Yayın Geliş Tarihi: 30.07.2015
Yayına Kabul Tarihi: 11.08.2015
Online Yayın Tarihi: 26.02.2016
http://dx.doi.org/10.16953/deusbed.12175

Dokuz Eylül Üniversitesi
Sosyal Bilimler Enstitüsü Dergisi
Cilt: 17, Sayı: 3, Yıl: 2015, Sayfa: 291-302
ISSN: 1302-3284 E-ISSN: 1308-0911

IMPORTANCE OF INITIAL VALUE IN EXPONENTIAL SMOOTHING METHODS

Sedat ÇAPAR*

Abstract

Exponential smoothing is a very popular forecasting method for a wide range of time series data. There are two problems with exponential smoothing. First one is choosing smoothing constant. And second one is how to get initial value. In this paper importance of initial value and effects of it on the forecast is investigated and a cross table is constructed to help forecasters.

Keywords: *Exponential Smoothing, Simple Exponential Smoothing, Initial Value.*

ÜSTEL DÜZLEŞTİRME YÖNTEMLERİNDE BAŞLANGIÇ DEĞERİNİN ÖNEMİ

Öz

Üstel düzleştirme çeşitli zaman serisi verileri için yaygın olarak kullanılan popüler bir tahmin yöntemidir. Üstel düzleştirme ile ilgili iki önemli problem mevcuttur. Birincisi, düzleştirme sabitinin değerine karar vermek. İkincisi de başlangıç değerini belirlemektir. Bu çalışmada başlangıç değerinin önemi ve tahmin üzerindeki etkisi araştırılmış ve araştırmacılara yardımcı olmak amacıyla bir çapraz tablo oluşturulmuştur.

Anahtar Kelimeler: *Üstel Düzleştirme, Basit Üstel Düzleştirme, Başlangıç Değeri.*

INTRODUCTION

The exponential smoothing methods were developed by Brown and Holt unaware of each other (Brown, 1959; 1964; Holt, 1957). Roberts G. Brown was working for the US Navy when he was first form exponential smoothing (Gass and Harris, 2000). By the way, Charles C. Holt worked independently of Brown and developed exponential smoothing for additive trends and seasonal data. Later, an Office of Naval Research memorandum was created to document Holt's original work (Holt, 1957). Many contributions were followed by different researchers (Pegels, 1969; Roberts, 1982; Abraham and Ledolter, 1983).

* Doç. Dr., Dokuz Eylül Üniversitesi, Fen Fakültesi, İstatistik Bölümü, sedat.capar@deu.edu.tr

Makridakis and Hibon compared different time series methods using 111 time series and concluded that simple methods like exponential smoothing did not performed worse than the advanced ones (Makridakis and Hibon, 1979). But their conclusion is not accepted by the majority of researchers. Then, M-Competition was launched by Makridakis to continue the empirical comparisons of time series (Makridakis et al., 1982). Another contribution came from Gardner and he published his first paper providing an exhaustive review of exponential smoothing (Gardner, 1985).

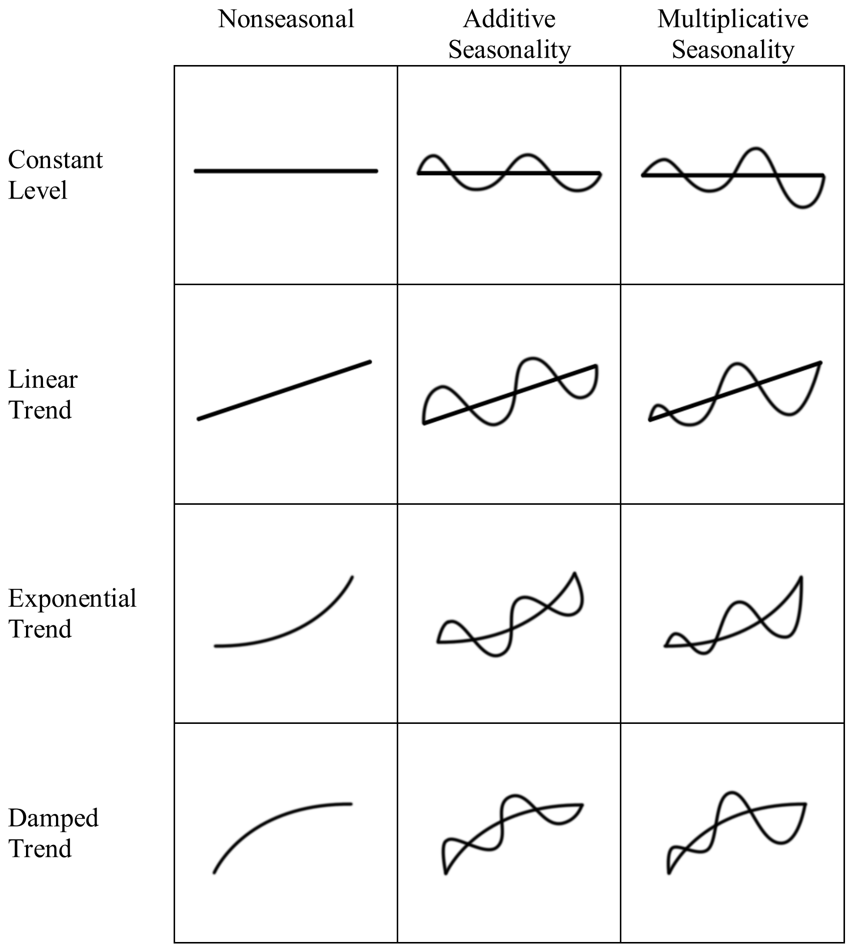
After these efforts, the popularity of exponential smoothing methods started to increase. Up to first paper of Gardner many researchers opinion about exponential smoothing was to ignore it since it was a special case of ARIMA (Gardner, 2006). However, many works done since 1985 showed that exponential smoothing methods are optimal for wide variety of time series models.

Makridakis and Hibbon continued with M2-Competition and M3-Competition. Many of researchers have repeated the conclusion of M-Competition by adding some new methods (Geurts and Kelly, 1986; Clemen, 1989) using M-Competition data (Lusk and Neves, 1984; Hill and Fildes, 1984; Koehler and Murphree, 1988) or using new data series (Armstrong and Collopy, 1992; 1993; Fildes et al., 1998). Al these studies showed the validity of four conclusions of M-Competition.

These works showed that simple methods like exponential smoothing is not worse than complicated once like ARIMA. Exponentinal smoothing methods are easy to calculate. Even not an expert person can calculate smoothing values and make forecast using simple computer programs like excel. On the other hand, the complicated methods like ARIMA need specialized computer programs and expert person to use them.

Besides of its ease many contributions were made for different forecast profiles. These profiles (Gardner, 1985) are given in Figure 1.

Figure 1: Forecast Profiles From Exponential Smoothing



There are a lot of methods for the forecast profiles above. Figure 2 contains standard equations of exponential smoothing (Gardner, 2006). These are extensions of the work of Brown, Holt and Winters (Brown, 1959; 1964; Holt, 1957; Winters, 1960). "For each type of trend, there are two sections of equations: the first give recurrence forms and the second gives equivalent error-correction forms. Recurrence forms were used in the original work by Brown and Holt and are still widely used in practice, but error-correction forms are simpler" (Gardner, 2006).

Figure 2: Standard Exponential Smoothing Equations

Trend	Seasonality		
	N None	A Additive	M Multiplicative
N None	$S_t = aX_t + (1-\propto)S_{t-1}$ $\hat{X}_t(m) = S_t$	$S_t = a(X_t - I_{t-p}) + (1-\propto)S_{t-1}$ $I_t = \delta(X_t - S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = S_t + I_{t-p+m}$	$S_t = a(X_t/I_{t-p}) + (1-\propto)S_{t-1}$ $I_t = \delta(X_t/S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = S_t I_{t-p+m}$
	$S_t = S_{t-1} + ae_t$ $\hat{X}_t(m) = S_t$	$S_t = S_{t-1} + ae_t$ $I_t = I_{t-p} + \delta(1-\alpha)e_t$ $\hat{X}_t(m) = S_t + I_{t-p+m}$	$S_t = S_{t-1} + ae_t/I_{t-p}$ $I_t = I_{t-p} + \delta(1-\alpha)e_t/S_t$ $\hat{X}_t(m) = S_t I_{t-p+m}$
A Additive	$S_t = aX_t + (1-\propto)(S_{t-1}+T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1}$ $\hat{X}_t(m) = S_t + mT_t$	$S_t = a(X_t - I_{t-p}) + (1-\propto)(S_{t-1} + T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1}$ $I_t = \delta(X_t - S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = S_t + mT_t + I_{t-p+m}$	$S_t = a(X_t/I_{t-p}) + (1-\propto)(S_{t-1} + T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1}$ $I_t = \delta(X_t/S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = (S_t + mT_t)I_{t-p+m}$
	$S_t = S_{t-1} + T_{t-1} + ae_t$ $T_t = T_{t-1} + \propto \gamma e_t$ $\hat{X}_t(m) = S_t + mT_t$	$S_t = S_{t-1} + T_{t-1} + ae_t$ $T_t = T_{t-1} + \propto \gamma e_t$ $I_t = I_{t-p} + \delta(1-\alpha)e_t$ $\hat{X}_t(m) = S_t + mT_t + I_{t-p+m}$	$S_t = S_{t-1} + T_{t-1} + ae_t/I_{t-p}$ $T_t = T_{t-1} + \propto \gamma e_t/I_{t-p}$ $I_t = I_{t-p} + \delta(1-\alpha)e_t/S_t$ $\hat{X}_t(m) = (S_t + mT_t)I_{t-p+m}$
DA Damped Additive	$S_t = aX_t + (1-\propto)(S_{t-1}+\emptyset T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)\emptyset T_{t-1}$ $\hat{X}_t(m) = S_t + \sum_{i=1}^{m} \emptyset^i T_t$	$S_t = a(X_t - I_{t-p}) + (1-\propto)(S_{t-1} + \emptyset T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)\emptyset T_{t-1}$ $I_t = \delta(X_t - S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = S_t + \sum_{i=1}^{m} \emptyset^i T_t + I_{t-p+m}$	$S_t = a(X_t/I_{t-p}) + (1-\propto)(S_{t-1} + \emptyset T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)\emptyset T_{t-1}$ $I_t = \delta(X_t/S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = (S_t + \sum_{i=1}^{m} \emptyset^i T_t)I_{t-p+m}$
	$S_t = S_{t-1} + \emptyset T_{t-1} + ae_t$ $T_t = \emptyset T_{t-1} + \propto \gamma e_t$ $\hat{X}_t(m) = S_t + \sum_{i=1}^{m} \emptyset^i T_t$	$S_t = S_{t-1} + \emptyset T_{t-1} + ae_t$ $T_t = \emptyset T_{t-1} + \propto \gamma e_t$ $I_t = I_{t-p} + \delta(1-\alpha)e_t$ $\hat{X}_t(m) = S_t + \sum_{i=1}^{m} \emptyset^i T_t + I_{t-p+m}$	$S_t = S_{t-1} + \emptyset T_{t-1} + ae_t/I_{t-p}$ $T_t = \emptyset T_{t-1} + \propto \gamma e_t/I_{t-p}$ $I_t = I_{t-p} + \delta(1-\alpha)e_t/S_t$ $\hat{X}_t(m) = (S_t + \sum_{i=1}^{m} \emptyset^i T_t)I_{t-p+m}$
M Multiplicative	$S_t = aX_t + (1-\propto)(S_{t-1}R_{t-1})$ $R_t = \gamma(S_t/S_{t-1}) + (1-\gamma)R_{t-1}$ $\hat{X}_t(m) = S_t R_t^m$	$S_t = a(X_t - I_{t-p}) + (1-\propto)S_{t-1}R_{t-1}$ $R_t = \gamma(S_t/S_{t-1}) + (1-\gamma)R_{t-1}$ $I_t = \delta(X_t - S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = S_t R_t^m + I_{t-p+m}$	$S_t = a(X_t/I_{t-p}) + (1-\propto)S_{t-1}R_{t-1}$ $R_t = \gamma(S_t/S_{t-1}) + (1-\gamma)R_{t-1}$ $I_t = \delta(X_t/S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = (S_t R_t^m)I_{t-p+m}$
	$S_t = S_{t-1}R_{t-1} + ae_t$ $R_t = R_{t-1} + \propto \gamma e_t/S_{t-1}$ $\hat{X}_t(m) = S_t R_t^m$	$S_t = S_{t-1}R_{t-1} + ae_t$ $R_t = R_{t-1} + \propto \gamma e_t/S_{t-1}$ $I_t = I_{t-p} + \delta(1-\alpha)e_t$ $\hat{X}_t(m) = S_t R_t^m + I_{t-p+m}$	$S_t = S_{t-1}R_{t-1} + ae_t/I_{t-p}$ $R_t = R_{t-1} + (\propto \gamma e_t/S_{t-1})/I_{t-p}$ $I_t = I_{t-p} + \delta(1-\alpha)e_t/S_t$ $\hat{X}_t(m) = (S_t R_t^m)I_{t-p+m}$
DM Damped Multiplicative	$S_t = aX_t + (1-\propto)(S_{t-1}R_{t-1}^{\emptyset})$ $R_t = \gamma(S_t/S_{t-1}) + (1-\gamma)R_{t-1}^{\emptyset}$ $\hat{X}_t(m) = S_t R_t^{\sum_{i=1}^{m} \emptyset^i}$	$S_t = a(X_t - I_{t-p}) + (1-\propto)S_{t-1}R_{t-1}^{\emptyset}$ $R_t = \gamma(S_t/S_{t-1}) + (1-\gamma)R_{t-1}^{\emptyset}$ $I_t = \delta(X_t - S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = S_t R_t^{\sum_{i=1}^{m} \emptyset^i} + I_{t-p+m}$	$S_t = a(X_t/I_{t-p}) + (1-\propto)(S_{t-1}R_{t-1}^{\emptyset})$ $R_t = \gamma(S_t/S_{t-1}) + (1-\gamma)R_{t-1}^{\emptyset}$ $I_t = \delta(X_t/S_t) + (1-\delta)I_{t-p}$ $\hat{X}_t(m) = (S_t R_t^{\sum_{i=1}^{m} \emptyset^i})I_{t-p+m}$
	$S_t = S_{t-1}R_{t-1}^{\emptyset} + ae_t$ $R_t = R_{t-1}^{\emptyset} + \propto \gamma e_t/S_{t-1}$ $\hat{X}_t(m) = S_t R_t^{\sum_{i=1}^{m} \emptyset^i}$	$S_t = S_{t-1}R_{t-1}^{\emptyset} + ae_t$ $R_t = R_{t-1}^{\emptyset} + \propto \gamma e_t/S_{t-1}$ $I_t = I_{t-p} + \delta(1-\alpha)e_t$ $\hat{X}_t(m) = S_t R_t^{\sum_{i=1}^{m} \emptyset^i} + I_{t-p+m}$	$S_t = S_{t-1}R_{t-1}^{\emptyset} + ae_t/I_{t-p}$ $R_t = R_{t-1}^{\emptyset} + (\propto \gamma e_t/S_{t-1})/I_{t-p}$ $I_t = I_{t-p} + \delta(1-\alpha)e_t/S_t$ $\hat{X}_t(m) = (S_t R_t^{\sum_{i=1}^{m} \emptyset^i})I_{t-p+m}$

EXPONENTIAL SMOOTHING

Exponential smoothing is probably the most widely used class of procedures for a wide variety of time series data in order to forecast the future. It weights past observations using exponentially decreasing weights. In other words, recent observations are given relatively more weight in forecasting than the older observations and these weights are automatically calculated by use of smoothing constants. There is no need to assign weights to each previous period.

In exponential smoothing, there are one or more parameters to be determined by the forecaster. These parameters assign the weights which are exponentially decreasing weights as the observations getting older. This is a desired situation because “future events usually depend more on recent data than on data from a long time ago” (Xie et al., 1997). This gives the power of adjusting an early forecast with the latest observation. In the case of moving averages, which is another technique of smoothing, the weights assigned to the observations are the same and equal to 1/ N so newest and oldest data have the same weights for forecasting.

“There are also other different types of forecasting procedures but exponential smoothing methods are widely used in industry. Their popularity is due to several practical considerations in short-range forecasting” (Gardner, 1985):

- equations are simple
- parameters have some intuitive meaning
- easy to compute
- accuracy can be obtained easily

There are 3 basic forms of exponential smoothing which are simple exponential smoothing, double exponential smoothing and triple exponential smoothing.

Simple Exponential Smoothing

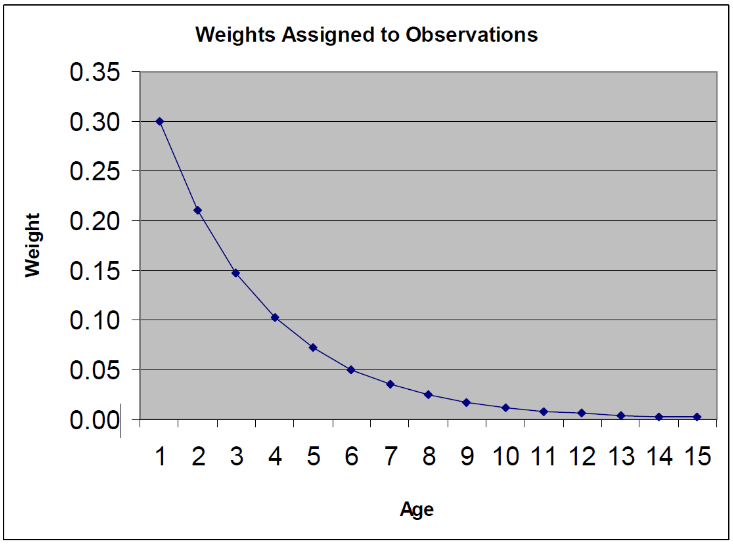
Simple exponential smoothing is suitable for forecasting data with no trend and no seasonal component. The smoothing equation for simple exponential smoothing in recurrence form is given by

$$S_t = \propto X_t + (1 - \propto) S_{t-1}$$

where S_t is the smoothing statistic (or smoothed value) and $\propto$ is the smoothing constant. It can be seen that the new smoothed value is the weighted sum of the current observation and the previous smoothed value. The weight of the most recent observation is $\propto$ and the weight of the most recent smoothed value is (1-$\propto$).

As a result, S_t is the weighted average of all past observations and the initial value of S_0. The weights are decreasing exponentially depending on the value of parameter $\propto$ (smoothing constant). Figure 3 shows the weights given to observations when $\propto$ value is 0.3. These weights appear to decline exponentially when connected by a smooth curve. This is why it is called “exponential smoothing”. More weights given to most recent observations and weights decrease geometrically with age.

Figure 3: Weights Assigned to Observations when ∝ is 0.3



Double Exponential Smoothing

Double exponential smoothing is preferred when there is a trend in the data. There are different ways to overcome the trend in the data. Holt (Holt et al., 1960) and Winters (Winters, 1960), uses two different parameters to smooth the level and trend of the series. The Brown models use one parameter to smooth both of them. The approach often used to determine updated estimates of *a* and *b* is known as double exponential smoothing. The specific formula for double exponential smoothing for two parameter is given by

$$S_t = \propto X_t + (1-\propto)(S_{t-1} + b_{t-1})$$

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1}$$

The first smoothing equation adjust S_t directly for the trend of the previous period, b_{t-1} , by adding it to the last smoothed value of S_{t-1}.

Triple Exponential Smoothing

Triple exponential smoothing is used if there is a trend and seasonality in the data. A third equation is introduced to care of the seasonality. The equations for triple exponential smoothing are given by

$$S_t = \propto \frac{X_t}{I_{t-L}} + (1-\propto)(S_{t-1} + b_{t-1})$$

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1}$$

$$I_t = \beta(\tfrac{X_t}{S_t}) + (1-\beta)I_{t-L}$$

INITIAL VALUE

As it seen from formulas exponential smoothing methods are recurring formulas. Each smoothed value is calculated as a weighted average of corresponding observation and previous smoothed value except S_0 which is called initial value or starting value. Recall simple exponential case, the formula is

$$S_t = \propto X_t + (1-\propto)S_{t-1}$$

then S_{t-1} can be written as

$$S_{t-1} = \propto X_{t-1} + (1-\propto)S_{t-2}$$

substituting S_{t-1} in first equation with its component S_t can be written as

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

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

and replacing S_{t-2} with its component we have

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

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

repeating the substitution for S_{t-3}, S_{t-4} and so on up to S_0 finally we have

$$S_t = \propto X_t + \propto (1-\propto)X_{t-1} + a(1-\propto)^2 X_{t-2} + a(1-\propto)^3 X_{t-3} + \propto (1-\propto)^4 X_{t-3} + \cdots + \propto (1-\propto)^{t-1} X_1 + (1-\propto)^t S_0$$

can also be written like this

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

As it seen from last formula, S_t is the weighted average of all past observations and the initial value S_0. The weights are decrease exponentially depending on the value of parameter $\propto$ (smoothing constant). The value of the parameters $\propto$ and S_0 must be given by the forecaster to calculate the smoothed values. Depending on the chosen value of these parameters, accuracy of simple exponential smoothing may vary.

Different methods for computing S_0 have been developed by a number of researchers. Brown's original suggestion is simply using the mean of the data for S_0. Other approaches are to use first observation or average of first 3 observations as S_0. Ledolter and Abraham (Ledolter and Abraham, 1984) recommended backcasting to obtain S_0. Another alternative with a limited number of data points is to use Bayesian methods to combine a prior estimate of the level with an average of the available data (Cohen, 1966), (Jonhson and Montgomery, 1974) and (Taylor, 1981).

There are numerous of different approaches to use for S_0. Therefore choosing S_0 is really important or if so knowing when it is important is studied in this paper. It is convenient to investigate effect of different initial values on real data from M-Competition. For this purpose first MNC44 is selected from M-Competition data and three different approaches are selected for S_0 as first observed value, average of first three observed values and average of all data and denoted by S_{0x_1}, $S_{0\bar{x}_3}$ and $S_{0\bar{x}}$ repectively. Number of observation is 126 and different values of α is used. Forecast values and their differences are given in Table 1. F_1, F_2 and F_3 denote the forecasts made when S_{0x_1}, $S_{0\bar{x}_3}$ and $S_{0\bar{x}}$ is used as initial value respectively.

Table 1: Forecasts and Their Differences for MNC44

∝	F_1	F_2	F_3	abs(F_1-F_2)	abs(F_1-F_3)	abs(F_2-F_3)
0.1	2099.319	2099.319	2099.319	0.0000006	0.0003007	0.0003013
0.2	2239.911	2239.911	2239.911	0.0000000	0.0000000	0.0000000
0.3	2265.154	2265.154	2265.154	0.0000000	0.0000000	0.0000000
0.4	2263.963	2263.963	2263.963	0.0000000	0.0000000	0.0000000
0.5	2257.560	2257.560	2257.560	0.0000000	0.0000000	0.0000000
0.6	2252.168	2252.168	2252.168	0.0000000	0.0000000	0.0000000
0.7	2249.447	2249.447	2249.447	0.0000000	0.0000000	0.0000000
0.8	2249.309	2249.309	2249.309	0.0000000	0.0000000	0.0000000
0.9	2251.071	2251.071	2251.071	0.0000000	0.0000000	0.0000000

The results in Table 1 shows that using different initial values do not have effect on the forecast made by exponential smoothing. This is of course due to large data size. Although small values of ∝ (ie, $\propto = 0.1, 0.2$ ...) gives less weight to related observation and more weight to previous smoothed values, there still seems to be no effect of S_0 on the forecast made by starting different values for S_0. We can conclude that S_0 loose its importance when *n* is big.

The effect of S_0 must also be investigated when *n* is small. QNM6 in M-competition has number of observations equal to 20. And it is used to see the effects of using different initial values when *n* is relatively small. Obtained results are shown in Table 2. This time it is possible to see some differences on forecasts for small values of ∝. However, the only significant difference is observed for ∝ is equal to 0.1. The differences are very small for ∝ values 0.2, 0.3 and 0.4. And it is possible to conclude that there is no difference on the forecasts when ∝ is bigger than 0.4.

Table 2: Forecasts and Their Differences for QNM6

∝	F_1	F_2	F_3	abs(F_1-F_2)	abs(F_1-F_3)	abs(F_2-F_3)
0.1	282.917	283.039	285.405	0.1221144	2.4885180	2.3664035
0.2	296.039	296.044	296.131	0.0045133	0.0919751	0.0874618
0.3	298.617	298.617	298.619	0.0001073	0.0021872	0.0020799
0.4	298.245	298.245	298.245	0.0000014	0.0000292	0.0000278
0.5	297.698	297.698	297.698	0.0000000	0.0000002	0.0000002
0.6	298.293	298.293	298.293	0.0000000	0.0000000	0.0000000
0.7	300.559	300.559	300.559	0.0000000	0.0000000	0.0000000
0.8	304.702	304.702	304.702	0.0000000	0.0000000	0.0000000
0.9	310.830	310.830	310.830	0.0000000	0.0000000	0.0000000

Now, it is expected that the effect of using different initial values will increase when n is getting smaller and smaller especially for the small values of ∝. If n is big then there is no effect. Now the question is that how big n is enough to think that there will be no difference.

If we look at to the last form we obtained for S_t, initial value has a weight of $(1-\propto)^t$. So two parameters ∝ and t define the weight of initial value S_0. t is the number or observations theoretically starts from 1 and go to infinity and ∝ is the smoothing constant whose value between 0 and 1 which is chosen by the forecaster. It is possible to calculate weights assigned to initial value S_0 choosing different ∝ values and number of observations.

So, starting ∝ from 0.1 and incrementing by 0.1 up to 0.9 and for t starting from 1 and incrementing by one up to 20 then incrementing arbitrarily the weights assigned to initial value S_0 is obtained and shown in Table 3. Now we can conclude the followings using the results from Table 3.

If the number of observations is big than weight assigned to initial value S_0 is so small even for different values of ∝. Therefore initial value will not effect the forecast too much, so the forecaster may chose first observation as the initial value without worrying about it.

If the number of observations is small, let say smaller than 20, than it becomes important with the chosen value of ∝. Since the weight assigned to initial value is affected both the value of smoothed constant and number of observations, the value of ∝ is gaining importance when number of observations is smaller than 20.

If high values for ∝ are chosen than weight assigned to initial value is still so small. For example when t is 5 and ∝ is 0.9 then weight is equal to 0.00001. On the contrary if small values for ∝ are chosen when t is also small then weight assigned to initial value is also high. Therefore the value assigned by forecaster for unknown initial value S_0 will be effect the accuracy of the forecast for small sized data when small values for ∝ is chosen.

Table 3: Weights Assigned to S_0 for Different Values of $\propto$ and *T*.

	$\propto$								
t	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	0.90000	0.80000	0.70000	0.60000	0.50000	0.40000	0.30000	0.20000	0.10000
2	0.81000	0.64000	0.49000	0.36000	0.25000	0.16000	0.09000	0.04000	0.01000
3	0.72900	0.51200	0.34300	0.21600	0.12500	0.06400	0.02700	0.00800	0.00100
4	0.65610	0.40960	0.24010	0.12960	0.06250	0.02560	0.00810	0.00160	0.00010
5	0.59049	0.32768	0.16807	0.07776	0.03125	0.01024	0.00243	0.00032	0.00001
6	0.53144	0.26214	0.11765	0.04666	0.01563	0.00410	0.00073	0.00006	0.00000
7	0.47830	0.20972	0.08235	0.02799	0.00781	0.00164	0.00022	0.00001	0.00000
8	0.43047	0.16777	0.05765	0.01680	0.00391	0.00066	0.00007	0.00000	0.00000
9	0.38742	0.13422	0.04035	0.01008	0.00195	0.00026	0.00002	0.00000	0.00000
10	0.34868	0.10737	0.02825	0.00605	0.00098	0.00010	0.00001	0.00000	0.00000
11	0.31381	0.08590	0.01977	0.00363	0.00049	0.00004	0.00000	0.00000	0.00000
12	0.28243	0.06872	0.01384	0.00218	0.00024	0.00002	0.00000	0.00000	0.00000
13	0.25419	0.05498	0.00969	0.00131	0.00012	0.00001	0.00000	0.00000	0.00000
14	0.22877	0.04398	0.00678	0.00078	0.00006	0.00000	0.00000	0.00000	0.00000
15	0.20589	0.03518	0.00475	0.00047	0.00003	0.00000	0.00000	0.00000	0.00000
16	0.18530	0.02815	0.00332	0.00028	0.00002	0.00000	0.00000	0.00000	0.00000
17	0.16677	0.02252	0.00233	0.00017	0.00001	0.00000	0.00000	0.00000	0.00000
18	0.15009	0.01801	0.00163	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000
19	0.13509	0.01441	0.00114	0.00006	0.00000	0.00000	0.00000	0.00000	0.00000
20	0.12158	0.01153	0.00080	0.00004	0.00000	0.00000	0.00000	0.00000	0.00000
30	0.04239	0.00124	0.00002	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
50	0.00515	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
100	0.00003	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
250	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
500	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

CONCLUSION

For the use of exponential smoothing methods it is required to assign some values to unknown parameters and initial value. Forecaster must first choose a proper value for smoothing constant and second set initial value to be able to calculate smoothed values and make forecast. These unknown values assigned by the forecaster effect the accuracy of forecast.

It is obvious that the accuracy of forecast directly affected by the value of smoothing constant since it adjusts the weights given to observations and initial value. Therefore it is always important to choose a proper value for smoothing constant.

However, in this paper it is shown that initial value is not always have an influence on the forecast made by exponential smoothing methods. If the number of observations is very large than the weight assigned to initial value is very low therefore how to choose an initial value is not a problem. Forecaster may choose first observation as the value of initial value.

On the other hand, it is important for small sized data but number of observations alone is not decisive. Number of observations and smoothed constant together have an important behavior on the weight assigned to initial value. If *t* is small but ∝ is high then the weight assigned to initial value is still small. Weight assigned to initial value is high only when both *t* and ∝ is small therefore choosing a proper initial value is important if both of them have small values.

The table constructed in this paper for different sample sizes and ∝ values can be used to determine weight assigned to initial value and practically decide whether or not the initial value will have an important effect on the forecast (Table 3).

REFERENCES

Abraham, B. and Ledolter, J. (1983). *Statistical methods for forecasting*. New York: John Wiley and Sons.

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

Armstrong, J. S. and Collopy, F. (1993). Causal forces: Structuring knowledge for time series extrapolation. *Journal of Forecasting,* 12 (2): 103-115.

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

Brown, R. G. (1964). *Smoothing, forecasting and prediction of discrete time series.* NJ: Prentice-Hall.

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

Cohen, G. D. (1966). Bayesian adjustment of sales forecasts in multi-item inventory control system. *Journal of Industrial Engineering,* 17: 474-479.

Fildes, R., Hibon, M., Makridakis, S. and Meade, N. (1998). Generalising about univariate forecasting methods: Further empirical evidence. *International Journal of Forecasting,* 14 (3): 339-358.

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

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

Gass, S. I. and Harris, C. M. (Eds.) (2000). *Encyclopedia of operations research and management science*. Dordrecht: Kluwer Academic Publishers.

Geurts, M. D. and Kelly, J. P. (1986). Forecasting retail sales using alternative models. *International Journal of Forecasting,* 2 (3): 261-272.

Hill, G. and Fildes, R. (1984). The accuracy of extrapolation methods: An automatic Box–Jenkins package SIFT. *Journal of Forecasting,* 3 (3): 319-323.

Holt, C. C. (1957). *Forecasting seasonals and trends by exponentially weighted moving averages, ONR Memorandum.* Pitssburgh, PA: Carnegie Institute of Technology.

Holt, C. C, Modigliani, F., Muth, J. F. and Simon, H. (1969). *Planning production, inventories, and the work force.* NJ: Prentice Hall.

Johnson, L. A. and Montgomery, D. C. (1974). *Operations research in production planning, scheduling, and inventory control.* New York: Wiley.

Koehler, A. B. and Murphree, E. S. (1988). A comparison of results from state space forecasting with forecasts from the Makridakis competition. *International Journal of Forecasting,* 4 (1): 45-55.

Ledolter, J. and Abraham, B. (1984). Some comments on the initialization of exponential smoothing. *Journal of Forecasting,* 3 (1): 79-84.

Lusk, E. J. and Neves, J. S. (1984). A comparative ARIMA analysis of the 111 series of the Makridakis competition. *Journal of Forecasting,* 3 (3): 329-332.

Makridakis, S., Hibon, M. and Moser, C. (1979). Accuracy of forecasting: An empirical investigation. *Journal of the Royal Statistical Society: Series A (General),* 142 (2): 97-145.

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

Pegels, C. C. (1969). Exponential forecasting: Some new variations. *Management Science.* 15 (5): 311-315.

Roberts, S. A. (1982). A general class of Holt–Winters type forecasting models. *Management Science,* 28 (7): 808-820.

Taylor, S. G. (1981). Initialization of exponential smoothing forecasts. *AIIE Transactions,* 13 (3): 199-205.

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

Xie, M., Hong, G. Y. and Wohlin, C. (1997) A study of the exponential smoothing technique in software reliability growth prediction. *Quality and Reliability Engineering International,* 13 (6): 347-353.