DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

EMPIRICAL INVESTIGATION OF THE PROPERTIES OF ATA FORECASTING METHOD

by

Beyza ÇETİN

January, 2021
İZMİR

EMPIRICAL INVESTIGATION OF THE PROPERTIES OF ATA FORECASTING METHOD

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

by

Beyza ÇETİN

January, 2021
İZMİR

M.Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled "**EMPIRICAL INVESTIGATION OF THE PROPERTIES OF ATA FORECASTING METHOD**" completed by **BEYZA ÇETİN** under supervision of **ASST. PROF. DR. İDİL YAVUZ** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

....

Asst. Prof. Dr. İdil YAVUZ

Supervisor

....

Prof. Dr. Çağdaş Hakan ALADAĞ

Jury Member

....

Prof. Dr. Güçkan YAPAR

Jury Member

Prof. Dr. Özgür ÖZÇELİK
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

Throughout my master's thesis, I would like to express my deepest gratitude to my supervisor Asst. Prof. Dr. İdil YAVUZ for her continuous support, selfless guidance, generous advice and helpful comments. In particular, I would like to thank to my supervisor for valuable insights, patience and incentives during difficult times of my research.

Besides to my advisor, I would like to thank to Prof. Dr. Güçkan YAPAR for all contributions. I am so grateful for his endless trust on me.

In addition, this thesis was supported by the project number 118F174 named as "Developing a New Estimation Method for Time Series Data", I would like to thank to TÜBİTAK for their contributions.

Special thanks to my fiance and my family for their kind patience and support. I am so grateful to them for their support.

Beyza ÇETİN

EMPIRICAL INVESTIGATION OF THE PROPERTIES OF ATA FORECASTING METHOD

ABSTRACT

Forecasting is important in all scientific fields such as industrial, commercial, medical and economic. There are many forecasting methods in the literature, but exponential smoothing is a very popular method due to its simplicity and accuracy. Simple exponential smoothing is used for data sets randomly distributed around a constant level. Holt's linear trend method is a method that helps to deal with linearly trended data. Despite the fact that exponential smoothing methods are widely used and have been in the literature for a long time, they have some problems that potentially affect the predictive accuracy of models. Ata is a new forecasting method that has been proposed to overcome these problems and to provide better forecasts. In this thesis, the forecasting accuracy of Ata and exponential smoothing will be compared for data sets with no trend or linear trend and damped trend. The results given in this study are obtained using simulated data sets with different sample sizes and variances and the forecast accuracy is compared using the mean squared forecast error. In line with these results, the forecast accuracy is calculated for both short and long term forecasting horizons. The results reveal that the Ata method outperforms exponential smoothing for most types of time series data for both short and long term forecasting horizons.

Keywords: Ata method, Exponential smoothing, Forecasting, Simulation, Smoothing parameter, Time series

ATA ÖNGÖRÜ YÖNTEMİNİN AMPİRİK ÖZELLİKLERİNİN İNCELENMESİ

ÖZ

Öngörü, endüstri, ticaret, sağlık ve ekonomi gibi tüm bilimsel alanlarda önemlidir. Literatürde birçok öngörü yöntemi vardır, ancak üstel düzleştirme, sadeliği ve doğruluğu nedeniyle çok popüler bir yöntemdir. Basit üstel düzleştirme, sabit bir seviye etrafında rastgele dağıtılmış veri kümeleri için kullanılır. Holt'un doğrusal eğilim yöntemi, doğrusal eğilim verileriyle başa çıkmaya yardımcı olan bir yöntemdir. Üstel düzleştirme yöntemlerinin yaygın olarak kullanılmasına ve uzun süredir literatürde olmasına rağmen, modellerin öngörü doğruluğunu potansiyel olarak etkileyen bazı problemleri vardır. Ata, bu sorunların üstesinden gelmek ve daha iyi tahminler sunmak için önerilen yeni bir öngörü yöntemidir. Bu yüksek lisans tezinde, Ata ve üstel düzleştirmenin öngörü doğrulukları, doğrusal eğilimi olmayan veya doğrusal eğilimi ve sönümlenmiş eğilimi olan veri kümeleri için karşılaştırılacaktır. Bu çalışmada verilen sonuçlar, farklı örneklem büyüklükleri ve varyansları ve tahmin hataları içeren simüle edilmiş veri setleri kullanılarak elde edilmiştir. Bu sonuçlar doğrultusunda hem kısa hem de uzun vadeli tahminler karşılaştırılmaktadır. Sonuçlar, Ata yönteminin hem kısa hem de uzun vadeli tahminler için her iki türde zaman serisi verisinde üstel düzleştirmeden daha iyi performans gösterdiğini ortaya koymaktadır.

Anahtar kelimeler: Ata metod, Üssel düzleştirme, Öngörü, Benzetim, Düzleştirme katsayısı, Zaman serisi

CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER ONE
INTRODUCTION

A time series is a series of measurements observed over periodic time intervals. Time series variables are series that show distribution according to a specific time such as day, week, month, season or year. The importance of time series studies is well known nowadays. Monthly product quantities exported in an enterprise, the number of accidents per week in a city, hourly temperature change in a region, annual import and export quantities of a country, annual investment and gross national product revenues, annual unemployment rates, monthly rainfall in a city are examples of time series. For example, it is important for the finance department to forecast the export volume of the next year by using the export amounts of the previous years during the budget preparation phase. In order to forecast the future, time series analysis is used to analyze the model of data points taken over time. In the analysis, forecasts are made for the future by using past observations. These analyzes are especially important for forecasting the future for demographics, economics, finance, operations management, marketing, risk management, medicine and industrial process control.

There are a lot of forecasting methods and models in the literature. Time series plots can reveal patterns such as trends, level shifts, periods or cycles, unusual observations, or a combination of patterns. Of existing models, the most suitable one for the available data should be selected. When choosing between models, the most accurate and simplest model is selected. At the same time, fully automatic and fast models should be preferred. Nowadays we have a lot of data at hand and our model should be able to handle big amounts of data.

Exponential smoothing (ES) method is the most well-known method in the field of time series forecasting. Previous studies in the literature have shown that there are two major univariate forecasting approaches: ES and ARIMA (De Gooijer & Hyndman, 2005). Despite the fact that there are many forecasting methods in time series literature, the most successful forecasting methods are based on the ES. ES is undisputedly one of the most widely used forecasting methods available due to its simplicity, adaptiveness and accuracy (Goodwin, 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 of 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 ES 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 time series have maximum three basic data components: level, trend, and seasonality. Estimates for the final values of these components are used to generate 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). Hence, there are 15 different ES models, the best known of which are simple exponential smoothing (SES) (no trend, no seasonality), Holt's linear model (additive trend, no seasonality) and Holt-Winters' additive model (additive trend, additive seasonality) (Goodwin, 2010). Correct selection of smoothing constants and initial values in an ES model is a very important step in successful forecasting. Gardner (2006) make selection the initial value and smoothing constant values for the ES models. Gardner (2006) has also made a comprehensive review and discussion of these models. Although there have been extensive studies on this subject in the literature, there has never been a consensus among estimators. There is no consistent guide in the forecasting literature on how to estimate smoothing constants and initial values.

Even though an enormous body of research is conducted on forecasting methods and new methods continue to emerge, still an undeniable proportion of these new forecasting methods are based on the two famous forecasting benchmarks: ES and ARIMA (De Gooijer & Hyndman, 2005). Among the two, ES is preferred even more frequently as a basis for new approaches, in its original form and in forecasting combinations compared to ARIMA due to its simplicity and proven accuracy in comparison to ARIMA (Goodwin, 2010). Despite the fact that ES is popular among forecasters due to its simplicity, for some time series it becomes significantly less accurate due to the notorious initialization problem or becomes hard to optimize for data sets with complex structures that involve many components. Ata method which is an alternative forecasting method has been proposed to overcome these issues (Yapar et al., 2017; Yapar et al., 2018; Yapar et al., 2019).

In many branches of statistics, whenever a new methodology is proposed as an alternative to existing methods, in order for the new methodology to gain acceptance in the field, it is expected to be superior with respect to practicality and accuracy compared to the existing methods. This can be usually proven by theoretically showing that the new method will work better (such as showing that a new estimator provides lower bias or has smaller variance) or showing empirically that the new approach will work better by comparing the variances or errors of the existing and proposed methods for large sets of simulated data sets. These data sets are typically generated for different values of pre-determined characteristics that are known to affect the workings of models for the area of interest (e.g. different variances or sample sizes etc).

In time series forecasting, however, over the years the researchers have abandoned this basic route. Recently, when a new forecasting method is proposed, the new method's performance is compared to the existing ones on large sets of real time series data sets. This new route has become the norm with the increasing popularity of the famous M-competitions (Makridakis et al., 1982; Makridakis & Hibon, 2000; Makridakis et al., 2018). M-competitions are forecasting competitions where parts of real time series data sets are presented to the competitors and after they submit their forecasts for pre-determined forecasting horizons, their forecasting errors are compared to forecasting methods that are considered as benchmarks. If a new method can outperform the benchmarks on these data sets, then they are considered as valuable additions to the forecasting literature. Even though the M-competitions are very informative and have led to significant improvements in the field, labeling new forecasting methods as better or worse just based on their performance on these real data sets has been shown to be necessary but not sufficient. For example, Theta method (Assimakopoulos & Nikolopoulos, 2000) which was the winner of the M3-competition (Makridakis & Hibon, 2000), has not been as successful in the M4-competition (Makridakis et al., 2018). A method can be very successful for a collection of data sets where it may not be as accurate for another collection. Since these real data sets are naturally of different scales, scale free error metrics are utilized in these competitions such as the symmetric mean absolute percentage error (sMAPE), the mean absolute scaled error (MASE) and the overall weighted average error (OWA) and the rankings of the methods are highly dependent on the choice of error metric as can be seen from (Makridakis et al., 2020a).

Various forms of Ata method have been shown to outperform their counter ES models for the data sets presented in these competitions, however in order to claim that their performances are truly superior to ES, the performances of these approaches should be compared for simulated time series data sets of various sizes and variability. Since the comparisons on this thesis will be based on these simulated data sets and the choice of error metric is not important since there is no issue of different scaling among the simulated data sets, the results can safely be generalized.

CHAPTER TWO
EXISTING METHODS

Let the observed values of a random variable over time be denoted by X_t for $t = 1, \ldots, n$. The aim is to estimate the future values of this random variable having observed the past values and the h step ahead forecast made at time t is denoted by $\hat{X}_t(h)$. Also h is called the forecasting horizon. 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. When choosing models in time series, the data is divided into two parts, in-sample and out-sample .The part of the data set observed and used in modeling is called in-sample and some of the data used to test forecasting accuracy is called an out-sample. The optimum parameters are obtained by minimizing the in-sample residuals and the forecasting accuracy can be measured using an error metric such as MSE, sMAPE, MAE, MASE etc. A suitable error metric is a very significant decision both for obtaining the correct model and reaching the correct parameters for the chosen model.

There are different types of forecast-error metrics: scale-dependent metrics such as the mean absolute error (MAE); 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 Naive 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 $(X_t - \hat{X}_t)$ 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 (Hyndman et al., 2006):

Mean square error:

$$MSE = n^{-1} \sum_{t=1}^{n} (X_t - \hat{X}_t)^2 \tag{2.2}$$

Root mean square error:

$$rMSE = \sqrt{n^{-1} \sum_{t=1}^{n} (X_t - \hat{X}_t)^2} \tag{2.3}$$

Mean absolute error:

$$MAE = n^{-1} \sum_{t=1}^{n} |X_t - \hat{X}_t| \tag{2.4}$$

Median absolute error:

$$MdAE = median|X_t - \hat{X}_t| \tag{2.5}$$

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

$$p_t = 100(X_t - \hat{X}_t)/X_t \tag{2.6}$$

Percentage errors have the advantage of being scale independent, so they are frequently used to compare forecast performance between different data series (Hyndman et al., 2006). The most widely used measures are:

Mean absolute percentage error:

$$MAPE = 100n^{-1} \sum_{t=1}^{n} |X_t - \hat{X}_t|/|X_t| \tag{2.7}$$

Symmetric mean absolute percentage error:

$$sMAPE = 100n^{-1} \sum_{t=1}^{n} |\frac{X_t - \hat{X}_t}{2}|/|X_t + \hat{X}_t| \tag{2.8}$$

Median absolute percentage error:

$$MdAPE = median(|p_t|) \tag{2.9}$$

Scaled errors: Relative error measures have problems (Hyndman & Koehler, 2006). One of the first problems 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 (Hyndman et al., 2006). Scaled errors were proposed by Hyndman & 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:

Mean absolute scaled error:

$$MASE = mean|q_t| \tag{2.10}$$

$$q_t = \frac{e_t}{\frac{1}{n-1}\sum_{i=1}^{n}|X_i - X_{i-1}|} \tag{2.11}$$

2.1 Simple Forecasting Methods

When data is observed over time, time series data can emerge on many subjects including finance and industry. In these cases, the main goal is to use a series of observations on some variables to forecast a future value. This is achieved by using some clusters of past observations to forecast future values. There are many studies working on 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 goal is then to get an estimate for X_{n+1}. It is assumed that the data do not display any clear trending behavior or any seasonality. Subsequently, the proposed method can be easily adapted for processing data containing such components. Under the model in in (2.1), the goal 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:

$$\hat{a} = F(X_1,\ldots,X_n) = \sum_{t=1}^{n} w_t X_t, \tag{2.12}$$

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 (2.12) 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.

There are a lot of ways to get estimators of form (2.12) and there is not an estimator that will be universally satisfactory, so researchers need a way to choose among all these potential smoothing methods. When choosing a method, researchers should consider some criteria to evaluate the advantages of each option that are important to them. In practice, weighting vectors that give more weight to recent observations should be preferred, so selecting up-to-date observations is crucial. In other words, weighting vectors with $w_j \geq w_i$ for $j > i$ are preferred.The average age (AA) of data used is a popular measure used to measure the ability of a smoothing method to use recent data. It is calculated as follows:

$$AA(\hat{a}) = \bar{k} = n - \sum_{t=1}^{n} t w_t. \tag{2.13}$$

One of the important criteria is the variance of the estimator at hand. Since the estimator in (2.12) is unbiased, its variance can be written as:

$$Var(\hat{a}) = E\left[\left(\sum_{t=1}^{n} w_t X_t - a\right)^2\right] = \sum_{t=1}^{n} w_t^2 \sigma^2 = V\sigma^2. \tag{2.14}$$

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

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

Even though it is desirable to keep both of the metrics in (2.13) and (2.14) 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).

$$\hat{X}_t(h) = \bar{X}_t, \qquad t = 1, \ldots, n \tag{2.15}$$

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 (2.14) 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 called the Naive method. For the naive method, the forecast is simply the last observed value of the time series (Aaker & Jacobson, 1987).

$$\hat{X}_t(h) = X_t, \qquad t = 1, \ldots, n \tag{2.16}$$

2.1.3 Moving Average Method

One of the techniques frequently used in forecasting economic data is the moving average method. Moving averages method is used extensively, especially for individual investors to direct their investments. The moving averages method aims to forecast the trend that may occur in the long term, without being affected by the short-term fluctuations in the considered variable. The most commonly used moving averages method is the simple moving averages method. A moving average of order m can be written as:

$$\hat{T}_t = \frac{1}{t} \sum_{j=-k}^{k} X_{t+j}, \tag{2.17}$$

where $m = 2k + 1$. The estimate of the trend-cycle at time t is obtained by averaging values of the time series within k periods of t (Hyndman & Athanasopoulos, 2018). The average eliminates some of the randomness in the data, leaving a smooth trend-cycle component since observations that are nearby in time are also likely to be close in value (Hyndman & Athanasopoulos, 2018).

2.2 Simple Exponential Smoothing

ES methods have been around since the 1950s. The method is the most popular forecasting method used in many areas. ES 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 ES while he was working for the US Navy during World War II (Anandalingam et al., 2000). Brown designed a monitoring system for fire control information to calculate the location of submarines. This monitoring model, designed by Brown, was basically aimed at simplifying exponential data. This model is still used in modern fire control equipment today. 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 ES at a conference (Brown, 1959).

The equations in ES methods for estimating the parameters and generating the forecasts are very easy to understand. While in moving averages the past observations are weighted equally, ES assigns exponentially decreasing weights as the observations get older. In other words, recent observations are given relatively more weight in forecasting than the older observations.

The simple exponential smoothing method (SES) 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:

$$componentform: \quad S_t = \alpha X_t + (1-\alpha)S_{t-1}, \tag{2.18}$$

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

The model can be written as alternatively:

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

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 (2.18) into itself successively, the model can be re-written as:

$$S_t = \alpha \sum_{k=0}^{t-1} (1-\alpha)^k X_t + (1-\alpha)^t S_0 \tag{2.21}$$

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 ES 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 ES process, the weight given to the data k periods ago is $\alpha(1-\alpha)^k$.

Although significant research has been done on ES models, some of its shortcomings and basic problems have not been solved. This continues to affect the quality of the estimates made using these models. As in all time series literature, there models assume that the future will be the continuation of the past and therefore, past observations are used when making predictions. Model initialization is the process of determining the necessary model parameters such as the basic value, the trend value, and the seasonal indices for the selected forecast model. It is necessary when use a model that forecasts a value for one period based on the forecast value for the period directly before it. Obviously an initial value is required to start the forecast. When the weight of the initial value is very large, α is small and the time series is relatively short. Despite ES models aim to give more weight to recent observations compared to older observations. Putting more weight on the initial value, which does not have a precise calculation method in the literature and is likely to be miscalculated, may lead to the wrong optimum parameter selection. The second problem with exponential models is that the optimization of models is still not well-understood. This situation leads to accuracy problems in estimations. Trying to find an optimal initial value complicates the optimization process. This becomes even more complicated in big data with trends and seasonality, and in this case when obtaining forecasts in a timely manner is of concern, delays in the process are inevitable. As a result, when smoothing a data set over time, the weights should be distributed to observations taking into account where along the time-line the value being smoothed resides. The most recent observation can receive more weight when there are fewer data points that are contributing to the smoothed value and a little less weight as we move along the time-line. ES models always assign the most recent value the same weight no matter where along the time-line smoothing is being carried out. All these problems prevent exponential smoothing from performing well in some cases.

2.3 Holt's Linear Trend Method

SES was extended by Holt (1957) to allow forecasting of data with a trend. SES includes a forecasting equation and two smoothing equations: for trend and level:

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

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

$$\hat{X}_t(h) = S_t + hT_t, \tag{2.24}$$

where X_t is the actual observation, $\hat{X}_t(h)$ is the h-step-ahead forecast, α and β are smoothing parameters, $0 < \alpha, \beta < 1, \beta < \alpha$. There are two smoothing parameters to estimate and initial 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 SES, the level equation here shows that S_t is a weighted average of observations X_t and the 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. For 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.

2.4 Damped Trend Method

Holt' s linear method shows a constant trend (increasing or decreasing) to the future indefinitely. Damped trend method was proposed by Gardner Jr & McKenzie (1985) with a modification of Holt' s linear method to allow the dampening of trends. The formulation of this method is as follows:

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

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

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

A multiplicative damped trend method, allowing for a multiplicative dampening effect on the trend, was later proposed by Taylor (2003) :

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

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

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

for $0 < \phi < 1$. In practice, ϕ usually varies between 0.8 and 1.

CHAPTER THREE
ATA METHOD

When large numbers of time series are involved in the modelling and forecasting process, development of accurate and robust forecasting methods for time series is very important. Despite the advantages of model selection algorithms, more accurate methods are always needed. Considering these needs, the Ata method has a simple and easy optimization. Therefore, it performs well in forecasting time series. Also, Ata method can be applied to non-seasonal or deseasonalized time series.

Ata method has similar form to ES models 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 (Yapar et al., 2019). Therefore the smoothing parameter for this method is a function of t unlike ES 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. This is a very important detail in the forecasting phase. 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 $p \in \{1,\ldots,n\}$, $q \in \{1,\ldots,p\}$ 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 model may be more appropriate when the trending behaviour is not linear but keeps getting stronger over time. 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 $p \in \{1,\ldots,n\}$, $q \in \{1,\ldots,p\}$ 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 $T_1 = 1$.

3.1 Simple Form of Ata Method

While the functional forms of Ata models are generally very similar to those of ES 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 and SES (Yapar et al., 2017). $Ata(p,0)$ 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 (Yapar et al., 2017).

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 $t > p$:

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

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

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

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.

3.2 The Trended Ata Methods

Ata method has been 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.

3.2.1 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)):

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

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

$$\hat{X}_t(h) = S_t + hT_t, \tag{3.12}$$

for $p \in \{1,\ldots,n\}$, $q \in \{1,\ldots,p\}$ 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 $n > p$:

$$S(t) = \left(\frac{p}{t}\right) X_t + \left(\frac{t-p}{t}\right) S_{t-1}, \tag{3.13}$$

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

As with SES, the level equation in 3.10 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 $S_{t1} + T_{t1}$. The trend equation shows that T_t is a 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.

3.2.2 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):

$$S_t = \left(\frac{p}{t}\right) X_t + \left(\frac{t-p}{t}\right)(S_{t-1} * T_{t-1}), \tag{3.14}$$

$$T_t = \left(\frac{q}{t}\right)\left(\frac{S_t}{S_{t-1}}\right) + \left(\frac{t-q}{t}\right) T_{t-1}, \tag{3.15}$$

$$\hat{X}_t(h) = S_t * T_t^h, \tag{3.16}$$

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

3.2.3 Damped Form of Ata Method

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 (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 $Ata_{damped}(p,q,\phi)$ 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:

$$S_t = \left(\frac{p}{t}\right)X_t + \left(\frac{t-p}{t}\right)(S_{t-1} + \phi T_{t-1}), \tag{3.17}$$

$$T_t = \left(\frac{q}{t}\right)(S_t - S_{t-1}) + \left(\frac{t-q}{t}\right)\phi T_{t-1}, \qquad t > p \geq q, \tag{3.18}$$

$$\hat{X}_t(h) = S_t + (\phi + \phi^2 + \phi^3 + \ldots + \phi^h)T_t, \tag{3.19}$$

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 $p \in \{1,\ldots,n\}$, $q \in \{1,\ldots,n\}$, $0 \leq \phi \leq 1$ 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.

CHAPTER FOUR
APPLICATION

4.1 Simulation Study

M-competitions are forecasting competitions where parts of real time series data sets are presented to the competitors and after they submit their forecasts for pre-determined forecasting horizons, their forecasting errors are compared to forecasting methods that are considered as benchmarks. If a new method can outperform the benchmarks on these data sets then they are considered to be valuable additions to the forecasting literature. Even though the M-competitions are very informative and have led to significant improvements in the field, labeling new forecasting methods as better or worse based solely on their performance on these real data sets has been shown to be a necessary but not sufficient condition (Cetin & Yavuz, 2020). For example, Theta method (Assimakopoulos & Nikolopoulos, 2000) which was the winner of the M3-competition (Makridakis & Hibon, 2000), has not been as successful in the M4-competition (Makridakis et al., 2018). A method can be very successful for a collection of data sets where it may not be as accurate for another collection. Since these real data sets are naturally of different scales, scale free error metrics are utilized in these competitions such as the symmetric mean absolute percentage error (sMAPE), the mean absolute scaled error (MASE) and the overall weighted average error (OWA) and the ranking of the methods is highly dependent on the choice of error metric as can be seen from (Makridakis et al., 2020b).

The Ata method was previously adapted to the M-competition data sets in order to see the forecasting performance. Simple and linearly trended versions of Ata method, i.e. $Ata(p,0)$ and $Ata(p,q)$, outperformed SES and Holt's linear trend method for the M-competition data sets (Yapar et al., 2018, 2019). As with real data sets, we have extended simulation studies to see the forecasting performance of the method and determine the impact of these factors on the performance of simulation data sets. In this study, data sets of various sample sizes and variability levels with no trend, linear trend and damped trend were generated. Some part of this data sets were used as in-sample, the appropriate ES and Ata models were optimized there and their forecasting accuracies were compared for the remaining parts of the data sets, i.e. the out-sample. Firstly, n_{total}=48, 68, 88 and 118 observations were generated where 18 values were used as the out-sample. Therefore the in-sample data sets consisted of n=30, 50, 70 and 100 observations. For trend and non-trend dataset, mean 0 and variance σ^2 normal distribution errors were added to constant process. The data sets with no trend were generated by:

$$X_t = a + \epsilon_t \tag{4.1}$$

where $\epsilon_t \sim N(0, \sigma^2)$, $t = 1, \ldots, n$, $a = 10$ and $\sigma = 1, 2, 3, 4$. Similarly the data sets with linear trend were generated by:

$$X_t = a + b * t + \epsilon_t \tag{4.2}$$

where $\epsilon_t \sim N(0, \sigma^2)$, $t = 1, \ldots, n$, $a = 10$, $\sigma = 1, 2, 3, 4$ and $b = 0.2, 0.5, 1, 2$. Similarly the data sets with damped trend were generated by:

$$X_t = a * F(t|\lambda = c) + \epsilon_t \tag{4.3}$$

$$F(t|\lambda = c) = P(X \leq t) \tag{4.4}$$

where $X \sim Exp(\lambda = c)$, $\epsilon_t \sim N(0, \sigma^2)$, $t = 1, \ldots, n$, $a = 10$, $\sigma = 0.5, 1, 1.5$ and $c = 1, 2, 3$. 50,000 data sets were simulated for each case. Since the simulated data sets all have the same scale MSE is used for optimizing the models and measuring forecasting accuracy. More formally, for optimizing the models and obtaining the parameter estimates the one-step-ahead in-sample MSE values are minimized and the average out-sample forecasting errors are calculated by first obtaining $e_{(t+h)} = X_{(t+h)} - \hat{X}_t(h)$ for each forecasting horizon. Then

$$MSE_h = \frac{\sum_{i=1}^{N} e^2_{(t+h),i}}{N} \tag{4.5}$$

is calculated for all forecasting horizons where N is the number of simulated data sets and $e_{(t+h),i}$ represents the forecasting error for the i^{th} time series for forecasting horizon *h*. In short, the squared forecasting errors are averaged over the 50,000 data sets for each forecasting horizon under each setting.

Time series plots of example simulated data sets can be seen below.

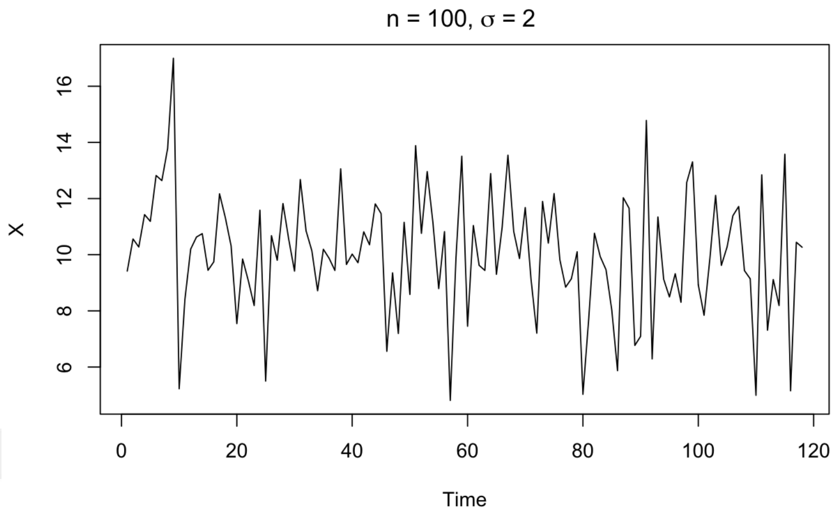

Figure 4.1 Time series graph with no trend

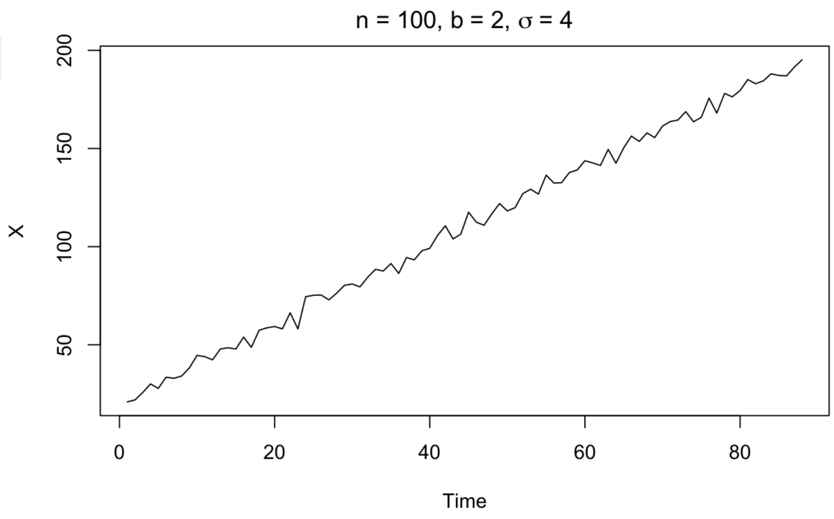

Figure 4.2 Time series graph with linearly trend

Time series graph with different standard deviations and different damped trend coefficient of the simulated data is given for n = 30 in Table 4.3, for n = 50 in Table 4.4, for n = 70 in Table 4.5, and for n = 100 in Table 4.6.

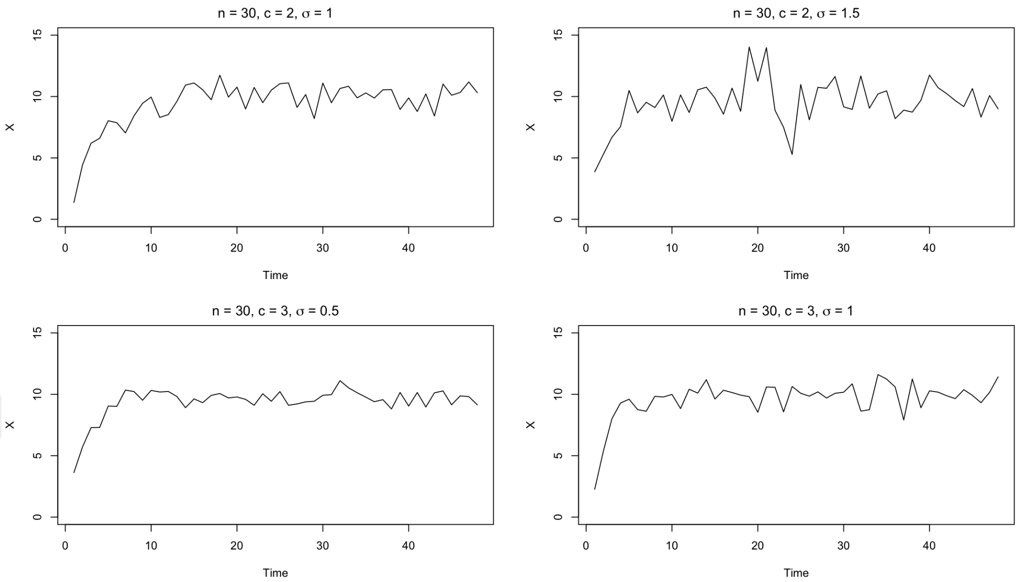

Figure 4.3 Time series graphs for n = 30 with damped trend

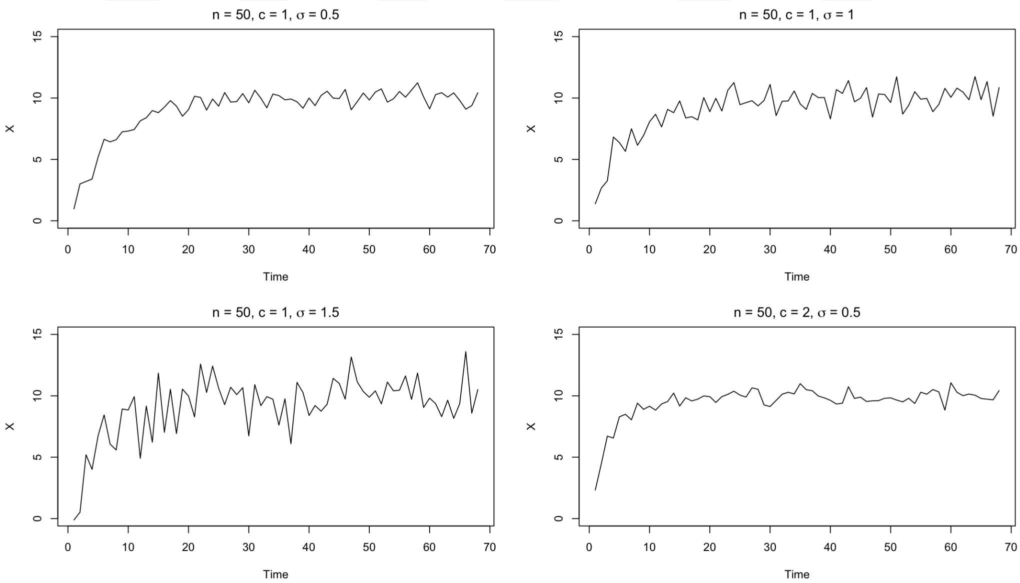

Figure 4.4 Time series graphs for n = 50 with damped trend

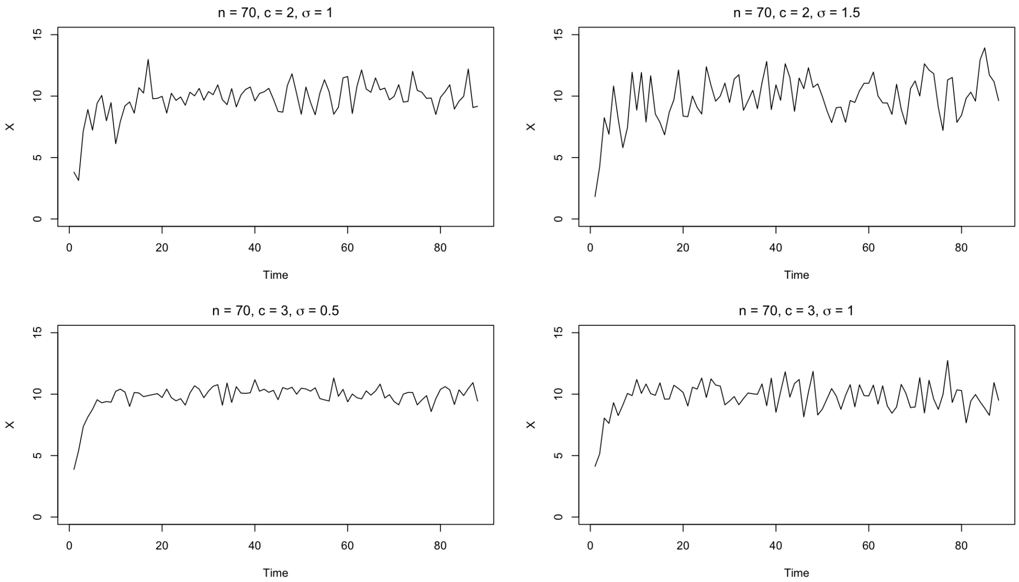

Figure 4.5 Time series graphs for n = 70 with damped trend

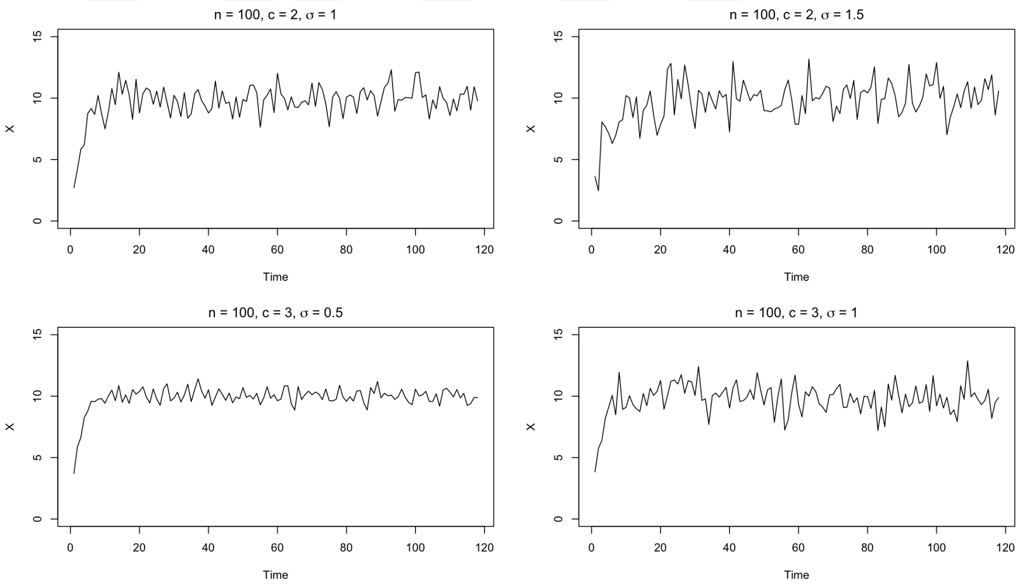

Figure 4.6 Time series graphs for n = 100 with damped trend

Table 4.1 Average forecasting errors (MSE) for data with no trend

	Forecasting Horizon	n=30		n=50		n=70		n=100	
		Ata(p,0)	SES	Ata(p,0)	SES	Ata(p,0)	SES	Ata(p,0)	SES
$\sigma = 1$	1-4	1.050	1.109	1.022	1.071	1.015	1.057	1.008	1.045
	1-6	1.049	1.109	1.023	1.072	1.015	1.057	1.008	1.045
	1-8	1.047	1.107	1.021	1.070	1.017	1.058	1.009	1.046
	1-12	1.048	1.108	1.021	1.070	1.016	1.058	1.011	1.047
	1-15	1.046	1.106	1.023	1.071	1.016	1.058	1.011	1.047
	1-18	1.046	1.105	1.023	1.072	1.017	1.059	1.012	1.048
$\sigma = 2$	1-4	4.179	4.409	4.106	4.299	4.099	4.265	4.039	4.179
	1-6	4.178	4.408	4.102	4.295	4.079	4.246	4.042	4.182
	1-8	4.177	4.408	4.105	4.298	4.079	4.245	4.044	4.186
	1-12	4.176	4.408	4.106	4.299	4.077	4.244	4.051	4.193
	1-15	4.181	4.413	4.104	4.298	4.073	4.242	4.051	4.194
	1-18	4.181	4.414	4.103	4.296	4.073	4.243	4.048	4.192
$\sigma = 3$	1-4	9.416	9.948	9.205	9.650	9.155	9.532	9.092	9.425
	1-6	9.432	9.966	9.217	9.656	9.146	9.521	9.103	9.434
	1-8	9.426	9.958	9.204	9.643	9.141	9.514	9.090	9.421
	1-12	9.429	9.961	9.208	9.644	9.147	9.520	9.095	9.423
	1-15	9.439	9.966	9.213	9.651	9.143	9.516	9.098	9.424
	1-18	9.437	9.969	9.215	9.653	9.145	9.518	9.106	9.432
$\sigma = 4$	1-4	16.740	17.693	16.420	17.182	16.200	16.874	16.199	16.743
	1-6	16.745	17.698	16.405	17.164	16.256	16.919	16.239	16.786
	1-8	16.745	17.689	16.409	17.169	16.257	16.919	16.240	16.794
	1-12	16.737	17.686	16.419	17.182	16.284	16.946	16.227	16.788
	1-15	16.737	17.683	16.405	17.167	16.258	16.922	16.207	16.774
	1-18	16.737	17.683	16.399	17.166	16.253	16.923	16.204	16.781

The resulting mean squared forecasting errors calculated as in equation (4.5) for data with no trend are summarized in Table 4.1. The rows labeled by 1-4, 1-6, ... represent the averaged mean squared forecasting errors for short to long term forecasting horizons. For example, 1-4 row gives $\sum_{h=1}^{4} MSE_h/4$ which is the average short term forecasting error and 1-18 row gives $\sum_{h=1}^{18} MSE_h/18$ which is the average long term forecasting error. For data with no trend, $ATA(p,0)$ produces smaller errors for all sample sizes (small to large) and variability levels for both short and long term forecasting horizons. Both methods get better with increasing sample size but the gap between $ATA(p,0)$ and SES gets even more evident for data sets with more variability.

Table 4.2 Average forecasting errors (MSE) for data with linear trend (n=30)

	Forecasting Horizon	b=0.2		b=0.5		b=1		b=2	
		Ata(p,q)	Holt	Ata(p,q)	Holt	Ata(p,q)	Holt	Ata(p,q)	Holt
$\sigma = 1$	1-4	1.284	1.272	1.283	1.272	1.276	1.273	1.278	1.283
	1-6	1.320	1.328	1.320	1.330	1.312	1.329	1.316	1.346
	1-8	1.360	1.396	1.361	1.400	1.354	1.398	1.358	1.413
	1-12	1.457	1.567	1.456	1.570	1.446	1.567	1.456	1.588
	1-15	1.539	1.718	1.534	1.723	1.529	1.723	1.540	1.748
	1-18	1.628	1.894	1.621	1.900	1.619	1.900	1.634	1.932
$\sigma = 2$	1-4	5.140	5.091	5.114	5.088	5.138	5.134	5.105	5.134
	1-6	5.273	5.312	5.260	5.324	5.286	5.381	5.251	5.382
	1-8	5.443	5.584	5.428	5.594	5.447	5.648	5.411	5.648
	1-12	5.822	6.252	5.809	6.286	5.820	6.337	5.788	6.338
	1-15	6.133	6.849	6.135	6.915	6.144	6.964	6.110	6.965
	1-18	6.490	7.550	6.489	7.618	6.502	7.686	6.468	7.689
$\sigma = 3$	1-4	11.578	11.513	11.617	11.510	11.581	11.512	11.556	11.569
	1-6	11.900	12.073	11.967	12.063	11.913	12.011	11.888	12.136
	1-8	12.259	12.719	12.337	12.694	12.293	12.649	12.254	12.752
	1-12	13.086	14.297	13.183	14.255	13.122	14.166	13.104	14.323
	1-15	13.829	15.756	13.907	15.670	13.851	15.551	13.833	15.743
	1-18	14.633	17.427	14.712	17.309	14.660	17.124	14.643	17.400
$\sigma = 4$	1-4	20.490	20.343	20.635	20.559	20.616	20.472	20.554	20.551
	1-6	21.077	21.267	21.180	21.483	21.209	21.355	21.146	21.559
	1-8	21.779	22.399	21.827	22.586	21.888	22.477	21.790	22.633
	1-12	23.272	25.051	23.272	25.301	23.357	25.149	23.280	25.403
	1-15	24.548	27.448	24.563	27.879	24.653	27.595	24.575	27.919
	1-18	25.923	30.216	25.998	30.855	26.096	30.389	26.010	30.850

The results for the data sets with linear trend are summarized in (Tables 4.2, 4.3, 4.4 and 4.5) for sample sizes 30, 50, 70 and 100 respectively. For $n = 30$ (Table 4.2), for slightly trended data sets and the shortest forecasting horizons (1-4) Holt's method performs slightly better than $Ata(p,q)$ but for longer forecasting horizons (1-6 to 1-18) $Ata(p,q)$ performs better than Holt's method for all trends and variability levels. This is true for b = 0.2, b = 0.5 and b = 1. For data with larger trend (b=2), $Ata(p,q)$ is better for all forecasting horizons and variability levels.

Table 4.3 Average forecasting errors (MSE) for data with trend (n=50)

	Forecasting Horizon	b=0.2		b=0.5		b=1		b=2	
		Ata(p,q)	Holt	Ata(p,q)	Holt	Ata(p,q)	Holt	Ata(p,q)	Holt
$\sigma = 1$	1-4	1.145	1.149	1.155	1.161	1.148	1.153	1.144	1.162
	1-6	1.158	1.174	1.166	1.180	1.161	1.176	1.159	1.188
	1-8	1.175	1.203	1.180	1.208	1.178	1.207	1.174	1.216
	1-12	1.208	1.269	1.209	1.272	1.207	1.265	1.202	1.279
	1-15	1.233	1.328	1.236	1.331	1.234	1.323	1.224	1.335
	1-18	1.263	1.397	1.264	1.399	1.262	1.387	1.252	1.400
$\sigma = 2$	1-4	4.647	4.648	4.619	4.643	4.603	4.614	4.583	4.606
	1-6	4.693	4.728	4.661	4.724	4.658	4.705	4.631	4.703
	1-8	4.746	4.830	4.722	4.829	4.719	4.811	4.685	4.809
	1-12	4.869	5.076	4.845	5.073	4.844	5.061	4.803	5.060
	1-15	4.968	5.297	4.956	5.298	4.946	5.286	4.902	5.290
	1-18	5.081	5.553	5.078	5.559	5.063	5.545	5.014	5.553
$\sigma = 3$	1-4	10.312	10.350	10.424	10.441	10.437	10.490	10.386	10.435
	1-6	10.448	10.558	10.555	10.662	10.521	10.646	10.526	10.653
	1-8	10.582	10.783	10.660	10.865	10.635	10.871	10.660	10.888
	1-12	10.867	11.324	10.939	11.408	10.898	11.412	10.914	11.417
	1-15	11.114	11.818	11.179	11.907	11.140	11.911	11.131	11.893
	1-18	11.365	12.368	11.428	12.476	11.402	12.475	11.382	12.445
$\sigma = 4$	1-4	18.415	18.446	18.547	18.582	18.511	18.572	18.485	18.577
	1-6	18.631	18.806	18.776	18.978	18.698	18.943	18.648	18.884
	1-8	18.841	19.202	18.962	19.354	18.898	19.321	18.867	19.268
	1-12	19.359	20.204	19.458	20.333	19.397	20.333	19.384	20.261
	1-15	19.819	21.124	19.886	21.226	19.812	21.201	19.786	21.119
	1-18	20.289	22.141	20.331	22.243	20.276	22.238	20.257	22.149

When we increase the sample size to $n = 50$ (Table 4.3), that $Ata(p,q)$ performs better than Holt's method for all trend strengths and variability levels for both short and long term intervals. For example, when $\sigma = 2$, the error of $Ata(p,q)$ in forecasting horizon 1-4 is smaller than the Holt' s trend method at all trend strengths ($4.647 < 4.648, 4.619 < 4.643, 4.603 < 4.614, 4.583 < 4.606$). Besides, the error of $Ata(p,q)$ in the forecasting horizon 1-18 at the same standard deviation value is smaller than the Holt's trend method's errors at all trend strengths.

Table 4.4 Average forecasting errors (MSE) for data with trend (n=70)

	Forecasting Horizon	b=0.2		b=0.5		b=1		b=2	
		Ata(p,q)	Holt	Ata(p,q)	Holt	Ata(p,q)	Holt	Ata(p,q)	Holt
$\sigma = 1$	1-4	1.103	1.107	1.107	1.113	1.101	1.110	1.099	1.112
	1-6	1.110	1.118	1.114	1.123	1.106	1.121	1.103	1.122
	1-8	1.116	1.132	1.120	1.137	1.110	1.132	1.108	1.135
	1-12	1.132	1.163	1.133	1.166	1.126	1.165	1.121	1.166
	1-15	1.144	1.190	1.144	1.194	1.138	1.193	1.134	1.195
	1-18	1.157	1.221	1.158	1.227	1.151	1.226	1.146	1.226
$\sigma = 2$	1-4	4.436	4.457	4.429	4.432	4.441	4.451	4.417	4.454
	1-6	4.473	4.507	4.459	4.482	4.463	4.493	4.430	4.489
	1-8	4.491	4.550	4.483	4.527	4.491	4.544	4.452	4.540
	1-12	4.549	4.674	4.548	4.652	4.542	4.654	4.515	4.664
	1-15	4.602	4.788	4.591	4.759	4.592	4.763	4.568	4.780
	1-18	4.656	4.912	4.646	4.883	4.640	4.879	4.615	4.901
$\sigma = 3$	1-4	9.918	9.983	9.980	10.003	9.924	9.988	9.968	10.038
	1-6	9.966	10.085	10.037	10.100	10.003	10.108	9.996	10.120
	1-8	10.042	10.223	10.098	10.216	10.072	10.223	10.045	10.227
	1-12	10.184	10.515	10.210	10.456	10.197	10.489	10.194	10.515
	1-15	10.298	10.782	10.316	10.692	10.306	10.731	10.317	10.779
	1-18	10.412	11.063	10.439	10.977	10.418	11.006	10.428	11.055
$\sigma = 4$	1-4	17.626	17.685	17.742	17.815	17.751	17.899	17.737	17.833
	1-6	17.812	17.946	17.845	17.981	17.790	18.021	17.789	17.989
	1-8	17.934	18.168	17.953	18.187	17.910	18.243	17.878	18.179
	1-12	18.200	18.686	18.153	18.640	18.144	18.712	18.141	18.693
	1-15	18.404	19.107	18.337	19.069	18.327	19.130	18.362	19.173
	1-18	18.621	19.598	18.557	19.582	18.557	19.645	18.561	19.673

When we keep increasing the sample size, it can be seen that the results do not change. $Ata(p,q)$ always produces smaller errors for all standard deviations under the sample size $n = 70$ (Table 4.4). When the standard deviation is small ($\sigma = 1$) or when the standard deviation is large ($\sigma = 4$), $Ata(p,q)$' s error is smaller than the error of the Holt' s trend method at all trend strengths and in all forecasting horizons.

Table 4.5 Average forecasting errors (MSE) for data with trend (n=100)

	Forecasting Horizon	b=0.2		b=0.5		b=1		b=2	
		Ata(p,q)	Holt	Ata(p,q)	Holt	Ata(p,q)	Holt	Ata(p,q)	Holt
$\sigma = 1$	1-4	1.067	1.072	1.071	1.078	1.073	1.083	1.064	1.075
	1-6	1.071	1.077	1.076	1.085	1.073	1.086	1.068	1.083
	1-8	1.076	1.085	1.080	1.092	1.075	1.091	1.070	1.089
	1-12	1.084	1.101	1.086	1.107	1.082	1.106	1.076	1.104
	1-15	1.090	1.114	1.092	1.120	1.088	1.119	1.080	1.117
	1-18	1.097	1.129	1.098	1.134	1.093	1.134	1.085	1.132
$\sigma = 2$	1-4	4.287	4.309	4.282	4.298	4.279	4.301	4.281	4.310
	1-6	4.300	4.335	4.288	4.312	4.295	4.325	4.284	4.326
	1-8	4.307	4.355	4.299	4.334	4.308	4.351	4.298	4.352
	1-12	4.327	4.405	4.332	4.394	4.336	4.404	4.331	4.415
	1-15	4.354	4.461	4.357	4.445	4.356	4.449	4.354	4.467
	1-18	4.377	4.514	4.383	4.499	4.380	4.504	4.371	4.516
$\sigma = 3$	1-4	9.642	9.686	9.632	9.677	9.592	9.658	9.690	9.771
	1-6	9.645	9.715	9.675	9.734	9.636	9.726	9.677	9.786
	1-8	9.679	9.769	9.711	9.802	9.676	9.794	9.707	9.847
	1-12	9.757	9.904	9.793	9.947	9.755	9.949	9.774	9.976
	1-15	9.810	10.015	9.840	10.057	9.803	10.060	9.817	10.085
	1-18	9.859	10.135	9.893	10.179	9.847	10.188	9.862	10.203
$\sigma = 4$	1-4	17.217	17.329	17.230	17.331	17.166	17.300	17.158	17.276
	1-6	17.249	17.426	17.284	17.425	17.170	17.355	17.222	17.378
	1-8	17.301	17.516	17.311	17.515	17.229	17.475	17.261	17.485
	1-12	17.411	17.755	17.402	17.738	17.350	17.734	17.351	17.705
	1-15	17.518	17.998	17.494	17.962	17.448	17.958	17.448	17.930
	1-18	17.621	18.249	17.586	18.211	17.554	18.212	17.552	18.187

For all other sample sizes ($n = 50, 70, 100$), it can be seen from (Tables 4.3, 4.4 and 4.5) that $Ata(p,q)$ performs better than Holt's method for all trend strengths and variability levels for both short and long term intervals. The advantage in forecasting accuracy that $Ata(p,q)$ provides becomes even more evident for increased variability in the data sets.

Table 4.6 p and α values for Ata(p,0) and SES

		Ata(p,0)				SES			
		p				α			
		30	50	70	100	30	50	70	100
$\sigma = 1$	$\bar{x}$	1.392	1.294	1.255	1.231	0.147	0.112	0.094	0.079
	s	1.375	0.877	0.681	0.602	0.128	0.091	0.074	0.061
	M	1.000	1.000	1.000	1.000	0.125	0.097	0.081	0.068
$\sigma = 2$	$\bar{x}$	1.388	1.294	1.258	1.235	0.146	0.096	0.094	0.079
	s	1.354	0.879	0.721	0.613	0.127	0.092	0.075	0.061
	M	1.000	1.000	1.000	1.000	0.125	0.096	0.081	0.068
$\sigma = 3$	$\bar{x}$	1.383	1.294	1.259	1.231	0.147	0.112	0.094	0.079
	s	1.266	0.859	0.694	0.620	0.127	0.091	0.074	0.061
	M	1.000	1.000	1.000	1.000	0.125	0.096	0.081	0.068
$\sigma = 4$	$\bar{x}$	1.382	1.297	1.259	1.232	0.147	0.112	0.095	0.079
	s	1.301	0.904	0.697	0.602	0.127	0.091	0.075	0.061
	M	1.000	1.000	1.000	1.000	0.125	0.097	0.082	0.068

The resulting average p and α, standard deviation of p and α, and median of p and α for data with no trend are summarized in Table 4.6. In here and other parameter tables, $\bar{x}$ represents the average of the parameters, s the standard deviaton of the parameters and M represents the median of the parameters.

The average optimum values of the parameters may also include some information, therefore we calculated the average p at optimum alpha levels for the 50.000 simulated data. When we compare with average p value of $ATA(p,0)$ and average α value of SES, we need to divide average p value by the corresponding sample sizes so that we can compare the weights both methods assign to the most recent observation. For example, in (Table 4.6), when the standard deviation is 3 and the sample size 70 the corresponding average p value is 1.259. When we divide the average p value by the corresponding sample size, we get the value of 0.0179 (1.259/70), the average α value is 0.094 that is, in the same sample size and standard deviation, the average p value divided by n is smaller than the average α. The conclusion here is that $ATA(p,0)$ gives smaller weight to the most recent than SES, but we can make sense of this, because for data with no trend the more we act like the average method, we will be more successful.

Table 4.7 p and q values for all sample sizes for b=0.2

b=0.2		Ata(p,q)							
		p				q			
		30	50	70	100	30	50	70	100
$\sigma = 1$	$\bar{x}$	3.239	3.304	3.368	3.421	1.281	1.361	1.416	1.467
	s	2.429	1.914	1.802	1.724	0.524	0.539	0.548	0.552
	M	3.000	3.000	3.000	3.000	1.000	1.000	1.000	1.000
$\sigma = 2$	$\bar{x}$	3.226	3.321	3.372	3.429	1.279	1.357	1.404	1.466
	s	2.400	1.983	1.834	1.726	0.526	0.540	0.544	0.550
	M	3.000	3.000	3.000	3.000	1.000	1.000	1.000	1.000
$\sigma = 3$	$\bar{x}$	3.241	3.306	3.372	3.427	1.273	1.358	1.411	1.460
	s	2.407	1.939	1.798	1.733	0.517	0.540	0.548	0.548
	M	3.000	3.000	3.000	3.000	1.000	1.000	1.000	1.000
$\sigma = 4$	$\bar{x}$	3.238	3.315	3.366	3.440	1.274	1.350	1.410	1.458
	s	2.478	1.976	1.777	1.746	0.521	0.534	0.549	0.548
	M	3.000	3.000	3.000	3.000	1.000	1.000	1.000	1.000

Table 4.8 α and β values for all sample sizes for b=0.2

b=0.2		Holt' s Trend Method							
		α				β			
		30	50	70	100	30	50	70	100
$\sigma = 1$	$\bar{x}$	0.029	0.028	0.027	0.027	0.006	0.004	0.002	0.001
	s	0.076	0.060	0.052	0.047	0.018	0.010	0.006	0.004
	M	0.000	0.001	0.001	0.002	0.000	0.000	0.000	0.000
$\sigma = 2$	$\bar{x}$	0.029	0.027	0.025	0.022	0.006	0.003	0.002	0.001
	s	0.075	0.060	0.051	0.042	0.018	0.010	0.006	0.004
	M	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
$\sigma = 3$	$\bar{x}$	0.029	0.025	0.023	0.021	0.006	0.003	0.002	0.001
	s	0.078	0.059	0.048	0.040	0.019	0.010	0.007	0.004
	M	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
$\sigma = 4$	$\bar{x}$	0.029	0.024	0.023	0.022	0.006	0.003	0.002	0.001
	s	0.079	0.058	0.049	0.040	0.018	0.010	0.006	0.004
	M	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000

When we compare average α and average β of Holt' s trend method with average p and average q values of the $Ata(p,q)$, we need to divide average p and average q values by the corresponding sample sizes for similar reasons. For example, in (Table 4.7), where the trend strength is 0.2, the standard deviation is 1 and the sample size 100 the corresponding average p value 3.421. If we divide average p value by the corresponding sample size, we get the value of 0.0342 (3.421/100), average the q value is 1.467, if we divide average q value by the corresponding sample size, we get the value of 0.0147 (1.467/100). When we examine Holt's trend (Table 4.8) in the same parameters, average α is 0.027, average β is 0.001. It is seen that $Ata(p,q)$ gives more weight to current observations when estimating the level and the trend than the Holt's trend method (0.0342 > 0.027, 0.0147 > 0.001).

Table 4.9 p and q values for all sample sizes for b=0.5

b=0.5		Ata(p,q)							
		p				q			
		30	50	70	100	30	50	70	100
$\sigma = 1$	$\bar{x}$	3.257	3.336	3.373	3.423	1.329	1.400	1.454	1.510
	s	2.394	1.932	1.810	1.681	0.538	0.546	0.553	0.552
	M	3.000	3.000	3.000	3.000	1.000	1.000	1.000	1.000
$\sigma = 2$	$\bar{x}$	3.257	3.299	3.364	3.427	1.289	1.369	1.419	1.470
	s	2.422	1.937	1.813	1.732	0.530	0.542	0.550	0.551
	M	3.000	3.000	3.000	3.000	1.000	1.000	1.000	1.000
$\sigma = 3$	$\bar{x}$	3.245	3.322	3.367	3.432	1.278	1.361	1.412	1.469
	s	2.376	1.987	1.831	1.740	0.523	0.542	0.548	0.550
	M	3.000	3.000	3.000	3.000	1.000	1.000	1.000	1.000
$\sigma = 4$	$\bar{x}$	3.220	3.319	3.367	3.446	1.279	1.360	1.410	1.462
	s	2.369	1.983	1.831	1.737	0.525	0.542	0.548	0.551
	M	3.000	3.000	3.000	3.000	1.000	1.000	1.000	1.000

Table 4.10 α and β values for all sample sizes for b=0.5

b=0.5		Holt' s Trend Method							
		α				β			
		30	50	70	100	30	50	70	100
$\sigma = 1$	$\bar{x}$	0.029	0.028	0.027	0.027	0.006	0.004	0.002	0.001
	s	0.075	0.060	0.053	0.047	0.018	0.010	0.006	0.004
	M	0.000	0.001	0.001	0.002	0.000	0.000	0.000	0.000
$\sigma = 2$	$\bar{x}$	0.030	0.027	0.025	0.023	0.007	0.003	0.002	0.001
	s	0.076	0.059	0.051	0.043	0.018	0.010	0.006	0.004
	M	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
$\sigma = 3$	$\bar{x}$	0.029	0.025	0.023	0.021	0.006	0.003	0.002	0.001
	s	0.078	0.059	0.049	0.040	0.019	0.010	0.006	0.004
	M	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
$\sigma = 4$	$\bar{x}$	0.030	0.025	0.023	0.022	0.006	0.003	0.002	0.001
	s	0.078	0.059	0.049	0.040	0.019	0.010	0.006	0.004
	M	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000

When we look at (Table 4.9), for example, where the trend strength is 0.5, the standard deviation is 4 and the sample size 50 the corresponding average p value is 3.319. If we divide the average p value by the corresponding sample size, we get the value of 0.0064 (3.319/50). For the same parameters, the average q value is 1.360 and in the same way, if we divide this value by the corresponding sample size, we get the value of 0.0272 (1.360/50). When we examine Holt's trend (Table 4.10) in the same parameters, it is average $\alpha = 0.025$, average $\beta = 0.003$. Similarly, in this table, average p values divided by n are higher than average α values and average q values divided by n are higher than average β values (1.360 > 0.025, 0.0272 > 0.003).

Table 4.11 p and q values for all sample sizes for b=1

b=1		Ata(p,q)							
		p				q			
		30	50	70	100	30	50	70	100
$\sigma = 1$	$\bar{x}$	3.345	3.353	3.394	3.417	1.482	1.537	1.586	1.634
	s	2.363	1.885	1.755	1.640	0.570	0.557	0.550	0.537
	M	3.000	3.000	3.000	3.000	1.000	2.000	2.000	2.000
$\sigma = 2$	$\bar{x}$	3.290	3.322	3.393	3.433	1.333	1.406	1.453	1.509
	s	2.488	1.942	1.786	1.709	0.542	0.553	0.553	0.551
	M	3.000	3.000	3.000	3.000	1.000	1.000	1.000	1.000
$\sigma = 3$	$\bar{x}$	3.255	3.305	3.369	3.438	1.301	1.377	1.435	1.486
	s	2.394	1.936	1.816	1.719	0.536	0.545	0.552	0.553
	M	3.000	3.000	3.000	3.000	1.000	1.000	1.000	1.000
$\sigma = 4$	$\bar{x}$	3.249	3.312	3.373	3.436	1.290	1.366	1.421	1.473
	s	2.410	1.957	1.802	1.718	0.531	0.542	0.550	0.551
	M	3.000	3.000	3.000	3.000	1.000	1.000	1.000	1.000

Table 4.12 α and β values for all sample sizes for b=1

b=1		Holt' s Trend Method							
		α				β			
		30	50	70	100	30	50	70	100
$\sigma = 1$	$\bar{x}$	0.029	0.028	0.027	0.027	0.006	0.003	0.002	0.001
	s	0.075	0.059	0.053	0.047	0.018	0.010	0.006	0.004
	M	0.000	0.001	0.001	0.002	0.000	0.000	0.000	0.000
$\sigma = 2$	$\bar{x}$	0.030	0.027	0.025	0.023	0.006	0.003	0.002	0.001
	s	0.078	0.059	0.050	0.042	0.018	0.010	0.006	0.004
	M	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
$\sigma = 3$	$\bar{x}$	0.029	0.025	0.023	0.021	0.006	0.003	0.002	0.001
	s	0.077	0.058	0.049	0.040	0.018	0.010	0.006	0.004
	M	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
$\sigma = 4$	$\bar{x}$	0.029	0.024	0.023	0.022	0.006	0.003	0.002	0.001
	s	0.077	0.058	0.048	0.040	0.018	0.010	0.007	0.004
	M	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000

Where the trend strength is 1, if we compare (Table 4.11) and (Table 4.12), in cases where the standard deviation is small and large, we can see that in small and large samples the values obtained by dividing the p value by the corresponding sample size are higher than average of α values and the values obtained by dividing the average q value by the corresponding sample size are higher than average of β values. From here, we can see that $Ata(p,q)$ gives more weight to current observations than the Holt's trend method.

Table 4.13 p and q values for all sample sizes for b=2

b=2		Ata(p,q)							
		p				q			
		30	50	70	100	30	50	70	100
$\sigma = 1$	$\bar{x}$	3.508	3.434	3.439	3.449	1.752	1.807	1.831	1.859
	s	2.504	1.932	1.741	1.624	0.524	0.484	0.454	0.428
	M	3.000	3.000	3.000	3.000	2.000	2.000	2.000	2.000
$\sigma = 2$	$\bar{x}$	3.361	3.358	3.377	3.405	1.479	1.539	1.583	1.629
	s	2.468	1.943	1.735	1.635	0.567	0.557	0.550	0.538
	M	3.000	3.000	3.000	3.000	1.000	2.000	2.000	2.000
$\sigma = 3$	$\bar{x}$	3.310	3.321	3.373	3.434	1.378	1.448	1.493	1.545
	s	2.493	1.913	1.781	1.695	0.554	0.562	0.555	0.549
	M	3.000	3.000	3.000	3.000	1.000	1.000	1.000	2.000
$\sigma = 4$	$\bar{x}$	3.290	3.329	3.377	3.424	1.333	1.406	1.457	1.508
	s	2.488	1.976	1.799	1.695	0.542	0.552	0.554	0.550
	M	3.000	3.000	3.000	3.000	1.000	1.000	1.000	1.000

Table 4.14 α and β values for all sample sizes for b=2

b=2		Holt' s Trend Method							
		α				β			
		30	50	70	100	30	50	70	100
$\sigma = 1$	$\bar{x}$	0.030	0.028	0.027	0.027	0.006	0.004	0.002	0.001
	s	0.077	0.060	0.053	0.047	0.018	0.010	0.006	0.004
	M	0.000	0.001	0.001	0.002	0.000	0.000	0.000	0.000
$\sigma = 2$	$\bar{x}$	0.030	0.027	0.025	0.022	0.007	0.003	0.002	0.001
	s	0.078	0.059	0.050	0.042	0.018	0.010	0.006	0.004
	M	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
$\sigma = 3$	$\bar{x}$	0.030	0.025	0.023	0.021	0.006	0.003	0.002	0.001
	s	0.078	0.058	0.048	0.040	0.019	0.010	0.007	0.004
	M	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
$\sigma = 4$	$\bar{x}$	0.030	0.025	0.022	0.022	0.006	0.003	0.002	0.001
	s	0.080	0.058	0.048	0.040	0.019	0.010	0.007	0.004
	M	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000

When we keep increasing the trend strength, we can see that in the (Table 4.13) and (Table 4.14) $Ata(p,q)$ gives more weight to current observations than the Holt's trend method in all parameters. This situation reveals that $Ata(p,q)$ performs better than Holt's trend method when forecasting according to yet another criterion.

Table 4.15 Average forecasting errors (MSE) for data with damped trend (n=30)

	Forecasting Horizon	c=1		c=2		c=3	
		Ata(p,q,ϕ)	Damped	Ata(p,q,ϕ)	Damped	Ata(p,q,ϕ)	Damped
$\sigma = 0.5$	1-4	0.282	0.312	0.267	0.352	0.267	0.365
	1-6	0.282	0.317	0.268	0.364	0.266	0.377
	1-8	0.283	0.323	0.268	0.374	0.266	0.388
	1-12	0.285	0.334	0.268	0.388	0.266	0.403
	1-15	0.287	0.341	0.268	0.397	0.266	0.412
	1-18	0.288	0.347	0.269	0.404	0.266	0.419
$\sigma = 1$	1-4	1.150	1.251	1.068	1.191	1.069	1.188
	1-6	1.154	1.277	1.070	1.206	1.069	1.197
	1-8	1.163	1.309	1.073	1.220	1.067	1.203
	1-12	1.169	1.362	1.074	1.239	1.065	1.214
	1-15	1.175	1.403	1.074	1.250	1.065	1.222
	1-18	1.181	1.442	1.076	1.260	1.065	1.228
$\sigma = 1.5$	1-4	2.591	2.777	2.435	2.658	2.414	2.612
	1-6	2.606	2.845	2.436	2.690	2.415	2.633
	1-8	2.620	2.912	2.440	2.723	2.411	2.651
	1-12	2.651	3.055	2.442	2.781	2.407	2.690
	1-15	2.670	3.156	2.443	2.824	2.408	2.720
	1-18	2.686	3.252	2.447	2.866	2.409	2.750

The results for the data sets with damped trend are summarized in (Tables 4.15, 4.16, 4.17 and 4.18) for sample sizes 30, 50, 70 and 100 respectively. For $n = 30$ (Table 4.15), in the shortest forecasting horizons (1-4) and longer forecasting horizons (1-6 to 1-18) $Ata(p,q,\phi)$ performs better than Damped trend method for all trends and variability levels.

Table 4.16 Average forecasting errors (MSE) for data with damped trend (n=50)

	Forecasting Horizon	c=1		c=2		b=3	
		Ata(p,q,ϕ)	Damped	Ata(p,q,ϕ	Damped	Ata(p,q,ϕ)	Damped
$\sigma = 0.5$	1-4	0.260	0.280	0.259	0.303	0.258	0.299
	1-6	0.260	0.282	0.259	0.307	0.257	0.302
	1-8	0.260	0.283	0.259	0.310	0.257	0.305
	1-12	0.261	0.286	0.259	0.314	0.258	0.309
	1-15	0.261	0.287	0.259	0.317	0.258	0.312
	1-18	0.261	0.288	0.259	0.318	0.258	0.313
$\sigma = 1$	1-4	1.048	1.092	1.034	1.085	1.034	1.078
	1-6	1.049	1.100	1.036	1.089	1.035	1.081
	1-8	1.050	1.106	1.035	1.091	1.034	1.082
	1-12	1.050	1.117	1.035	1.095	1.034	1.085
	1-15	1.050	1.123	1.035	1.097	1.034	1.086
	1-18	1.050	1.129	1.036	1.100	1.033	1.086
$\sigma = 1.5$	1-4	2.361	2.442	2.330	2.409	2.327	2.401
	1-6	2.364	2.460	2.332	2.419	2.329	2.409
	1-8	2.364	2.474	2.334	2.426	2.327	2.412
	1-12	2.366	2.504	2.336	2.441	2.327	2.420
	1-15	2.365	2.523	2.334	2.445	2.326	2.425
	1-18	2.364	2.539	2.336	2.454	2.324	2.428

When we increase the sample size to $n = 50$ (Table 4.16) that $Ata(p,q,\phi)$ performs better than Damped trend method for all trend strengths and variability levels for both short and long term intervals.

Table 4.17 Average forecasting errors (MSE) for data with damped trend (n=70)

	Forecasting Horizon	c=1		c=2		c=3	
		Ata(p,q,ϕ)	Damped	Ata(p,q,ϕ)	Damped	Ata(p,q,ϕ)	Damped
$\sigma = 0.5$	1-4	0.257	0.272	0.257	0.285	0.256	0.278
	1-6	0.257	0.272	0.257	0.287	0.256	0.280
	1-8	0.257	0.272	0.256	0.288	0.255	0.280
	1-12	0.257	0.273	0.256	0.289	0.256	0.283
	1-15	0.257	0.274	0.256	0.290	0.256	0.284
	1-18	0.257	0.274	0.256	0.291	0.255	0.284
$\sigma = 1$	1-4	1.029	1.056	1.025	1.053	1.022	1.047
	1-6	1.029	1.058	1.026	1.054	1.021	1.047
	1-8	1.028	1.059	1.024	1.054	1.021	1.047
	1-12	1.028	1.063	1.025	1.057	1.022	1.049
	1-15	1.028	1.064	1.023	1.056	1.022	1.051
	1-18	1.027	1.066	1.022	1.056	1.023	1.052
$\sigma = 1.5$	1-4	2.311	2.358	2.306	2.349	2.300	2.339
	1-6	2.309	2.360	2.308	2.354	2.298	2.339
	1-8	2.306	2.363	2.306	2.353	2.297	2.342
	1-12	2.311	2.377	2.306	2.357	2.298	2.347
	1-15	2.314	2.385	2.305	2.358	2.299	2.350
	1-18	2.314	2.391	2.307	2.363	2.301	2.354

When we keep increasing the sample size, it can be seen that the results do not change. $Ata(p,q,\phi)$ always produces smaller errors for all standard deviations under the sample size $n = 70$ (Table 4.17). When the standard deviation is small ($\sigma = 1$) or when the standard deviation is large ($\sigma = 4$), $Ata(p,q,\phi)$' s error is smaller than the error of the Damped trend method at all trend strengths and in all forecasting horizons.

Table 4.18 Average forecasting errors (MSE) for data with damped trend (n=100)

	Forecasting Horizon	c=1		c=2		c=3	
		Ata(p,q,ϕ)	Damped	Ata(p,q,ϕ)	Damped	Ata(p,q,ϕ)	Damped
$\sigma = 0.5$	1-4	0.256	0.263	0.253	0.272	0.254	0.273
	1-6	0.255	0.263	0.253	0.272	0.253	0.273
	1-8	0.255	0.263	0.253	0.273	0.253	0.274
	1-12	0.255	0.263	0.253	0.274	0.254	0.275
	1-15	0.255	0.264	0.253	0.274	0.254	0.275
	1-18	0.255	0.264	0.254	0.275	0.254	0.276
$\sigma = 1$	1-4	1.018	1.032	1.016	1.030	1.016	1.031
	1-6	1.019	1.035	1.016	1.031	1.016	1.032
	1-8	1.020	1.038	1.014	1.031	1.016	1.033
	1-12	1.020	1.040	1.015	1.033	1.014	1.032
	1-15	1.020	1.041	1.015	1.033	1.014	1.033
	1-18	1.020	1.043	1.014	1.031	1.014	1.033
$\sigma = 1.5$	1-4	2.301	2.324	2.284	2.305	2.276	2.301
	1-6	2.294	2.320	2.284	2.307	2.273	2.299
	1-8	2.295	2.324	2.282	2.306	2.274	2.300
	1-12	2.296	2.331	2.283	2.308	2.276	2.304
	1-15	2.297	2.336	2.283	2.309	2.276	2.305
	1-18	2.297	2.339	2.285	2.312	2.279	2.308

For all other sample sizes ($n = 50, 70, 100$), it can be seen from (Tables 4.16, 4.17 and 4.18) that $Ata(p,q,\phi)$ performs better than Damped trend method for all trend strengths and variability levels for both short and long term intervals. The advantage in forecasting accuracy that $Ata(p,q,\phi)$ provides becomes even more evident for increased variability in the data sets.

Table 4.19 p, q and ϕ values for all sample sizes for c=1

c=1		Ata(p,q,ϕ)											
		p				q				ϕ			
		30	50	70	100	30	50	70	100	30	50	70	100
$\sigma = 0.5$	$\bar{x}$	2.588	2.462	2.408	2.571	1.670	1.671	1.681	1.598	0.846	0.846	0.846	0.900
	s	1.521	1.062	0.898	0.979	0.565	0.535	0.532	0.526	0.022	0.017	0.017	0.004
	M	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	0.850	0.850	0.850	0.900
$\sigma = 1$	$\bar{x}$	2.458	2.183	2.308	2.418	1.404	1.568	1.404	1.300	0.841	0.709	0.839	0.898
	s	1.704	0.909	0.996	1.134	0.828	0.709	0.698	0.607	0.080	0.103	0.066	0.041
	M	2.000	2.000	2.000	2.000	1.000	2.000	1.000	1.000	0.850	0.750	0.850	0.900
$\sigma = 1.5$	$\bar{x}$	2.349	2.050	2.172	2.309	1.285	1.429	1.286	1.200	0.833	0.702	0.829	0.891
	s	1.877	1.015	1.154	1.288	0.926	0.805	0.802	0.719	0.133	0.162	0.119	0.088
	M	2.000	2.000	2.000	2.000	1.000	1.000	1.000	1.000	0.850	0.750	0.850	0.900

Table 4.20 α, β and ϕ values for all sample sizes for c=1

c=1		Damped Trend Method											
		α				β				ϕ			
		30	50	70	100	30	50	70	100	30	50	70	100
$\sigma = 0.5$	$\bar{x}$	0.121	0.097	0.080	0.036	0.011	0.006	0.004	0.001	0.857	0.860	0.865	0.914
	s	0.194	0.153	0.130	0.069	0.035	0.020	0.013	0.005	0.024	0.022	0.021	0.008
	M	0.006	0.004	0.002	0.000	0.000	0.000	0.000	0.000	0.856	0.860	0.866	0.915
$\sigma = 1$	$\bar{x}$	0.045	0.030	0.022	0.008	0.006	0.003	0.002	0.001	0.879	0.876	0.876	0.914
	s	0.117	0.077	0.060	0.028	0.024	0.013	0.008	0.005	0.030	0.024	0.022	0.010
	M	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.882	0.879	0.878	0.914
$\sigma = 1.5$	$\bar{x}$	0.025	0.016	0.012	0.005	0.005	0.003	0.002	0.001	0.885	0.880	0.879	0.913
	s	0.083	0.053	0.039	0.021	0.021	0.012	0.007	0.004	0.035	0.028	0.025	0.014
	M	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.889	0.882	0.881	0.914

When we compare α and β values of Damped trend method with p and q values of the $Ata(p,q,\phi)$, we need to divide p and q values by the sample sizes. For example, in (Table 4.19), where the trend strength is 1, the standard deviation is 1.5 and the sample size 50 the p value is $2.050/50 = 0.041$, the q value is $1.429/50 = 0.0286$. When we examine Damped trend (Table 2.25) in the same parameters, it is $\alpha = 0.016, \beta = 0.003$. It is seen that $Ata(p,q,\phi)$ gives more weight to current observations than the Damped trend method.

Table 4.21 p, q and ϕ values for all sample sizes for c=2

c=2		Ata(p,q,ϕ)											
		p				q				ϕ			
		30	50	70	100	30	50	70	100	30	50	70	100
$\sigma = 0.5$	$\bar{x}$	2.312	2.225	2.221	2.192	1.732	1.734	1.730	1.733	0.721	0.721	0.754	0.754
	s	1.184	0.767	0.663	0.587	0.530	0.521	0.516	0.513	0.041	0.040	0.034	0.035
	M	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	0.700	0.700	0.750	0.750
$\sigma = 1$	$\bar{x}$	2.279	2.183	2.174	2.142	1.574	1.568	1.542	1.536	0.709	0.709	0.743	0.744
	s	1.317	0.909	0.775	0.672	0.790	0.709	0.707	0.683	0.109	0.103	0.092	0.088
	M	2.000	2.000	2.000	2.000	1.000	2.000	1.000	1.000	0.750	0.750	0.750	0.750
$\sigma = 1.5$	$\bar{x}$	2.159	2.050	2.041	1.998	1.444	1.429	1.407	1.406	0.702	0.702	0.732	0.732
	s	1.482	1.015	0.910	0.814	0.917	0.805	0.787	0.760	0.169	0.162	0.150	0.146
	M	2.000	2.000	2.000	2.000	1.000	1.000	1.000	1.000	0.750	0.750	0.750	0.750

Table 4.22 α, β and ϕ values for all sample sizes for c=2

c=2		Damped Trend Method											
		α				β				ϕ			
		30	50	70	100	30	50	70	100	30	50	70	100
$\sigma = 0.5$	$\bar{x}$	0.283	0.178	0.143	0.104	0.049	0.022	0.010	0.005	0.812	0.814	0.828	0.831
	s	0.266	0.202	0.179	0.144	0.085	0.049	0.030	0.018	0.025	0.024	0.030	0.031
	M	0.211	0.109	0.062	0.027	0.000	0.000	0.000	0.000	0.800	0.800	0.818	0.824
$\sigma = 1$	$\bar{x}$	0.084	0.046	0.027	0.017	0.010	0.005	0.003	0.002	0.829	0.827	0.837	0.838
	s	0.165	0.099	0.068	0.046	0.029	0.015	0.009	0.006	0.032	0.029	0.029	0.029
	M	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.818	0.819	0.836	0.838
$\sigma = 1.5$	$\bar{x}$	0.040	0.022	0.014	0.009	0.006	0.003	0.002	0.001	0.842	0.838	0.844	0.844
	s	0.110	0.064	0.042	0.029	0.021	0.011	0.007	0.005	0.038	0.034	0.032	0.032
	M	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.837	0.835	0.844	0.844

Where the trend strength is 2, if we compare (Table 4.21) and (Table 4.22), in cases where the standard deviation is small and large, we can see that in small and large samples p values are higher than α values and q values are higher than β values.

Table 4.23 p, q and ϕ values for all sample sizes for c=3

c=3		Ata(p,q,ϕ)											
		p				q				ϕ			
		30	50	70	100	30	50	70	100	30	50	70	100
$\sigma = 0.5$	$\bar{x}$	2.237	2.153	2.122	2.191	1.711	1.703	1.709	1.737	0.623	0.623	0.623	0.743
	s	1.187	0.762	0.542	0.606	0.538	0.524	0.517	0.508	0.063	0.061	0.061	0.037
	M	2.000	2.153	2.000	2.000	2.000	1.703	2.000	2.000	0.600	0.623	0.600	0.750
$\sigma = 1$	$\bar{x}$	2.172	2.093	2.052	2.136	1.589	1.591	1.582	1.545	0.613	0.611	0.612	0.733
	s	1.224	0.849	0.669	0.683	0.731	0.689	0.659	0.688	0.135	0.131	0.128	0.092
	M	2.000	2.000	2.000	2.000	2.000	2.000	2.000	1.000	0.600	0.600	0.600	0.750
$\sigma = 1.5$	$\bar{x}$	2.018	1.930	1.876	1.988	1.457	1.454	1.444	1.413	0.618	0.616	0.618	0.723
	s	1.373	0.982	0.810	0.804	0.847	0.774	0.732	0.755	0.198	0.191	0.187	0.149
	M	2.000	2.000	2.000	2.000	1.000	1.000	1.000	1.000	0.650	0.650	0.650	0.750

Table 4.24 α, β and ϕ values for all sample sizes for c=3

c=3		Damped Trend Method											
		α				β				ϕ			
		30	50	70	100	30	50	70	100	30	50	70	100
$\sigma = 0.5$	$\bar{x}$	0.277	0.154	0.099	0.107	0.056	0.023	0.012	0.006	0.806	0.807	0.807	0.830
	s	0.280	0.197	0.144	0.148	0.098	0.050	0.030	0.019	0.017	0.017	0.017	0.031
	M	0.194	0.154	0.026	0.027	0.001	0.023	0.000	0.000	0.800	0.807	0.800	0.822
$\sigma = 1$	$\bar{x}$	0.071	0.036	0.023	0.018	0.010	0.004	0.003	0.002	0.819	0.819	0.821	0.836
	s	0.153	0.089	0.060	0.048	0.030	0.014	0.009	0.006	0.027	0.026	0.027	0.029
	M	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.801	0.802	0.805	0.835
$\sigma = 1.5$	$\bar{x}$	0.036	0.019	0.012	0.009	0.006	0.003	0.002	0.001	0.833	0.831	0.832	0.842
	s	0.105	0.059	0.038	0.029	0.020	0.010	0.007	0.005	0.039	0.035	0.035	0.032
	M	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.815	0.819	0.822	0.842

When we keep increasing the trend strength, we can see that in the (Table 4.23) and (Table 4.24) $Ata(p,q,\phi)$ gives more weight to current observations than the Damped trend method in all parameters.

4.2 Real Data

It is known that the most important factor in finance is the accurate forecasting as in many other fields and therefore, when it is an economic activity, finding the right forecasting technique is a critical process. Forecasting interest rates is essential for the correct planning of resources. For example, one may be interested in forecasting the U.S. annualized monthly inflation rates for services. Annualized monthly inflation rates were forecasted for services using data available publicly on the Bureau of Labor Statistics website (Statistics, 2021). While forecasting, ES and Ata method were used and the forecasting results were compared under different error criteria. To mimic the simulation studies, the data from March 2010 to December 2019 are used where the first 100 observations (until June 2018) are used as the in-sample and the forecasting accuracies of the methods are compared on the remaining 18 observations (the out-sample) (Cetin & Yavuz, 2020). The inflation rates are calculated by

$$x_t = \frac{CPI_t - CPI_{t-1}}{CPI_{t-1}} \times 1200, \tag{4.6}$$

where CPI_t is the consumer price index at time t.

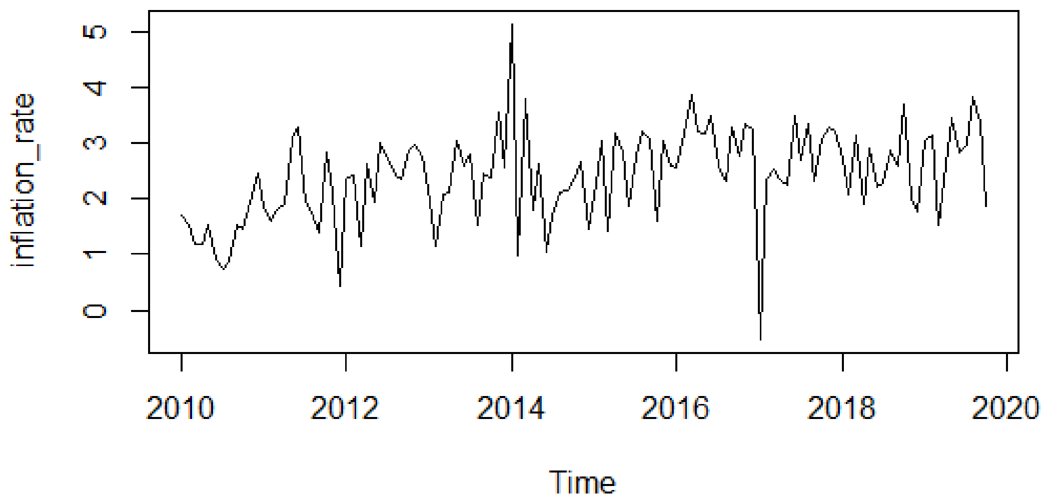

Figure 4.7 Time series plot of the annualized monthly inflation rates for services

The time series plot for the series calculated using equation (4.6) is given in Figure 4.8. It was observed that the data showed a slight trend behavior. Holt's linear trend method and $ATA(p,q)$ were employed for obtaining forecasts. The models were optimized by minimizing the in-sample one-step ahead MSE and the optimum parameters for Holt's linear trend method and $ATA(p,q)$ were found as ($p = 4$; $q = 1$) and ($\alpha = 0.000104$; $\beta = 0.0001$) respectively.

The resulting models according to these parameters can be written as:

$$\begin{aligned} S_t &= 0.000104X_t + (1-0.000104)(S_{t-1}+T_{t-1}), \\ T_t &= 0.0001(S_t - S_{t-1}) + (1-0.0001)T_{t-1}; \end{aligned}$$

$$S_t = \begin{cases} \left(\frac{4}{t}\right)X_t + \left(\frac{t-4}{t}\right)(S_{t-1}+T_{t-1}) & t > 4 \\ X_t & t \leq 4 \end{cases},$$

$$T_t = \begin{cases} \left(\frac{1}{t}\right)(S_t - S_{t-1}) + \left(\frac{t-1}{t}\right)T_{t-1} & t > 1 \\ 0 & t = 1 \end{cases},$$

respectively where $\hat{X}_t(h) = S_t + hT_t$ for both models.

Forecasts were obtained using these models and the forecast errors produced by each method are given in Table 4.25 with respect to various error criteria. The errors presented here are the averages for forecasting horizons 1-18.

Table 4.25 Average outsample forecast errors for inflation rates

Error Criteria	Ata(p, q)	Holt(α, β)
MSE	0.503	0.584
rMSE	0.710	0.765
sMAPE	0.208	0.222
MASE	0.710	0.769
MAE	0.552	0.599
MAPE	0.254	0.279

More detailed averages are prensented in Table 4.26 for MSE. Here the short and long term forecasting performances of the models can be compared.

Table 4.26 Average MSE for inflation rates for various forecast horizons

	Forecasting Horizon					
	1-4	1-6	1-8	1-12	1-15	1-18
Ata(p, q)	0.241	0.273	0.518	0.551	0.456	0.503
Holt(α, β)	0.328	0.322	0.620	0.659	0.545	0.584

$ATA(4,1)$ was able to produce more accurate forecasts on average compared to Holt's method with respect to all error criteria as can be seen from Table 4.25. It is also clear that Table 4.26 Ata produces better forecasts in short and long term forecasting horizons according to MSE error criteria. This application shows that ATA can be used whenever an ES based method is applicable, is easy to implement and provides better forecasts.

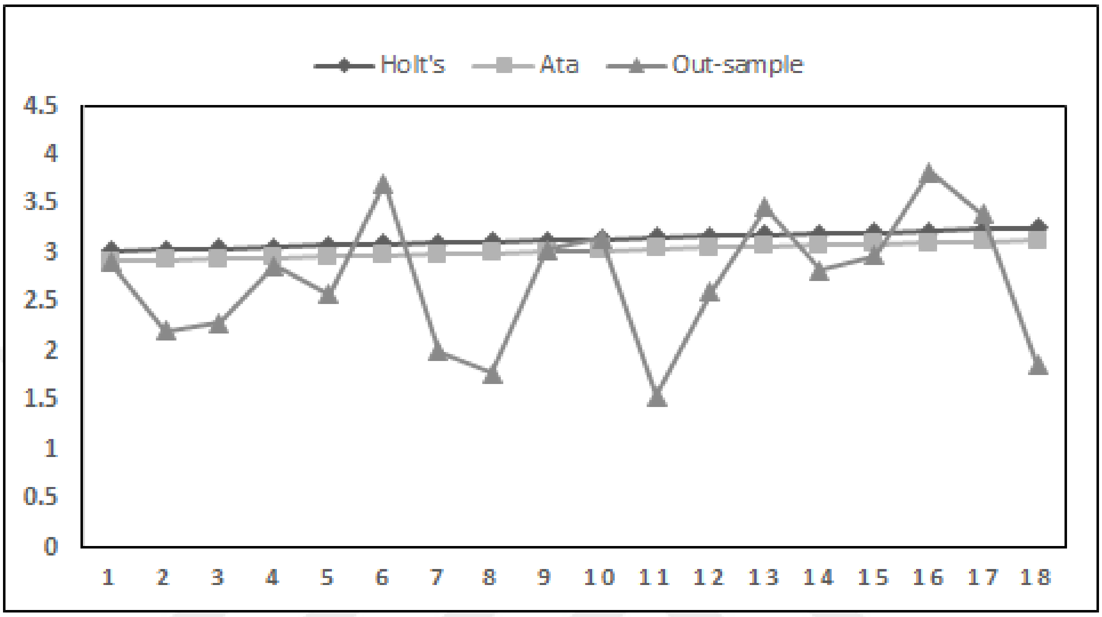

Figure 4.8 Time series plot of the observed and fitted out sample values

CHAPTER FIVE
SUMMARY AND CONCLUSION

Many forecasting methods have been proposed and studied in the literature so far. There is no method that is the best in all data sets according to different error criteria. 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. It is important that time series methods used in many fields produce accurate forecasts. When comparing forecasting methods, comparisons are usually made on the basis of some features. These features are flexibility, speed, accuracy and simplicity. Studies conducted show that when a new method has one or more of these features, this method can be considered as an alternative to existing methods.

ES 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. Like ES, Ata can be generalized to all types of time series data as explained by (Yapar et al., 2019) therefore is at least as flexible as ES. Ata' s weighting scheme takes into account the number of observations being used when obtaining a smoothed value, so Ata method can be argued that it is more adaptive. Also, since Grid search is used for optimizing both ES and Ata and Ata only considers discrete values for its parameters, the optimization for Ata is much faster compared to ES counter models. Another reason why Ata is faster is the fact that Ata models do not require initialization. Hence, Ata is faster and more stable since it does not suffer from initilization problems. Similar to ES and unlike ARIMA, no expert opinion is needed when building an Ata model thus the models are easy to implement and do not change from user to user. This is an advantage, especially considering the need for automated forecasting approaches today, with the increase of big data and streaming data.

Despite the fact that Ata has many advantages if these advantages can not be transformed into more accurate forecasts, we can not advise forecasters to prefer Ata safely. In accordance with this purpose, Ata was compared to its closest and most popular counter model, ES, for simulated data sets with no trend, linear trend and damped trend. Ata performed better than ES models under all situations for all simulated data types when forecasting the near and the distant future. The comparison between Ata method and ES are carried out based on the simple, linearly trended and damped versions of the two approaches. It can not be denied that the Ata method works better even if the sample sizes, variances or trend coefficients of the data sets changes. These comparisons in addition to the empirical performances of the methods prove that based on accuracy, simplicity, speed and interpretability Ata method is better than ES.

Ata method has been very successful in forecasting time series according to the simulation result, and has contributed to the literature in this regard. Considering the fact that, with the help of machine learning, combining forecasts helps improve the forecasting accuracy (Makridakis et al., 2018, 2020c), using appropriate Ata models in these combinations will definitely improve forecasting performance of the approaches. On individual data sets, the researchers can safely and easily apply Ata to obtain better forecasts. In future studies, ATA will also compared for higher order models, for data with level shift or outliers.

REFERENCES

Aaker, D. A., & Jacobson, R. (1987). The sophistication of ‘naive’modeling. *international Journal of Forecasting*, *3*(3-4), 449–451.

Anandalingam, G., Gass, S., & Harris, C. (2000). Simulated annealing. *Encyclopedia of Operations Research and Management Science, 2nd edn. Kluwer Academic Publishers, Boston*, 748–751.

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

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

Cetin, B., & Yavuz, I. (2020). Comparison of forecast accuracy of ata and exponential smoothing. *Journal of Applied Statistics*, 1–11.

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. (2010). The holt-winters approach to exponential smoothing: 50 years old and going strong. *Foresight: The International Journal of Applied Forecasting*, 30–33.

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

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

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. Australia: OTexts.

Hyndman, R. J., et al. (2006). Another look at forecast-accuracy metrics for intermittent demand. *Foresight: The International Journal of Applied Forecasting*, *4*(4), 43–46.

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.

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. (2000). The m3-competition: results, conclusions and implications. *International Journal of Forecasting*, *16*(4), 451–476.

Makridakis, S., Hyndman, R. J., & Petropoulos, F. (2020a). Forecasting in social settings: the state of the art. *International Journal of Forecasting*, *36*(1), 15–28.

Makridakis, S., Hyndman, R. J., & Petropoulos, F. (2020b). Forecasting in social settings: The state of the art. *International Journal of Forecasting*, *36*(1), 15–28.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The m4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, *34*(4), 802–808.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020c). The m4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, *36*(1), 54–74.

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*. New York: Wiley.

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

Statistics, U. B. O. L. (2021). *Inflation for services data*. Retrieved January 05, 2020, from https://www.bls.gov.

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.

Yapar, G., Capar, S., Selamlar, H. T., & Yavuz, I. (2018). Modified holt's linear trend method. *Hacettepe Journal of Mathematics and Statistics*, *47*(5), 1394–1403.

Yapar, G., Selamlar, H. T., Capar, S., & Yavuz, I. (2019). Ata method. *Hacettepe Journal of Mathematics and Statistics*, *48*(6), 1838–1844.

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

APPENDICES

APPENDIX-1: R codes for simulation of ATA(p,0) and SES

```
library(forecast)
nin=30
nout=18
n=nin+nout
nsim=50000
ATAp=NULL
Expoalpha=NULL
foreataa=NULL
foreexpoo=NULL
insample=NULL
outsample=NULL
for (k in 1:nsim){
x=rep(10,n)+rnorm(n,0,3)
xin=x[1:nin]
insample=rbind(insample,xin)
xout=x[(nin+1):n]
outsample=rbind(outsample,xout)
S=s=NULL
for(p in 1:nin){
for(t in 1:nin){
s[t]=ifelse(t>p,(p/t)*x[t]+(t-p)/t*s[t-1],x[t])
}
S=rbind(S,s)
}
xmat=matrix(rep(x[2:nin],nin),nin,nin-1,byrow=TRUE)
mse=apply((xmat-S[,1:(nin-1)])^2, 1, sum)/(nin - 1)
nmin = sum(ifelse(mse == min(mse), 1, 0))
```

```
p = ifelse(nmin == 1, sum(ifelse(mse == min(mse), 1, 0) * (1 : nin)), nin)
ATAp = rbind(ATAp, p)
xints = ts(xin)
xfore = HoltWinters(x, beta = FALSE, gamma = FALSE)
Expoalpha = rbind(Expoalpha, xfore$alpha)
foreexpo = rep(xfore$fitted[nin - 1, 1], nout)
foreexpoo = rbind(foreexpoo, foreexpo)
foreata = rep(S[p, nin], nout)
foreataa = rbind(foreataa, foreata)
}
squarederror = (outsample - foreataa)^2
atahorizon = apply(squarederror, 2, mean)
ata14 = sum(squarederror[, 1 : 4])/(4 * nsim)
ata16 = sum(squarederror[, 1 : 6])/(6 * nsim)
ata18 = sum(squarederror[, 1 : 8])/(8 * nsim)
ata112 = sum(squarederror[, 1 : 12])/(12 * nsim)
ata115 = sum(squarederror[, 1 : 15])/(15 * nsim)
ata118 = sum(squarederror[, 1 : 18])/(18 * nsim)
squarederrorexpo = (outsample - foreexpoo)^2
expohorizon = apply(squarederrorexpo, 2, mean)
ses14 = sum(squarederrorexpo[, 1 : 4])/(4 * nsim)
ses16 = sum(squarederrorexpo[, 1 : 6])/(6 * nsim)
ses18 = sum(squarederrorexpo[, 1 : 8])/(8 * nsim)
ses112 = sum(squarederrorexpo[, 1 : 12])/(12 * nsim)
ses115 = sum(squarederrorexpo[, 1 : 15])/(15 * nsim)
ses118 = sum(squarederrorexpo[, 1 : 18])/(18 * nsim)
```

APPENDIX-2: R codes for simulation of ATA(p,q) and Holt

```
library(forecast)
nin=30
nout=18
n=nin+nout
nsim=50000
ATAp=NULL
Expoalpha=NULL
foreata=NULL
foreexpoo=NULL
insample=NULL
outsample=NULL
tt=c(1:n)
beta=2
vv=1
a=10
for (k in 1:nsim){
x=a+beta*tt+rep(10,n)+rnorm(n,0,vv)
xin=x[1:nin]
insample=rbind(insample,xin)
xout=x[(nin+1):n]
outsample=rbind(outsample,xout)
S=s=b=B=NULL
msemin=10000
pmin=10000
qmin=10000
b[1]=0
sson=NULL
for (p in 1:nin){
for (q in 0:p){
```

```
s[1]=x[1]
for(t in 2:nin){
s[t]=ifelse(t>p,(p/t)*x[t]+(t-p)/t*s[t-1]+b[t-1],x[t])
b[t]=ifelse(t>q,q/t*(s[t]-s[t-1])+((t-q)/(t))*b[t-1],(x[t]-x[t-1]))}
mse=sum((xin[2:nin]-s[1:nin-1])^2)/(nin-1)
if(mse<msemin){
pmin=p
qmin=q
msemin=mse
sson=s
}
}
}
forecast=s[nin]+c(1:18)*b[nin]
xfore=HoltWinters(x,beta=TRUE,gamma=FALSE)
Expoalpha=rbind(Expoalpha,xfore$alpha)
foreexpo=rep(xfore$fitted[nin-1,1],nout)
foreexpoo=rbind(foreexpoo,foreexpo)
}
squarederror=(outsample-forecast)^2
atahorizon=apply(squarederror,2,mean)
ata14=sum(squarederror[,1:4])/(4*nsim)
ata16=sum(squarederror[,1:6])/(6*nsim)
ata18=sum(squarederror[,1:8])/(8*nsim)
ata112=sum(squarederror[,1:12])/(12*nsim)
ata115=sum(squarederror[,1:15])/(15*nsim)
ata118=sum(squarederror[,1:18])/(18*nsim)
squarederrorexpo=(outsample-foreexpoo)^2
expohorizon=apply(squarederrorexpo,2,mean)
holt14=sum(squarederrorexpo[,1:4])/(4*nsim)
holt16=sum(squarederrorexpo[,1:6])/(6*nsim)
```

$holt18 = sum(squarederrorexpo[,1:8])/(8*nsim)$

$holt112 = sum(squarederrorexpo[,1:12])/(12*nsim)$

$holt115 = sum(squarederrorexpo[,1:15])/(15*nsim)$

$holt118 = sum(squarederrorexpo[,1:18])/(18*nsim)$

APPENDIX-3: R codes for simulation of ATA(p,q,ϕ) and Damped

```
library(ATAforecasting)
library(forecast)
nin=30
nout=18
n=nin+nout
nsim=50000
ATAp=NULL
ATAq=NULL
ATAphi=NULL
Expopar=NULL
foreata=NULL
foreexpoo=NULL
insample=NULL
outsample=NULL
t=c(1:n)
tt=t/6
a=10
beta=4
sigma=0.25
for (k in 1:nsim){
x=pexp(tt,2)
x=x*10
x=x+rnorm(30,0,0.5)
x=ts(x)
xin=x[1:nin]
insample=rbind(insample,xin)
xout=x[(nin+1):n]
outsample=rbind(outsample,xout)
xints<-ts(xin)
```

xoutts<-ts(xout)

ata<-ATA(xints,xoutts,parP ="opt",parQ ="opt",parPHI = "opt", h=nout, model.type="A",accuracy.type="MSE",plot.out=FALSE,level.fixed = FALSE)

foreata<-rbind(foreata,ata$forecast)

ATAp=c(ATAp,ata$p)

ATAq=c(ATAq,ata$q)

ATAphi=c(ATAphi,ata$phi)

xfore=holt(xints,damped=TRUE,h=18)

foreexpo=xfore$mean

foreexpoo=rbind(foreexpoo,foreexpo)

Expopar=rbind(Expopar,xfore$model$par[1:3])

}

squarederror=(outsample-foreata)^2

$atahorizon = apply(squarederror, 2, mean)$

$ata16 = sum(squarederror[, 1 : 6])/(6 * nsim)$

$ata18 = sum(squarederror[, 1 : 8])/(8 * nsim)$

$ata112 = sum(squarederror[, 1 : 12])/(12 * nsim)$

$ata118 = sum(squarederror[, 1 : 18])/(18 * nsim)$

$squarederrorexpo = (outsample - foreexpoo)^2$

$expohorizon = apply(squarederrorexpo, 2, mean)$

$damped14 = sum(squarederrorexpo[, 1 : 4])/(4 * nsim)$

$damped16 = sum(squarederrorexpo[, 1 : 6])/(6 * nsim)$

$damped18 = sum(squarederrorexpo[, 1 : 8])/(8 * nsim)$

$damped112 = sum(squarederrorexpo[, 1 : 12])/(12 * nsim)$

$damped115 = sum(squarederrorexpo[, 1 : 15])/(15 * nsim)$

$damped118 = sum(squarederrorexpo[, 1 : 18])/(18 * nsim)$

APPENDIX-4: R codes for the annualized monthly inflation rates

```
library(forecast)
veri<-read.csv("veriler.csv", stringsAsFactors = FALSE, header = TRUE)
veri=veri[2:nrow(veri),]
x<-veri$rate
x<-x[650:767]
n=length(x)
nout=18
nin=n-nout
xin=x[1:nin]
xout=x[(nin+1):n]
s=b=NULL
msemin=10000
pmin=10000
qmin=10000
b[1]=0
sson=NULL
bson=NULL
for (p in 1:nin)
{
for (q in 1:p)
{
s[1]=x[1]
for(t in 2:nin)
{
s[t]=ifelse(t>p,(p/t)*x[t]+((t-p)/t)*(s[t-1]+b[t-1]),x[t])
b[t]=ifelse(t>q,(q/t)*(s[t]-s[t-1])+((t-q)/t)*b[t-1],(x[t]-x[t-1]))
}
mse=sum((xin[2:nin]-(s[1:(nin-1)]+b[1:(nin-1)]))^2)/(nin-1)
if(mse < msemin){
```

$pmin = p$

$qmin = q$

$msemin = mse$

$sson = s$

$bson = b$

}

}

}

$forecast = sson[nin] + c(1:18) * bson[nin]$

$foreata = rbind(foreata, forecast)$

$mseata = mean((xout - foreata)^2)$

$xfore = holt(xints, h = 18)$

$foreexpo = xfore\$mean$

$foreexpoo = rbind(foreexpoo, foreexpo)$