DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

TIME SERIES FORECASTING OF INTERMITTENT DEMAND BY USING ATA METHOD AND COMPUTATIONAL INTELLIGENCE

by
Tuğçe EKİZ YILMAZ

July, 2018
İZMİR

TIME SERIES FORECASTING OF INTERMITTENT DEMAND BY USING ATA METHOD AND COMPUTATIONAL INTELLIGENCE

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
Tuğçe EKİZ YILMAZ

July, 2018
İZMİR

M. Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"TIME SERIES FORECASTING OF INTERMITTENT DEMAND BY USING ATA METHOD AND COMPUTATIONAL INTELLIGENCE"** completed by **TUĞÇE EKİZ YILMAZ** under supervision of **PROF. DR. GÜÇKAN YAPAR** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

....

Prof. Dr. Güçkan YAPAR

Supervisor

.... Assoc. Prof. Dr. Filiz KARDİYEN

Jury Member

.... Assist. Prof. Dr. İdil Yavuz

Jury Member

Prof. Dr. Latif SALUM

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude and thanks to my supervisor, Prof. Dr. Güçkan YAPAR for his support, direction, encouragement that he gave to me, positivity and for being cheerful all the time. Specifically; during my research, I wish to thank my supervisor for his extraordinary patience and invaluable psychological help. I am honored to be a student of him.

Special thanks are to my priceless sisters, Özge EKİZ and Deniz EKİZ for their eternal reliance and support, whenever I needed. Besides, I am also grateful to YILMAZ family for their respect and positively manner during my studying.

This thesis is devoted to my valuable family Nurgül – Cengiz EKİZ and my lovely husband Serkan YILMAZ due to their endless love, trust, encouragement and irreplaceable entities in all my life.

Tuğçe EKİZ YILMAZ

TIME SERIES FORECASTING OF INTERMITTENT DEMAND BY USING ATA METHOD AND COMPUTATIONAL INTELLIGENCE

ABSTRACT

Intermittent demand forecasting is substantially important for firms and commercial activities. Recently, many researchers have focused on the methods which are used to forecast for intermittent demand and developed varied forecasting techniques. The outstanding method among these developed techniques is Croston method which is developed by inspired from exponential smoothing methods. Primarily, exponential smoothing (ES) method which is universally used for more than 50 years is still one of the most prevalent forecasting techniques. Moreover, this model's superiority and success is proved by the famous forecasting competition (M-competitions). Despite its supremacy and widespread use in many areas, ES models have some drawbacks such as the initial value problem that negatively affect the accuracy of forecasts. On the other hand, Croston method is widely used in forecasting of intermittent demand and inventory (stock) control. Since these demands usually include zero values, using the groundbreaking method developed by Croston in this data becomes inevitable. Nevertheless, there are some shortcomings proved such as biasedness in this method. In this study, our aim is to compare the consequences of combining Croston and ATA method as new method for an alternative of exponential smoothing and the other goal is to adjust ATA method's properties to Croston method. Hence, ATA method will be used to intermittent demand forecasting and also proposed this method will be compared statistically to all mentioned methods.

Keywords: Croston's method, demand forecasting, exponential smoothing, forecasting, initial value, intermittent demand, inventory (stock) control, M-competition

İŞLEMSEL ZEKA VE ATA METOT KULLANILARAK KESİKLİ TALEPLERİN ZAMAN SERİSİ TAHMİNLEMESİ

ÖZ

Kesikli taleplerin öngörüsü firmalar ve ticari faaliyetler için son derece önemlidir. Son zamanlarda birçok araştırmacı kesikli talepler için öngörü elde etmede kullanılan yöntemler üzerine yoğunlaşmıştır ve farklı öngörü yöntemleri geliştirmişlerdir. Bu geliştirilen yöntemler içerisinde öne çıkan metot üssel düzleştirme (ÜD) yönteminden esinlenilerek geliştirilen Croston metodudur. Öncelikle, 50 yılı aşkın bir süredir yaygın olarak kullanılan üssel düzleştirme metodu mevcut en yaygın öngörü tekniklerinden biridir. Hatta bu yöntemin üstünlüğü ve başarısı ünlü tahminleme yarışmalarınca (M-yarışması) kanıtlanmıştır. Fakat bunun üstünlüğü ve birçok alanda yaygın kullanımı olsa da, ÜD modelleri başlangıç değeri gibi öngörü doğruluğunu olumsuz yönde etkileyen bazı dezavantajlara sahiptir. Öte yandan Croston metodu kesikli taleplerin öngörüsünde ve envanter kontrolünde yaygın olarak kullanılır. Bu talepler genellikle sıfır değerini içerdiği için, bu verilerde Croston tarafından geliştirilen çığır açan yöntemi kullanmak kaçınılmaz olur. Fakat bu metodun da yanlılık gibi kanıtlanmış bazı kusurları bulunmaktadır. Bu çalışmada amacımız üssel düzleme yöntemine alternatif olarak geliştirilen yeni ATA metodu ve Croston yöntemini karşılaştırmaktır ve diğer amacımız ATA metodun özelliklerini Croston metoduna uyarlamaktır. Sonuç olarak ATA metot kesikli taleplerin tahmin edilmesinde kullanılacaktır ve hatta önerilen bu yöntem bahsedilen tüm yöntemler ile istatistiksel olarak karşılaştırılacaktır.

Anahtar kelimeler: Croston metodu, talep öngörüsü, üssel düzleştime, öngörü, başlangıç değer, kesikli talep, envanter (stok) kontrolü, M-yarışması

CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER ONE
INTRODUCTION

Intermittent demands appear sporadically with some periods demonstrating no demand. It is often referred to as sporadic demand. For intermittent demand items the observed demand during some periods is zero interspersed by periods with low or high, regular or irregular non-zero demand (Syntetos, 2001). Intermittent demand is random demand with a large proportion of zero values and does not contain any nonnegative integer values (Silver, 1981; Shenstone & Hyndman 2005; Willemain et al., 2004). Data for such items are composed of time series of non-negative integer values where some values are zero (Vasumathi & Saradha, 2013). Demand that is intermittent is often “lumpy” or “erratic” meaning that there is great variability among the nonzero values (Willemain et al., 2004). This demand occur various realms. Willemain et al. (2004) describe intermittent demand in several areas, in particular heavy machinery and respective spare parts, aircraft service parts, maritime spare parts, electronics, etc.; while Syntetos & Boylan (2005) discovered intermittent demands in the automotive spare parts (Kourentzes, 2013). Similarly, Ghobbar & Friend (2003) characterized a significance of intermittent demands in the field of aviation, just as Johnston et al. (2003) mentioned that spare parts often exhibited intermittent demand and were particularly prevailing in the aerospace, automotive, military and IT (information technology) sectors (Altay et al., 2012). In the military for instance, spare parts constitute such a big percentage of inventory the US Defense Logistics Agency (DLA) identified 24 strategic initiatives to better mitigate shortages of critical parts inventory (Altay et al., 2012). Johnston et al. (2003) and Kourentzes (2013) specified that such items can account for approximately 60% of the total stock value.

Intermittent demand creates significant problems in the manufacturing and supply realm as far as forecasting and inventory controls are concerned. It is not only the variability of the demand sizes, but also the variability of the demand pattern that make this demand so difficult to forecast (Syntetos & Boylan, 2001; Syntetos, 2001). Some of these drawbacks are that historical data of spare parts demand are usually very restricted (Hua et al., 2007) and in some industries, spare parts inventory level is largely a function of how equipment is used and how it is maintained (Kennedy et al., 2002). The most commonly cited problems by firms in the development of reliable forecasts are the relatively high percentage of items which have experienced erratic or lumpy demand. Hence, in order to promote inventory holding and replenishment decisions and overcome these shortcomings, accurate demand forecast are inevitable (Kourentzes, 2013).

1.1 Literature Review

There are several approaches to deal with intermittent demand data such as traditionally exponential smoothing. Conventional forecasting methods are often based on assumptions that are accused in inappropriate for items with an erratic demand pattern. However, a first systematic approach to cope with intermittent demand data was presented by Croston (1972) and later it was corrected by Rao (1973). The third chapter introduces a forecasting model that compares Croston's method, which was developed specifically for forecasting erratic demand, with includes more conventional methods including exponential smoothing and Box – Jenkins methodology. Croston (1972) highlighted the inadequacies of exponentially weighted moving averages to estimate the underlying demand pattern. For these item lines, the observed demand during many periods would be zero (Johnston & Boylan, 1996). Croston (1972) proposed the decomposition of such data into two separate series i.e. corresponding to the nonzero demand sizes ($\hat{Z}$) and the inter-demand intervals ($\hat{n}$) (Petropoulos et al., 2016; Vasumathi & Saradha, 2013).

Intermittent time series present a special challenge because they display variability in demand size in addition to demand interval. Therefore, forecasting and stock control of intermittent demand attracted a substantial amount of academic researches (Altay et al., 2012; Syntetos, 2001; Syntetos & Boylan, 2005; Willemain et al., 2004; etc.). Most work on intermittent demand forecasting is based on Croston's innovative article. Also, these researches on intermittent demand assume independence between successive demand sizes and successive demand intervals, and independence between sizes and intervals, because most of the forecasting and inventory control theory in particular Croston method is established on the assumption that successive demand values are independent (Altay et al., 2012; Syntetos, 2001). In other respects, Croston's method was alleged to be unbiased but despite its theoretical superiority modest benefits were registered in the literature (Syntetos et al., 2009). Nevertheless, Syntetos & Boylan (2001) proved that the Croston's method is positively biased (i.e. over-forecasting mean demand) and then Levén & Segerstedt (2004) proposed the Modified Croston's method by using earlier work by Segerstedt (2000). This method was only a modification of the original Croston's method. However; the modification to Croston's method proposed by Levén & Segerstedt (2004) was found by Syntetos & Boylan (2005) to be more biased than the original estimator. They subsequently proposed an "approximately" unbiased estimator: SBA (Syntetos-Boylan Approximation, Syntetos & Boylan, 2005). Subsequently, Shale et al. (2006) derived the expected bias in Croston's method and proposed an "exact" correction factor, SBJ. The first estimation procedure (SBA) performs a deflating factor to the Croston estimates in order to take away the bias. But a little bias though still remains, on the opposite side. This procedure was further evaluated by Teunter & Sani (2009) and it was found to perform very well. Moreover, according to their findings, Croston's original method presents smaller bias if few demands are zero, whereas SBA modification has a smaller bias if many demands are zero (Petropoulos et al., 2013). Furthermore, the modification considerably enhances the performance of the original Croston method, the Syntetos-Boylan variation (Teunter & Duncan, 2009). However, there are cases and some (negative) bias remains while the Syntetos-Boylan Approximation (SBA) may be more biased than the original Croston method (Wallström & Segerstedt, 2010; Teunter & Sani, 2009; Teunter et al., 2011). Hence, Teunter et al. (2011) proposed a novel method which is TSB (Teunter-Syntetos-Babai) method. This method that is originally not proposed for intermittent demand, but does update the forecast after every period, is the naïve method that uses the last observed demand as the forecast for future periods (Teunter et al., 2011). Nevertheless, Teunter et al. (2011) suggested the use of their new method as an alternative to Croston and SBA methods, but their findings are only relied upon simulated data (Petropoulos et al., 2013). Although there are several modifications of original Croston method, all of them built on the same concept that is separating the non-zero demands from the intervals or the probability to have a non-zero demand (Petropoulos et al., 2016). Thus in this study, we will empirically investigate the Croston method which is an essential and first systematic method for an intermittent demand data in addition to SBA and

SBJ methods that are widely used after Croston method in literature thoroughly for M4-competition inventory data set.

1.2 Scope of Research

In this study, we focused on forecasting for intermittent demand. Some methods which are used in these demands were mentioned. Moreover, we dealt with a novel ATA method (Yapar et al., 2017) that is developed for alternative to exponential smoothing method and compared the consequences concerning with Croston method and its two variations: SBA and SBJ. Our aim is to compare the solutions of Croston method in addition to its two variations SBA and SBJ as well as ATA method for intermittent demand data.

This study consists of six chapters. In the first chapter, a review of the literature of the intermittent demand forecasting and the outline and also the aim of the study is given. In the second chapter, intermittent demand is thoroughly mentioned. Fundamental forecasting approaches and Croston method is negotiated in chapter three. One of the selected forecasting methods is the relatively well-known Croston method, which separately applies exponential smoothing to the interval between demands and the size of the demands. In the fourth chapter, a new forecasting method ATA is analyzed and its applications and consequences for intermittent demand are given. In the fifth chapter, ATA method and Croston method as well as its two variations are compared and the results obtaining from some error metrics are given. Finally; in the sixth chapter, all conclusions of these methods are given and compared according to some error metrics in the M4-competition inventory data set and thus the best method for this demand is selected.

CHAPTER TWO
INTERMITTENT DEMAND DATA

Modern inventory control systems may involve thousands of items, many of which show very low levels of demand. Furthermore, such items may be requested only on an occasional basis. When events corresponding to positive demands occur solely sporadically, we refer to demand as intermittent. Silver et al. (1998) defined intermittent demand as "*infrequent in the sense that the average time between consecutive transactions is considerably larger than the unit time period, the latter being the interval of forecast updating"* (p. 127). Intermittent demand for consumable spare parts tends to be divided into two categories, namely erratic and lumpy. Both patterns are characterized by infrequent demand transactions. While a lumpy demand pattern is distinguished by low demand sizes, an erratic demand pattern is determined by variable demand sizes (Eaves, 2002). Besides, lumpy demand is always sporadic while the opposite is not necessarily true. Vereecke & Verstraeten (1994) are defined lumpy demand items as "*items whose demand frequency is less than 4 times a year* (p. 379)". So, it may be take a part in a classification table of demand pattern. But it is not necessary. These different parts of items are also part of intermittent demand and they are pictured in Figure 2.4. Items with intermittent demand also known as volatile, variable, sporadic or unpredictable demand have many zero or low volume values interspersed with random spikes of demand that are often many times larger than the average. Intermittent demand comes about when a product experiences some periods of zero demand. Often in these situations, when small demand occurs sometimes highly variable in size (Snyder et al., 2012). Therefore intermittence (sporadicity) refers to the demand incidences and not to the demand size when demand occurs. Nevertheless sporadicity has often been associated with lumpiness in the academic literature (Ward, 1978; Schultz, 1987; Dunsmuir & Snyder, 1989). Moreover lumpiness refers to intermittence coupled with erratic (irregular) demand sizes when demand occurs (Syntetos, 2001).

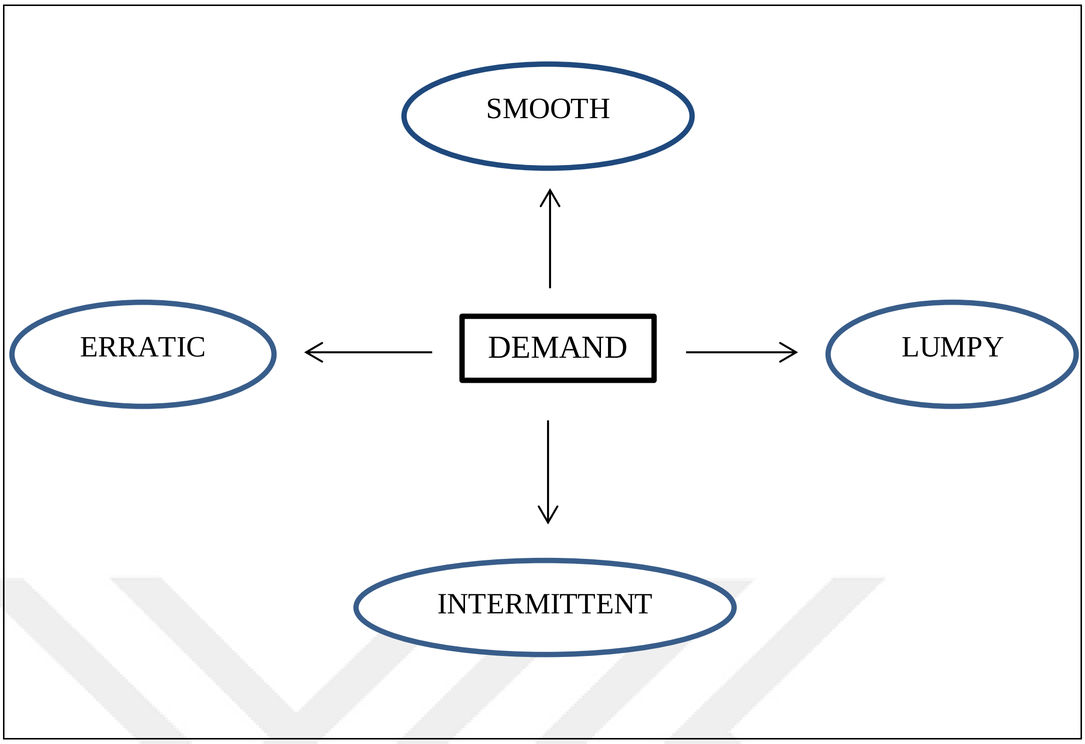

Figure 2.1 Classification of demand patterns

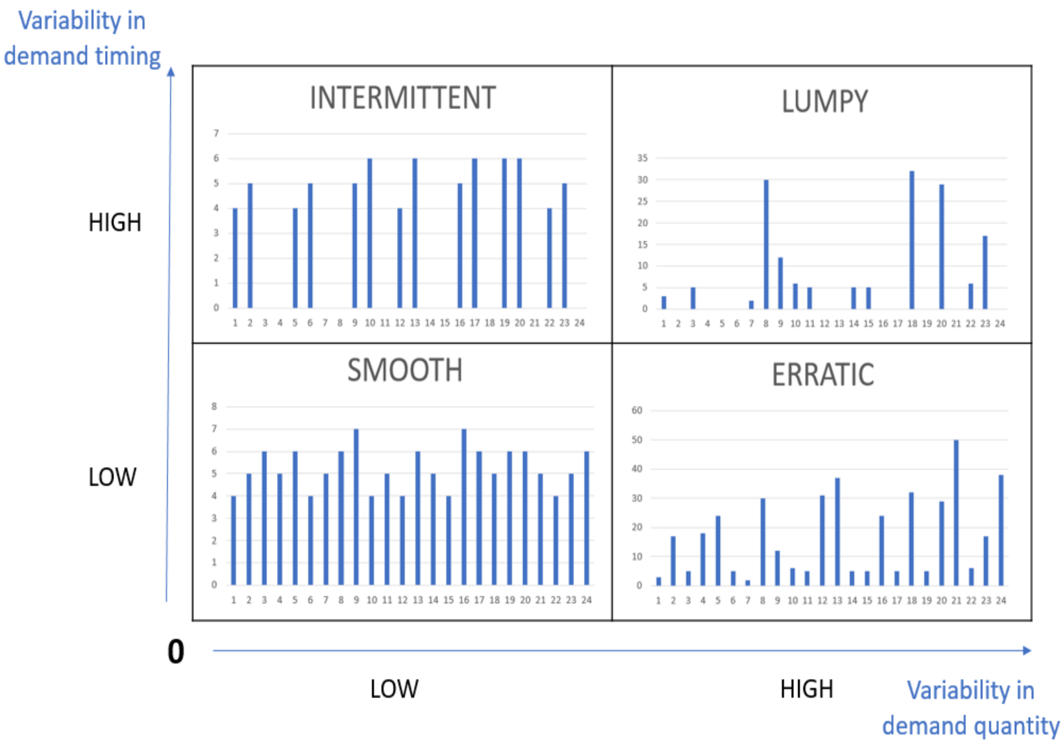

Figure 2.2 Graphics of demand patterns

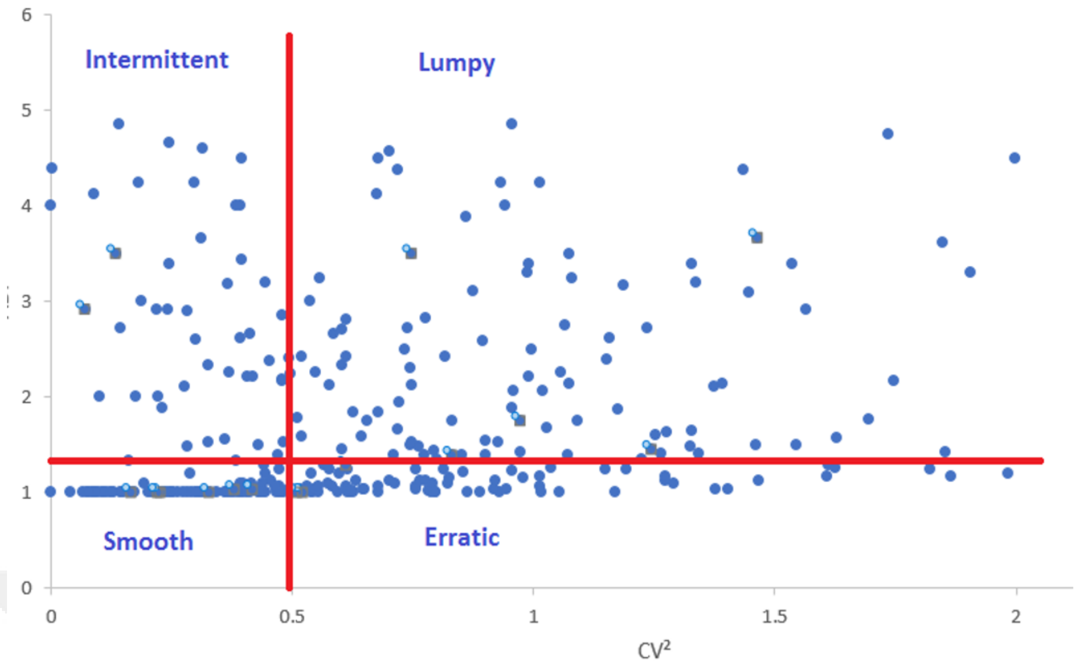

Figure 2.3 Scatter plot of demand patterns

The quality of a forecast strongly depends on the demand history characteristics. In order to determine the characteristics of a demand history, two coefficients are used:

- The first coefficient is the Average Demand Interval (ADI), it measures the regularity of a demand in time by computing the average interval between two demands.
- The second coefficient is the square of the Coefficient of Variation (CV^2), it measures the variation in the demand quantities.

Table 2.1 Coefficients of demands

	Average demand interval	Square of the coefficient of variation
Smooth demand	< 1.32	< 0.49
Intermittent demand	≥ 1.32	< 0.49
Erratic demand	< 1.32	≥ 0.49
Lumpy demand	≥ 1.32	≥ 0.49

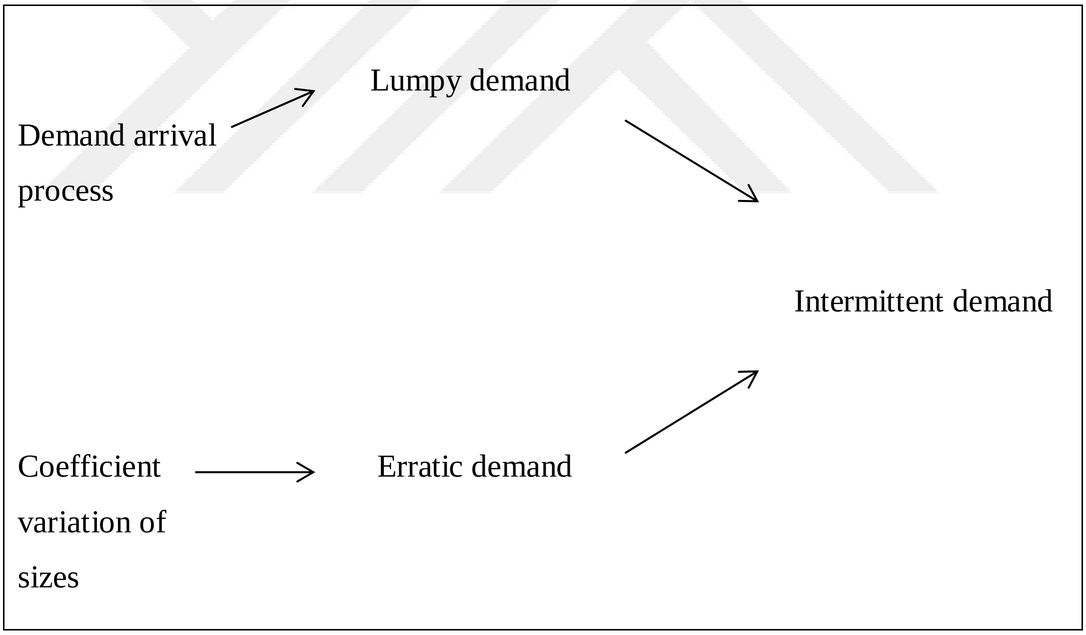

Figure 2.4 Classification of non-normal demand pattern

- An intermittent demand item is an item whose demand is zero in some time periods.
- An erratic demand item is an item whose demand size is highly variable.
- A lumpy demand item is an item whose demand is zero in some time periods. Furthermore demand is highly variable when it occurs (Syntetos, 2001).

Intermittent demand appears at random, with some periods having no demand. Moreover when a demand occurs the request is very often for more than a single unit. Since the early work of Brown (1959), the problem of forecasting for fast moving inventory items has attracted a substantial amount of academic research. As such, intermittent demand creates significant problems in the manufacturing and supply environment as far as forecasting and inventory control are concerned. It is not only the variability of the demand pattern but also, in many cases, the variability of the demand size that makes intermittent demand so difficult to forecast (Syntetos & Boylan, 2001). Nevertheless, forecasting for items with intermittent demand has received far less attention, though such items typically account for considerable proportions of stock value and revenues (Syntetos et al., 2015).

Intermittent demand is often experienced in industries such as aviation, automotive, defense and manufacturing; it also typically occurs with products nearing the end of their life cycle. Some companies operating in these areas observe intermittent demand for over half the products in their inventories. Indeed, such items may collectively account for approximately 60% of the total stock value (Johnston et al., 2003) and are especially prevalent in the aerospace, automotive, military and IT (information technology) sectors (Syntetos & Boylan, 2010). A survey by Deloitte (2011) benchmarked the service businesses of many of the world's largest manufacturing companies with combined revenues reaching more than $1.5 trillion; service operations accounted for an average of 26% of revenues. Thus small improvements in management of intermittent demand items may be translated to substantial cost savings; research in this realm has direct relevance to a wide range of companies and industries (Syntetos et al., 2015).

Selecting the right periodic inventory system for intermittent demand is a problem facing many organizations (Sani & Kingsman, 1997). Intermittent demand makes it difficult to accurately estimate the safety stock and service level inventory requirements needed for successful supply chain planning (Willemain et al., 2004). Since forecast of such demand have been so unreliable, most companies forecast inventory requirements relying primarily on subjective business knowledge, forecast only a fraction of their higher volume inventory, or use simple rule of thumb estimates for inventory control. The result is that billions of dollars are wasted every year because of either excess inventory costs or poor customer services due to stock-outs. In such situations there must be a clear financial incentive to inventory control and retaining proper stock levels, and therefore to forecasting accurately demand for these items.

Forecasting intermittently demanded, slow-moving product items or parts is a problem that is specifically well-known to managers in service parts organizations and companies in the capital goods industries, among others. There are two common mistakes these firms make when forecasting intermittent demand. Firstly, they focus on estimates of per period demand when they should really center on estimates of the inventory stocking requirements necessary meet to their desired service levels. Secondly, they use forecasting methods that are inappropriate for intermittent demand. Conventional forecasting methods such as exponential smoothing, Box-Jenkins methodology and moving averages that are designed for normal, high volume demand just do not work very well with intermittent demand. Furthermore, these traditional forecasting techniques fail since they try to identify recognizable patterns in the demand data such as trend and seasonality. However, intermittent demand data do not exhibit such regular patterns. Instead demand that is intermittent is often also lumpy meaning that there is great variability among the nonzero values (Willemain et al, 2004). On the other hand, traditional technologies neglect the special role of zero values when analyzing demand and tend to simply smooth over the zero and occasional non-zero values found in intermittent data. But, the zero values are so important. If the frequency and timing of zero values are not calculated correctly, there is no generated accurate demand forecast or proper estimate the required inventory level for an intermittent item. However, new empirical methods like Croston method and statistical bootstrapping do not make assumptions about patterns in the data, rather they account for the zero values found in intermittent demand and can produce accurate forecast and inventory stocking estimates. As summarized by Willemain et al. (2004), there are five types of forecasting methods for intermittent demand. That is;

- Assessing or relaxing the standard assumptions for non-intermittent demand,
- Non-extrapolative approaches, such as reliability theory and expert systems,
- Variants of the Poisson model,
- Simple statistical smoothing methods,
- Croston's variant of exponential smoothing (ES).

Ghobbar & Friend (2003) evaluated some of the above methods by applying them to forecast the intermittent demand of aircraft's spare parts. They compared and evaluated 13 methods, that is, additive Winter, multiplicative Winter, seasonal regression model, component service life, weighted calculation of demand rates, weighted regression demand forecasters, Croston, single exponential smoothing, exponentially weighted moving average, trend-adjusted exponential smoothing, weighted moving averages, Holt linear trend model and adaptive response rate single exponential smoothing. They found that exponentially weighted moving average and Croston method outperformed other forecasting methods for intermittent demand (Hua et al., 2007). However, exponentially weighted moving average method does not properly function in this demand including zero values. Thus, in forecasting intermittent demand Croston method provides by far the best for forecast accuracy and robustness among traditional methods (Smart, 2009).

When evaluating for intermittent demand forecasting, there is some properties for needed that we consider. Firstly, the forecasting method should accurately forecast the entire distribution of lead time demand values (i.e. total demand over a lead time), rather than produce just a single number representing the average demand per period. Secondly, the solution should provide optimal service level inventory requirements for satisfying total lead time demand. Lastly, the solution should accurately reflect the asymmetrical (i.e. non-normal) nature of the lead time demand distribution.

An important paper on intermittent demand is the 1972 paper of Croston who demonstrated that using simple exponential smoothing forecasts to set inventory levels could lead to excessive stock levels. He argues that exponential smoothing places the most weight on more recent data and therefore gives estimates that are the highest just after a demand, and the lowest just before a demand (Eaves, 2002). Moreover, Croston’s method (1972) is a method that establishes demand estimates taking into account both demand size and the interval between demand occurrences. More information about Croston method and its applications will be thoroughly elaborated in Section 3.3.

CHAPTER THREE
FORECASTING APPROACHES FOR INTERMITTENT DEMAND DATA

Forecasting is a process in which one is dealt with a sequence of numbers that is a discrete time series. Some of the numbers or maybe all the numbers in this sequence have been observed and are known. The forecasting problem is estimating the probability distribution from which further numbers in the sequence will be drawn (Brown, 1962).

Most forecasting problems involve the use of time series data. A time series is simply a set of observations that is measured consecutive or chronological time periods (Paul, 2011).

All forecasting techniques can be categorized into two prevalent classes: qualitative and quantitative. Firstly, qualitative forecasting is an estimation methodology that uses expert judgment rather than numerical analysis. This type of forecasting depends on the knowledge of highly experienced employees and consultants to supply intuition into future outcomes. In this situation, there are some methods such as executive opinions, Delphi technique, sales force composite and consumer or market surveys. Secondly, quantitative forecasting techniques make formal use of historical data and a forecasting model (Montgomery et al., 2015). Namely, types of this method are based on the ground of mathematical (numerical) models. They rely heavily on mathematical computations.

Quantitative forecasting models can be classed into two categories. These are time series models and associative models. Associative models (often called causal models) suppose that the variable being forecasted is associated to other variables in the environment. On the other hand, time series models concern about past patterns of data and attempt to predict to future relied on the underlying patterns included within those data. Although there are several time series models such as naïve method, simple mean, simple moving average, weighted moving average, exponential smoothing, Box-Jenkins methodology etc., exponential smoothing method is one of the most prevalent techniques in forecasting due to its simplicity, robustness and accuracy as automatic forecasting procedures. Furthermore, the major conclusion of M3-competition was that simple methods such as exponential smoothing outperformed the sophisticated ones, in other words statistically sophisticated or complex methods did not provide more accurate forecasts than simpler ones (Makridakis & Hibon, 2000).

3.1 Box-Jenkins Methodology

The Box-Jenkins methodology is very useful in forecasting future values of a time series having values that are statistically dependent on or related to each other. This method consists of four basic steps: identification, estimation, diagnostic checking and forecasting. In the first step; a tentative model is identified by analysis of the historical data. In the second step; the unknown parameters of the tentative model are estimated. In the third step; diagnostic checks are accomplished to test the adequacy of the model, and if need be, to suggest potential improvements. When a time series model has been developed, a fourth step, called forecasting, generates predictions of future values of the time series (O'Connell & Bowerman, 1979).

The Box-Jenkins methodology presents a more systematic approach to building, analyzing and forecasting with time series models. However, the Box-Jenkins methodology has some shortcomings. First, at least 50 and preferably 100 observations are needed to establish a good Box-Jenkins model. Thus the Box-Jenkins methodology has frequently been implemented to time series that generate hourly, daily or weekly observations. Another drawback of the Box-Jenkins methodology is that there are no available automatic procedures to either update the estimates of the model parameters as new data are observed or to identify and adapt to changes in the underlying process generating the time series. A last disadvantage of the Box-Jenkins methodology is that, since the model building process can be rather complex and since a large amount (50 to 100 observations) of reliable data must be obtained, it can be somewhat time consuming and costly use. We are not managing that a forecasting methodology should not be used because it is somewhat expensive and time consuming.

3.2 Exponential Smoothing Method

Exponential smoothing was proposed in the late 1950s by Brown (1959), Holt (1957) and Winters (1960). It has gained successful achievements among forecasting methods. Their popularity is because of several practical considerations in short-range forecasting (Gardner, 1985). Furthermore, exponential smoothing is probably the most widely used class of procedures for smoothing discrete time series in order to forecast the immediate future (Montgomery et al., 1990).

Exponential smoothing is a forecasting method that weights the observed time series values unequally. Moreover, these weights diminish exponentially as the observations get older i.e. $\alpha, \alpha(1-\alpha), \alpha(1-\alpha)^2$. For example, if the smoothing constant α is 0.1, then these coefficients are 0.1, 0.09, and 0.081 and so on (Bowerman & O'Connell, 1979). More recent observations are weighted more heavily than more remote observations. The unequal weighting is accomplished by using one or more smoothing constant, which determine how much weight is given to each observation. Exponential smoothing has been found to be the most effective when the parameters describing the time series may be changing slowly over time. The main aim in these methods involves using a sequence of observations on some variable to predict a future value of it (Bowerman & O'Connell, 1993).

Exponential smoothing (ES) is a special sort of moving average that does not require keeping a long historical record in the active file and thus cuts down on the data processing time required (Brown, 1959).

Exponential smoothing methods are the most prevailing techniques for obtaining forecasts because of their simplicity, robustness and accuracy. However; their popularity in time series analysis is not only as a result of simplicity but also their proven superiority against more elaborated approaches (Makridakis & Hibon, 2000). This technique is known as a method that is able to produce good forecasting results. There are number of empirical studies and forecasting competitions that prove the usefulness of the exponential smoothing method. For example; Makridakis et al. (1993) performed the M2-competition to determine the post sample accuracy of various forecasting methods. Their results suggest that exponential smoothing method in particular; damped and single smoothing has the most accurate results (Osman & King, 2015).

Exponential smoothing models are of more or less three components: level, trend and seasonality. Furthermore, trend component has five types which are none, additive, multiplicative, damped additive and damped multiplicative. Similarly, seasonality component has three types which are none, additive and multiplicative. Hence; there are 15 different ES models, particularly simple exponential smoothing (SES) (no trend, no seasonality) and Holt's linear model (additive trend, no seasonality) and Holt-Winter's additive model (additive trend, additive seasonality) are the best known models in forecasting approach. In this study, compared the consequences of these best known models, just simple exponential smoothing and Holt-Winters method was mentioned in this section. In addition, Table 3.1 demonstrates that some basic exponential smoothing models' equations including best known models.

Table 3.1 Standard additive methods for exponential smoothing

Trend	Seasonality: None	Seasonality: Additive
N None	***SES*** $S_t = \alpha X_t + (1-\alpha)S_{t-1}$ $\hat{X}_t(h) = S_t$	$S_t = \alpha(X_t - I_{t-p}) + (1-\alpha)S_{t-1}$ $I_t = \beta(X_t - S_t) + (1-\beta)I_{t-p}$ $\hat{X}_t(h) = S_t + I_{t-p+h}$
N None	$S_t = S_{t-1} + \alpha e_t$ $\hat{X}_t(h) = S_t$	$S_t = S_{t-1} + \alpha e_t$ $I_t = I_{t-p} + \beta(1-\alpha)e_t$ $\hat{X}_t(h) = S_t + I_{t-p+h}$
A Additive	***Holt's Linear Model*** $S_t = \alpha X_t + (1-\alpha)(S_{t-1} + T_{t-1})$ $T_t = \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1}$ $\hat{X}_t(h) = S_t + hT_t$	$S_t = \alpha(X_t - I_{t-p}) + (1-\alpha)(S_{t-1} + T_{t-1})$ $T_t = \beta(S_t - S_{t-1}) + (1-\beta)T_{t-1}$ $I_t = \beta(X_t - S_t) + (1-\beta)I_{t-p}$ $\hat{X}_t(h) = S_t + hT_t + I_{t-p+h}$
A Additive	$S_t = S_{t-1} + T_{t-1} + \alpha e_t$ $T_t = T_{t-1} + \alpha\beta e_t$ $\hat{X}_t(h) = S_t + hT_t$	***Holt-Winters Additive Model*** $S_t = S_{t-1} + T_{t-1} + \alpha e_t$ $T_t = T_{t-1} + \alpha\beta e_t$ $I_t = I_{t-p} + \beta(1-\alpha)e_t$ $\hat{X}_t(h) = S_t + hT_t + I_{t-p+h}$

3.2.1 Simple Exponential Smoothing

The simplest of the exponentially smoothing method is fundamentally called simple exponential smoothing. In some books, it is called single exponential smoothing. The simple exponential smoothing model is one of the most popular forecasting methods that we use to forecast the next period for a time series that have no trend and seasonality pattern. This popularity can be referred to its simplicity, its computational efficiency, the ease of adjusting its responsiveness to changes in the process being forecast, and its reasonable accuracy.

The idea of the exponential smoothing is to smooth the original series the way the moving averages does and to use the smoothed series in forecasting future values of the variable of interest.

This method is used when data pattern is approximately horizontal i.e. there is neither cyclic variation that is seasonal cycle nor pronounced trend in the historical data. Formally, the simple exponential smoothing equation takes the form of;

$$S_t = S_{t-1} + \alpha(X_t - S_{t-1}) \tag{3.1}$$

or

$$S_t = \alpha X_t + (1-\alpha)S_{t-1} \tag{3.2}$$

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

where S_t is the smoothed value for period *t*, $\hat{X}_t(h)$ is the forecast value for *h* periods ahead from origin *t*, X_t is the actual value for period *t* and α is the smoothed constant where $0 \le \alpha \le 1$.

The forecast $\hat{X}_t(h)$ is based on weighting the most recent observation S_t with a weight α and weighting the most recent smoothed value S_{t-1} with a weight of 1-α.

Even though the simple exponential smoothing method is the most prevalent forecasting technique in literature, it has some drawbacks that affect the quality of the forecasts obtaining using this method. The one of the most widespread shortcoming in using *SES* is that selecting an initial value. So as to initiate the algorithm, we need an initial forecast, an actual value and a smoothing constant. Since S_1 is not known, we can either set the first estimate that is equal to the first observation and to use $X_1 = S_0$ or use the average of the first five or six observations for the initial smoothed value. The other disadvantage is that this method's aim is to assign comparatively more weight to recent observations compared to older ones (Yapar et al., 2017).

Smoothing constant α is a selected number between zero and one, i.e. $0 \leq \alpha \leq 1$. When $\alpha = 1$, the original and smoothed version of the series is identical. On the other hand; when $\alpha = 0$ the series is smoothed flat.

Rewriting the equation (3.1) to draw attention one of the neat things about *SES* model,

$$S_t - S_{t-1} = \alpha(X_t - S_{t-1}) \tag{3.4}$$

change in forecasting value is proportionate to the forecast error. That is:

$$S_t = S_{t-1} + \alpha e_t \tag{3.5}$$

where residual $e_t = X_t - S_{t-1}$ is forecast error for time period *t*.

By continuing to substitute previous forecasting values back to the starting point of the data in equation (3.2), we receive:

$$S_t = \alpha X_t + (1-\alpha)\left[\alpha X_{t-1} + (1-\alpha)S_{t-2}\right]$$

$$S_t = \alpha X_t + \alpha(1-\alpha)X_{t-1} + (1-\alpha)^2 S_{t-2}$$

$$S_t = \alpha X_t + \alpha(1-\alpha)X_{t-1} + \alpha(1-\alpha)^2 X_{t-2} + (1-\alpha)^3 S_{t-3}$$

$$S_t = \alpha X_t + \alpha(1-\alpha)X_{t-1} + \alpha(1-\alpha)^2 X_{t-2} + \alpha(1-\alpha)^3 X_{t-3} + (1-\alpha)^4 S_{t-4}$$

The forecast equation in general form is

$$S_t = \alpha X_t + \alpha(1-\alpha)X_{t-1} + \alpha(1-\alpha)^2 X_{t-2} + \cdots + \alpha(1-\alpha)^{t-2} X_2$$

$$+ \alpha(1-\alpha)^{t-1} X_1 + (1-\alpha)^t S_0$$

$$S_t = \alpha \sum_{i=0}^{t-1} (1-\alpha)^i X_t + (1-\alpha)^t S_0 \tag{3.6}$$

where S_0 is the initial value of the variable X at time period *t* from knowledge of the actual series' values X_t, X_{t-1}, X_{t-2} and so on back in time to the first known value of the time series X_1. Therefore, S_t represents a weighted moving average of all past observations with weights decreasing exponentially.

The series of weights used in producing the smoothed value S_t is α, $\alpha(1-\alpha)$, $\alpha(1-\alpha)^2$,…. These weights decline toward zero exponentially, so it can be said that for sufficiently large α recent observations get more weight. Thus as we go back in the series, each value has a smaller weight than following values in terms of its effect on the forecast. Then, the weights from *SES* satisfy the following conditions.

- $w_t \in [0,1] \qquad t = 1, \dots, n$
- $\sum_{t=1}^{n} w_t = 1$
- $w_1 \le w_2 \le \dots \le w_n$ (3.7)

Weights assigned by simple exponential smoothing are non-negative and increasingly go on from first one to last one. These weights can be represented as $\alpha(1-\alpha)^k$ for the weight that is given k^{th} periods in the exponential smoothing process. If α is small, more weight is given to observations from the more distant past. On the other hand; if α is large, more weight is given to the more recent observations. Hence; for different α values, the corresponding weights are given in Table 3.2.

Table 3.2 The corresponding weights for different smoothing levels

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

After the model is specified, then its performance characteristics should be verified or confirmed by comparison of its forecast with historical data. In this situation, we can utilize the error measures such as MAPE (mean absolute percentage error), MSE (mean square error) and MAD (mean absolute deviation).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|e_i|}{X_i}\cdot 100\% \tag{3.8}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}e_i^2 \tag{3.9}$$

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|e_i| \tag{3.10}$$

In addition to these error measures, there is a new popular metric that is used for measuring a smoothing method's ability to utilize fresh data is the average age and it is denoted by AA (Yapar, 2016). The first connection between the simple moving average (SMA) and the exponential moving average (EMA) along with the $\frac{2}{(p+1)}$ conversion formula was presented in Brown's book (1962). Furthermore, simple exponential smoothing (SES) and simple moving averages (SMA) are often used in practice to handle intermittent demand (Syntetos & Boylan, 2005). Later, the conversion formula became broadly adapted by technical analysts after the publication of Jack Hutson's article (1984).

The AA for SES is:

$$\bar{k} = 0\alpha + 1\alpha(1-\alpha) + 2\alpha(1-\alpha)^2 + \cdots$$

$$\bar{k} = \alpha\sum_{k=o}^{\infty}k(1-\alpha)^k$$

$$\bar{k} = \frac{1-\alpha}{\alpha} \tag{3.11}$$

One way to define an exponential smoothing system that is equivalent to a p period moving average is to say that the smoothing constant is selected to give the same average age of the data (Brown, 1962).

$$\frac{1-\alpha}{\alpha} = \frac{p-1}{2} \qquad \Rightarrow \qquad \alpha = \frac{2}{p+1} \tag{3.12}$$

Now, we can easily calculate the expected value and variance of the smoothed value S_t Firstly, for sufficiently large *t*, the expected value of S_t is:

$$E(S_t) = \left(\alpha \sum_{k=0}^{\infty} (1-\alpha)^k X_{t-k} + (1-\alpha)^t S_0 \right)$$

$$= \alpha \sum_{k=0}^{\infty} (1-\alpha)^k E(X_{t-k})$$

$$= E(X) \alpha \sum_{k=0}^{\infty} (1-\alpha)^k$$

$$= \alpha \frac{\alpha}{\alpha}$$

$$= \alpha \tag{3.13}$$

where $(1-\alpha)^k S_0 \to 0$ when $k \to \infty$ and $0 \leq \alpha \leq 1$.

So, S_t is unbiased estimator of the constant α when $t \to \infty$. Hence, S_t can be utilized for future forecasts. Then the variance of S_t is

$$V(S_t) = V\left(\alpha \sum_{k=0}^{\infty} (1-\alpha)^k X_{t-k} + (1-\alpha)^t S_0\right)$$

$$= \alpha^2 \sum_{k=0}^{\infty} V(X_{t-k})$$

$$= V(X)\alpha^2 \sum_{k=0}^{\infty} \left((1-\alpha)^k\right)^2$$

$$= \sigma_\varepsilon^2 \alpha^2 \frac{1}{1-(1-\alpha)^2}$$

$$= \frac{\alpha^2}{2\alpha - \alpha^2} \sigma_\varepsilon^2$$

$$= \frac{\alpha}{2-\alpha} \sigma_\varepsilon^2 \qquad (3.14)$$

For large sample sizes, S_t is unbiased estimator since $\sum_{t=1}^{n} w_t = 1$ and has the average age $AA = \bar{k} = \frac{1-\alpha}{\alpha}$ and the variance $V = \frac{\alpha}{2-\alpha}$. In addition, Table 3.3 demonstrates the weights assigned to observations for diversified smoothing levels while using *SES*.

Table 3.3 Weights attached to actual values by several smoothing parameters

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

3.2.2 Holt's Linear Trend Method

Holt (1957) widened simple exponential smoothing to authorize forecasting of data with trend. To account for a trend component in the time series, Holt's linear trend model (or it is called sometimes double exponential smoothing) incorporates a second smoothing constant, β. This extended method involves three equations created a forecast. First one is that smooth the time series in equation 3.15, second one is that to smooth the trend and last one is to combine these equations to obtain the forecast in equation 3.16 and 3.17 respectively.

$$S_t = \alpha X_t + (1-\alpha)(S_{t-1} + T_{t-1}) \tag{3.15}$$

$$T_t = \beta\left(S_t - S_{t-1}\right) + (1-\beta)T_{t-1} \tag{3.16}$$

$$\hat{X}_t(h) = S_t + hT_t \tag{3.17}$$

where X_t is the actual observation, $\hat{X}_t(h)$ is the *h*-step-ahead forecast, α and β are smoothing parameters for level and trend respectively, $0 < \alpha, \beta < 1$. S_t is the smoothed constant process value and T_t is the smoothed trend value for period *t*.

As with single exponential smoothing, selecting starting values for S_t and T_t is critical, as well as values for α and β. Recall that these processes are judgmental and constants closer to a value of 1.0 are chosen when less smoothing is desired (and more weight placed on recent values), whereas constants closer to 0.0 when more smoothing is desired (and less weight placed on recent values).

Hence, we mentioned above that well-known special case of exponential smoothing methods i.e. simple exponential smoothing and Holt's linear trend method. In addition to these methods, Yapar et al. (2016) proposed Modified Holt's linear trend method. This method's aim is to eliminate some important issues faced ES models that are initialization and optimization problems. Then, Holt's linear trend method is modified as follows:

$$S_t = \left(\frac{p}{t}\right)X_t + \left(\frac{t-p}{t}\right)\left(S_{t-1} + T_{t-1}\right) \tag{3.18}$$

$$T_t = \left(\frac{q}{t}\right)\left(S_t - S_{t-1}\right) + \left(\frac{t-q}{t}\right)T_{t-1} \tag{3.19}$$

where $n \geq p \geq q \geq 0$.

$$\hat{X}_t = S_t + hT_t \tag{3.20}$$

for $p \in \{1,...,n\}$ and $q \in \{0,1,...,n\}$. The first equation is smooth equation in 3.18 and the second one is trend equation in 3.19 and finally the last one is forecast equation in 3.20. In intermittent demand forecasting, if there are trend component in non-zero values of intermittent demand data, then this model can be used.

Summarizing the properties of exponential smoothing method is following:

- Exponential smoothing methods are recursive, that is, rely upon all observations in the time series. The weight on each observation diminishes exponentially more distant in the past.
- Smoothing constants are utilized to assign weights between 0 and 1 to the most recent observations. The closer constant is to 0.0, the more smoothing that occurs and the lighter weight assigned to the most recent observation; while the closer constant is to 1.0, the less smoothing that happens and the heavier the weight assigned to the most recent observation.
- When no discernible trend is exhibited in the data, single exponential smoothing is appropriate; when a trend is present in the time series, Holt's linear trend model is necessary.
- Exponential smoothing methods require you to generate starting forecasts for the first period in the time series. Deciding on those initial forecasts, as well as on the values of your smoothing constants α and β are arbitrary. You need to base your judgments on your experience in the business, in addition to some experiments.
- Exponential smoothing methods are best used for short-term forecasting.
- Exponential smoothing models do not forecast well when the time series pattern (e.g., level of sales) is suddenly, drastically and permanently altered by some event or change of course or action. When these situations take place, a new model will be necessary.

Consequently, the standard method used for forecasting intermittent demand is single exponential smoothing (Croston, 1972; Willemain et al., 1994). In addition; the most adopted technique for short-term forecasting is probably the single exponential smoothing according to Brown (1959, 1962). It is simple to use compared to other sophisticated forecast methods, given that it requires only two pieces of data which are the last forecast and the observation of the latest period. It is also flexible in that the used smoothing constant, can be easily changed. In practice, exponential smoothing is often used when dealing with intermittent demand. Though exponential smoothing places more weight on the most recent data, resulting, in the case of intermittence, in a series of estimates that are highest just after a demand occurrence and lowest just before demand occurs again (Syntetos & Boylan; 2001). Today *SES* is still a widespread method in different computer systems and forecasting programs. However, the ability for *SES* to forecast on an intermittent demand, when for an item the forecasting periods often have zero demand, has been questioned (Wallström & Segerstedt; 2010). Moreover; in 1972 Croston developed a method for forecasting in intermittent demand situations which he showed to be superior to single exponential smoothing and also he proved the unsuitability of exponential smoothing as a forecasting method when dealt with intermittent demands. Furthermore Johnston (1980) suggested the inadequacies of exponential smoothing on intermittent data become apparent when the mean interval between transactions is greater than two time periods. Besides, Croston (1972) showed that his method is unbiased, whereas the single exponential smoothing is not. In addition, he showed that his method has lower variance than that of single exponential smoothing (Sani & Kingsman, 1997). Furthermore, these facts were ratified by Johnston (1980) and Willemain et al. (1994). Thus, the conventional time series methods such as exponential smoothing method did not perform well and not suitable for forecasting intermittent demand (Kourentzes, 2013). So, we will investigate thoroughly Croston's method which is more appropriate for intermittent demand forecasting.

3.3 Croston's Method

Forecasting for intermittent demand has long been recognized as a very difficult business task. Many authors have recognized and contributed to the additional problems of forecasting caused by infrequent and irregular demand patterns. In an early and important paper, Croston (1972) highlighted the inadequacies of exponentially weighted moving averages to estimate the underlying demand pattern. For these item lines, the observed demand during many periods would be zero, interspersed by occasional periods with irregular non-zero demand (Johnston & Boylan, 1996). Hence, traditional forecasting methods particularly exponential smoothing method is often inappropriate for products with erratic (or intermittent) demand (Snyder, 2002).

Croston (1972) proposed a method that builds demand estimates taking into account both demand size and the interval between demand incidences, then later it was verified by Rao (1973). Croston method is a classical method that specifically dealing with intermittent demand, it was developed to base upon the simple exponential smoothing. The main change that Croston introduced was that the forecast is renewed only when there is demand and not when the forecast time interval has passed like ordinary exponential smoothing. Croston's method not just focused on the size of the order, his model also takes the time between orders into consideration. This made the model suitable for forecasting items that have intermittent demand patterns (Levén & Segerstedt; 2004). It was advocated by Croston (1972) and then later as corrected by Rao (1973) that it is proved that exponential smoothing as a forecasting method when dealing with intermittent demands is not appropriate.

Croston method is a forecasting approach which was enhanced to supply a more accurate estimate for items with intermittent demands (Vasumathi & Saradha, 2013). In this case, Willemain et al. (1994) suggested that Croston's method provides more accurate forecasts of the mean demand per period than exponential smoothing. Indeed, this method developed by Croston for forecasting in intermittent demand is superior to single exponential smoothing (Sani & Kingsman, 1997), because this data can be included zero values.

The Croston's method comprises of two main parts which decompose the original intermittent series into two separate series. The first one contains all non-zero demand sizes, while the second one involves the respective intervals between two successive non-zero demands. The Croston method does differentiate between these two elements. Using exponential smoothing, the demand size and the demand interval are separately updated after each period with a positive demand. The ratio of the former over the latter then provides a forecast for the future demand per period (Teunter et al., 2011). When demand occurs every period namely time periods contain no zero demand, Croston's method is as same as the conventional exponential smoothing method. Most of the forecasting and inventory control theory like Croston's method is built on the assumption that consecutive demand values are independent (Altay et al., 2012).

Under a situation of infrequent (intermittent) transactions, it is often preferable to forecast two separate components of the demand process:

- The interval between consecutive transactions (n).
- The magnitude of individual transactions (Z).

The aim being to estimate the mean demand per period as Z/n, where Z is the mean demand size and n is the mean interval between transactions (Eaves, 2002).

The method developed by Croston (1972), and corrected by Rao (1973), separately applies exponential smoothing to the interval between demands and the size of the demands (Willemain et al., 2004). When the demand is stable, the aim of this method is to estimate the mean demand per period. Updating only occurs at moments of positive demand; if a period has no demand, the method simply increments the count of time periods since the last demand.

Croston's method notations are given below:

X_t: demand for an item at time t

n_t: number of periods since last transactions (or number of periods between nonzero demands)

$\hat{n}_t$: estimate of time intervals

$\hat{Z}_t$: estimate of demand

α: smoothing parameter for number of demand

β: smoothing parameter for number of periods

$\hat{X}_t$: forecast value for one period ahead

- If there are no transactions (i.e. $X_t = 0$), then;

$$\hat{Z}_t = \hat{Z}_{t-1} \tag{3.21}$$

$$\hat{n}_t = \hat{n}_{t-1} \tag{3.22}$$

- If there are any transactions (i.e. $X_t \neq 0$), in other words demand is greater than zero, then;

$$\hat{Z}_t = \alpha X_t + (1-\alpha)\hat{Z}_{t-1} \tag{3.23}$$

$$\hat{n}_t = \beta n_t + (1-\beta)\hat{n}_{t-1} \tag{3.24}$$

Then forecast equation is given as follows:

$$\hat{X}_t = \frac{\hat{Z}_t}{\hat{n}_t} \tag{3.25}$$

Combining the estimates of size and interval provides an estimate of the mean demand per period.

An initial value problem of this method is unimportant. Because contrary to exponential smoothing method, choosing an initial value is rather easy. Croston method chooses a first non-zero demand from an intermittent demand as an initial value of demand. Similarly, an initial value of a time interval is the time periods up to first non-zero demand.

It is necessary to examine and evaluate the performances of the forecasts. Before considering accuracy measures for the purpose of comparing alternative methods in an intermittent context, it would be wise to determine which ones can be computed, taking into account that there is some zero demand in time periods. Silver et al. (1998) point out; “no single measure is universally best”. Still, evaluations of forecasting performances are done using only one measure of the forecasting errors. The most common measures are mean absolute deviation (MAD) or mean squared error (MSE) (Wallström & Segerstedt; 2010). This metrics are well suited to intermittent demand series because they never give infinite or undefined results. Nonetheless all relative to the series accuracy measures (e.g., MAPE and MdAPE that is meant respectively absolute percentage error and median absolute percentage error) must be excluded because actual demand may be zero. But the symmetric mean absolute percentage error (sMAPE) can be calculated because the sum of actual and forecast value of a demand cannot be zero. Lastly, an absolute error measures are calculated as a function of the forecast errors alone: there are no ratio calculations with respect to the series itself or to other forecasting methods.

One possible exception to the exclusion of absolute measures may be the mean error (ME), also known as the mean signed error. The mean error is defined as:

$$ME = \frac{\sum_{t=1}^{n} X_t - \hat{Z}_t}{n} = \frac{\sum_{t=1}^{n} e_t}{n} \tag{3.26}$$

where X_t is an actual demand at time *t*, $\hat{Z}_t$ is an estimate of demand in period *t*, e_t is a forecast error in period *t* and *n* is the number of demand time periods. The symmetric MAPE (sMAPE) is defined as below:

$$sMAPE = \frac{2}{n}\sum_{t=1}^{n}\frac{\left|\hat{Z}_t - X_t\right|}{\hat{Z}_t + X_t}100\% \tag{3.27}$$

The method of Croston (1972) for forecasting intermittent demand series is increasing in popularity. The method is incorporated in statistical forecasting software packages (e.g., Forecast Pro) and demand planning modules of component based enterprise and manufacturing solutions (e.g., Industrial and Financial Systems). It is also included in integrated real-time sales and operations planning processes (Boylan & Syntetos; 2007).

This method's forecasting accuracy and inventory performance has been widely accepted by Willemain et al. (1994), Johnston & Boylan (1996), Gardner (2005) and Boylan & Syntetos (2006). Moreover Croston method was alleged to be unbiased but despite its theoretical superiority modest benefits were registered in the literature (Syntetos et al., 2009). Nevertheless, Croston method has been criticized by Willemain et al (1994), while its assumption of the independence between demands and intervals has been challenged (Shenstone & Hyndman, 2005; Kourentzes, 2013). Also, Syntetos & Boylan (2001) proved that Croston's method is biased with the magnitude of bias being positively correlated with the value of the smoothing parameter (Petropoulos & Kourentzes, 2014). Then Levén & Segerstedt (2004) proposed to Modified Croston's method by using earlier work by Segerstedt (2000). Such method was only the modification of the original Croston's method. However; proposed a modification to Croston's method by Levén & Segerstedt (2004) was found by Syntetos & Boylan (2005) to be more biased than the original estimator. They subsequently proposed an approximately unbiased estimator: SBA (Syntetos-Boylan Approximation, Syntetos & Boylan, 2005). Such modified version is recognized as the Syntetos – Boylan Approximation (SBA). Afterwards, Shale et al. (2006) derived the expected bias in Croston's method and proposed an "exact" correction factor, SBJ. The first estimation procedure (SBA) performs a deflating factor to the Croston estimates in order to take away the bias. But a little bias though still remains, on the opposite side. This procedure was further evaluated by Teunter & Sani (2009) and it was found to perform very well. Moreover, according to their findings, Croston's original method presents smaller bias if few demands are zero, whereas SBA modification has a smaller bias if many demands are zero (Petropoulos et al., 2013). Furthermore, the modification considerably enhances the performance of the original Croston method, the Syntetos-Boylan variation (Teunter & Duncan, 2009). However, there are cases and some (negative) bias remains while the Syntetos – Boylan Approximation (SBA) may be more biased than the original Croston method (Wallström & Segerstedt, 2010; Teunter & Sani, 2009; Teunter et al., 2011). Hence, Teunter et al. (2011) proposed a novel method which is TSB (Teunter-Syntetos-Babai) method. This method that is originally not proposed for intermittent demand, but does update the forecast after every period, is the naïve method that uses the last observed demand as the forecast for future periods (Teunter et al., 2011). Nevertheless, Teunter et al. (2011) suggested that the use of their new method as an alternative to Croston and SBA methods, but their findings are only relied upon simulated data (Petropoulos et al., 2013). Although there are several modifications of original Croston method, all of them built on the same concept that is separating the non-zero demands from the intervals or the probability to have a non-zero demand (Petropoulos et al., 2016). Thus, we will investigate the Croston method as well as its two variations SBA and SBJ methods in chapter 5 and also ATA method which will mention in the next chapter.

CHAPTER FOUR
ATA METHOD AND ITS APPLICATIONS

The development of accurate and robust forecasting methods for univariate time series is very important when large numbers of time series are involved in the modeling and forecasting process. When considered all methods for forecasting time series, there are still drawbacks for accurate extrapolation. In this case, forecasting competition has played a critical role in moving towards the forecasting of large numbers of time series. Recently, proposed new technique ATA method will appeal a wide range of attention by its simplicity, easy optimization and surprisingly good performance (Yapar et al., 2017; Selamlar, 2017). The novel method ATA can be performed non-seasonal time series or seasonalized time series which is applied by classical decomposition technique.

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

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

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

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

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

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

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

$$S_t = \left(\frac{p}{t}\right)X_t + \left(\frac{t-p}{t}\right)(S_{t-1} \times T_{t-1}) \tag{4.4}$$

$$T_t = \left(\frac{q}{t}\right)\left(\frac{S_t}{S_{t-1}}\right) + \left(\frac{t-q}{t}\right)T_{t-1} \tag{4.5}$$

$$\hat{X}_t(h) = S_t \times T_t^h \tag{4.6}$$

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

It is worth pointing out that ATA method does not require to initial values for level and trend patterns. Because this method is developed by inspiration of the ES models, it can be also adjusted to ES models which can be classified into 30 different models (Hyndman et al., 2008).

4.1 Simplest form of $ATA(p, 0)$

This form occurs when $q = 0$, *ATA*(*p*, *q*) reduces to a simple model that has similar form to simple exponential smoothing, i.e. for $t > p$:

$$S_t = \left(\frac{p}{t}\right) X_t + \left(\frac{t-p}{t}\right) S_{t-1}, \qquad t > p \tag{4.7}$$

$$S_t = X_t, \qquad t \le p \tag{4.8}$$

$$\hat{X}_t(h) = S_t \tag{4.9}$$

where p is the smoothing parameter and it organizes the smoothing process. The model will be called the simple form of $ATA(p,0)$. S_t can be interpreted as a weighted average of past observations. Furthermore, the smoothed value S_t can be written as the recursive form:

$$S_t = \sum_{k=0}^{t-(p+1)} \frac{\binom{t-k-1}{p-1}}{\binom{t}{p}} X_{t-k} + \frac{1}{\binom{t}{p}} S_p\,, \tag{4.10}$$

where S_p is the starting or initial value for $ATA(p,0)$ which can be simply the p^{th} observation or the average of the oldest p observations. It can be easily seen that the smoothed value at time t is a weighted average of all past observations and the initial value S_p. From the equation (4.10), it is evident that the initial value of S_p relies on the smoothing parameter of p. Thus, the smoothing parameter and initial value are optimized at the same time (Selamlar, 2017).

4.1.1 Weights of ATA (p, 0)

The weights of the *ATA* method which has some basic characteristic are as same as that of *SES* model. These properties as follows are given in equation (3.7).

- $w_t \in [0,1] \quad t = 1, \cdots, n$
- $w_1 \le w_2 \le \ldots \le w_n$

- $$\sum_{t=1}^{n} w_t = 1. \tag{4.11}$$

The weights attached to observations are consecutively given in Table 4.1.

Table 4.1 The weights attached to the observations by $ATA(p,0)$

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

4.1.2 Average age of ATA (*p, 0*)

The average age (*AA*) of a model is the measure of the model's ability to use fresh data. It is represented by $\bar{k}$. The smaller $\bar{k}$ is far better to compare the model. The equation of the average age metric is defined by Brown (1959) as follows:

$$AA(\hat{a}) = \bar{k} = n - \sum_{t=1}^{n} t w_t \tag{4.12}$$

where w_t is the weigh for given to the t^{th} observation while trying to obtain the forecast. For the simplest model of *ATA,* the average age is follows:

$$AA_{ATA(p,0)} = \frac{n-p}{p+1} \tag{4.13}$$

When compared to the simplest form of *SES* and *ATA* methods at the same α level (where $\alpha = p/n$), it is rather clear that the average age of *ATA* is smaller than that of *SES*, i.e. $\bar{k}_{ATA} < \bar{k}_{SES}$. Because,

$$\bar{k}_{ATA} = \frac{n-p}{p+1} < \frac{(1-\alpha)}{\alpha} = \frac{n-p}{p} = \bar{k}_{SES} \tag{4.14}$$

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

Among the important metrics, the variance must be considered in order to compare forecasting methods. The formulation of the variance is as below:

$$Var(\hat{a}) = E\left[\left(\sum_{t=1}^{n} w_t X_t - a\right)^2\right] = \sum_{t=1}^{n} w_t^2 \sigma^2 = V\sigma^2 \tag{4.15}$$

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

$$V_{ATA(p,0)} = \sum_{t=1}^{n} w_t^2 \tag{4.16}$$

where the weights are given in Table 4.1. If the weighted values are substituted in equation (4.16), we obtain

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

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

$$= \left(\frac{p}{n}\right)^2 {}_3F_2((1, p-n, p-n), (1-n, 1-n), 1) \quad (4.17)$$

From the equation (4.17), it can be seen that the variance of the *ATA* (*p, 0*) estimator involves the Generalized Hyper-Geometric series:

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

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

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

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

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

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

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

As it is mentioned before, *ATA* method does not require an initial value which from the equation as follows:

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

$$S_p = X_t, \qquad \text{where} \quad t \le p$$

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

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

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

4.1.5 Smoothing Parameter of ATA (p, 0)

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

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

$$S_t = \left(\frac{1}{t}\right) X_t + \left(\frac{1}{t}\right) X_{t-1} + \ldots + \left(\frac{1}{t}\right) X_2 + \left(\frac{1}{t}\right) S_1, \tag{4.18}$$

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

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

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

$$S_t = \frac{2}{t} X_t + \frac{2}{t}\frac{(t-2)}{(t-1)} X_{t-1} + \frac{2}{t}\frac{(t-3)}{(t-1)} X_{t-2} + \ldots + \frac{2}{t}\frac{2}{(t-1)} X_3 + \frac{2}{t(t-1)} S_2, \tag{4.19}$$

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

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

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

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

$$S_t = \left(\frac{3}{t}\right)X_t + \left(\frac{t-3}{t}\right)S_{t-1}\,, \qquad t>3. \tag{4.20}$$

If we expand the formula iteratively, we get

$$S_4 = \left(\frac{3}{4}\right)X_4 + \left(\frac{1}{4}\right)S_3 \tag{4.21}$$

$$S_5 = \left(\frac{3}{5}\right)X_5 + \left(\frac{2}{5}\right)\left(\frac{3}{4}\right)X_4 + \left(\frac{2}{5}\right)\left(\frac{1}{4}\right)S_3 \tag{4.22}$$

$$S_6 = \left(\frac{3}{6}\right)X_6 + \left(\frac{3}{6}\right)\left(\frac{3}{5}\right)X_5 + \left(\frac{3}{6}\right)\left(\frac{2}{5}\right)\left(\frac{3}{4}\right)X_4 + \left(\frac{3}{6}\right)\left(\frac{2}{5}\right)\left(\frac{1}{4}\right)S_3 \tag{4.23}$$

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

$$\begin{aligned} S_{10} &= 0.300X_{10} + 0.233X_9 + 0.175X_8 + 0.125X_7 + 0.083X_6 \\ &\quad + 0.050X_5 + 0.025X_4 + 0.008S_3 \end{aligned} \tag{4.24}$$

where $S_3 = X_3$.

4.2 The Trended ATA Methods

ES models consist of a family of models which assume that the time series has roughly three underlying data components: level, trend, seasonality. Each model contains one of five types of trend (none, additive, multiplicative, damped additive and damped multiplicative) and one of three types of seasonality (none, additive and multiplicative). When different combinations of trend and seasonality are regarded, 15 different ES models can be generated. The best known of these are *SES* (no trend, no seasonality), Holt's linear model (additive trend, no seasonality), and Holt-Winters' additive model (additive trend, additive seasonality). Proposed by Hyndman et al. (2008) ETS state space models consider 30 different forecasting models according to error, trend and seasonality components. Because Holt's additive method's performance is quite better than Holt Winters' seasonal model, we will not deal with seasonal models. So, we will consider with trended (additive, multiplicative, damped additive and damped multiplicative) models using ATA method with deseasonalized data.

4.2.1 ATA (p, q) with Additive Trend

For the trended series $X_t,\ t = 1, \dots, n$, 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 components. This method has two smoothing equations which account for level S_t and trend T_t and also forecasting equations as follows:

$$S_t = \left(\frac{p}{t}\right) X_t + \left(\frac{t-p}{t}\right)\left(S_{t-1} + T_{t-1}\right) \tag{4.25}$$

$$T_t = \left(\frac{q}{t}\right)\left(S_t - S_{t-1}\right) + \left(\frac{t-q}{t}\right) T_{t-1} \tag{4.26}$$

$$\hat{X}_t(h) = S_t + hT_t \tag{4.27}$$

for $p \in \{1, \ldots, n\}$, $q \in \{0,1, \ldots, p\}$ and $n > p \geq q$. For $n \leq p$ and $S_t = X_t$, for $n \leq q$ and $T_t = X_t - X_{t-1}$ and then $T_1 = 0$. S_t denotes an estimate of the level of the series and also 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 *SES*, i.e. for $n > p$:

$$S(t) = \left(\frac{p}{t}\right) X_t + \left(\frac{t-p}{t}\right) S_{t-1} \tag{4.28}$$

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

As with *SES*, the level equation in (4.25) shows that S_t is a weighted average of the observation at time *t* and the with in-sample one-step-ahead forecast for time $t-1$, here given by $S_{t-1} + T_{t-1}$. The trend equation shows that T_t is a 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 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.

4.2.2 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 which account for level S_t and trend T_t component.

$$S_t = \left(\frac{p}{t}\right) X_t + \left(\frac{t-p}{t}\right) \left(S_{t-1} * T_{t-1}\right) \tag{4.29}$$

$$T_t = \left(\frac{q}{t}\right) \left(\frac{S_t}{S_{t-1}}\right) + \left(\frac{t-q}{t}\right) T_{t-1} \tag{4.30}$$

$$\hat{X}_t(h) = S_t * T_t^h \tag{4.31}$$

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

CHAPTER FIVE

COMPARISON OF ATA AND CROSTON METHOD FOR INTERMITTENT DEMAND DATA

In this study we have dealt with an intermittent demand including 1,500 different data set. Also each set has monthly 78 in-sample and 6 out-sample observations from M4-competition inventory data set which is retrieved from https://www.m4.unic.ac.cy/the-dataset/. There are 1,324 data set with starting zero but 176 data set with initiating non-zero. Each observation of this data has been carried out at R-Studio for using ATA method with "ATAforecasting" package and for using Croston method and its derivation SBA and SBJ (that is; Syntetos-Boylan-Approximation and Shale-Boylan-Johnston methods) with "tsintermittent" package. The performance of these methods is compared for all inventory data set. Moreover some of the most suitable accuracy measures for an intermittent demand which are MAE (mean absolute error), MSE (mean squared error) and sMAPE (symmetric mean absolute percentage error) are conducted on this data set since this data include zero values. Because intermittent data include zero values, other accuracy measures such as MAPE cannot be calculated. Since denominator never include zero values. Consequently, the forecast values and result of accuracy measures are examined.

In this chapter, we have discussed all inventory data particularly first data whose name is I0001, because first data starts zero. Table 5.1 presents first data set from inventory data sequentially row by row. Furthermore Table 5.2 shows in-sample periods, non-zero demand sizes and inter-demand intervals for I0001 data. In addition Table 5.3 manifests that the R-codes for Croston method both default codes in R and used this data set. Moreover, Table 5.4 demonstrates in-sample forecasts for Croston, SBA, SBJ and ATA methods using I0001 data from M4-competition data set taking into consideration forecast accuracy measure MAE. Also Table 5.5 expresses that one-step ahead forecast value for I0001 data with its error measures, weights and initial values for demand and inter-demand intervals using MAE. Similarly; Table 5.6 shows in-sample forecasts for Croston, SBA, SBJ and ATA methods using I0001 data from M4-competition data set taking into consideration forecast accuracy measure MSE and also Table 5.7 indicates that one-step ahead forecast value using MSE for I0001 data with its error measures, weights and initial values for demand and inter-demand intervals.

5.1 Formulation of Croston and ATA Methods

For intermittent demand, used techniques are different from conventional methods such as exponential smoothing. Although when there are any transaction, the formula are as the same as that of exponential smoothing, while there are no transaction then utilized formula can be changed. Because, the formula differ from if there are any demand. Hence, used formula for forecasting of intermittent demand are given as the below with regard to Croston and ATA methods respectively.

For Croston Method:

- If there are any transactions, then the demand and inter-demand interval equations for Croston method are below in equation 5.1 and 5.2 respectively.

$$\hat{Z}_t = \alpha X_t + (1-\alpha)\hat{Z}_{t-1} \tag{5.1}$$

$$\hat{n}_t = \beta n_t + (1-\beta)\hat{n}_{t-1} \tag{5.2}$$

- If there are no transactions (i.e. $X_t = 0$), then the demand and inter-demand interval equations for Croston method are below in equation 5.3 and 5.4 respectively.

$$\hat{Z}_t = \hat{Z}_{t-1} \tag{5.3}$$

$$\hat{n}_t = \hat{n}_{t-1} \tag{5.4}$$

Finally, the forecast equation of per period demand is below for Croston method in equation 5.5:

$$\hat{X}_t = \frac{\hat{Z}_t}{\hat{n}_t} \tag{5.5}$$

For ATA Method:

- If there are any transactions, then the demand equations are below in equation 5.6 and 5.7 and inter-demand interval equations for ATA method are below in equation 5.8 and 5.9 respectively.

$$\hat{Z}_t = \left(\frac{p}{t}\right) X_t + \left(\frac{t-p}{t}\right) \hat{Z}_{t-1} \;, \qquad \text{for} \quad t > p \qquad (5.6)$$

$$\hat{Z}_t = X_t \qquad \text{for} \quad t \leq p \qquad (5.7)$$

$$\hat{n}_t = \left(\frac{q}{t}\right) n_t + \left(\frac{t-q}{t}\right) \hat{n}_{t-1} \qquad \text{for} \quad t > q \qquad (5.8)$$

$$\hat{n}_t = n_t \qquad \text{for } t \leq q \qquad (5.9)$$

- If there are no transactions (i.e. $X_t = 0$), then the demand and inter-demand interval equations for ATA method are below in equation 5.10 and 5.11 respectively.

$$\hat{Z}_t = \hat{Z}_{t-1} \qquad (5.10)$$

$$\hat{n}_t = \hat{n}_{t-1} \qquad (5.11)$$

Finally, the forecast equation of per period demand is below for ATA method in equation 5.12:

$$\hat{X}_t = \frac{\hat{Z}_t}{\hat{n}_t} \qquad (5.12)$$

5.2 Initialization for Croston and ATA Methods

In literature, the most prevalent forecasting technique is simple exponential smoothing method for its simplicity, robustness etc., but it has some drawbacks that affect the quality of the forecasts obtaining using this method. The one of the most widespread shortcoming in using *SES* is that selecting an initial value. So as to get started the algorithm, we need an initial forecast, an actual value and a smoothing constant. Since S_1 is not known, we can either set the first estimate equal to the first observation and then to use $X_1 = S_0$ or use the average of the first five or six observations for the initial smoothed value. The other disadvantage is that this method's aim is to assign comparatively more weight to recent observations compared to older ones (Yapar et al., 2017).

For Croston method, an initial value problem is quite easy, contrary to exponential smoothing method. Croston method chooses a first non-zero demand from an intermittent demand as an initial value of demand. Similarly, an initial value of a time interval is the time periods from starting points to first non-zero demands, namely it is the time periods up to first non-zero demands.

It is worth pointing out that ATA method does not require to initial values for level and trend patterns. Because this method is developed by inspiration of the ES models, it can be also adjusted to ES models which can be classified into 30 different models (Hyndman et al., 2008). It can be easily seen that the smoothed value at time t is a weighted average of all past observations and the initial value. From the equation (5.6), it is evident that the initial value relies on the smoothing parameter of p. Thus, the smoothing parameter and initial value are optimized at the same time (Selamlar, 2017). ATA method uses the p^{th} value as the initial value where p is the optimum smoothing parameter. Here X_t is the value of original series, $\hat{Z}_t$ is the smoothed value of demand at time t for $t > p$ in equation 5.6 and for $t \leq p$ in equation 5.7. Likewise; $\hat{n}_t$ is the smoothed value of inter-demand interval at time t for $t > q$ in equation 5.8 and for

$t \leq q$ in equation 5.9. Finally; $\hat{X}_t(h)$ is the h step ahead forecast value in equation 5.12.

Table 5.1 Original demand sizes with 78 in-sample and 6 out-sample (with red filling) observations row by row from left to right sequentially for I0001 data

0	0	15	0	50	0	0	0	0	0	0	0
0	130	20	0	0	17	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	150	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	10	0	0	0	0	0	0
0	0	35	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0

Table 5.2 In-sample periods and non-zero demand sizes and inter-demand intervals for I0001 data

Period (in-sample)	Non – zero demand	Inter – demand intervals
3	15	3
5	50	2
14	130	9
15	20	1
18	17	3
34	150	16
54	10	20
63	35	9

In R-Studio, in order to obtain optimum smoothing parameters α and β; we use "tsintermittent" package for Croston Method and used codes in R-Studio is given Table 5.3;

Table 5.3 R-output of Croston method and its variations

R-output	
Original code in default	crost(data , h, w, init, nop, type, cost, init.opt, outplot opt.on, na.rm)
Croston method's code	fit1<-crost(data, h=6, w=NULL, init = c("naive"), nop = c(2), type = c("croston"), cost = c("mae"), init.opt = c(FALSE), outplot = c(TRUE), opt.on = c(FALSE), na.rm = c(TRUE))

In original data for I0001, $X_1 = X_2 = 0$ and the 1st non-zero demand is occurred in 3rd period (that is $X_3 = 15$). So, the Croston method, SBA and SBJ approximations give the first in-sample forecast at 4th point. Namely, they foresee the third period's demand at the fourth period. Some of the applications for I0001 data are given below using optimum smoothing parameters calculated by R-Studio.

For optimum smoothing parameter $\alpha = 0.0000000054$ which is obtained in R-Studio on "tsintermittent" package for accuracy measure MAE used by for Croston Method for I0001 data, some of estimate demands are below;

Using equation 5.1; $\hat{Z}_4 = \alpha X_3 + (1-\alpha)\hat{Z}_3$, where $X_3 = \hat{Z}_3 = 15$.

Continuing the above equation;

$\hat{Z}_4 = (5.4 * 10^{-9}) * 15 + 0.9999999946 * 15 = 15$,

Using equation 5.3; $\hat{Z}_5 = \hat{Z}_4 = 15$, since at 4th point does not include a demand in the original data, 5th period's demand equal to that of 4th point.

Using equation 5.1; $\hat{Z}_6 = \alpha X_5 + (1-\alpha)\hat{Z}_5$

Continuing the above equation;

$\hat{Z}_6 = (5.4 * 10^{-9}) * 50 + 0.9999999946 * 15 = 15$

Lastly using equation 5.3; $\hat{Z}_{79} = \hat{Z}_{78} = \cdots = \hat{Z}_{64}$, since the last non-zero observation is occurred at 63rd period.

$$\hat{Z}_{79} = \hat{Z}_{64} = (5.4*10^{-9})*35 + 0.9999999946*15 = 15$$

For optimum smoothing parameter $\beta = 0.4999$, which is calculated in R-Studio on "tsintermittent" package for accuracy measure MAE used by for Croston Method for I0001 data, some of estimate inter-demand intervals are below;

Using equation 5.2; $\hat{n}_4 = \beta n_3 + (1-\beta)\hat{n}_3$, where $\hat{n}_3 = n_3 = 3$ is an initial value of inter-demand interval.

Using equation 5.2; $\hat{n}_4 = 0.4999*3 + 0.5001*3 = 3$,

Using equation 5.4; $\hat{n}_5 = \hat{n}_4 = 3$, since at 4th point does not include a demand in the original data, 5th period's time interval equal to that of 4th point.

Similarly using equation 5.2 again;

$$\hat{n}_6 = 0.4999*2 + 0.5001*3 = 2.500088$$

Lastly using equation 5.4; $\hat{n}_{79} = \hat{n}_{78} = \cdots = \hat{n}_{64}$, since the last non-zero observation is occurred at 63rd period.

$$\hat{n}_{79} = \hat{n}_{64} = 0.4999*9 + 0.5001*14.7954 = 11.8982$$

In order to obtain forecast value for Croston method, we use equation 5.5;

$$\hat{X}_4 = \frac{\hat{Z}_4}{\hat{n}_4} = \frac{15}{3} = 5 = \hat{X}_5$$

$$\hat{X}_6 = \frac{\hat{Z}_6}{\hat{n}_6} = \frac{15}{2.500088} = 5.9998$$

Finally $\hat{X}_{79} = \frac{\hat{Z}_{79}}{\hat{n}_{79}} = \frac{15}{11.8982} = 1.260693$

The formulations of sMAPE and MAE for single forecast value are below:

$$sMAPE = \frac{\hat{X}_t - X_t}{\left(\hat{X}_t + X_t\right)/2} 100\% \tag{5.9}$$

$$MAE = \left|\hat{X}_t - X_t\right| \tag{5.10}$$

Using the formula 5.9 for Croston Method for 1st out – sample from I0001 data:

$$sMAPE_1 = \frac{1.2607 - 0}{(1.2607 + 0)/2} = \frac{1.2607}{1.2607/2} 100\% = 200 \tag{5.11}$$

Using the formula 5.10 for Croston Method for 1st out – sample from I0001 data:

$$MAE_1 = |1.2607 - 0| = 1.2607 \tag{5.12}$$

Table 5.4 In-sample forecasts for I0001 data with periods for all methods using MAE

PERIOD	ORIGINAL	CROSTON	SBA	SBJ	ATA	PERIOD	ORIGINAL	CROSTON	SBA	SBJ	ATA
1	0	-	-	-	-	40	0	1.5637	0.4688	0.00000004	11.6612
2	0	-	-	-	-	41	0	1.5637	0.4688	0.00000004	11.6612
3	15	-	-	-	5.0000	42	0	1.5637	0.4688	0.00000004	11.6612
4	0	5.0000	2.5000	0.00000018	5.0000	43	0	1.5637	0.4688	0.00000004	11.6612
5	50	5.0000	2.5000	0.00000018	11.6667	44	0	1.5637	0.4688	0.00000004	11.6612
6	0	5.9998	3.7500	0.00000028	11.6667	45	0	1.5637	0.4688	0.00000004	11.6612
7	0	5.9998	3.7500	0.00000028	11.6667	46	0	1.5637	0.4688	0.00000004	11.6612
8	0	5.9998	3.7500	0.00000028	11.6667	47	0	1.5637	0.4688	0.00000004	11.6612
9	0	5.9998	3.7500	0.00000028	11.6667	48	0	1.5637	0.4688	0.00000004	11.6612
10	0	5.9998	3.7500	0.00000028	11.6667	49	0	1.5637	0.4688	0.00000004	11.6612
11	0	5.9998	3.7500	0.00000028	11.6667	50	0	1.5637	0.4688	0.00000004	11.6612
12	0	5.9998	3.7500	0.00000028	11.6667	51	0	1.5637	0.4688	0.00000004	11.6612
13	0	5.9998	3.7500	0.00000028	11.6667	52	0	1.5637	0.4688	0.00000004	11.6612
14	130	5.9998	3.7500	0.00000028	12.6829	53	0	1.5637	0.4688	0.00000004	11.6612
15	20	2.6089	0.8333	0.00000007	12.9687	54	10	1.5637	0.4688	0.00000004	10.0250
16	0	4.4442	7.5000	0.00000060	12.9687	55	0	1.0138	0.3750	0.00000003	10.0250
17	0	4.4442	7.5000	0.00000060	12.9687	56	0	1.0138	0.3750	0.00000003	10.0250
18	17	4.4442	7.5000	0.00000060	12.2845	57	0	1.0138	0.3750	0.00000003	10.0250
19	0	4.7057	2.5000	0.00000020	12.2845	58	0	1.0138	0.3750	0.00000003	10.0250
20	0	4.7057	2.5000	0.00000020	12.2845	59	0	1.0138	0.3750	0.00000003	10.0250
21	0	4.7057	2.5000	0.00000020	12.2845	60	0	1.0138	0.3750	0.00000003	10.0250
22	0	4.7057	2.5000	0.00000020	12.2845	61	0	1.0138	0.3750	0.00000003	10.0250
23	0	4.7057	2.5000	0.00000020	12.2845	62	0	1.0138	0.3750	0.00000003	10.0250
24	0	4.7057	2.5000	0.00000020	12.2845	63	35	1.0138	0.3750	0.00000003	9.6916
25	0	4.7057	2.5000	0.00000020	12.2845	64	0	1.2607	0.8333	0.00000007	9.6916
26	0	4.7057	2.5000	0.00000020	12.2845	65	0	1.2607	0.8333	0.00000007	9.6916
27	0	4.7057	2.5000	0.00000020	12.2845	66	0	1.2607	0.8333	0.00000007	9.6916
28	0	4.7057	2.5000	0.00000020	12.2845	67	0	1.2607	0.8333	0.00000007	9.6916
29	0	4.7057	2.5000	0.00000020	12.2845	68	0	1.2607	0.8333	0.00000007	9.6916
30	0	4.7057	2.5000	0.00000020	12.2845	69	0	1.2607	0.8333	0.00000007	9.6916
31	0	4.7057	2.5000	0.00000020	12.2845	70	0	1.2607	0.8333	0.00000007	9.6916
32	0	4.7057	2.5000	0.00000020	12.2845	71	0	1.2607	0.8333	0.00000007	9.6916
33	0	4.7057	2.5000	0.00000020	12.2845	72	0	1.2607	0.8333	0.00000007	9.6916
34	150	4.7057	2.5000	0.00000020	11.6612	73	0	1.2607	0.8333	0.00000007	9.6916
35	0	1.5637	0.4688	0.00000004	11.6612	74	0	1.2607	0.8333	0.00000007	9.6916
36	0	1.5637	0.4688	0.00000004	11.6612	75	0	1.2607	0.8333	0.00000007	9.6916
37	0	1.5637	0.4688	0.00000004	11.6612	76	0	1.2607	0.8333	0.00000007	9.6916
38	0	1.5637	0.4688	0.00000004	11.6612	77	0	1.2607	0.8333	0.00000007	9.6916
39	0	1.5637	0.4688	0.00000004	11.6612	78	0	1.2607	0.8333	0.00000007	9.6916

Table 5.5 One-step ahead forecast, weights and initials for demand and intervals and error measures for I0001 data using MAE

	CROSTON	SBA	SBJ	ATA
forecasts	1.26069300000	0.83333340000	0.00000007234	9.69160000000
weight Z	0.000000005	0.000000003	0.009790862	0.026315790
weight n	0.499912200	1.000000000	0.999999982	0.013157895
initial Z	15	15	15	5
initial n	3	3	3	1
sMAPE	196.530	197.370	200.000	**188.790**
MAE	7.845	6.891	**5.693**	14.061
MSE	568.955	577.406	588.520	**560.167**

Table 5.6 In-sample forecasts for I0001 data with periods for all methods using MSE

PERIOD	ORIGINAL	CROSTON	SBA	SBJ	ATA	PERIOD	ORIGINAL	CROSTON	SBA	SBJ	ATA
1	0	-	-	-	-	40	0	2.6164	4.7259	4.9271	11.6612
2	0	-	-	-	-	41	0	2.6164	4.7259	4.9271	11.6612
3	15	-	-	-	5.0000	42	0	2.6164	4.7259	4.9271	11.6612
4	0	5.0000	3.1992	3.0112	5.0000	43	0	2.6164	4.7259	4.9271	11.6612
5	50	5.0000	3.1992	3.0112	11.6667	44	0	2.6164	4.7259	4.9271	11.6612
6	0	7.7589	9.0470	7.4825	11.6667	45	0	2.6164	4.7259	4.9271	11.6612
7	0	7.7589	9.0470	7.4825	11.6667	46	0	2.6164	4.7259	4.9271	11.6612
8	0	7.7589	9.0470	7.4825	11.6667	47	0	2.6164	4.7259	4.9271	11.6612
9	0	7.7589	9.0470	7.4825	11.6667	48	0	2.6164	4.7259	4.9271	11.6612
10	0	7.7589	9.0470	7.4825	11.6667	49	0	2.6164	4.7259	4.9271	11.6612
11	0	7.7589	9.0470	7.4825	11.6667	50	0	2.6164	4.7259	4.9271	11.6612
12	0	7.7589	9.0470	7.4825	11.6667	51	0	2.6164	4.7259	4.9271	11.6612
13	0	7.7589	9.0470	7.4825	11.6667	52	0	2.6164	4.7259	4.9271	11.6612
14	130	7.7589	9.0470	7.4825	12.6829	53	0	2.6164	4.7259	4.9271	11.6612
15	20	3.4724	7.2221	7.1795	12.9687	54	10	2.6164	4.7259	4.9271	10.0250
16	0	9.6868	11.9499	9.3844	12.9687	55	0	1.7405	1.8343	1.9970	10.0250
17	0	9.6868	11.9499	9.3844	12.9687	56	0	1.7405	1.8343	1.9970	10.0250
18	17	9.6868	11.9499	9.3844	12.2845	57	0	1.7405	1.8343	1.9970	10.0250
19	0	8.4414	7.4691	6.9655	12.2845	58	0	1.7405	1.8343	1.9970	10.0250
20	0	8.4414	7.4691	6.9655	12.2845	59	0	1.7405	1.8343	1.9970	10.0250
21	0	8.4414	7.4691	6.9655	12.2845	60	0	1.7405	1.8343	1.9970	10.0250
22	0	8.4414	7.4691	6.9655	12.2845	61	0	1.7405	1.8343	1.9970	10.0250
23	0	8.4414	7.4691	6.9655	12.2845	62	0	1.7405	1.8343	1.9970	10.0250
24	0	8.4414	7.4691	6.9655	12.2845	63	35	1.7405	1.8343	1.9970	9.6916
25	0	8.4414	7.4691	6.9655	12.2845	64	0	2.8206	2.4095	2.2646	9.6916
26	0	8.4414	7.4691	6.9655	12.2845	65	0	2.8206	2.4095	2.2646	9.6916
27	0	8.4414	7.4691	6.9655	12.2845	66	0	2.8206	2.4095	2.2646	9.6916
28	0	8.4414	7.4691	6.9655	12.2845	67	0	2.8206	2.4095	2.2646	9.6916
29	0	8.4414	7.4691	6.9655	12.2845	68	0	2.8206	2.4095	2.2646	9.6916
30	0	8.4414	7.4691	6.9655	12.2845	69	0	2.8206	2.4095	2.2646	9.6916
31	0	8.4414	7.4691	6.9655	12.2845	70	0	2.8206	2.4095	2.2646	9.6916
32	0	8.4414	7.4691	6.9655	12.2845	71	0	2.8206	2.4095	2.2646	9.6916
33	0	8.4414	7.4691	6.9655	12.2845	72	0	2.8206	2.4095	2.2646	9.6916
34	150	8.4414	7.4691	6.9655	11.6612	73	0	2.8206	2.4095	2.2646	9.6916
35	0	2.6164	4.7259	4.9271	11.6612	74	0	2.8206	2.4095	2.2646	9.6916
36	0	2.6164	4.7259	4.9271	11.6612	75	0	2.8206	2.4095	2.2646	9.6916
37	0	2.6164	4.7259	4.9271	11.6612	76	0	2.8206	2.4095	2.2646	9.6916
38	0	2.6164	4.7259	4.9271	11.6612	77	0	2.8206	2.4095	2.2646	9.6916
39	0	2.6164	4.7259	4.9271	11.6612	78	0	2.8206	2.4095	2.2646	9.6916

Table 5.7 One-step ahead forecast, weights and initials for demand and intervals and error measures for I0001 data using MSE

	CROSTON	SBA	SBJ	ATA
forecasts	2.8206	2.4095	2.2646	9.6916
weight Z	0.0703	0.4924	0.4343	0.0128
weight n	0.7494	0.7203	0.5691	0.0128
initial Z	15	15	15	5
initial n	3	3	3	1
sMAPE	194.85	193.49	193.82	**188.79**
MAE	**9.3554**	9.7181	9.4827	14.0608
MSE	572.3682	554.6642	**551.4314**	560.1674

In Table 5.4 and Table 5.6; in-sample forecast values for I0001 data are given using MAE and MSE metrics for optimization respectively. Then, Tables 5.5 and 5.7 give some information that one-step ahead forecast, weights and initials for both demands and intervals and also some error metrics for I0001 data. On the other hand; Table 5.8 and Table 5.9 indicate that out-sample sMAPE values for all inventory data set which include 1500 different data that each one includes 78 monthly observations, using MAE and MSE for optimization of the parameters respectively. We have clearly seen that ATA method is better than the others as far as out-sample sMAPE values for all data set are concerned.

Table 5.8 Out-sample sMAPE values for all inventory data set of M4-competition (from I0001 to I1500 data) using MAE metrics

	1	**2**	**3**	**4**	**5**	**6**	**1--2**	**1--4**	**1--6**
Croston	1.9715	1.9667	1.9645	1.9557	1.9625	1.9731	1.9691	1.9646	1.9657
SBA	1.9757	1.9715	1.9687	1.9643	1.9677	1.9765	1.9736	1.9701	1.9707
SBJ	1.9908	1.9911	1.9861	1.9875	1.9901	1.9919	1.9910	1.9889	1.9896
ATA	**1.9420**	**1.9424**	**1.9306**	**1.9268**	**1.9283**	**1.9558**	**1.9422**	**1.9355**	**1.9377**

Table 5.9 Out-sample sMAPE values for all inventory data set of M4-competition (from I0001 to I1500 data) using MSE metrics

	1	**2**	**3**	**4**	**5**	**6**	**1--2**	**1--4**	**1--6**
Croston	1.9596	1.9590	1.9510	1.9462	1.9486	1.9628	1.9593	1.9540	1.9545
SBA	1.9637	1.9640	1.9559	1.9505	1.9538	1.9673	1.9639	1.9585	1.9592
SBJ	1.9667	1.9596	1.9620	1.9560	1.9567	1.9693	1.9682	1.9636	1.9634
ATA	**1.9419**	**1.9421**	**1.9302**	**1.9260**	**1.9286**	**1.9535**	**1.9420**	**1.9351**	**1.9371**

When examining the tables from Table 5.10 to Table 5.13, there are average out-sample sMAPE values for all inventory data set of M4-competition using all methods mentioned in this thesis. Besides, these tables contain both intermittent and lumpy demands' consequences for 1500 different data set. We have seen that there are nearly no differences between intermittent and lumpy demand consequences. Because these results are too close to each other. However; considering the best results for average out-sample sMAPE values, this is unarguably consequences of ATA method, because ATA method's results are quite smaller than the others.

Table 5.10 Average out-sample sMAPE values for all inventory data set of M4-competition using ATA method

ATA METHOD						
Satır Etiketleri	Ortalama smape1	Ortalama smape2	Ortalama smape3	Ortalama smape4	Ortalama smape5	Ortalama smape6
Intermittent	1.95	1.95	1.93	1.93	1.93	1.96
Lumpy	1.93	1.94	1.93	1.92	1.93	1.94
Genel Toplam	**1.94**	**1.94**	**1.93**	**1.93**	**1.93**	**1.96**

Table 5.11 Average out-sample sMAPE values for all inventory data set of M4-competition using Croston method

CROSTON METHOD						
Satır Etiketleri	Ortalama smape1	Ortalama smape2	Ortalama smape3	Ortalama smape4	Ortalama smape5	Ortalama smape6
Intermittent	1.98	1.97	1.96	1.96	1.96	1.98
Lumpy	1.97	1.97	1.97	1.96	1.96	1.96
Genel Toplam	**1.97**	**1.97**	**1.96**	**1.96**	**1.96**	**1.97**

Table 5.12 Average out-sample sMAPE values for all inventory data set of M4-competition using SBA method

SBA METHOD						
Satır Etiketleri	Ortalama smape1	Ortalama smape2	Ortalama smape3	Ortalama smape4	Ortalama smape5	Ortalama smape6
Intermittent	1.98	1.97	1.97	1.96	1.97	1.98
Lumpy	1.97	1.97	1.97	1.96	1.97	1.97
Genel Toplam	**1.98**	**1.97**	**1.97**	**1.96**	**1.97**	**1.98**

Table 5.13 Average out-sample sMAPE values for all inventory data set of M4-competition using SBJ method

SBJ METHOD						
Satır Etiketleri	Ortalama smape1	Ortalama smape2	Ortalama smape3	Ortalama smape4	Ortalama smape5	Ortalama smape6
Intermittent	1.99	1.99	1.99	1.99	1.99	1.99
Lumpy	1.99	1.99	1.98	1.99	1.99	1.99
Genel Toplam	**1.99**	**1.99**	**1.99**	**1.99**	**1.99**	**1.99**

CHAPTER SIX
SUMMARY AND CONCLUSION

In this thesis, we were concerned with producing forecasts for 1,500 different data sets with an intermittent demand. Also each set has monthly 78 in-sample and 6 out-sample observations from M4-competition inventory data set which can be downloaded from https://www.m4.unic.ac.cy/the-dataset/. There are 1,324 data sets with starting zero values but 176 data sets with initiating non-zero. In order to obtain accurate forecasts for these data sets, we study the essential forecasting approaches such as Box-Jenkins methodology and exponential smoothing method in addition to Croston method and its two variations: SBA and SBJ and finally ATA method. Some of the most suitable approaches for an intermittent demand are chosen so as to obtain the best consequences. These are Croston method and its two variations SBA and SBJ and also ATA method. Though Box-Jenkins methodology and exponential smoothing methods are by far the most popular forecasting approaches, they give big errors on intermittent demand data. So, we do not choose these methods. Actually though both Croston and ATA method are developed by inspired from exponential smoothing method; these two methods give better results than exponential smoothing since ES is generally appropriate for continuous time series, but intermittent data is discrete.

In this thesis, we have fundamentally dealt with two methods: Croston method that is the most comprehensive method for an intermittent demand data and ATA method that is new proposed method when inspired of exponential smoothing in the literature. These two methods, Croston and ATA have similar formula for intermittent demands; however used smoothing parameter and initial values in forecasting are different from each other. Firstly; ATA method uses p value for smoothing constant of a demand and also utilizes q value for smoothing constant of an inter-demand interval. Moreover according to p and q values, ATA method identifies the initial value's position of data. Secondly; Croston method uses α and β values for smoothing constant of demand and inter-demand interval respectively. These values are determined in R-Studio or other statistical programs. In addition; Croston method uses the first non-zero demand as an initial value of a demand and also uses the time interval up to first non-zero demand as an initial value of inter-demand interval. Likewise; ATA method utilizes the first non-zero demand as an initial value of a demand and also uses the time interval between starting point to first non-zero demand as an initial value of inter-demand interval too.

In chapter 5; when Croston method is compared to its two variations, when the in sample optimization is carried out based on MAE and the out-sample accuracy is measured in terms of sMAPE, Croston method is better than the others. Contrary to MAE; when MSE is used for obtaining the optimum parameters, SBA and SBJ are better than the Croston method again for sMAPE. Since these three methods' results have changed with respect to some error metrics, all of them have been examined in this thesis. However; ATA method's results do not change according to the error metric used for optimization, i.e. MAE or MSE, and its results are quite better than those of the other methods.

Consequently; ATA method, Croston method and its two variations: SBA and SBJ have been compared relied on popular metrics that are commonly used when compared forecasting methods in previous chapter. Furthermore, some periods' demand and inter-demand interval estimates and some error metrics are calculated by hand using equations of ATA and Croston method for optimum smoothing parameter obtained by R-Studio. In-sample forecasts for I0001 data using these methods for optimization parameters as to MAE and MSE metrics are given in Table 5.4 and Table 5.6 respectively. Besides one-step ahead forecasts, weights, initial values and some error metrics for MAE and MSE metrics are given in Table 5.5 and Table 5.7 respectively. The results obtained from this comparison in chapter 5 say that ATA method gives more accurate forecasts than while comparing the Croston and its variation methods. Because while considering out-sample sMAPE values for all inventory data set of M4-competition regarding both Table 5.8 and 5.9, ATA method's results are quite smaller than the other methods: Croston method, SBA and SBJ. Furthermore; when examining tables from Table 5.10 to Table 5.13, there are average out-sample sMAPE values for all inventory data set of M4-competition utilizing all methods mentioned in this thesis: ATA, Croston, SBA and SBJ as for that both intermittent and lumpy demand. Then we have clearly seen that there are no considerable differences between intermittent and lumpy demand consequences. However; while analyzing thoroughly Tables 5.10, 5.11, 5.12 and 5.13; ATA method’s results are better than the others, because these consequences are rather small. Therefore; we have seen that among these examined methods, ATA method gives the best results for both MAE and MSE metrics taking into account the smallest sMAPE values for the M4-competition inventory data set.

REFERENCES

Altay, N., Litteral, L. A., & Rudisill, F. (2012). Effects of correlation on intermittent demand forecasting and stock control. *Int. J. Production Economics, 135,* 275-283.

Boylan, J. E., & Syntetos, A. A. (2006). Accuracy and accuracy-implication metrics for intermittent demand. *The International Journal of Applied Forecasting, 4,* 39-42.

Boylan, J. E., & Syntetos, A. A. (2007). The accuracy of a modified Croston procedure. *Int. J. Production Economics, 107,* 511-517.

Bowerman, B. L., & O'Connell, R. T. (1979). *Forecasting & time series.* California: Wadsworth.

Bowerman, B. L., & O'Connell, R. T. (1993). *Forecasting and time series: an applied approach* (3rd edition). California: Wadsworth.

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

Brown, R. G. (1962). *Smoothing, forecasting and prediction of discrete time series* (4th edition), Englewood Cliffs, New Jersey: Prentice-Hall.

Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly, 23*(3), 289-303.

Deloitte (2011). *The Service Revolution in Global Manufacturing Industries.* New York: Deloitte Research.

Dunsmuir, W. T. M., & Snyder, R. D. (1989). Control of inventories with intermittent demand. *European Journal of Operational Research*, *40,* 16-21.

Eaves, A. H. C. (2002). *Forecasting for the ordering and stock-holding of consumable spare parts.* Phd Thesis, University of Canterbury, New Zealand.

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

Gardner, E. S. (2005). Exponential smoothing: The state of the art-Part II. *International Journal of Forecasting, 22*(4), 637-666.

Ghobbar, A. A., & Friend, C. H. (2003). Evaluation of forecasting methods for intermittent parts demand in the field of aviation: A predictive model. *Computers & Operations Research*, *30*, 2097-2114.

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

Hua, Z. S., Zhang, B., Yang, J., & Tan, D. S. (2007). A new approach of forecasting intermittent demand for spare parts inventories in the process industry. *The Journal of the Operational Research Society*, *58*(1), 52-61.

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

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

Johnston, F. R. (1980). An interactive stock control system with a strategic management role. *The Journal of Operational Research Society, 31,* 1069-1084.

Johnston, F. R., & Boylan, J. E. (1996). Forecasting with items with intermittent demand. *The Journal of the Operational Research Society*, *47*(1), 113-121.

Johnston, F. R., Boylan, J. E., & Shale, E., A. (2003). An examination of the size of orders from customers, their characterization and the implications for inventory control for slow moving items. *The Journal of the Operational Research Society, 54,* 833-837.

Kourentzes, N. (2013). Intermittent demand forecasts with neural networks. *Int. J. Production Economics, 143,* 198-206.

Levén, E., & Segerstedt, A. (2004). Inventory control with a modified Croston procedure and Erlang distribution. *Int. J. Production Economics, 90,* 361-367.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K, & Simmons, L., F. (1993). The M2-competition: A real-time judgmentally based forecasting study. *International Journal of Forecasting*, *9*(1), 5-22.

Makridakis, S., & Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting, 16,* 451-476.

Montgomery, D. C., Johnson, L. A., & Gardner, J. S. (1990). *Forecasting & time series analysis* (2nd edition). Singapore: McGraw-Hill.

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting* (2nd edition). New Jersey: John Wiley & Sons.

Osman, A. F., & King, M. L. (2015). *A new approach to forecasting based on exponential smoothing with independent regressors,* Retrieved January 5, 2018, from http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Petropoulos, F., Nikolopoulos, K., Spithourakis, G. P., & Assimakopoulos, V. (2013). Empirical heuristics for improving intermittent demand forecasting. *Ind. Manag. Data Syst., 113*(5), 683-696.

Petropoulos, F., & Kourentzes, N. (2014). Forecast combinations for intermittent demand. *Journal of the Operational Research Society, 66,* 914-924.

Petropoulos, F., Kourentzes, N., & Nikolopoulos, K. (2016). Another look at estimators for intermittent demand. *Int. J. Production Economics, 181,* 154-161.

Sani, B., & Kingsman, B. G. (1997). Selecting the best periodic inventory control and demand forecasting methods for low demand items. *The Journal of the Operational Research Society, 48*(7), 700-713.

Schultz, C. R. (1987). Forecasting and inventory control for sporadic demand under periodic review. *Journal of Operational Research Society, 38,* 453-458.

Shale, E. A., Boylan, J. E. & Johnston, F. R. (2006). Forecasting for intermittent demand: the estimation of an unbiased estimate. *Journal of the Operational Research Society, 57*(5), 588-592.

Shenstone, L., & Hyndman, R. J. (2005). Stochastic models underlying Croston's method for intermittent demand forecasting. *Journal of Forecasting, 24*(6), 389-402.

Silver, E., A. (1981). Operations research in inventory management: a review and critique. *Operations Research, 29,* 628-645.

Silver, E. A., Pyke, D. F., & Peterson, R. (1998). *Inventory management and production planning and scheduling* (3rd ed.). New York: John Wiley & Sons.

Smart, C. (2009). *Understanding intermittent demand forecasting solutions*. Retrieved February 2, 2018 from http://demand-planning.com/2009/10/08/understanding-intermittent-demand-forecasting-solutions/

Snyder, R. (2002). Forecasting sales of slow and fast moving inventories. *European Journal of Operational Reseach, 140,* 684-699.

Snyder, R. D., Ord, J. K., & Beaumont, A. (2012). Forecasting the intermittent demand for slow-moving inventories: A modelling approach. *International Journal of Forecasting, 28*(2), 485-496.

Syntetos, A. A. (2001). *Forecasting of intermittent demand*. Phd Thesis, Buckinghamshire Chilterns University College, Brunel University, London.

Syntetos, A. A., & Boylan, J. E. (2001). On the bias of intermittent demand estimates. *Int. J. Production Economics, 71,* 457-466.

Syntetos, A. A., & Boylan, J. E. (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting, 21,* 3 03-314.

Syntetos, A. A., Babai, M. Z., Dallery, Y., & Teunter, R. (2009). Periodic control of intermittent demand items: Theory and empirical analysis, *The Journal of the Operational Research Society, 60*(5), 611-618.

Syntetos, A. A., & Boylan, J. E. (2010). On the variance of intermittent demand estimates. *Int. J. Production Economics, 128,* 546-555.

Syntetos, A. A., Babai, M. Z., & Gardner, Jr. E. S. (2015). Forecasting intermittent inventory demands: Simple parametric methods vs. bootstrapping. *Journal of Business Research, 68,* 1746-1752.

Teunter, R. H., & Duncan, L. (2009). Forecasting intermittent demand: A comparative study. *The Journal of the Operational Research Society, 60*(3), 321-329.

Teunter, R. H., & Sani, B. (2009). On the bias of Croston's forecasting method. *European Journal of Operational Research, 194*(1), 177-183.

Teunter, R. H., Syntetos, A. A., & Babai, M. Z. (2011). Intermittent demand: Linking forecasting to inventory obsolescence, *European Journal of Operational Research, 214,* 606-615.

Vasumathi, B., & Saradha, A. (2013). Forecasting intermittent demand for spare parts. *International Journal of Computer Applications, 75*(11), 12-16.

Vereecke, A., & Verstraeten, P. (1994). An inventory management model for an inventory management model for an inventory consisting of lumpy items, slow movers and fast movers. *Int. J. Production Economics, 35,* 379-389.

Wallström, P., & Segerstedt, A. (2010). Evaluation of forecasting error measurements and techniques for intermittent demand. *Int. J. Production Economics, 128,* 625-636.

Ward, J. B. (1978). Determining re-order points when demand is lumpy. *Management Science, 24,* 623-632.

Willemain, T. R., Smart, C. N., Shocker, J. H., & Desautels, P. A. (1994). Forecasting intermittent demand in manufacturing: a comparative evalution of Croston's method. *International Journal of Forecasting, 10,* 529-538.

Willemain, T. R., Smart, C. N., & Schwarz, H. F. (2004). A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting, 20,* 375-387.

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

Yapar, G. (2016). Modified simple exponential smoothing, *Hacettepe University Journal of Mathematics and Statistics*, Doi:10.15672/HJMS.201614320580.

Yapar, G., Capar, S., Selamlar, H. T., & Yavuz, I. (2016). Modified Holt's linear trend method. *Hacettepe University Journal of Mathematics and Statistics* Doi: 10.15672/HJMS.2017.493.

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.