Exponential Smoothing Method to Forecast Numbers of Neoplasms Patients in Libya.

Mohamed AH Milad

University of Al-Marqab, Faculty of Economics & Commerce, Department of Data Analysis and Computer

agga2425@yahoo.com and mamilad@elmergib.edu.ly

Abstract

Forecasting models have played significant roles in many applications for over a century. If the error terms in models are normally distributed, the models can produce more accurate forecasting results. The current study develops reliable techniques for a forecasting model for the number of Neoplasms patients in Libya. Simple, Holt's linear trend, Brown's linear tread, and Damped-trend exponential smoothing methods were used to extract information that helps to enhance the forecasting accuracy of patient cases. The data obtained cover the period between 2010 and 2019. Based on the results of the current study, the findings revealed that Brown's linear tread exponential smoothing method offers more probabilistic information which improves the forecasting of the Neoplasms patients in Libya. According to this model, the study predicts new cases in the next two years and they are increasing. The study recommends working on appropriate measures to stop this increase.

Keywords: Model selection; Exponential smoothing; Neoplasms patients; Forecast

Abbreviations

SES refers to simple exponential something. HLTES refers to Holt's linear trend exponential smoothing. BLTES is Brown's linear trend. DTES refers to Damped trend exponential smoothing. ACF is Autocorrelation coefficients function. PACF refers to Partial Autocorrelation.

Introduction

Over the past years, our beloved country has gone through a clear decline in all sectors, which requires a comprehensive vigil in all fields and economic activities, and this happens with the concerted efforts of researchers in all specialties to conduct studies and research that would reduce what afflicted the country pollution, diseases and pests that affected the health, agricultural, economic and industrial aspects as well. Therefore, this study came to deal with its importance at the developmental level, as it means the human age, which falls on the responsibility of building, ages, and keeping pace with progress and civilization development. One of the foundations of building health is to ward off all diseases, including tumors, which cause a high rate of deaths compared to other diseases, and given an increase in the number of people with this disease recently, this study came to reveal this phenomenon, which has increased in Libya due to the acute shortage of health care and treatment due to poor services in most health centers. The emergence and utilization of exponential smoothing techniques for prediction were in the mid-1950s by Brown in 1956, Hole in 1957, and Maggie in 1958. Since then, these methods have had many applications in various aspects of working life. The exponential smoothing methods depend on giving weighted averages so that the modern data contain more averages of weights than the old data, and this is more logical and corresponds to the forecasting goal, which made these methods more accurate and thus more used in practice (Ostertagová & Ostertag, 2011, 2012; Yang et al., 2015). According to (Gardner, 1985) the exponential smoothing methods may be seasonal or non-seasonal. Four non-seasonal methods are Simple, Holt's linear trend, Brown's linear tread, and Damped-trend exponential smoothing methods besides three seasonal methods which are simple seasonal, winter additive, and winter multiplicative exponential smoothing methods. Furthermore, the selection of exponential smoothing methods depends on the type of component that exists in the time series. The main purpose of this study is to forecast the numbers of Neoplasms Patients in Libya and compare the proposed approaches with each other empirically. Using these methods, I forecast the Neoplasms patients for Libya under scenarios related to the poor services in most health centres. Finally, I discuss policy options to address the expected Neoplasms patients' increase.

Material and Methods

Material. This section describes the case study which is regarded as an effective research approach to investigate and compare the proposed models with each other. This case study is adopted based on the steps discussed below.

Data collection. The efficiency and reliability of the proposed forecasting model, which used data of the Neoplasms patients in Libya covering a period of 10 years have been tested and demonstrated. In particular, the period covered by the data is from 2010 to 2019. The source of this information was Bureau of statistics and census Libya. The graphical plot of the series is presented in Figure 1.

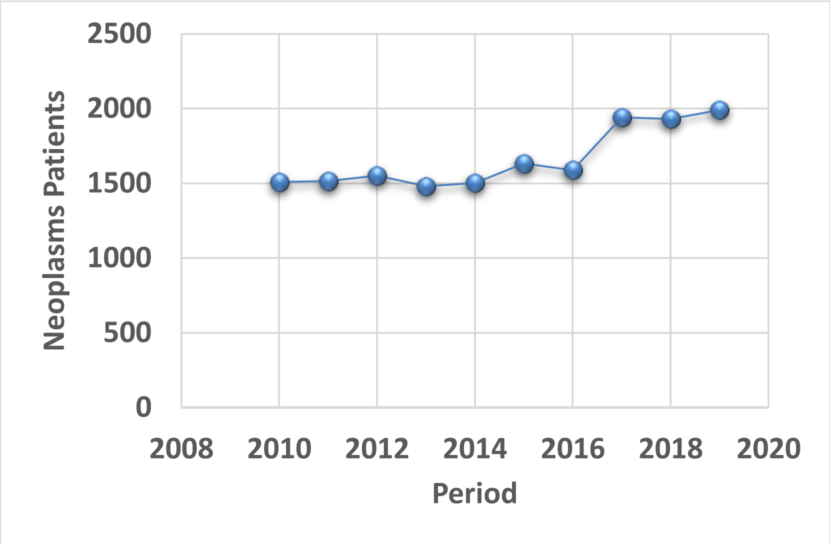

Figure 1: Time series of Neoplasms patients (2010 –1919)

Evaluation of the forecasting performance indices. This study employed the parameter significance test and three statistical indices to measure and evaluate the proposed approach and it's forecasting accuracy. The root square error (RMSE), mean absolute percentage error (MAE), and Bayesian information criterion (BIC) are used to measure the goodness-of-fit of models, with smaller values indicating a better forecasting performance (Ball, 2001; Wang & Hu, 2015).
These indices are defined as follows:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} e_t^2}{n}} \quad (1)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |e_t| \quad (2)$$

$$BIC = n\log(1 - R^2) + k\ log(n) \quad (3)$$

In Equations (1), and (2) e_t is the residual at time t, y_t is the observation at time t, and n is the length of the series. In Equation (3), k is the number of parameters fitted in the model, n is the number of observations and R^2 is the proportion of the variance explained by the model.

Methods. This section describes the methods proposed in this study, including the exponential smoothing methods such as Simple, Holt's linear trend, Brown's linear trend, and Damped-trend exponential smoothing.

I. Simple Exponential Smoothing (SES)

The SES model is based on the premise that the time-series level must fluctuate around a constant level or change slowly over time (Ostertagová & Ostertag, 2012). As stated by (Soubhik Chakraborty, 2013), simple exponential smoothing is easily applied, and it produces a smoothed statistic once two observations are available. This model is appropriate for a series in which there is no trend or seasonality. Let the observed time series up to time period t upon a variable y be denoted by $y_1, y_2, \ldots, y_t$. Assume that we want to determine the forecast of $y_t(1)$ of the next value $y_{t+1}$of the series, which is yet to be observed. Given data up to time period t-1, the forecast for the next time t is denoted by y_{t-1}. When the observation y_t becomes available, the forecast error is expressed as $y_t - y_{t-1}(1)$.

As mentioned, the forecast for the next period using the forecast error is calculated by taking the forecast for the previous period and adjusting it due to the simple method or single exponential smoothing. The simple exponential smoothing has a single-level parameter and can be described by the following equations:

Level $$L(t) = \alpha Y(t) + (1-\alpha)L(t-1) \quad (4)$$

Forecast $$\hat{Y}_t(k) = L(t) \quad (5)$$

Where α is the level smoothing weight that lies between 0 and 1, $L(t)$ is the old smoothed value or forecast for period t, $Y(t)$ = is a new observation or actual value of the series in period t, and $\hat{Y}_t(k)$ is the forecast for k periods ahead, i.e., the forecast of y_{t+k} for a certain subsequent time period $t+k$ based on all data points up to time period t. The ARIMA model equivalent to the simple exponential smoothing model is the ARIMA (0, 1, 1) model with zero-order of autoregressive, one order of differencing, one order of moving average, and no constant (Fomby, 2008; Ramasubramanian).

II. Holt's linear trend exponential smoothing (HLTES)

This model is appropriate for a series in which the trend is linear and there is no seasonality. Its smoothing parameters are level and trend, which are not constrained by the values of the other. Holt's model is more general than Brown's model but it may take longer to compute the larger series. Holt's exponential smoothing has level and trend parameters and three equations; the forecast can be identified as follows:

Level $$L(t) = \alpha Y(t) + (1-\alpha)L(t-1) + T(t-1 \quad (6)$$

Trend $$T(t) = \gamma(L(t) - L(t-1)) + (1-\gamma)T(t-1) \quad (7)$$

Forecast $$\hat{Y}_t(k) = L(t) + kT(t) \quad (8)$$

where $L(t)$ denotes the level of the series at time t, $T(t)$ denotes the trend (addition) of the series at time t, $Y(t)$ is a new observation or actual value of series in period t, k is the number of periods to be forecast, $\hat{Y}_t(k)$ is the forecast for k periods into the future, α is the level smoothing weight for the data $(0 < \alpha < 1)$, and γ is the trend smoothing weight estimate $(0 < \gamma < 1)$.

Holt's exponential smoothing is similar to an ARIMA model with zero orders of autoregressive, two orders of differencing, and two orders of moving average (SPSS, 2013). (Chatfield et al., 2001) and (Taylor & Bunn, 1999) both confirmed that Holt's linear trend method is optimal for an ARIMA (0, 2, 2) model.

III. Brown's linear trend (BLTES)

This model is appropriate for a series in which a linear trend and no seasonality are observed. Its smoothing parameters are level and trend, which are assumed to be equal. Thus, Brown's model is a special case of Holt's model. Brown's exponential smoothing has level and trend parameters and can be described by the following three equations:

$$L(t) = \alpha Y(t) + (1-\alpha)L(t-1) \tag{9}$$

$$T(t) = \alpha(L(t) - L(t-1)) + (1-\alpha)T(t-1) \tag{10}$$

$$\hat{Y}_t(k) = L(t) + ((k-1) + \alpha^{-1})T(t) \tag{11}$$

Where $L(t)$ is the exponentially smoothed value of $\mathrm{y_t}$ at time t, $T(t)$ is the double exponentially smoothed value of y_t at time t, α is the smoothing constant ($0 < \alpha < 1$), and $\hat{Y}_t(k)$ is the forecast for period t. The ARIMA model equivalent to the linear, exponential smoothing model is the ARIMA (0, 2, 2) model (Fomby, 2008).

IV. Damped trend exponential smoothing (DTES)

Damped-trend exponential smoothing has level and damped trend parameters and can be described by the following equations:

$$L(t) = \alpha\, Y(t) + (1-\alpha)(L(t-1) + \emptyset\, T(t-1)), \tag{12}$$

$$T(t) = \gamma\big(L(t) - L(t-1)\big) + (1-\gamma)\, \emptyset\, T(t-1), \tag{13}$$

$$\hat{y}_t(k) = L(t) + \sum_{i=1}^{k} \emptyset^i\, T(t) \tag{14}$$

Where L(t) and T(t) are the level and trend components of the series. Two smoothing parameters exist, namely, h_1 and h_2, for the level and trend, and an autoregressive-damping parameter φ_i is used to control the growth rate of the forecasts. It is functionally equivalent to an ARIMA (1, 1, 2) (SPSS, 2013).

Results and Discussion

Results: The first step in developing the exponential smoothing model is to check the stationary pattern of the time series. The data were collected from 2010 to 2019, as shown in Figure 1. Corresponding to the sequence of value for a single variable in ordinary data analysis, each case (row) in the data represents an observation at a different time. The observations must be taken at similarly spaced intervals. According to Figure 1, the oscillations (differences) between the observations at the beginning of the series were low during the period from 2010 to 2016, whereas increased oscillations in subsequent years were observed. This observation indicates the instability of the variance. Furthermore, a clear increase in the time series was

observed, which indicates the heterogeneity of the mean. The dataset does not have stationary invariance; that is, a natural logarithm transformation is required to give the dataset a constant variance. Finally, the dataset is not stationary in mean; that is, a difference is required to provide a constant mean.

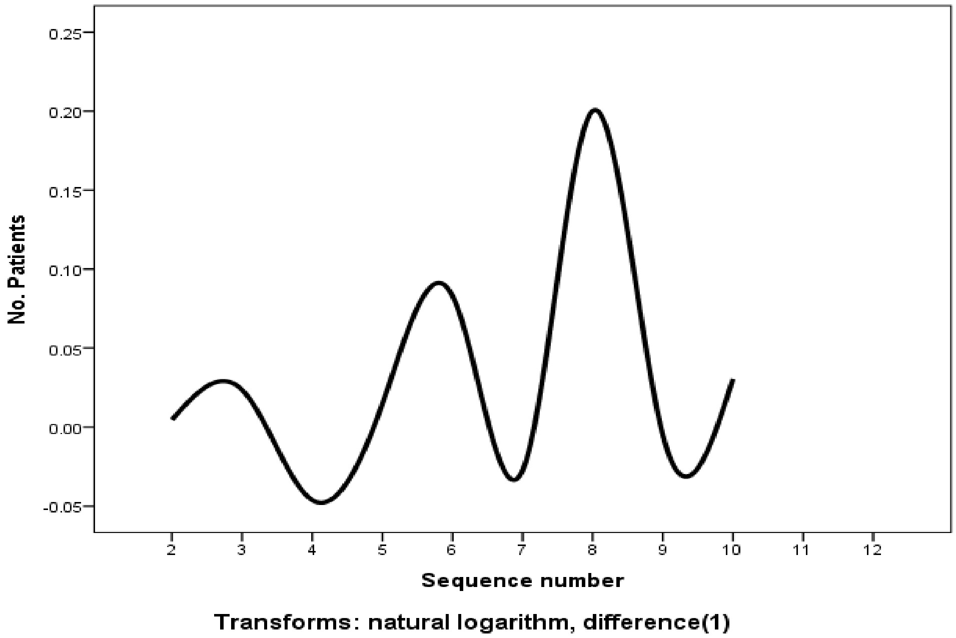

Figure 2: Time series after transforms and differencing of order one.

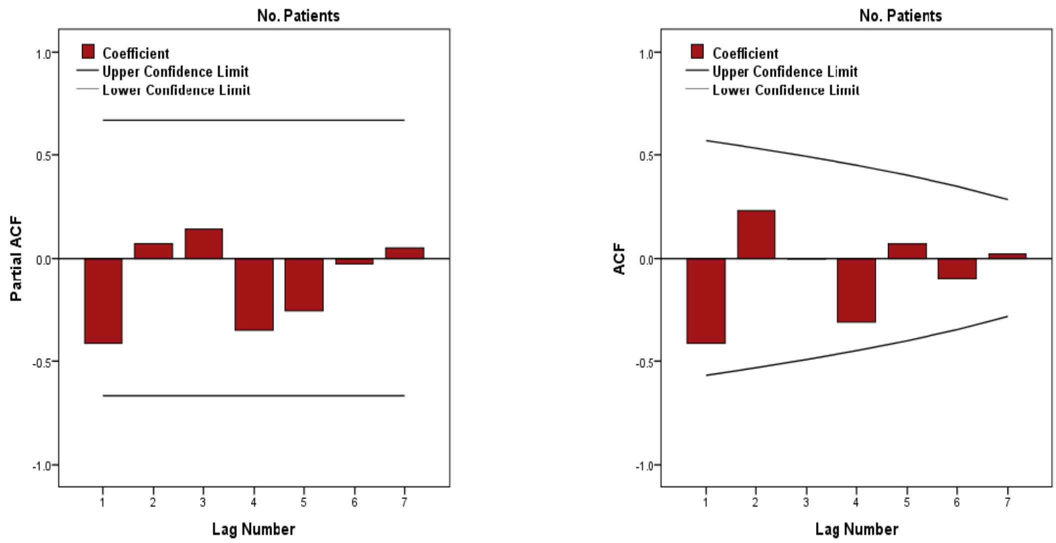

Figure 3: ACF and PACF of residuals of Neoplasms Patients.

At this point, I examine the data regarding ACF and PACF. As evident in Figure 3, they are insignificant, indicating that the model can interpret the general trend. The study reported in this paper compares all the above-mentioned models following various steps to test the significance of the estimated parameters and measures the forecasting error.

First step - Testing the significance of the estimated parameters. This step tested the significance of the estimated coefficients of Neoplasms patients' models.

Table 1. Results of testing the significance of the estimated parameters of the exponential smoothing models

Model	Parameters	Estimation	Standard error	p-value
SES	Alpha (level)	0.897	0.236	0.026
HLTES	Alpha (Level)	0.001	0.294	0.997
	Gamma (Trend)	0.000	109.97	1.000
BLTES	Alpha (level and Trend)	0.461	0.133	0.007
DTES	Alpha (level)	0.000	0.406	1.000
	Gamma (Trend)	0.001	6969.7	1.000
	Phi (Trend damping factor)	0.999	0.017	0.000

Source: Own data calculations

Table 1 presents the p-values for the estimated coefficients of all Neoplasms patients in Libya. The p-values for the estimated coefficients of the SES model and BLTES model are less than 0.05, indicating that they are highly significant. However, the p-value for at least one of the estimated coefficients of the HLTES model and DTES model are greater than 0.1. This finding implies that the values are insignificant at the 0.05 level.

As shown in Table 2, the four models, namely, the SES model, BLTES model, HLTES model and DTES model are compared. The comparison among these models focused on various measures of error. The results of the forecasting performance of these four models are summarized in Table 2 below:

Table 2: Statistical measures of forecast error for the Neoplasms Patients in Libya

Models	BIC	RMSE	MAE
SES	9.925	127.427	70.982
HLTES	9.968	125	85.28
BLTES	9.817	120.7029	70.428
DTES	10.14	122.83	85.72

Source: Own data calculations

The results of the comparison between the four abovementioned models in terms of their forecasting performance in Table 2 include the BIC, RMSE, and MAE. Figure 4 also illustrates the results generated from such robustness assessment of the different methods. The forecasting performance of these different models upon further comparison is indicated by each bar that shows the number of best forecasts obtained by its corresponding model in terms of a specified accuracy measure.

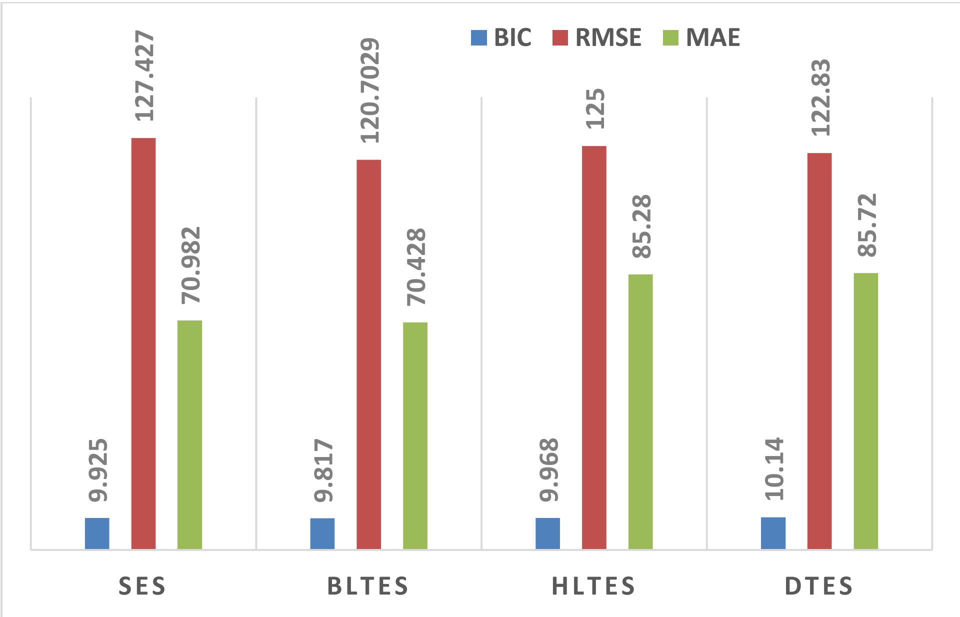

Figure 4: Results of the comparison of the forecasting performance among the different models.

The author interpreted and discussed the relevant issues according to the results illustrated in Table 2 and Figure 4

Discussion. The previously stated experimental design and methodologies employed in the current study aimed to show the experimental forecasts of the Neoplasms Patients in Libya. Subsequently, the forecasting performance of the models was evaluated by testing the significance of the estimated parameters and three main measurements criteria. The results presented in Table 2 and Figure 4 revealed that the BIC, RMSE and MAE values of BLTES are 9.817, 120.7029, and 70.428, respectively, for the time series of the Neoplasms patients in Libya. Such results clearly indicate that the BIC, RMSE, and MAE values are lower than those of other models. A further comparison of all models used in the present study shows that the BLTES model achieved the best performance among all models because its fit was the best. The ACF and PACF of the residuals are presented in Figure 5. After fitting the model, the residuals should only be white noise to obtain a good forecasting model. In examining the residuals, insignificant values are expected for these statistics.

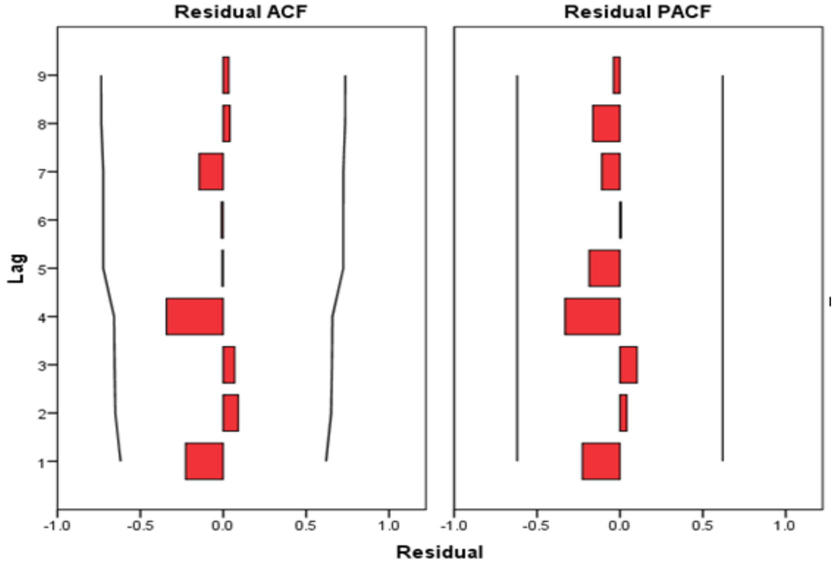

Figure 5: ACF and PACF of the residuals of Neoplasms Patients from the fitted BLTES model.

As shown in Figure 5, the ACF and PACF of the residual errors are insignificant, thereby implying that the BLTES model is good and suitable for forecasting Neoplasms Patients in Libya.

Forecasting Neoplasms patients in Libya

Once a model is identified, its parameters are estimated and its diagnostics are checked. The model can then be used to forecast the future values of a time series. Fitted exponential smoothing models can be used to run forecasts for future values of the time series by using the forecast BLTES function in the forecast package. Figure 6 presents the actual values for the years of 2010 to the years 2019 and the predicted values for the next 2 years using our BLTES model.

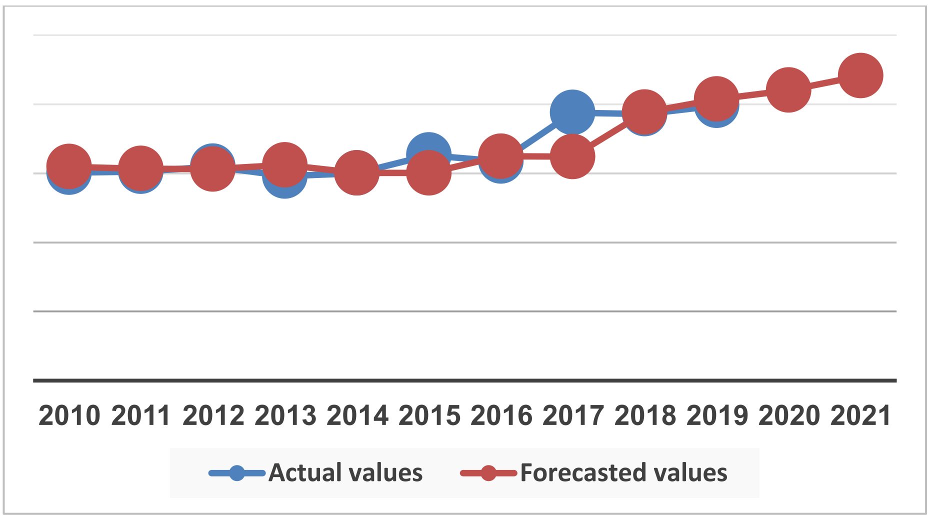

Figure 4.16: Forecast for the production of Neoplasms Patients in Libya

The selected model demonstrates excellent performance as reflected in its explained variability and predictive power. Therefore, the results of the model show an increase in the numbers of Neoplasms Patients in Libya during the next two years. As shown in the above Figure, this increase showed the same behaviour of the original series.

Conclusions

The present study proposed and evaluated a method for forecasting of Neoplasms Patients in Libya. The proposed models and the exponential smoothing (SES, HLTES, BLTES, and DTES) models, were evaluated by comparing them to each other based on the time series of Neoplasms Patients in Libya. This study has a useful contribution to the literature because it represents the first empirical study that applied the exponential smoothing method in this research area. The results provided evidence of the importance and value of such BLTES method as a powerful forecasting method that improves or increases the prediction accuracy of the numbers of Neoplasms Patients as well as enhances forecasting methods in the Libyan context. As observed from the results an increase in the Neoplasms patients in Libya. For this purpose, I recommend taking the results of this research, which showed an increase in the number of Neoplasms patients over time, which requires taking appropriate measures to stop this growth. Future research would benefit better from this research by focusing on other methods, which use data from a broad sample of Neoplasms in the Libya context and by comparing their findings with the research results.

الملخص

لعبت نماذج التنبؤ أدوارًا مهمة في العديد من التطبيقات لأكثر من قرن. إذا كانت حدود الخطأ في النموذج موزعة توزيعاً طبيعياً يمكن للنموذج إنتاج نتائج تنبؤ أكثر دقة. وقد طورت الدراسة الحالية تقنيات موثوقة لنموذج تنبؤ لعدد مرضى الأورام في ليبيا. تم استخدام أسلوب التمهيد الأسي البسيط و طريقة التمهيد الأسي الخطي لهولت وطريقة التمهيد الأسي الخطي لبراون و طريقة التمهيد الأسي الخطي المتباطئ لاستخراج المعلومات التي تساعد على تعزيز دقة التنبؤ بحالات المرضى. تغطي البيانات التي تم الحصول عليها الفترة ما بين سنة 2010 الي سنة 2019. واستناداً إلى نتائج الدراسة الحالية ، كشفت النتائج أن طريقة التمهيد الأسي الخطي لبراون تقدم معلومات احتمالية أكثر مما يحسن التنبؤ بمرضى الأورام في ليبيا. وفقًا لهذا النموذج ، تتنبأ الدراسة بحالات جديدة في العامين المقبلين وهي في تزايد. وتوصي الدراسة بالعمل على الإجراءات المناسبة لوقف هذه الزيادة.

الكلمات الدالة: التنبؤ، التمهيد الأسي ، حالات الأورام

REFERENCES

Ball, R. D. (2001). Bayesian methods for quantitative trait loci mapping based on model selection: approximate analysis using the Bayesian information criterion. *Genetics, 159*(3), 1351-1364.

Chatfield, C., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2001). A new look at models for exponential smoothing. *Journal of the Royal Statistical Society: Series D (The Statistician), 50*(2), 147-159.

Fomby, T. B. (2008). Exponential smoothing models: Retrieved from Southern Methodist University website: http://www. google. com/url.

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

Ostertagová, E., & Ostertag, O. (2011). *The simple exponential smoothing model.* Paper presented at the The 4th International Conference on Modelling of Mechanical and Mechatronic Systems, Technical University of Košice, Slovak Republic, Proceedings of conference.

Ostertagová, E., & Ostertag, O. (2012). Forecasting using simple exponential smoothing method.

Ramasubramanian, V. Time-series analysis, modelling and forecasting using SAS software.

Soubhik Chakraborty, S. S., Swarima Tewari and Mita Pal. (2013). an intersting application of simple exponential smoothing in music analysis. *International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI), 2*, 37-44.

SPSS, I. (2013). IBM SPSS Modeler 16 Algorithms Guide, IBM IBM SPSS.

Taylor, J. W., & Bunn, D. W. (1999). Investigating improvements in the accuracy of prediction intervals for combinations of forecasts: A simulation study. *International Journal of Forecasting, 15*(3), 325-339.

Wang, J., & Hu, J. (2015). A robust combination approach for short-term wind speed forecasting and analysis–Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model. *Energy, 93*, 41-56.

Yang, D., Sharma, V., Ye, Z., Lim, L. I., Zhao, L., & Aryaputera, A. W. (2015). Forecasting of global horizontal irradiance by exponential smoothing, using decompositions. *Energy, 81*, 111-119.