

The Comparison of Various Exponential Smoothing Models and Their Significance in Modern Time Series Forecasting

¹Mohammed Salisu Alfa, ²Ahmed Ibrahim and ²Mohammed Idris Umar

¹Department of Statistics, The Federal Polytechnic, Bida. Niger State. ²Department of Statistics, Nasarawa State University, Keffi. Nasarawa State.

Corresponding author: alfams001@gmail.com

Abstract

Exponential Smoothing model is one of the most used forecasting models in modern time series forecasting. This is a technique which is applied to time series data particularly with seasonality, used to yield smoothed data for presentation and or to make forecasts. Exponential smoothing models have been found to be amongst the most effective forecasting models mostly used for linear seasonal data series. It is made up of three types namely Simple exponential smoothing (Type I) which is used when the time series has no trend, Double exponential smoothing (Type II) which is used in handling a time series that displays a slowly changing linear trend and the third is winters' exponential smoothing method (Type III), used in predicting seasonal data. It is a powerful forecasting method which can be used as an alternative to the popular Box-Jenkins ARIMA methods. It also produces accurate and reliable forecasts which predict for the future. The forecasts show projected and actual demand which allows demand planning to be done effectively. The better the analysis of the main characteristics of the time series, the more accurate the exponential smoothing forecasts are likely to be. Exponential smoothing is easy to understand and apply; what is needed in this method are the forecast for the latest time period and relevant parameter constants such as α , β and γ which are associated with the level, trend and seasonality respectively.

Keywords: Exponential smoothing; Time series; Forecasting; Data; Seasonal.

Introduction

Forecasting comprises making projections about future performance on the basis of historical and current data. When the result of an action is of significance, but cannot be known in advance with precision, forecasting may reduce decision risk by supplying additional information about the possible outcome. In the work of Uddin et al., (2017), forecasting is most often part of a larger process of managing and planning and a necessary ingredient to provide accurate estimates of the future for the greater process. All the forecasting methods are categorized under the type time series univariate methods, once data have been captured for the time series to be used for forecast, the next step for the analyst is to select a model for forecasting. Various statistical and graphic techniques may be at the disposal of the analyst to be used in the selection process. The best place to commence with any time series forecasting analysis is to graph the sequence plots of the time series to be forecasted. After selecting a model, the next step is its specification. The process of specifying a forecasting model involves choosing the variables to be involved,

selecting the form of the equation of relationship, and estimating the values of the parameters in that equation. Forecasting is part of a larger process of planning and management and also a necessary ingredient that provide accurate estimates of the future for the larger process (Rahman et al., 2016).

Generally, the performance evaluation criteria are widely used in evaluating the results of time series forecasting as described in various literature, (Woschnagg & Cipan, 2004) and (Dawson et al., 2007). Looking at these error measures as provided by (Hyndman, 2014) and (Shcherbakov & Brebels, 2013). After the model is specified, its performance characteristics is then verified or validated by comparison of its forecasts with historical data for the process it was designed to forecast. Error measures such as Mean square error (MSE), Root mean square (RMSE), Mean average error (MAE) and Mean absolute percentage error (MAPE) may be used for validating the model. The selection of an error measure has an important effect on the conclusions about which of a set of forecasting methods is most accurate. Time-series forecasting assumes that a time series is a combination of a pattern and some random error. The aim is to distinct the pattern from the error by understanding the pattern's trend, its long-term increase or decrease, and its seasonality, the change caused by seasonal factors such as fluctuations in use and demand. Numerous methods of time series forecasting are presented such as the Moving Averages method, Linear Regression with Time, Exponential Smoothing etc. This paper focuses on the Exponential Smoothing technique as applied to time series.

Exponential smoothing is a time series forecasting method for univariate data that can be extended to support data with a systematic trend or seasonal component. It is a powerful forecasting method that can be used as an alternative to the popular Box-Jenkins ARIMA family. Exponential smoothing is a procedure for continually revising a forecast in the light of more recent experience. The method according to Kalekar & Bernard, (2004) produces forecasts that are weighted averages of past observations where the weights of older observations exponentially decrease. In other words, recent observations are given relatively more weight in forecasting than the older observations.

Exponential smoothing is a recursive time series procedure whereby forecasts are updated for each new incoming data value. The technique has been widely used in forecasting, particularly in business and inventory modelling. Exponential smoothing produces accurate forecasts (Makatjane & Moroke, 2016). Forecasts produced using this method is accurate and reliable then predict for the next period. The forecast shows projected demand and actual demand. This allows demand planning to be done effectively, therefore resulting in accurate inventory levels. Smoothing is a technique applied to time series to remove the fine-grained variation between time steps. The hope of smoothing is to remove noise and better expose the signal of the underlying causal processes. The smoothing techniques are the members of time series forecasting methods or algorithms, which use the weighted average of a past observation to predict the future values or forecast the new value. These techniques are well suited for time-series data having fewer deviations with time.

Exponential smoothing is the most widely used class of techniques for smoothing discrete time series in order to forecast the immediate future. The idea of exponential smoothing is to smooth the original series the way the moving average does and to use the smoothed series in forecasting future values of the variable of interest. In exponential smoothing, however, we want to allow the more recent values of the series to have greater influence on the forecast of future values than the more distant observations.

In the view of Kalinova, (2015), Exponential smoothing is the most widely used class of measures for smoothing discrete time series data in order to forecast the immediate future. The idea of exponential smoothing is to smooth the original series the way the moving average does and to use the smoothed series in forecasting future values of the variable of interest. In exponential smoothing, however, we want to allow the more recent values of the series to have greater influence on the forecast of future values than the more distant observations. Exponential smoothing is a simple and logical approach to forecasting, whereby the forecast is constructed from an exponentially weighted average of past observations. The largest weight is given to the present observation, less weight to the immediately preceding observation, even less weight to the observation before that, and so on.

Literature Review

Exponential smoothing was first suggested in the statistical literature without reference to any previous work by Robert Goodell Brown in (1956) and then expanded by Charles C. Holt in 1957. Exponential smoothing is a broadly accurate principle for smoothing time series data using the exponential window function. The controlling input of the exponential smoothing calculation is defined as the smoothing factor or the smoothing constant.

Exponential Smoothing Forecasting is generally used to make short term forecasts, but longer-term forecasts using this technique can be quite unreliable. More recent observations given larger weights by exponential smoothing methods, and the weights decrease exponentially as the observations become more distant. When the parameters describing the time series are changing slowly over time, then these methods are most effective. Forecasting involves the estimation of values we don't know by using values that we know there are many forecasting methods, and exponential smoothing is just one of them. Exponential smoothing is a technique used to detect significant changes in data by considering the most recent data. Also known as averaging, this method is used in making short-term forecasts (Rahman et al., 2016).

Given that there are many other ways to make forecasts, what makes exponential smoothing better in certain cases compared to others? Also, what makes it not ideal for certain scenarios?

The formulation of exponential smoothing forecasting methods arose in the 1950's from the original work of Brown (1959) and Holt (1962) who were working on creating forecasting models for inventory control systems. The growth and practice of the exponential forecasting methods dates back from the works of R. G. Brown in the 1940's the outcomes

of which were published in 1959. These were further developed and expanded by C. C. Holt in 1957 and Peter Winters in 1960. In 1960s Pegles (1969) developed the first taxonomy for the classification of the time series exponential smoothing forecasting methods. In the 1980's Gardner (1985; 1987) presented some interesting techniques aimed at smoothing of the error residuals in the achieved forecasts. Gardner (1985) also further expanded the opportunities for classifying the exponential smoothing forecasting methods.

In the work of Rahman et al., (2016) on revenue forecasting using Holt–Winters Exponential Smoothing with additive and multiplicative models concluded that the additive model was superior than the multiplicative model. The Holt–Winters method is a statistical forecasting method in time series univariate methods. The forecasting engages make projection about future performance on the basis of past and recent data (Uddin et al., 2017). Exponential Smoothing, which is almost like moving average in time series, is another kind of data analysis over time (Rangamati & Ahmed, 2018).

Forecasting involve the estimation of values we don't know by using values that we know. There are numerous forecasting methods, and exponential smoothing is just one of them. Exponential smoothing is a technique used to detect significant changes in data by considering the most recent data. Also known as averaging, this method is used in making short-term forecasts. Supposed that there are many other ways to make forecasts, what makes exponential smoothing better in certain cases compared to others? Also, what makes it not ideal for certain situations? This will lead us to discuss the advantages and disadvantages of Exponential Smoothing as an attempt to answer the two questions.

In answering the first part, it is easy to learn and apply since only three pieces of data are needed for exponential smoothing methods. One, it needs the forecast for the most recent time period. Two, it needs the actual value for that time period and three, it needs the value of the smoothing constant, a weighting factor that reflects the weight given to the most recent data values. Secondly, it produces accurate forecasts. An exponential smoothing method produces a forecast for one period ahead. Using the trend projection technique, forecasts for more periods ahead can then be generated. The forecast is considered accurate as it accounts for the difference between actual projections and what actually occurred. Thirdly, it gives more significance to recent observations. Observed data is the sum of two or more components, one being the random error which is the difference between the observed value and the true value. In a smoothing technique, the random variation is neglected. In that regard, it is much easier to see the underlying phenomenon. With respect to the second part, it produces forecasts that lag behind the actual trend. The lag is a side effect of the smoothing process. There's a purpose for this method having "smoothing" in its name this is because it disregards the ups and downs associated with random variation. As such, seeing this on a graph shows you a smoother line or curve. However, ignoring the random variation also allows one to see the underlying phenomenon, which helps when presenting data and making a forecast of future values. Secondly, it cannot handle trends well. Exponential smoothing is best used for forecasts

that are short-term and in the absence of seasonal or cyclical variations (Supriatna et al., 2017). As a result, forecasts are not accurate when data with cyclical or seasonal variations are present. As such, this kind of averaging would not work as expected if there is a trend in the series. Methods like this are only accurate when a reasonable amount of continuity between the past and future can be anticipated. As such, it is best suitable for short-term forecasting as it assumes future patterns and trends will look like current patterns and trends. While this type of assumption may sound realistic in the short term, it creates problems as the forecast goes. Nonetheless, there are variations of exponential smoothing that can handle trend patterns which is the Holt's method that can calculate strong trend patterns while winter's method can cover a strong trend and seasonal pattern variations.

Methodology

There are various steps to explore this important model which can

Step 1: Determine whether the model fits your data. Examine the smoothing plot to determine whether your model fits your data

Step 2: Compare the fit of your model to other models

Step 3: Determine whether the forecasts are accurate

Smoothing data removes random variation and shows trends and cyclic components. Inherent in the collection of data used over time is some form of random variation. There exist methods for reduction of terminating the effect due to random variation. An often-used technique in industry is usually the "smoothing".

Usually in most of the analysis carried out the ALPHA (α) is the smoothing parameter that defines the weighting and should be greater than 0 and less than 1. α equal 0 sets the current smoothed point to the previous smoothed value and α equal 1 sets the current smoothed point to the current point which specifies the coefficient for the level smoothing.

Beta (β) specifies the coefficient for the trend smoothing while Gamma (γ) specifies the coefficient for the seasonal smoothing. There is also a parameter for the type of seasonality:

Additive seasonality, where each season changes by a constant number. β is the numeric value between 0 and 1 which controls the trend calculation.

Linear exponential smoothing (LES) uses a moving average to create a forecast from a time series. The forecast is first created for the similar period as the existing data and then into the future where there is no data. The difference between the existing data and the forecast data can be checked for accuracies. Time series forecasting is a technique for predicting future events by analyzing past trends, based on the assumption that future trends will hold similar to historical trends (Smyl, 2019). Forecasting involves using models fit on historical data to predict future values. A time series is white noise if the variables are independent and identically distributed with a mean of zero. This means that all variables have the same variance δ^2 and each value has a zero correlation with all other values in the series. It should be noted that white noise has zero mean, constant variance, and is uncorrelated in time. As

its name suggests, white noise has a power spectrum which is uniformly spread across all allowable frequencies.

Smoothing constants

It is well known that the smoothing constants determine how fast the weights of the series declines. The values may be chosen either subjectively (individually) or objectively (empirically). Values of a smoothing constant close to one place almost all weight on the most recent observations while values of a smoothing constant close zero allow the distant past observations to have a large influence in which α is associated with the level of the series, β is associated with the trend, and γ is associated with the seasonality factors.

While selecting the smoothing constant subjectively, one uses his own experience with this and similar series. Similarly, specifying the smoothing constant by oneself allows him to tune the forecast to one's own beliefs about the future of the series. If you believe that the instrument that generates the series has recently gone through some fundamental changes, a smoothing constant value of 0.9 is used and this will cause distant observations to be overlooked. If, however, one thinks the series is fairly stable and only going through random fluctuations, a value of 0.1 is used.

To select the value of the smoothing constants objectively, one searches for values that are best in some sense. Our program searches for that values that minimize the size of the combined forecast errors of the currently available series. Three or more of the five methods of summarizing the amount of error in the forecasts such as the mean square error (MSE), the root mean square error (RMSE), the mean absolute error (MAE), the mean absolute percent error (MAPE) and the correlation coefficient (R) which measures the degree of relationship between these measures are applied. The forecast error is the difference between the forecast of the current period made at the last period and the value of the series at the current period. This is given as

$$e_t = X_t - F_{t-1} \quad (1.0)$$

$$\text{Where } e_t = X_t - \hat{X}_t$$

Using this formulation, the three error-size criterion can be defined as follows:-

$$MSE = \frac{1}{n} \sum e_t^2$$

$$MAE = \frac{1}{n} |e_t|$$

$$MAPE = \frac{100}{n} \sum \left| \frac{e_t}{X_t} \right|$$

Or

$$MSE = \sum_{t=1}^n [X_t - \hat{X}_t]^2 \quad (2.0)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |X_t - \hat{X}_t| \quad (3.0)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n |X_t - \hat{X}_t| * 100 \quad (4.0)$$

To find the value of the smoothing constants objectively, we then select one of these measure and search for those values of α and β that minimizes this function. The program is used to conduct a search for the appropriate values using an efficient algorithm.

Simple exponential smoothing

This model is used when the time series data has no trend and no seasonality. The simple exponential smoothing model is given by the following model equations:

$$y_t = \mu_t + a_t \quad (5.0)$$

The smoothing equation is given by:

$$L_t = \alpha y_t + (1 - \alpha) L_{t-1} \quad (6.0)$$

The h-step-ahead prediction equation is:

$$\hat{y}_{t+h} = L_t, \quad h = 1, 2, \dots \quad (7.0)$$

That is, one forecast y h-steps ahead by using the last available estimated (smoothed) level state L_t .

The simple exponential smoothing is mostly used for univariate time series data that do not have a trend or seasonal cycle. It is also referred to as single exponential smoothing. It's the simplest form of exponential smoothing and a great place to start. Simple exponential smoothing is used to estimates only the level component. Because it models one component, it uses only one weighting parameter, alpha (α). This value determines the degree of smoothing by changing how quickly the level component adjusts to the most recent data. Alpha values ranges from 0 to 1, inclusive. Lower values produce smoother fitted lines because they give more weight to past observations, averaging out fluctuations over time. Higher values create a more pointed line because they weigh current data more highly, which reduces the degree of averaging by the older data. In its reaction to changing conditions sounds like a positive attribute, setting an excessively high alpha smoothing constant can produce erratic forecasts because the model responds to random fluctuations (noise). Equally, an alpha that is too low causes a lag (a delay or gap) between changing conditions and when they impact the forecasts. A weight of 1 causes the most recent observation to have all the weight, while all previous observations have no impact on the model. Consequently, the forecast values for $\alpha = 1$ are simply the current value. In terms of forecasting, simple exponential smoothing produces a constant set of values. All forecasts

equal the last value of the level component. Accordingly, these forecasts are appropriate only when the time series data have no trend or seasonality.

Double (Brown) exponential smoothing

Double exponential smoothing is also referred to as Holt exponential smoothing method (Supriatna et al., 2017). This model should be used when the time series data has trend but no seasonality

The Double exponential smoothing model is given by the model equation:

$$y_t = \mu_t + \beta_t t + a_t \quad (8.0)$$

The smoothing equations are:

$$L_t = \alpha y_t + (1 - \alpha) L_{t-1} \quad (9.0)$$

$$T_t = \alpha (L_t - L_{t-1}) + (1 - \alpha) T_{t-1} \quad (10.0)$$

The h-step-ahead prediction equation is

$$\hat{y}_{t+h} = L_t + ((h-1) + 1/\alpha) T_t, \quad h = 1, 2, \dots \quad (11.0)$$

That is, you forecast y h-steps ahead by taking the last available estimated level state and multiplying the last available trend (slope), T_t , by $((h-1) + 1/\alpha)$

Double exponential smoothing can model trend components and level components for univariate time series data. Trends are slopes in the data. This method models dynamic gradients because it updates the trend component for each observation. To model trends, Double Exponential Smoothing includes an additional parameter, beta (β). As with alpha, beta can be between 0 and 1, inclusive. Higher values place more weight on recent observations, allowing the trend component to react more quickly to changes in the trend. Forecasts for this method change at a constant rate which is equal to the final value of the trend component. A popular extension for this method adds a dampening (reducing) component to the forecasts, causing the forecasts to level out over time to avoid excessively optimistic long-term forecasts.

Triple Exponential Smoothing

This is also referred to as Holt-Winter exponential smoothing or Seasonal exponential smoothing. The model is used when the time series data has no trend but has seasonality

The seasonal exponential smoothing model is given by the model equation:

$$y_t = \mu_t + S_{t,p} + a_t \quad (12.0)$$

The smoothing equations are:

$$L_t = \alpha (y_t - S_{t-p}) + (1 - \alpha) L_{t-1} \quad (13.0)$$

$$S_t = \delta (y_t - L_t) + (1 - \delta) S_{t-p} \quad (14.0)$$

The h-step-ahead prediction equation is

$$\hat{y}_{t+h} = L_t + S_{t-p+h}, \quad h = 1, 2, \dots \quad (15.0)$$

Meaning that, you forecast y h -steps ahead by taking the last available estimated level state and then add the last available smoothed seasonal factor, S_{t-p+h} , that matches the month of the forecast horizon.

Triple exponential smoothing can model seasonality, trend, and level components for univariate time series data. Seasonal cycles are patterns in the data that occur over a standard number of observations. This method adds gamma (γ) parameter to account for the seasonal component. For this method, one must specify the period for the seasonal cycle. In triple exponential smoothing, seasonality can be multiplicative or additive. Multiplicative seasonality has a pattern where the magnitude increases when the data increase. Additive seasonality reflects a seasonal pattern that has a constant scale even as the observations change.

Winters method-additive version with seasonality

This model should be used when the time series data have both the trend and the seasonality.

The seasonal exponential smoothing model is given by the model equation:

$$y_t = \mu_t + \beta_t t + S_{t,p} + a_t \quad (16.0)$$

The smoothing equations are:

$$L_t = \alpha(y_t - S_{t-p}) + (1-\alpha)(L_{t-1} + T_{t-1}) \quad (17.0)$$

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1} \quad (18.0)$$

$$S_t = \delta(y_t - L_t) + (1-\delta)S_{t-p} \quad (19.0)$$

The h -step-ahead prediction equation is:

$$\hat{y}_{t+h} = L_t + hT_t + S_{t-p+h}, \quad h = 1, 2, \dots \quad (20.0)$$

That is, you forecast y h -steps ahead by using the last available estimated level state and incrementing it h times using the last available trend (slope) T_t while at the same time adding the last available smoothed seasonal factor, S_{t-p+h} , that matches the month of the forecast horizon.

A stationary time series is one whose properties do not depend on the time at which the series is observed. Thus, time series with trends, or with seasonality, are not stationary — the trend and seasonality will affect the value of the time series at different times. A Holt-Winters model is defined by its three order parameters, alpha, beta, and gamma. Alpha specifies the coefficient for the level smoothing. Beta specifies the coefficient for the trend smoothing while Gamma specifies the coefficient for the seasonal smoothing. Double exponential smoothing employs a level component and a trend component at each period.

Double exponential smoothing uses two weights, (also called smoothing parameters), to update the components at each period. Methods designated generally as exponential smoothing are remarkably well known in down to earth time series smoothing and forecasting (Makatjane & Moroke, 2016).

Results and Discussion

Forecasting involve the estimation of values we don't know by using values that we know. There are numerous forecasting methods, and exponential smoothing is just one of them (Sulandari & Utami, 2016). Exponential smoothing is a technique used to detect significant changes in data by considering the most recent data. It is also known as averaging. This method is used in making short-term forecasts. Supposed that there are many other ways to make forecasts, what makes exponential smoothing better in certain cases compared to others? Also, what makes it not ideal for certain situations? This will lead us to discuss the advantages and disadvantages of Exponential Smoothing as an attempt to answer the two questions.

In answering the first part, it is easy to learn and apply since only three pieces of data are needed for exponential smoothing methods. One, it needs the forecast for the most recent time period. Two, it needs the actual value for that time period and three, it needs the value of the smoothing constant, a weighting factor that reflects the weight given to the most recent data values. Secondly, it produces accurate forecasts. An exponential smoothing method produces a forecast for one period ahead. Using the trend projection technique, forecasts for more periods ahead can then be generated. The forecast is considered accurate as it accounts for the difference between actual projections and what actually occurred. Thirdly, it gives more significance to recent observations. Observed data is the sum of two or more components, one being the random error which is the difference between the observed value and the true value. In a smoothing technique, the random variation is neglected. In that regard, it is much easier to see the underlying phenomenon. With respect to the second part, it produces forecasts that lag behind the actual trend. The lag is a side effect of the smoothing process. There's a purpose for this method having "smoothing" in its name this is because it disregards the ups and downs associated with random variation. As such, seeing this on a graph shows you a smoother line or curve. However, ignoring the random variation also allows one to see the underlying phenomenon which helps when presenting data and making a forecast of future values. Exponential smoothing is best used for forecasts that are short-term and in the absence of seasonal or cyclical variations. As a result, forecasts are not accurate when data with cyclical or seasonal variations are present. As such, this kind of averaging would not work as expected if there is a trend in the series. Methods like this are only accurate when a reasonable amount of continuity between the past and future can be anticipated. As such, it is best suitable for short-term forecasting as it assumes future patterns and trends will look like current patterns and trends. While this type of assumption may sound realistic in the short term, it