DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

MODELLING AND FORECASTING TIME SERIES DATA USING ATA METHOD

by
Hanife TAYLAN SELAMLAR

July, 2017
İZMİR

MODELLING AND FORECASTING TIME SERIES DATA USING 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 Doctor of Philosophy in Statistics

by
Hanife TAYLAN SELAMLAR

July, 2017
İZMİR

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled "**MODELING AND FORECASTING TIME SERIES USING ATA METHOD**" completed by **HANİFE TAYLAN SELAMLAR** under supervision of **PROF. DR. GÜÇKAN YAPAR** and and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

....

Prof. Dr. Güçkan YAPAR

Supervisor

....

Prof. Dr. Esin FİRUZAN

Thesis Committee Member

....

Prof. Dr. Ali Kemal ŞEHİRLİOĞLU

Thesis Committee Member

....

Prof. Dr. Erol EĞRİOĞLU

Examining Committee Member

....

Doç. Dr. Filiz KARDİYEN

Examining Committee Member

Prof. Dr. Emine İlknur CÖCEN
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

Completion of this doctoral dissertation was possible with the support of several people. First of all, I would like to express my deepest gratitude to my supervisor Prof. Dr. Güçkan YAPAR for this continuous support, selfless guidance, generous advice and helpful comments. Especially, I wish to thank my supervisor, also for his valuable insighths, patience and encouragement during the hard times throughout my research.

Besides my advisor, I would like to thank Prof. Dr. Esin FİRUZAN and Prof. Dr. Ali Kemal ŞEHİRLİOĞLU, for their contributions and valuable help. I'm also grateful for the insights and efforts put forth.

Special thanks are to my precious sisters, for their kind patience to all my complaints and their support, whenever I needed. I am also grateful to Hasan-Erdoğan GÜLER, Emir-Ece SEÇİLMİŞ and Zeynep-Mehmet YILMAZ for their valuable precious moments that inspired my academic life and look forward to the future with hope.

This thesis is dedicated to my parents Hacer and Erdogan TAYLAN and my lovely husband Burak SELAMLAR because of their endless love, confidence, encouragement and support in my life. Words cannot express my thanks to my lovely family.

Hanife TAYLAN SELAMLAR

MODELLING AND FORECASTING TIME SERIES DATA USING ATA METHOD

ABSTRACT

It is difficult to make predictions especially about the future and making good predictions is not always easy. However, better predictions remain the foundation of all science therefore the development of accurate, robust and reliable forecasting methods is very important. Numerous number of forecasting methods have been proposed and studied in the literature. There are still two dominant major forecasting methods: Box-Jenkins ARIMA and Exponential Smoothing (ES), methods are derived or inspired from them.

After more than 50 years of widespread use, exponential smoothing is still one of the most practically relevant forecasting methods available due to their simplicity, robustness and accuracy as automatic forecasting procedures especially in the famous M-Competitions. The well-known fact in these competitions is ES has a proven success against more complex ARIMA models. Despite its success and widespread use in many areas, ES models have some shortcomings that negatively affect the accuracy of forecasts. Therefore, a new forecasting method in this thesis will be proposed to cope with these shortcomings and it will be called ATA method. This new method is obtained from traditional ES models by modifying the smoothing parameters therefore both methods have same structural forms and ATA can be easily adapted to all of the individual ES models. The two methods will be compared on popular metrics that are commonly used for evaluating performance of forecasting techniques. It will be shown that ATA models have better performance in terms of accuracy, simplicity, speed and interpretability.

The performance of ATA method will be compared not only to ES but also to other most successful competitors according to different performance criterions on the famous M3-Competition data set since it is still the most recent and comprehensive time-series data collection available.

Keywords: Exponential smoothing, time series, smoothing parameter, initial value, optimization, accuracy, M3-competition.

ATA YONTEMİ KULLANILARAK ZAMAN SERİSİ VERİLERİNİN MODELLENMESİ VE TAHMİNLENMESİ

ÖZ

Özellikle geleceğe ilişkin tahminlerde bulunmak ve iyi tahminler yapmak her zaman kolay değildir. Bununla birlikte, daha iyi tahminler tüm bilimlerin temelini oluşturur, dolayısıyla doğru, dayanıklı ve güvenilir tahmin yöntemlerinin geliştirilmesi çok önemlidir. Literatürde çok sayıda tahminleme yöntemi önerilmiş ve çalışılmıştır. Literatürde hala Box-Jenkins ARIMA ve Üssel Düzleştirme (ÜD) Yöntemi olmak üzere iki temel yöntem bulunmaktadır ve diğer tahminleme yöntemleri de bu iki temel yöntemden türetilen ya da onlardan ilham alan yöntemlerdir.

50 yıldan uzun süren yaygın kullanımdan sonra üssel düzleştirme, özellikle ünlü tahminleme yarışmalarında (M-competition) otomatik tahmin prosedürleri gibi basitlik, dayanıklılık ve doğruluklarından dolayı mevcut en prestijli tahmin yöntemlerinden biridir. Bu yarışmalardaki iyi bilinen gerçek, ÜD'nin daha karmaşık ARIMA modellerine karşı kanıtlanmış bir başarısı olduğudur. ÜD modellerinin başarısı ve yaygın kullanımı birçok alanda olmasına rağmen, tahminlerin doğruluğunu olumsuz yönde etkileyen bazı eksikliklere sahiptir. Bu nedenle, bu eksikliklerin üstesinden gelmek için bu tezde yeni bir tahmin metodu önerilecektir ve buna ATA metodu denilecektir. Bu yeni yöntem, düzleştirme parametrelerini değiştirerek geleneksel ÜD modellerinden elde edilir, bu nedenle her iki yöntem de aynı yapısal formlara sahiptir ve ATA her ÜD modeline kolayca uygulanabilir. Her iki yöntem de, tahmin tekniklerinin performansını değerlendirmek için yaygın olarak kullanılan popüler metriklerde karşılaştırılacaktır. ATA modellerinin doğruluk, basitlik, hız ve yorumlanabilirlik bakımından daha iyi bir performansa sahip oldukları gösterilecektir.

ATA yönteminin performansı, ünlü M3-Yarışması veri setindeki farklı performans kriterlerine göre yalnızca ÜD değil, aynı zamanda diğer en başarılı rakiplerle karşılaştırılacaktır çünkü bu veri seti en yeni ve kapsamlı zaman serisi veri koleksiyonudur.

Anahtar kelimeler: Üssel düzleştirme, zaman serisi, düzleştirme katsayısı, başlangıç değeri, optimizasyon, doğruluk, M3-yarışması.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE
INTRODUCTION

Exponential smoothing (ES) method is the most well-known tool in the field of time series forecasting. In recent decades, numerous time series forecasting models have been proposed. Forecasting is an essential activity in various branches of science and in many areas of industrial, commercial and economic activity. Forecasts can be obtained by using purely judgmental, explanatory and extrapolative methods or any combination of these three but extrapolative methods are reliable, objective, inexpensive, quick, and easily automated. Previous studies have demonstrated that there are still two major univariate forecasting approaches: exponential smoothing (ES) and ARIMA (De Gooijer & Hyndman (2005)). There are many forecasting methods but the most successful forecasting methods are based on the concept of exponential smoothing (ES). Exponential smoothing is inarguably one of the most widely used forecasting methods available due to its simplicity, adaptiveness and accuracy (Goodwinet al. (2010)). The formulation of the first ES method by Brown in the late 1950s (Brown (1959)) was followed by Holt (1957) and Winters (1960) for trended and seasonal data sets. Later, damped trend model was proposed Gardner Jr & McKenzie (1985) to help deal with over-trending.

The main idea behind ES is to assign recent observations more weight compared to the distant past when obtaining forecasts. Forecasts produced using ES methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older. In other words, recent observations are given relatively more weight in forecasting than the older observations. The popularity of exponential smoothing can also be attributed to its proven record against more sophisticated approaches (Makridakis et al. (1984); Makridakis & Hibon (2000); Koning et al. (2005)).

ES models assume that the time series have up to three underlying data components: level, trend and seasonality. Estimates for the final values of these components are used to construct the forecasts. ES models can include 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). Pegels (1969) proposed a taxonomy of ES methods, which was extended and modified later by Gardner Jr & McKenzie (1985); Hyndman et al. (2002); Taylor (2003) and Hyndman & Athanasopoulos (2014). Thus, there are 15 different ES models, the best known of which are simple exponential smoothing (SES) (no trend, no seasonality), Holt's linear model (additive trend, no seasonality) and Holt-Winters' additive model (additive trend, additive seasonality) Goodwinet al. (2010). For this reason we can say that ES is not a simple model but rather a family of models.

There are many studies on the numerical and theoretical comparison of Box-Jenkins and ES methods. For several decades, ES has been considered an ad hoc approach to forecasting, with no proper underlying stochastic formulation. The state space framework described by Hyndman et al. (2002) brings exponential smoothing into the same class as ARIMA models. ES is widely applicable and has a sound stochastic model behind the forecast therefore the "ad hoc approach" argument is no longer true. Hyndman et al. (2002) introduced a state space framework that subsumes all the exponential smoothing models and allows for the computation of prediction intervals, likelihood and model selection criteria. They also proposed an automatic forecasting strategy based on this model framework. The ETS state space models (Hyndman et al. (2002); Hyndman et al. (2008); Hyndman & Athanasopoulos (2014)) brought exponential smoothing to a higher level by providing the method with a solid theoretical background. They extended the earlier classifications so that now there are 30 potential ES models for various types of trend, seasonality and errors.

The implementation of ES requires the user to specify the type of the model, the smoothing constants and initial values. In order to use any ES model for forecasting, the forecaster has to provide the value of both the initial conditions, that is, the level, trend and seasonal components at the start of the series and the smoothing and damped parameters for each component; all of which are unknown. The performance of any ES model depends mainly on the smoothing constant reflecting the relative weight assigned to the most current observations and initial values. The weights assigned to components of an ES model can take on values between 0 and 1. Considerable effort has been focused on finding the appropriate values of smoothing constants and initial values. The user must choose parameters, either fixed or adaptive, as well as initial values and loss functions. The smoothing constants and initial values for any ES method can be estimated by minimizing the sum of the squared errors. Standard ES methods are usually fitted in two steps, by choosing fixed initial values followed by an independent search for parameters. Chatfield & Yar (1988) describe how the use of different approaches for the derivation of initial values for the smoothed level, trend and seasonal components can give rise to substantially different optimized parameter values which can lead to substantially different forecasts. Makridakis & Hibon (1991) measured the effect of different initial values and loss functions. In contrast, the new state-space methods are usually fitted using maximum likelihood, a procedure that makes the choice of initial values less of a concern because they are redefined simultaneously with the smoothing parameters during the optimization process. Unfortunately, maximum likelihood may require significant computation times, as discussed in Hyndman et al. (2002). For example, in monthly seasonal models with a damped trend, there are 13 initial values and 4 parameters, so optimization is done in 17- dimensional space and this optimization is not an easy task.

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

The first serious empirical study of forecasting performance emerged in the 1969 by Reid. Several systematic studies have been published by Newbold & Granger

(1974), Makridakis & Hibon (1974). However, the main competition between these two major forecasting methods is on their post-sample forecasting accuracy. The 1001 and 3003 time series used respectively in the M-competition (Makridakis et al. (1982) and the M3-competition Makridakis & Hibon (2000)) have become recognized collections of test data for the evaluation of forecasting methods. In the M3 forecasting competition (Makridakis & Hibon (2000); Koning et al. (2005)), ES methods obtained the best results in the competition. For the ETS models they used the forecast package (Hyndman et al. (2008);Hyndman & Athanasopoulos (2014)) in the programming language R (R Core Team, 2014), thus a fully automatic software for fitting ETS models is available. The first conclusion of these competitions is that "statistically sophisticated or complex methods did not provide more accurate forecasts than simpler ones". Some standard and simple combinations of ES methods were used in these competitions and their performances verified this result. According to statements by the participants of the M-competition, the Box-Jenkins methodology (ARMA models) required the most time (on the average more than one hour per series). To propose a simple, accurate, robust and automatic forecasting method as an alternative to ES methods is not easy task after the results of Hyndman et al. (2002). In this study, they applied an automatic forecasting strategy to the M-competition data Makridakis et al. (1982) and the M3 competition data Makridakis & Hibon (2000). The automatic forecasting procedure proposed in that paper tries each of the 24 state space models on a given time series and selects the "best" method using the Akaike Information Criterion. They showed that the methodology is particularly good at short term forecasts (up to about 6 periods ahead), and especially accurate for seasonal short-term series (beating all other methods in the competitions).

Several other studies that are based on the automatic forecasting procedures exist. Particularly for seasonal time series the forecast package offers the TBATS model, which was proposed by De Livera et al. (2011). TBATS uses a trigonometric parsimonious representation of seasonality, instead of conventional seasonal indices, and also incorporates ARMA errors. In addition, the function also automatically performs Box-Cox transformation of the time series, if necessary. Multiple

Aggregation Prediction Algorithm (MAPA) was suggested by Kourentzes et al. (2014) and the model produces temporally aggregated versions of a time series by first using statistical tests to identify the appropriate order of differencing. They proposed a framework that mitigates the issue of model selection, while improving the forecasting accuracy, by taking advantage of temporal aggregation and forecast combination. Another significant addition to the competition literature named THETA, a decomposition based method proposed by Assimakopoulos & Nikolopoulos (2000), later shown to be equivalent to a simple exponential model with drift by Hyndman & Billah (2003), stood out in the M3-competition. Assimakopoulos & Nikolopoulos (2000) removed the curvature of the original time series and called the resulting series that maintained only the mean and slope of the original series Theta-lines. They decomposed the original series into two or more Theta-lines and extrapolated these lines separately to obtain forecasts that in the end are combined to produce forecasts for the 3003 series in the M3-competition. As confirmed once again in Assimakopoulos & Nikolopoulos (2000), it is well known that combining forecasts (Bates & Granger (1969); Clemen (1989)) under certain circumstances improves forecasting accuracy (Armstrong (1989); Armstrong (2001); Makridakis & Winkler (1983); Makridakis et al. (1982)). Because of this, the research since the beginning of this competition mainly focuses on certain transformations, decompositions, rules and combinations of ES and ARIMA (a few examples are Clemen (1989); Cleveland et al. (1990); Adya et al. (2000); Bergmeir et al. (2016)) to improve the forecasting performance rather than proposing new forecasting methods.

To sum up, ES is popular since it is accurate, simple, fast and inexpensive compared to the Box-Jenkins method. The main objective of this study is to introduce a new forecasting method as an alternative to ES. The proposed method is developed from ES models by modifying the smoothing parameters.

1.1 Research Scope and Objectives

For a model to be considered as an alternative method to ES, it should be simpler, more accurate, faster than ES and should not be a special case of it. The main objective of this thesis is to introduce a new forecasting method as an alternative to ES methodology that has the best performance among the methods presented by the intended competitors in Makridakis & Hibon (2000). This thesis is set out to propose a new forecasting method to obtain the initial value and the smoothing parameter simultaneously. The proposed method is named ATA, which is based on ES methodology but the smoothed constant is modified. The proposed method can be easily adapted for higher order ES models. It can be formulated as:

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

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

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

for $p \in \{1, \ldots, n\}$, $q \in \{0, 1, \ldots, p\}$ and $t > p \geq q$. For $t \leq p$ let $S_t = X_t$, for $t \leq q$ let $T_t = X_t - X_{t-1}$ and let $T_1 = 0$, where X_t is the actual observation of the series, S_t denotes an estimate of the level of the series at time t, T_t denotes an estimate of the growth (trend) value of the series at time t, p is the smoothing parameter for the level and q is the smoothing parameter for the trend. Even though ATA can be adapted to all ES models, in this thesis we focus on just the ATA model with linear and exponential trend components. Furthermore, for the sake of simplicity, the simplest form of the proposed method will be given along with a comparison of its main features compared to its counter ES models.

A competition has been organized by the International Institute of Forecasters (IIF) to determine the best method of predicting the future. For this competition, there are three types of data sets (https://forecasters.org/resources/time-series-data/) respectively named M-Competition (1001 series), M2-Competition and

M3-Competition (3003 series). In this thesis, the M3 competition data set is used because the 3003 series is still the most recent, comprehensive time-series data collection available. Moreover the data is still competitive and its results are verified. 24 forecasting methods have joined the contest and they obtained the forecasts for the 3003 series based on their own methodologies (Makridakis & Hibon (2000)). The results have demonstrated that, the best performing methods in the competition are based on ES and Box-Jenkins methodology, and the simple methods yield better results than complex methods. Any forecasting method may have many desirable futures but the ultimate goal is to predict future events accurately. Therefore, after the method is proposed, the performance of the method will be tested empirically using the same competition data sets mentioned above.

Table 1.1 Average symmetric MAPE across different forecast horizons: all 3003 series

Method	Forecasting horizons										Averages					
	1	2	3	4	5	6	8	12	15	18	1 to 4	1 to 6	1 to 8	1 to 12	1 to 15	1 to 18
Naive2	10.5	11.3	13.6	15.1	15.1	15.9	14.5	16.0	19.3	20.7	12.62	13.57	13.76	14.24	14.81	15.47
Single	9.5	10.6	12.7	14.1	14.3	15.0	13.3	14.5	18.3	19.4	11.73	12.71	12.84	13.13	13.67	14.32
Holt	9.0	10.4	12.8	14.5	15.1	15.8	13.9	14.8	18.8	20.2	11.67	12.93	13.11	13.42	13.95	14.60
Dampen	8.8	10.0	12.0	13.5	13.7	14.3	12.5	13.9	17.5	18.9	11.05	12.04	12.14	12.44	12.96	13.63
Winter	9.1	10.5	12.9	14.6	15.1	15.9	14.0	14.6	18.9	20.2	11.77	13.01	13.19	13.48	14.01	14.65
Comb S-H-D	8.9	10.0	12.0	13.5	13.7	14.2	12.4	13.6	17.3	18.3	11.10	12.04	12.13	12.40	12.91	13.52
B-J automatic	9.2	10.4	12.2	13.9	14.0	14.8	13.0	14.1	17.8	19.3	11.42	12.41	12.54	12.80	13.35	14.01
Autobox1	9.8	10.1	13.1	15.1	16.0	16.8	14.2	15.4	19.1	20.4	12.30	13.67	13.78	14.00	14.56	15.23
Autobox2	9.5	10.4	12.2	13.8	13.8	14.9	13.2	15.2	18.2	19.9	11.48	12.44	12.63	13.10	13.70	14.41
Autobox3	9.7	11.2	12.9	14.6	15.8	16.5	14.4	16.1	19.2	21.2	12.08	13.43	13.64	14.01	14.57	15.33
Robust-trend	10.5	11.2	13.2	14.7	15.0	15.9	15.1	17.5	22.2	24.3	12.38	13.40	13.73	14.57	15.42	16.30
ARARMA	9.7	10.9	12.6	14.2	14.6	15.6	13.9	15.2	18.5	20.3	11.83	12.92	13.12	13.54	14.09	14.74
Automat ANN	9.0	10.4	11.8	13.8	13.8	15.5	13.4	14.6	17.3	19.6	11.23	12.38	12.58	12.96	13.48	14.11
Flores/Pearce1	9.2	10.5	12.6	14.5	14.8	15.3	13.8	14.4	19.1	20.8	11.68	12.79	13.03	13.31	13.92	14.70
Flores/Pearce2	10.0	11.0	12.8	14.1	14.1	14.7	12.9	14.4	18.2	19.9	11.96	12.77	12.81	13.04	13.61	14.29
PP-autocast	9.1	10.0	12.1	13.5	13.8	14.7	13.1	14.3	17.7	19.6	11.20	12.21	12.40	12.80	13.34	14.01
ForecastPro	8.6	9.6	11.4	12.9	13.3	14.3	12.6	13.2	16.4	18.3	10.64	11.69	11.86	12.14	12.60	13.19
SmartFcs	9.2	10.3	12.0	13.5	14.0	15.1	13.0	14.9	18.0	19.4	11.23	12.34	12.49	12.94	13.48	14.13
Theta-sm	9.8	11.3	12.6	13.6	14.3	15.0	12.7	14.0	16.2	18.3	11.81	12.76	12.77	13.04	13.40	13.88
Theta	8.4	9.6	11.3	12.6	13.8	14.7	12.0	13.2	16.5	18.5	10.46	11.72	11.80	12.11	12.58	13.19
RBF	9.9	10.5	12.4	13.4	13.2	14.2	12.8	14.1	17.3	17.8	11.56	12.28	12.42	12.77	13.25	13.75
ForecastX	8.7	9.8	11.6	13.1	13.2	13.9	12.6	13.9	17.8	18.7	10.82	11.73	11.89	12.22	12.81	13.49
AAM1	9.8	10.6	11.2	12.6	13.0	13.5	14.1	14.9	18.0	20.4	11.04	11.76	12.43	13.04	13.77	14.63
AAM2	10.0	10.7	11.3	12.9	13.2	13.7	14.3	15.1	18.4	20.7	11.21	11.95	12.62	13.21	13.97	14.85
ETS	8.8	9.8	12.0	13.5	13.9	14.7	13.0	14.1	17.6	18.9	11.04	12.13	12.32	12.66	13.14	13.77
ATA Method	8.4	9.4	11.1	12.5	12.8	13.3	11.7	13.0	16.3	17.3	10.35	11.25	11.36	11.72	12.21	12.78
Rank	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

The new forecasting method first appeared in Yapar (2016) where he proposed a simple modification of the exponential smoothing model, generalized here as the ATA method, which produces surprising results in terms of forecasting accuracy and simplicity. The ATA method eliminates the initialization problem and is easier to optimize compared to its counter ES models. This method is shown to have better forecasting accuracy for the M-competition data when equivalent parameterizations of simple exponential smoothing and ATA models are compared on post sample forecasting powers. In this thesis, the ideas from Yapar (2016) will be extended to handle different trend types and various model selection and forecast combination options will be discussed for ATA. The best performance achieved by applying some basic model selection and combination rules for ATA are given in Table 1.1 and it can be seen that for the M3-competition data proposed approach outperforms all other competitors consistently. This table alone underline underlines the importance of the contribution made in this dissertation and the details on how these results were obtained will be explained step by step in the following sections.

1.2 Research Outline

In this thesis, the new forecasting method ATA will be proposed and its accuracy performance will be given along with other properties. In addition, the results will be compared to the benchmarks competition methods in forecasting literature. The thesis consists of six chapters. In the first chapter, a review of the related literature of the time series forecasting and the scope and aim of this thesis are given.

Existing forecasting methods from the literature one discussed in Cahpter 2 starting from the simplest ones to more sofisticated approaches. In addition, metrics commonly used in measuring forecasting accuracy and methods for combining forecasts are given here. A brief introduction to the M-Competitions is made.

In the third chapter, the simple form of ATA is given and its features are discussed in detail. These are the weights attached to the observations, the average age and the variance, the weight of initial value, the smoothing parameters, its optimization and parameter space. The simple form of the ATA and the SES method are compared based on their perfromance on the M3-Competition data set and their features.

The trended ATA methods will be given for additive, multiplicative and damped trend in the fourth chapter. An application of these methods will be explained with a data set from the M3-Competitions. Also, the proposed models' accuracy will be compared to its counter ES model using M3 competition data set in the last part of this chapter.

In the fifth chapter, the best forecasting performance of ATA method is given. First some, selection and combination rules are defined for various time intervals. Then, we present our results for all series according to sMAPE and MASE error metrics. The out sample performance of different types of series are calculated and discussed in the following subsections.

CHAPTER TWO
METHODOLOGY AND DATA

There are many different situations where forecasts are required in order to make good decisions. Forecasting is a decision making tool used by many businesses to help in budgeting, planning and estimating future growth. In the simplest term, forecasting is the use of historic data to determine the direction of future patterns. According to Armstrong (2001) forecasting is defined as the prediction of an actual value in a future time period. Makridakis S. & Hyndman (1998) state that forecasting supplies information of what may occur in the future. Thus, it is used to estimate when an event is probable to happen so that proper action can be taken. Despite the wide range of these situations that require forecasts, there are two broad types of forecasting techniques qualitative (judgemental) and quantitative forecasting (Makridakis S. & Hyndman (1998) and Armstrong (2001)). They are known as objective and subjective techniques. According to Makridakis S. & Hyndman (1998) quantitative forecasting can be applied under three conditions:

- Quantitative information availability about the past,
- Information can be expressed in numerical data,
- Assumption of continuity, which is the statement that characteristics of the past patterns will continue in the future.

On the other hand, qualitative forecasting is applied in case of lack of quantitative information, but sufficient qualitative knowledge and experience exists. Finally, when neither quantitative information nor qualitative knowledge is available a satisfactory forecast cannot be performed. Both quantitative and qualitative techniques differ extensively in accuracy, cost and complexity. Quantitative techniques are divided in to two categories: explanatory (causal) models and time series models. The first category investigates the cause and effect relationship between the forecasted variable and one or more independent variables. Time series models predict the future value of a variable based upon its past values without attempting to estimate the external factors that affect this behaviour. The most common forecasting techniques are classified in the following figure (Armstrong (2001)).

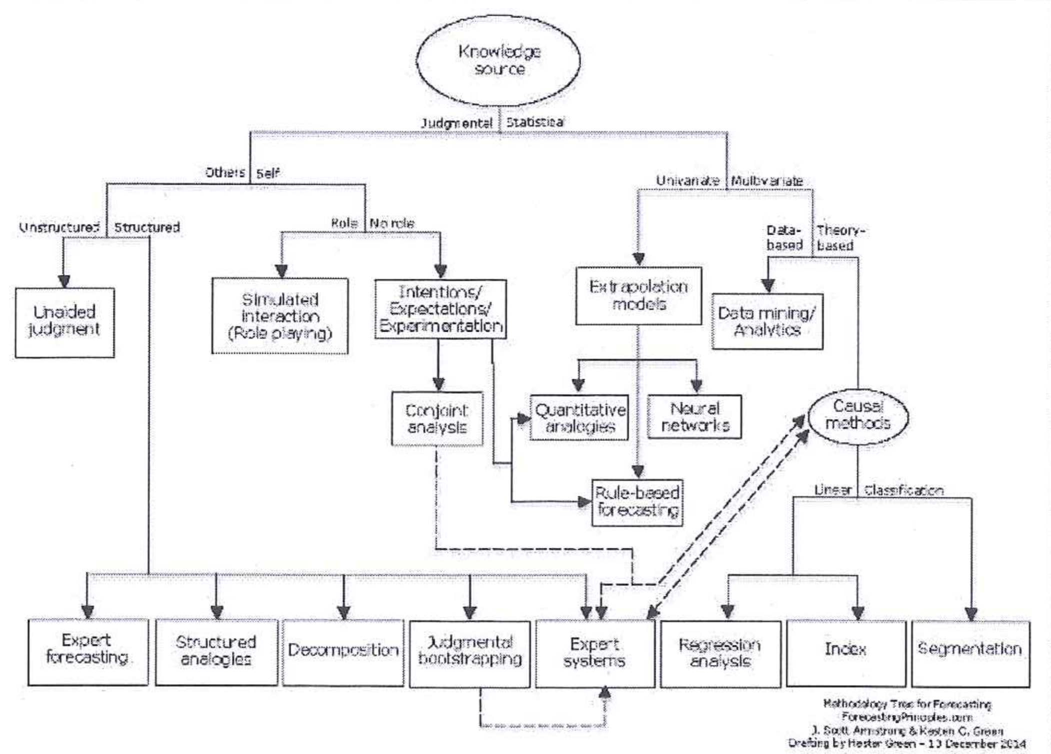

Figure 2.1 The forecasting techniques and their interactions. Principles of Forecasting, Armstrong (2001)

If the series at hand is seasonal the seasonal component can be removed from the original data. Then, the resulting values are called the "seasonally adjusted" or "deseasonalized" data. The ATA method which is proposed in this thesis is applied to non seasonal or deseasonalized time series, where the deseasonalisation is performed via the multiplicative classical decomposition.

2.1 Some Basic Forecasting Techniques

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

$$X_t = a + e_t, \tag{2.1}$$

where e_t is some random noise with mean zero and variance σ^2. Under the model in (2.1), the aim is then reduced to finding a good estimator for the constant a so that it can be used to forecast future values. The general form of this estimator should involve some sort of an average of the observed values. It can be notated as:

$$\hat{a} = F(X_1, \ldots, X_n) = \sum_{t=1}^{n} w_t X_t, \tag{2.2}$$

where w_t are a collection of weights called weighting vector such that $w_t \in [0, 1]$ for $t = 1, \ldots, n$ and $\sum_{t=1}^{n} w_t = 1$. The estimators of form (2.2) will be unbiased. In order to deal with sequential updating, the term a_n is sometimes used to indicate the smoothed value at time n, therefore a_n and $\hat{a}$ are synonyms.

Since there are a lot of ways to obtain estimators of form (2.2) and there is not an estimator that will be universally satisfactory, researchers need a way to choose among all these potential smoothing methods. When making a choice, some criteria to judge the relative merits of each alternative that are important to the researchers are needed. Most importantly, researchers have a preference for fresh data and therefore in practice weighting vectors that assign more weight to recent observations are preferred. In other words, weighting vectors with $w_j \geq w_i$ for $j > i$ are preferred. One popular metric that is used for measuring a smoothing method's ability to utilize fresh data is the average

age (AA) of the data used:

$$AA(\hat{a}) = \bar{k} = n - \sum_{t=1}^{n} t w_t. \tag{2.3}$$

Another important metric to consider is the variance of the estimator at hand as usual. Since the estimator in (2.2) is unbiased, its variance can be written as:

$$Var(\hat{a}) = E\left[\left(\sum_{t=1}^{n} w_t X_t - a\right)^2\right] = \sum_{t=1}^{n} w_t^2 \sigma^2 = V\sigma^2. \tag{2.4}$$

The following paragraphs summarises the main findings of basic forecasting methods as follows:

1. $w_t \in [0, 1] \qquad t = 1, \ldots, n$

2. $\sum_{t=1}^{n} w_t = 1$

3. $w_1 \leqslant w_2 \leqslant \ldots \leqslant w_n$

   - $AA(\hat{a}) = \bar{k} = n - \sum_{t=1}^{n} t w_t$.
   - $Var(\hat{a}) = E\left[(\sum_{t=1}^{n} w_t X_t - a)^2\right] = \sum_{t=1}^{n} w_t^2 \sigma^2 = V\sigma^2$.

Even though it is desirable to keep both of the metrics in (2.3) and (2.4) minimal simultaneously, it is not an achievable goal. Consider two extreme weighting schemes which will result in boundary values of these metrics. The first scheme is the average method which assigns equal weights to all observations over time and the second method is the naive method.

2.1.1 Average Method

For the average method, the forecast for all future values is equal to the mean of historical data $\{X_1, \ldots, X_n\}$ (Johnston et al. 1999).

$$\hat{a} = \hat{X}_t(h) = \bar{X}_t, \qquad t = 1, \ldots, n \tag{2.5}$$

1. $w_1 = w_2 = \ldots w_t = \frac{1}{n}$,

2. $AA = \bar{k} = \frac{n-1}{2}$,

3. $V = \frac{1}{n}$.

Here the estimator $\hat{a}$ is simply the simple average and it is well known that for the conditions $w_t \in [0, 1]$ and $\sum_{t=1}^{n} w_t = 1$ the variance in (2.4) is minimized since $V = \frac{1}{n}$. On the contrary, AA attains its largest value under this weighting scheme which is equal to $\frac{n-1}{2}$.

2.1.2 Naive Method

Another approach is the simplest, but widely used forecasting approach which is called the Naive method. For the naive method, the forecast is simply the last observed value of the time series (Aaker and Jacobson, 1987).

$$\hat{a} = \hat{X}_t(h) = X_t, \qquad t = 1, \ldots, n \tag{2.6}$$

Under the naive weighting scheme where all the observations other than the latest one are discarded, i.e. $w_n = 1$ and $w_t = 0$ for $t = 1, \ldots, n-1$, the estimator is simply equal to the latest observation.

1. $w_1 = w_2 = \dots w_{n-1} = 0, \qquad w_n = 1$

2. $AA = \bar{k} = 0,$

3. $V = 1.$

AA under this scenario will be equal to zero thus minimized but this time the variance of the estimator will be maximized since V now reaches its maximum value which is equal to 1.

2.1.3 Moving Average Method

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

$$\hat{a} = \frac{X_n + X_{n-1} + \dots + X_{n-p+1}}{p} = \frac{1}{p}\sum_{j=0}^{p-1} X_{n-j}, \qquad (2.7)$$

for $p \leq n$. For this model the weights are $w_j = 0$ for $j \leq n - p$ and $w_j = \frac{1}{p}$ for $n - p + 1 \leq j \leq n$.

1. $w_1 = w_2 = \dots = w_{n-p} = 0 \qquad w_{n-p+1} = \dots = w_n = \frac{1}{p}$

2. $AA = \bar{k} = \frac{p-1}{2},$

3. $V = \frac{1}{p}.$

The average age of the data used in a moving average is defined as below:

$$\bar{k} = \frac{0 + 1 + 2 + \ldots + p - 1}{p} = \frac{p-1}{2} \tag{2.8}$$

So, this model has $AA = \frac{p-1}{2}$ and $V = \frac{1}{p}$ Brown (1962). The moving average for time t is the mean of the p most recent observations. The constant number p is specified at the outset. The smaller the number p, the more weight is given to recent periods. The greater the number p, the less weight is given to more recent periods. A large p is desirable when there are wide, infrequent fluctuations in the series. It is a disadvantage to have to carry all the past data necessary to compute the moving average. A small p is most desirable when there are sudden shifts in the level of series.

So far, this section has focused on some basic forecasting methods. Another classic and well known forecasting approach that allows the researchers to utilize more data is the ES. In this thesis, the ATA method, which is an extrapolative forecasting method alternative to exponential smoothing methods, will be evaluated. Therefore, the procedure and basic properties of ES will be discussed in the following section.

2.2 Simple Exponential Smoothing (SES)

Exponential smoothing methods originated from the works of Brown (1959), Holt (1957) and Winters (1960). The method was independently developed 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, 2000). Brown was assigned to design a tracking system for firecontrol information to compute the location of submarines. Brown's tracking model was essentially simple exponential smoothing of continuous data. This model is still used in modern fire control equipment. During the early 1950s, Brown extended simple exponential to discrete data and developed methods for 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 (Brown, 1959).

The SES method is a classical and well-known approach where the smoothing constant is denoted by $\alpha \in [0, 1]$. Let X_t denote the observed value of a time series at time t and $\hat{a} = \hat{X}_t(h)$ be the forecast for h periods ahead from origin t. The integer $h(> 0)$ is called the forecasting horizon or lead time. Therefore, the model can be written as:

$$componentform: \quad S_t = \alpha X_t + (1 - \alpha) S_{t-1}, \tag{2.9}$$

$$errorform: \quad S_t = S_{t-1} + \alpha(X_t - S_{t-1}), \tag{2.10}$$

$$\hat{a} = \hat{X}_t(h) = S_t, \tag{2.11}$$

where S_t is the smoothed value at time t which is as mentioned earlier in the previous chapter a synonym for $\hat{a}$. Substituting the model in (2.9) into itself successively, the model can be re-written as:

$$S_t = \alpha \sum_{k=0}^{t-1} (1 - \alpha)^k X_t + (1 - \alpha)^t S_0 \tag{2.12}$$

so S_t represents a weighted moving average of all past observations with weights decreasing exponentially, S_0 is the initial value. It can be seen that for large α recent observations get more weight. Taken together, the weights from SES satisfy the required conditions on weights similar to naive and average.

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

Weights assigned by simple exponential smoothing are non-negative and sum to unity. If α is small, more weight is given to observations from the more distant past. If α is large, more weight is given to the more recent observations. In the exponential smoothing process, the weight given data k periods ago is $\alpha(1-\alpha)^k$. For different smoothing levels, the weights are given to the Table 2.1.

Table 2.1 The weights of observations for different parameters

		weight of X_t		
Observation	Formulation	$\alpha = 0.1$	$\alpha = 0.4$	$\alpha = 0.8$
X_n	α	0.1	0.4	0.8
X_{n-1}	$\alpha(1-\alpha)^1$	$0.1(0.9)^1$	$0.4(0.6)^1$	$0.8(0.2)^1$
X_{n-2}	$\alpha(1-\alpha)^2$	$0.1(0.9)^2$	$0.4(0.6)^2$	$0.8(0.2)^2$
X_{n-3}	$\alpha(1-\alpha)^3$	$0.1(0.9)^3$	$0.4(0.6)^3$	$0.8(0.2)^3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_2	$\alpha(1-\alpha)^{n-1}$,	$0.1(0.9)^{n-1}$	$0.4(0.6)^{n-1}$	$0.8(0.2)^{n-1}$
X_1	$\alpha(1-\alpha)^n$	$0.1(0.9)^n$	$0.4(0.6)^n$	$0.8(0.2)^n$
Initial value	$(1-\alpha)^n$	$(0.9)^n$	$(0.6)^n$	$(0.2)^n$

The AA for SES is:

$$\begin{aligned}\bar{k} &= 0\alpha + 1\alpha(1-\alpha) + 2\alpha(1-\alpha)^2 + \dots \\ &= \alpha\sum_{k=0}^{\infty} k(1-\alpha)^k \\ &= \frac{(1-\alpha)}{\alpha}.\end{aligned} \tag{2.13}$$

One way to define an exponential smoothing that is equivalent to a p period moving average is to say that the smoothing constant is selected to give the same average age of the data.

$$\frac{1-\alpha}{\alpha} = \frac{p-1}{2} \quad or \quad \alpha = \frac{2}{p+1} \tag{2.14}$$

Now, it is possible to calculate the expected value and variance of the smoothed value S_t. For sufficiently large t, the expected value of S_t is

$$\begin{aligned}
E(S_t) &= E\left(\alpha\sum_{k=0}^{\infty}(1-\alpha)^k X_{t-k} + (1-\alpha)^t S_0\right) \qquad (1-\alpha)^t S_0 \to 0 \quad when \quad t\to\infty \\
&= \alpha\sum_{k=0}^{\infty}(1-\alpha)^k E(X_{t-k}) \\
&= E(X)\alpha\sum_{k=0}^{\infty}(1-\alpha)^k, \quad 0 \le \alpha \le 1 \\
&= a\alpha\frac{1}{1-(1-\alpha)} \\
&= a\frac{\alpha}{\alpha} \\
&= a
\end{aligned} \tag{2.15}$$

so S_t is unbiased estimator of the constant a when $t\to\infty$. Therefore, S_t can be used for future forecasts. The variance of S_t is

$$\begin{aligned}
V(S_t) &= V\left(\alpha\sum_{k=0}^{\infty}(1-\alpha)^k X_{t-k} + (1-\alpha)^t S_0\right) \\
&= \alpha^2\sum_{k=0}^{\infty} V(X_{t-k}) \\
&= V(X)\alpha^2\sum_{k=0}^{\infty}\left((1-\alpha)^k\right)^2, \quad 0 \le \alpha \le 1 \\
&= \sigma_\epsilon^2\alpha^2\frac{1}{1-(1-\alpha)^2} \\
&= \frac{\alpha^2}{2\alpha-\alpha^2}\sigma_\epsilon^2 \\
&= \frac{\alpha}{2-\alpha}\sigma_\epsilon^2
\end{aligned} \tag{2.16}$$

S_t is an unbiased estimator since $\sum_{t=1}^{n} w_t = 1$. For large sample sizes, $AA = \hat{k} = \frac{1-\alpha}{\alpha}$ and $V = \frac{\alpha}{2-\alpha}$ have been defined Brown (1962). Table 2.2 shows the weights attached to observations for three different values of α when forecasting using simple exponential smoothing. In addition, the weights of initial value, average age and the variance are shown in the table for 10 observations. The weight given to starting value is 0.349 which is bigger than all weights given to other values when $\alpha = 0.1$.

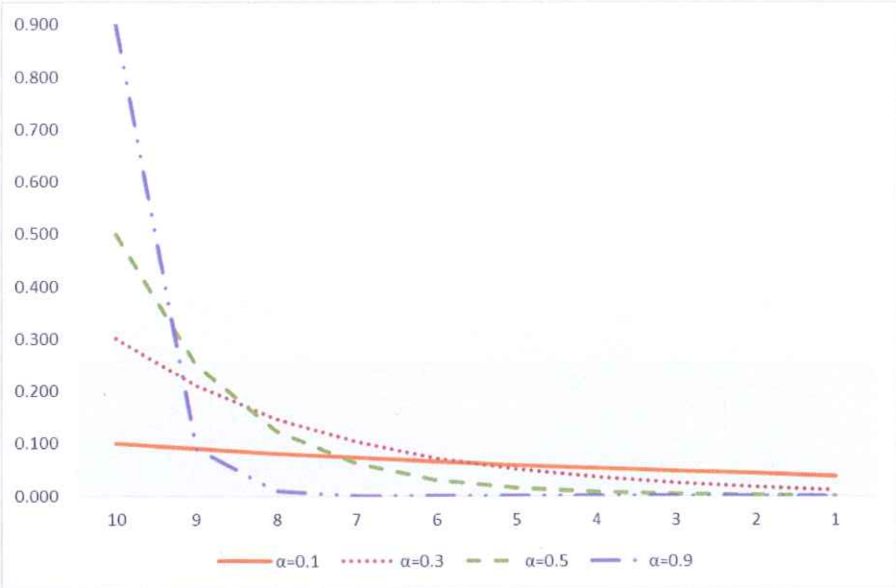

Figure 2.2 Weights assigned to observations by various smoothing parameters of SES

Table 2.2 Weights assigned to observations by various smoothing parameters of SES

	weights by α for SES			
Observation	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.9$
X_{10}	0.100	0.300	0.500	0.900
X_9	0.090	0.210	0.250	0.090
X_8	0.081	0.147	0.125	0.009
X_7	0.073	0.103	0.063	0.001
X_6	0.066	0.072	0.031	0.000
X_5	0.059	0.050	0.016	0.000
X_4	0.053	0.035	0.008	0.000
X_3	0.048	0.025	0.004	0.000
X_2	0.043	0.017	0.002	0.000
X_1	0.039	0.012	0.001	0.000
Weight of initial	0.349	0.028	0.001	0.000
Average Age (AA)	9.000	2.333	1.000	0.111
Sum of squared weights (V)	0.046	0.176	0.333	0.818

The weight of last observation is much smaller than the weight given to the starting value which contradicts the main idea of exponential smoothing. When $\alpha = 0.1$, the weight given to starting value is too small as expected. The sum of the weights even for small α will be approximately one for any reasonable sample size. For example, sum of the weights attached to the most recent 10 observations are 0.999 respectively and almost zero weight after tenth observation when $\alpha = 0.5$. That's why most of practitioners recommend smoothing constant smaller than 0.3. As a matter of fact, if we carry out a sequence of trials on some set of actual data and find that smoothing constant is higher than 0.3, we should check the validity of using a constant model, there may be a significant autocorrelation, a significant trend, seasonal pattern or level shift. In order to filter out the major part of the noise in the input, it is customary to use small α value. In general, the weight attached to observations must be small. However, in the case that α is small or the time series is relatively short, the weight may be large enough to have a noticeable effect on the resulting forecasts. When researchers use $\alpha = 0.1$, the initial value has a weight of 0.349 after ten iterations, which means that it is still given more weight than any other term even from the most current observation. So small smoothing constant is desired but small smoothing constants put excessive weight on the initial value. The weight on the initial value is $(1 - \alpha)^t$ and therefore will decrease faster if a higher value of α is used.

It can be seen from the table that SES for smaller α values assigns more weight to fresher data points while assigning less weight to older data points. This can also be seen from the row "AA" of the table which shows the average ages of the model under different parameter settings. The sum of squared weights is expected to increase as α. In contrast, the average age is decreasing.

N35 data set is a member of M3-competition which will be explained in section 2.10 and its time series plot is shown in Figure 2.3. As shown in graph above, the first 14 data is called "in sample" which is used for obtaining parameter estimation and also the last six data is called "out sample" which is used for comparing forecast performance.

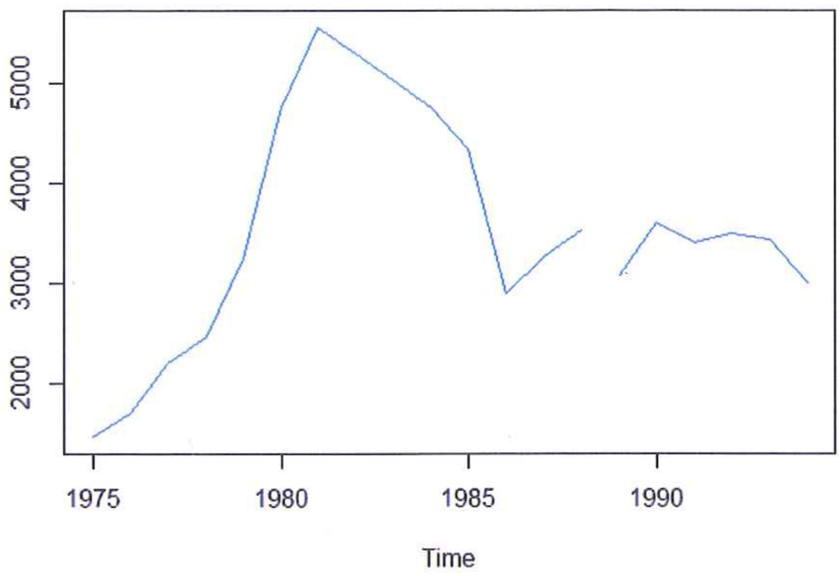

Figure 2.3 Time series plot of N35 data set.

Table 2.3 presents the application of SES for four different smoothing parameters. The last four columns show the smoothed values of data using the 0.1, 0.3, 0.5 and 0.99 alpha values for times $t = 1$ to $t = 14$ then the forecasts for $h = 1, 2, 3, 4, 5, 6$. For the first row, the initial level S_0 is set to X_1 for different alpha values. In the last coloumn both the smoothing parameter and the initial level are estimated. Using an optimization tool, the optimum smoothing parameter α^*, is estimated by minimizing sum of squared error (SSE). Additional, the accuracy measures are calculated and shown for both cases (in and out sample), which help compare forecasting performance of the methods, then again these measures will be explained in 2.8.

Table 2.3 Forecasting data N35 (M3 competition) using simple exponential smoothing with four different values for the smoothing parameter α

		Data	The smoothed values S_t			
Year	t	X_t	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.99^*$
-	0	-	1461.6	1461.6	1461.6	1461.6
1975	1	1461.6	1461.6	1461.6	1461.6	1461.6
1976	2	1692.5	1461.6	1461.6	1461.6	1461.6
1977	3	2193.8	1484.7	1530.8	1577.0	1692.5
1978	4	2459.7	1555.6	1729.7	1885.4	2193.8
1979	5	3246.8	1646.0	1948.7	2172.6	2459.7
1980	6	4748.9	1806.1	2338.1	2709.7	3246.8
1981	7	5559.5	2100.3	3061.4	3729.3	4748.9
1982	8	5292.4	2446.3	3810.8	4644.4	5559.5
1983	9	5029.4	2730.9	4255.3	4968.4	5292.4
1984	10	4753.6	2960.7	4487.5	4998.9	5029.4
1985	11	4344.6	3140.0	4567.3	4876.2	4753.6
1986	12	2897.4	3260.5	4500.5	4610.4	4344.6
1987	13	3256.4	3224.2	4019.6	3753.9	2897.4
1988	14	3525.2	3227.4	3790.6	3505.2	3256.4
	h		Forecasts			
1989	1	3070.2	3257.2	3711.0	3515.2	3525.2
1990	2	3601.6	3257.2	3711.0	3515.2	3525.2
1991	3	3407.4	3257.2	3711.0	3515.2	3525.2
1992	4	3500.6	3257.2	3711.0	3515.2	3525.2
1993	5	3437.8	3257.2	3711.0	3515.2	3525.2
1994	6	3007.0	3257.2	3711.0	3515.2	3525.2
Accuracy measures(in-sample)						
SSE			42832494	20900520	13374282	6587505
RMSE			1749.1	1221.8	977.4	685.9
MAE			1334.4	943.4	720.2	527.7
MAPE			31.6	24.9	19.9	15.1
MASE			2.3	1.7	1.3	0.9
Accuracy measures(out-sample)						
RMSE			234.7	433.8	283.2	289.6
MAE			225.9	373.5	206.5	213.2
MAPE			6.7	11.6	6.6	6.8
MASE			0.39	0.65	0.36	0.37

2.2.1 Initialization and Optimization Smoothing Parameter

Parameter selection is an important problem of the for all ES models. The value of the smoothing constant and starting value must be initialized to start the recurrence formula of S_t . There are different methods for choosing both smoothing constant and starting value but there is no proven evidence favoring any particular method.

First the smoothing constant should be chosen. It is certain that α value should fall into the interval between 0 and 1. There are two extreme cases when α is zero or one. If the coefficient $\alpha = 0$, then the current observation is ignored entirely and next period's forecast will be the same as the last period's forecast $S_t = S_{t-1}$ (which in turn is computed from the smoothed observation before it, and so on; thus all smoothed values will be equal to the initial smoothed value S_0), and if the coefficient is one then the next period's forecast will be the same as the current period's data and all the previous observations are ignored entirely, $S_t = X_t$ (naive method). In-between values will produce intermediate results. However, it is obvious that when α is close to 1 more weights are put on the recent observations and when is close to 0 more weights are put on the earlier observations. So it is crucial to choose a proper α value.

There are many theoretical and empirical arguments for selecting an appropriate smoothing value (Gardner (1985)). Gardner reports that an α smaller than 0.30 is usually recommended (Gardner (1985)). However, some studies recommend α values above 0.30 since they frequently yielded the best forecasts (Montgomery & Johnson (1976); Makridakis et al. (1982)). In practice, the smoothing parameter is often chosen by a grid search of the parameter space; that is, different solutions for α are tried starting, for example, from $\alpha = 0.1$ to $\alpha = 0.9$, with increments of 0.1. Then the α value which produces the smallest sum of squares (or mean squares) for the residuals is chosen as the smoothing constant. In addition, besides the ex post mean squared error criterion, there are other statistical measures (for example mean absolute error, or mean absolute percentage) that can be used to determine the optimum α value.

After the smoothing parameter is determined the starting value should be chosen. The weight of S_0 may be quite large when a small α is chosen and the time series relatively short. Then the choice of the starting value becomes more important. This is known as the "initialization problem". Depending on the chosen value of α , starting value can effect the quality of forecasts for many observations.

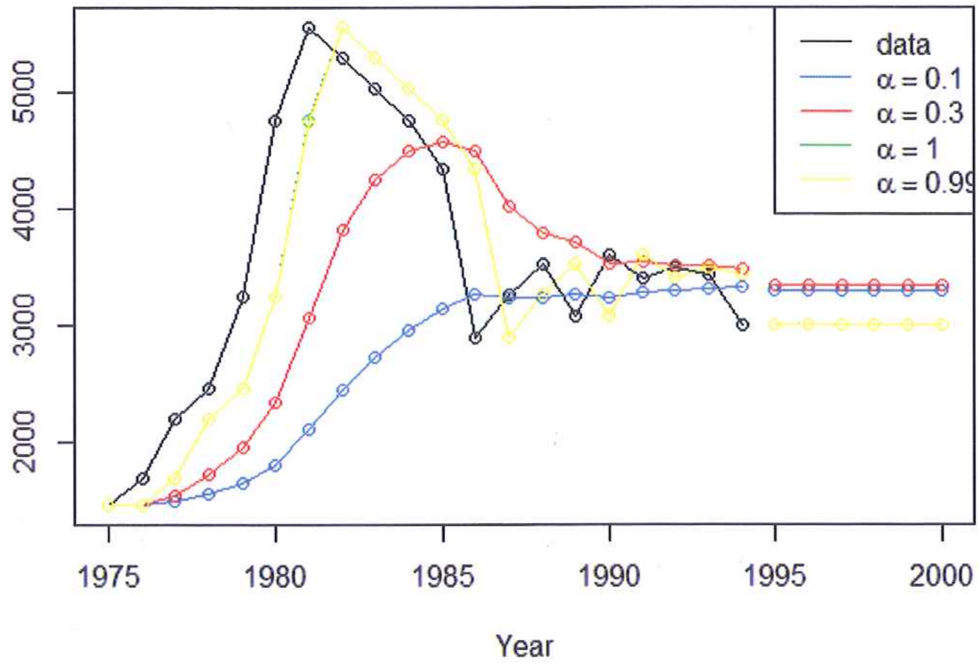

Figure 2.4 SES results of N35 data set

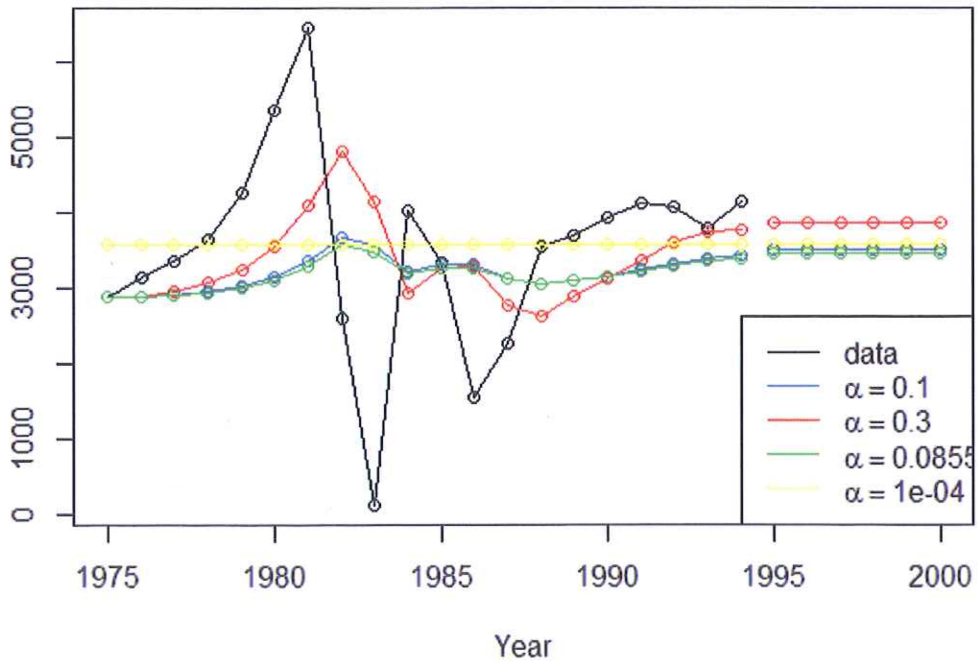

Figure 2.5 SES results of N137 data set

Figure 2.4 and Figure 2.5 compares the results obtained from the simple exponential smoothing method of N35 and N137 data sets by different smoothing and initial values. The smoothing values are "0.1", "0.3" and "optimal" and, the inital values are "simple" and "optimal" which were described in "forecast" package of R. The "black" line represents that the observations of data sets. The "blue", "red" and "green" lines respectively demonstrate the smoothed values with alpha equal to 0.1, 0.3 and "optimal" and initial value determined by the option "simple" in R. In addition, the "yellow" line illustrates the smoothed values with "optimal" alpha and initial value. When forecasting with ES, first of all we identify the choice of initial value. Then, we obtain fitted values with different smoothing parameters between 0 and 1. We obtain the optimum smoothing parameter according to accuracy measures explained in section 2.8. If we want to select an alternative initial value, we must apply this procedure again. It is obvious that the mentioned problems are confronted during the application.

We can see that the optimum smoothing parameter for N35 is equal to 0.99 in the graph above. This means that the one step forecast is the same as the last period's observation and all the previous observations are ignored as done when using Naive method. This is an undesirable situation and one of the major problems of the SES method.

Another problem with this application is that the optimum smoothing parameter is affected by the choice of the initial value. Looking at Figure 2.5, it is apparent that optimum alpha for yellow line is equal to 0.085 and equal to zero for the green line. This means that the forecasting performance of SES is therefore affected by the starting value. Looking at the literature about forecasting, it's clear that the estimate of future values are considered to be the average of historical values. Therefore, the weight given to the optimal smoothing parameter being equal to zero as shown in Figure 2.5, means that the future values are only tied to the initial value owing to implementation principle of the model is not meaningful. In conclusion, its clear that this situation is contrary to the procedure and concept of ES.

2.3 Holt's Linear Trend Method

Holt (1957) extended simple exponential smoothing to allow forecasting of data with a trend. This method involves a forecasting equation and two smoothing equations (one for the level and one for the trend):

$$S_t = \alpha X_t + (1-\alpha)(S_{t-1} + T_{t-1}), \tag{2.17}$$

$$T_t = \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1}, \tag{2.18}$$

$$\hat{X}_t(h) = S_t + hT_t, \tag{2.19}$$

where X_t is the actual observation, $\hat{X}_t(h)$ is the h-step-ahead forecast, α and β are smoothing parameters, $0 < \alpha, \beta < 1$. There are two smoothing parameters to estimate and starting values for both the level and trend must be provided. The parameters and initial values can be estimated by minimizing the one step MSE, MAE, MAPE or some other criterion for measuring in-sample forecast error.

As with simple exponential smoothing, the level equation here shows that S_t is a weighted average of observation X_t and the with in-sample one-step-ahead forecast for time t, here given by $S_{t-1} + T_{t-1}$. The trend equation shows that T_t is a weighted average of the estimated trend at time t based on $S_t - S_{t-1}$ and T_{t-1}, the previous estimate of the trend. The weights α and β can be chosen by minimizing the value of MSE or some other criterion. Optimization we could evaluate the MSE over a grid of values of α and β and then select the combination of α and β which correspond to the lowest MSE. The forecast function is no longer flat but trending. The h-step-ahead forecast is equal to the last estimated level plus h times the last estimated trend value. Hence the forecasts are a linear function of h.

One interesting special case of this method occurs when $\beta = 0$, which is known as "SES with drift," which is closely related to the "Theta method" of forecasting due to Assimakopoulos and Nikolopoulos(2000). The connection between these methods was demonstrated by Hyndman and Billah (2003). The method as follow:

$$S_t = \alpha X_t + (1 - \alpha)(S_{t-1} + T), \quad (2.20)$$

$$\hat{X}_t(h) = S_t + hT. \quad (2.21)$$

2.4 Exponential Trend Method

A variation of Holt's linear trend method can be achieved by allowing the level and the slope to be multiplied rather than added:

$$S_t = \alpha X_t + (1-\alpha)(S_{t-1} * T_{t-1}), \tag{2.22}$$

$$T_t = \beta(S_t/S_{t-1}) + (1-\beta)T_{t-1}, \tag{2.23}$$

$$\hat{X}_t(h) = S_t * T_t^h. \tag{2.24}$$

where X_t is the actual observation, $\hat{X}_t(h)$ is the h-step-ahead forecast, α and β are smoothing parameters, $0 < \alpha, \beta < 1$. The local growth, T_t, by smoothing successive differences, (S_t/S_{t-1}) of the local level, S_t

2.5 Damped Trend Method

Damped trend method was proposed by Gardner Jr & McKenzie (1985) with a modification of Holt's linear method to allow the "damping" of trends. The formulation of this method as follows:

$$S_t = \alpha X_t + (1-\alpha)(S_{t-1} + \phi T_{t-1}), \tag{2.25}$$

$$T_t = \beta(S_t - S_{t-1}) + (1-\beta)\phi T_{t-1}, \tag{2.26}$$

$$\hat{X}_t(h) = S_t + (\phi + \phi^2 + \phi^3 + \ldots + \phi^h)T_t. \tag{2.27}$$

The damped trend method can be applied with multiplicative trend factor. Taylor (2003) suggested a damping parameter to the exponential trend method resulting in a multiplicative damped trend method:

$$S_t = \alpha X_t + (1-\alpha)(S_{t-1} * T_{t-1}^{\phi}), \tag{2.28}$$

$$T_t = \beta(S_t/S_{t-1}) + (1-\beta)T_{t-1}^{\phi}, \tag{2.29}$$

$$\hat{X}_t(h) = S_t + T_t^{(\phi+\phi^2+\phi^3+\ldots+\phi^h)}. \tag{2.30}$$

This method will produce less conservative forecasts than the additive damped trend method when compared to Holt's linear method.

2.6 A Taxonomy of Exponential Smoothing Method

The classification of ES models was proposed by Pegels (1969). It was later extended by Gardner Jr & McKenzie (1985) to include methods with additive damped trend and by Taylor (2003) to include methods with multiplicative damped trend. There are fifteen exponetial smoothing methods which have five types of trend components (none, additive, additive damped, multiplicative, multiplicative damped) and three types of seasonal components (none, additive, multiplicative) which are given in Table 2.4.

Table 2.4 Classification of ES methods.

	Seasonal Component		
Trend Component	N (None)	A (Additive)	M (Multiplicative)
N (None)	N,N	N,A	N,M
A (Additive)	A,N	A,A	A,M
Ad (Additive damped)	Ad,N	Ad,A	Ad,M
M (Multiplicative)	M,N	M,A	M,M
Md (Multiplicative damped)	Md,N	Md,A	Md,M

Some of these methods we have already seen:

$$\begin{aligned}
(N, N) &= simple \quad exponential \quad smoothing \\
(A, N) &= Holts \quad linear \quad method \\
(M, N) &= Exponential \quad trend \quad method \\
(Ad, N) &= additive \quad damped \quad trend \quad method \\
(Md, N) &= multiplicative \quad damped \quad trend \quad method \\
(A, A) &= additive \quad Holt - Winters \quad method \\
(A, M) &= multiplicative \quad Holt - Winters \quad method \\
(Ad, M) &= Holt - Winters \quad damped \quad method
\end{aligned}$$

Table 2.5 Formulae for recursive calculations and point forecasts.

Trend	Seasonal N	Seasonal A	Seasonal M
N	$\hat{y}_{t+h\|t} = \ell_t$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}$ $s_t = \gamma(y_t/\ell_{t-1}) + (1-\gamma)s_{t-m}$
A	$\hat{y}_{t+h\|t} = \ell_t + hb_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + hb_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + hb_t)s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + b_{t-1})) + (1-\gamma)s_{t-m}$
$\mathbf{A_d}$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t$ $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t + \phi_h b_t + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = (\ell_t + \phi_h b_t)s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)(\ell_{t-1} + \phi b_{t-1})$ $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1} + \phi b_{t-1})) + (1-\gamma)s_{t-m}$
M	$\hat{y}_{t+h\|t} = \ell_t b_t^h$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h\|t} = \ell_t b_t^h + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t b_t^h s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}$ $s_t = \gamma(y_t/(\ell_{t-1}b_{t-1})) + (1-\gamma)s_{t-m}$
$\mathbf{M_d}$	$\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h}$ $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$	$\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h} + s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $s_t = \gamma(y_t - \ell_{t-1}b_{t-1}^{\phi}) + (1-\gamma)s_{t-m}$	$\hat{y}_{t+h\|t} = \ell_t b_t^{\phi_h} s_{t-m+h_m^+}$ $\ell_t = \alpha(y_t/s_{t-m}) + (1-\alpha)\ell_{t-1}b_{t-1}^{\phi}$ $b_t = \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}^{\phi}$ $s_t = \gamma(y_t/(\ell_{t-1}b_{t-1}^{\phi})) + (1-\gamma)s_{t-m}$

We note that the level, trend and seasonal components are respectively denoted by symbols l, b, s instead of S, T, I in Table 2.5. Therefore, the observations at time t are denoted by y_t instead of X_t

Table 2.5 gives the recursive formulae for applying all possible fifteen exponential smoothing methods (Hyndman & Athanasopoulos (2014)). Each cell includes the forecasting equation for generating h-step-ahead forecasts and the smoothing equations for applying the method. In each case, l_t denotes the series level at time t, b_t denotes the slope at time t, s_t denotes the seasonal component of the series at time t, and m denotes the number of seasons in a year; α, β^*, γ and ϕ are smoothing parameters and $h_+^m = [(h-1) mod \quad m] + 1$.

Hyndman & Athanasopoulos (2014) presents some strategies for selecting initial values for some of the most commonly applied exponential smoothing methods, for instance the method (N,N) uses $l_0 = y_1$ and also the method (M,N) or (Md,N) $l_0 = y_1, b_0 = \frac{y_1}{y_2}$.

2.7 Innovation State Space Methods Underlying Exponential Smoothing

Hyndman et al. (2002) introduced a state space framework that subsumes all the exponential smoothing models and allows for the computation of prediction intervals, likelihood and model selection criteria. Hyndman et al. (2008) describe two possible innovative state space models for each of the 15 exponential smoothing methods, one corresponding to a model with additive errors and the other to a model with multiplicative errors. Now, there are 30 potential models described in this classification. The notation ETS(*,*,*) is used to identify these exponential smoothing models, where the triplet (*,*,*) stands for possible error (E), trend (T) and seasonal (S) combinations respectively. All ETS models can be written in innovations state space form Hyndman et al. (2002) and also shown in Table 2.6.

Table 2.6 The state space models and their formulation

ADDITIVE ERROR MODELS

Trend	Seasonal N	Seasonal A	Seasonal M
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
A_d	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$
M	$y_t = \ell_{t-1}b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$	$y_t = \ell_{t-1}b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1})$
M_d	$y_t = \ell_{t-1}b_{t-1}^{\phi} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi} + \alpha\varepsilon_t$ $b_t = b_{t-1}^{\phi} + \beta\varepsilon_t/\ell_{t-1}$	$y_t = \ell_{t-1}b_{t-1}^{\phi} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi} + \alpha\varepsilon_t$ $b_t = b_{t-1}^{\phi} + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}^{\phi}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1}^{\phi} + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1}^{\phi})$

MULTIPLICATIVE ERROR MODELS

Trend	Seasonal N	Seasonal A	Seasonal M
N	$y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A_d	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M	$y_t = \ell_{t-1}b_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$	$y_t = (\ell_{t-1}b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M_d	$y_t = \ell_{t-1}b_{t-1}^{\phi}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^{\phi}(1 + \beta\varepsilon_t)$	$y_t = (\ell_{t-1}b_{t-1}^{\phi} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi} + \alpha(\ell_{t-1}b_{t-1}^{\phi} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1}^{\phi} + \beta(\ell_{t-1}b_{t-1}^{\phi} + s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1}^{\phi} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}^{\phi}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^{\phi}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^{\phi}(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

Estimating ETS Models: An alternative to estimating the parameters by minimizing the sum of squared errors, is to maximize the "likelihood". The likelihood is the probability of the data arising from the specified model. So a large likelihood is associated with a good model. For an additive error model, maximizing the likelihood gives the same results as minimizing the sum of squared errors. However, different results will be obtained for multiplicative error models.

Model Selection: A great advantage of the ETS statistical framework is that information criteria can be used for model selection. The Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) can be used here to determine which of the 30 ETS models is most appropriate for a given time series. For ETS models, Akaike's Information Criterion (AIC) is defined as:

$$AIC = -2log(Likelihood) + 2p \tag{2.31}$$

where $p =$ # the total number of parameters and initial states that have been estimated. To sum up, Hyndman et al. (2008) proposed an automated procedure and the software developed for ETS is available in the R "forecast" package.

2.8 Forecast Accuracy

The forecasting accuracy should be tested according to different perspectives. First is the goodness of fit, which shows how well the model is able to reproduce the actual known data. On the other hand, the out of sample perspective shows the predictive accuracy to unknown data. In order to measure the out of sample accuracy, the full amount of data is separated into a training and test set. The training set is used for the estimating the parameters of the forecasting model. First, the model is formulated, then the data of the training set are initialized and the parameters of the model are optimised by the most appropriate method (depending on the model) and according to the values of the data. Then, the model is ready to generate forecasts for the test data set. The out of sample forecast accuracy is then determined by comparing the forecasts with the actual data, which have not been used for the model development (Makridakis S. & Hyndman (1998)).

The forecasting error can be calculated as:

$$e_t = Y_t - F_t \tag{2.32}$$

with e_t is the forecasting error, Y_t the actual value and F_t the forecast for period t.

There are four types of forecast-error metrics: scale-dependent metrics such as the mean absolute error (MAE or MAD); percentage-error metrics such as the mean absolute percent error (MAPE); relative-error metrics, which average the ratios of the errors from a designated method to the errors of a naïve method; and scale-free error metrics, which express each error as a ratio to an average error from a baseline method.

Scale-dependent measures: Accuracy measures that are based on e_t are therefore scale-dependent and cannot be used to make comparisons between series that are on different scales. These are useful when comparing different methods applied to the same set of data. The most commonly used scale-dependent measures are based on the absolute errors or squared errors:

Mean square error:

$$MSE = mean(e_t^2) \tag{2.33}$$

Root mean squared error:

$$RMSE = \sqrt{MSE} \tag{2.34}$$

Mean absolute error:

$$MAE = mean(|e_t|) \tag{2.35}$$

Median absolute error:

$$MdAE = median(|e_t|) \tag{2.36}$$

RMSE and MSE have been popular, largely because of their theoretical relevance in statistical modelling. In performance of M-competition, MSE was used by Makridakis et al. (1982). Inappropriate using of MSE was largely discussed (Chatfield & Yar (1988); Armstrong & Collopy (1992)). It has been noted that they are more sensitive to outliers than MAE or MdAE (Armstrong (2001)). When comparing forecast methods on a single data set, the MAE is popular as it is easy to understand and compute.

Measures based on percentage errors: The percentage error is given by

$$p_t = 100e_t/Y_t. \tag{2.37}$$

Percentage errors have the advantage of being scale independent, so they are frequently used to compare forecast performance between different data series. The most widely used measures are:

Mean absolute percentage error:

$$MAPE = mean(|p_t|) \tag{2.38}$$

Median absolute percentage error:

$$MdAPE = median(|p_t|) \tag{2.39}$$

But measurements based on percentage errors have the disadvantage of being infinite or undefined if there are zero values in a series. Furthermore, measures can have an extremely skewed distribution when actual values are close to zero (Hyndman & Koehler (2006)).

The MAPE has another disadvantage: it puts a heavier penalty on positive errors than on negative errors. This observation has led to the use of the "symmetric" MAPE (sMAPE) in the M3-competition (Makridakis & Hibon (2000)). It is defined as the Symmetric mean absolute percentage error:

$$sMAPE = mean(200|Y_t - F_t|/(Y_t + F_t)) \tag{2.40}$$

Symmetric median absolute percentage error:

$$sMdAPE = median(200|Y_t - F_t|/(Y_t + F_t)). \tag{2.41}$$

However, if the actual value is zero, the forecast is likely to be close to zero. Thus 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.

Measures based on relative errors: An alternative to percentages for the calculation of scale independent measurements involves dividing each error by the error obtained using some benchmark method of forecasting. Let

$$r_t = e_t/e_t^* \tag{2.42}$$

denote the relative error where e_t^* is the forecast error obtained from the benchmark method. Usually the benchmark method is the naive method where F_t is equal to the last observation. Then, the relative measures are:
Mean relative absolute error:

$$MRAPE = mean(|p_t|) \tag{2.43}$$

Median relative absolute error:

$$MdRAPE = median(|p_t|) \tag{2.44}$$

These relative-error metrics were suggested in studies by Armstrong and Collopy (1992) for evaluating forecast accuracy across multiple series. A serious deficiency of relative error measures is that e_t^* can be small. In fact, r_t has infinite variance because e_t^* has positive probability density at 0. One common special case is when e_t and e_t^* are normally distributed, in which case r_t has a Cauchy distribution (Hyndman & Koehler (2006)).

Scaled errors: Relative error measures have problems (Hyndman & Koehler (2006)). One of the first problem is that relative errors have a statistical distribution with undefined mean and infinite variance. Second, they can only be computed when there are several forecasts on the same series, and so cannot be used to measure out of sample forecast accuracy at a single forecast horizon. Scaled errors were proposed by Hyndman and Koehler (2006) as an alternative to using percentage errors when comparing forecast accuracy across series on different scales. They proposed scaling the errors based on the training MAE from a simple forecast method. For a non-seasonal time series, a useful way to define a scaled error uses naive forecasts:

$$q_t = \frac{e_t}{\frac{1}{n-1}\sum_{i=1}^{n}|Y_i - Y_{i-1}|} \tag{2.45}$$

Because the numerator and denominator both involve values on the scale of the original data, a scaled error is independent of the scale of the data. A scaled error is less than one if it arises from a better forecast than the average naive forecast computed on the training data. Conversely, it is greater than one if the forecast is worse than the average naive forecast computed in sample. For seasonal time series, a scaled error can be defined using seasonal naive forecasts:

$$q_t = \frac{e_t}{\frac{1}{n-m}\sum_{i=m+1}^{n}|Y_i - Y_{i-m}|} \tag{2.46}$$

m is a seasonal period such as 4 for quarterly seasonality or 12 for monthly seasonality. And so, the mean absolute scaled error is simply

$$MASE = mean|q_t| \tag{2.47}$$

Similarly, the mean squared scaled error (MSSE) or the median absolute scaled error (MdASE) can be defined where the errors (on the training data and test data) are squared instead of using absolute values. In studies Hyndman (2003) recommended that the MASE can be used to compare forecast methods on a single or multiple series and also it never gives infinite or undefined values.

2.9 Combined Forecasting

Combinations of forecasts were introduced by Bates and Granger (1969) and it is a very common way to improve the forecasting accuracy. The forecasts that are combined can be based on different data or different techniques. The main idea of combining forecasts lies in the fact that different forecasting methods contain useful and independent information. It is well known that combining forecasts (Bates & Granger (1969); Clemen (1989)) under certain circumstances improves forecasting accuracy (Armstrong (1989); Armstrong (2001); Makridakis & Winkler (1983); Makridakis et al. (1982)). According to Armstrong (2001) the areas of expert forecasting and econometric forecasting have proved good evidence about the improvement of forecasting accuracy through combining individual forecasts. Moreover, combining forecasts has been very useful when it is difficult to select the most accurate forecasting method. It has been shown a good way of hedging the risk in situations of very expensive forecasting errors (Armstrong (2001)). Makridakis (1989) states that the accuracy of an individual forecast is sensitive to several factors that may affect the accuracy.

The research since the beginning of M-competition mainly focuses on certain transformations, decompositions, rules and combinations of ES and ARIMA to improve the forecasting performance rather than proposing new forecasting methods.

Combining can be expressed mathematically as follows:

$$F_c(t) = \sum_{i=1}^{k} w_i F_i(t) \tag{2.48}$$

$$with \quad \sum_{i=1}^{k} w_i = 1 \tag{2.49}$$

where there are k forecasts that are combined. $F_c(t)$ is the combined forecast at time t, $F_i(t)$ is the result of forecast i $(0 \leq i \leq k)$ and w_i is the weight of forecast i $(0 \leq w_i \leq 1)$. Researchers (e.g. Newbold and Bos, 1994, Russel and Adam, 1987 and de Menezez et al., 2000) agree that the most common methods to estimate the values of the combining weights are simple average-equal weights, inversely proportional weights, regression-based weights and weights based on the absolute error. In this thesis we will focus on just the simple average-equal weights case.

Simple average equal weights case: The simplest way to combine individual forecasts is to assign them equal weights. Hence:

$$F_ct = \frac{\sum_{i=1}^{k} F_it}{k} \tag{2.50}$$

$$thus \quad w_i = \frac{1}{k} \tag{2.51}$$

Note that, simple average combination will be used in this thesis to improve the forecasting performance of the proposed method on the competition data set.

2.10 The Forecasting M-Competitions

There are many studies on the numerical and theoretical comparison of Box-Jenkins and ES methods. Several empirical studies have been published in turn by Reid (1969); Newbold & Granger (1974); Makridakis & Hibon (1979); Makridakis et al. (1982); Makridakis et al. (1993); Makridakis & Hibon (2000).

Newbold & Granger (1974) compared the forecasting performance of three methods over a large sample of economics time series. These methods are Box Jenkins, Holt-Winters and stepwise autoregression. The study explored the possibility of combining individual forecasts in the production of an overall forecast and presented empirical results which indicated that such a procedure can frequently be profitable.

111 time series, were used for investigating why some methods achieve greater accuracy than others for different types of the data (Makridakis & Hibon (1979)), which is a subgroup selected from M1-competition (1001 series). Surprisingly, it has been shown that for these simpler methods perform well in comparison to the more complex and statistically sophisticated ARMA models. This study was the first large scale empirical evaluation of time series forecasting methods and highly controversial at the time.

Then, the M-Competition was established by Makridakis in 1982 in a paper which studied the post-sample accuracy of several time series forecasting methods

(Makridakis et al. (1982)). The number of series was increased to 1001 and the data were subdivided into various categories (micro, macro, industry, demography, finance, other.). The participants tested the accuracy of 24 methods on 1001 series with various horizons which were six for yearly data, eight for quarterly data and eighteen for monthly data. The objective of the competition was to investigate how different techniques differ from each other and how information can be provided so that forecasters can be able to make practical choices under different circumstances (Makridakis et al. (1984)). According to Makridakis et al. (1984) the most significant conclusions of the competition were the following:

- *"Statistically sophisticated or complex methods do not necessarily provide more accurate forecasts than simpler ones.*
- *The relative ranking of the performance of the various methods varies according to the accuracy measure being used.*
- *The accuracy when various methods are being combined outperforms, on average, the individual methods being combined and does very well in comparison to other methods. Combined forecasting reduces the forecasting error significantly.*
- *The accuracy of the various methods depends upon the length of the forecasting horizon involved.*
- *The nature of the series, such as the period (e.g. monthly, yearly) and data types (e.g. financial, demand) affects the forecasting accuracy of different techniques."*

The M2 Competition consisted of distributing 29 actual series (23 of these series came from four companies and six were of macro economic nature) to five forecasters. The purpose of the M2-Competition is to determine the post sample accuracy of various forecasting methods. It is an empirical study organized in such a way as to avoid the major criticism of the M-Competition that forecasters in real situations can use additional information to improve the predictive accuracy of quantitative methods (Makridakis et al. (1993)). Although the forecasters had additional information about the series being predicted the results show few or no differences in post-sample forecasting accuracy when compared to those of the M-Competition or the earlier Makridakis and Hibon empirical study.

In Makridakis & Hibon (2000), third competition was launched. It explains the reasons for conducting the competition and summarizes its results and conclusions. In addition, the paper compares such results / conclusions with those of the previous two M-Competitions as well as with those of other major empirical studies. 3003 series consist of 6 different types of series and 4 different time intervals between successive observations. The time series are classified according to time interval in Table 2.7 and types of the data in Table 2.8

Table 2.7 The classificaton of the 3003 time series data

Time interval between successive observations	Types of time series data						
	Micro	Industry	Macro	Finance	Demographic	Other	Total
Yearly	146	102	83	58	245	11	645
Quarterly	204	83	336	76	57		756
Monthly	474	334	312	145	111	52	1428
Other	4			29		141	174
Total	828	519	731	308	413	204	3003

Table 2.8 Average size of the data

Types of time series data	Time interval between successive observations				
	Yearly	Quarterly	Monthly	Other	Total
Demographic	27	56	123		57
Finance	36	52	124	95	87
Industry	43	60	140		108
Macro	23	47	131		80
Micro	20	44	93	104	68
Other	37		83	72	73
Total	28	49	117	77	79

Table 2.7 and Table 2.8 show that the series have been obtained on a quota basis:

- All time series are positive.
- 3003 series have different time intervals (yearly, quarterly, monthly, other).
- Forecasting results are examined for different six subgroup (seasonal, non-seasonal, yearly, quarterly, monthly or other).
- 3003 series have various types of time series data (micro, macro, industry, etc.).
- Series length between 20 and 140.
- The time horizons of forecasting are 6 periods for yearly, 8 periods for quarterly, 18 periods for monthly, 8 periods for other data.
- For 3003 series, the 5 accuracy measures utilized.

The different methods as shown in Table 2.9 have been classified in the following categories:

- Naive, simple methods
- Explicit trend methods
- Decomposition
- ARIMA/ARARMA models
- Expert systems
- Neural networks

The results of this competition indicate that the THETA method obtained the best forecasting performance among the other methods. Moreover, ForecastPRO, ForecastX, Comb S-H-D, Dampen, RBF, ETS, THETAsm and B-J automatic respectively followed the performance of THETA.

Table 2.9 A list of the 24 methods that have been used in the competition with competitors and a short description

Method	Competitors	Description
Naıve ¨ /simple		
1. Naıve2	M. Hibon	Deseasonalized Naıve (Random Walk)
2. Single	M. Hibon	Single Exponential Smoothing
Explicit trend models		
3. Holt	M. Hibon	Automatic Holt's Linear Exponential Smoothing (two parameter model)
4. Robust-Trend	N. Meade	Non-parametric version of Holt's linear model with median based estimate of trend
5. Winter	M. Hibon	Holt–Winter's linear and seasonal exponential smoothing (two or three parameter model)
6. Dampen	M. Hibon	Dampen Trend Exponential Smoothing
7. PP-autocast	H. Levenbach	Damped Trend Exponential Smoothing
8. Theta-sm	V. Assimakopoulos	Successive smoothing plus a set of rules for dampening the trend
9. Comb S-H-D	M. Hibon	Combining three methods: Single/Holt/Dampen
Decomposition		
10. Theta	V. Assimakopoulos	Specific decomposition technique, projection and combination of the individual components
ARIMA/ARARMA model		
11. B–J automatic	M. Hibon	Box–Jenkins methodology of 'Business Forecast System' a
12. Autobox1	D. Reilly	Robust ARIMA univariate Box–Jenkins a
13. Autobox2	with/without	Intervention Detection a
14. Autobox3		
15. AAM1	G. Melard,	Automatic ARIMA modelling with/without
16. AAM2	J.M. Pasteels	intervention analysis
17. ARARMA	N. Meade	Automated Parzen's methodology with Auto regressive filter Expert system a
18. ForecastPro	R. Goodrich, E. Stellwagen	Selects from among several methods: Exponential Smoothing/Box Jenkins/Poisson and negative binomial models/Croston's Method/Simple Moving Average
19. SmartFcs	C. Smart	Automatic Forecasting Expert System which conducts a forecasting tournament among four exponential smoothing and two moving average methods
20. RBF	M. Adya, S. Armstrong, F. Collopy, M. Kennedy	Rule-based forecasting: using three methods — random walk, linear regression and Holt's, to estimate level and trend, involving corrections, simplification, automatic feature identification and re-calibration
21. Flores/Pearce1	B. Flores,	Expert system that chooses among four methods
22. Flores/Pearce2	S. Pearce	based on the characteristics of the data a
23. ForecastX	J. Galt	Runs tests for seasonality and outliers and selects from among several methods: Exponential Smoothing, Box–Jenkins and Croston's method Neural networks
24. Automat ANN	K. Ord,S. Balkin	Automated Artificial Neural Networks for forecasting purposes

The aims of the M3-Competition were to clear up the accuracy issues of several forecasting techniques and to extend the results of the previous competition. The extension involved the use of additional techniques (mainly from the area of artificial intelligence), more practitioners and more series (3003). The findings of the M3 confirmed the conclusions of the previous competitions. Particularly:

I. "Sophisticated techniques are slightly better than simpler techniques for time series with limited variability.
II. The performance of a forecasting method differs according to the used accuracy measure.
III. The accuracy of combined forecasting is usually as good as or better than the accuracy of the individual methods that were combined as well as than the accuracy of other methods.
IV. The best method for a series depends on the forecasting horizon, type of the data and the category of the data.
V. Some specific new methods not used in the M- Competition perform consistently better than the others specific circumstances.
VI. The performance of the different methods does not significantly differ for short, medium an long term."

The three competitions have played a very important role in the forecasting research the last three decades and the results of the M-competitions were similar to those of the earlier Makridakis and Hibon study. Their results provided a basis for future forecasting research. In September 2010, Makridakis et al. launched a fourth competition. According to the competition team, the purpose of the M4 Competition is to further study the accuracy and the utility of several forecasting techniques. For this reason, the number of the series, the categories and the forecasting techniques are increased. The results of the competition have not been published yet.

CHAPTER THREE
ATA METHOD

The development of accurate and robust forecasting methods for univariate time series is very important when large numbers of time series are involved in the modelling and forecasting process. Despite the advantages of model selection algorithms, there is still a need for accurate extrapolation methods. Forecasting competition have played an important role in moving towards the forecasting of large numbers of time series, with the objectives of identifying high performing methods. The ATA method will attract a great deal of attention by simplicity, easy optimization and surprisingly good performance. The ATA method can be applied to non-seasonal or deseasonalized time series, where the deseasonalisation is performed via the multiplicative classical decomposition. The simplest form of ATA method will be introduced in this chapter but the higher order forms will be presented in the next chapter.

The ATA method has similar form to ES but the smoothing parameters are modified so that when obtaining a smoothed value at a specific time point the weights among the observations are distributed taking into account how many observations can contribute to the value being smoothed. Therefore the smoothing parameter for this method is a function of t unlike exponential smoothing where no matter where the value you are smoothing resides on the time line, the observations receive weights only depending on their distances from the value being smoothed. For the series X_t, $t = 1, \ldots, n$, the general additive ATA method which we will denote by $ATA(p, q)$ throughout the thesis can be written as:

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

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

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

for $p \in \{1, \ldots, n\}$, $q \in \{0, 1, \ldots, p\}$ and $t > p \geq q$. For $t \leq p$ let $S_t = X_t$, for $t \leq q$ let $T_t = X_t - X_{t-1}$ and let $T_1 = 0$ where X_t is the actual observation of the series, S_t denotes an estimate of the level of the series at time t, T_t denotes an estimate of the growth (trend) value of the series at time t, p is the smoothing parameter for the level and q is the smoothing parameter for the trend. A multiplicative version of the same model $ATA_{mult}(p, q)$ which can be given 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 $p \in \{1, \ldots, n\}$, $q \in \{0, 1, \ldots, p\}$ and $t > p \geq q$. For $t \leq p$ let $S_t = X_t$, for $t \leq q$ let $T_t = X_t / X_{t-1}$ and let $T_1 = 1$.

Note that there are two smoothing parameter (p and q) to estimate but no initial values are needed for level and trend, and also ATA methods can be adapted for each of 30 different ES models classified by Hyndman et al. (2008). In this thesis the focus will be on modelling the trend component for additive and multiplicative form of the ATA method. The ATA method can be applied to non seasonal or deseasonalized time series, where the deseasonalisation is usually performed via the multiplicative classical decomposition.

3.1 Simplest Form of $ATA(p, 0)$

It is worth pointing out that when $q = 0$, $ATA(p, q)$ reduces to a simple model that has similar form to SES, i.e. for $t > p$:

$$S(t) = \left(\frac{p}{t}\right) X_t + \left(\frac{t-p}{t}\right) S_{t-1}, \qquad t > p, \tag{3.7}$$

$$S_t = X_t, \qquad t \leq p, \tag{3.8}$$

$$\hat{a} = \hat{X}_t(h) = S_t. \tag{3.9}$$

where p is the smoothing parameter and it regulates the smoothing process. The model will be called simple form of $ATA(p,0)$ henceforth. S_t can be interpreted as a weighted average of past observations. The model in (3.7) is applied recursively to all successive observations in the series as below:

$$\begin{aligned}
S_{t-1} &= \left(\frac{p}{t-1}\right) X_{t-1} + \left(\frac{t-p-1}{t-1}\right) S_{t-2}, \\
S_t &= \left(\frac{p}{t}\right) X_t + \left(\frac{p}{t}\right)\left(\frac{t-p}{t-1}\right) X_{t-1} + \left(\frac{t-p}{t}\right)\left(\frac{t-p-1}{t-1}\right) S_{t-2}, \\
S_{t-2} &= \left(\frac{p}{t-2}\right) X_{t-2} + \left(\frac{t-p-2}{t-2}\right) S_{t-3}, \\
S_t &= \left(\frac{p}{t}\right) X_t + \left(\frac{p}{t}\right)\left(\frac{t-p}{t-1}\right) X_{t-1} + \left(\frac{p}{t}\right)\left(\frac{t-p}{t-1}\right)\left(\frac{t-p-1}{t-2}\right) X_{t-2} \\
&\quad + \left(\frac{t-p}{t}\right)\left(\frac{t-p-1}{t-1}\right)\left(\frac{t-p-2}{t-2}\right) S_{t-3}, \\
&\vdots \\
S_t &= \left(\frac{p}{t}\right) X_t + \left(\frac{p}{t}\right)\left(\frac{t-p}{t-1}\right) X_{t-1} + \dots \\
&\quad + \frac{\binom{p}{p-1}}{\binom{t}{p}} X_{p+1} + \left(\frac{t-p}{t}\right)\left(\frac{t-p-1}{t-1}\right)\dots\left(\frac{1}{p+1}\right) S_p.
\end{aligned}$$

Therefore, the smoothed value S_t can be written in the recursive form:

$$S_t = \sum_{k=0}^{t-(p+1)} \frac{\binom{t-k-1}{p-1}}{\binom{t}{p}} x_{t-k} + \frac{1}{\binom{t}{p}} S_p, \tag{3.10}$$

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 now easily be seen that the smoothed value at time t is a weighted average of past observations and the initial value S_p. Looking at equation (3.10), it is apparent that the initial value of S_p depends on the smoothing parameter of p. Hence, the smoothing parameter and initial value are optimized simultaneously. This process will be examined in details section (3.1.6). In the following sections, the features of $ATA(p,0)$ will be examined comparatively with

SES.

3.1.1 Weights of $ATA(p,0)$

We have already seen that simple exponential smoothing model is the appropriate model to forecast a constant process. SES model has some desirable features such as attaching exponentially decreasing weights to the observations. The weights must have some significant properties such as below:

- $w_t \in [0,1] \qquad t = 1,\ldots,n$
- $w_1 \leqslant w_2 \leqslant \ldots \leqslant w_n$
- $\sum_{t=1}^{n} w_t = 1.$

The smoothing parameter p is smaller or equal to t so the first condition is satisfied. The weights attached to observations are sequentially given in Table 3.1. It is clear that $w_i \leqslant w_j, \quad i \leqslant j$

Table 3.1 The comparison of weights attached to the observations by SES and $ATA(p,0)$

weight of X_t	
$ATA(p,0)$	SES
$w_n = \left(\frac{p}{n}\right)$	$w_n = \alpha$
$w_{n-1} = \left(\frac{p}{n}\right)\left(\frac{n-p}{n-1}\right)$	$w_{n-1} = \alpha\,(1-\alpha)^1$
$w_{n-2} = \left(\frac{p}{n}\right)\left(\frac{n-p}{n-1}\right)\left(\frac{n-p-1}{n-2}\right)$	$w_{n-2} = \alpha\,(1-\alpha)\,2$
$\vdots$	$\vdots$
$w_{p+1} = \frac{\binom{p}{p-1}}{\binom{n}{p}},$	$w_2 = \alpha\,(1-\alpha)^{n-1},$
$w_p = \ldots = w_1 = 0$	$w_1 = \alpha\,(1-\alpha)^n$

Comparing both columns of Table 3.1, it can be seen that the second condition is provided $w_1 \leqslant w_2 \leqslant \ldots \leqslant w_n$. It should be showed that the sum of the weights must

and also the variance is

$$V(X) = m\frac{(N+1)(N-M)}{(M+1)(M+2)}\left(1 - \frac{m}{M+1}\right). \tag{3.13}$$

Correspondingly, the weights given to the observations follow a Negative Hyper-Geometric distribution with parameters $(n, p, 1)$ instead of (N, M, m). From the distribution we can see that the weighting scheme sum to unity and as a result last condition is provided. The results in this section indicate that all desirable properties of weights attached to observations are satisfied by ATA method like SES. The next section, therefore, moves on to the discussion of other properties of ATA in comparison to SES.

3.1.2 Average Age of $ATA(p,0)$, $(\bar{k})$

The average age $(\bar{k})$ of a model is the measure of the model's ability to utilize fresh data and is a well known metric for comparing forecasting models. The smaller the $\bar{k}$ the better. Brown (1959) defines the average age as in equation (2.3) where w_t is the weight given to the t^{th} observation when trying to obtain the forecast. Utilizing the expected value of the distribution shown in equation (3.11), the average age of $ATA(p,0)$ can then be easily found as:

$$AA_{ATA(p,0)} = \frac{n-p}{p+1}. \tag{3.14}$$

If SES and ATA are compared at the same α level $(\alpha = \frac{p}{n})$, then $\bar{k}_{ATA} < \bar{k}_{SES}$ since

$$\frac{n-p}{p+1} < \frac{(1-\alpha)}{\alpha} = \frac{n-p}{p}$$

Therefore, when the average age of data used from SES and $ATA(p,0)$ are compared at the same α level, it is obvious that $ATA(p,0)$ is always younger than SES at the same smoothing constant.

3.1.3 The Sum of the Squared Weights of $ATA(p,0)$, V

The variance of the estimates, should be taken into account to compare forecasting methods and the variance can be calculated as in equation (2.4) (Brown (1959)), which is another well known metric like the average age of the model. In order to calculate the variance of the ATA estimator, it is needed to calculate the sum of squared weights, V, once again. With a slight re-arrangement of the numerators in the weights, the sum can be written as:

$$\begin{aligned} V_{ATA(p,0)} &= \sum_{t=1}^{n} w_t^2 \\ &= \left(\frac{p}{n}\right)^2 + \left(\frac{p}{n}\right)^2 \left(\frac{n-p}{n-1}\right)^2 + \left(\frac{p}{n}\right)^2 \left(\frac{n-p}{n-1}\right)^2 \left(\frac{n-p-1}{n-2}\right)^2 + \dots \\ &\quad + \left(\frac{p}{n}\right)^2 \left(\frac{n-p}{n-1}\right)^2 \left(\frac{n-p-1}{n-2}\right)^2 \dots \left(\frac{1}{p}\right)^2 \\ &= \left(\frac{p}{n}\right)^2 \left[1 + \sum_{i=0}^{n-p-1} \prod_{j=0}^{i} \left(\frac{n-p-j}{n-1-j}\right)^2\right] \\ &= \left(\frac{p}{n}\right)^2 {}_3F_2\left((1, p-n, p-n), (1-n, 1-n), 1\right) \qquad (3.15) \end{aligned}$$

From equation (3.15) it can be seen that the variance of the $ATA(p,0)$ estimator involves the Generalized Hyper-Geometric series:

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

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

$$\bar{k}_{ATA} = \bar{k}_{SES} \Longrightarrow \frac{n-p}{p+1} = \frac{(1-\alpha)}{\alpha} \Longrightarrow \alpha = \frac{(p+1)}{(n+1)},$$

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

$$\bar{k}_{ATA} = \bar{k}_{SES} \Longleftrightarrow \alpha = \frac{(p+1)}{(n+1)},$$

then the variance of ATA is always smaller than the variance of SES $(V_{ATA} < V_{SES})$. When the same smoothing constant is used for both models, since ATA always has a smaller AA, its V value will be greater than that of SES as shown by Yager (2008).

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

ATA method does not need an initial value which from the formula below:

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

$$S_p = X_t \qquad t \leq p$$

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

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

where S_0 refers to the initial value. Weights attached to initial values S_p and S_0 at time t are respectively equal to $\frac{1}{\binom{t}{p}}$ for $ATA(p,0)$ and $(1-\alpha)^t$ for SES. For SES, most practitioners work with α values between 0.01 and 0.3. However, also known as the "initialization problem", when either t or α is small, SES attaches more weight to initial value than even the most current observation. The choice of starting value then becomes particularly important for SES.

It is conceptually wrong to treat the smoothing constant α alone without paying any attention to the sample size t. As a matter of fact, there are two extreme forecasting method. One treats all past observations equally (i.e. average method) and other one attaches weight one only the last observation (i.e. naive method). ES conceptually must be between these two by attaching more weight to most current observation and exponentially decreasing appropriate weights to old observations. The main idea of ES is to assign more weight to recent observations and therefore an ES model should assign the most recent observation at least a weight of $\frac{1}{n}$. This can be achieved only if the search for the smoothing parameter is limited to the interval $[\frac{1}{n}, 1]$.

To compare, the smoothing constant of SES as $\alpha = p/n$ is set to make the smoothing constants of the two models equal. At the same smoothing constant level, $ATA(p, 0)$ assigns less weight to the initial value even for small α. The weights assigned to the initial values by these two models are $(1 - \alpha)^n$ and $(t - p)!p!/n!$ respectively. To visualize, the weights of initial values for both of these models for relatively short time series are plotted in (Figure 3.1) for $p = 2$ and $n = 20$ resulting in $\alpha = p/n = 0.1$. It is obvious that the weight given to initial value for ATA method approaches to zero faster than ES at any iteration.

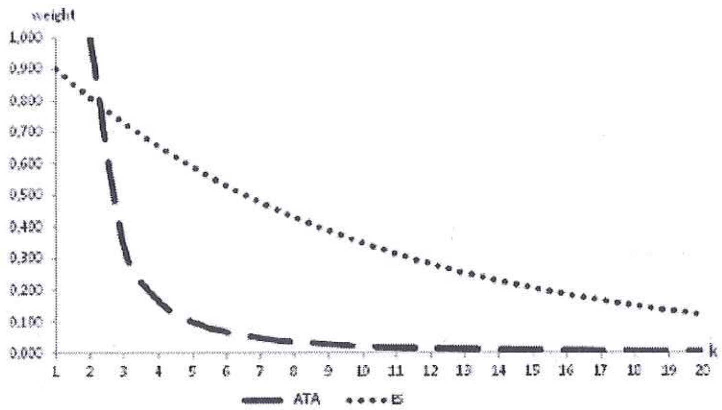

Figure 3.1 Weights of initial value when $p = 2, t = 20; \alpha = \frac{p}{n} = 0.1$

3.1.5 Smoothing Parameter of $ATA(p, 0)$

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

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

$$S_t = \left(\frac{1}{t}\right) X_t + \left(\frac{1}{t}\right) X_{t-1} + \cdots + \left(\frac{1}{t}\right) X_2 + \left(\frac{1}{t}\right) S_1, \qquad (3.16)$$

where $S_1 = X_1$ since $S_p = X_p$ when $t \leq p$ and the h step forecast ($\hat{X}_t(h) = \bar{X}_t$) is equal to the simple average of all past observations, in other words, all observations contribute equally to the forecast. This very important estimator, which intuitively should be the starting point for any method, can never be formed by SES.

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

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

$$S_t = \frac{2}{t}X_t + \frac{2}{t}\frac{(t-2)}{(t-1)}X_{t-1} + \frac{2}{t}\frac{(t-3)}{(t-1)}X_{t-2} + \cdots + \frac{2}{t}\frac{2}{(t-1)}X_3 + \frac{2}{t(t-1)}S_2, \quad (3.17)$$

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 $\frac{2}{t}$ which again can never be achieved by SES since it always assigns exponentially decreasing weights to observations no matter the parameter choice. For $p \geq 3$ 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 more distant past at the same smoothing constant.

Not only is $ATA(p, 0)$ more flexible but also it is more adaptive to the data hand. No matter where along the time line smoothing is being carried out, the weights attached to the observations by SES stay fixed for different sample sizes. The observations are assigned the weights $\alpha, \alpha(1-\alpha), \alpha(1-\alpha)^2, \ldots$ regardless of the sample size at hand. This is not the case for ATA as the weights change with respect to the sample size as the weights are:

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

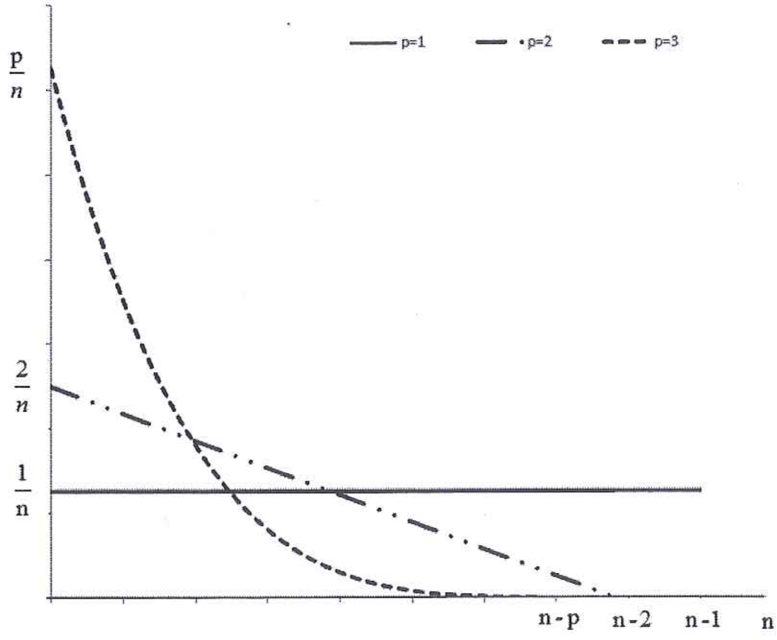

Figure 3.2 Weights attached to the observations by ATA for different p values.

respectively. See Figure Figure 3.2 for illustration of weights assigned to observations for various p levels.

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

$$S_t = \left(\frac{3}{t}\right) X_t + \left(\frac{t-3}{t}\right) S_{t-1} \quad t > 3. \tag{3.18}$$

If we expand the formula iteratively, we get

$$S_4 = \left(\frac{3}{4}\right) X_4 + \left(\frac{1}{4}\right) S_3, \tag{3.19}$$

$$S_5 = \left(\frac{3}{5}\right) X_5 + \left(\frac{2}{5}\right)\left(\frac{3}{4}\right) X_4 + \left(\frac{2}{5}\right)\left(\frac{1}{4}\right) S_3, \tag{3.20}$$

$$S_6 = \left(\frac{3}{6}\right) X_6 + \left(\frac{3}{6}\right)\left(\frac{3}{5}\right) X_5 + \left(\frac{3}{6}\right)\left(\frac{2}{5}\right)\left(\frac{3}{4}\right) X_4 + \left(\frac{3}{6}\right)\left(\frac{2}{5}\right)\left(\frac{1}{4}\right) S_3, \tag{3.21}$$

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

$$\begin{aligned} S_{10} &= 0.300X_{10} + 0.233X_9 + 0.175X_8 + 0.125X_7 + 0.083X_6 \\ &\quad +0.050X_5 + 0.025X_4 + 0.008S_3 \end{aligned} \tag{3.22}$$

where $S_3 = X_3$. The weights attached to observations and initial value by ATA and SES for $t = 10$, $p = 3$ and $\alpha = 0.3$ are given in following table:

Table 3.2 Weights assigned to observations by $ATA(p, 0)$ and SES for an example data set for $p = 3, \alpha = 0.3$ and $n = 10$

Observation	ATA	SES	Difference
X_{10}	0.300	0.300	0.000
X_9	0.233	0.210	0.023
X_8	0.175	0.147	0.028
X_7	0.125	0.103	0.022
X_6	0.083	0.072	0.011
X_5	0.050	0.050	0.000
X_4	0.025	0.035	-0.010
X_3	-	0.025	
X_2	-	0.017	
X_1	-	0.012	
Weight of inital value	0.008	0.028	
Average age ($\hat{k}$)	1.692	1.985	
Variance (V)	0.201	0.176	

As it seen in Table 3.2, both methods give the same weight $\alpha = 0.3$ to the most current observation X_{10} but ATA gives more weight to the recent observations X_9, X_8, X_7, X_6 and same weight to X_5 and less weight to X_4 and zero weight to the oldest three observation X_3, X_2, X_1 and X_3 is taken as the initial value with weight 0.008 while SES gives weight 0.028 for any selected initial value. In fact the average

age of the data used in SES is 1.985 but 1.692 in ATA that means average age of the data for ATA is younger than average age of the SES at the same α value. It can be seen that exponentially decreasing weights are also achieved by ATA and sum of the weights given for observed observations and initial value is one.

To figure out the difference between SES and ATA according to attached weights to the observations at the same smoothing constant $\alpha = p/n$ is given in the following figures. In Figure 3.3, different $p = (2, 3, 5, 10)$ and $n = (20, 40, 60)$ values are selected and then for each p and t combination, $\alpha = p/t$ is attached as a weight to most current observation for both SES and ATA. After the initialization of smoothing constant, the weights attached to other observations are calculated sequentially for both method and their differences are plotted. In Figure 3.4, four different $\alpha = (0.1, 0.2, 0.3, 0.5)$ values are selected for SES and their corresponding p values for ATA are adjusted to obtain the same α for given n. The weights attached to other observations are calculated same as in Figure 3.3 sequentially for both methods and their differences are plotted. Both figures show that ATA attaches more weight to the recent past observations than SES but less weight to old past observations at the same smoothing parameter level. It is obvious that for any fix p value the sharp distinction between the methods will decrease while the number of observations is increased. Locations of the maximum variations are shifting to the most current observation while p is increasing for any fix n value.

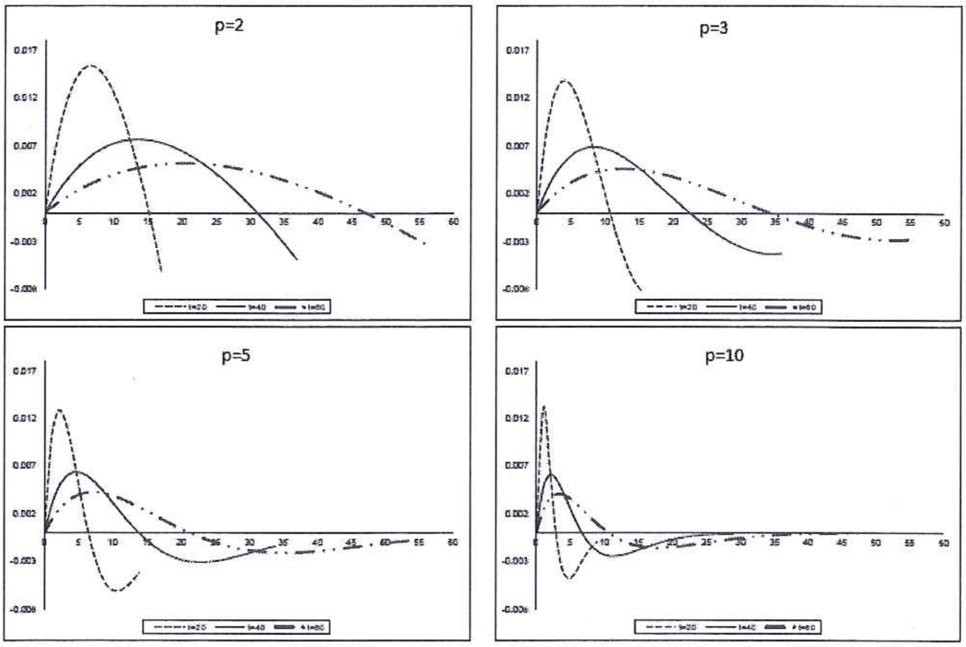

Figure 3.3 Difference of weights ($p = 2, 3, 5, 10$ and $t = 20$,$t = 40$,$t = 60$)

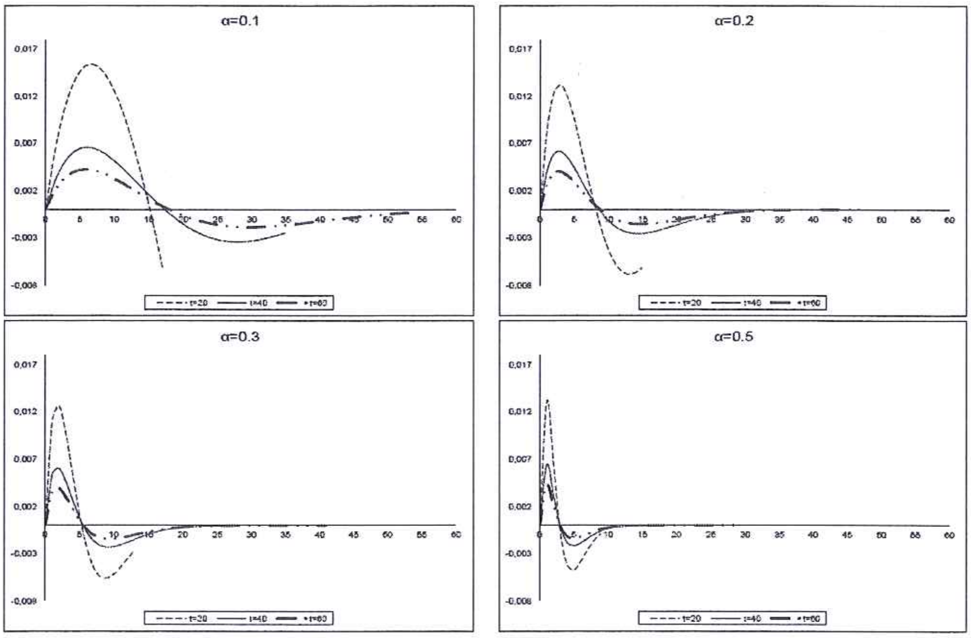

Figure 3.4 Difference of weights for different $\alpha = (0.1, 0.2, 0.3, 0.5)$ values

As a result, the ATA model is more flexible and intuitive compared to SES since it allows for more meaningful weighting schemes when searching for an optimal parameter while potentially reducing the number of iterations needed for reaching that optimal smoothing parameter.

3.1.6 Optimization of $ATA(p, 0)$

Even though both SES and $ATA(p, 0)$ methods require smoothing constants, the optimization process is easier for $ATA(p, 0)$. $ATA(p, 0)$ does not need an initial value unlike SES. When the optimal smoothing parameter is found, the initial value is found simultaneously. With ATA there is no limitation on the number of observations to forecast, only one observation is enough. The smoothing constants and initial values for ES are commonly estimated by minimizing a predetermined error measure like the mean squared error (MSE), the mean absolute error (MAE) or the mean absolute percentage error (MAPE). Several solutions on the issue of finding an initial value are suggested in the literature for SES (Brown (1962); Montgomery & Johnson (1976); Makridakis & Wheelwright (1978); Bowerman & O'Connell (1979)). After discussing various theoretical and empirical arguments for selecting an appropriate smoothing constant, Gardner (1985) concludes that it is best to estimate an optimum α from the data. This is generally done by a search of the parameter space for where $\alpha \in [0, 1]$. In practice, various α values starting from 0.01 with increments 0.01 are tried and the α value that produces the minimum error is chosen. The number of iterations required to find the optimum smoothing constant for ES is then 100 for any data set. The number of iterations for higher order smoothing models (level, trend and season) to find the optimum smoothing constant combinations will be huge (100^3). On the other hand, for $ATA(p, 0)$, the search for the optimal smoothing constant is much easier since the constant depends on the choice of p and $p \in \{1, 2, \ldots, n\}$. Therefore the total number of iterations needed is only n and it will be n^3 for higher order smoothing models. If the predetermined initial values are not optimum, the process must be applied again for ES unlike ATA. As a result, when the data size is less than 100, the total number of iterations for $ATA(p, 0)$ will always be less than those for ES.

The process is easily implemented into computer, it does not require large amounts of historical data and new forecasts are easy to obtain. It is obvious that ATA satisfies all these desirable properties besides being even computationally simpler than SES.

3.1.7 Parameter Space of $ATA(p, 0)$

ATA(p,0) does not violate the concept of exponential smoothing since the smallest weight attached to the last observation is $\frac{1}{n}$ but SES may attain smoothing parameter less than $\frac{1}{n}$.

$$\begin{aligned} \alpha_{ATA} &= \frac{p}{n} \\ \frac{1}{n} &\leqslant \alpha_{ATA} \leqslant 1 \\ \underbrace{0 \leqslant \cdots \leqslant}_{improper\ region} \frac{1}{n} &\leqslant \alpha_{SES} \leqslant 1 \end{aligned} \tag{3.23}$$

If the goal is to estimate the final value for a component and it is known that it is changing in time, it makes sense to give greater emphasis to the most recent history versus the more distant past, due to the fact that the component is changing and therefore the recent history should more accurately reflect current conditions. For this reason it is not surprising that ATA is more accurate than traditional ES methods. Also philosophically, ATA does never violate the concept of exponential weighting scheme, from the point of view that the recent data is more representative of the future therefore should be assigned more weight. This is guaranteed with ATA since the smallest weight attached to the most recent observation by ATA is $\frac{1}{n}$ but this is not the case for SES.

3.2 Comparison of $ATA(p,0)$ and SES

While the functional forms of ATA models are generally very similar to those of exponential smoothing models, there are distinctive features of ATA that separate it from ES. $ATA(p,0)$ can be thought of as an approach that lies in between moving averages (MA) and simple exponential smoothing (SES). $ATA(p,0)$ attaches weights to only the most recent $(n-p)$ observations and zero weights to the other p observations like MA and the weights decrease exponentially like SES for some p ($p \leq 3$). The weighting scheme $ATA(p,0)$, however, is more flexible and intuitive than SES.

In addition, the other important difference lies in the weights assigned to observations by ATA and SES. $ATA(p,0)$ can be parameterized so that all past observations receive equal weights while this is not possible for any SES model. Also when the $ATA(p,0)$ and SES that assign equal weights to the most recent observation are compared, it can be seen that $ATA(p,0)$ tends to assign more weight to the other recent observations while assigning less weight to the distant past compared to SES. While all ES models require initialization and the initial values affect the quality of forecasts especially for small values of n and α, ATA does not require initialization and the optimization of the other parameters are simpler and faster since the parameter values are restricted to integers.

Forecasting models can also be compared on some well known metrics like the average age of the method and the variance of the forecast. As summarized in sections 3.1.2 and 3.1.3, at the same smoothing constant level, the average age of $ATA(p,0)$ is smaller than SES,

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

therefore $ATA(p,0)$ should be preferred by researchers since it utilizes fresher data. In order for the two models to have equal average ages the smoothing constant of SES should be given the value $\alpha = \frac{p+1}{n+1}$. When smoothing constants for the two models are chosen in this fashion to make the average ages equal, $ATA(p,0)$ is still preferable since then $V_{ATA} < V_{SES} = \dfrac{\alpha}{2-\alpha}$.

To present the discussions about the features of ATA in a more organized way, the weights that $ATA(p, 0)$, SES and MA assign to the observations when trying to obtain one step forecasts are given in Table 3.3 and Table 3.4 for sample sizes 12 and 30 respectively. The tables also contain the average ages ($\hat{k}$), variance (V) and the initial values and weights assigned to observations.

Table 3.3 Weights assigned to observations by ATA, ES and MA for $n = 12$ to obtain $\hat{X}_{13}$

Observation	$p=1$ and $\alpha=1/12$			$p=2$ and $\alpha=2/12$			$p=3$ and $\alpha=3/12$			$p=6$ and $\alpha=6/12$		
	ATA	ES	MA	ATA	ES	MA	ATA	ES	MA	ATA	ES	MA
X_{12}	0.083	0.083	0.083	0.167	0.167	0.167	0.250	0.250	0.250	0.500	0.500	0.500
X_{11}	0.083	0.076	0.083	0.152	0.139	0.167	0.205	0.188	0.250	0.273	0.250	0.500
X_{10}	0.083	0.070	0.083	0.136	0.116	0.167	0.164	0.141	0.250	0.136	0.125	-
X_9	0.083	0.064	0.083	0.121	0.096	0.167	0.127	0.105	0.250	0.061	0.063	-
X_8	0.083	0.059	0.083	0.106	0.080	0.167	0.095	0.079	-	0.023	0.031	-
X_7	0.083	0.054	0.083	0.091	0.067	0.167	0.068	0.059	-	0.006	0.016	-
X_6	0.083	0.049	0.083	0.076	0.056	-	0.045	0.044	-	-	0.008	-
X_5	0.083	0.045	0.083	0.061	0.047	-	0.027	0.033	-	-	0.004	-
X_4	0.083	0.042	0.083	0.045	0.039	-	0.014	0.025	-	-	0.002	-
X_3	0.083	0.038	0.083	0.030	0.032	-	0.000	0.019	-	-	0.001	-
X_2	0.083	0.035	0.083	-	0.027	-	-	0.014	-	-	0.000	-
X_1	-	0.032	0.083	-	0.022	-	-	0.011	-	-	0.000	-
Initial value	X_1	?	-	X_2	?	-	X_3	?	-	X_4	?	-
weight of initial	0.083	0.352	-	0.015	0.112	-	0.005	0.032	-	0.001	0.000	-
AA	5.500	6.776	5.500	3.348	4.327	2.500	2.259	2.873	1.500	0.863	1.000	0.500
V	0.083	0.162	0.083	0.116	0.102	0.167	0.164	0.144	0.250	0.347	0.333	0.500

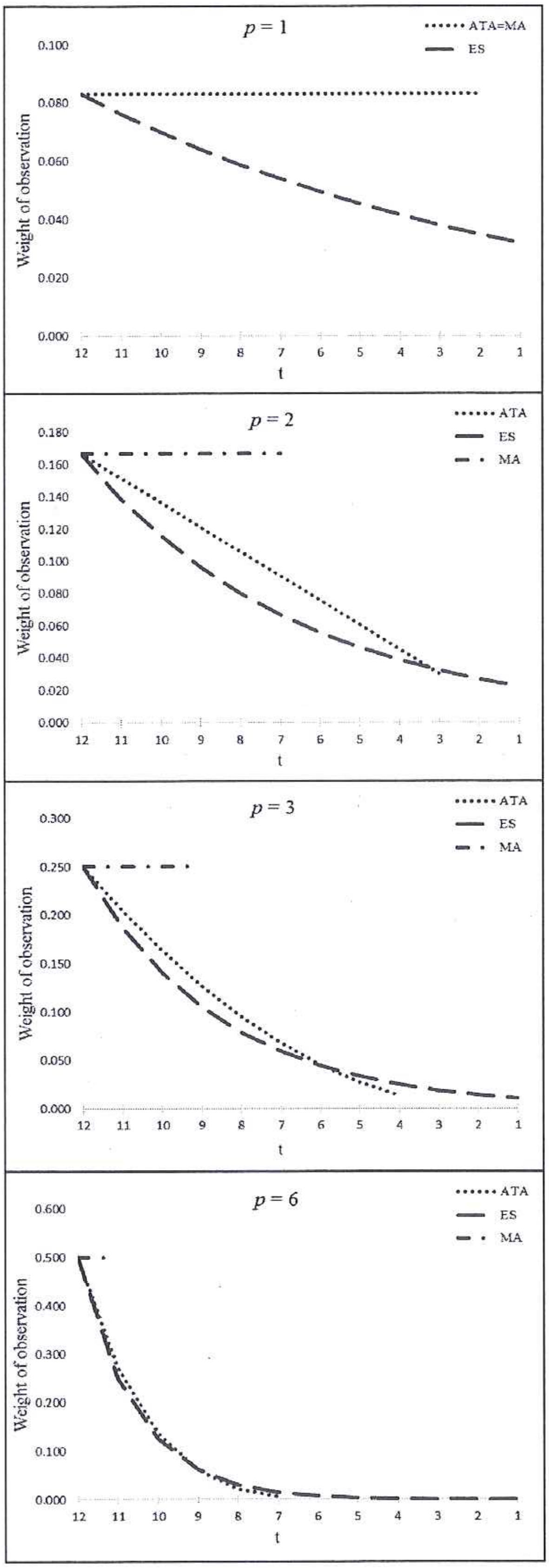

Figure 3.5 Weights assigned to observations by $ATA(p, 0)$, SES and MA for $n = 12$ and various p values.

Table 3.4 Weights assigned to observations by ATA, ES and MA for $n = 30$ to obtain $\hat{X}_{31}$

Observation	$p=1$ and $\alpha=1/30$			$p=2$ and $\alpha=2/30$			$p=3$ and $\alpha=3/30$			$p=15$ and $\alpha=6/30$		
	ATA	ES	MA	ATA	ES	MA	ATA	ES	MA	ATA	ES	MA
X_{30}	0.033	0.033	0.033	0.067	0.067	0.067	0.100	0.100	0.100	0.500	0.500	0.500
X_{29}	0.033	0.032	0.033	0.064	0.062	0.067	0.093	0.090	0.100	0.259	0.250	0.500
X_{28}	0.033	0.031	0.033	0.062	0.058	0.067	0.086	0.081	0.100	0.129	0.125	-
X_{27}	0.033	0.030	0.033	0.060	0.054	0.067	0.080	0.073	0.100	0.062	0.063	-
X_{26}	0.033	0.029	0.033	0.057	0.051	0.067	0.074	0.066	0.100	0.029	0.031	-
X_{25}	0.033	0.028	0.033	0.055	0.047	0.067	0.068	0.059	0.100	0.013	0.016	-
X_{24}	0.033	0.027	0.033	0.053	0.044	0.067	0.062	0.053	0.100	0.005	0.008	-
X_{23}	0.033	0.026	0.033	0.051	0.041	0.067	0.057	0.048	0.100	0.002	0.004	-
X_{22}	0.033	0.025	0.033	0.048	0.038	0.067	0.052	0.043	0.100	0.001	0.002	-
X_{21}	0.033	0.025	0.033	0.046	0.036	0.067	0.047	0.039	0.100	0.000	0.001	-
X_{20}	0.033	0.024	0.033	0.044	0.033	0.067	0.042	0.035	-	0.000	0.000	-
X_{19}	0.033	0.023	0.033	0.041	0.031	0.067	0.038	0.031	-	0.000	0.000	-
X_{18}	0.033	0.022	0.033	0.039	0.029	0.067	0.033	0.028	-	0.000	0.000	-
X_{17}	0.033	0.021	0.033	0.037	0.027	0.067	0.030	0.025	-	0.000	0.000	-
X_{16}	0.033	0.021	0.033	0.034	0.025	0.067	0.026	0.023	-	0.000	0.000	-
X_{15}	0.033	0.020	0.033	0.032	0.024	-	0.022	0.021	-	-	0.000	-
X_{14}	0.033	0.019	0.033	0.030	0.022	-	0.019	0.019	-	-	0.000	-
$1X_{13}$	0.033	0.019	0.033	0.028	0.021	-	0.016	0.017	-	-	0.000	-
X_{12}	0.033	0.018	0.033	0.025	0.019	-	0.014	0.015	-	-	0.000	-
X_{11}	0.033	0.018	0.033	0.023	0.018	-	0.011	0.014	-	-	0.000	-
X_{10}	0.033	0.017	0.033	0.021	0.017	-	0.009	0.012	-	-	0.000	-
X_9	0.033	0.016	0.033	0.018	0.016	-	0.007	0.011	-	-	0.000	-
X_8	0.033	0.016	0.033	0.016	0.015	-	0.005	0.010	-	-	0.000	-
X_7	0.033	0.015	0.033	0.014	0.014	-	0.004	0.009	-	-	0.000	-
X_6	0.033	0.015	0.033	0.011	0.013	-	0.002	0.008	-	-	0.000	-
X_5	0.033	0.014	0.033	0.009	0.012	-	0.001	0.007	-	-	0.000	-
X_4	0.033	0.014	0.033	0.007	0.011	-	0.001	0.006	-	-	0.000	-
X_3	0.033	0.013	0.033	0.005	0.010	-	-	0.006	-	-	0.000	-
X_2	0.033	0.013	0.033	-	0.010	-	-	0.005	-	-	0.000	-
X_1	-	0.012	0.033	-	0.009	-	-	0.005	-	-	0.000	-
Initial Value	X_1	?	-	X_2	?	-	X_3	?	-	X_15	?	-
Weight of initial	0.033	0.362	-	0.002	0.126	-	0.000	0.042	-	-	0.000	-
AA	14.500	18.150	14.500	9.333	12.107	7.000	6.750	8.491	4.500	0.937	1.000	0.500
V	0.032	0.015	0.033	0.045	0.034	0.067	0.062	0.053	0.100	0.339	0.333	0.500

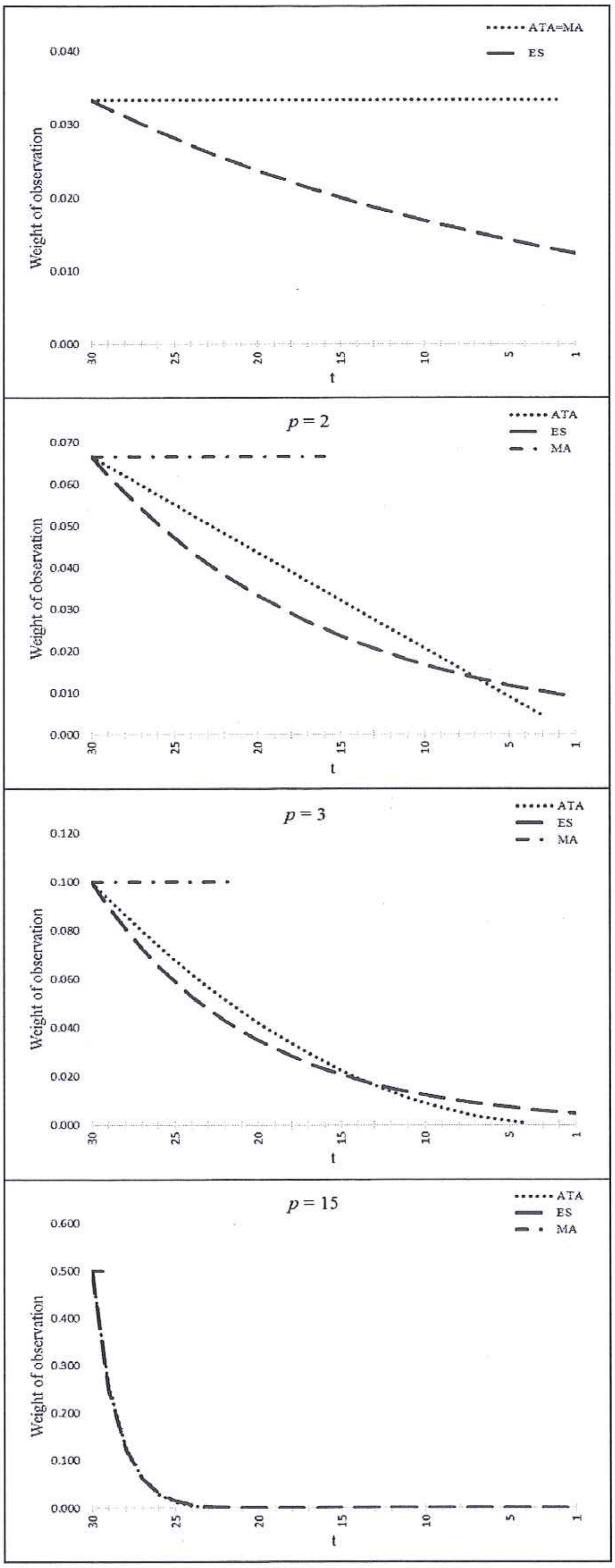

Figure 3.6 Weights assigned to observations by $ATA(p,0)$, SES and MA for $n = 12$ and various p values.

From the tables it can be seen that for $p = 1$ the MA and $ATA(p, 0)$ models assign the same weights to observations therefore they have equal average ages, however, SES has larger average age with a slightly smaller variance. For $p = 2$ the weights of the oldest two observations are zero for $ATA(p, 0)$ and the average age and variance of the model is now between those of MA and ES with ES having the largest. For $p = 3$, now the weights of the oldest three observations are zero for $ATA(p, 0)$ and the average age and variance of the model are again between those of MA and ES with ES having the largest. It is worth drawing attention to the differences between the weights attached to the initial values by the ES and $ATA(p, 0)$. For all smoothing levels ES assigns a relatively much larger weight to the initial value compared to $ATA(p, 0)$. The fact that ES is not adaptive to the data can be seen by looking at the $p = 6$ columns of Table 3.3 and the $p = 15$ columns of Table 3.3. Since $\alpha = 0.5$ for both of these cases, even though the weights should be distributed among 12 observations in the first case and 30 observations in the second case, ES assigns 0.5 to the most recent, 0.273 to the second most recent etc., exactly the same weights for both data sets. even though we have a much larger data set when $n = 30$ compared to $n = 12$, ES keeps assigning the same weights to the most recent, the second most recent observations etc. $ATA(p, 0)$ on the other hand takes into account the amount of data that can be utilized and is able to distribute the weights in a fashion that favors the recent observations. When the tables are studied closely, it can be seen that $ATA(p, 0)$ always assigns more weight to recent observations and less weight to older observations.

3.2.1 M3 Competition Results

To compare $ATA(p, 0)$ and SES on their forecasting accuracies, both methods are applied to the M3-competition data (Makridakis & Hibon (2000)) since this collection is the most recent and comprehensive time-series data collection available with verified results. This collection consists of 3003 data sets from various fields. Data sets are of various lengths, with different kinds of trend and seasonality components and each data set consists of in sample and out sample data points. When comparing the methods, the optimum smoothing parameters are obtained by minimizing an in-sample error measure and then the forecasts up to 18 steps ahead (the number of steps as specified in the M3 competition) are computed to obtain the average out sample errors for both models. The data sets are deseasonalized by the classical decomposition method of the ratio-to-moving averages, if necessary and reseasonalized forecasts are produced for as many steps ahead as required.

First, to stay consistent with the rest of the literature, the symmetric mean absolute percentage errors (sMAPE) were used. For all data sets, the required numbers of forecasts (for the pre-determined forecasting horizons) were computed and out sample sMAPEs were averaged across all 3003 series for each forecasting horizon. The results are given in Table 3.5.

Table 3.5 Average symmetric MAPE across different forecast horizons: all 3003 series

	Forecasting horizons										Averages					
Method	1	2	3	4	5	6	8	12	15	18	1-4	1-6	1-8	1-12	1-15	1-18
SES	9.5	10.6	12.7	14.1	14.3	14.9	13.3	14.5	18.3	19.4	11.73	12.68	12.82	13.12	13.66	14.31
$ATA(p,0)$	8.9	10.0	12.1	13.7	13.9	14.7	12.8	13.9	17.3	18.9	11.16	12.21	12.34	12.64	13.13	13.77

When the methods are compared based on sMAPE as in Table 3.5, it can be seen that $ATA(p,0)$ produces smaller average errors for all individual forecasting horizons. The errors are averaged for short and long term horizons on the right side of the table so that the differences between the errors can be more clearly seen. Overall $ATA(p,0)$'s average sMAPE is 13.77 compared to 14.31 for SES which is significantly larger.

Second comparison can be made based on the mean absolute scaled errors (MASE) which were summurized for both models in Table 3.6. When the comparison is based on this metric, $ATA(p,0)$ still performs better than SES on each forecasting horizon and on average for both short and long term forecasting horizons.

Table 3.6 Average MASE across different forecast horizons: all 3003 series

	Forecasting horizons										Averages					
Method	1	2	3	4	5	6	8	12	15	18	1-4	1-6	1-8	1-12	1-15	1-18
SES	0.78	1.03	1.36	1.63	1.82	2.05	1.49	0.97	1.25	1.38	1.20	1.45	1.44	1.35	1.32	1.33
$ATA(p,0)$	0.75	1.00	1.33	1.61	1.80	2.04	1.47	0.97	1.20	1.37	1.17	1.42	1.42	1.33	1.30	1.30

The results obtained from the forecasting analysis of M3 competition data set are summarized in Table 3.7. That is, SES and $ATA(p,0)$ have been applied to 3003 series and the optimal smoothing parameter values have been obtained. Simple exponential smoothing methods have been performed in two different ways. The initial values have been selected either as the first observation optimized in forecast package of R programming. Then, it has been observed that when the initial value is X_1 for SES, the value of the average optimal smoothing parameter is 0.7419. The smoothing level α equals 0.7257, when the initial value is optimal. In addition, for $ATA(p,0)$ the average smoothing level is 0.6307 as shown in Table 3.7. When the alpha values of the two models have been compared, one can see that ATA has smaller variance. As a result, ATA yields more efficient estimates than SES. To sum up, when the simple versions of the two models are compared theoretically, ATA is superior since it is simpler and has more desirable properties. Also, when compared emprically ATAoutperforms SES on the M3-Competition data sets on all forecasting horizons for all the popular error metrics.

Table 3.7 Average optimal smoothing parameter for SES and $ATA(p,0)$: all 3003 series

	Average smoothing parameter ($\bar{\alpha}_{opt}$)		
Method	SES	SES	ATA(p,0)
Initial value	X_1	Optimal	X_p
All 3003 series	0.7419	0.7257	0.6307
Yearly series	0.9177	0.9071	0.8266
Quarterly series	0.8713	0.8623	0.7582
Monthly series	0.5666	0.5421	0.4388
Other series	0.9628	0.9624	0.9226

CHAPTER FOUR
THE TRENDED ATA METHODS

ES models consist of a family of models which assume that the time series has up to three underlying data components: level, trend and seasonality. In ES the goal is to obtain estimates for the level, trend and seasonal pattern and then to use these final values to forecast the future. Each model contains one of the five types of trend (none, additive, additive damped, multiplicative and multiplicative damped) and one of the three types of seasonality (none, additive and multiplicative). When different combinations of trend and seasonality are considered, 15 different ES models can be formed. The best known of these are SES (no trend, no seasonality), Holt's linear model (additive trend, no seasonality) and Holt-Winters' additive model (additive trend, additive seasonality). The ETS state space models underlying ES have been proposed by Hyndman et al. (2008) which consider 30 different forecasting models according to error, trend and seasonality components. In this thesis, we will not consider seasonal models since Holt's additive method's performance for deseasonalized data is better than Holt Winters' seasonal method. Therefore, we will deal with trended (additive, multiplicative, additive damped and multiplicative damped) models using ATA method with deseasonalized data. Most of the competitors in the M3-competition cope with the seasonality problem by this fashion.

The forecasting methods that performed the best in Makridakis & Hibon (2000) mostly use data after the data is deseasonalized by the classical decomposition method. It is especially worth noting Holt's linear trend method on the deseasonalized data performs better than Holt Winters' when the accuracies are compared on empirical data. This itself proves that it is sufficient to use a decomposition method on seasonal data instead of employing a complex method to deal wit seasonality. Taking into account all of these, in this thesis the focus is on modeling the trend component and handling seasonality patterns by utilizing classical decomposition. Therefore, ATA method will be expanded to higher order ES methods for additive, multiplicative and damped trend components. In the following sections the different trended ATA models will be introduced and demonstrated by using the series N96 from the M3 data set which is a yearly series. The data consists of 20 data points where 14 of these are in sample (training) and the remainin 6 are considered as out sample (test) observations. The proposed model's accuracy will be compared to its counter ES model using M3 competition data set in the last part of this chapter.

4.1 ATA(p,q) with Additive Trend

For the trended series X_t, $t = 1, \ldots, n$, the model which is denoted by $ATA(p,q)$ can be used $ATA(p,q)$ corresponds to a Holt's linear model with some modifications on level and trend parameters. This method can be given by a forecast equation and two smoothing equations (one for the level (S_t) and one for the trend (T_t):

$$S_t = \left(\frac{p}{t}\right) X_t + \left(\frac{t-p}{t}\right) (S_{t-1} + T_{t-1}) \tag{4.1}$$

$$T_t = \left(\frac{q}{t}\right) (S_t - S_{t-1}) + \left(\frac{t-q}{t}\right) T_{t-1} \tag{4.2}$$

$$\hat{X}_t(h) = S_t + hT_t, \tag{4.3}$$

for $p \in \{1, \ldots, n\}$, $q \in \{0, 1, \ldots, p\}$ 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 $T_1 = 0$. S_t denotes an estimate of the level of the series at time t, and T_t denotes an estimate of the trend (slope) of the series at time t. It is worth pointing out that when $q = 0$ $ATA(p,q)$ reduces to a simple model that has similar form to simple ES, i.e. for $n > p$:

$$S(t) = \left(\frac{p}{t}\right) X_t + \left(\frac{t-p}{t}\right) S_{t-1}, \tag{4.4}$$

and $S_(t) = X_t$ for $n \leq p$.

As with simple exponential smoothing, the level equation in 4.1 shows that S_t is a weighted average of the observation at time t and the with in-sample one-step-ahead forecast for time $t-1$, here given by $S_{t-1} + T_{t-1}$. The trend equation shows that T_t is a

Table 4.1 Parameter space for $ATA(p, q)$

Trend smoothing parameter(q)	Level smoothing parameter (p) 1	2	3		...	p	...	n
0	**(1,0)**	**(2,0)**	**(3,0)**	**(4,0)**	...	**(p,0)**	...	**(n,0)**
1	**(1,1)**	**(2,1)**	**(3,1)**	**(4,1)**	...	**(p,1)**	...	**(n,1)**
2	(1,2)	**(2,2)**	**(3,2)**	**(4,2)**	...	**(p,2)**	...	**(n,2)**
3	(1,3)	(2,3)	**(3,3)**	**(4,3)**	...	**(p,3)**	...	**(n,3)**
4	(1,4)	(2,4)	(3,4)	**(4,4)**	...	**(p,4)**	...	**(n,4)**
⋮					⋮	...	⋮	⋮
						(p*,q*)		
⋮					⋮	...	⋮	⋮
n	(1,n)	(2,n)	(3,n)	(4,n)	...	**(p,n)**	...	**(n,n)**

weighted average of the estimated trend at time t which is simply $S_t - S_{t-1}$ and T_{t-1}, the previous estimate of the trend. $\hat{X}_t(h)$ denotes the h-step-ahead forecast which is equal to the last estimated level plus h times the last estimated trend value.

The parameter space of ATA method, as shown in Table 4.1 allows for all possible parameter value combinations to be considered as candidates. The value of p represents level parameter and also the value of q represents trend parameter where $p \in \{1, \ldots, n\}$, $q \in \{0, 1, \ldots, p\}$ and $n > p \geq q$. However, $(p*, q*)$ demonstrated in Table 4.1 refers to optimal smoothing parameters for the procedure of ATA method. From the Table 4.1, an algorithm can be obtained to estimate the smoothing parameters. The algorithm can be summarized as follows:

Step 1 If needed the time series is deseasonalized via the classical decomposition method, assuming a multiplicative relationship holds for seasonal component.

Step 2 Estimate p and q by minimizing the in-sample symmetric mean absolute one-step-ahead forecast errors, i.e. $sMAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|X_t - \hat{X}_t|}{|X_t + \hat{X}_t|} * 200$ where X_t is the actual value and $\hat{X}_t$ is the one-step-ahead forecasted value assuming $t > p \leq q$.

Step 3 The forecasts are reseasonalized, if they were deseasonalized in Step 1.

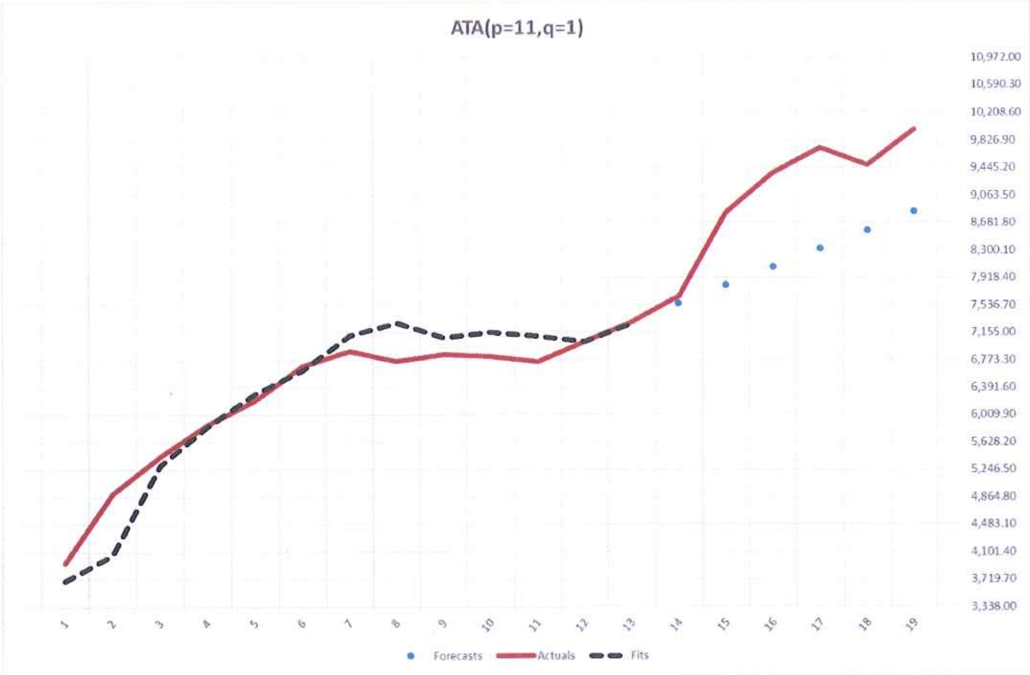

Figure 4.1 The fitted and forecast values of $ATA(p, q)$ with optimal smoothing parameters

Figure 4.1 presents the observed data, fitted and forecast values at the optimal p and q smoothing levels. According to the graph, the red line refers to original series of N96 competition data set and the black line represents the one step forecast values obtained from $ATA(p, q)$. In addition, the six step ahead forecasts are calculated and market as blue points. The optimal smoothing parameters were found to be 1 for the trend and 11 for the level parameters with the constraint $p \leq q$. The original data set has a slightly increasing trend and the value of $q = 1$ is suitable to represent this.

In Table 4.2 demonstrates the results that are obtained from applying $ATA(p, q)$ on the N96 competition data set. The smoothing parameters are estimated by minimizing the in sample sMAPE one step forecast ahead.

Table 4.2 An application of $ATA(p, q)$ linear trend method on the N96 time series data set from the M3-competition

		Data	The smoothed and forecast values		
Year	t	X	Level (S_t)	Trend (T_t)	Forecast ($\hat{X}_t$)
1975	1	3709.24	3709.24	0.00	
1976	2	3947.02	3947.02	118.89	3709.24
1977	3	4907.50	4907.50	399.42	4065.91
1978	4	5425.42	5425.42	429.05	5306.92
1979	5	5866.84	5866.84	431.52	5854.47
1980	6	6211.48	6211.48	417.04	6298.36
1981	7	6689.54	6689.54	425.76	6628.52
1982	8	6896.62	6896.62	398.42	7115.30
1983	9	6749.48	6749.48	337.80	7295.04
1984	10	6847.42	6847.42	313.82	7087.28
1985	11	6823.48	6823.48	283.11	7161.24
1986	12	6740.24	6770.77	255.13	7106.59
1987	13	7023.82	7024.14	254.99	7025.90
1988	14	7303.28	7298.11	256.35	7279.13
	h				$\hat{X}_t(h)$
1989	1	7661.38			7554.45
1990	2	8816.56			7810.80
1991	3	9366.04			8067.15
1992	4	9715.20			8323.50
1993	5	9485.74			8579.84
1994	6	9974.00			8836.19
Accuracy measures					
		in-sample		out-sample	
		MSE	110935	MSE	1127035
		MAE	238	MAE	974
		MAPE	4.31	MAPE	11.76
		sMAPE	4.19	sMAPE	10.99

4.2 ATA(p,q) with Multiplicative Trend

For the series X_t, $t = 1, \ldots, n$, the model which we will denote by $ATA_{mult}(p, q)$ allows for smoothing and forecasting of data with an exponential trend. This method can be given by the following a forecast equation and two smoothing equations (one for the level (S_t) and one for the trend (T_t):

$$S_t = \left(\frac{p}{t}\right) X_t + \left(\frac{t-p}{t}\right) (S_{t-1} * T_{t-1}), \tag{4.5}$$

$$T_t = \left(\frac{q}{t}\right) \left(\frac{S_t}{S_{t-1}}\right) + \left(\frac{t-q}{t}\right) T_{t-1}, \tag{4.6}$$

$$\hat{X}_t(h) = S_t * T_t^h, \tag{4.7}$$

for $p \in \{1, \ldots, n\}$, $q \in \{0, 1, \ldots, p\}$ 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 $T_1 = 1$.

The trend equation shows that T_t represents an estimated growth rate (in relative terms rather than absolute) which is multiplied rather than added to the estimated level. The trend component now has an exponential effect on the forecast value rather than linear such that $\hat{X}_t(h)$ is equal to the final estimate of the level times the final estimate of the trend to the power of h.

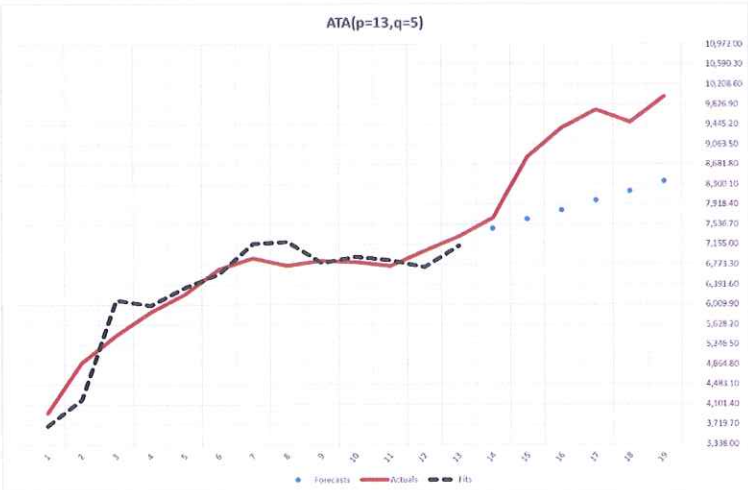

Figure 4.2 The fitted and forecast values of $ATA_{mult}(p, q)$ with optimal smoothing parameters

Table 4.3 An application of $ATA_{mult}(p,q)$ exponential trend method on the N96 time series data set from the M3-competition.

		Data	The smoothed and forecast values		
Year	t	X	Level (S_t)	Trend (T_t)	Forecast ($\hat{X}_t$)
Year	t	X	Level	Trend	Forecast
1975	1	3709.24	3709.24	1.00	
1976	2	3947.02	3947.02	1.06	3709.24
1977	3	4907.50	4907.50	1.24	4200.04
1978	4	5425.42	5425.42	1.11	6101.71
1979	5	5866.84	5866.84	1.08	5998.00
1980	6	6211.48	6211.48	1.06	6344.17
1981	7	6689.54	6689.54	1.07	6599.78
1982	8	6896.62	6896.62	1.05	7176.77
1983	9	6749.48	6749.48	1.01	7218.42
1984	10	6847.42	6847.42	1.01	6809.45
1985	11	6823.48	6823.48	1.00	6927.52
1986	12	6740.24	6740.24	1.00	6856.17
1987	13	7023.82	7023.82	1.01	6724.82
1988	14	7303.28	7290.73	1.02	7127.59
	h				$\hat{X}_t(h)$
1989	1	7661.38			7458.92
1990	2	8816.56			7630.99
1991	3	9366.04			7807.03
1992	4	9715.20			7987.14
1993	5	9485.74			8171.39
1994	6	9974.00			8359.90
Accuracy measures					
		in-sample		out-sample	
		MSE	115510	MSE	1866014
		MAE	266	MAE	1267
		MAPE	4.70	MAPE	15.87
		sMAPE	4.65	sMAPE	14.54

The results that are obtained from applying $ATA_{mult}(p,q)$ on the N96 competition data set are given in Table 4.3. Otherwise, the smoothing parameters are estimated by minimizing the in sample one step ahead sMAPE. For demonstration purposes the optimal smoothing parameters are set to $p = 13$ and $q = 5$. Figure 4.2 presents the results obtained from the application of $ATA_{mult}(p, q)$ as it was done in previous section for additive model. Nevertheless, it can be highlighted that the trend parameter for the multiplicative model has a larger value compared to the trend parameter for the additive model. The difference between the additive and multiplicative trend to calculate in the column T_t. In the $ATA(p, q)$ method with additive trend, T_t is added to the corresponding level term (S_{t-1}) in the calculations; in the $ATA(p, q)$ method with multiplicative trend, T_t is multiplied with the corresponding level term (S_{t-1}) in the calculations.

4.3 Level Fixed Trended Methods

We have seen that ATA method can be adapted to all ES models. There are numerous variations of standard exponential smoothing models that have been proposed to increase their forecasting performance by defining some rules and putting some constraints on the smoothing parameters. Although ES methods incorporate five different types of trend and three different type of seasonal components what determines the forecasting performance is the level of the series. Without knowing the level of the series, finding the right trend and seasonal component does not mean anything. Therefore, fixing the trend parameter to zero, we will first find the optimum level and then using this optimum level we will search for the appropriate trend value. Since, if a data set is not trended, the value of the trend parameter will be zero, this approach gives us the opportunity of making a selection between a simple and a trended model. For the fixed level parameter, the proposed models are described as additive, multiplicative and damped trend in the following sections.

4.3.1 Additive Forms of ATA(p,q) with Fixed Level*

In this section, a version of ATAwhich we denoted by $ATA(p*, q)$ will be given as an alternative Holt's linear trend method. The goal is again to forecast future values of a time series from its own current and past values. That is, given a series of equally spaced observations, X_t, $t = 1, 2, \ldots, n$ on some quantity, forecasts for $h = 1, 2, \ldots$ should be obtained. The method estimates the local growth, T_t, by smoothing successive differences, $(S_t - S_{t-1})$ of the local level, S_t. The forecast function is the sum of level and projected growth. Now, Holt's linear trend method, which was defined in equations (2.17)-(2.19), will be modified as below:

$$S_t = \left(\frac{p}{t}\right) X_t + \left(\frac{t-p}{t}\right)(S_{t-1} + T_{t-1}), \tag{4.8}$$

$$T_t = \left(\frac{q}{t}\right)(S_t - S_{t-1}) + \left(\frac{t-q}{t}\right) T_{t-1},\ t > p \geq q, \tag{4.9}$$

$$\hat{X}_t(h) = S_t + hT_t, \tag{4.10}$$

with constraint $S_t = X_t$ for $t \leq p$, $T_t = X_t - X_{t-1}$ for $t \leq q$ and $T_1 = 0$ where $p \in \{0, 1, \ldots, n\}$, $q \in \{0, 1, \ldots, n\}$ and $p \geq q$. Note that there are two smoothing parameters (p and q) to estimate but no starting values are needed for level and trend. The parameters can be estimated by minimizing the one step ahead MSE, MAE, MAPE or some other criterion for measuring the in-sample forecast error. Also notice that for $q = 0$ the $ATA(p*, q)$ model defined in equations (4.8)-(4.10) reduces to an ATA model with additive trend with no trend component which is an alternative to simple exponential smoothing. The steps involved to obtain point forecasts for $ATA(p*, q)$ are summarized below:

Step 1 Deseasonalize the data by the classical decomposition method of the ratio-to-moving averages, if necessary.

Step 2 Obtain the optimal value for the parameter $p = p^*$ by minimizing the in-sample mean absolute one-step-ahead forecast errors, i.e. $sMAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{|X_t - \hat{X}_t|}{|X_t + \hat{X}_t|} * 200$ where X_t is the actual value and $\hat{X}_t$ is the one-step-ahead forecasted value assuming $q = 0$.

$p^* = argminp(sMAPE|q = 0), \qquad p = 1, \ldots, n$

Step 3 Using the optimal p^* parameter found in Step 2, find an optimal value for the parameter q ($q \leq p^*$) by minimizing the in-sample one-step-ahead sMAPE again. (p^*, q).

Step 4 Produce reseasonalized forecasts using the proposed method with optimized parameters (p^*, q) for as many steps ahead as required.

Table 4.4 Parameter space of $ATA(p^*, q)$

	Level parameter (p)						
Trend parameter(q)	1	2	3	⋮	p	⋮	n
0	**(1,0)**	**(2,0)**	**(3,0)**	**...**	**(p*,0)**	...	(n,0)
1	(1,1)	(2,1)	(3,1)	...	**(p*,1)**	...	(n,1)
2	(1,2)	(2,2)	(3,2)	...	**(p*,2)**	...	(n,2)
3	(1,3)	(2,3)	(3,3)	...	**(p*,3)**	...	(n,3)
⋮					**⋮**		
q					**(p*,q)**		
⋮					⋮		
n	(1,n)	(2,n)	(3,n)	...	(p^*, n)	...	(n,n)

The implementation of the algorithm and the parameter space of $ATA(p^*, q)$ is given in Table 4.4.

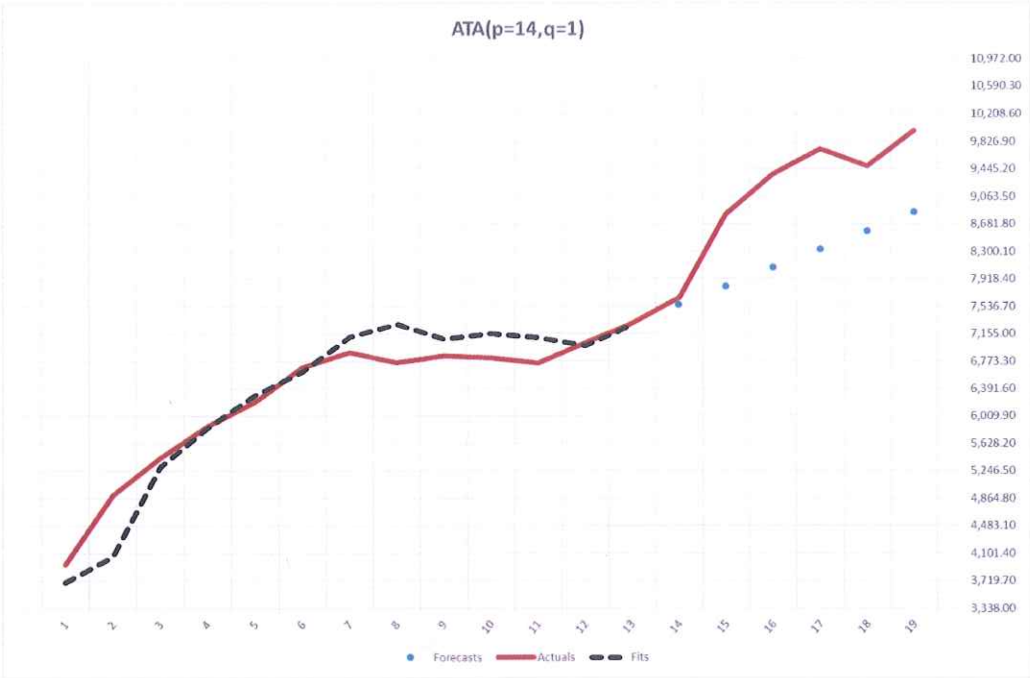

Figure 4.3 The fitted and forecast values of $ATA(p^*, q)$ with optimal smoothing parameters

The smoothed values and optimum parameters obtained by modelling the N96 data set by $ATA(p^*, q)$ are given in Table 4.5. The smoothing parameters are estimated by minimizing the sMAPE for the within-sample one-step forecast errors. Also, the level and trend parameters are found to be $p = 14$ and $q = 1$. When $ATA(p^*, q)$ and $ATA(p, q)$ are compared, the difference of the level parameter can be clearly observed. It is interpreted that a significant difference will occur due to the dramatical trend effects in the long term. The advantage of this model is that it removes the misleading effect that the trend component may have in determining the level of the model. Because, a level that is not correctly determined will lead to errors in determining the slope.

Table 4.5 contains the results that are obtained from applying on the $ATA(p^*, q)$ analysis for N96 competition data set. The table shows us the smoothed values for trend and level components and also the fitted and the forecast values for training and test sets respectively.

Table 4.5 An application of $ATA(p^*, q)$ exponential trend method on the N96 time series data set from M3-competition.

		Data	The smoothed and forecast values		
Year	t	X	Level (S_t)	Trend (T_t)	Forecast ($\hat{X}_t$)
Year	t	X	Level	Trend	Forecast
1975	1	3709.24	3709.24	0.00	
1976	2	3947.02	3947.02	118.89	3709.24
1977	3	4907.50	4907.50	399.42	4065.91
1978	4	5425.42	5425.42	429.05	5306.92
1979	5	5866.84	5866.84	431.52	5854.47
1980	6	6211.48	6211.48	417.04	6298.36
1981	7	6689.54	6689.54	425.76	6628.52
1982	8	6896.62	6896.62	398.42	7115.30
1983	9	6749.48	6749.48	337.80	7295.04
1984	10	6847.42	6847.42	313.82	7087.28
1985	11	6823.48	6823.48	283.11	7161.24
1986	12	6740.24	6740.24	252.58	7106.59
1987	13	7023.82	7023.82	254.97	6992.82
1988	14	7303.28	7301.53	256.59	7278.79
	h				$\hat{X}_t(h)$
1989	1	7661.38			7558.12
1990	2	8816.56			7814.71
1991	3	9366.04			8071.31
1992	4	9715.20			8327.90
1993	5	9485.74			8584.49
1994	6	9974.00			8841.08
Accuracy measures					
		in-sample		out-sample	
		MSE	111010	MSE	1118507
		MAE	240	MAE	970
		MAPE	4.34	MAPE	11.70
		sMAPE	4.23	sMAPE	10.94

4.3.2 Multiplicative Forms of $ATA_{mult}(p^*, q)$ with Level Fixed

The multiplicative form of ATA will be given in this section which can be called $ATA_{mult}(p*, q)$ as an to alternative Holt's exponential trend method. The notation as given in the previous section so Holt's exponential trend method, which was defined in equations (2.22)-(2.24), will be modified as below:

$$S_t = \left(\frac{p}{t}\right) X_t + \left(\frac{t-p}{t}\right) (S_{t-1} * T_{t-1}), \tag{4.11}$$

$$T_t = \left(\frac{q}{t}\right) \frac{S_t}{S_{t-1}} + \left(\frac{t-q}{t}\right) T_{t-1}, \qquad t > p \geq q, \tag{4.12}$$

$$\hat{X}_t(h) = S_t + hT_t, \tag{4.13}$$

with constraint $S_t = X_t$ for $t \leq p$, $T_t = X_t/X_{t-1}$ for $t \leq q$ and $T_1 = 1$ where $p \in \{0, 1, \ldots, n\}$, $q \in \{0, 1, \ldots, n\}$ and $p \geq q$. Note that there are two smoothing parameters (p and q) to estimate but no starting values are needed for level and trend. The parameters can be estimated by minimizing the one step MSE, MAE, MAPE or some other criterion for measuring the in-sample forecast error. Also notice that for $q = 0$ the $ATA_{mult}(p^*, q)$ model defined in equations (4.11)-(4.13) reduces to ATA model with multiplicative trend with no trend component which is an alternative to simple exponential smoothing. Point forecasts can be obtained by foıllowing the same process as was done for $ATA(p^*, q)$.

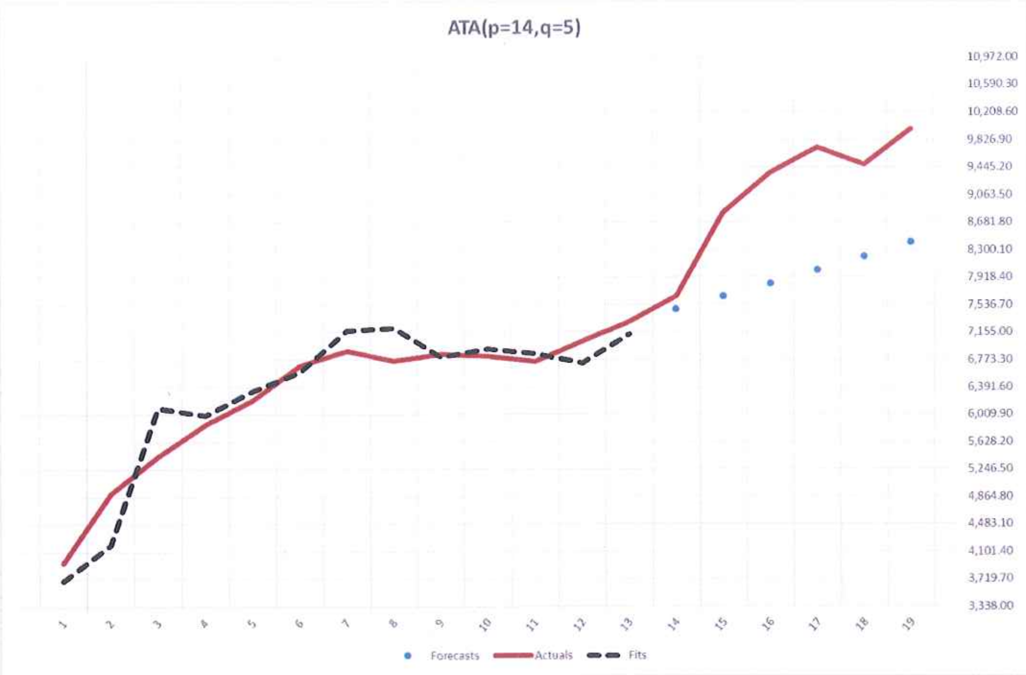

Figure 4.4 The fitted and forecast values of $ATA_{mult}(p^*, q)$ with optimal smoothing parameters

The smoothed and forecast values obtained by modeling the N96 data set by $ATA_{mult}(p^*, q)$ are given in Figure 4.4. The level and trend parameter are found to be $p = 14$ and $q = 5$. Moreover, the simple and trended versions of the multiplicative model can be obtained from one single model and when optimization is carried out the selection is done simultaneously.

Table 4.6 presents the results that are obtained from applying $ATA_{mult}(p^*, q)$ on the N96 data set. The smoothing parameters are estimated by minimizing the in sample one step ahead sMAPE. In addition to sMAPE, three different accuracy measures are calculated for training and test sets.

Table 4.6 An application of $ATA_{mult}(p^*, q)$ exponential trend method on the N96 time series data set from M3-competition.

		Data	The smoothed and forecast values		
Year	t	X	Level (S_t)	Trend (T_t)	Forecast ($\hat{X}_t$)
Year	t	X	Level	Trend	Forecast
1975	1	3709.24	3709.24	1.00	
1976	2	3947.02	3947.02	1.06	3709.24
1977	3	4907.50	4907.50	1.24	4200.04
1978	4	5425.42	5425.42	1.11	6101.71
1979	5	5866.84	5866.84	1.08	5998.00
1980	6	6211.48	6211.48	1.06	6344.17
1981	7	6689.54	6689.54	1.07	6599.78
1982	8	6896.62	6896.62	1.05	7176.77
1983	9	6749.48	6749.48	1.01	7218.42
1984	10	6847.42	6847.42	1.01	6809.45
1985	11	6823.48	6823.48	1.00	6927.52
1986	12	6740.24	6740.24	1.00	6856.17
1987	13	7023.82	7023.82	1.01	6724.82
1988	14	7303.28	7303.28	1.02	7127.59
	h				$\hat{X}_t(h)$
1989	1	7661.38			7476.42
1990	2	8816.56			7653.67
1991	3	9366.04			7835.11
1992	4	9715.20			8020.86
1993	5	9485.74			8211.02
1994	6	9974.00			8405.68
Accuracy measures					
		in-sample		out-sample	
		MSE	115510	MSE	1780932
		MAE	266	MAE	1236
		MAPE	4.70	MAPE	15.42
		sMAPE	4.65	sMAPE	14.16

4.3.3 Damped Form of $ATA_{damped}(p^*, q)$ with Level Fixed

The forecasts generated by Holt's linear method display a constant trend (increasing or decreasing) indefinitely into the future. Even more extreme are the forecasts generated by the exponential trend method which include exponential growth or decline. Empirical evidence indicates that these methods tend to over-forecast, especially for longer forecast horizons. Motivated by this observation, Gardner & McKenzie (1985) introduced a parameter that "dampens" the trend to a flat line some time in the future. Methods that include a damped trend have proven to be very successful and are arguably the most popular individual methods when forecasts are required automatically for many series (Hyndman & Athanasopoulos (2014)).

The damped form of ATA will be given in this section which will be denoted by $ATA_{damped}(p^*, q)$ as an alternative Additive damped trend method. Now, Additive damped trend method, which was defined in equations (2.25)-(2.27), will be modified as below:

$$S_t = \left(\frac{p}{t}\right) X_t + \left(\frac{t-p}{t}\right) (S_{t-1} + \phi T_{t-1}), \tag{4.14}$$

$$T_t = \left(\frac{q}{t}\right) (S_t - S_{t-1}) + \left(\frac{t-q}{t}\right) \phi T_{t-1}, \qquad t > p \geq q, \tag{4.15}$$

$$\hat{X}_t(h) = S_t + (\phi + \phi^2 + \phi^3 + \ldots + \phi^h) T_t, \tag{4.16}$$

with constraint $S_t = X_t$ for $t \leq p$, $T_t = X_t - X_{t-1}$ for $t \leq q$ and $T_1 = 0$ where $p \in \{0, 1, \ldots, n\}$, $q \in \{0, 1, \ldots, n\}$, $0 \leq \phi \leq 1$ and $p \geq q$. Note that there are three smoothing parameters (p, q and ϕ) to estimate and no starting values are needed for level and trend. Note that, the fixed damped parameter values are used in this thesis as 0.80, 0.85, 0.9 or 0.95 as these values are commonly utilized by most researchers. The parameters can be estimated by minimizing the one step MSE, MAE, MAPE or some other criterion for measuring in-sample forecast error. Also notice that for $q = 0$ the $ATA(p^*, q)$ model defined by equations (4.14)-(4.16) reduces to an additive ATA trended model with no trend component which is an alternative to simple exponential smoothing. The steps involved to obtain point forecasts using $ATA(p^*, q)$ are summarized below:

Step 1 Deseasonalize the data by the classical decomposition method of the ratio-to-moving averages, if necessary.

Step 2 Obtain the optimal value for the parameter $p = p^*$ by minimizing the in-sample mean absolute one-step-ahead forecast errors, i.e. $sMAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{|X_t - \hat{X}_t|}{|X_t + \hat{X}_t|} * 200$ where X_t is the actual value and $\hat{X}_t$ is the one-step-ahead forecasted value assuming $q = 0$.

$p^* = opt \quad p \quad for \quad q = 0, \qquad p = 1, \ldots, n$

Step 3 Define the fixed damped parameter as $\phi = 0.90$ where $0 \leq \phi \leq 1$.

Step 3 Using the optimal p^* parameter found in Step 2, find an optimal value for the parameter q $(q \leq p^*)$ by minimizing the in-sample one-step-ahead sMAPE again. (p^*, q).

Step 4 Produce reseasonalized forecasts using the proposed method with optimized parameters (p^*, q) for as many steps ahead as required.

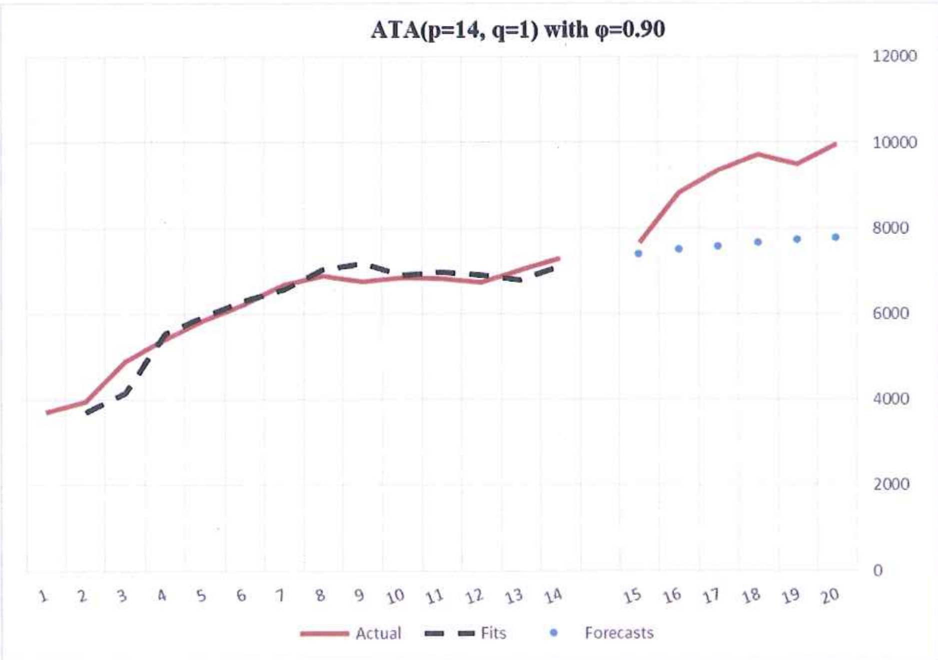

Figure 4.5 The fitted and forecast values of $ATA_{damped}(p^*, q)$ with optimal smoothing parameters

In Figure 4.5 demonstrates the results from the model $ATA_{damped}(p^*, q)$. It is clearly observed that the model reduces the effect of the over trend.

Table 4.7 An application of $ATA_{damped}(p^*, q)$ exponential trend method on the N96 time series data set from M3-competition.

		Data	The smoothed and forecast values		
Year	t	X	Level (S_t)	Trend (T_t)	Forecast ($\hat{X}_t$)
Year	t	X	Level	Trend	Forecast
1975	1	3709.24	3709.24	0.00	
1976	2	3947.02	3947.02	237.78	3709.24
1977	3	4907.50	4907.50	711.65	4161.02
1978	4	5425.42	5425.42	579.20	5547.99
1979	5	5866.84	5866.84	489.34	5946.70
1980	6	6211.48	6211.48	408.48	6307.24
1981	7	6689.54	6689.54	399.18	6579.11
1982	8	6896.62	6896.62	321.22	7048.81
1983	9	6749.48	6749.48	192.16	7185.72
1984	10	6847.42	6847.42	157.94	6922.42
1985	11	6823.48	6823.48	111.95	6989.57
1986	12	6740.24	6740.24	70.09	6924.23
1987	13	7023.82	7023.82	97.00	6803.32
1988	14	7303.28	7303.28	114.75	7111.12
	h				$\hat{X}_t(h)$
1989	1	7661.38			7406.56
1990	2	8816.56			7499.51
1991	3	9366.04			7583.16
1992	4	9715.20			7658.45
1993	5	9485.74			7726.21
1994	6	9974.00			7787.20
Accuracy measures					
		in-sample		out-sample	
		MSE	78662	MSE	2847741
		MAE	217	MAE	1560
		MAPE	3.75	MAPE	16.49
		sMAPE	3.84	sMAPE	18.22

4.4 Comparison of ATA Trended Methods

In this section, the forecasting performance of all ATA trended methods are compared based on their predictive performance on the N96 data from the M3-competition. The data spans the period 1975–1994. The data from the period of 1989–1994 is a test set, and the data up to and including the year 1988 is the training set. The smoothed data and forecasts obtained from all trended methods are compared at their optimal smoothing parameter levels in Figure 4.6. The parameters of the methods are estimated for all methods by minimizing sMAPE over the training set. In Table 4.8 the estimation results and error measures over the training and the test sets are presented.

For the simplest form of ATA method, the smoothing parameter is estimated to be 14 indicating that the last observed value is equal to the forecast value. This is expected as the series is clearly trending over time and the method requires the largest possible adjustment in each step to capture this trend.

Table 4.8 Optimum smoothing parameters and accuracy measures for different ATA models on the N96 data set

Method	$ATA(p,0)$	$ATA(p,q)$	$ATA_{mult}(p,q)$	$ATA(p^*,q)$	$ATA_{mult}(p^*,q)$	$ATA_{damped}(p^*,q)$
Parameter						
p	14	11	14	14	14	14
q	0	1	5	1	5	1
in-sample						
MSE	156124	110935	115510	111010	115510	78662
MAE	315.59	237.89	265.91	240.14	265.91	217
MAPE	6.11	4.31	4.70	4.34	4.70	3.75
sMAPE	5.78	4.19	4.65	4.23	4.65	3.84
out-sample						
MSE	4139327	1127035	1866014	1118507	1780932	2847741
MAE	1886.50	974.50	1267.26	970.22	1236.03	1560
MAPE	25.90	11.76	15.87	11.70	15.42	16.49
sMAPE	22.53	10.99	14.54	10.94	14.16	18.22

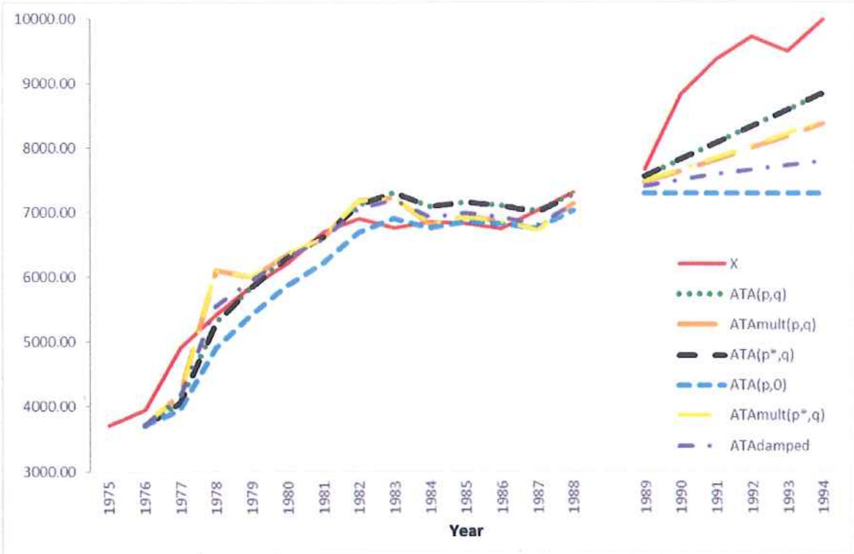

Figure 4.6 Forecasts of ATA trend methods to N96 data set in M3 competition.

The sMAPE measures calculated from the training set show that $ATA(p, q)$ the additive trended ATA method, provides the best fit to the data followed by the additive form of ATA with fixed level. $ATA(p, 0)$ generates the largest with in-sample one-step errors. In Figure 4.6, the forecasts generated by the methods can be examined. Pretending that we have not seen the data over the test-set we would conclude that all forecasts are quite plausible especially from the methods that account for the trend in the data.

Comparing the forecasting performance of the methods over the training set in Table 4.8, $ATA(p, q)$ is the most accurate method according to the MAE, MAPE and sMAPE, while $ATA(p^*, q)$ is most accurate according to the MSE. Similarly, $ATA(p^*, q)$ outperforms the others when the performance on the test sets are compared. Conflicting results like this are very common when performing forecasting competitions between methods. As forecasting tasks can vary by many dimensions (length of forecast horizon, size of test set, forecast error measures, frequency of data, etc.), it is unlikely that one method will be better than all others for all forecasting scenarios.

4.5 Results From The M3-Competition

Since the M-3 competition data (Makridakis & Hibon (2000)) is still the most recent and comprehensive time-series data collection available, the performance of the proposed combination will be evaluated by applying the proposed methods to this collection. The results from this competition are verified by the M-3 data set with forecasts for all competing methods are available in the International Journal of Forecasting's website. In this section, the results of five benchmark methods namely single, holt, winter, comb(S-H-D) and ETS are obtained from Makridakis & Hibon (2000), and are compared one by one to their counter ATA models.

Before the ATA method is applied, the data sets were deseasonalized by the classical multiplicative decomposition method as explained in previous sections, when necessary. The parameters are optimized by minimizing the in-sample one-step-ahead sMAPE and to stay consistent with rest of the literature forecasts up to 18 steps ahead (the number of steps as specified in the M3-competition) are computed and again sMAPE for all forecast horizons are computed and averaged across all 3003 series. The data sets are of various lengths, with different kinds of trend and seasonality components, time intervals between successive observation are yearly, quarterly, monthly or other and each data set consists of in and out sample data points. Yearly data have an average size of 28 and also the average size equals to 49 for quarterly, 117 for monthly and 77 for other time intervals between successive observation.

When the trended models of ATA are applied to the 3003 series, the smaller values of trend parameters performed more accurately than the larger values such as $q = 1$ especially for short data size and forecast horizons. When $q = 1$ is chosen all past trends are treated or weighted equally which helps eliminate possible over trending problems. On the other hand, for data of larger size the trend behaviour can be detected better and the parameter may have been larger values. When the performance of trended ATA models are obtained, it is observed that $ATA(p^*, q)$ gives more accurate results for data of larger size. Moreover, key aspects of the trend is hidden in the simple form of the ATA model. If the level of the observations is not correctly estimated, accurate estimation of the trend parameter may not be possible. The $ATA(p, 0)$ results obtained from the M3 competition will support this idea. Therefore, in this thesis pre determined model parameters as defined in the following list items are used to obtain the accurate forecasts. Results from seven different applications of the ATA method will be considered here.

1. $ATA(p, 0)$ is an alternative to SES method where p is the optimum value for $q = 0$.
2. $ATA(p, 1)$ where p is optimized for $q = 1$
3. $ATA(p^*, q)$ is an alternative to Holt's linear trend method where q is optimized for $p = p^*$
4. $ATA_{damped}(p^*, q)$ is an alternative to damped trend method where q is optimized for $p = p^*$ with fixed damped trend ($\phi = 0.95$)
5. ATA_{comb}(S-H-D) is an alternative to M3-competitiors of Comb(S-H-D) where q is optimized for $p = p^*$
6. $ATA - select$ where a simple model selection of the two models in (1) and (2) is carried out based on in-sample sMAPE
7. $ATA - comb$ where a simple average of the forecasts from the two models in (1) and (2) is used as a forecast.

When the patterns of trend component are examined on a time series data set like in Figure 4.7, the impact of ATA trend models on the predictive performance is clearly seen. There is an increasing trend, so enlarging the value of the trend parameter for short-length data will lead us to inconsistent estimates.

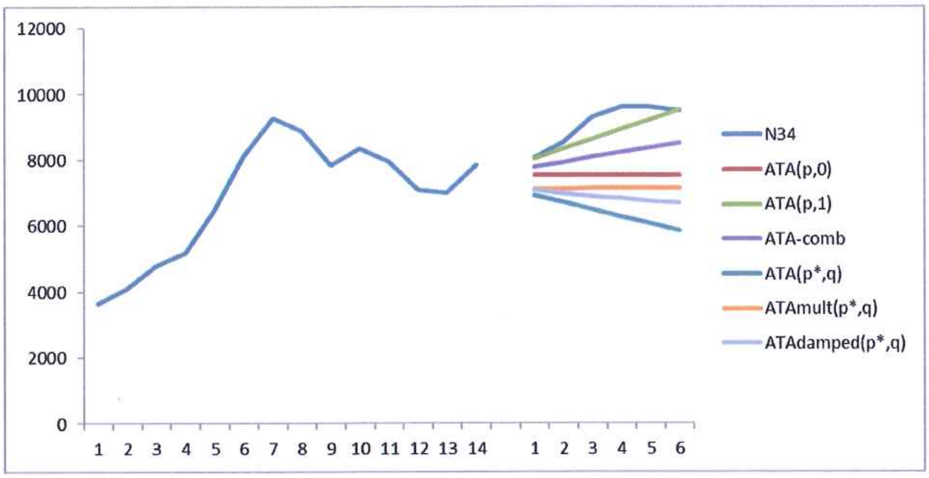

Figure 4.7 The h-step forecasts for all ATA trend methods

Reseasonalized forecasts are produced when necessary for all versions for as many steps ahead as required. The results are given in Table 4.9- Table 4.13. When the ATA models given in the list are each compared to their counter ES models, it can be seen that they all perform better both for individual forecast horizons and an average except for the case when Comb(S-H-D) is compared to ATA_{comb}(S-H-D) on the forecast horizon 18. (ATA_{comb}(S-H-D)' sMAPE 18.4 and Comb(S-H-D)' sMAPE 18.3)

The results from all 3003 data sets can be summarized as in Table 4.9. Here when the methods are compared based on the average sMAPE for forecasting horizons 1-18, $ATA - comb$ stands out from the rest, ranking first. It performs better than not just the pure approaches like ETS and Comb(S-H-D), it outperforms all existing methods. It is worth noting that $ATA - comb$ is more accurate than ETS and Comb(S-H-D) for all individual forecasting horizons, not just on average.

The success of ATA is evident even when just the results from the simplest version of the method ($ATA(p^*, 0)$) are studied. This simple version performs better than SES for all forecasting horizons and on average and its average sMAPE for horizons $1-18$ is the same as ETS' sMAPE (13.77). The ATA with linear trend component ($ATA(p^*, q)$) is again more accurate than its competitor Holt on all forecasting horizons.

ATA method does not perform as well when the results are averaged just for the seasonal series. This can be attributed to the fact that the ATA models we considered here do not model seasonality like the other competitors. Still, $ATA - select$ ranks second among all the competitors when the average sMAPE for horizons $1 - 18$ are considered. For non-seasonal data, ATA models inarguably perform as well as Theta and much better than the other methods.

For annual data $ATA - select$ performs better than all other methods for both short term and long term forecasting horizons except the first horizon. In addition, it is worth noting that these results show a 3% improvement compared to slightly the nearest competitor, the RBF model. When compared the SES model performs better than $ATA(p, 0)$ but when the average values of the optimal smoothing parameters(α) are examined, it is observed that the value equals to 0.9177 for SES and 0.8266 for ATA. These findings support the idea that ATA obtains the forecast values with smaller variance. Also, for SES a smoothing value larger than 0.5 is considered to be problematic. On the other hand, when the $ATA(p^*, q)$ and $Holt$ models are compared, it is apparent that the performance of the ATA model improves by 8.33% on sMAPE on average. These results are evidence that most of the yearly data sets have meaningful trends. However, it is also supported this idea that the average symmetric MAPE is 16.7 for $ATA(p, 1)$ and 17.6 for $ATA(p^*, q)$. It is a significant finding for forecasting in literature since the best performance of ATA can be reached with only two smoothing trend parameters ($q = 0, 1$)

The quarterly data has up to trend and seasonality patterns as explained previous sections. For these data, $ATA - comb$ ranks second right after Theta when sMAPE is averaged for horizons $1 - 4$, $1 - 6$ and $1 - 8$. Note that the model is better than other popular models according to simplicity, optimization speed and accuracy. For monthly data it outperforms all other methods when errors are averaged for horizons $1 - 18$. Further, the out sample results obtained by applying the ATA method are 5.88% better than the single method. $ATA(p^*, q)$ performs noticeably well for the 174 other data sets.

In this chapter a combination of forecasts from the ATA method is proposed and the

Table 4.9 Average symmetric MAPE across different forecast horizons: all 3003 series

	Forecasting horizons										Averages					
Method	1	2	3	4	5	6	8	12	15	18	1-4	1-6	1-8	1-12	1-15	1-18
Naive2	10.5	11.3	13.6	15.1	15.1	15.9	14.5	16.0	19.3	20.7	12.62	13.57	13.76	14.24	14.81	15.47
Single	9.5	10.6	12.7	14.1	14.3	15.0	13.3	14.5	18.3	19.4	11.73	12.71	12.84	13.13	13.67	14.32
Holt	9.0	10.4	12.8	14.5	15.1	15.8	13.9	14.8	18.8	20.2	11.67	12.93	13.11	13.42	13.95	14.60
Winter	9.1	10.5	12.9	14.6	15.1	15.9	14.0	14.6	18.9	20.2	11.77	13.01	13.19	13.48	14.01	14.65
Dampen	8.8	10.0	12.0	13.5	13.8	14.3	12.5	13.9	17.5	18.9	11.07	12.05	12.17	12.45	12.98	13.64
Comb (S-H-D)	8.9	10.0	12.0	13.5	13.7	14.2	12.4	13.6	17.3	18.3	11.10	12.04	12.13	12.4	12.91	13.52
ETS	8.8	9.8	12.0	13.5	13.9	14.7	13.0	14.1	17.6	18.9	11.04	12.13	12.32	12.66	13.14	13.77
$ATA(p,0)$	8.9	10.0	12.1	13.7	13.9	14.7	12.8	13.9	17.3	18.9	11.16	12.21	12.34	12.64	13.13	13.77
$ATA(p,1)$	8.4	9.7	11.5	12.9	13.6	14.2	12.9	15.4	18.9	20.9	10.64	11.72	11.94	12.66	13.32	14.09
$ATA(p^*,q)$	8.6	9.7	11.7	13.5	13.7	14.5	12.5	13.6	17.3	18.7	10.89	11.96	12.08	12.37	12.87	13.50
$ATA_{damped}(p^*,q)$	8.6	9.6	11.6	13.2	13.5	14.2	12.4	13.7	17.0	18.6	10.76	11.77	11.92	12.24	12.75	13.39
ATA_{comb}(S-H-D)	8.6	9.6	11.5	13.1	13.2	13.9	12.2	13.5	16.8	18.4	10.70	11.65	11.78	12.11	12.61	13.25
$ATA-select$	8.7	9.6	11.5	12.9	13.1	13.7	12.1	13.7	17.3	18.7	10.69	11.58	11.70	12.09	12.64	13.28
$ATA-comb$	8.5	9.6	11.4	12.8	13.0	13.6	12.0	13.1	16.3	17.4	10.56	11.47	11.58	11.94	12.40	12.94

Table 4.10 Average symmetric MAPE across different forecast horizons: 645 annual series

	Forecasting horizons						Averages	
Method	1	2	3	4	5	6	1-4	1-6
Naive2	8.5	13.2	17.8	19.9	23.0	24.9	14.85	17.88
Single	8.5	13.3	17.6	19.8	22.8	24.8	14.82	17.82
Holt	8.3	13.7	19.0	22.0	25.2	27.3	15.77	19.27
Winter	8.3	13.7	19	22	25.2	27.3	15.77	19.27
Dampen	8.0	12.4	17	19.3	22.3	24.0	14.19	17.18
Comb (S-H-D)	7.9	12.4	16.9	19.3	22.2	23.7	14.11	17.07
ETS	9.3	13.6	18.3	20.8	23.4	25.8	15.48	18.53
$ATA(p,0)$	9.1	13.5	17.6	19.9	22.8	25.1	15.04	18.00
$ATA(p,1)$	8.3	12.2	16.8	18.6	21.5	23.3	13.95	16.78
$ATA(p^*,q)$	8.3	12.4	17.0	20.0	23.1	25.1	14.40	17.62
$ATA_{damped}(p^*,q)$	8.2	12.1	16.4	19.1	21.9	23.5	13.94	16.87
ATA_{comb}(S-H-D)	8.2	12.0	16.1	18.6	21.3	22.9	13.73	16.51
$ATA-select$	8.3	11.5	15.6	17.7	20.5	22.0	13.28	15.94
$ATA-comb$	8.4	12.3	16.5	18.3	21.0	22.7	13.87	16.54

Table 4.11 Average symmetric MAPE across different forecast horizons: 756 quarterly series.

	Forecasting horizons							Averages		
Method	1	2	3	4	5	6	8	1-4	1-6	1-8
Naive2	5.4	7.4	8.1	9.2	10.4	12.4	13.7	7.55	8.82	9.95
Single	5.3	7.2	7.8	9.2	10.2	12.0	13.4	7.38	8.63	9.72
Holt	5.0	6.9	8.3	10.4	11.5	13.1	15.6	7.67	9.21	10.67
Winter	5.0	7.1	8.3	10.2	11.4	13.2	15.3	7.65	9.21	10.61
Dampen	5.1	6.8	7.7	9.1	9.7	11.3	12.8	7.18	8.29	9.33
Comb S-H-D	5.0	6.7	7.5	8.9	9.7	11.2	12.8	7.03	8.16	9.22
ETS	5.0	6.6	7.9	9.7	10.9	12.1	14.2	7.32	8.71	9.94
$ATA(p,0)$	5.2	7.1	7.8	9.7	10.1	11.8	13.5	7.45	8.62	9.71
$ATA(p,1)$	5.3	6.8	7.6	9.1	9.9	11.0	12.4	7.19	8.28	9.24
$ATA(p^*,q)$	5.0	6.9	7.6	9.5	10.2	11.9	13.7	7.23	8.51	9.69
$ATA_{damped}(p^*,q)$	5.0	6.8	7.5	9.4	10.0	11.8	13.5	7.17	8.42	9.56
ATA_{comb}(S-H-D)	5.0	6.8	7.4	9.2	9.7	11.4	13.0	7.12	8.26	9.33
$ATA-select$	5.2	7.0	7.7	9.2	9.6	11.1	12.3	7.28	8.30	9.21
$ATA-comb$	5.1	6.8	7.5	9.0	9.6	10.9	12.3	7.10	8.13	9.07

Table 4.12 Average symmetric MAPE across different forecast horizons: 1428 monthly series.

	Forecasting horizons										Averages					
Method	1	2	3	4	5	6	8	12	15	18	1-4	1-6	1-8	1-12	1-15	1-18
Naive2	15.0	13.5	15.7	17.0	14.9	14.4	15.6	16.0	19.3	20.7	15.30	15.08	15.26	15.55	16.16	16.89
Single	13.0	12.1	14.0	15.1	13.5	13.1	13.8	14.5	18.3	19.4	13.53	13.44	13.60	13.83	14.51	15.32
Holt	12.2	11.6	13.4	14.6	13.6	13.3	13.7	14.8	18.8	20.2	12.95	13.11	13.33	13.77	14.51	15.36
Winter	12.5	11.7	13.7	14.7	13.6	13.4	14.1	14.6	18.9	20.2	13.17	13.28	13.52	13.88	14.62	15.44
Dampen	11.9	11.4	13.0	14.2	12.9	12.6	13.0	13.9	17.5	18.9	12.63	12.67	12.85	13.10	13.77	14.59
Comb S-H-D	12.3	11.5	13.2	14.3	12.9	12.5	13.0	13.6	17.3	18.3	12.83	12.79	12.92	13.11	13.75	14.48
ETS	11.5	10.6	12.3	13.4	12.3	12.3	13.2	14.1	17.6	18.9	11.93	12.05	12.43	12.96	13.64	14.45
$ATA(p,0)$	11.5	10.8	12.6	13.8	12.6	12.5	12.9	13.9	17.3	18.9	12.20	12.33	12.78	12.98	13.67	14.49
$ATA(p,1)$	11.0	10.9	12.2	13.4	12.8	12.8	13.8	15.4	18.9	20.9	11.86	12.16	13.19	14.07	15.01	15.33
$ATA(p^*,q)$	11.6	10.9	12.5	13.8	12.4	12.3	12.7	13.6	17.3	18.7	12.18	12.37	12.82	13.16	13.93	15.06
$ATA_{damped}(p^*,q)$	11.5	10.9	12.5	13.7	12.4	12.3	12.6	13.7	17.0	18.6	12.13	12.19	12.40	12.77	13.44	14.24
ATA_{comb}(S-H-D)	11.5	10.8	12.5	13.7	12.4	12.2	12.5	13.5	16.8	18.4	12.11	12.17	12.36	12.71	13.36	14.15
$ATA-select$	11.6	11.0	12.6	13.8	12.4	12.3	12.7	13.7	17.3	18.7	12.22	12.26	12.65	12.84	13.53	14.31
$ATA-comb$	11.1	10.7	12.1	13.1	12.0	11.9	12.4	13.1	16.3	17.4	11.75	11.83	12.07	12.50	13.08	13.76

Table 4.13 Average symmetric MAPE across different forecast horizons: 174 other series.

	Forecasting horizons							Averages		
Method	1	2	3	4	5	6	8	1-4	1-6	1-8
Naive2	2.2	3.6	5.4	6.3	7.8	7.6	9.2	4.38	5.49	6.30
Single	2.1	3.6	5.4	6.3	7.8	7.6	9.2	4.36	5.48	6.29
Holt	1.9	2.9	3.9	4.7	5.8	5.6	7.2	3.32	4.13	4.81
Winter	1.9	2.9	3.9	4.7	5.8	5.6	7.2	3.32	4.13	4.81
Dampen	1.8	2.7	3.9	4.7	5.8	5.4	6.6	3.28	4.06	4.61
Comb S-H-D	1.8	2.8	4.1	4.7	5.8	5.3	6.2	3.36	4.09	4.56
ETS	2.0	3.0	4.0	4.4	5.4	5.1	6.3	3.37	3.99	4.51
$ATA(p,0)$	2.1	3.5	5.4	6.3	7.8	7.5	9.1	4.34	5.45	6.26
$ATA(p,1)$	1.9	2.9	4.1	4.8	6.0	5.7	7.1	3.46	4.26	4.87
$ATA(p^*,q)$	1.7	2.6	3.7	4.3	5.4	4.9	6.2	3.09	3.77	4.30
$ATA_{damped}(p^*,q)$	1.8	2.7	4.1	4.7	5.8	5.3	6.3	3.33	4.06	4.55
ATA_{comb}(S-H-D)	1.8	2.8	4.2	4.7	5.9	5.3	6.2	3.38	4.10	4.56
$ATA-select$	1.8	2.8	4.0	4.6	5.7	5.3	6.6	3.30	4.03	4.60
$ATA-comb$	1.9	3.0	4.5	5.1	6.3	5.8	6.8	3.62	4.40	4.93

proposed approach's forecasting performance is investigated. Even though the models from the ATA method have similar form to their counter ES models, the proposed combination's predictive performance is much better for the M3 data sets. The optimum parameter values, forecasts and errors for the proposed method can be reached from the website *https://atamethod.wordpress.com*.

The results presented in this section do not reflect the end performance of ATA. In this section we competed with just one ATA model (linearly trended with additive errors) and combined forecasts from only two parameterizations of it. Incorporating other types of trend using different accuracy measures will surely increase the method's performance which will be discusses in chapter 5. The fact that this simple combination can perform better than existing methods is fascinating and this further strengthens the idea that simplicity is indeed a prerequisite for forecasting accuracy.

CHAPTER FIVE
BEST PERFORMANCE OF ATA METHOD

In the previous chapters, ATA method has been proposed as an alternative to ES models and we have seen there is one to one correspondence between them. The results presented in the previous chapters do not reflect the end performance of ATA and on the contrary those results were just the initial findings. ahen the two approasches are compared model to model on an individual base, ATA models consistently outperforms their counter ES models. There are numerous ways to increase the performance of ES methods. Most of the attempts to increase the performance are based on combinations of different ES models, model selection rules, transformations and sophisticated data preprocessing techniques. These methods were employed by some of the competitors in the M3 competition.

In order to improve our methods performance, in this section we will implement similar ideas to our methodology. More specifically, we will propose certain model selection rules for data of different time intervals and we eill consider combinations of forecasts from different parameterizations of our model. The details on these ideas are given in the following subsection. Later on we will compare the outsample performance of these newly proposed methods to existing ones on the M3 competition data based on the metrics sMAPE and MASE.

Before the ATA method is applied, the data sets were deseasonalized by the classical multiplicative decomposition method, when necessary. The parameters are optimized by minimizing the in-sample one-step-ahead sMAPE which is defined in equation (2.40) and to stay consistent with rest of the literature forecasts up to 18 steps ahead (the number of periods ahead as 6 for yearly, 8 for quarterly/other and 18 for monthly series) are computed and again sMAPE and also MASE which is defined in equation (2.47) for all forecast horizons are computed and averaged across all 3003 series.

5.1 Combination and Selection Criterion of ATA Models

As it was stated before the accuracy when various methods are being combined outperforms, on average, the individual methods being combined and the accuracy of the various methods depends upon the length of the forecasting horizon involved. Therefore, we will concentrate on the extra improvement in accuracy since the “best” method varies according to the accuracy measure being used and the type of data (micro, industry, macro, etc.) involved. Such differentiation becomes clearer if the data are further subdivided into yearly, quarterly, monthly and ‘other’ since size and characteristics of each data set are different. The method "Comp S-H-D" in the M3 Competition is the simple arithmetic average of three methods: Single, Holt and Dampen Trend Exponential Smoothing. Clearly, the combination is more accurate than the three individual methods being combined for practically all forecasting horizons. Therefore, similar attempts will be tried in this section for ATA method in order to improve forecasting accuracy. As Makridakis & Hibon (2000) states that any improvement in out-sample forecasting accuracy can result in savings of many millions of dollars, less wasted resources, and/or better service.

Different ATA models can easily be constructed by using different smoothing parameter values. These derivative models highlight different aspects of the original data, as different smoothing parameters helps in strengthening or attenuating the signals of different time series components. In each series, the appropriate ATA method is fitted and its respective time series components are forecasted. This approach achieves a better estimation of the different time series components, through the various smoothing parameters and reduces the importance of model selection through forecast combination. In this section, different modelling strategies were employed for different types of data. An empirical evaluation of the proposed framework shows significant improvements in forecasting accuracy by taking advantage of, the easy optimization process.

More specifically, for annual data a model selection was carried out between the models $ATA(p^*, 0)$, $ATA(p, 1)$ and $ATA_{damped}(p, 1)$ based on their in sample one step ahead sMAPE values where p^* is the optimum value of p for $q = 0$. For quarterly data, a simple combination of forecasts from the models $ATA(p^*, 0)$, $ATA(p, 1)$ and $ATA_{damped}(p, 1)$ selection and $ATA(p^*, q)$ was used by obtaining the average of the three forecasts from these models. For monthly data, in addition to $ATA(p^*, 0)$, $ATA(p, 1)$ and $ATA(p, 2)$ also contributed to the combination for forecast horizons 1-4. for horizons 1-5, a simple combination of forecasts from the models $ATA(p^*, 0)$, $ATA(p, 1)$ and for forecast horizons 7-10 a simple combination of forecasts from the models $ATA(p^*, 0)$, $ATA(p, 1)$ and $ATA(p^*, 0)$. So, for monthly data, the average of forecasts from these four models was used. Finally for the other data sets, we carried out a simple model selection based on in-sample sMAPE again, between $ATA(p^*, q)$ as before and a multiplicative version of the same model $ATA_{mult}(p^*, q)$. To summarize, all selection and combination criteria can be given in a table as below:

Table 5.1 Model selection and combination criteria

Model selection and combination criteria		
Data	Forecast horizon	Methods
Yearly	1-6	select$(ATA(p, 0)$; select$(ATA(p, 1); ATA_{damped}(p, 1)))$
Quarterly	1-8	combine$(ATA(p, 0)$; select$(ATA(p, 1); ATA_{damped}(p, 1)); ATA(p^*, q))$
Monthly	1-4	combine$(ATA(p, 0); ATA(p, 1); ATA(p, 2))$
	5-6	combine$(ATA(p, 0); ATA(p, 1))$
	7-10	combine$(ATA(p, 0); ATA(p^*, q); ATA(p, 2))$
	11-18	combine$(ATA(p, 0); ATA(p^*, q); ATA(p, 1))$
Other	1-8	select$(ATA(p^*, q); ATA_{mult}(p^*, q))$

5.2 M3-Competition Results

The M3 competition data set consists of 645 yearly, 756 quarterly, 1428 monthly and 174 other series. The original data, as well as the forecasts of the methods that participated in the competition, are available in the R package Mcomp and website (International Institute of Forecasters-IIF). Although the M3 competition took place some time ago, the original submissions to the competition are still competitive and valid benchmarks. So, the results obtained for the M3 competition data are still of interest and will be discussed in detail.

In this thesis, R package is used to obtained forecast values from ATA methods. The algorithm for all models are constructed manually in R. All details and R codes for all ATA models will be given in the appendix.

The strategy explained in the previous section is used when modeling the M3 competition data and in the following tables the results obtained in this fashion are labelled as $ATA - best$. Reseasonalized forecasts are produced when necessary for all versions for as many steps ahead as required. The results are given in Table 5.2- Table 5.7.

It can be clearly seen that ATA out-performs all its competitors when the short term and long term forecasting horizons are averaged for both metrics. It is especially interesting that ATA outperforms ETS which performs a model selection from 30 possible ES models (models include different types of trend (multiplicative, damped, multiplicative damped) and seasonal components (additive, multiplicative) in addition to different types of errors (additive and multiplicative errors)) utilizing AIC as a selection criterion. Moreover, according to the sMAPE results, the performance of the proposed $ATA - best$ method yields an error that is 5.61% smaller than that of the Theta method, best method in the original M3-competition. It can be seen taht, when MASE is used, again $ATA - best$ performs better than the other competitior, yielding on error that is 11.11% smaller than Theta. It is also worth noting that the results from

Table 5.2 Average symmetric MAPE across different forecast horizons: all 3003 series

	Forecasting horizons										Averages					
Method	1	2	3	4	5	6	8	12	15	18	1-4	1-6	1-8	1-12	1-15	1-18
Ata	8.3	9.4	11.0	12.3	12.6	13.2	11.7	12.9	16.2	17.3	10.27	11.14	11.27	11.62	12.10	12.68
Theta	8.4	9.6	11.3	12.5	13.2	13.9	12.0	13.2	16.4	18.4	10.45	11.49	11.62	11.96	12.45	13.05
ForecastPro	8.6	9.6	11.4	13.0	13.4	14.3	12.7	13.3	16.4	18.3	10.66	11.72	11.90	12.18	12.64	13.23
ForecastX	8.7	9.8	11.6	13.1	13.2	13.8	12.6	14.0	17.8	18.8	10.82	11.72	11.88	12.22	12.81	13.50
Comp S-H-D	8.9	10.0	12.0	13.5	13.7	14.0	12.4	13.6	17.3	18.3	11.10	12.02	12.11	12.39	12.90	13.51
Dampen	8.8	10.0	12.0	13.5	13.8	14.3	12.5	13.9	17.5	18.9	11.07	12.05	12.17	12.45	12.98	13.64
RBF	9.9	10.5	12.4	13.4	13.2	14.1	12.8	14.1	17.3	17.8	11.56	12.26	12.40	12.76	13.24	13.74
ETS	8.8	9.8	12.0	13.5	13.9	14.7	13.0	14.1	17.6	18.9	11.04	12.13	12.32	12.66	13.14	13.77
B-J automatic	9.2	10.4	12.2	13.9	14.0	14.6	13.0	14.1	17.8	19.3	11.42	12.39	12.52	12.78	13.33	13.99
SmartFcs	9.2	10.3	12.0	13.5	14.0	15.0	13.0	14.9	18.0	19.4	11.23	12.32	12.47	12.93	13.47	14.11
PP-aotocast	9.1	10.1	12.2	13.7	14.0	14.6	13.3	14.4	17.9	19.8	11.27	12.28	12.49	12.92	13.47	14.14
Flores/Pearce2	10.0	11.0	12.8	14.1	14.1	14.6	12.9	14.4	18.2	20.2	11.96	12.76	12.80	13.03	13.60	14.30
Single	9.5	10.6	12.7	14.1	14.3	14.9	13.3	14.5	18.3	19.4	11.73	12.68	12.82	13.12	13.66	14.31
Theta-sm	9.5	10.6	12.6	13.8	14.8	15.6	13.4	14.4	17.7	19.3	11.62	12.80	12.98	13.31	13.75	14.34
AutoBox2	9.5	10.4	12.2	13.9	13.9	14.8	13.3	15.2	18.4	20.0	11.53	12.47	12.67	13.14	13.75	14.46
Flores/Pearce1	9.2	10.5	12.6	14.5	14.8	15.2	13.8	14.4	19.1	21.0	11.68	12.78	13.03	13.31	13.92	14.72
Ararma	9.7	10.9	12.6	14.2	14.6	15.5	13.9	15.2	18.5	20.3	11.85	12.91	13.12	13.54	14.09	14.74
Holt	9.0	10.5	12.9	14.8	15.6	16.2	14.4	15.3	19.5	21.1	11.79	13.16	13.39	13.75	14.33	15.03
Winter	9.1	10.5	13.1	14.9	15.6	16.3	14.5	15.2	19.6	21.1	11.90	13.26	13.48	13.83	14.41	15.11
AutoBox1	9.8	11.1	13.1	15.1	16.0	16.7	14.2	15.4	19.1	20.4	12.30	13.64	13.76	13.99	14.54	15.21
Naive2	10.5	11.3	13.6	15.1	15.1	15.8	14.5	16.0	19.3	20.7	12.62	13.55	13.74	14.22	14.80	15.46
AutoBox3	9.8	11.3	13.1	14.8	16.1	16.6	14.8	16.7	19.8	22.0	12.24	13.62	13.87	14.28	14.88	15.71
Auto-ANN	9.0	10.5	11.9	13.9	13.9	15.6	13.5	14.8	17.5	19.8	11.32	12.47	12.66	13.05	13.59	14.23
Robust-Trend	10.5	11.2	13.3	14.8	15.2	15.9	15.3	18.0	23.2	25.7	12.46	13.49	13.84	14.77	15.70	16.70

the THETA model that we present here are obtained from the forecasts provided on the IIF website and are a little different from the M3 competition paper (Makridakis & Hibon (2000)). When the results shown in Table 5.2 and Table 5.3 are examined, it can be clearly seen that ATA is consistently more accurate across all horizons regardless of accuracy measures. These results are very impressive and the best forecasting methods in litarature can never achieves these.

5.2.1 Best Performance of ATA for Yearly Data

In this section, the results are shown for all methods on just the yearly data. the forecasts of $ATA - best$ are obtained and the two error metrics are calculated. It can be seen that the $ATA - best$ using different modelling strategies performs better than the all original M3 competition methods, consistently outperforming them in all measures. The difference that the ATA method makes can be even more evidently seen in Table 5.4

Table 5.3 Average MASE across different forecast horizons: all 3003 series

	Forecasting horizons										Averages					
Method	1	2	3	4	5	6	8	12	15	18	1-4	1-6	1-8	1-12	1-15	1-18
Ata	0.68	0.85	1.10	1.32	1.46	1.62	1.15	0.85	1.05	1.20	0.99	1.17	1.16	1.10	1.08	1.09
Theta	0.75	0.95	1.28	1.53	1.73	1.94	1.34	0.94	1.19	1.35	1.04	1.24	1.23	1.15	1.13	1.14
ForecastPro	0.68	0.89	1.19	1.48	1.67	1.87	1.28	0.85	1.07	1.24	1.06	1.30	1.29	1.20	1.17	1.17
ForecastX	0.70	0.90	1.17	1.42	1.57	1.74	1.26	0.91	1.16	1.29	1.05	1.25	1.24	1.17	1.16	1.17
Comp S-H-D	0.70	0.90	1.21	1.46	1.63	1.80	1.24	0.89	1.14	1.27	1.07	1.28	1.27	1.19	1.17	1.18
Dampen	0.70	0.91	1.23	1.51	1.69	1.87	1.27	0.91	1.16	1.31	1.09	1.32	1.30	1.22	1.20	1.21
RBF	0.82	0.98	1.27	1.49	1.62	1.81	1.30	0.89	1.11	1.21	1.14	1.33	1.32	1.24	1.21	1.21
B-J automatic	0.73	0.95	1.28	1.56	1.74	1.93	1.32	0.92	1.18	1.33	1.13	1.36	1.35	1.26	1.24	1.24
SmartFcs	0.79	0.98	1.25	1.52	1.66	1.85	1.28	0.95	1.16	1.30	1.14	1.34	1.32	1.24	1.22	1.23
PP-aotocast	0.72	0.94	1.27	1.54	1.72	1.91	1.33	1.01	1.23	1.42	1.12	1.35	1.34	1.27	1.26	1.27
Flores/Pearce2	0.88	1.05	1.31	1.54	1.69	1.87	1.32	0.96	1.19	1.50	1.19	1.39	1.37	1.28	1.26	1.27
Single	0.78	1.03	1.36	1.63	1.82	2.05	1.49	0.97	1.25	1.38	1.20	1.45	1.44	1.35	1.32	1.33
Theta-sm	0.75	0.95	1.28	1.53	1.73	1.94	1.34	0.94	1.19	1.35	1.13	1.36	1.35	1.27	1.25	1.25
AutoBox2	0.77	0.95	1.23	1.48	1.62	1.83	1.34	1.13	1.23	1.97	1.11	1.31	1.31	1.26	1.26	1.30
Ararma	0.73	0.97	1.29	1.60	1.86	2.16	1.29	0.89	1.12	1.30	1.15	1.44	1.40	1.30	1.27	1.26
Flores/Pearce1	0.75	0.96	1.26	1.55	1.74	1.93	1.38	0.94	1.26	1.53	1.13	1.36	1.36	1.28	1.26	1.28
Holt	0.71	0.93	1.26	1.56	1.76	1.96	1.33	0.91	1.15	1.32	1.12	1.37	1.35	1.26	1.24	1.24
Winter	0.72	0.96	1.73	1.56	1.77	1.97	1.33	0.93	4.18	1.31	1.24	1.45	1.42	1.33	1.44	1.42
AutoBox1	0.80	1.05	1.38	1.70	1.94	2.23	1.37	0.92	1.16	1.33	1.23	1.52	1.48	1.36	1.33	1.32
Naive2	0.82	1.06	1.39	1.67	1.85	2.09	1.53	1.02	1.29	1.44	1.24	1.48	1.48	1.39	1.37	1.37
AutoBox3	0.78	1.01	1.29	1.56	1.75	1.95	1.37	0.99	1.20	1.40	1.16	1.39	1.38	1.29	1.27	1.28
Auto-ANN	0.72	0.95	1.22	1.50	1.69	1.95	1.36	0.96	1.16	1.34	1.10	1.34	1.33	1.25	1.24	1.24
Robust-Trend	0.77	0.95	1.22	1.44	1.61	1.78	1.31	0.99	1.26	1.45	1.10	1.30	1.29	1.24	1.24	1.25

for the annual series. The method is again the most accurate here.

The literature on forecasting competition has highlighted that the performance of a forecasting method differs according to the used accuracy measure. The $ATA-best$ method apparently outperforms all other methods for both average sMAPE and MASE metrics. However, all other methods move positions when the metric changes in performance order. For instance, Robust trend has moved up from 5 to 2. This is an interesting result because the Robust Trend method for all series has been illustrated the worst performance.

5.2.2 Best Performance of ATA for Quarterly Data

The accuracy performance of quarterly series is demonstrated and interpreted in this section. For quarterly series in Table 5.5, it is noticeable that the Theta and ATA models produce significantly better forecasts for all horizons compared to the other

Table 5.4 The accuracy perfomance of all methods for the 645 annual series at different (sMAPE and MASE) metrics

mean sMAPE			mean MASE		
	Averages			Averages	
Methods	1-4	1-6	Methods	1-4	1-6
Ata	13.08	15.601	Ata	2.01	2.55
RBF	13.75	16.42	Robust-Trend	2.03	2.62
ForecastX	13.80	16.48	RBF	2.10	2.72
AutoBox2	13.69	16.59	AutoBox2	2.14	2.75
Theta	14.06	16.97	ForecastX	2.17	2.77
Robust-Trend	13.87	17.03	Theta	2.19	2.81
Comp S-H-D	14.11	17.07	Comp S-H-D	2.23	2.88
PP-aotocast	14.12	17.12	Flores/Pearce1	2.26	2.94
Flores/Pearce1	14.22	17.20	SmartFcs	2.36	3.00
ForecastPro	14.19	17.27	Theta-sm	2.29	3.01
Dampen	14.30	17.36	Flores/Pearce2	2.42	3.02
SmartFcs	14.95	17.71	PP-aotocast	2.31	3.02
B-J automatic	14.78	17.72	ForecastPro	2.30	3.03
Single	14.82	17.81	Dampen	2.32	3.03
Flores/Pearce2	15.31	17.84	Auto-ANN	2.33	3.03
Naive2	14.85	17.88	B-J automatic	2.45	3.16
Theta-sm	14.60	17.92	Single	2.48	3.17
Ararma	15.17	18.35	Naive2	2.48	3.17
ETS	15.48	18.53	AutoBox3	2.51	3.18
Auto-ANN	15.39	18.56	Winter	2.41	3.18
Winter	16.19	20.02	Holt	2.41	3.18
Holt	16.19	20.02	Ararma	2.50	3.48
AutoBox3	17.48	20.87	AutoBox1	2.72	3.68
AutoBox1	17.57	21.58	ETS	-	2.83

approaches. However, the first four successful models retain their performance against different metrics but others can not. In addition, it is surprisingly that the single model performs better than Holt and Winter, eventhough the data set contains seasonal and trend patterns.

5.2.3 Best Performance of ATA for Monthly Data

Table 5.6 shows the results from all methods on the monthly data, ordered by their average sMAPE and MASE. It is known that the monthly series are seasonal, trended and have long size, so their patterns are more complex. From Table 5.6 it can be seen that even though we do not model the seasonality properly, the method still performs better than ARIMA and ETS on both short and long term forecasting horizons. The ATA method also consistently outperforms all the original methods from the M3 in all measures for monthly data. While the performance of other successful methods can changes according to sMAPE and MASE, this is not the case for ATA.

5.2.4 Best Performance of ATA for Other Data

Table 5.7 shows the results for all methods on the other data, ordered by means of sMAPE and MASE. Across the complete forecast horizon, the ATA method again outperforms all other methods for all error metrics.

Table 5.5 The accuracy perfomance of 756 quarterly series at different (sMAPE and MASE) metrics

mean sMAPE				mean MASE			
	Averages				Averages		
Methods	1-4	1-6	1-8	Methods	1-4	1-6	1-8
Ata	6.96	7.96	8.89	Ata	0.79	0.93	1.07
Theta	7.00	8.04	8.96	Theta	0.81	0.95	1.09
Comp S-H-D	7.03	8.16	9.22	Comp S-H-D	0.79	0.95	1.10
Dampen	7.18	8.29	9.36	Dampen	0.81	0.96	1.13
PP-aotocast	7.12	8.28	9.39	PP-aotocast	0.80	0.96	1.13
ForecastX	7.12	8.35	9.54	Robust-Trend	0.86	1.02	1.15
RBF	7.69	8.67	9.57	ForecastX	0.81	0.97	1.15
Single	7.38	8.63	9.72	RBF	0.90	1.03	1.17
Robust-Trend	7.63	8.86	9.79	Flores/Pearce1	0.84	1.01	1.18
ForecastPro	7.28	8.60	9.82	Ararma	0.86	1.01	1.18
Theta-sm	7.59	8.75	9.82	AutoBox2	0.87	1.02	1.19
ETS	7.32	8.71	9.94	B-J automatic	0.84	1.01	1.19
Naive2	7.55	8.82	9.95	ForecastPro	0.83	1.01	1.20
Flores/Pearce1	7.48	8.78	9.95	Theta-sm	0.87	1.04	1.21
AutoBox2	7.79	8.98	10.00	Winter	0.85	1.03	1.22
SmartFcs	8.02	9.16	10.15	Holt	0.85	1.04	1.23
Ararma	8.03	9.16	10.19	SmartFcs	0.91	1.06	1.23
Auto-ANN	7.80	9.10	10.20	Single	0.86	1.05	1.23
B-J automatic	7.79	9.10	10.26	Naive2	0.87	1.06	1.24
Flores/Pearce2	8.57	9.54	10.43	Auto-ANN	0.87	1.06	1.24
Winter	7.65	9.34	10.84	Flores/Pearce2	0.97	1.11	1.26
Holt	7.67	9.34	10.94	AutoBox3	0.90	1.09	1.27
AutoBox1	7.95	9.52	10.96	AutoBox1	0.93	1.13	1.33
AutoBox3	8.14	9.71	11.19	ETS	-	-	1.18

Table 5.6 The accuracy perfomance of 1428 monthly series at different (sMAPE and MASE) metrics

mean sMAPE							mean MASE						
	Averages							Averages					
Methods	1-4	1-6	1-8	1-12	1-15	1-18	Methods	1-4	1-6	1-8	1-12	1-15	1-18
Ata	11.63	11.73	11.93	12.29	12.90	13.60	Ata	0.59	0.63	0.67	0.73	0.78	0.84
Theta	11.54	11.75	12.09	12.50	13.14	13.89	ForecastPro	0.59	0.63	0.66	0.72	0.78	0.85
ForecastPro	11.75	11.81	12.07	12.48	13.13	13.90	Theta	0.60	0.64	0.68	0.74	0.79	0.86
ETS	11.93	12.05	12.43	12.96	13.64	14.45	ForecastX	0.62	0.65	0.69	0.76	0.82	0.89
Comp S-H-D	12.83	12.74	12.88	13.09	13.73	14.47	Comp S-H-D	0.65	0.68	0.72	0.77	0.83	0.90
ForecastX	12.32	12.28	12.44	12.82	13.61	14.47	Ararma	0.66	0.70	0.74	0.79	0.84	0.91
Dampen	12.63	12.63	12.81	13.08	13.75	14.58	Dampen	0.64	0.68	0.72	0.78	0.84	0.91
RBF	13.49	13.14	13.36	13.64	14.19	14.76	Holt	0.64	0.68	0.72	0.78	0.84	0.91
B-J automatic	12.78	12.70	12.86	13.19	13.95	14.80	RBF	0.73	0.73	0.77	0.81	0.86	0.91
SmartFcs	12.16	12.53	12.85	13.49	14.20	15.01	B-J automatic	0.63	0.67	0.71	0.77	0.84	0.91
Auto-ANN	12.32	12.52	12.93	13.48	14.21	15.03	SmartFcs	0.65	0.68	0.72	0.79	0.85	0.92
Flores/Pearce2	13.26	13.18	13.31	13.52	14.30	15.19	AutoBox1	0.67	0.70	0.74	0.79	0.85	0.92
Single	13.53	13.39	13.56	13.81	14.49	15.30	Auto-ANN	0.63	0.67	0.71	0.79	0.85	0.93
PP-aotocast	13.16	13.22	13.52	13.88	14.53	15.33	Theta-sm	0.69	0.72	0.76	0.82	0.88	0.95
Theta-sm	13.38	13.66	13.95	14.17	14.68	15.38	Flores/Pearce2	0.69	0.72	0.76	0.81	0.87	0.95
AutoBox2	13.54	13.51	13.76	14.17	14.90	15.73	AutoBox3	0.66	0.70	0.75	0.82	0.89	0.96
Holt	13.02	13.18	13.48	14.05	14.87	15.79	Single	0.70	0.73	0.77	0.84	0.90	0.97
AutoBox1	13.27	13.37	13.67	14.07	14.91	15.81	PP-aotocast	0.71	0.75	0.80	0.86	0.92	0.99
Ararma	13.42	13.55	13.96	14.39	15.06	15.83	Flores/Pearce1	0.71	0.77	0.82	0.86	0.92	1.01
Winter	13.26	13.39	13.72	14.23	15.03	15.93	Robust-Trend	0.77	0.80	0.85	0.91	0.97	1.04
Flores/Pearce1	13.74	13.92	14.21	14.28	15.03	15.99	Naive2	0.77	0.80	0.84	0.90	0.97	1.04
AutoBox3	13.13	13.58	14.04	14.65	15.50	16.59	AutoBox2	0.74	0.78	0.82	0.90	0.97	1.08
Naive2	15.30	15.08	15.26	15.55	16.16	16.89	Winter	0.91	0.86	0.86	0.89	1.15	1.17
Robust-Trend	15.49	15.50	16.03	16.84	17.85	18.93	ETS	-	-	-	-	-	0.87

Table 5.7 The accuracy perfomance of 174 other series at different (sMAPE and MASE) metrics

mean sMAPE				mean MASE			
	Averages				Averages		
Methods	1-4	1-6	1-8	Methods	1-4	1-6	1-8
Ata	3.04	3.68	4.14	Ata	1.27	1.52	1.72
Ararma	3.17	3.87	4.38	AutoBox2	1.35	1.63	1.86
Theta	3.20	3.93	4.41	Robust-Trend	1.36	1.65	1.88
AutoBox2	3.19	3.86	4.41	Theta	1.36	1.67	1.90
ETS	3.37	3.99	4.51	ForecastPro	1.36	1.66	1.92
Comp S-H-D	3.36	4.09	4.56	ForecastX	1.44	1.71	1.92
Robust-Trend	13.87	17.03	4.58	AutoBox3	1.42	1.71	1.97
ForecastPro	3.31	4.00	4.60	Winter	1.38	1.71	1.99
Dampen	3.28	4.06	4.61	Holt	1.38	1.71	1.99
PP-aotocast	3.29	4.07	4.62	Ararma	1.43	1.77	2.01
ForecastX	3.42	4.10	4.64	SmartFcs	1.59	1.84	2.03
AutoBox3	3.39	4.09	4.71	Dampen	1.42	1.77	2.04
Auto-ANN	3.26	4.07	4.80	Comp S-H-D	1.48	1.81	2.04
Winter	3.32	4.13	4.81	PP-aotocast	1.44	1.78	2.05
Holt	3.32	4.13	4.81	AutoBox1	1.61	1.86	2.08
SmartFcs	3.68	4.33	4.86	Auto-ANN	1.39	1.74	2.08
Flores/Pearce2	3.67	4.43	4.89	Theta-sm	1.57	1.93	2.20
Theta-sm	3.66	4.37	4.93	Flores/Pearce1	1.63	1.95	2.23
AutoBox1	3.76	4.38	4.93	B-J automatic	1.54	1.94	2.26
B-J automatic	3.52	4.38	5.06	Flores/Pearce2	1.74	2.06	2.29
Flores/Pearce1	3.71	4.47	5.09	RBF	2.06	2.41	2.66
RBF	4.38	5.12	5.60	Naive2	2.07	2.64	3.09
Single	4.36	5.48	6.29	Single	2.07	2.64	3.09
Naive2	4.38	5.49	6.30	ETS	-	-	1.79

Table 5.8 Mean sMAPE by time interval between successive observations

	Time interval between successive observation				
Methods	Monthly	Other	Quartly	Yearly	All
Ata	13.598	4.139	8.888	15.601	12.682
Theta	13.892	4.410	8.956	16.974	13.051
ForecastPro	13.898	4.604	9.815	17.271	13.234
ForecastX	14.466	4.638	9.537	16.480	13.502
Comp S-H-D	14.466	4.561	9.216	17.072	13.508
Dampen	14.576	4.609	9.361	17.360	13.640
RBF	14.760	5.598	9.565	16.424	13.740
ETS	14.450	4.510	9.940	18.530	13.770
B-J automatic	14.796	5.062	10.260	17.726	13.995
SmartFcs	15.007	4.860	10.153	17.706	14.115
PP-aotocast	15.328	4.617	9.395	17.128	14.144
Flores/Pearce2	15.186	4.893	10.431	17.843	14.299
Single	15.300	6.295	9.717	17.817	14.313
Theta-sm	15.380	4.927	9.821	17.922	14.344
AutoBox2	15.731	4.414	10.004	16.593	14.460
Flores/Pearce1	15.986	5.087	9.954	17.205	14.718
Ararma	15.826	4.383	10.186	18.356	14.738
Holt	15.795	4.811	10.938	20.021	15.030
Winter	15.926	4.811	10.840	20.021	15.105
AutoBox1	15.811	4.935	10.961	21.588	15.214
Naive2	16.891	6.302	9.951	17.880	15.462
AutoBox3	16.590	4.713	11.192	20.877	15.710
Auto-ANN	15.031	4.800	10.199	18.565	14.226
Robust-Trend	18.931	4.578	9.789	17.033	16.699

Table 5.9 Mean MASE by time interval between successive observations

	Time interval between successive observation				
Methods	Monthly	Other	Quartly	Yearly	All
Ata	0.841	1.717	1.072	2.550	1.098
Theta	0.858	1.904	1.087	2.806	1.138
ETS	0.870	1.790	1.180	2.830	1.160
ForecastX	0.894	1.925	1.155	2.769	1.172
ForecastPro	0.848	1.920	1.204	3.026	1.174
Comp S-H-D	0.896	2.045	1.105	2.876	1.180
RBF	0.910	2.658	1.173	2.720	1.208
Dampen	0.908	2.036	1.126	3.032	1.208
SmartFcs	0.919	2.034	1.226	2.996	1.229
Holt	0.909	1.993	1.225	3.182	1.239
Auto-ANN	0.926	2.083	1.241	3.033	1.241
B-J automatic	0.914	2.259	1.188	3.165	1.245
Robust-Trend	1.036	1.877	1.152	2.623	1.253
Theta-sm	0.950	2.204	1.211	3.006	1.255
Ararma	0.907	2.008	1.178	3.481	1.262
Flores/Pearce2	0.950	2.295	1.255	3.016	1.267
PP-aotocast	0.994	2.048	1.128	3.016	1.267
AutoBox3	0.962	1.969	1.272	3.177	1.282
Flores/Pearce1	1.008	2.226	1.184	2.938	1.284
AutoBox2	1.082	1.860	1.185	2.754	1.303
AutoBox1	0.924	2.082	1.331	3.679	1.322
Single	0.974	3.091	1.229	3.171	1.325
Naive2	1.037	3.089	1.238	3.172	1.370
Winter	1.165	1.993	1.217	3.182	1.416

CHAPTER SIX
SUMMARY AND CONCLUSION

Many forecasting methods have been proposed and studied in the literature. None of these proposed methods is consistently the best according to all different forecasting criterion. The relative ranking of the performance of the different forecasting methods varies according to the accuracy measure being used and length of the forecasting horizon involved. An obvious fact accepted by forecasters is statistically sophisticated or complex methods do not necessarily provide more accurate forecasts than simpler ones and the accuracy when various methods are being combined outperforms, on average, the individual methods being combined. Exponential smoothing methods are the most widely used and well-known forecasting method in the literature since they are very accurate, simple and easy to manage but the basic reason for this popularity is their proven success against other more complex methods.

Despite the fact that these models are well established, they still have some shortcomings that affect the quality of forecasts they produce. We propose a family of models which are given the name ATA method, that help overcome these shortcomings especially related to optimization and initialization, since there is no consensus among forecasters on the value of smoothing parameters and selection of initial values. The formulation of ATA is similar to ES but the weights attached to observations are different and these variations help eliminate these shortcomings.

In this paper, the two approaches, exponential smoothing and ATA method, have been compared based on popular metrics that are commonly used while comparing forecasting methods. While some of these comparisons are based on the simple versions of the two approaches, the generality of these results hold for models with more components. Empirical performances of both approaches on the M3-competition data sets have been given and it was verified that ATA has better forecasting accuracy than all other competitors in addition to the fact that it is a simple and interpretable model. Contrary to expectations, the relative ranking of the performance of the ATA method didn't change according to the accuracy measure being used and length of the forecasting horizon involved. In short, best should be best in every forecasting setting, and ATA method accomplishes this.

We have also proposed a very simple but fast automatic forecasting strategy based on the model framework and this strategy is also applied to the M3 competition data set and performed even better. We note that we have not done any preprocessing of the data, identification of outliers or level shifts, or used any other strategy and transformations to improve the forecasts. These results are based on a simple application of the algorithm to the data. Of course, the performance of ATA method could be improved further if we used some sophisticated data preprocessing techniques as was done by some of the competitors in the M3 competition. The fact that even the simple version of ATA and the straight forward algorithm for combinations and selections we proposed perform so well is very promising and the results given in these thesis is not the final performance of the ATA method since the method can be improved by allowing for more sophisticated model selection strategies and more involved combinations of forecasts.

REFERENCES

Adya, M., Armstrong, J. S., Collopy, F., & Kennedy, M. (2000). An application of rule-based forecasting to a situation lacking domain knowledge. *International Journal of Forecasting, 16*(4), 477--484.

Armstrong, J. S. (1989). Combining forecasts: The end of the beginning or the beginning of the end? *International Journal of Forecasting, 5*(4), 585--588.

Armstrong, J. S. (2001). *Principles of forecasting: a handbook for researchers and practitioners*, volume 30. Springer Science & Business Media.

Armstrong, S. J., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: empirical comparisons (with discussion). *International Journal of Forecasting, 8*(1), 69--80.

Assimakopoulos, V., & Nikolopoulos, K. (2000). The theta model: a decomposition approach to forecasting. *International Journal of Forecasting, 16*(4), 521--530.

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

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society, 20*(4), 451--468.

Bergmeir, C., Hyndman, R. J., & Benítez, J. M. (2016). Bagging exponential smoothing methods using stl decomposition and box--cox transformation. *International Journal of Forecasting, 32*(2), 303--312.

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

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

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

Chatfield, C., & Yar, M. (1988). Holt-winters forecasting: some practical issues. *The Statistician*, 129--140.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*(4), 559--583.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, *6*(1), 3--73.

De Gooijer, J. G., & Hyndman, R. J. (2005). 25 years of iif time series forecasting: a selective review. *Tinbergen Institute Discussion Papers No. TI*, 05--068.

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

Gardner, E. S. (2006). Exponential smoothing: The state of the art—part ii. *International Journal of Forecasting*, *22*(4), 637--666.

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

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

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

Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Springer-Verlag.

Hyndman, R. J. (2003). *International Journal of Forecasting*, *19*(2), 287--290.

Hyndman, R. J., & Athanasopoulos, G. (2014). *Forecasting: principles and practice*. OTexts.

Hyndman, R. J., & Billah, B. (2003). Unmasking the theta method. *International Journal of Forecasting*, *19*(2), 287--290.

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

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

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

Koning, A. J., Franses, P. H., Hibon, M., & Stekler, H. O. (2005). The m3 competition: Statistical tests of the results. *International Journal of Forecasting, 21*(3), 397--409.

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

Makridakis, S., & Hibon, M. (1979). Accuracy of forecasting: an empirical investigation. *Journal of Royal Statistical Society, 142*(2), 97--145.

Makridakis, S., & Hibon, M. (1991). Exponential smoothing: The effect of initial values and loss functions on post-sample forecasting accuracy. *International Journal of Forecasting, 7*(3), 317--330.

Makridakis, S., & Hibon, M. (2000). The m3-competition: results, conclusions and implications. *International Journal of Forecasting, 16*(4), 451--476.

Makridakis, S., & Winkler, R. L. (1983). Averages of forecasts: Some empirical results. *Management Science, 29*(9), 987--996.

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

Makridakis, S. G., Hibon, M., Chatfield, S., Lawrence, M., Mills, T., Ord, K., & Simmons, L. F. (1993). The m-2 competition: a real time judgementally based forecasting study. *International Journal of Forecasting, 9*(2), 5--23.

Makridakis, S. G., & Wheelwright, S. C. (1978). *Forecasting: Methods and Applications*. New Jersey: Wiley.

Makridakis S., C. S, W., & Hyndman, J. R. (1998). *Forecasting: Methods and Application*, volume 3. New Jersey;Willey.

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

Newbold, P., & Granger, C. W. (1974). Experience with forecasting univariate time series and the combination of forecast. *Journal of Royal Statistical Society*, *137*(2), 131--165.

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

Reid, D. J. (1969). A comparative study of time series prediction techniques on economic data. *International Journal of Forecasting*, *19*(2), 303--308.

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

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

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

Yapar, G. (2016). Modified simple exponential smoothing. *Hacettepe University Journal of Mathematics and Statistics*, *Early access*, 0.

APPENDICES

A.1: R codes

APPENDICES ONE

APPENDIX

R code for $ATA(p,q)$

```
library(xlsx)
library(matrixcalc)
library(biganalytics)
library(matrixStats)
data=read.xlsx("d:/m3seasonallyadjusteddata.xlsx",1)
data=data.frame(data)
n=matrix(0,nrow=3003, ncol=1)
Sonuc<-array(NA,c(4,3003,127))
tsonforecast<-sontrend<-matrix(0,nrow=1, ncol=3003)
minp1<-tminsMAPE<-matrix(0,nrow=1, ncol=3003)

for (k in 1:3003)
m3003=data[k,]
m3003=m3003[m3003!=9999999999]
m3003=ts(m3003)
lengths(m3003)
n[k,]<-sum(lengths(m3003))
m=n[k]
p=m-1

for (y in 0:p){
q=y+1
tsMAPE<-tforecast<-trend<-matrix(NA,nrow=m, ncol=3003)
tfit<-tlevel<-ttrend<-terror<-matrix(NA,nrow=m, ncol=m)
for (x in q:m){
for (i in 1:m){
```

```
if (i<=y) {
if (i==1){
tlevel[x,i]=(m3003[i])
ttrend[x,i]=0
tfit[x,i]<-(tlevel[x,i]+ttrend[x,i])
}
else{
tlevel[x,i]=(m3003[i])
ttrend[x,i]=(m3003[i]-m3003[i-1])
tfit[x,i]<-(tlevel[x,i]+ttrend[x,i])
}}
else if ((y<i) & (i<=x)){
if (y==0){
tlevel[x,i]=(m3003[i])
ttrend[x,i]=0
tfit[x,i]=m3003[i]+0
}
else {
tlevel[x,i]=(m3003[i])
ttrend[x,i]<-((y/i)*(tlevel[x,i]-tlevel[x,i-1]))+((1-(y/i))*(ttrend[x,i-1]))
tfit[x,i]<-(tlevel[x,i]+ttrend[x,i])
}
}
else if ((y<i) & (i>x))
tlevel[x,i]<-((x/i)*m3003[i])+((1-(x/i))*(tlevel[x,i-1]+ttrend[x,i-1]))
ttrend[x,i]<-((y/i)*(tlevel[x,i]-tlevel[x,i-1]))+((1-(y/i))*(ttrend[x,i-1]))
tfit[x,i]<-(tlevel[x,i]+ttrend[x,i])
}

terror[x,i]<-(abs(m3003[i+1]-tfit[x,i])/(abs(m3003[i+1]+tfit[x,i])/2))*100
tforecast[,k]=tlevel[,m]
```

```
trend[,k]=ttrend[,m]
tsMAPE[,k]<-rowMeans(terror, na.rm = TRUE)
tminsMAPE[,k]<- min(tsMAPE[,k], na.rm = TRUE)
minp1[,k]<- max(which(tsMAPE[,k]==tminsMAPE[,k]))
tsonforecast[,k]<-tforecast[minp1[,k],k]
sontrend[,k]<-trend[minp1[,k],k]

Sonuc[1,k,y+1]=minp1[,k]
Sonuc[2,k,y+1]=tminsMAPE[,k]
Sonuc[3,k,y+1]=tsonforecast[,k]
Sonuc[4,k,y+1]=sontrend[,k]

   }}}

   }
proc.time()
library(xlsx)
write.xlsx(Sonuc[1,,], "d:/ATA(p).xlsx")
library(xlsx)
write.xlsx(Sonuc[1,,], "d:/ATA(sMAPE).xlsx")
library(xlsx)
write.xlsx(Sonuc[1,,], "d:/ATA(forecast).xlsx")
library(xlsx)
write.xlsx(Sonuc[1,,], "d:/ATA(trend).xlsx")
```

R code for $ATA_{mult}(p, q)$

```
library(xlsx)
library(matrixcalc)
library(biganalytics)
library(matrixStats)
```

```
data=read.xlsx("d:/m3seasonallyadjusteddata.xlsx",1)
data=data.frame(data)
n=matrix(0,nrow=3003, ncol=1)
Sonuc<-array(NA,c(4,3003,127))
tsonforecast<-sontrend<-matrix(0,nrow=1, ncol=3003)
minp1<-tminsMAPE<-matrix(0,nrow=1, ncol=3003)

for (k in 1:3003)
m3003=data[k,]
m3003=m3003[m3003!=9999999999]
m3003=ts(m3003)
lengths(m3003)
n[k,]<-sum(lengths(m3003))
m=n[k]
p=m-1

for (y in 0:p){
q=y+1
tsMAPE<-tforecast<-trend<-matrix(NA,nrow=m, ncol=3003)
tfit<-tlevel<-ttrend<-terror<-matrix(NA,nrow=m, ncol=m)
for (x in q:m){
for (i in 1:m){
if (i<=y) {
if (i==1){
tlevel[x,i]=(m3003[i])
ttrend[x,i]=1
tfit[x,i]<-(tlevel[x,i]*ttrend[x,i])
}
else{
tlevel[x,i]=(m3003[i])
ttrend[x,i]=(m3003[i]/m3003[i-1])
```

```
tfit[x,i]<-(tlevel[x,i]*ttrend[x,i])
}}
else if ((y<i) & (i<=x)){
if (y==0){
tlevel[x,i]=(m3003[i])
ttrend[x,i]=1
tfit[x,i]=m3003[i]*0
}
else {
tlevel[x,i]=(m3003[i])
ttrend[x,i]<-((y/i)*(tlevel[x,i]/tlevel[x,i-1]))+((1-(y/i))*(ttrend[x,i-1]))
tfit[x,i]<-(tlevel[x,i]*ttrend[x,i])
}
}
else if ((y<i) & (i>x))
tlevel[x,i]<-((x/i)*m3003[i])+((1-(x/i))*(tlevel[x,i-1]*ttrend[x,i-1]))
ttrend[x,i]<-((y/i)*(tlevel[x,i]/tlevel[x,i-1]))+((1-(y/i))*(ttrend[x,i-1]))
tfit[x,i]<-(tlevel[x,i]*ttrend[x,i])
}

terror[x,i]<-(abs(m3003[i+1]-tfit[x,i])/(abs(m3003[i+1]+tfit[x,i])/2))*100
tforecast[,k]=tlevel[,m]
trend[,k]=ttrend[,m]
tsMAPE[,k]<-rowMeans(terror, na.rm = TRUE)
tminsMAPE[,k]<- min(tsMAPE[,k], na.rm = TRUE)
minp1[,k]<- max(which(tsMAPE[,k]==tminsMAPE[,k]))
tsonforecast[,k]<-tforecast[minp1[,k],k]
sontrend[,k]<-trend[minp1[,k],k]

Sonuc[1,k,y+1]=minp1[,k]
Sonuc[2,k,y+1]=tminsMAPE[,k]
```

```
Sonuc[3,k,y+1]=tsonforecast[,k]
Sonuc[4,k,y+1]=sontrend[,k]

  }}}

  }
proc.time()
library(xlsx)
write.xlsx(Sonuc[1,,], "d:/ATAmult(p).xlsx")
library(xlsx)
write.xlsx(Sonuc[1,,], "d:/ATAmult(sMAPE).xlsx")
library(xlsx)
write.xlsx(Sonuc[1,,], "d:/ATAmult(forecast).xlsx")
library(xlsx)
write.xlsx(Sonuc[1,,], "d:/ATAmult(trend).xlsx")
```

R code for $ATA(p, q = constant)$

```
library(xlsx)
library(matrixcalc)
library(biganalytics)
library(matrixStats)

data=read.xlsx("d:/m3seasonallyadjusteddata.xlsx",1)
data=data.frame(data)
n=matrix(0,nrow=3003, ncol=1)
Sonuc<-matrix(1,nrow=7, ncol=3003)
sonforecast<-tsonforecast<-sontrend<-matrix(NA,nrow=1, ncol=3003)
minp1<-minsMAPE<-minq2<-tminsMAPE<-matrix(NA,nrow=1, ncol=3003)
for (k in 1:3003){
m3003=data[k,]
```

```
m3003=m3003[m3003!=9999999999]
m3003=ts(m3003)
lengths(m3003)
n[k,]<-sum(lengths(m3003))
m=n[k]
a=1
tsMAPE<-tforecast<-trend<-matrix(NA,nrow=m, ncol=3003)
tfit<-tlevel<-ttrend<-terror<-te1<-matrix(NA,nrow=m, ncol=m)
for (y in a:m){
for (i in 1:m){
if (i<=a) {
if (i==1){
tlevel[y,i]=(m3003[i])
ttrend[y,i]=0
tfit[y,i]<-(tlevel[y,i]+ttrend[y,i])
}
else{
tlevel[y,i]=(m3003[i])
ttrend[y,i]=(m3003[i]-m3003[i-1])
tfit[y,i]<-(tlevel[y,i]+ttrend[y,i])
}}
else if ((a<i) & (i<=y)){
if (a==0){
tlevel[y,i]=(m3003[i])
ttrend[y,i]=0
tfit[y,i]=m3003[i]+0
}
else {
tlevel[y,i]=(m3003[i])
ttrend[y,i]<-((a/i)*(tlevel[y,i]-tlevel[y,i-1]))+((1-(a/i))*(ttrend[y,i-1]))
tfit[y,i]<-(tlevel[y,i]+ttrend[y,i])
```

```
}


}
else if ((a<i) & (i>y)){
tlevel[y,i]<-((y/i)*m3003[i])+((1-(y/i))*(tlevel[y,i-1]+ttrend[y,i-1]))
ttrend[y,i]<-((a/i)*(tlevel[y,i]-tlevel[y,i-1]))+((1-(a/i))*(ttrend[y,i-1]))
tfit[y,i]<-(tlevel[y,i]+ttrend[y,i])
}

terror[y,i]<-(abs(m3003[i+1]-tfit[y,i])/(abs(m3003[i+1]+tfit[y,i])/2))*100
tforecast[,k]=tlevel[,m]
trend[,k]=ttrend[,m]
tsMAPE[,k]<-rowMeans(terror, na.rm = TRUE)
tminsMAPE[,k]<- min(tsMAPE[,k], na.rm = TRUE)
minq2[,k]<- max(which(tsMAPE[,k]==tminsMAPE[,k]))
tsonforecast[,k]<-tforecast[minq2[,k],k]
sontrend[,k]<-trend[minq2[,k],k]

Sonuc[1,k]=minp1[,k]
Sonuc[2,k]=minsMAPE[,k]
Sonuc[3,k]=sonforecast[,k]
Sonuc[4,k]=minq2[,k]
Sonuc[5,k]=tminsMAPE[,k]
Sonuc[6,k]=tsonforecast[,k]
Sonuc[7,k]=sontrend[,k]


   }}

   }
proc.time()
library(xlsx)
```

```
write.xlsx(Sonuc, "d:/ATA(p=opt,q=sbt)xlsx")
```

R code for $ATA_{mult}(p, q = constant)$

```
library(xlsx)
library(matrixcalc)
library(biganalytics)
library(matrixStats)

data=read.xlsx("d:/m3seasonallyadjusteddata.xlsx",1)
data=data.frame(data)
n=matrix(0,nrow=3003, ncol=1)
Sonuc<-matrix(1,nrow=7, ncol=3003)
sonforecast<-tsonforecast<-sontrend<-matrix(NA,nrow=1, ncol=3003)
minp1<-minsMAPE<-minq2<-tminsMAPE<-matrix(NA,nrow=1, ncol=3003)
for (k in 1:3003){
m3003=data[k,]
m3003=m3003[m3003!=9999999999]
m3003=ts(m3003)
lengths(m3003)
n[k,]<-sum(lengths(m3003))
m=n[k]
a=1
tsMAPE<-tforecast<-trend<-matrix(NA,nrow=m, ncol=3003)
tfit<-tlevel<-ttrend<-terror<-te1<-matrix(NA,nrow=m, ncol=m)
for (y in a:m){
for (i in 1:m){
if (i<=a) {
if (i==1){
tlevel[y,i]=(m3003[i])
```

```
ttrend[y,i]=1
tfit[y,i]<-(tlevel[y,i]*ttrend[y,i])
}
else{
tlevel[y,i]=(m3003[i])
ttrend[y,i]=(m3003[i]/m3003[i-1])
tfit[y,i]<-(tlevel[y,i]*ttrend[y,i])
}}
else if ((a<i) & (i<=y)){
if (a==0){
tlevel[y,i]=(m3003[i])
ttrend[y,i]=1
tfit[y,i]=m3003[i]*0
}
else {
tlevel[y,i]=(m3003[i])
ttrend[y,i]<-((a/i)*(tlevel[y,i]/tlevel[y,i-1]))+((1-(a/i))*(ttrend[y,i-1]))
tfit[y,i]<-(tlevel[y,i]*ttrend[y,i])
}

}
else if ((a<i) & (i>y)){
tlevel[y,i]<-((y/i)*m3003[i])+((1-(y/i))*(tlevel[y,i-1]*ttrend[y,i-1]))
ttrend[y,i]<-((a/i)*(tlevel[y,i]/tlevel[y,i-1]))+((1-(a/i))*(ttrend[y,i-1]))
tfit[y,i]<-(tlevel[y,i]*ttrend[y,i])
}

terror[y,i]<-(abs(m3003[i+1]-tfit[y,i])/(abs(m3003[i+1]+tfit[y,i])/2))*100
tforecast[,k]=tlevel[,m]
trend[,k]=ttrend[,m]
tsMAPE[,k]<-rowMeans(terror, na.rm = TRUE)
```

```
tminsMAPE[,k]<- min(tsMAPE[,k], na.rm = TRUE)
minq2[,k]<- max(which(tsMAPE[,k]==tminsMAPE[,k]))
tsonforecast[,k]<-tforecast[minq2[,k],k]
sontrend[,k]<-trend[minq2[,k],k]

Sonuc[1,k]=minp1[,k]
Sonuc[2,k]=minsMAPE[,k]
Sonuc[3,k]=sonforecast[,k]
Sonuc[4,k]=minq2[,k]
Sonuc[5,k]=tminsMAPE[,k]
Sonuc[6,k]=tsonforecast[,k]
Sonuc[7,k]=sontrend[,k]


  }}

  }
proc.time()
library(xlsx)
write.xlsx(Sonuc, "d:/ATAmult(p=opt,q=sbt)xlsx")
```

R code for $ATA_{damped}(p, q = constant)$

```
  library(xlsx)
library(matrixcalc)
library(biganalytics)
library(matrixStats)

data=read.xlsx("d:/m3seasonallyadjusteddata.xlsx",1)
data=data.frame(data)
data1=read.xlsx("d:/s.index3003.xlsx",1)
```

```
data1=data.frame(data1)
n=matrix(NA,nrow=3003, ncol=1)
n1=matrix(NA,nrow=19, ncol=3003)
Sonuc<-array(NA,c(23,3003,9))
sonforecast<-tsonforecast<-sontrend<-matrix(NA,nrow=1, ncol=3003)
minp1<-minsMAPE<-minq2<-tminsMAPE<-matrix(NA,nrow=1, ncol=3003)

for (k in 1:3003){
m3003=data[k,]
m3003=m3003[m3003!=9999999999]
m3003=ts(m3003)
lengths(m3003)
n[k,]<-sum(lengths(m3003))
m=n[k]
n1=data1
a=1
d1=0.80
d=1
tsMAPE<-tforecast<-trend<-matrix(NA,nrow=m, ncol=3003)
tfit<-tlevel<-ttrend<-terror<-te1<-matrix(NA,nrow=m, ncol=m)
for (y in a:m){
for (i in 1:m){
if (i<=a) {
if (i==1){
tlevel[y,i]=(m3003[i])
ttrend[y,i]=0
tfit[y,i]<-(tlevel[y,i]+ttrend[y,i]*d1)
}
else{
tlevel[y,i]=(m3003[i])
ttrend[y,i]=(m3003[i]-m3003[i-1])
```

```
tfit[y,i]<-(tlevel[y,i]+ttrend[y,i]*d1)
}}
else if ((a<i) & (i<=y)){
if (a==0){
tlevel[y,i]=(m3003[i])
ttrend[y,i]=0
tfit[y,i]=m3003[i]+0*d1
}
else {
tlevel[y,i]=(m3003[i])
ttrend[y,i]<-((a/i)*(tlevel[y,i]-tlevel[y,i-1]))+((1-(a/i))*(ttrend[y,i-1])*d1)
tfit[y,i]<-(tlevel[y,i]+ttrend[y,i]*d1)
}
}
else if ((a<i) & (i>y)){
tlevel[y,i]<-((y/i)*m3003[i])+((1-(y/i))*(tlevel[y,i-1]+ttrend[y,i-1]*d1))
ttrend[y,i]<-((a/i)*(tlevel[y,i]-tlevel[y,i-1]))+((1-(a/i))*(ttrend[y,i-1]*d1))
tfit[y,i]<-(tlevel[y,i]+ttrend[y,i]*d1)
}
terror[y,i]<-(abs(m3003[i+1]-tfit[y,i])/(abs(m3003[i+1]+tfit[y,i])/2))*100
tforecast[,k]=tlevel[,m]
trend[,k]=ttrend[,m]
tsMAPE[,k]<-rowMeans(terror, na.rm = TRUE)
tminsMAPE[,k]<- min(tsMAPE[,k], na.rm = TRUE)
minq2[,k]<- max(which(tsMAPE[,k]==tminsMAPE[,k]))
tsonforecast[,k]<-tforecast[minq2[,k],k]
sontrend[,k]<-trend[minq2[,k],k]

Sonuc[1,k,d]=minp1[,k]
Sonuc[2,k,d]=minq2[,k]
Sonuc[3,k,d]=tminsMAPE[,k]
```

```
Sonuc[4,k,d]=tsonforecast[,k]
Sonuc[5,k,d]=sontrend[,k]

if (n1[1,k]==6){
Sonuc[6,k,d]=(tsonforecast[1,k]+sontrend[1,k]*d1)*n1[2,k]
Sonuc[7,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12))*n1[3,k]
Sonuc[8,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13))*n1[4,k]
Sonuc[9,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14))*n1[5,k]
Sonuc[10,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15))
*n1[6,k]
Sonuc[11,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15+d16))
*n1[7,k] }

else if (n1[1,k]==8){
Sonuc[6,k,d]=(tsonforecast[1,k]+sontrend[1,k]*d1)*n1[2,k]
Sonuc[7,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12))*n1[3,k]
Sonuc[8,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13))*n1[4,k]
Sonuc[9,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14))*n1[5,k]
Sonuc[10,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15))
*n1[6,k]
Sonuc[11,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15+d16))
*n1[7,k]
Sonuc[12,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15
+d16+d17))*n1[8,k]
Sonuc[13,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15
+d16+d17+d18))*n1[9,k]}

else if (n1[1,k]==18){
Sonuc[6,k,d]=(tsonforecast[1,k]+sontrend[1,k]*d1)*n1[2,k]
Sonuc[7,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12))*n1[3,k]
Sonuc[8,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13))*n1[4,k]
```

```
Sonuc[9,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14))*n1[5,k]
Sonuc[10,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15))
*n1[6,k]
Sonuc[11,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15+d16))
*n1[7,k]
Sonuc[12,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15+
d16+d17))*n1[8,k]
Sonuc[13,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15+
d16+d17+d18))*n1[9,k]
Sonuc[14,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15+
d16+d17+d18+d19))*n1[10,k]
Sonuc[15,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15+
d16+d17+d18+d19+d110))*n1[11,k]
Sonuc[16,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15+
d16+d17+d18+d19+d110+d111))*n1[12,k]
Sonuc[17,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15+
d16+d17+d18+d19+d110+d111+d112))*n1[13,k]
Sonuc[18,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15+
d16+d17+d18+d19+d110+d111+d112+d113))*n1[14,k]
Sonuc[19,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15+
d16+d17+d18+d19+d110+d111+d112+d113+d114))*n1[15,k]
Sonuc[20,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15+
d16+d17+d18+d19+d110+d111+d112+d113+d114+d115))*n1[16,k]
Sonuc[21,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15+
d16+d17+d18+d19+d110+d111+d112+d113+d114+d115+d116))*n1[17,k]
Sonuc[22,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15+
d16+d17+d18+d19+d110+d111+d112+d113+d114+d115+d116+d117))*n1[18,k]
Sonuc[23,k,d]=(tsonforecast[1,k]+sontrend[1,k]*(d1+d12+d13+d14+d15+
d16+d17+d18+d19+d110+d111+d112+d113+d114+d115+d116+
d117+d118))*n1[19,k]}
```

```
}}

}
proc.time()
library(xlsx)
write.xlsx(Sonuc[,,1], "d:/adddamped(ATA(p=opt,q=sbt,0.80)).xlsx")
```

R code for $ATA(p^*, q)$

```
library(xlsx)
library(matrixcalc)
library(biganalytics)
library(matrixStats)

data=read.xlsx("d:/m3seasonallyadjusteddata.xlsx",1)
data=data.frame(data)
n=matrix(0,nrow=3003, ncol=1)
Sonuc<-matrix(1,nrow=7, ncol=3003)
sonforecast<-tsonforecast<-sontrend<-matrix(0,nrow=1, ncol=3003)
minpstar<-minsMAPE<-minq2<-tminsMAPE<-matrix(0,nrow=1, ncol=3003)

for (k in 1:3003){
m3003=data[k,]
m3003=m3003[m3003!=9999999999]
m3003=ts(m3003)
lengths(m3003)
n[k,]<-sum(lengths(m3003))
m=n[k]
ssMAPE<-sforecast<-matrix(0,nrow=m, ncol=3003)
sfit<-slevel<-serror<-se1<-matrix(0,nrow=m, ncol=m)
```

```
for (x in 1:m){
for (i in 1:m){
if (i<=x) {
slevel[x,i]=(m3003[i])
sfit[x,i]=slevel[x,i]
}
else {
slevel[x,i]<-((x/i)*m3003[i])+((1-(x/i))*(slevel[x,i-1]))
sfit[x,i]=slevel[x,i]
}

serror[x,i]<-abs(m3003[i+1]-sfit[x,i])/(abs(m3003[i+1]+sfit[x,i])/2))*100
sforecast[,k]=slevel[,m]
ssMAPE[,k]<-rowMeans(serror, na.rm=TRUE)
minsMAPE[,k]<- min(ssMAPE[,k], na.rm=TRUE)
minpstar[,k]<- max(which(ssMAPE[,k]==minsMAPE[,k]))
sonforecast[,k]<-sforecast[minpstar[,k],k]


   }}

   p1=minpstar[,k]
a=p1
tsMAPE<-tforecast<-trend<-matrix(0,nrow=a+1, ncol=3003)
tfit<-tlevel<-ttrend<-terror<-te1<-matrix(0,nrow=a+1, ncol=m)

for (y in 0:a){
for (i in 1:m){
if (i<=y) {
if (i==1){
tlevel[y+1,i]=(m3003[i])
```

```
ttrend[y+1,i]=0
tfit[y+1,i]<-(tlevel[y+1,i]+ttrend[y+1,i])
}
else{
tlevel[y+1,i]=(m3003[i])
ttrend[y+1,i]=(m3003[i]-m3003[i-1])
tfit[y+1,i]<-(tlevel[y+1,i]+ttrend[y+1,i])
}}
else if ((y<i) & (i<=a)){
if (y==0){
tlevel[y+1,i]=(m3003[i])
ttrend[y+1,i]=0
tfit[y+1,i]=m3003[i]+0
}
else {
tlevel[y+1,i]=(m3003[i])
ttrend[y+1,i]<-((y/i)*(tlevel[y+1,i]-tlevel[y+1,i-1]))+((1-(y/i))*(ttrend[y+1,i-1]))
tfit[y+1,i]<-(tlevel[y+1,i]+ttrend[y+1,i])
}
}
else if ((y<i) & (i>a)){
tlevel[y+1,i]<-((a/i)*m3003[i])+((1-(a/i))*(tlevel[y+1,i-1]+ttrend[y+1,i-1]))
ttrend[y+1,i]<-((y/i)*(tlevel[y+1,i]-tlevel[y+1,i-1]))+((1-(y/i))*(ttrend[y+1,i-1]))
tfit[y+1,i]<-(tlevel[y+1,i]+ttrend[y+1,i])
}

terror[y+1,i]<-abs(m3003[i+1]-tfit[y,i])/(abs(m3003[i+1]+tfit[y,i])/2))*100
tforecast[,k]=tlevel[,m]
trend[,k]=ttrend[,m]
tsMAPE[,k]<-rowMeans(terror, na.rm = TRUE)
tminsMAPE[,k]<- min(tsMAPE[,k], na.rm = TRUE)
```

```
minq2[,k]<- max(which(tsMAPE[,k]==tminsMAPE[,k]))
tsonforecast[,k]<-tforecast[minq2[,k],k]
sontrend[,k]<-trend[minq2[,k],k]

Sonuc[1,k]=minpstar[,k]
Sonuc[2,k]=minsMAPE[,k]
Sonuc[3,k]=sonforecast[,k]
Sonuc[4,k]=minq2[,k]-1
Sonuc[5,k]=tminsMAPE[,k]
Sonuc[6,k]=tsonforecast[,k]
Sonuc[7,k]=sontrend[,k]


   }}

   }
proc.time()
library(xlsx)
write.xlsx(Sonuc, "d:/ATA(pstar,q).xlsx")
```