DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

FINANCIAL DATA ANALYSIS BY EXPONENTIAL SMOOTHING AND ATA METHOD

by
Selma ŞALK

November, 2019
İZMİR

FINANCIAL DATA ANALYSIS BY EXPONENTIAL SMOOTHING AND ATA METHOD

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Master of Science in Statistics, Statistics Program

by
Selma ŞALK

November, 2019
İZMİR

M. Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"FINANCIAL DATA ANALYSIS BY EXPONENTIAL SMOOTHING AND ATA METHOD"** completed by **SELMA ŞALK** under supervision of **PROF. DR. GÜÇKAN YAPAR** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

....

Prof. Dr. Güçkan YAPAR

Supervisor

Doç. Dr. Sinan Saraçlı

....

Jury Member

Prof. Dr. Cenk Özler

....

Jury Member

Prof. Dr. Kadriye ERTEKİN

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

Special thanks are to my supervisor, my teacher for life, Prof. Dr. Güçkan YAPAR. He has always had a positive impact on my life since we met. His teaching, support, friendship always shed light on my way. I'm grateful that he supported me in this study and he is always with me in life. He will be one of the most special person for me forever.

Moreover, I want to thank to my family for their love, trust, encouragement and being with me all the time.

Selma ŞALK

FINANCIAL DATA ANALYSIS BY EXPONENTIAL SMOOTHING AND ATA METHOD

ABSTRACT

Time series occurs by collecting the data in a particular category in a given time period. Accurate analysis of financial data, which is a sort of time series, has a great importance for financial institutions to make predictions for the future. Exponential smoothing method is one of the most used method in time series analysis. Exponential smoothing methods have been used widely for many years due to their simplicity and success in prediction results. The success of the method has proved many times in the famous M-competitions. However, the selection of initial value and smoothing constant according to subjective choices for exponential smoothing method adversely affect the accuracy of this method. The ATA method, which is a new method developed as an alternative to the exponential smoothing method, eliminates these disadvantages of the exponential smoothing method. In this study, the M4 results of exponential smoothing method and ATA method will be compared, especially their performance in financial data will be evaluated.

Keywords: Time series, financial data, exponential smoothing, ATA method, M-competition

ÜSSEL DÜZLEŞTİRME VE ATA METODU İLE FİNANSAL VERİ ANALİZİ

ÖZ

Zaman serileri belirli bir kategorideki verilerin belli zaman diliminde elde edilmesiyle oluşur. Zaman serilerinin bir çeşidi olan finansal verilerin doğru analiz edilmesi, finansla ilgili kurumların geleceğe yönelik öngörü yapmaları için büyük önem arz etmektedir. Zaman serisi analizinde en çok kullanılan yöntemlerden biri üssel düzleştirme metodudur. Üssel düzleştirme metotları basit olması ve öngörü sonuçlarındaki başarısı ile uzun yıllardır yaygın bir şekilde kullanılmaktadır. Ünlü M-yarışmalarında da metodun başarısı defalarca kanıtlanmıştır. Ancak üssel düzleştirme metotları için başlangıç değeri ve düzleştirme katsayısı seçiminin sübjektif tercihlere göre yapılması bu metodun doğruluğunu olumsuz yönde etkilemektedir. Üssel düzleştirme yöntemine alternatif olarak geliştirilen yeni bir metot olan ATA metodu, üssel düzleştirme metodunun bu dezavantajlarını ortadan kaldırmaktadır. Bu çalışmada, üssel düzleştirme metodu ve ATA metodunun M4 yarışması sonuçları karşılaştırılacak, özellikle finansal verilerdeki performansları değerlendirilecektir.

Anahtar kelimeler: Zaman serileri, finansal veri, üssel düzleştirme, ATA metot, M-yarışması

CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER ONE

INTRODUCTION

Time series analysis principle concerned with collecting and analysing data about a specific field, interpreting and presenting the results based on these analyses and making prediction for future. The analysis helps us to make forecasts about the future, make more effective decisions and engage in more accurate actions.

The term ‘financial analysis’ contains both ‘analysis and interpretation’. Simplification of financial data according to methodical classification given in the financial statements explains the term analysis. Interpretation means describing the meaning and importance of the data. Analysis and interpretation are two complementary techniques. Analysis is useless without interpretation, and interpretation is difficult or even impossible without analysis (Analysis of financial statements, n.d.).

Financial data analysis is a process that aims to determine the best estimates of the future situation of the business and aims to estimate the operational results of the operation by benefiting from the current and past financial conditions. This analysis determines the strengths and weaknesses of the enterprises, and evaluates the performance of the firms in a certain time interval (time series analysis) and helps them to make estimations and decisions for the future (Analysis of financial statements, n.d.).

In order to understand and interpret the events in a particular area; politics, medicine, economy, finance etc., we analyse the data about the event with statistical techniques and methods. So we make an estimation about the future for the event and try to make the right decision. We can use forecasting methods for this.

There are many forecasting methods but Autoregressive Integrated Moving Averages (ARIMA) and Exponential Smoothing (ES) are the two dominant major forecasting techniques and the other methods are usually derived from them. Between these two, ES methods are applied more frequently due to their simplicity, robustness and accuracy as automatic forecasting procedures especially in the famous M-competitions (Makridakis & Hibon, 2000).

Autoregressive (AR) and Moving Average (MA) appeared in the early 1900's. After that, work of Box & Jenkins in 1970 integrated these techniques into one approach and finally created ARIMA (De Gooijer & Hyndman, 2006). The Box & Jenkins approach allowed for non-stationary time series trends to be modelled (Shumway & Stoffer, 2016). Non-stationary data can be made stationary by the process known as differencing. In some time series models there is a need to set for seasonality (Smith & Agrawal, 2015).

ES was proposed in the late 1950s (Brown, 1959; Holt, 1957; Winters, 1960) and has inspired the development most successful forecasting methods. Pegels (1969) provided a simple but useful classification of the trend and the seasonal patterns depending on whether they are additive (linear) or multiplicative (nonlinear). Muth (1960) was the first to propose a statistical basis for simple exponential smoothing (SES) by proving that it provided the optimal forecasts for a random walk plus noise. Further studies for ES were provided by Box & Jenkins (1970, 1976), Roberts (1982) and Abraham and Ledolter (1983, 1986), who showed that some linear ES forecasts arise as special cases of ARIMA models. But these results failed to reach any nonlinear ES methods. (De Gooijer & Hyndman, 2006).

Forecasts generated by using ES methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older. In other saying, the more recent observation is the higher weighted (Hyndman & Athanasopoulos, 2018). That means, ES methods give larger weights to more recent observations, and the weights decrease exponentially as the observations become more distant. These methods are most powerful when the parameters defining the time series are changing slowly over time (Wan, n.d.).

ES models are based on the assumption that time series have three main components: level, trend and seasonality. Hyndman et al.'s (2002) taxonomy, which is extended by Taylor in 2003, provides a helpful categorization in describing the various ES methods. Each method consists of one of the five types of trend: none, additive, damped additive, multiplicative or damped multiplicative; and one of the three types of seasonality: none, additive or multiplicative. So, there are 15 different ES models, the best known of which are:

1. Simple Exponential Smoothing (SES) with no trend and no seasonality,
2. Holt's linear model with additive trend and no seasonality,
3. Holt-Winter's additive model with additive trend and additive seasonality (De Gooijer & Hyndman, 2006).

In this thesis, we propose a new forecasting method, which is called ATA method, as an alternative for ES method. ES method has some deficiencies like initial value problem and smoothing parameter selection. ATA method eliminates the initialization problem and is more practical to optimize compared to its counterpart ES models (Yapar, 2016).

This thesis consists of four chapters. The chapter one is introduction. In the chapter two, some forecasting methods and ES methods are mentioned. In the third chapter, ATA method is described, ATA method and ES method are applied to M4 competition finanacial data set and the two forecasting methods are compared. Finally in the fourth chapter, the conclusions are interpreted and summarized.

1.1 Purpose of This Thesis

The main objective of this thesis is to introduce a new forecasting method as an alternative to ES. The proposed method, which is called ATA method, is developed from ES models by modifying the smoothing parameters. This thesis shows the good performance of ATA method better than ES method by application to M4 competition financial data set.

1.2 Research Scope and Objectives

In this thesis, M4 competition inventory data set is used for application of forecasting methods, which is retrieved from M4 competition web site. The forecasts are obtained by application of forecasting methods with R-software. You can obtain the R-package for ATA method from GitHub web site. After the application of forecasting methods to M4 competition dataset, the sMAPE, MASE and MAPE values are calculated for the accuracy of forecasts. Comparisons between methods are made over these sMAPE, MASE and MAPE values. At the end of the thesis, in section 3.3.1, M4 financial data set is analysed by ES method and ATA method using R package of these methods and obtained forecast values. Then the sMAPE, MASE and MAPE values are calculated by excel and the average of the sMAPE, MASE and MAPE results are given in the tables for every frequency of data.

CHAPTER TWO

TIME SERIES FORECASTING METHODS

"If a man gives no thought about what is distant, he will find sorrow near at hand." says Confucius (Armstrong, 2001). Forecasting is important in many areas of life. Individuals wants to make prediction about their marriage, career and income. Companies increase their earnings according to their estimates about their new products, factories and financial conditions. Government agencies makes prediction about economy, environmental impacts and effects of proposed social programs. Unrealistic forecasts can lead to devastating consequences for them (Armstrong, 2001). Therefore; individuals, companies and governments try to make accurate predictions by using scientific techniques to make the right decisions about the future.

Forecasting techniques has improved over time. Time series forecasting has wide literature and many number of applications about forecasting. For companies, budgeting, production planning, inventory management, marketing, sales and distribution all depend on accurate time series forecasts. Exponential smoothing, Box & Jenkins ARIMA methods, Kalman filters, Census X11, regression methods are the traditional time series forecasting methods (Duncan, Gorr, & Szczypula, 2001). But ES, which was proposed in the late 1950's, and ARIMA, which was the work of Box & Jenkins in 1970, are the popular ones. In this section, we will examine basic time series' estimation methods.

2.1 Times Series and Financial Data

The term "time series" itself, denotes a data storing format, which consists of the two necessary components, one is time unit and the other one is the corresponding value assigned for the given time unit. In time series, values must denote the same meaning and correlate among the nearby values. It is the rule that at the same time there can be at most one value for each time unit (Ostashchuk, 2017).

In theory, there are two fundamental ways to record time series' data. One is that data are measured and recorded continuously along the time intervals. Electrical signals from sensors, earth shakings, various indicators from medicine, are giving us a continuous measurement of corresponding physical quantity. This kind of processes produces a continuous time series (Ostashchuk, 2017). There is an example for continuous time series from (Disaster and Emergency Management Presidency, Turkey) in Figure 2.1.

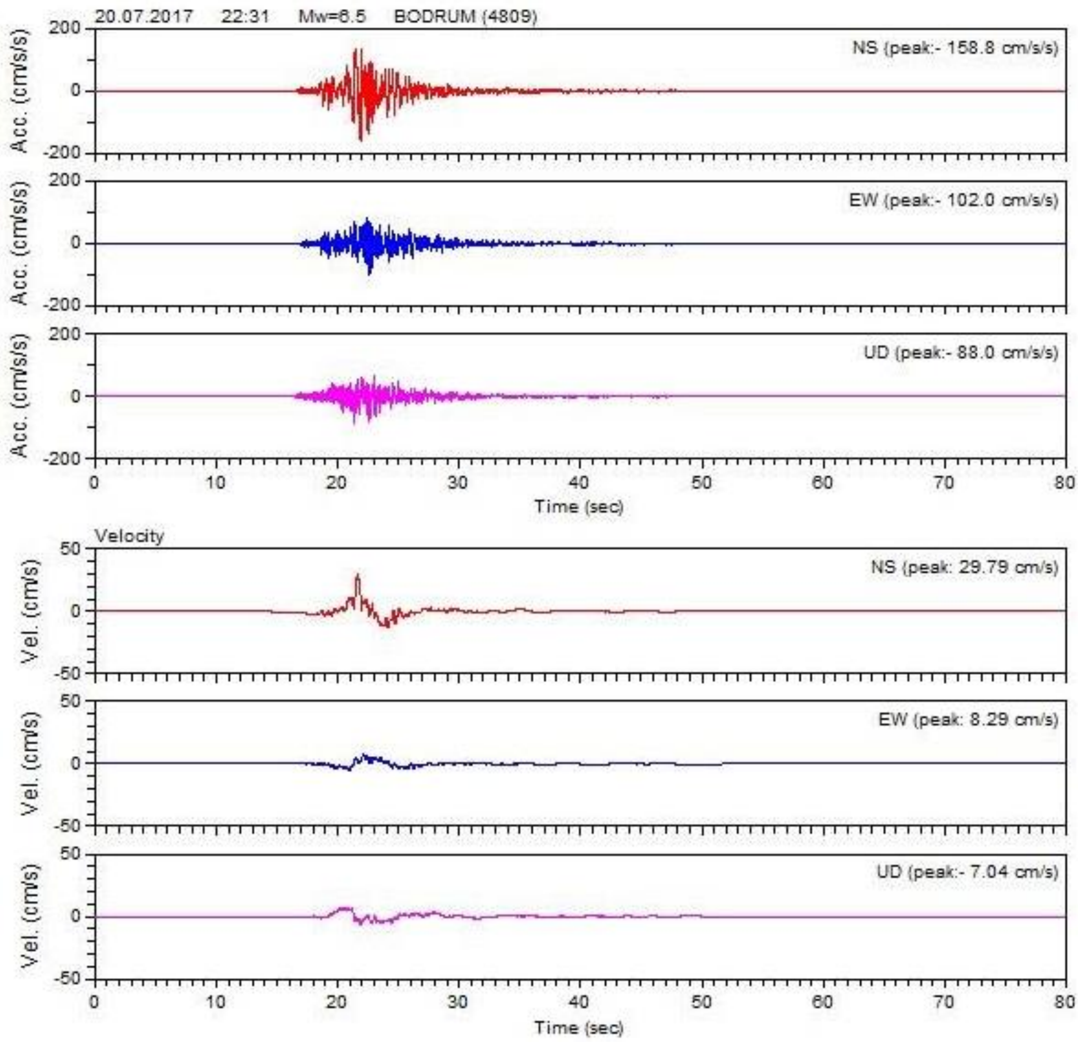


Figure 2.1 Acceleration and velocity of an earthquake in Bodrum (DEMP, 2017)

Figure 2.1 denotes the earthquake for 80 seconds in Bodrum on date 20.07.2017 at time 22:31. It gives us a data that measured continuously in 80 seconds. There is no cut in this 80 seconds so this is an example of continuous time series.

The second way is measuring the values just for the specific times, what may occur periodically or occasionally according to various conditions, but anyway, result will be a discrete set of values, formally called discrete time series. This is very common case and frequently observed in practice. Especially in economy sector, most of the indicators are measured periodically with the specific periods, therefore economic indicators represent an appropriate example of discrete time series. Hence, we can say that financial data is a discrete time series (Ostashchuk, 2017). There is an example in Figure 2.2, shows that Special Consumption Tax Income in year 2018 in Turkey:

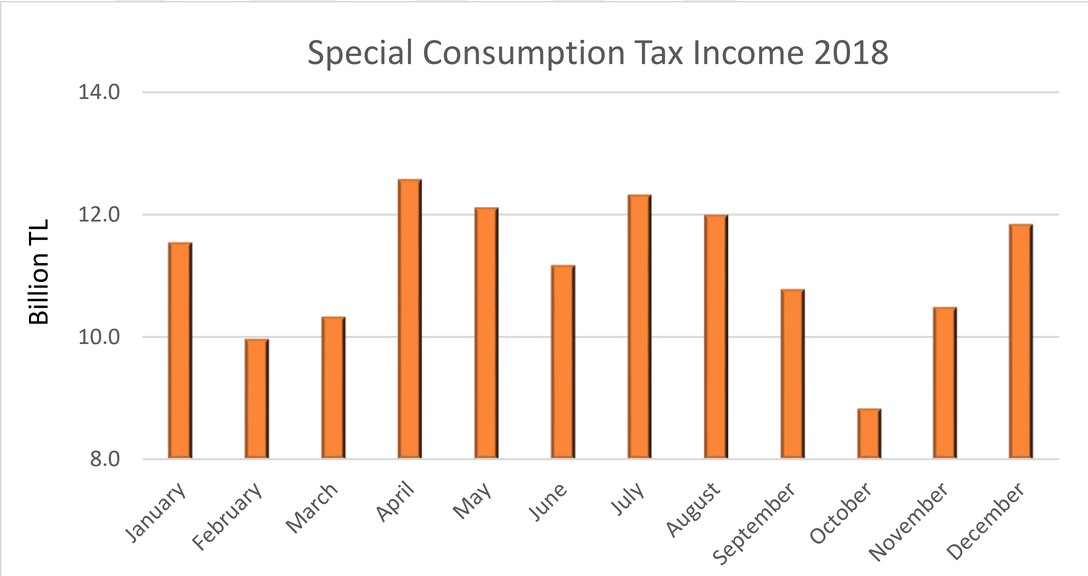


Figure 2.2 Special consumption tax income in Turkey 2018 (MTF, 2019)

In the market system, demand and supply of money and capital creates the financial market system. Bond market, stock market, commodity market and exchange market are four components of the financial market. Shares, currency, bonds, credits and commodities are the occupations of the financial market. The most important information is the price from financial market: price of currency, price of bond, price of share, price of commodity etc. Observation of the prices in certain time produces time series. This time series, which is based on prices, is an example of financial time series (Arlt & Arltova, 2001). Figure 2.3 denotes the consumer price index in Turkey in year 2018 according to Turkish Standardization Institute (TSI, 2019) :

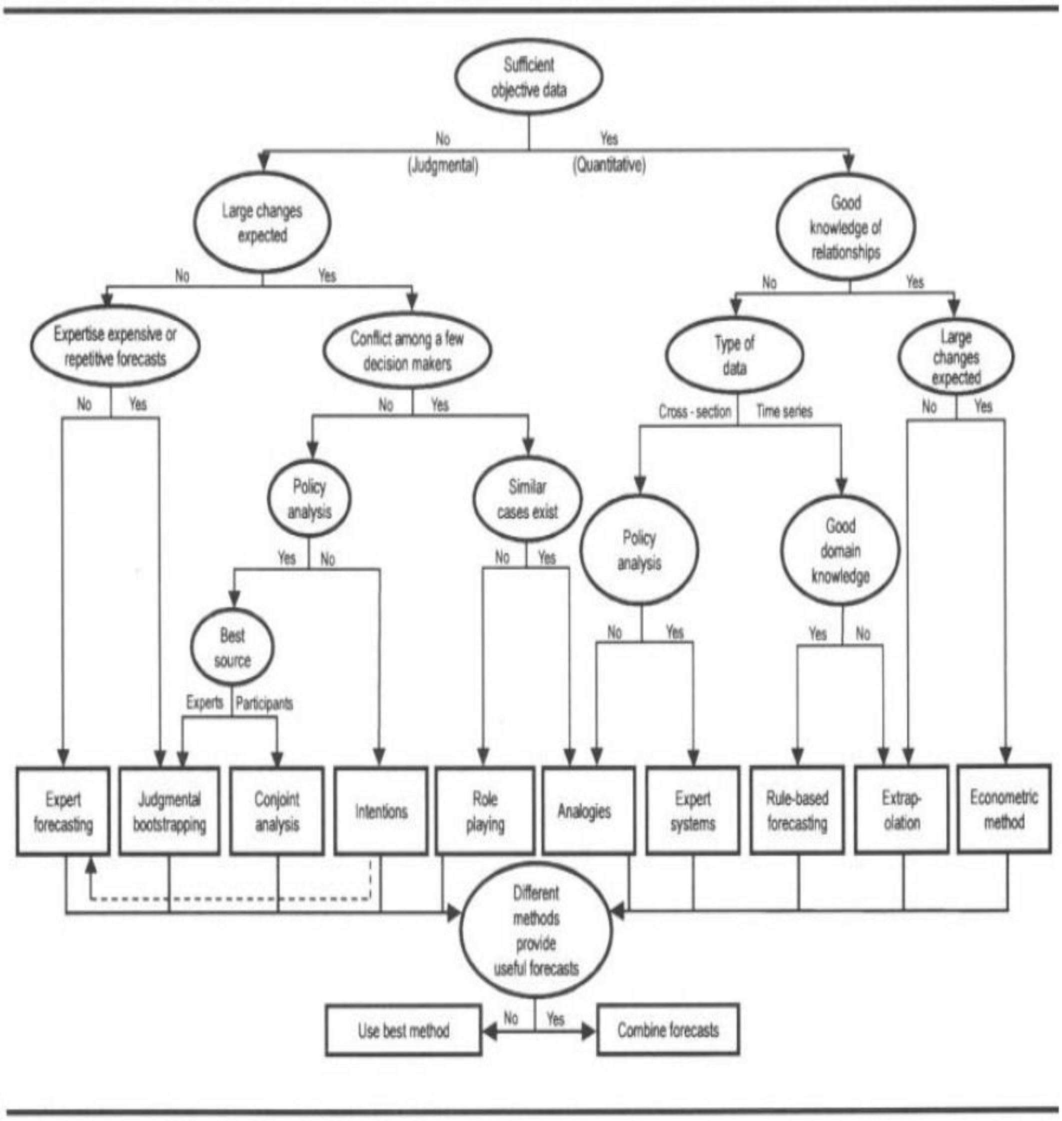
Figure 2.3 Consumer price index in Turkey 2018 (TSI, 2019)

Financial time series have some characteristic features and shapes that are created by financial markets when compared with other economic time series. A high frequency of individual values is a key feature of financial time series. This causes a great deal of vulnerability over time. Trend and cycle play an important role for this vulnerability, the seasonal part does not has an important role. There are typical properties of financial time series. Normality and linearity of log returns of the financial time series are the two main assumption of the financial time series (Arlt & Arltova, 2001).

2.2 Some Simple Forecasting Methods

Time series occur in many different areas when something is observed over time, for example industry and finance. We try to predict the future values of these time series with some methods. If the time series to be predicted is seasonal, the seasonal component can be purified from original data. After the removing, the resulting values will be “seasonally adjusted” or “deseasonalized” data. The main goal is to make predictions on future value using this time series. This is achieved by collecting some past observations (Taylan Selamlar, 2017). There is an example of selection tree for forecasting method:

Table 2.1 Selection tree for forecasting methods (Armstrong, 2001)



2.2.1 Average Method

Average method is a forecasting method that forecasting values are adequate to the average (or “mean”) of data of time series. If we denote the observations of the time series by $Y_1, Y_2, \dots, Y_T$, then we can write the forecasts as:

$$\hat{Y}_{T+h|T} = \bar{Y} = \frac{(Y_1 + Y_2 + \cdots + Y_T)}{T} \tag{2.1}$$

$\hat{Y}_{T+h|T}$ denotes the short-hand for the forecast of Y_{T+h} based on the data $Y_1, Y_2, \dots, Y_T$ (Hyndman & Athanasopoulos, 2018).

2.2.2 Naïve Method

In the naïve method, forecasting values are equal to the last observation of the time series. That is:

$$\hat{Y}_{T+h|T} = Y_T \tag{2.2}$$

For economic and financial time series, naïve method performs very well. Since a naïve prediction is optimal when data follows a random gait, these are also called random gait forecastings. (Hyndman & Athanasopoulos, 2018).

2.2.3 Seasonal Naïve Method

It is useful to use a similar method for seasonal data. For this estimate, we ensure that each estimate is equal to the last observed value in the same season of the year (for example the September value in the last year will be the September forecast of this year). The forecasting value for time $T + h$ can be written as:

$$\hat{Y}_{T+h|T} = Y_{T+h-m(k+1)} \tag{2.3}$$

where, m is the seasonal period, and k is the integer part of $(h-1)/m$. It is not as complicated as it actually looks. For example if we want to use monthly data, the last observation of July will be the forecasting value. For quarterly data, equivalent rules will be used. It does not matter which seasonal period, the forecasting value will be the last observation of the same seasonal period (Hyndman & Athanasopoulos, 2018).

2.2.4 Drift Method

In this method, a modified version of naive method, with increasing or decreasing the forecasts over time, where the average change seen in historical data is the quantity of change over time (it is called as drift). Hence the forecasting value for time $T+h$ is denoted by:

$$\hat{Y}_{T+h|T} = Y_T + \frac{h}{T-1}\sum_{t=2}^{T}(Y_T - Y_{T-1}) = Y_T + h\left(\frac{Y_T - Y_1}{T-1}\right) \tag{2.4}$$

This equation means create a line among the first and last observations, and extrapolate it in to the future (Hyndman & Athanasopoulos, 2018).

2.2.5 Simple Moving Average Method

As we described in section *2.1.1*, average method assumes that the best predictor of the future value is the average of all observations that has happened until now. But there is a method that gives more importance to the last observations. This is called *simple moving average method*, and its equation for predicting the value of Y at time $T+1$ based on data up to time T is:

$$\hat{Y}_{T+1} = \frac{(Y_{\mathrm{T}} + Y_{\mathrm{T}-1} + \cdots + Y_{T-m+1})}{m} = \frac{1}{m}\sum_{j=0}^{m-1} Y_{T-j} \tag{2.5}$$

In the simple moving average method, each of the past m observations has a weight of 1/m in the averaging formula (Nau, 2014). If m is smaller, the more weight is given to recent observations. If m gets larger, the less weight is given to recent observations.

2.3 Exponential Smoothing Method

Exponential smoothing methods arise from the works of Brown (Brown, 1959), (Brown, 1964), Holt (Holt, 1957) and Winters (Winters, 1960). The method was improved by Brown and Holt. Roberts G. Brown originated the exponential smoothing while he was working for the US Navy during World War II (Gass & Harris, 2001). In those years, Brown was commissioned to design a monitoring system for fire control information to calculate the location of submarines. Actually, the tracking model developed by Brown consisted of simple exponentially smoothing of continuous data. During the early 1950s, Brown extended simple exponential to discrete data and developed methods for including trends and seasonality. In 1956, Brown presented his work on exponential smoothing at a conference and this formed the basis of Brown's first book (Çapar, 2009).

In 1957, Charles C. Holt was working on forecasting methods on production, inventories and labour force. It seems that Holt and Brown worked independently from each other and they knew nothing about each-other's study. Holt published a paper "Forecasting trends and seasonal by exponentially weighted moving averages" with the support of the Office of Naval Research. He described double exponential smoothing (Trubetskoy, 2016).

Three years later, in 1960, Peter R. Winters, who was a student of Charles C. Holt, developed the algorithm by adding seasonality and published his study "Forecasting sales by exponentially weighted moving averages". His study became known as triple exponential smoothing or Holt-Winters method (Trubetskoy, 2016).

Different types of exponential smoothing methods for forecasting have been come out since at least 1950's. But model selection techniques within the exponential smoothing family arose much later. This family includes methods with different error type, trend type and seasonality type. Some of these are in the Table 2.2 as below:

Table 2.2 Extended exponential smoothing statements (Bicen, Kayikci & Aras, 2015; Hyndman, n.d.)

		Seasonality		
		None	*Additive*	*Multiplicative*
Trend	*None*	$S_t = \alpha X_t + (1-\alpha)S_{t-1}$ $\hat{X}_t(m) = S_t$	$S_t = \alpha(X_t - I_{t-p}) + (1-\alpha)S_{t-1}$ $I_t = \delta(X_t - S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = S_t + I_{t-p+m}$	$S_t = \alpha(X_t/I_{t-p}) + (1-\alpha)S_{t-1}$ $I_t = \delta(X_t/S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = S_t\, I_{t-p+m}$
	Additive	$S_t = \alpha X_t + (1-\alpha)(S_{t-1} + T_{t-1})$ $T_t = \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1}$ $\hat{X}_t(m) = S_t + mT_t$	$S_t = \alpha(X_t - I_{t-p}) + (1-\alpha)(S_{t-1} + T_{t-1})$ $T_t = \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1}$ $I_t = \delta(X_t - S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = S_t + mT_t + I_{t-p+m}$	$S_t = \alpha(X_t/I_{t-p}) + (1-\alpha)(S_{t-1} + T_{t-1})$ $T_t = \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1}$ $I_t = \delta(X_t/S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = (S_t + mT_t)I_{t-p+m}$
	Multiplicative	$S_t = \alpha X_t + (1-\alpha)(S_{t-1}R_{t-1})$ $R_t = \beta(S_t/S_{t-1}) + (1-\beta)R_{t-1}$ $\hat{X}_t(m) = S_t\, R_t^m$	$S_t = \alpha(X_t - I_{t-p}) + (1-\alpha)(S_{t-1}R_{t-1})$ $R_t = \beta(S_t/S_{t-1}) + (1-\beta)R_{t-1}$ $I_t = \delta(X_t - S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = S_t\, R_t^m + I_{t-p+m}$	$S_t = \alpha(X_t/I_{t-p}) + (1-\alpha)(S_{t-1}R_{t-1})$ $R_t = \beta(S_t/S_{t-1}) + (1-\beta)R_{t-1}$ $I_t = \delta(X_t/S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = (S_t\, R_t^m)I_{t-p+m}$
	Additive damped	S_t $= \alpha X_t + (1-\alpha)(S_{t-1} + \varnothing T_{t-1})$ T_t $= \beta(S_t - S_{t-1}) + (1-\beta)\varnothing T_{t-1}$ $\hat{X}_t(m) = S_t + \sum_{i=1}^{m} \varnothing^i T_t$	$S_t = \alpha(X_t - I_{t-p}) + (1-\alpha)(S_{t-1} + \varnothing T_{t-1})$ $T_t = \beta(S_t - S_{t-1}) + (1-\beta)\varnothing T_{t-1}$ $I_t = \delta(X_t - S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = S_t + \sum_{i=1}^{m} \varnothing^i T_t + I_{t-p+m}$	$S_t = \alpha(X_t/I_{t-p}) + (1-\alpha)(S_{t-1}$ $+ \varnothing T_{t-1})$ $T_t = \beta(S_t - S_{t-1}) + (1-\beta)\varnothing T_{t-1}$ $I_t = \delta(X_t/S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = \left(S_t + \sum_{i=1}^{m} \varnothing^i T_t\right) I_{t-p+m}$
	Multiplicative damped	$S_t = \alpha X_t + (1-\alpha)\left(S_{t-1}R_{t-1}^{\varnothing}\right)$ $R_t = \beta(S_t/S_{t-1}) + (1-\beta)R_{t-1}^{\varnothing}$ $\hat{X}_t(m) = S_t\, R_t^{\sum_{i=1}^{m} \varnothing^i}$	$S_t = \alpha(X_t - I_{t-p}) + (1-\alpha)\left(S_{t-1}R_{t-1}^{\varnothing}\right)$ $R_t = \beta(S_t/S_{t-1}) + (1-\beta)R_{t-1}^{\varnothing}$ $I_t = \delta(X_t - S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = S_t\, R_t^{\sum_{i=1}^{m} \varnothing^i} + I_{t-p+m}$	$S_t = \alpha(X_t/I_{t-p}) + (1-\alpha)\left(S_{t-1}R_{t-1}^{\varnothing}\right)$ $R_t = \beta(S_t/S_{t-1}) + (1-\beta)R_{t-1}^{\varnothing}$ $I_t = \delta(X_t/S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = \left(S_t\, R_t^{\sum_{i=1}^{m} \varnothing^i}\right) I_{t-p+m}$

α	*Smoothing parameter for the level of the series*
β	*Smoothing parameter for the trend*
δ	*Smoothing parameter for seasonal indices*
ϕ	*Damping parameter*
m	*Number forecast period*
p	*Number of periods in the seasonal cycle*
S_t	*Smoothed level of the series in period(t)*
X_t	*Observed value of the time series in period(t)*
$\hat{X}_t(m)$	*Forecast for m periods ahead from (t)*
T_t	*Smoothed additive trend in period(t)*
R_t	*Smoothed multiplicative trend in period(t)*
I_t	*Smoothed seasonal index in period(t)*

Researchers made a lot of contributions to Brown’s and Holt’s original studies on exponential smoothing. These contributions have been successful, so exponential smoothing family was extended. These contributions made for different forecast profiles (Çapar, 2009). These profiles are given in the figure below:

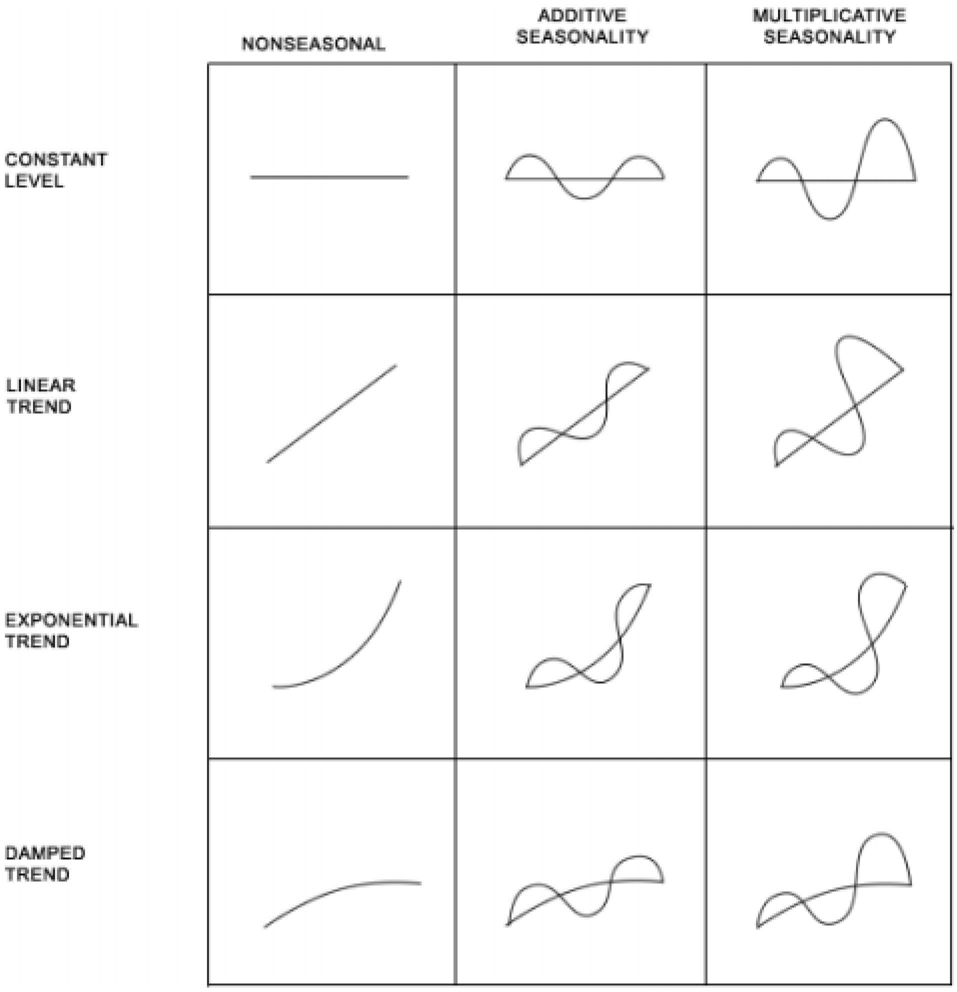


Figure 2.4 Forecast profiles from exponential smoothing (Gardner, 1985)

Exponential smoothing is a method that forecasts are constantly updated by taking into account the recent changes and jumps in the data. When making a forecast for future observation (time $T+1$) with exponential smoothing method, the forecast for the last observation (time T) and some part of the error that obtained from the last observation forecast are used (Kiren, n.d.). This can be stated as:

$$\hat{X}_{T+1} = \hat{X}_T + \alpha(e_t) \tag{2.6}$$

where α is the smoothing parameter in this equation.

$$S_T = S_{T-1} + \alpha(X_T - S_{T-1}) \tag{2.7}$$

S_T : Forecast of the future observation $(T + 1)$

S_{T-1} : Forecast of the last observation (T)

α : Smoothing parameter or weight, where α is a constant and $\alpha \in [0,1]$.

$(X_T - S_{T-1})$: Error for the forecast of the last period

$$\hat{X}_{T+1} = S_T = \alpha X_T + (1-\alpha)S_{T-1} \tag{2.8}$$

If we substitute the equation of the S_{T-1} in the equation 2.8 and continue this until to obtain the term S_0, we get the equation below:

$$\hat{X}_{T+1} = S_T = \alpha X_T + \alpha(1-\alpha)X_{T-1} + \alpha(1-\alpha)^2 X_{T-2} + \cdots + \alpha(1-\alpha)^{T-1}X_1 + \alpha(1-\alpha)^T S_0 \tag{2.9}$$

Equation can be express like this:

$$\hat{X}_{T+1} = S_T = \alpha \sum_{j=0}^{T} (1-\alpha)^j X_{T-j} \tag{2.10}$$

The equation 2.10 is the general exponential smoothing equation, where S_T is the smoothed value at time T and it decreases with weights exponentially.

As it is seen in the equation 2.9, values of the observations are being multiplied by the weight α. Due to α is a constant between 0 and 1, past observations are getting smaller weights. It means that recent observations are more important in the equation.

Best known methods in this family are simple exponential smoothing, Holt's linear method and Holt-Winter's multiplicative method (Svensson, 2018). In this part we will focus on these three familiar methods in the exponential smoothing family.

2.3.1 Simple Exponential Smoothing

The simplest of the exponentially smoothing methods is called as simple exponential smoothing (SES). Using this method is suitable for forecasting data with no trend or seasonal pattern (Hyndman & Athanasopoulos, 2018). SES is a very simple useful way for making forecasts. There is an automatic optimization rule for selecting the value needed in the initialization phase along with the smoothing parameter in the SES function (Svensson, 2018). So it is considered that simple exponential smoothing is best applied to time series that do not exhibit prevalent trend and do not exhibit seasonality. The smoothing constant (or weight) α is used to control the speed which the updated forecast will adapt to local level (or mean) of the time series (Stokes, 2011).

Let X_T denote observed value of a time series at time T and $\hat{X}_T(h)$ be the forecast for h periods from the origin T. Here h is an integer greater than 0 and it represents the time between the initiation and completion of the process. So h can be called forecasting horizon or lead time (Taylan Selamlar, 2017). Therefore, SES model can be expressed as:

$$S_T = \alpha X_T + (1-\alpha) S_{T-1} \tag{2.11}$$

$$S_T = S_{T-1} + \alpha (X_T - S_{T-1}) \tag{2.12}$$

$$\hat{X}_T(h) = S_T \tag{2.13}$$

where S_T is the smoothed value at time T. Here equation (2.11) is the component form and equation (2.12) is the error form. Smoothing constant is denoted by $\alpha \in [0,1]$.

Substituting the equation (2.11) into itself until obtaining the initial term S_0, we can re-write the equation as:

$$S_T = \alpha \sum_{k=0}^{T-1} (1-\alpha)^k X_T + (1-\alpha)^T S_0 \tag{2.14}$$

So S_T represents a weighted moving average of all past observations with weights decreasing exponentially. As it can be seen in the equation, for larger α, recent observations gets more weight. Similar to naïve and average methods, the weights from SES satisfy the conditions below:

1. $w_T \in [0,1]$ where $T = 1,2,\dots,n$

2. $\sum_{T=1}^{n} w_T = 1$

3. $w_1 \leq w_2 \leq \cdots \leq w_n$

It can be seen that sum of the weights assigned by exponential smoothing is unity. If is small α, exponential smoothing process gives more weight to more distant past observation. In the exponential smoothing equation, weight of k period ago observation is equal to $\alpha(1-\alpha)^k$. Now we give an example for 10 observations of the time series X_T. For different smoothing levels, we can see the weights assigned for each observation given in the Table 2.3 as below:

Table 2.3 Weights of 10 observations for different smoothing parameters of SES

		Weights of X_T			
Observation	Formulation	α =0.1	α =0.4	α =0.8	α =0.9
X_{10}	α	0.1	0.4	0.8	0.9
X_9	$\alpha(1-\alpha)^1$	0.09	0.24	0.16	0.09
X_8	$\alpha(1-\alpha)^2$	0.081	0.144	0.032	0.009
X_7	$\alpha(1-\alpha)^3$	0.0729	0.0864	0.0064	0.0009
X_6	$\alpha(1-\alpha)^4$	0.06561	0.05184	0.00128	0.00009
X_5	$\alpha(1-\alpha)^5$	0.059049	0.031104	0.000256	0.000009
X_4	$\alpha(1-\alpha)^6$	0.0531441	0.0186624	0.0000512	0.0000009
X_3	$\alpha(1-\alpha)^7$	0.04782969	0.01119744	0.00001024	0.00000009
X_2	$\alpha(1-\alpha)^8$	0.043046721	0.006718464	0.000002048	0.000000009
X_1	$\alpha(1-\alpha)^9$	0.0387420489	0.0040310784	0.0000004096	0.0000000009
Initial Value	$(1-\alpha)^{10}$	0.3486784401	0.0060466176	0.0000001024	0.0000000001

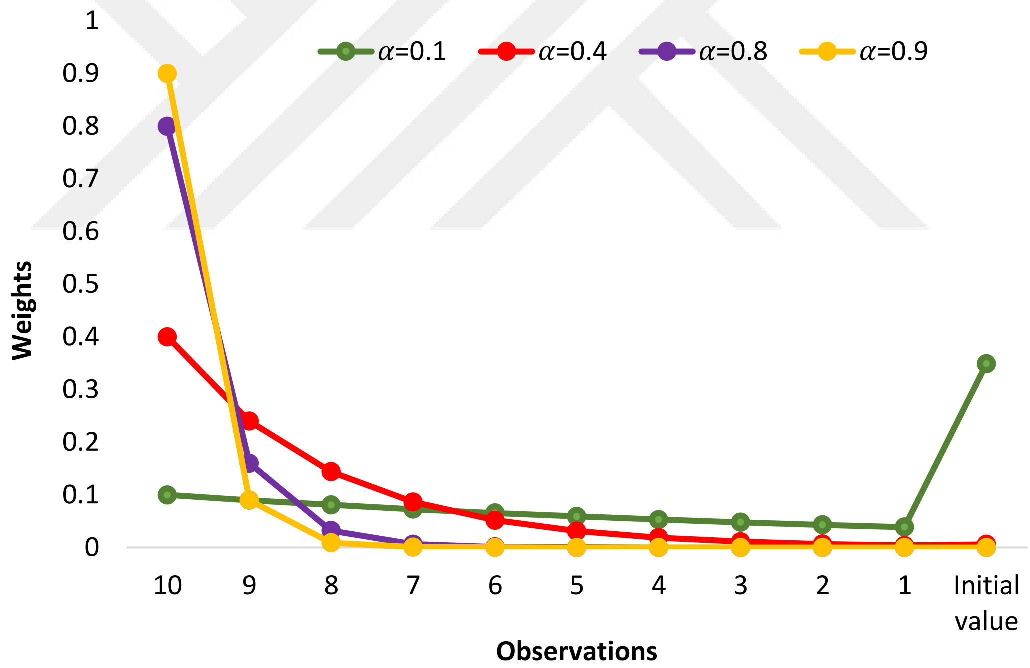


Figure 2.5 Weights assigned by different smoothing parameters of SES

As it is seen in the Table 2.3 and figure 2.5, while smoothing constant $\alpha = 0.1$, distant observations get higher weights with regard to larger values of α. Parallel to this, when smoothing constant $\alpha = 0.9$, a large value, recent observations get higher value. So we can say that if smoothing constant takes a large value, recent observations get higher importance in SES. Therefore, analysis by SES dominantly shapes with recent observations.

Table 2.4 Oil production data (Hyndman, n.d.)

Year	Time period t	Observed values y_t	$\alpha = 0.2$	$\alpha = 0.6$ Level ℓ_t	$\alpha = 0.89^*$
–	0	–	446.7	446.7	447.5*
1996	1	446.7	446.7	446.7	446.7
1997	2	454.5	448.2	450.6	453.6
1998	3	455.7	449.7	453.1	455.4
1999	4	423.6	444.5	438.4	427.1
2000	5	456.3	446.8	447.3	453.1
2001	6	440.6	445.6	444.0	441.9
2002	7	425.3	441.5	434.6	427.1
2003	8	485.1	450.3	459.9	478.9
2004	9	506.0	461.4	483.0	503.1
2005	10	526.8	474.5	504.9	524.2
2006	11	514.3	482.5	509.6	515.3
2007	12	494.2	484.8	501.9	496.5
	h			Forecasts $\hat{y}_{T+h\|T}$	
2008	1	–	484.8	501.9	496.5
2009	2	–	484.8	501.9	496.5
2010	3	–	484.8	501.9	496.5

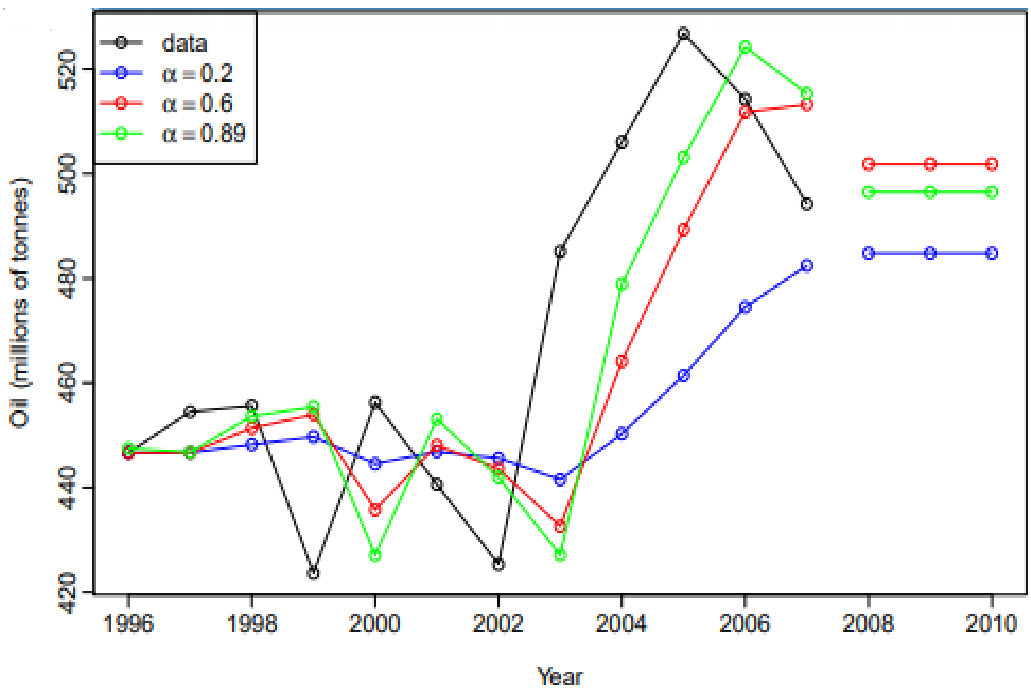


Figure 2.6 Analysing oil production data by SES for different α (Hyndman, n.d.)

In Table 2.4, there are oil production data by years and forecasts by analysing data with SES for this data. Here l_0 represents the initial value. Figure 2.6 shows the forecasting values of this analysis for different α. As it is seen in the figure, for the largest $\alpha = 0.89$, forecasts are approaching the real value of observations by the year 2004. For the smallest $\alpha = 0.2$, forecasts are diverging the real value of observations by the year 2004. It can be understood that, as we have mentioned before largest α gives more weight to recent observations and this increases the accuracy of the forecast with SES.

Table 2.5 2008-2017 Tax revenues 2008-2017 in Turkey (MTF, 2019)

Tax Revenues in Turkey				
n	**Year**	**Billion TL**	**Smoothed Value (α=0.1)**	**Smoothed Value (α=0.9)**
1	2008	168		
2	2009	172	168.01	168.10
3	2010	211	168.45	172.01
4	2011	254	172.66	206.70
5	2012	279	180.78	249.10
6	2013	326	190.58	275.81
7	2014	353	204.14	321.13
8	2015	408	218.98	349.38
9	2016	459	237.86	401.97
10	2017	537	259.97	453.30

As we see in Table 2.5, the greater $\alpha = 0.9$ gives better smoothed values than the smaller $\alpha = 0.1$. We can say that giving more weight to recent observations enhance the accuracy of our forecasting.

2.3.1.1 Smoothing Parameter and Initial Value

Most of the time, it is complicated to choose a smoothing parameter and initial value for simple exponential smoothing method. Because selection of the smoothing parameter and initial value will affect the accuracy of the forecast by simple exponential smoothing method. In other words, choosing the correct smoothing parameter and initial value is very important to enhance the accuracy of forecasting by SES.

According to common opinion, firstly smoothing parameter should be chosen. The accuracy of SES rely on the value of smoothing constant. The value of smoothing constant should be chosen in such a way that forecasts are more accurate. The level of forecasting accuracy is measured with regards to forecasting error which denotes the difference between actual value and forecasted value. The aim is to identify the smoothing constant that minimizes the forecast error (Karmaker, 2017).

The smoothing constant determines the sensitivity of the forecast. As we mentioned before, large values of α give more weight to recent observations. This makes the forecast more susceptible to more recent observations, whereas smaller values of α have a damping effect. Because smaller values of α give more weight to distant past observations (Ravinder, 2013).

Most academic publications make general recommendations for the size of the smoothing constant. For example, both Schroeder, Rungtusanatham, & Goldstein (2013) and Jacobs & Chase (2013) suggest values of α between 0.1 and 0.3. Heizer & Render (2011) and Stevenson (2012) advocate a wider range: 0.05 to 0.50. Chopra & Meindl (2013) prescribe α values no larger than 0.20 (Ravinder, 2013).

Most textbooks also recommend that smoothing constants should be chosen with statistical measures like Mean Absolute Deviation (MAD), Mean Squared Error (MSE), Mean Absolute Percent Error (MAPE), or some other summary metric, so that forecasts will be more accurate (Ravinder, 2013). In practice, smoothing parameter is usually chosen by a grid search to minimize the *ex post* MSE. It is certain that value of α should be in the interval between 0-1. For SES, Gardner (1985) says that more restricted range 0.10-0.30 is typical in practice, it is widely held that a more complex model should be used if the best value of α falls above 0.30 during the model-fitting process (Gardner, 1985).

After choosing smoothing parameter the second aim is to choose starting value. This is calling as “initialization problem”. Several methods have been developed to calculate the initial value S_0. Bu there is no specific method that is supported by researchers generally. Brown’s (1959) original suggestion is popular in practice, which is simply using the mean of the data for calculating S_0 (Gardner, 1985). Ledolter and Abraham (1984) recommend back casting to obtain S_0 (Çapar, 2009).

2.3.1.2 Measuring Forecast Accuracy

For a forecasting method, its performance characteristics should be verified or validated of its forecast with historical data for the process it was designed forecast. This is no consensus among researchers as to which measure is best for determining the appropriate forecasting method. It can be accuracy as a criterion that determines the best forecasting method; so, accuracy is the most important concern in evaluating the quality of a forecast. The aim of the forecast is minimizing error (Ostertagova & Ostertag, 2012).

The forecasting error can be calculated as:

$$e_t = Y_t - F_t \tag{2.15}$$

where e_t is the forecasting error, Y_t is the actual value and F_t the forecast period t (Taylan Selamlar, 2017).

Mainly common indicators used to evaluate accuracy are “mean absolute error (MAE)”, “mean squared error (MSE)”, “root mean squared error (RMSE)” or “mean absolute percentage error (MAPE)”:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t| \tag{2.16}$$

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2 \tag{2.17}$$

$$RMSE = \sqrt{MSE} = \sqrt{\left(\frac{1}{n}\sum_{t=1}^{n} e_t^2\right)} \tag{2.18}$$

where, Y_t is the actual value at the time t, e_t is the residual at the time t, n is the total number of the time periods (Ostertagova & Ostertag, 2012).

MAE, measures the difference between the forecasting value and the real value. MAE is the average over the verification sample of the absolute values of the differences between forecast and the corresponding observation, where all errors are assigned equal weights. If a good forecast has been made, MAE will be close to zero. If the value of MAE is large, it means the forecast is not successful (Anonymus, n.d.).

MSE is also a measure of overall accuracy that gives an additional weight for large error. For evaluating exponential smoothing and other methods MSE is a generally accepted technique (Ostertagova & Ostertag, 2012).

The square root of MSE, RMSE, measures the difference between forecast and corresponding observed value, squared this difference and averaged over the sample. Finally, the square root of the average is taken. The RMSE gives a relatively high weight to large errors since the errors are squared before they are averaged. This means the RMSE is most useful when large errors are particularly undesirable (Anonymus, n.d.).

There are some other measures based on percentage error. The percentage error is defined as:

$$p_t = \frac{100e_t}{Y_t} \tag{2.19}$$

The most widely used measures are mean absolute percentage error and median absolute percentage error. Percentage error have the advantage of being scale independent, so they are frequently used to compare forecast performance between different data series. They are defined as,

Mean absolute percentage error:

$$MAPE = mean\ (|p_t|) = \frac{1}{n}\sum_{t=1}^{n}\frac{|e_t|}{Y_t}.100 \tag{2.20}$$

Median absolute percentage error:

$$MdAPE = median\ (|p_t|) \tag{2.21}$$

The MAPE is simple to calculate. It measures the size of the error and express this in percentage terms. The MAPE is easy to understand because most of people feel comfortable thinking in percentage terms. However, the accuracy and reliability of MAPE is being discussed in academic people (Swanson, Tayman, & Bryan, 2015).

But the measurement MAPE gives us a disadvantage that it puts a heavier penalty on positive errors than on negative errors. This status has led to the use of the “symmetric” MAPE (sMAPE) in the M3-competition (Taylan Selamlar, 2017). It is defined as,

Symmetric mean absolute percentage error:

$$sMAPE = mean\left(200\frac{|Y_t - F_t|}{(Y_t + F_t)}\right) \tag{2.22}$$

Symmetric median absolute percentage error:

$$sMdAPE = median\left(200\frac{|Y_t - F_t|}{(Y_t + F_t)}\right) \tag{2.23}$$

If the actual value is zero, the forecast is likely to be close to zero. Hence the measurement will still involve division by a number close to zero. Also, the value of sMAPE can be negative, so it is not really a measure of “absolute percentage errors” at all (Taylan Selamlar, 2017).

Hyndman & Koehler (2006) proposed a new measurement that is suitable for all situations, by scaling error based on the *in-sample* MAE from the naïve (random walk) forecast method. The scaled error is defined as:

$$q_t = \frac{e_t}{\frac{1}{n-1}\sum_{i=2}^{n}|Y_i - Y_{i-1}|} \tag{2.24}$$

which is clearly independent of the scale of the data. A scale error is less than one of it arises from a better forecast than the average one-step naïve forecast computed in-sample. Conversely it is greater than of if the forecast is worse than the average one-step naïve forecast computed in-sample (Hyndman & Koehler, 2006).

The mean absolute scaled error is simply:

$$MASE = mean\ (|q_t|) \tag{2.25}$$

Billah et al. (2005) used a similar error measure when they computed the absolute value of the forecast error as a percentage of the insample standard deviation. However, this approach has the disadvantage that the denominator grows with the sample size for non-stationary series containing a unit root. Scaling by the insample naïve MAE only assumes that series has no more than one unit root, which is almost always true for real data. When MASE < 1, the proposed method gives, on average, smaller errors than the one-step errors from the na¨ıve method. If multi-step forecasts are being computed, it is possible to scale by the in-sample MAE computed from multi-step naïve forecasts (Hyndman & Koehler, 2006).

2.3.2 Holt's Linear Trend Method

Holt's two parameter exponential smoothing (Holt, 1957) is best applied to time series that have a prevalent linear trend but does not exhibit seasonal behaviour. The smoothing constant α is again used to control speed of adaptation to local level but a second smoothing constant β is introduced to control degree of local trend carried through to multi-step-ahead forecast periods. The recursive form of the equation can be written as:

$$\mathrm{L}_T = \alpha X_T + (1-\alpha)(L_{T-1} + T_{T-1}) \tag{2.26}$$

$$\mathrm{T}_T = \beta(L_T + L_{T-1}) + (1-\beta)T_{T-1} \tag{2.27}$$

$$\widehat{\mathrm{X}}_T(\mathrm{h}) = S_T + hT_T \tag{2.28}$$

where X_T is the time series in period T, L_T is the level equation, T_T is the trend equation, α is the smoothing parameter for level, β is the smoothing parameter for trend, where $0 < \alpha, \beta < 1$, $\widehat{\mathrm{X}}_T(\mathrm{h})$ is the h-step-ahead smoothed forecast value for X (Stokes, 2011).

For level and trend, starting values is required for Holt's exponential smoothing method. The parameters and starting values can be estimated by minimizing the MSE, MAE, MAPE or some other criterion for measuring in-sample forecast error (Yapar, Capar, Taylan Selamlar, & Yavuz, 2017).

2.3.3 Holt-Winters' Seasonal Method

Holt's method is widen to capture seasonality by Holt (1957) and Winters (1960). The Holt-Winters seasonal method includes the forecast equation and three smoothing equations; L_T is the equation for level, T_T is the equation for trend and S_T is the equation for seasonal component, with matching smoothing parameters α, β and γ. The frequency of the seasonality is denoted by m, i.e., the number of seasons in one year. For instance we use for quarterly data $m = 4$, and for monthly data $m = 12$ (Hyndman & Athanasopoulos, 2018).

There are two versions of this method that the seasonal component changes in these versions. When the seasonal variations are roughly constant through the series the additive method is preferred. If the seasonal variations are changing proportional to the level of the series the multiplicative method is preferred. In the additive method, the seasonal component is expressed in absolute terms on the observed series scale, and in the level equation the series is seasonally adjusted by subtracting the seasonal component. Within each year, the seasonal component will add up to approximately zero. In the multiplicative method, seasonal component expressed relative (percentage), and the series is seasonally adjusted by the seasonal component. The seasonal component will be collected approximately in each year. (Hyndman & Athanasopoulos, 2018).

2.3.3.1 Holt Winters' Additive Seasonal Method

If the seasonal component is in the additive case, the series shows steady seasonal fluctuations, regardless of the overall level of the series (Kalekar, 2004). The additive method is preferred when the seasonal variations are roughly constant through the series. Let's estimate L_T is the smoothed estimate of the level at time T, T_T is the smoothed estimate of the change in the trend value at time T, S_T is the smoothed estimate of the appropriate seasonal component at time T. In additive method:

$$\mathrm{L}_T = \alpha(X_T - S_{T-p}) + (1-\alpha)(L_{T-1} + T_{T-1}) \tag{2.29}$$

$$\mathrm{T}_T = \beta(L_T - L_{T-1}) + (1-\beta)T_{T-1} \tag{2.30}$$

$$\mathrm{S}_T = \gamma(X_T - L_T) + (1-\gamma)S_{T-p} \tag{2.31}$$

$$\widehat{\mathrm{X}}_T(\mathrm{h}) = (L_T + hT_T)S_T \tag{2.32}$$

where α, β and γ are the smoothing parameters, L_T is the smoothed level at time T as given in equation (2.29), T_T is the change in the trend at time T as in equation (2.30), S_T is the seasonal smoothing parameter at time T as in equation (2.31), and p is the number of seasons per year, $\widehat{X}_T(h)$ is the h-step-ahead smoothed forecast value for X as in equation (2.32) (Tirkeş, Güray, & Çelebi, 2017).

2.3.3.2 Holt-Winters' Multiplicative Model

It is similar to Holt-Winters' additive method. This seasonal multiplicative method multiplies the trended forecast by the seasonality, producing the Holt-Winters' multiplicative forecast.

Multiplicative model is used when the data exhibit multiplicative seasonality. Multiplicative method is preferred when the seasonal variations are changing proportionally the level of the series. The component form for the multiplicative method is:

$$L_T = \alpha \frac{X_T}{S_{T-p}} + (1-\alpha)(L_{T-1} + T_{T-1}) \tag{2.33}$$

$$T_T = \beta(L_T - L_{T-1}) + (1-\beta)T_{T-1} \tag{2.34}$$

$$S_T = \gamma \frac{X_T}{L_T} + (1-\gamma)S_{T-p} \tag{2.35}$$

$$\widehat{X}_T(h) = (L_T + hT_T)S_T \tag{2.36}$$

where α, β and γ are the smoothing parameters, L_T is the smoothed level at time T as given in equation (2.33), T_T is the change in the trend at time T as in equation (2.34), S_T is the seasonal smoothing parameter at time T as in equation (2.35), and p is the number of seasons per year, $\widehat{X}_T(h)$ is the h-step-ahead smoothed forecast value for X as in equation (2.36) (Tirkeş et al., 2017).

2.4 ARIMA Models

ARIMA models give another approach to time series forecasting. Exponential smoothing and ARIMA models are the two most widely used approaches to time series forecasting, and provide supplementary approaches to the problem. While exponential smoothing models are based on a description of the trend and seasonality in the data, ARIMA models aim to describe the autocorrelations in the data. So ARIMA models are more complicated than exponential smoothing methods (Hyndman & Athanasopoulos, 2018).

ARIMA models as also known as Box-Jenkins. ARIMA model is a class of statistical models for analysing and forecasting time series data. ARIMA is an acronym that represents AutoRegressive Integrated Moving Average. If we want to explain the meaning of the acronym:

- *AR:* AutoRegressive. A model that uses the dependent relationship between an observation and some number of lagged observations.
- *I:* Integrated. The use of differencing of raw observations (e.g. subtracting an observation from an observation at the previous time step) in order to make the time series stationary.
- *MA:* Moving Average. A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations (Brownlee, 2017).

For forecasting a time series, ARIMA models are the most general class of models in theory, which can be made to be “stationary” by differencing (if necessary), perhaps in connected with nonlinear transformations such as logging or deflating (if necessary). A random variable that is a time series is stationary if its statistical properties are all constant over time. A stationary series has no trend, its variations around its mean have a constant amplitude, and it wiggles in a consistent fashion (Nau, 2018).

ARIMA models are mathematical models of persistence, or autocorrelation, in a time series. ARIMA models allow us not only to uncover the hidden patterns in the data but also to generate forecasts and predict a variable’s future values from its past values. ARIMA models can be expressed by a series of equations. One subset of ARIMA models is called autoregressive, or AR models. An AR model describes a time series as a linear function of its past values plus a noise term ε_t. The order of the AR model shows the number of past values included. The simplest AR model is the first-order autoregressive, or AR (1) model. The equation for this model is given by:

$$X_t = \Phi X_{t-1} + Z_t \tag{2.37}$$

where $t = 1, 2, \ldots, N$, X_t is a stationary zero-mean time series and Φ is the first-order autoregressive coefficient. We can see that the AR (1) model has the form of a regression model in which z_t is regressed on its previous value, and the error term Z_t is analogous to the regression residuals and represents a "white noise" (uncorrelated with mean 0 and variance σ^2) process (Fu, 2010).

The moving average (MA) model is another form of ARIMA model in which the time series is described as a linear function of its prior errors plus a noise term ε_t. The first-order moving average, or MA (1), model is given by:

$$X_t = Z_t - \theta Z_{t-1} \tag{2.38}$$

where $t = 1, 2, \ldots, N$, z_t is a stationary zero-mean time series, Z_t, Z_{t-1} are the error terms at time t and $t - 1$, and θ is the first-order moving average coefficient (Fu, 2010).

A general autoregressive moving average (ARMA) model, ARMA (p, q), is given by:

$$X_t = Z_t - \theta Z_{t-1} \tag{2.39}$$

The integrated ARMA (ARIMA) is a broadening of the class of ARMA that includes differencing. The first stage is to identify the model. Identification consists of specifying the appropriate model (AR, MA, ARMA, or ARIMA) and order of model. The second stage is to estimate the order of the model. At this stage, the coefficients are estimated, so that the sum of squared residuals is minimized. The final stage is model diagnostics. One of the important elements in this stage is to make sure that the residuals of the candidate model are random and normally distributed. And the other one is to ensure that the estimated parameters are statistically significant (Nau, 2018).

A nonseasonal ARIMA model is classified as an "ARIMA(p,d,q)" model, where:

- p is the number of autoregressive terms,
- d is the number of nonseasonal differences needed for stationary, and
- q is the number of lagged forecast in the prediction equation.

The forecasting equation is constructed as follows. First, let x denote the d^{th} difference of X, which means:

If $= 0$: $x_t = X_t$

If $= 1$: $x_t = X_t - X_{t-1}$

If $= 2$: $x_t = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) = X_t - 2X_{t-1} + X_{t-2}$

Note that the difference of X (the $d = 2$ case) is not the difference from 2 periods ago. Rather, it is the *first-difference-of-the-first-difference*, which is the discrete analog of a second derivative, i.e., the local acceleration of the series rather than its local trend (Nau, 2018).

In terms of x, the general forecasting equation is:

$$\hat{x}_t = \mu + \Phi_1 x_{t-1} + \cdots + \Phi_p x_{t-p} - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q} \qquad (2.40)$$

Here the moving average parameters (θ’s) are defined so that their signs are negative in the equation, following the convention introduced by Box and Jenkins. Φ’s are slope coefficients and e’s are error terms (Nau, 2018).

CHAPTER THREE

ATA METHOD AND M4-COMPETITIONS

3.1 ATA Method

There are some reasons why a new forecasting method is needed. For example, in other forecasting methods, selection of initial value and smoothing parameter is a problem. Because the results of forecasting vary according to the selected initial value and the smoothing parameter. Accuracy of the results change according to selected initial value and the smoothing parameter. This selection of these variables directly affects the values of forecast. So the correct selection of initial value and smoothing parameter. For the accuracy of forecasting, the selection of these variables should be bounded by certain rules.

Among the time series forecasting methods, exponential smoothing is considered as one of the most succesful forecasting method. In addition of its success, ES is a very simple method, so this makes ES the widely used forecasting method. But its accuracy can be affected by the selection technique of initial value and smoothing parameter. The selection of these variables is a subjective decision so this is a problem for the accuracy of forecast. But ES is also suffers from basic problems, for example giving weights to the observations. Exponential smoothing fails to account for the amount of data points that can contribute to the forecast when assigning weights to historical data (Yapar et al., 2017).

ATA method, which is proposed by Yapar (2016) and Yapar et al. (2017), is a new forecasting method as an alternative for other forecasting techniques (Yapar, 2016; Yapar et al., 2017). Remarkable features of the ATA method are simplicity, easy optimization and amazing performance (Yapar et al., 2017). These features will make the ATA method easily accepted in the literature. This new forecasting method will have widespread use in the future.

The ATA method has similar manner to exponential smoothing method but there is significantly distinctness in ATA method which eliminates the initialization problem and is more practical to optimize compared to its counterpart ES models (Yapar, 2016). Furthermore, the smoothing parameters in ATA method are modified so that while acquiring a smoothed value at a specific time point the weights among observations are distributed by considering how many observations can contribute to the smoothed value (Taylan Selamlar, 2017).

For the series X_t $t = 1,2,\ldots,n$ the general additive model which will be denoted by ATA(p, q) can be written as:

$$S_t = \left(\frac{p}{t}\right) X_t + \left(\frac{t-p}{t}\right)(S_{t-1} + T_{t-1}) \quad for \tag{3.1}$$

$$T_t = \left(\frac{q}{t}\right)(S_t - S_{t-1}) + \left(\frac{t-q}{t}\right) T_{t-1} \quad for \tag{3.2}$$

$$\widehat{X}_t(h) = S_t + hT_t \tag{3.3}$$

For $p \in \{1,\ldots,n\}$, $q \in \{0,1,\ldots,n\}$ and $n > p \geq q$. For $n \leq p$ and $S_t = X_t$; for $n \leq q$ and $T_t = X_t - X_{t-1}$ and then $T_1 = 0$. Here X_t is the value of original series, T_t is the trend component and S_t is the smoothed value at time *t*. *p* is the smoothing parameter for level, *q* is the smoothing parameter for trend and $\widehat{X}_t(h)$ is the *h* step ahead forecast value (Taylan Selamlar, 2017).

Similarly, a multiplicative version of the same model $ATA_{mult}(p,q)$ which can be written as:

$$S_t = \left(\frac{p}{t}\right) X_t + \left(\frac{t-p}{t}\right)(S_{t-1} \times T_{t-1}) \tag{3.4}$$

$$T_t = \left(\frac{q}{t}\right)\left(\frac{S_t}{S_{t-1}}\right) + \left(\frac{t-q}{t}\right) T_{t-1} \tag{3.5}$$

$$\hat{X}_t(h) = S_t \times T_t^h \tag{3.6}$$

for $p \in \{1,\dots,n\}$, $q \in \{0,1,\dots,n\}$ and $t > p \geq q$. For $t \leq p$ and $S_t = X_t$, for $t \leq q$ and $T_t = X_t / X_{t-1}$ and then $T_1 = 1$.

It is worth pointing out that ATA method does not require to initial values for level and trend patterns. Because this method is developed by inspiration of the ES models, it can be also adjusted to ES models which can be classified into 30 different models (Hyndman et al., 2018).

3.1.1 The Trended ATA Method

3.1.1.1 ATA (p, q) with Additive Trend

For the series X_t, *t = 1, 2, … , n* the general additive model which is denoted by *ATA(p, q)*, is similar to Holt's linear model with some modifications on level and trend parameters. We can formulate *ATA(p, q)* as below:

$$S_t = \left(\frac{p}{t}\right) X_t + \left(\frac{t-p}{t}\right)(S_{t-1} + T_{t-1}) \tag{3.7}$$

$$T_t = \left(\frac{q}{t}\right)(S_t - S_{t-1}) + \left(\frac{t-q}{t}\right) T_{t-1} \tag{3.8}$$

$$\hat{X}_t(h) = S_t + hT_t \tag{3.9}$$

for $p \in \{1, 2, \dots, n\}$, $q \in \{0, 1, \dots, p\}$ and $n > p \geq q$. For $n \leq p$ let $S_t = X_t$, for $n \leq q$ let $T_t = X_t - X_{t-1}$ and let $T_1 = 0$. Here, the sign S_t denotes an estimate of the level at time *t*, and the sign T_t denotes an estimate of the trend (slope) of the series at time *t*. p is the smoothing parameter for level, q is the smoothing parameter for trend and $\hat{X}_t(h)$ is the h step ahead forecast value (Taylan Selamlar, 2017).

3.1.1.2 ATA (p, q) with Multiplicative Trend

For the series X_t, *t = 1, 2, … , n,* the model which is denoted by $ATA_{mult}(p, q)$ is using for smoothing and forecasting the data which has an exponential trend. The formulation of this method is here below with a forecast equation and the smoothing equations for level (S_t) and trend (T_t) :

$$\mathrm{S}_t = \left(\frac{p}{t}\right) X_t + \left(\frac{t-p}{t}\right)(S_{t-1} * T_{t-1}) \tag{3.10}$$

$$\mathrm{T}_t = \left(\frac{q}{t}\right)\left(\frac{S_t}{S_{t-1}}\right) + \left(\frac{t-q}{t}\right) T_{t-1} \tag{3.11}$$

$$\hat{\mathrm{X}}_t(\mathrm{h}) = S_t * T_t^h \tag{3.12}$$

for $p \in \{1, 2, \dots, n\}$, $q \in \{0, 1, \dots, p\}$ and $n > p \geq q$. For $n \leq p$ let $S_t = X_t$, for $n \leq q$ let $T_t = X_t / X_{t-1}$ and let $T_1 = 1$. As it seen, the trend component T_t has an exponential effcet on the forecast value rather than linear such that $\hat{\mathrm{X}}_t(\mathrm{h})$ is equal to the final estimate of the trend to the power of *h* (Taylan Selamlar, 2017).

3.1.2 The Simplest Form of ATA (p, q)

When *q = 0,* ATA (p, q) gives a simple model which has a similar form to simple exponential smoothing equation:

$$\mathrm{S}_t = \left(\frac{p}{t}\right) X_t + \left(\frac{t-p}{t}\right) S_{t-1} \quad for \quad t > p \tag{3.13}$$

$$\mathrm{S}_t = X_t \qquad\qquad for \quad t \leq p \tag{3.14}$$

$$\hat{\mathrm{X}}_t(\mathrm{h}) = S_t \tag{3.15}$$

where *p* is the smoothing parameter and it organizes the smoothing process. We will call this model as simple form of *ATA (p, 0).* The given equation of S_t which is given in (3.13) can be interpreted as weighted average of the past observations. Additionally, the smoothed value S_t can be rewritten as follows:

$$S_t = \sum_{k=0}^{t-(p+1)} \frac{\binom{t-k-1}{p-1}}{\binom{t}{p}} X_{t-k} + \frac{1}{\binom{t}{p}} S_p \tag{3.16}$$

where S_p is the starting or initial value for *ATA (p, 0),* which can be simply the p^{th} observation or the average of the oldest p observations. It can be easily seen from the equation that the smoothed value at time t is the weighted average of all past observations and the starting value S_p . From the equation (3.16), it is evident that the initial value of S_p relies on the smoothing parameter of p. Thus, the smoothing parameter and initial value are optimized at the same time (Taylan Selamlar, 2017).

3.1.3.1 Weights of ATA (p, 0)

It can be said that the weights of ATA method have the similar properties with that of simple exponential smoothing model. These properties are given below (Ekiz Yılmaz, 2018):

1. $w_t \in [0, 1]$ where $t = 1, 2, \ldots, n$
2. $w_1 \leq w_2 \leq \cdots \leq w_n$
3. $\sum_{t=1}^{n} w_t = 1$.

The weights of the observations given in the Table 3.1 as below (Ekiz Yılmaz, 2018):

Table 3.1 The weights attached to the observations by $ATA(p,0)$ (Ekiz Yılmaz, 2018)

Weight of X_t by $ATA(p,0)$
$w_n = \left(\frac{p}{n}\right)$
$w_{n-1} = \left(\frac{p}{n}\right)\left(\frac{n-p}{n-1}\right)$
$w_{n-2} = \left(\frac{p}{n}\right)\left(\frac{n-p}{n-1}\right)\left(\frac{n-p-1}{n-2}\right)$
$\vdots$
$w_{p+1} = \frac{\binom{p}{p-1}}{\binom{n}{p}}$
$w_p = \cdots = w_1 = 0$

3.1.3.2 Average Age of ATA (p, q=0)

The ability of a model of fresh data is measured by the average age (AA) of that model. The average age is denoted by $\bar{k}$. The smaller $\bar{k}$ is much better to compare the model. The equation of the average age metric is defined by Brown (1959) as follows:

$$\mathrm{AA}\ (\hat{a}) = \bar{k} = n - \sum_{t=1}^{n} t w_t \tag{3.17}$$

where w_t is the weight for the given t^{th} observation while trying to obtain the forecast. For the simple form of *ATA*, the average age is as follows:

$$AA_{ATA(p,0)} = \bar{k}_{ATA} = \frac{n-p}{p+1} \tag{3.18}$$

If we compare the average age of simple form of ATA with that of simple exponential smoothing, we should look the average age of SES with the same α level which is $\alpha = {}^{p}/_{n}$,

$$AA_{SES} = \bar{k}_{SES} = \frac{1-\alpha}{\alpha} = \frac{1-{}^{p}/_{n}}{{}^{p}/_{n}} = \frac{n-p}{p} \tag{3.19}$$

If we compare these two formulation, we will see that average age of ATA is smaller than average age of SES at the same α level. Because:

$$\bar{k}_{ATA} = \frac{n-p}{p+1} < \frac{1-\alpha}{\alpha} = \frac{n-p}{p} = \bar{k}_{SES} \tag{3.20}$$

3.1.3.3 The Sum of Squared Weights of ATA (p,0)

When we comparing the forecasting methods, variance can also be used as an important metric. The error component form of X_t is $X_t = a + e_t$, where e_t is some random noise with mean zero and variance σ ; and $\hat{a} = \sum_{t=1}^{n} w_t X_t$, where w_t are weights . Since the estimator of the equation of $\hat{a}$ will be unbiased, its variance can be written as given below:

$$\mathrm{Var}(\hat{a}) = E\left[\left(\sum_{t=1}^{n} w_t X_t - a\right)^2\right] = \sum_{t=1}^{n} w_t^2 \sigma_t^2 = V\sigma^2 \tag{3.21}$$

In order to calculate the variance of the *ATA* estimator, it is required to calculate the sum of squared weights denoted by *V*. The sum can be written as:

$$V_{ATA(p,0)} = \sum_{t=1}^{n} w_t^2 \tag{3.22}$$

where the weights are given in Table 3.1. If the weighted values are substituted in equation (3.22), we obtain:

$$V_{ATA(p,0)} = \left(\frac{p}{n}\right)^2 + \left(\frac{p}{n}\right)^2 \left(\frac{n-p}{n-1}\right)^2 + \cdots + \left(\frac{p}{n}\right)^2 \left(\frac{n-p}{n-1}\right)^2 \left(\frac{n-p-1}{n-2}\right)^2 \cdots \left(\frac{1}{p}\right)^2$$

$$= \left(\frac{p}{n}\right)^2 \left[1 + \sum_{i=0}^{n-p-1} \prod_{j=0}^{i} \left(\frac{n-p-j}{n-1-j}\right)^2\right] \quad (3.23)$$

$$= \left(\frac{p}{n}\right)^2 {}_3F_2((1, p-n, p-n), (1-n, 1-n), 1) \quad (3.24)$$

From the equation (3.24), it can be seen that the variance of the *ATA* (p, *0*) estimator involves the Generalized Hyper-Geometric series:

$${}_3F_2((1, p-n, p-n), (1-n, 1-n), 1)$$

This model can be easily adjusted to incorporate higher order components. Note that when:

$$\bar{k}_{ATA} = \bar{k}_{SES} \Rightarrow \frac{n-p}{p+1} = \frac{1-\alpha}{\alpha} \Rightarrow \alpha = \frac{p+1}{n+1} \quad (3.25)$$

and also we obtained that the average age of *ATA* is equivalent to average age of *SES:*

$$\bar{k}_{ATA} = \bar{k}_{SES} \Leftrightarrow \alpha = \frac{p+1}{n+1} \quad (3.26)$$

then the variance of *ATA* is always smaller than the variance of *SES* at the same α level ($V_{ATA(p,0)} < V_{SES}$). When the same smoothing constant is utilized for both models, since *ATA* always has a smaller average age (*AA*), the sum of squared weights value of *ATA* will be greater than that of *SES* (Taylan Selamlar, 2017).

3.1.3.4 The Weight of the Initial Value of the ATA (p, 0)

Unlike the some of other forecasting methods, *ATA* method does not require an initial value. The smoothing equation is as:

$$\mathrm{S}_t = \sum_{k=0}^{t-(p+1)} \frac{\binom{t-k-1}{p-1}}{\binom{t}{p}} X_{t-k} + \frac{1}{\binom{t}{p}} S_p \tag{3.27}$$

$$\mathrm{S}_p = X_t \quad when \quad t \leq p \tag{3.28}$$

where the smoothed values for time points before *p* are equal to the actual observations themselves. Also, the same recursive formula for *SES* is:

$$S_t = \alpha \sum_{k=0}^{t-1} (1-\alpha)^k X_t + (1-\alpha)^t S_0 \tag{3.29}$$

where S_0 refers to the initial value. Weights attached to initial values S_p and S_0 at time *t* are successively equal to $\frac{1}{\binom{t}{p}}$ for *ATA (p,0)* and $(1-\alpha)^t$ for *SES*. For *SES*, the most scientist study on the issue with α values between 0.01 and 0.3. However; it is also known as the initialization problem, when either *t* or α is small, *SES* attaches more weight to initial value than the most recent observation. Hence, the choice of starting value becomes quite important for *SES* (Ekiz Yılmaz, 2018).

3.1.3.5 The Smoothing Parameter of ATA (p, 0)

The main aim of choosing a smoothing parameter is that obtain forecasted values which more represent the observation values and also minimum error. For instance, the main idea in *SES* which is the one of the most prevalent techniques is that the recent observation is more symbolized of the next values and hence more emphasis should be given the last observations. Similarly, in average method the starting point for a grid search should be weighting all past observations equally and also in naïve method gives greater emphasis to recent observations slowly until finishing up by weighting the last observation by 1. This would warranty that the weight of the initial value stays less than or equal to the weight of the most recent observations (Ekiz Yılmaz, 2018). This can easily be obtained with an *ATA* model with $p=1$ for any t as below:

$$S_t = \left(\frac{1}{t}\right) X_t + \left(\frac{t-1}{t}\right) S_{t-1} \tag{3.30}$$

$$S_t = \left(\frac{1}{t}\right) X_t + \left(\frac{1}{t}\right) X_{t-1} + \cdots + \left(\frac{1}{t}\right) X_2 + \left(\frac{1}{t}\right) S_1 \tag{3.31}$$

where $S_1 = X_1$ since $S_p = X_p$ when $t \le p$ and the h step forecast $(\hat{X}_t(h) = \overline{X}_t)$ is equal to the simple average of all past observations, namely, the whole observations contribute equally to the forecast (Ekiz Yılmaz, 2018).

For *p=2* the ATA smoothed value at time *t* can be written as:

$$S_t = \left(\frac{2}{t}\right) X_t + \left(\frac{t-2}{t}\right) S_{t-1} \tag{3.32}$$

$$S_t = \frac{2}{t} X_t + \frac{2}{t}\frac{(t-2)}{(t-1)} X_{t-1} + \frac{2}{t}\frac{(t-2)}{(t-1)} X_{t-2} + \cdots + \frac{2}{t}\frac{2}{(t-1)} X_3 + \frac{2}{t(t-1)} S_2$$

where $S_2 = X_2$. When $p = 2$ for any *t*, the *ATA* model produces weights that decrease linearly with slope $2/t(t-1)$ and intercept $2/t$ which again can never be achieved by *SES* since it always assigns exponentially decreasing weights to observations no matter the parameter choice. Likewise, for $p \ge 3$ the weights start to decrease exponentially as the observations get older as in *SES* but not exactly at the same rate. In this case, *ATA* gives greater emphasis than *SES* to the most recent history and less emphasis than *SES* to the most distant past at the same smoothing constant.

Not only *ATA* (*p, 0*) is more flexible but also it is more adaptive to the data set. The observations are assigned the weights $\alpha, \alpha(1-\alpha), \alpha(1-\alpha)^2, \ldots$ regardless of the sample sizes at hand. This is not the case for *ATA* as the weights change with respect to the sample size as the weights are below respectively:

$$\left(\frac{p}{n}\right), \left(\frac{p}{n}\right)\left(\frac{n-p}{n-1}\right), \left(\frac{p}{n}\right)\left(\frac{n-p}{n-1}\right)\left(\frac{n-p-1}{n-2}\right), \ldots \tag{3.33}$$

For the sake of simplicity and demonstration choose $p=3$, then the smoothing formula will be;

$$S_t = \left(\frac{3}{t}\right) X_t + \left(\frac{t-3}{t}\right) S_{t-1}, \qquad if \quad t > 3 \tag{3.34}$$

If we expand the formula iteratively, we get:

$$S_4 = \left(\frac{3}{4}\right) X_4 + \left(\frac{1}{4}\right) S_3 \tag{3.35}$$

$$S_5 = \left(\frac{3}{5}\right) X_5 + \left(\frac{2}{5}\right)\left(\frac{3}{4}\right) X_4 + \left(\frac{2}{5}\right)\left(\frac{1}{4}\right) S_3 \tag{3.36}$$

$$S_6 = \left(\frac{3}{6}\right) X_6 + \left(\frac{3}{6}\right)\left(\frac{3}{5}\right) X_5 + \left(\frac{3}{6}\right)\left(\frac{2}{5}\right)\left(\frac{3}{4}\right) X_4 + \left(\frac{3}{6}\right)\left(\frac{2}{5}\right)\left(\frac{1}{4}\right) S_3 \tag{3.37}$$

and go on. If we keep on going up to $t=10$ then we end up following smoothing equation with three digit calculations.

$$\begin{aligned} S_{10} = {} & 0.300X_{10} + 0.233X_9 + 0.175X_8 + 0.125X_7 + 0.083X_6 + 0.050X_5 \\ & + 0.025X_4 + 0.008S_3 \end{aligned} \tag{3.38}$$

where $S_3 = X_3$.

3.2 The M-Competitions and M4-Competition

The M-Competitions are experimental studies that the experts are competing with the forecasts that obtained with the forecasting method supported by them. The performance of that forecasting methods are compared in the M-Competitions (Makridakis Competition) for the large number of major time series. Forecasts from each expert are evaluated after being received and compared with those of other experts and compared with some simple methods used as benchmarks. Forecasting competitions guarantee both neutrality and expertise knowledge (M-Competitions, n.d.).

3.2.1 History of M-Competitions

The beginning of the M-competition dates back to 1979, Makridakis and Hibon selected 111 time series, including data of business firms, industry and macro data as yearly, quarterly, monthly. It was the establishment of the idea of forecasting competitions. There were 22 methods and the major finding was simple methods did better than sophisticated ones (Spiliotis, Makridakis, & Assimakopoulos, 2018).

The first M-competition created in 1982. In this competition, 1001 time series and 15 forecasting methods were used. Based on the papers written by the researchers about the results of the competition, the main conclusions of the M-Competition were:

1. Simple methods give more accurate results than statistically complex methods.
2. The performance of the methods changes relative to the accuracy measure used.
3. Generally, when several methods are combined accuracy increases compared to individual methods.
4. The length of the forecasting horizon affects the accuracy of the used method (M-Competitions, n.d.).

Makridakis and Hibon (Makridakis and Hibon, 1979) published in the Journal of the Royal Statistical Society (JRSS) an article before the first M-Competition, proving that simple methods give more accurate results compared to the more complex and statistically sophisticated methods. It was criticized by statisticians about that these arguments could not be possible. Consecutive M2 competition and M3 competition were motivated by their criticism, as a result of these competitions Makridakis and Hibon's study was proved. (M-Competitions, n.d.).

In the M2-Competition, which was the second competition, bigger data set was used. The M2 competition, which includes six macroeconomic series and is held in real time, was held jointly with four companies. The conclusions of the M2 were printed in 1993 in a study. According to this study, the results were similar to M-Competition (M-Competitions, n.d.). Finally, in most cases forecasters did not improve on the accuracy of mechanical methods. The best methods were found to be simplest: Dampen and Simple exponential smoothing and combining of smoothing methods (Makridakis et al. 1993).

The third competition called the M-3 Competition or M3-Competition, was intended to both replicate *and* extend the features of the M-competition and M2-Competition, through the inclusion of more methods and researchers (particularly researchers in the area of neural networks) and more time series. A total of 3003-time series was used. The data set was included yearly, quarterly, monthly, daily, and other time series. The sort of time series were: micro, industry, macro, finance, demographics, and other. sMAPE, median symmetric APE, average ranking, median RAE and percentage better were the five accuracy measure that used in the competition (M-Competitions, n.d.). This M3-Competition confirmed the original conclusions of M-Competition using a new and much enlarged set o data. In addition, it demonstrated, once more that simple methods that developed by practicing forecasters (e.g. Brown's Single and Gardner's Dampen Trend Exponential Smoothing) do as well, or in many cases better than statistically sophisticated ones like ARIMA and ARARMA models (Makridakis & Hibon, 2000).

3.2.2 M4-Competition

The fourth competition was M4-Competition. It started in Jan 1 2018, ended in May 31 2018. The M4-competition improved the results of the previous three competitions, used the most accurate methods and extended data set for different estimates for various types of forecasting. The main aim was how to improve forecasting accuracy and classify the most suitable methods for each condition. 100,000 real-life time series used in M4-competition. (M-Competitions, n.d.).

The M4-Competition series, as those of the M1 and M3, aim at representing the real world as much as possible. The series were selected randomly from a database of 900.000. The 100,000 time series of the dataset come mainly from the Economic, Finance, Demographics and Industry areas, while also including data from Tourism, Trade, Labor and Wage, Real Estate, Transportation, Natural Resources and the Environment. The minimum number of observations is 13 for yearly, 16 for quarterly, 42 for monthly, 80 for weekly, 93 for daily and 700 for hourly series (M-Competitions, n.d.).

Economic data or economic statistics are data (quantitative measures) describing an actual economy, past or present. A collection of such data in table form comprises a data set. Methodological economic and statistical elements of the subject include measurement, collection, analysis, and publication of data (Wikipedia, n.d.).

Industry data helps in knowing the current position of a business in the market. This data can give us overall analysis of an industry’s economic, legal, and market status (iResearch Services, 2018).

Demographic data is statistical data collected about the characteristics of the population, e.g. age, gender and income for example. It is usually used to research a product or service and how well it is selling, who likes it and/or in what areas it is most popular (Bigwave media, n.d.).

As we mentioned before a financial data, which is the main subject of this study, consists of pieces or sets of information related to the financial health of a business. The pieces of data are used by internal management to analyse business performance and determine whether tactics and strategies must be altered. People and organizations outside a business will also use financial data reported by the business to judge its credit worthiness, decide whether to invest in the business, and determine whether the business is complying with government regulations (Study.com, n.d.).

Table 3.2 The number of data set of M4 competition (M-Competitions, n.d.)

Frequency	Demographic	Finance	Industry	Macro	Micro	Other	Total
Yearly	1,088	6,519	3,716	3,903	6,538	1,236	23,000
Quarterly	1,858	5,305	4,637	5,315	6,020	865	24,000
Monthly	5,728	10,987	10,017	10,016	10,975	277	48,000
Weekly	24	164	6	41	112	12	359
Daily	10	1,559	422	127	1,476	633	4,227
Hourly	0	0	0	0	0	414	414
Total	8,708	24,534	18,798	19,402	25,121	3,437	100,000

The conclusions according to sMAPE values of benchmark methods and the most successful five methods are given in order in the below table:

Table 3.3 Conclusions of benchmarks and most successful five methods in M4

Method	sMAPE						
User ID	Yearly	Quarterly	Monthly	Weekly	Daily	Hourly	Total
118 (Hybrid)	13.176	9.679	12.126	7.817	3.170	9.328	11.374
72 (Combination (S & ML))	13.712	9.809	12.487	6.814	3.037	9.934	11.695
245 (Combination (S & ML))	13.528	9.733	12.639	7.625	3.097	11.506	11.720
69 (Combination (S))	13.673	9.816	12.737	8.627	2.985	15.563	11.836
237 (Combination (S))	13.943	9.796	12.747	6.919	2.452	9.611	11.845
Theta	14.593	10.311	13.002	9.093	3.053	18.138	12.309
Com	14.848	10.175	13.434	8.944	2.980	22.053	12.555
ARIMA	15.168	10.431	13.443	8.653	3.193	12.045	12.661
Damped	15.198	10.237	13.473	8.866	3.064	19.265	12.661
ETS	15.356	10.291	13.525	8.727	3.046	17.307	12.725
SES	16.396	10.600	13.618	9.012	3.045	18.094	13.087
Naive2	16.342	11.012	14.427	9.161	3.045	18.383	13.564
Holt	16.354	10.907	14.812	9.708	3.066	29.249	13.775
Naive	16.342	11.610	15.256	9.161	3.045	43.003	14.208
sNaive	16.342	12.521	15.988	9.161	3.045	13.912	14.657
RNN	22.398	17.027	24.056	15.220	5.964	14.698	21.152
MLP	21.764	18.500	24.333	21.349	9.321	13.842	21.653

The significant findings of M4-Competition according to Makridakis (2018) are stated as:

- 12 of the 17 most accurate methods consisted of a combination of statistical approaches. So, it could be said that the combined methods was the king of the competition.
- The "hybrid" approach, which is a combination of both statistical and ML (machine learning) features, was very successful than it was thought. This method made the most accurate forecasts. Depending on the accuracy measure sMape, it was close to 10% more accurate than the combination (Comb) benchmark of the competition, which is a great development.
- The second most successful method was a combination of seven statistical methods and one ML, average weights were calculated with an ML algorithm trained to minimize estimation error by removal tests.

- Unfortunately, six ML methods performed badly in the competition. These methods were even worse than Naive2 method (Makridakis, Spiliotis, & Assimakopoulos, 2018).

According to above findings, we can say that combinations of the methods and the hybrid models, give more accurate results than individual ML methods or statistical methods (Makridakis et al., 2018).

The other section 3.3 will be about appliaction of ATA-Method to M4-Competition data set and its conclusions. We will apply this forecasting method for all types of data but we will focus more on financial data.

3.3 Application of ATA Method to M4-Competition Data Set

ATA method is an innovative new forecasting technique where the forms of the models are similar to exponential smoothing methods. But the smoothing parameters of ATA method which depends on the sample size are optimized on a discrete space and initialization is both easier as it is done simultaneously when the parameters are optimized and is less influential since the weights assigned to initial values approach quickly. In addition, ATA can be applied to all time series settings easily and provides better forecasting performance due to its flexibility (Taylan, 2018).

Obtaining the point forecasts, the data sets were tested for stationarity using the Augmented Dickey-Fuller test using the function *ndiffs* from the *forecast* package in R-Studio. Then, the data sets were tested for seasonality using Seasonality Test function from the *4Thetamethod.R* code download from the competitions GitHub repository for $\alpha = 0.20$ and seasonality periods 4 for quarterly, 12 for monthly, 7 for daily. After that, this data sets that could be classified as seasonal by this test are de-seasonalized by the classical multiplicative decomposition method. The hourly and weekly data were treated slightly different as they have not been put through the seasonality test and all have been de-seasonalized using the classical multiplicative decomposition method for periods 168 and 52 respectively. In addition, there were various arbitrary models that were used for different data types such as for yearly data $ATA_{add}(p, 1, \emptyset)$ was used to obtain forecasts where $\emptyset \in \{0.60, 0.61, \cdots, 1\}$. Similarly; for all other data types, the simple models $AT(p,\ 0,\ \emptyset)$ and $ATA_{add}(p, 1, 1)$ were obtained and averaged. Finally, the forecasts were re-seasonalized using the seasonality indexes from the classical multiplicative decomposition to obtain final point forecasts. If any negative forecasts were obtained they were set equal to zero (Taylan, 2018).

In the Table 3.4, ATA method, ES method, ARIMA, THETA method are applied to M4-Competition data set with R software. With R code of each method, we obtain the forecasting results for the data set. After that, we use symmetric mean absolute percentage error (sMAPE) as accuracy measure for the forecasting results. Then we calculate the average of this sMAPE's for all observations in the Table 3.4 given below:

Table 3.4 Conclusion of 4 methods based on average of sMAPE in M4 data set

Method	Data	Demographic	Finance	Industry	Macro	Micro	Other	Total
ATA	Daily	6.355	**3.501**	3.890	2.607	2.398	3.234	3.095
	Hourly	-	-	-	-	-	12.851	12.851
	Monthly	4.822	**14.142**	13.741	13.790	14.447	13.020	12.936
	Quarterly	10.164	11.804	9.272	9.956	10.551	7.018	10.292
	Weekly	12.306	4.284	6.933	12.311	12.367	11.368	8.540
	Yearly	10.555	**14.948**	17.464	14.735	10.895	14.411	13.930
	Total	6.700	**13.109**	13.151	12.853	11.872	10.181	**12.098**
ES	Daily	6.340	**3.521**	3.918	2.616	2.361	2.925	3.046
	Hourly	-	-	-	-	-	17.307	17.307
	Monthly	4.860	**14.627**	13.675	14.285	16.130	12.906	13.525
	Quarterly	10.402	11.791	9.401	9.982	10.352	7.088	10.291
	Weekly	22.847	3.671	7.490	15.334	10.096	14.845	8.727
	Yearly	10.647	**16.525**	19.036	15.607	12.715	15.450	15.356
	Total	6.817	**13.739**	13.460	13.298	13.021	11.055	**12.725**
ARIMA	Daily	7.044	3.747	3.948	2.939	2.441	3.068	3.193
	Hourly	-	-	-	-	-	13.980	13.980
	Monthly	4.834	14.367	13.851	14.326	15.844	13.028	13.443
	Quarterly	10.428	12.085	9.519	10.046	10.462	7.331	10.431
	Weekly	20.655	4.649	9.777	13.016	9.870	12.552	8.653
	Yearly	11.027	15.840	19.235	15.842	12.488	15.094	15.168
	Total	6.847	13.525	13.623	13.381	12.866	10.616	**12.669**
THETA	Daily	6.536	3.503	3.889	2.592	2.402	2.941	3.053
	Hourly	-	-	-	-	-	18.138	18.138
	Monthly	4.928	14.114	13.703	13.843	14.695	13.094	13.002
	Quarterly	10.121	11.772	9.280	9.971	10.609	7.297	10.311
	Weekly	22.458	3.870	13.292	16.297	10.215	16.551	9.093
	Yearly	10.454	15.908	17.859	15.064	11.794	14.806	14.593
	Total	6.776	13.341	13.213	12.960	12.219	11.000	**12.309**

In the table 3.4, ATA method gives the better result in total with 12.098. This result means that ATA method is more accurate forecasting technique than the other techniques. After ATA method, other methods can be listed due to their accuracy in total as Theta is the second, ARIMA third, ES method is the fourth one.

In the Table 3.5, similarly we obtain the average of mean absolute scaled error (MASE) results given below:

Table 3.5 Conclusion of 4 methods based on average of MASE in M4 data set

Method	Data	Demographic	Finance	Industry	Macro	Micro	Other	Total
ATA	Daily	9.898	**3.502**	3.831	3.291	2.713	3.564	3.277
	Hourly	-	-	-	-	-	2.238	2.238
	Monthly	0.910	**0.990**	1.027	0.988	0.879	0.919	0.962
	Quarterly	1.271	1.228	1.162	1.269	1.272	1.007	1.231
	Weekly	1.640	3.116	1.824	2.890	1.930	2.469	2.578
	Yearly	2.580	**3.399**	3.217	3.148	2.888	2.919	3.117
	Total	1.208	**1.855**	1.556	1.519	1.608	2.312	**1.631**
ES	Daily	9.753	**3.411**	3.854	3.318	2.693	3.561	3.239
	Hourly	-	-	-	-	-	1.783	1.783
	Monthly	0.896	**0.976**	0.982	0.951	0.910	0.832	0.947
	Quarterly	1.247	1.152	1.145	1.180	1.166	0.930	1.160
	Weekly	2.745	2.571	1.939	2.831	2.038	5.308	2.527
	Yearly	2.621	**3.612**	3.417	3.344	3.517	3.091	3.433
	Total	1.202	**1.880**	1.568	1.515	1.760	2.302	**1.676**
ARIMA	Daily	11.801	3.619	3.916	3.793	2.801	3.635	3.390
	Hourly	-	-	-	-	-	0.942	0.942
	Monthly	0.883	0.954	0.991	0.933	0.872	0.850	0.930
	Quarterly	1.272	1.161	1.148	1.178	1.166	0.944	1.164
	Weekly	2.528	3.125	2.851	2.508	1.660	3.197	2.555
	Yearly	2.586	3.507	3.463	3.378	3.471	2.997	3.397
	Total	1.196	1.861	1.585	1.514	1.736	2.178	**1.663**
THETA	Daily	9.952	3.489	3.826	3.242	2.703	3.533	3.262
	Hourly	-	-	-	-	-	2.455	2.455
	Monthly	0.931	0.994	1.025	0.997	0.890	0.925	0.970
	Quarterly	1.267	1.226	1.167	1.263	1.277	1.036	1.232
	Weekly	2.742	3.082	4.618	2.648	1.758	3.531	2.637
	Yearly	2.650	3.708	3.362	3.391	3.244	3.070	3.382
	Total	1.233	1.938	1.586	1.570	1.706	2.398	**1.696**

According to average MASE results given in the Table 3.5, ATA method has the smallest MASE in total with 1.631. This means that ATA method gives the more accurate result than the other methods. Other methods can be ordered as follows; ARIMA is the second one, ES is the third one and Theta is the last one, acoording to MASE results. In detail, we can say that ATA method performs better than ES method.

In the Table 3.6, similarly we obtain the average of mean absolute percentage error (MAPE) results given below:

Table 3.6 Conclusion of 4 methods based on average of MAPE in M4 data set

Method	Data	Demographic	Finance	Industry	Macro	Micro	Other	Total
ATA	Daily	5.873	**6.196**	5.275	2.585	2.538	3.158	4.262
	Hourly	-	-	-	-	-	14.541	14.541
	Monthly	4.935	**16.364**	18.466	16.065	16.757	16.068	15.465
	Quarterly	11.991	13.462	10.792	11.866	11.697	7.197	11.810
	Weekly	13.127	4.244	7.518	13.716	11.723	12.475	8.583
	Yearly	10.561	**18.328**	19.363	16.044	13.328	15.552	16.170
	Total	7.167	**15.531**	16.451	14.817	13.794	11.076	**14.248**
ES	Daily	5.862	**4.570**	5.396	2.602	2.511	2.931	3.632
	Hourly	-	-	-	-	-	23.612	23.612
	Monthly	4.949	**17.387**	18.376	17.126	19.788	16.114	16.597
	Quarterly	12.242	13.623	11.128	12.090	11.636	7.076	11.960
	Weekly	18.959	3.640	8.038	13.984	10.100	13.809	8.274
	Yearly	10.845	**19.757**	20.112	16.826	14.843	16.661	17.332
	Total	7.281	**16.296**	16.637	15.584	15.489	12.503	**15.104**
ARIMA	Daily	6.652	4.420	5.376	2.881	2.581	3.082	3.632
	Hourly	-	-	-	-	-	18.354	18.354
	Monthly	4.941	16.929	18.483	16.426	18.436	15.776	16.056
	Quarterly	11.955	14.209	11.066	12.960	11.837	7.631	12.319
	Weekly	17.284	4.316	9.607	11.983	9.931	11.746	8.147
	Yearly	11.255	19.821	20.993	18.767	14.762	16.519	17.811
	Total	7.262	16.230	16.853	15.850	14.929	11.952	**15.018**
THETA	Daily	6.061	6.285	5.274	2.573	2.546	2.934	4.264
	Hourly	-	-	-	-	-	21.689	21.689
	Monthly	5.054	16.576	18.262	16.170	17.088	15.805	15.581
	Quarterly	11.697	13.449	10.684	12.047	11.768	7.507	11.833
	Weekly	18.600	3.830	14.832	14.195	10.148	13.977	8.495
	Yearly	10.373	18.805	18.810	16.556	13.816	15.821	16.446
	Total	7.174	15.753	16.209	15.025	14.076	12.054	**14.402**

As we can see in the Table 3.6, ATA method also has the smallest MAPE value with 14.248. It is easy to understand that ATA method also gives good results due to MAPE values. After ATA method, Theta, ARIMA and ES have good results in order.

On the other hand, ATA method gives good performance with the combination with ARIMA. In the given Tables 3.7, 3.8 and 3.9 below, we show the results of combinations of ATA method and ES method with ARIMA:

Table 3.7 Conclusion of combination methods based on average of sMAPE results in M4 data set

Method	Data	Demographic	Finance	Industry	Macro	Micro	Other	Total
ATA/ARIMA	Daily	6.294	**3.326**	3.880	2.726	2.387	3.026	2.998
	Hourly	-	-	-	-	-	11.942	11.942
	Monthly	4.613	**13.749**	13.254	13.388	14.537	12.504	12.653
	Quarterly	10.038	11.466	9.091	9.698	10.066	6.838	9.987
	Weekly	11.240	3.937	7.482	11.252	10.575	10.401	7.607
	Yearly	10.376	**14.733**	17.526	14.504	10.991	14.206	13.847
	Total	6.511	**12.789**	12.859	12.527	11.811	9.869	**11.859**
ES/ARIMA	Daily	6.345	**3.554**	3.900	2.754	2.383	2.978	3.076
	Hourly	-	-	-	-	-	14.377	14.377
	Monthly	4.690	**14.009**	13.270	13.528	15.244	12.580	12.917
	Quarterly	10.176	11.495	9.153	9.702	10.105	6.843	10.027
	Weekly	21.684	3.913	8.344	13.928	9.710	13.253	8.439
	Yearly	10.461	**15.581**	18.420	15.095	12.110	14.888	14.691
	Total	6.631	**13.152**	13.061	12.725	12.416	10.417	**12.205**

Table 3.8 Conclusion of combination methods based on average of MASE results in M4 data set

Method	Data	Demographic	Finance	Industry	Macro	Micro	Other	Total
ATA/ARIMA	Daily	9.376	**3.443**	3.795	3.499	2.703	3.570	3.255
	Hourly	-	-	-	-	-	1.437	1.437
	Monthly	0.862	**0.934**	0.970	0.916	0.845	0.842	0.908
	Quarterly	1.232	1.144	1.118	1.172	1.157	0.939	1.148
	Weekly	1.509	2.831	2.075	2.596	1.692	2.744	2.345
	Yearly	2.487	**3.286**	3.208	3.121	2.963	2.863	3.093
	Total	1.155	**1.777**	1.513	1.450	1.584	2.174	**1.575**
ES/ARIMA	Daily	9.493	**3.434**	3.813	3.537	2.712	3.578	3.259
	Hourly	-	-	-	-	-	1.250	1.250
	Monthly	0.863	**0.938**	0.959	0.906	0.863	0.809	0.909
	Quarterly	1.233	1.125	1.119	1.148	1.137	0.904	1.132
	Weekly	2.633	2.765	2.340	2.626	1.804	4.029	2.476
	Yearly	2.538	**3.469**	3.362	3.290	3.407	2.986	3.334
	Total	1.166	**1.822**	1.538	1.473	1.704	2.190	**1.627**

Table 3.9 Conclusion of combination methods based on average of MAPE results in M4 data set

Method	Data	Demographic	Finance	Industry	Macro	Micro	Other	Total
ATA/ARIMA	Daily	5.844	**4.936**	5.288	2.687	2.528	3.016	3.777
	Hourly	-	-	-	-	-	14.799	14.799
	Monthly	4.730	**16.143**	17.919	15.566	16.995	15.375	15.222
	Quarterly	11.683	13.344	10.606	12.063	11.332	7.065	11.672
	Weekly	10.412	3.834	7.849	11.384	10.328	10.418	7.449
	Yearly	10.512	**18.433**	19.440	16.578	13.390	15.516	16.316
	Total	6.953	**15.352**	16.129	14.716	13.820	10.971	**14.108**
ES/ARIMA	Daily	5.898	**4.403**	5.354	2.719	2.528	2.986	3.584
	Hourly	-	-	-	-	-	19.497	19.497
	Monthly	4.789	**16.670**	17.956	16.019	18.377	15.550	15.769
	Quarterly	11.843	13.459	10.776	12.194	11.428	7.022	11.794
	Weekly	18.081	3.810	8.616	12.761	9.753	12.353	8.006
	Yearly	10.683	**19.298**	19.968	17.178	14.400	16.218	17.081
	Total	7.069	**15.809**	16.297	15.111	14.707	11.794	**14.589**

In the Tables 3.7, 3.8 and 3.9, combination of ATA/ARIMA gives smaller average of sMAPE, MASE and MAPE than the combination of ES/ARIMA in total results. This means that in total results, combination of ATA/ARIMA performs better than ES/ARIMA.

3.3.1 Financial Data Analysing by Exponential Smoothing Method and ATA Method

In Table 3.10 below, we compare the average sMAPE, MASE and MAPE results for ATA, ES, ARIMA, THETA and combinations of ATA/ARIMA and ES/ARIMA for M4 financial data set:

Table 3.10 Conclusion of methods in M4 financial data set

Accuracy Measure	Data	ATA	ES	ARIMA	THETA	ATA/ARIMA	ES/ARIMA
sMAPE	Daily	3.501	**3.521**	3.747	3.503	3.326	3.554
	Monthly	14.142	**14.627**	14.367	14.114	13.749	14.009
	Quarterly	11.804	11.791	12.085	11.772	11.466	11.495
	Weekly	4.284	3.671	4.649	3.870	3.937	3.913
	Yearly	14.948	**16.525**	15.840	15.908	14.733	15.581
	Total	13.109	**13.739**	13.525	13.341	12.789	**13.152**
MASE	Daily	3.502	**3.411**	3.619	3.489	3.443	3.434
	Monthly	0.990	**0.976**	0.954	0.994	0.934	0.938
	Quarterly	1.228	1.152	1.161	1.226	1.144	1.125
	Weekly	3.116	2.571	3.125	3.082	2.831	2.765
	Yearly	3.399	**3.612**	3.507	3.708	3.286	3.469
	Total	1.855	**1.880**	1.861	1.938	1.777	**1.822**
MAPE	Daily	6.196	4.570	4.420	6.285	4.936	4.403
	Monthly	16.364	**17.387**	16.929	16.576	16.143	16.670
	Quarterly	13.462	**13.623**	14.209	13.449	13.344	13.459
	Weekly	4.244	3.640	4.316	3.830	3.834	3.810
	Yearly	18.328	**19.757**	19.821	18.805	18.433	19.298
	Total	15.531	**16.296**	16.230	15.753	15.352	**15.809**

If we want to compare ATA method and ES method in M4 financial data according to Table 3.10:

1) ATA method gives better performance than ES in daily, monthly and yearly data due to sMAPE results. ES method is better than ATA method in quarterly and weekly data. But in total, ATA method is better than ES method.
2) According to MASE results, ATA method makes accurate forecasts than ES in daily, monthly and yearly data. Otherwise, ES is better than ATA method in quarterly and weekly data. But in total ATA method makes accurate forecasts.
3) Similarly for the MAPE results, it can be seen that ATA method has the good conclusions in monthly, quarterly and yearly data. ES method is better than ATA method in daily and weekly data. But in total, ATA method is has good performance than ES method.

4) For the results of this three accuracy measures, combination of ATA/ARIMA gives better performance than combination of ES/ARIMA in total results.
5) ATA method is also the most accurate method according to other three benchmark methods: ES, ARIMA and THETA in total results.

This means that in financial data analysis ATA method performs better than ES method. ATA method gives more accurate forecasting results than ES. This results support our suggestion that ATA method can be used as an alternative forecasting method to ES. In addition it gives more accurate results than ES.

CHAPTER FOUR

SUMMARY AND CONCLUSION

Forecasting has become very critical in many areas in today's world. From economy to medicine, forecasting methods are used for making better decisions for future. So that to make right decisions the forecasting method which is used should be the most accurate one.

There are many forecasting techniques in the literature. However, the accuracy of these methods varies depending on types and components of data. Many studies show that simple forecasting methods generally give more accurate results than complex ones.

Although there are many forecasting methods in literature, the most widely used and well known is exponential smoothing method based on its simplicity and accuracy. Even though it has positive features, there some problems about ES methods. One of these problems is initial value and the second one is smoothing parameter. Although ES is widely used method, there is no rule for choosing initial value and smoothing parameter that generally accepted by forecasters. The accuracy of forecasts obtained by ES affect from these two values.

In this thesis, we propose a new forecasting method, which is called as ATA method, to overcome these disadvantages of the well-known forecasting method ES. ATA method does not require an initial value so it solves the initial value problem of ES. Another deficiency about initial value of ES is weights. More weights are assigned to initial value of ES than ATA method. As a consequence of this initial value affects the forecasts in ES more than ATA method. But it is pathetic that there is no general rule of choosing initial value for ES, this is only a subjective decision. ATA method brings solution to this problem. Moreover, ATA method eliminates the selection of smoothing parameter of ES. Its smoothing parameter depends on the equation of the time series.

In chapter 3, we compared ATA method with other forecasting methods by applying them to M4-competitions data set. The forecasts made by R software and then sMAPE, MASE and MAPE values calculated for forecasts as accuracy measure. Then when we compared results, we saw that ATA method performs better than the other forecasting methods.

In addition, the combination of ATA/ARIMA gave better result than the combination of ES/ARIMA. This shows that the combination of ATA and ARIMA works better than combination of ES and ARIMA.

Finally, in accordance with the purpose of this study, ATA method, ES method, ARIMA, THETA and the combinations of ATA/ARIMA and ES/ARIMA were applied to M4-competition financial data set. sMAPE, MASE and MAPE values for this forecasting were calculated and two methods compared based on these values. We saw that ATA method performed better than ES method in total results for this three accuracy measures. In addition to its strong performance against ES, ATA method gives accurate results than ARIMA and THETA. Also the combination of ATA/ARIMA gives the best result in total among the results of ATA, ES, ARIMA, THETA and ES/ARIMA in M4 financial data.

As a consequence; there is something that these studies tell us, the new forecasting method that we proposed ATA makes very successful forecasts. The combination of ATA and ARIMA also looks great and makes the results more accurate. This paper and analysis proves that ATA method will be a very useful and valuable forecasting method in the literature. With respect to its good performance, ATA method will replace the well-known forecasting method ES because of its deficiencies.

REFERENCES

Analysis of financial statements. (n.d.). Retrieved October 22, 2018, from http://ncert.nic.in/ncerts/l/leac204.pdf

Anonymus. (n.d.). *Eumetrain*. Retrieved December 12, 2018, from http://www.eumetrain.org/data/4/451/english/msg/ver_cont_var/uos3/uos3_ko1.htm

Arlt, J., & Arltova, M. (2001). Financial time series and their features. *Acta Oeconomica Pragensia, 9*(4), 7-20. Retrieved December 24, 2018, from https://nb.vse.cz/~arlt/publik/AA_FTSTF_01.pdf

Armstrong, J. (2001). Introductıon. In J. Armstrong, *Principles of forecasting* (1-12). Dordrecht: Kluwer Academic Publishers.

Bigwave media. (n.d.). *3 uses of demographic data*. Retrieved March 24, 2019, from Bigwave media web site: https://www.bigwavemedia.co.uk/blog/uses-of-demographic-data

Brownlee, J. (2017, January 9). *How to create an ARIMA model for time series forecasting in python*. Retrieved January 17, 2019, from Machine learning mastery: https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/

Çapar, S. (2009). *Improvement for exponential smoothing.* Phd Thesis, Dokuz Eylül University, İzmir.

De Gooijer, J., & Hyndman, R. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, *22*(3), 443-473.

DEMP. (2017). *Presidency, Disaster and Emergency Management*. DEMP: Retrieved February 17, 2019, from https://deprem.afad.gov.tr/tarihteBuAy?id=32

Duncan, G., Gorr, W., & Szczypula, J. (2001). Forecasting analogous time series. In J. Armstrong, *Principles of Forecasting*, 195-213. Dordrecht: Kluwer Academic Publishers.

Ekiz Yılmaz, T. (2018). *Time series forecasting of intermittent demand by using ATA method and computational intelligence.* Master Thesis, Dokuz Eylül University, Izmir.

Fu, W. (2010). *ARIMA model for forecasting Poisson data: application to long-term earthquake predictions.* Master Thesis, University of Nevada, Las Vegas.

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

Gass, S., & Harris, C. (2001). *Encyclopedia of operations research and management science.* Dordrecht, The Netherlands: Springer US.

Hyndman, R. J. (n.d.). *Otexts*. Retrieved February 15, 2018 from https://otexts.com/fpp2/ses.html

Hyndman, R., & Athanasopoulos, G. (2018). *Forecasting principles and practice* (2 ed.). Australia: Monash University. Retrieved October 15, 2018, from https://otexts.org/fpp2/

Hyndman, R., & Koehler, A. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679-688.

iResearch Services. (2018, April 3). *Industry overview services: How is industry analysis done?* Retrieved March 24, 2019, from iResearch Services web site: https://www.iresearchservices.com/industry-overview-services-how-is-industry-analysis-done/

Kalekar, P. (2004, December 6). *Time series forecasting using Holt-Winters exponential smoothing.* Retrieved December 27, 2018, from OmniTI Labs: https://labs.omniti.com/people/jesus/papers/holtwinters.pdf

Karmaker, C. (2017). Determination of optimum smoothing constant of single exponential smoothing method: A case study. *Internatonal Journal of Research in Industrial Engineering, 6*, 184-192.

Kiren, Ö. (n.d.). *Üstel düzleştirme yöntemi.* İzmir. Retrieved December 5, 2018, from www.deu.edu.tr/userweb/ozlem.kiren/USTEL_DUZLESTIRME_YONTEMI_web.ppt

Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting* (16), 451-476. Retrieved March 10, 2019, from https://pdfs.semanticscholar.org/8461/b79f9747a0caee85522c49bd4655c64e10fb.pdf

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K., & Simmons, L. (1993). The M2-Competition: A real time judgmentally based forecasting study. *International Journal of Forecasting*(9), 5-22. Retrieved March 10, 2019, from https://www.academia.edu/8859671/The_M2-competition_A_real-time_judgmentally_based_forecasting_study

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competititon: Results, findings, conclusion and way forward. *International Journal of Forecasting, 34*(4), 802-808.

M-Competitions. (n.d.). Retrieved March 9, 2019, from M4Competition: https://www.mcompetitions.unic.ac.cy/

MTF. (2019). *Ministry of Treasure and Finance*. Retrieved April, 2019 from https://en.hmb.gov.tr/#

Nau, R. (2014). *Forecasting with moving averages.* North Carolina, US. Retrieved November 29, 2018, from https://people.duke.edu/~rnau/Notes_on_forecasting_with_moving_averages- -Robert_Nau.pdf

Nau, R. (2018). *Introduction to ARIMA: nonseasonal models.* North Carolina, US. Retrieved January 20, 2019, from https://people.duke.edu/~rnau/411arim.htm

Ostashchuk, O. (2017). *Times series data prediction and analysis.* Master Thesis, Czech Technical University, Prague.

Ostertagova, E., & Ostertag, O. (2012, December). *Forecasting using exponential smoothing method.* Retrieved 12 10, 2018, from Research Gate: https://www.researchgate.net/publication/256086712_Forecasting_Using_Simple_Exponential_Smoothing_Method

Ravinder, H. (2013). Determining the optimal values of exponential smoothing constant-does solver really work? *American Journal of Business Education, 6*, 347-360.

Shumway, R., & Stoffer, D. (2016). *Time series analysis and its applications* (4 ed.). New York: Springer.

Smith, M., & Agrawal, R. (2015). Acomparison of time series model forecasting methods on patent groups. *Modern Al and Cognitive Science Conference*, 167-173.

Spiliotis, E., Makridakis, S., & Assimakopoulos, V. (2018). *The M4 competition in progress.* University of Nicosia, Cyprus.

Stokes, H. (2011). *Exponential smoothing forecasting using SCAB34S and SCA workbench.* University of Illinois, Chicago.

Study.com. (n.d.). *What is financial data? Definition & concept*. Retrieved March 24, 2019, from Study.com web site: https://study.com/academy/lesson/what-is-financial-data-definition-lesson-quiz.html

Svensson, M. (2018). *An evaluation of methods for combibing univariate time series forecasts.* Bachelor's Thesis, Lund University, Lund.

Swanson, D., Tayman, J., & Bryan, T. (2015). *Mape-r: a rescaled measure of accuracy for cross-sectıonal forecasts.* Retrieved December 12, 2018, from https:// cssd.ucr.edu/Papers/PDFs/MAPE-R%20EMPIRICAL%20V24%20Swanson%20Tayman%20Bryan.pdf

Taylan Selamlar, H. (2017). *Modelling and forecasting time series data.* Phd Thesis, Dokuz Eylül Üniversity, Izmir.

Taylan, A. S. (2018, June 11). Retrieved May 10, 2019 from GitHub web site: https://github.com/M4Competition/M4-methods/blob/master/255%20-%20alsabtay/M4-Method-Description-ATA-1.pdf

Tirkeş, G., Güray, C., & Çelebi, N. (2017). Demand forecasting: a comparison between the Holt-Winters, trend analysis and decomposition models. *Technical Gazette, 4*(2), 503-509.

Trubetskoy, G. (2016, January 29). *Holt-Winters forecasting for dummies (or developers) - Part I. New Yok, USA.* Retrieved December 3, 2018, from https://grisha.org/blog/2016/01/29/triple-exponential-smoothing-forecasting/

TSI. (2019). *Turkish Standardization Institute*. Retrieved March 18, 2019 from https://www.tse.org.tr/

Wan, A. (n.d.). *Exponential smoothing methods*. Retrieved October 17, 2018, from City University of Hong Kong: http://personal.cb.cityu.edu.hk/msawan/teaching/ms6215/Exponential%20Sm oothing%20Methods.pdf

Wikipedia. (n.d.). *Economic data*. Retrieved March 24, 2019, from Wikipedia web site: https://en.wikipedia.org/wiki/Economic_data

Yapar, G. (2016). Modified simple exponential smoothing. *Hacettepe Journal of Mathematics and Statistics*, *47*(3), 741-754.

Yapar, G., Capar, S., Taylan Selamlar, H., & Yavuz, İ. (2017). *Modified Holt's linear trend method.* Retrieved December 25, 2018, from ATA Method: https://atamethod.files.wordpress.com/2018/02/hjms-pdf.pdf

Yapar, G., Yavuz, I., & Selamlar, H. (2017). Why and how does exponential smoothing fail? An in depth comparison of ATA-simple and simple exponential smoothing. *Turkish Journal of Forecasting*, *01*(1), 30-39.