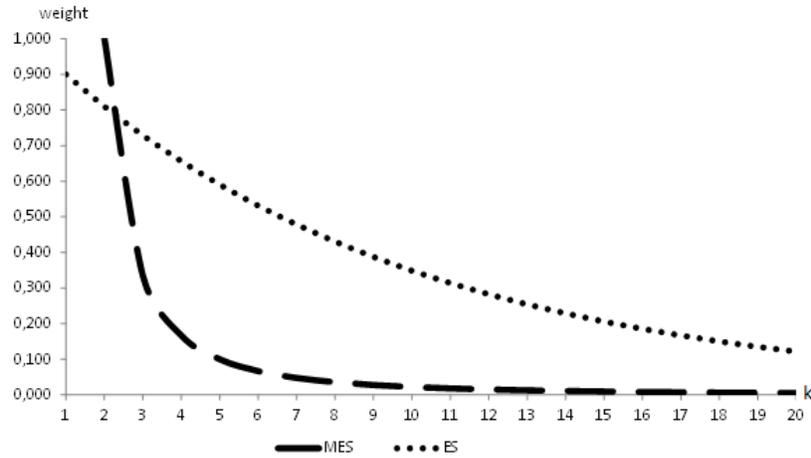
Figure 2. Weights attached to initial value by MSES and SES for different iterations $k = 1, 2, \ldots, n$.

least a weight of $\frac{1}{n}$. This can be achieved only if the search for the smoothing parameter is limited to the interval $[\frac{1}{n}, 1]$.

To compare, we set the smoothing constant of SES as $\alpha = m/n$ to make the smoothing constants of the two models equal. At the same smoothing constant level, MSES assigns less weight to the initial value even for small α. The weights assigned to the initial values by these two models are $(1-\alpha)^n$ and $(n-m)!m!/n!$ respectively. To visualize, the weights of initial values for both of these models for relatively short time series are plotted in Figure 2 for $m = 2$ and $n = 20$ resulting in $\alpha = m/n = 0.1$.

As discussed in Section 1, a model's potential to provide a good smoothing estimate is generally affected by the smoothing method's ability to use recent data while keeping its variance small. If we compare the average age of data used from SES and MSES at the same α level, it is obvious that MSES is always younger than SES method at the same smoothing constant $(AA_{MSES} < AA_{SES})$, since

$$\frac{n-m}{m+1} < \frac{1 - \frac{m}{n}}{\frac{m}{n}} \ \text{ or } \ \frac{1}{m+1} < \frac{1}{m}.$$

When the same smoothing constant is used for both models, since MSES always has a smaller AA, its V value will be greater than that of SES as shown by Yager (2008).

Another way of defining a SES system that is equivalent to MSES is using a smoothing constant that results in equal average ages for both models. In other words, the smoothing constant for SES should satisfy:

$$\frac{1-\alpha}{\alpha} = \frac{n-m}{m+1},$$

so $\alpha = (m+1)/(n+1)$.

The V values of the two approaches when they have equal average ages, can be visualized as in Figure 3 below. This Figure was produced for $n = 15$ for demonstration, changing the value of m and letting $\alpha = \frac{m+1}{n+1}$. It can be seen from Figure 3 that MSES attains a smaller variance at the same smoothing constant. The variance of MSES stays
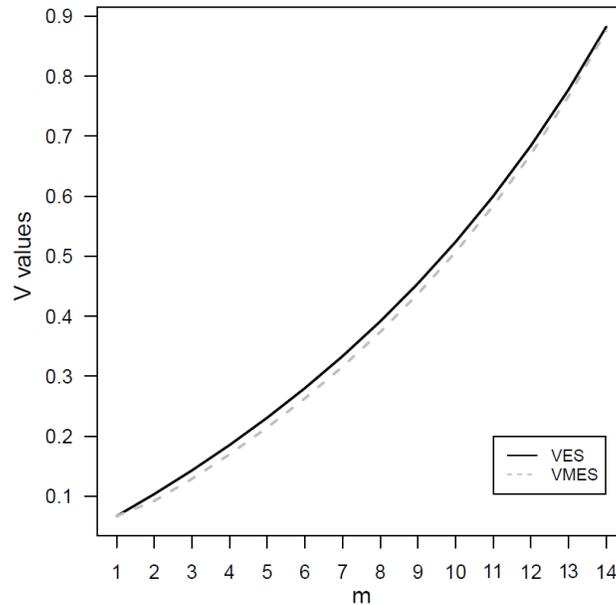
Figure 3. V values of MSES and SES for $n = 15$, $m \in \{1, 2, \ldots, 14\}$ and $\alpha = \frac{m+1}{n+1}$.

smaller than that of SES at all parameter values. This makes MSES more *flexible* than SES as Yager (2008) defines it.

Other major advantages of SES method are its relatively good short-term accuracy, simplicity and low cost. The process is easily implemented into computer, it does not require large amounts of historical data and new forecasts are easy to obtain. It is obvious that MSES satisfies all these desirable properties besides being even computationally simpler than SES.

To summarize, MSES attaches more weight to the recent observations than SES does but less weight to older observations at the same α level. It is obvious that for any fixed m value, the sharp distinction between the two methods will diminish when the number of observations is increased. It is clear that MSES is more adaptive than SES when there is level-shift in the data since it puts more weight on the most recent observations at the same smoothing constant. This helps MSES adapt faster to the data. The difference between the two approaches is more significant when α and n are small since under these conditions SES performs worse because it assigns even more weight to the initial value.

A final comparison can be made on the issue of finding an initial value for both models.Several solutions are suggested in the literature for SES (Brown, 1962; Montgomery and Johnson, 1976; Makridakis and Wheelwright, 1978; Bowerman and O'Connell, 1979). MSES, on the other hand, does not need an initial value. In other words, when we find the optimum m, the initial value is chosen simultaneously.

4. M-competition

To illustrate the proposed method and compare it to SES we applied both models to the first data set "YAF2" from the 1001 series of the M-competition data (Makridakis

Table 1. Weights assigned to observations by various MSES and SES models for YAF2 data

t	$m = 1$ and $\alpha = 1/22$			$m = 2$ and $\alpha = 2/22$			$m = 3$ and $\alpha = 3/22$		
	MSES	SES	Diff	MSES	SES	Diff	MSES	SES	Diff
22	0.045	0.045	0.000	0.091	0.091	0.000	0.136	0.136	0.000
21	0.045	0.043	0.002	0.087	0.083	0.004	0.123	0.118	0.005
20	0.045	0.041	0.004	0.082	0.075	0.007	0.111	0.102	0.009
19	0.045	0.040	0.005	0.078	0.068	0.010	0.099	0.088	0.011
18	0.045	0.038	0.007	0.074	0.062	0.012	0.088	0.076	0.012
17	0.045	0.036	0.009	0.069	0.056	0.013	0.078	0.066	0.012
16	0.045	0.034	0.011	0.065	0.051	0.014	0.068	0.057	0.011
15	0.045	0.033	0.012	0.061	0.047	0.014	0.059	0.049	0.010
14	0.045	0.031	0.014	0.056	0.042	0.014	0.051	0.042	0.009
13	0.045	0.030	0.015	0.052	0.039	0.013	0.043	0.036	0.007
12	0.045	0.029	0.016	0.048	0.035	0.013	0.036	0.031	0.005
11	0.045	0.027	0.018	0.043	0.032	0.011	0.029	0.027	0.002
10	0.045	0.026	0.019	0.039	0.029	0.010	0.023	0.023	0.000
9	0.045	0.025	0.020	0.035	0.026	0.009	0.018	0.020	-0.002
8	0.045	0.024	0.021	0.030	0.024	0.006	0.014	0.018	-0.004
7	0.045	0.023	0.022	0.026	0.022	0.004	0.010	0.015	-0.005
6	0.045	0.022	0.023	0.022	0.020	0.002	0.006	0.013	-0.007
5	0.045	0.021	0.024	0.017	0.018	-0.001	0.004	0.011	-0.007
4	0.045	0.020	0.025	0.013	0.016	-0.003	0.002	0.010	-0.008
3	0.045	0.019	0.026	0.009	0.015	-0.006	0.000	0.008	-0.008
2	0.045	0.018	0.027	0.000	0.014	-0.014	0.000	0.007	-0.007
1	0.000	0.017	-0.017	0.000	0.012	-0.012	0.000	0.006	-0.006
Weight of initial	0.045	0.359	-0.314	0.004	0.123	-0.119	0.001	0.040	-0.039
Average Age (AA)	10.500	21.000	-10.500	6.667	10.000	-3.333	4.750	6.333	-1.583

et al., 1982). The data set is measured yearly, non-seasonal, consists of 22 observations and the number of required forecasts is 6.

We applied MSES with parameters $m \in \{1, 2, 3\}$ and the equivalent SES models with $\alpha = \frac{m}{n}$ which correspond to $\alpha \in \{1/22, 2/22, 3/22\}$. In Table 1 we provide the weights assigned to observations by these six models side-by-side.

The first column in Table 1 is the time index and the following three columns are the weights obtained from MSES with $m = 1$ and from SES with $\alpha = 1/22$ and their differences calculated by subtracting the weights of the SES model from those of the MSES model. The following three columns are similar with $m = 2$ and $\alpha = 2/22$ and the final three columns are calculated similarly with $m = 3$ and $\alpha = 3/22$.

It can be seen from the table that MSES assigns more weight to fresher data points while assigning less weight to older data points. This can also be seen from the row "AA" of the table which shows the average ages of the two models under different parameter settings. Since MSES utilizes fresher data, its average age is always smaller than that of SES at the same smoothing constant level. The difference in weights can be seen clearly in Figure 4. For $m = 1$ the MSES always assigns more weight to the observations regardless of the age of the data point and by doing so less weight is given to the initial value hence the data is utilized better. For $m = 2$ MSES assigns more weight to recent observations and less weight to distant observations. This pattern gets stronger as m is
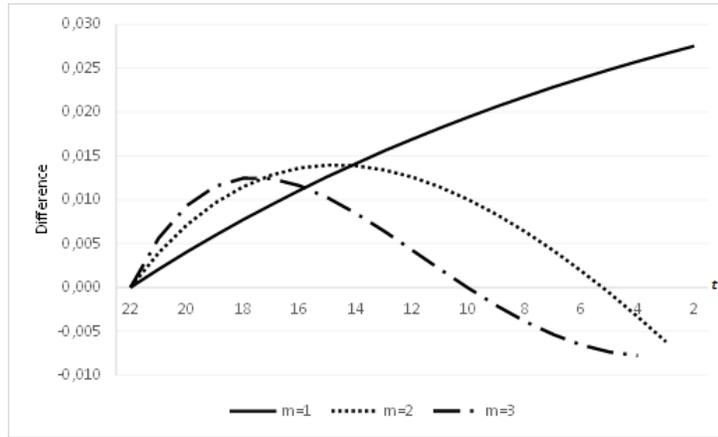
Figure 4. Difference between the weights assigned to x_t by MSES and SES $(w_{t,MSES} - w_{t,SES})$.
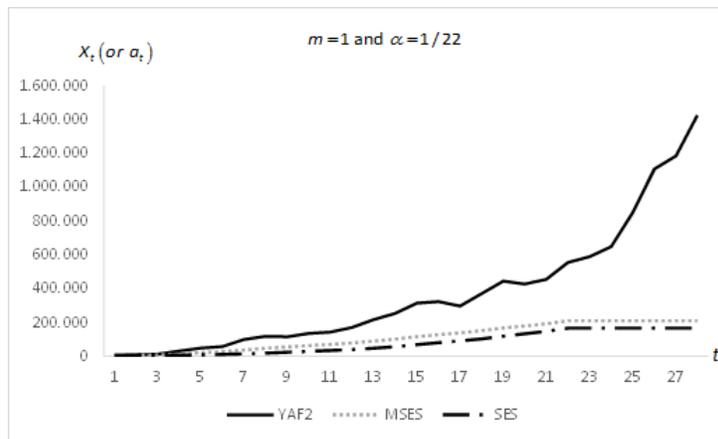

Figure 5. YAF2 along with the smoothed values from MSES($m = 1$) and SES($\alpha = 1/22$).

increased to 3 since the difference in weights for the recent observations get bigger and more of the distant observations get smaller weight.

The smoothed values are plotted with the YAF2 data set in Figures 5, 6 and 7 for the MSES models with $m \in \{1, 2, 3\}$ and the equivalent SES models with $\alpha \in \{1/22, 2/22, 3/22\}$ respectively. It can be seen from these figures that even though the original series clearly involves a trend component, MSES models are able to provide better approximations in all three cases.

In Table 2 we compare the in-sample performance of the models based on the mean absolute percentage error (MAPE) and compare their forecasting accuracies using MAPE to be consistent with the rest of the literature. It can be seen from the table that whenever the two models are at the same smoothing constant level MSES produces smaller error. Also, since the MAPE values for all six forecast horizons are smaller for MSES when the
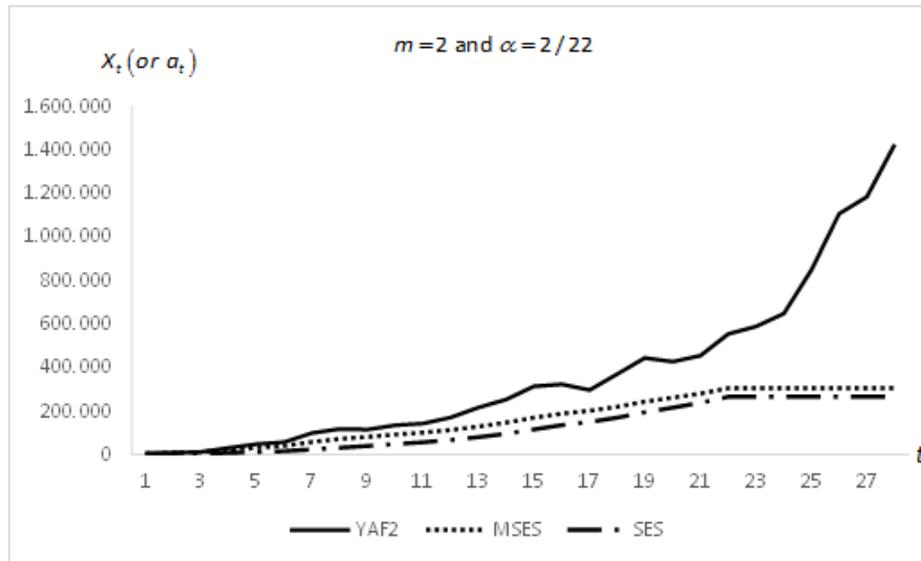
Figure 6. YAF2 along with the smoothed values from MSES($m = 2$) and SES($\alpha = 2/22$).
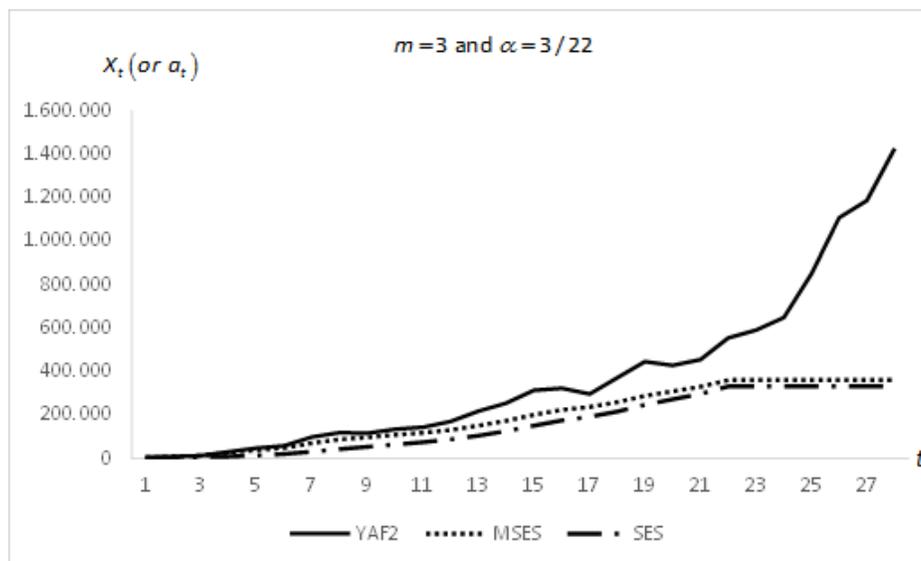

Figure 7. YAF2 along with the smoothed values from MSES($m = 3$) and SES($\alpha = 3/22$).

two models are at the same smoothing constant level, it can be said that MSES produces more accurate forecasts.

Table 2. MAPE for MSES and SES on YAF2 data

	$m = 1$ and $\alpha = 1/22$		$m = 2$ and $\alpha = 2/22$		$m = 3$ and $\alpha = 3/22$	
	MSES	SES	MSES	SES	MSES	SES
in-sample	0.63	0.79	0.47	0.68	0.38	0.60
out-sample forecast horizon						
1	0.65	0.72	0.48	0.55	0.39	0.44
2	0.68	0.74	0.53	0.59	0.44	0.49
3	0.75	0.81	0.64	0.69	0.58	0.61
4	0.81	0.85	0.72	0.76	0.67	0.70
5	0.82	0.86	0.74	0.78	0.70	0.72
6	0.85	0.88	0.79	0.81	0.75	0.77

Table 3. Average MAPE across different forecast horizons (1001 series) : MSES with $m \in \{1, 2, 3\}$ and equivalent SES with $\alpha = \frac{m}{n}$

	Forecasting horizons										Average of forecasting horizons					
Model	1	2	3	4	5	6	8	12	15	18	1-4	1-6	1-8	1-12	1-15	1-18
MSES(m=1)	37.65	37.61	39.76	36.16	41.28	37.42	37.27	44.48	69.03	73.22	37.80	38.31	38.56	39.90	42.45	46.31
SES	39.81	40.51	43.04	39.15	44.44	40.47	40.21	47.78	71.65	80.60	40.63	41.24	41.37	42.61	45.17	49.27
MSES(m=2)	27.40	28.23	30.14	27.65	31.59	30.01	28.86	33.18	50.06	53.20	28.36	29.17	29.39	30.25	31.98	34.63
SES	30.15	31.27	33.32	30.60	34.91	32.59	31.64	36.86	55.13	60.72	31.33	32.14	32.31	33.24	35.18	38.20
MSES(m=3)	20.99	22.68	24.48	22.82	25.98	25.87	24.10	26.59	37.99	41.21	22.74	23.80	24.00	24.56	25.76	27.60
SES	23.85	25.39	27.25	25.25	28.93	28.01	26.54	30.01	43.79	48.06	25.43	26.45	26.64	27.34	28.81	31.08

Any forecasting method may have some desirable features but its ultimate performance should be evaluated based on its capability of predicting future events accurately. To check the performance of the proposed method and compare it to SES, we applied both models to the 1001 series of the M-competition data under various settings. For all settings the data sets were first deseasonalized if necessary using the seasonal indices provided in the M-competition data file.

In the first setting, the two models were compared for pre-determined values of the parameter m of MSES where $m \in \{1, 2, 3\}$. The corresponding α values for SES were calculated depending on the sample size of each individual data set as $\alpha = \frac{m}{n}$. Then, at the same smoothing constant level forecasts were computed from both models up to 18 steps ahead (as determined in the M-competition data file) and both models' performance was evaluated with respect MAPE as in Table 3. It can be seen from the table that on average MSES models produced smaller errors for all forecast horizons compared to the equivalent SES models at the same smoothing constant level. When the errors are averaged for both the short term and long term forecast horizons this difference in the error terms is still significantly large.

In the second setting, the two models were compared for pre-determined values of the parameter α of SES where $\alpha \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. The corresponding m values for MSES were calculated depending on the sample size of each individual data set as $m = \alpha \times n$. Then, again at the same smoothing constant level forecasts were computed

Table 4. Average MAPE across different forecast horizons (1001 series) : SES with $\alpha \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ and equivalent MSES with $m = \alpha \times n$

Model	Forecasting horizons										Average of forecasting horizons					
	1	2	3	4	5	6	8	12	15	18	1-4	1-6	1-8	1-12	1-15	1-18
SES ($\alpha = 0.1$)	18.05	20.21	21.72	20.86	23.75	24.40	20.60	21.67	29.15	32.79	20.21	21.50	21.31	21.25	21.93	23.11
MSES	16.39	18.73	20.25	19.61	22.27	23.41	19.51	20.11	26.10	29.67	18.74	20.11	19.96	19.86	20.41	21.38
SES ($\alpha = 0.2$)	12.52	15.12	16.67	16.63	19.53	21.13	17.98	17.54	21.91	25.52	15.24	16.93	17.06	17.05	17.54	18.31
MSES	11.94	14.62	16.17	16.18	19.05	20.80	17.90	17.35	21.46	25.07	14.73	16.46	16.66	16.71	17.21	17.96
SES ($\alpha = 0.3$)	10.95	13.59	15.20	15.54	18.67	20.40	17.84	16.85	21.13	24.63	13.82	15.72	16.05	16.14	16.68	17.45
MSES	10.72	13.35	14.99	15.37	18.50	20.22	17.83	16.84	21.12	24.61	13.61	15.52	15.89	16.01	16.57	17.36
SES ($\alpha = 0.4$)	10.16	12.75	14.44	15.05	18.34	20.10	17.98	16.62	20.94	24.43	13.10	15.14	15.61	15.76	16.33	17.13
MSES	10.08	12.65	14.35	14.99	18.29	20.04	17.99	16.62	20.95	24.44	13.02	15.07	15.55	15.72	16.29	17.10
SES ($\alpha = 0.5$)	9.70	12.18	13.95	14.75	18.18	19.94	18.17	16.54	20.89	24.33	12.65	14.78	15.36	15.56	16.15	16.97
MSES	9.61	12.07	13.85	14.68	18.11	19.86	18.19	16.54	20.90	24.34	12.55	14.70	15.29	15.51	16.11	16.94

from both models up to 18 steps ahead (as determined in the M-competition data file) and both models' performance was evaluated with respect MAPE as in Table 4. The α levels used in this application are highly recommended levels in the literature. When we apply SES models with these recommended α levels and the equivalent MSES models, again the MSES produces smaller errors for all forecast horizons. When the MAPE values are averaged for different forecast horizons as on the right side of the table, the patterns can be seen more clearly. On average, both for short term and long term forecast horizons, MSES produces more reliable forecasts than SES.

5. Conclusion

In this paper, an alternative smoothing technique to SES was introduced by modifying the smoothing constant. By means of this modification, we saw that MSES has attractive features as a forecasting method and it outperforms SES at the same smoothing constant level when applied to the M-competition data which includes time series with non-seasonal, monthly and quarterly data. This happens due to the fact that MSES assigns more weight to the recent past compared to SES and therefore produces forecasts that are more in sync with the recent past. Also whenever α and n are small SES suffers the initialization problem since the initial value is given a big weight where MSES is immune to this phenomenon since for all parameter values the initial value's weight will be less than or equal to the most recent observation's weight.

The application of the proposed model to the M-competition data in Section 5 demonstrates that MSES performs well for both short term and long term forecasts. The proposed model is as simple as SES, does not need initialization, is faster to optimize and performs better.

This model can be extended further to allow for more complicated trending behaviors and to model the seasonality that may be present in the data sets. After this extension, the model can be applied to the M3-competition data (Makridakis and Hibon, 2000). Surely the model can benefit from data pre-processing techniques, outlier detection and other strategies to improve the forecasts. These are left as future work.

References

Bailey, W. N., 1935. Generalized hypergeometric series. University Press Cambridge.

Bowerman, B. L., O'Connell, R. T., 1979. Time series and forecasting. Duxbury Press.

Brown, R. G., 1959. Statistical forecasting for inventory control. McGraw/Hill.

Brown, R. G., 1962. Smoothing, forecasting and prediction of discrete time series. Prentice-Hall.

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

Gardner, E. S., 2006. Exponential smoothing: The state of the art—part ii. International journal of forecasting 22 (4), 637–666.

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

Goodwin, P., et al., 2010. The holt-winters approach to exponential smoothing: 50 years old and going strong. Foresight 19, 30–33.

Holt, C., 1957. Forecasting trends and seasonal by exponentially weighted averages. International Journal of Forecasting 20 (1), 5–10.

Hyndman, R. J., Athanasopoulos, G., 2014. Forecasting: principles and practice. OTexts.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., Grose, S., 2002. A state space framework for automatic forecasting using exponential smoothing methods. International Journal of Forecasting 18 (3), 439–454.

Johnson, N. L., Kotz, S., 1977. Urn models and their application: an approach to modern discrete probability theory. Vol. 77. Wiley New York.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., Winkler, R., 1982. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. Journal of forecasting 1 (2), 111–153.

Makridakis, S., Hibon, M., 2000. The m3-competition: results, conclusions and implications. International journal of forecasting 16 (4), 451–476.

Makridakis, S. G., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., Winkler, R., 1984. The forecasting accuracy of major time series methods. Wiley.

Makridakis, S. G., Wheelwright, S. C., 1978. Forecasting: Methods and Applications. Wiley.

Montgomery, D. C., Johnson, L. A., 1976. Forecasting and time series analysis. McGraw-Hill.

Pegels, C. C., 1969. On startup or learning curves: An expanded view. AIIE Transactions 1 (3), 216–222.

Taylor, J. W., 2003. Exponential smoothing with a damped multiplicative trend. International journal of Forecasting 19 (4), 715–725.

Winters, P. R., 1960. Forecasting sales by exponentially weighted moving averages. Management Science 6 (3), 324–342.

Yager, R. R., 2008. Time series smoothing and owa aggregation. IEEE Transactions on Fuzzy Systems 16 (4), 994–1007.