Global and Stochastic Analysis
Special Issue: *25th International Conference of Forum for Interdisciplinary Mathematics*

MUK PUBLICATIONS
Open Access Publisher

EXPONENTIAL SMOOTHING MODELING AND FORECASTING FOR INCIDENCE OF TUBERCULOSIS IN INDIA

M. J. THIRUNAVUKKARASU

Abstract. The Indian annual tuberculosis disease occurrence is predicted for suitable plan to prevent sickness epidemic and different smoothing techniques to predict the future quantity of disease, especially for periodic disease. Four smoothing model were used for disease forecasting on India tuberculosis yearly data. The time series were downward movement and forecasting performance was compared in the base of the actual and prediction value. A total of 5119 (per 100,000 people) tuberculosis incidence in India were identified for the duration of the study time, the average TB rate was 204.76, SD 16.546, minimum 167 and maximum 217 tuberculosis incidence. This Holt's linear model ($\alpha = 0.8, \beta = 1.0$), passed the parameter ($p < 0.01$) and residual ($p > 0.05$) tests, with highest value and lowest value of RMSE, MAPE, MAE, and NBIC 99%, 0.772, 0.310, 0.618, 0.906 respectively. The result point out that Holt's linear model was the most suitable for prevalence of tuberculosis in India.

1. Introduction

Tuberculosis (TB) is a most important health problem in India has the high-burden countries. It causes ill-health amongst millions of persons every year and had the highest amount of cases India accounted for 27% of global Tuberculosis notifications in 2014 and 12% of notifications were from the private segment. In six of 41 countries that reported data, higher than 50% of notifications were from community referrals in areas where community engagement activities were in place [1]. The World Health Organisation (WHO) figures for 2015 provide an estimated incidence figure of 2.2 million cases of TB for India out of a global occurrence of 9.6 million. In India 2014, there were an estimated of which approximately 1609547 new and relapse cases, 1055386 were men, 554161 were women, 140 000 were children and 95709 (6%) cases aged below 15 years, female and male ratio is 1.9 [2]. The amount of TB deaths is timely identification and correct treatment, almost all people with TB can be cured.

2. MATERIALS AND METHODS

The current study was a time-series data on numbers of tuberculosis for the year from 1990 to 2014 in India has been collected from the web site www.who.int/tb/

2000 *Mathematics Subject Classification.* Primary 62N05; Secondary 90B25.
Key words and phrases. Stochastic, Tuberculosis, high-burden, prevent disease, epidemic, criteria.

data maintained by the Department of The World Health Organization (WHO) [3]. The data analysed using Statistical Package for Social Sciences (23 version of SPSS as it is licensed with the SVMCH & RC) and fit the best suitable exponential smoothing model for the tuberculosis data. The performance criteria were used to determine if the model was accurately specified. Forecasting of the occurrence of yearly TB cases be also done using the best fit.

2.1. Time sequence forecasting techniques model. Forecasting methods based on simple exponential smoothing (SES) and moving averages are also used for forecasting. A time series when there is no trend or seasonal pattern, but the mean of the time sequence Y_t is slowly changing over time. calculate by smoothing weighted averages and provide short-term forecasts (2.1) [4].

$$F_t^{'} = \alpha Y_t + (1-\alpha)F_{t-1}^{'}, \quad \hat{F}_{t+1} = F_t^{'} \tag{2.1}$$

Here, $\hat{F}_{t+1}$ refers to the estimate value for the forthcoming period, α refers to smoothing coefficient (the range of $0 < \alpha < 1$); Y_t refers to true index value during time t or new observation and $F_t^{'}$ refers to former smoothed value. The most important point to consider at this point is the determination of α so the errors are minimized.

Holt's parameter two exponential smoothing model is an extension of single exponential smoothing model to take into account a possible linear trend and provide short-range forecasts model. This model is used when a sequence has no seasonality and exhibits some form of trend. The Holt's linear model is indicated in equation (2.2) [5].

$$\begin{aligned} L_t^{'} &= \alpha Y_t + (1-\alpha)(L_{t-1}^{'} + T_{t-1}) \\ \hat{F}_{t+1} &= L_t^{'} + hT_t. \end{aligned} \tag{2.2}$$

Here, $T_t = \beta(L_t^{'} - L_{t-1}^{'}) + (1-\beta)T_{t-1}$. In addition, β and $1-\beta$ are the parameter of the method and take a value 0 and 1. Brown's smoothing system characteristic of SES is suitable only for stochastic series with level and trend pattern is indicated in equation (2.3) [6].

$$L_t^{'} = \alpha Y_t + (1-\alpha)L_{t-1}^{'} \tag{2.3}$$

Here, $T_t = \alpha(L_t^{'} - L_{t-1}^{'}) + (1-\alpha)T_{t-1}$, α are the parameters of the method and take a smooth value 0 and 1. The $h-$step ahead prediction equation is $\hat{Y}_{t+h} = L_t^{'} + ((h-1) + \frac{1}{\alpha})T_t$, $h = 1, 2, \ldots, n$. This $h-$steps ahead by taking the last available estimated level state and multiplying the last available trend (slope), T_t, by $((h-1)+\frac{1}{\alpha})$. Damped exponential smoothing constant linear trend seems to be inappropriate, especially for longer time horizons. As a result, empirical evidence indicates that SES as well as Holt's model tend to over-or under forecast. If we choose to expand Holt's model on the damping factor ϕ, is indicated in equation

(2.4) [7].

$$\begin{aligned} L_t^{'} &= \alpha Y_t + (1-\alpha)(L_{t-1}^{'} + \phi T_{t-1}) \\ T_t &= \beta(L_t^{'} - L_{t-1}^{'}) + (1-\beta)\phi T_{t-1} \\ F_{t+h} &= L_t^{'} + \sum_{i=1}^{h} \phi^i T_t \end{aligned} \tag{2.4}$$

Here, $\phi-$damping factor with a value between 0 and 1. Under the Holt's model with additive damped trend the forecast.

2.2. Comparison between the smoothing models in simulation performance. Four indexes were used in comparisons of the fitting effectiveness of the four types of Exponential Smoothing model: Root Mean Square Error (RMSE)= $\left[\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2/n\right]^{1/2}$, Mean Absolute Error (MAE)= $\sum_{i=1}^{n}\left|(Y_i - \hat{Y}_i)\right|/n$, and Mean Absolute Percentage Error (MAPE)= $\frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100$, where Y_t is the real incidence of tuberculosis number at time t, $\hat{Y}_t$ is the forecasted incidence of tuberculosis at time t, and n is the amount of prediction. NBIC$(p,q) = In\ v^*(p,q) + (p+q)[In(n)/n]$; where p and q are the order of AR and MA processes respectively and n is the number of observations in the time series and v^* is the estimate of white noise variance σ^2. Good fitness performance is demonstrated with these four indices showing as small a value as possible.

3. RESULTS AND DISCUSSION

Stochastic model used to predict the future distribution of tuberculosis cases. Times series analysis is a quantitative method for establish pattern in data collected regularly over time. These patterns to estimate future values and cope with uncertainty about the future. For the forecasting, we obtained the data for the Indian countries.

Table 1. Indian incidence number of tuberculosis(per 100,000 people) from 1990 to 2014

Year	TB	Year	TB	Year	TB	Year	TB	Year	TB
1990	217	1995	216	2000	216	2005	209	2010	185
1991	216	1996	217	2001	216	2006	205	2011	180
1992	216	1997	217	2002	215	2007	201	2012	176
1993	216	1998	217	2003	214	2008	196	2013	171
1994	216	1999	217	2004	212	2009	191	2014	167

Source: The World Health Organization (WHO)www.who.int/tb/data.

The data were from 1990 to 2014 which means that there were 25 yearly data points in Table 1. According to WHO guidelines, (5.9%) were pediatric tuberculosis cases and (94.1%) were adults in Table 2. The smoothing method can remove the casual fluctuations of occurrence of Tuberculosis series and identify the major

TABLE 2. Distributionof TB cases by age and sex, 2014

Gender	Age (in years) 0 -14	≥ 15	Total
Female	(3.1)%	(31.3)%	(34.4)%
Male	(2.8)%	(62.8)%	(65.6)%
Total	(5.9)%	(94.1)%	(100)%

trends, but the method doesn't consider the impact of long-term data. The observed values are measured in smoothing method and different weights are given according to the proximity of the period, which can make the predict value closer to the observed values. Exponential coefficients including lies between 0.1 and 1.0 are calculated respectively. The prediction of parameter value among Holt's, Damped, Brown's and exponential smoothing model are the principle for selection is the minimum error value. For the single exponential smoothing model (2.1), $\alpha = 1$. For the Holt's model (2.2), $\alpha = 0.8$ and $\beta = 1.0$. For the Brown's exponential smoothing model (2.3), $\alpha = 0.928$. For the Damped exponential smoothing (2.4), $\alpha = 0.869$, $\beta = 1$ and $\phi = 1$ are displayed in Table 3.

TABLE 3. Estimate parameter and performance value for four types of time series model.

Model	Parameter	SE	t	Sig.	R^2(%)	RMSE	MAPE	MAE	Normalized BIC
SES	$\alpha = 1$	.212	4.728	.003	.967	3.014	1.097	2.080	2.335
Holt's	$\alpha = 0.8$	.214	3.734	.000	.998	0.772	0.305	0.618	0.096
	$\beta = 1.0$	.459	2.179	.000					
Brown's	$\alpha = 0.928$	.093	9.936	.000	.998	0.779	0.316	0.629	0.371
Damped	$\alpha = 0.869$	.225	3.853	.014	.998	0.786	0.310	0.623	0.260
	$\beta = 1$	.526	1.901	.040					
	$\phi = 1$	.054	18.683	.006					

To conclude which forecasting model provides the best prediction of tuberculosis cases (TB). The India TB data points are used for comparing with the forecast by computing the performance value among the actual and forecast values. Seen from Table 3, RMSE of Holt's linear model (0.772) is smaller than that of Brown's model (0.779), Damped model (0.786) and exponential smoothing model (3.014). MAPE of Holt's linear model (0.305) is smaller than that of Brown's model (0.316), Damped model (0.310) and exponential smoothing model (1.097). MAE of Holt's linear model (0.618) is smaller than that of Brown's model (0.629). Damped model (0.623) and exponential smoothing model (2.08). Normalized BIC of Holt's linear model (0.096) is smaller than that of Brown's model (0.371), Damped model (0.26) and exponential smoothing model (2.335). R^2 value of all exponential smoothing models is 97% and there is significant. The better prediction model for tuberculosis is Holt's, Brown's and Damped model. Therefore, relative to the selected SES model is higher fluctuation then other three models and is only suitable for the short term forecast analysis of time series with smooth fluctuation and larger

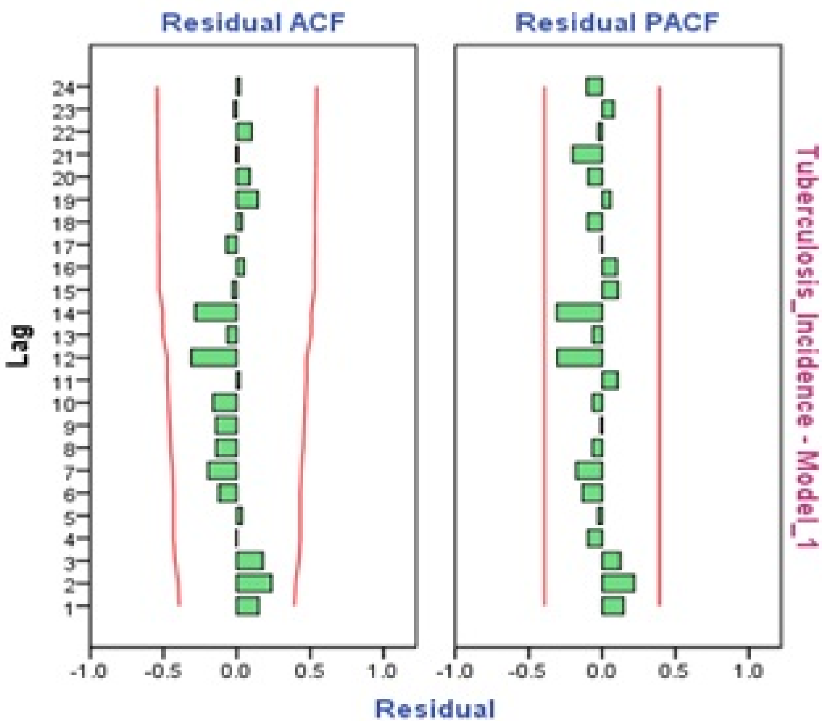

FIGURE 1. Residual plot of ACF and PACF of fitted Holt's linear model

error for higher fluctuation. This Holt's linear model ($\alpha = 0.8, \beta = 1.0$), passed the parameter ($p < 0.01$) and residual ($p > 0.05$) tests, with high and lowest RMSE, MAPE, MAE, and NBIC 99%, 0.772, 0.310, 0.618, 0.906 respectively. The Holt's linear model is more applicable to Incidence of tuberculosis trend forecast to prevent disease epidemic. This is followed by Holt's, Damped, Brown's and SES are another way of the residual values ACF and PACF values were found to be within the 95% confidence bound the residuals due to the Holt's linear model possessed a white noise in Figure 1. In order to forecast for five years forward, we use all the 25 data points in Table 1 and therefore in equation (2.2) the smoothing parameter values used $\alpha = 0.8$ and $\beta = 1.0$. The trend in Incidence of Tuberculosis based on Holt's linear Exponential smoothing model is depicted in the figure 2. Decreasing trends was observed for the prediction of after that six years tuberculosis value in India. The prediction of tuberculosis based on Holt's models is the best forecasting model since it has the smallest variance are given prediction value in the Table 4. For the situation with larger data fluctuations, the decision-makers should dialectically use Holt's linear model output and adjust weight of prediction value according to the actual situation such as considering the time, policies and other related factor.

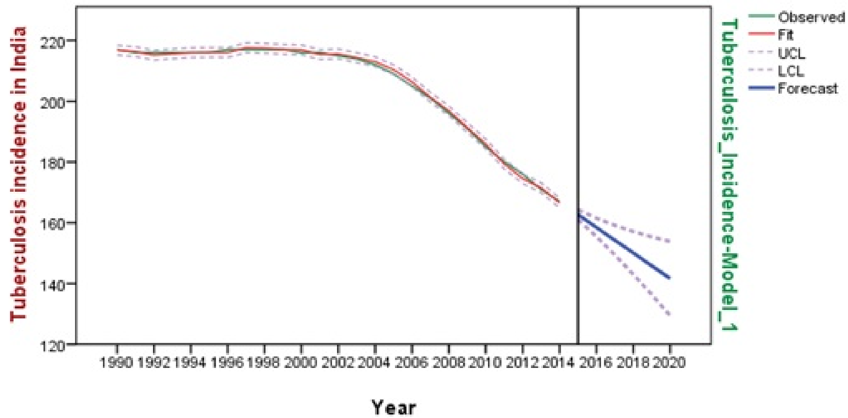

FIGURE 2. Trends in incidence of tuberculosis based on Holt's linear model

TABLE 4. Prediction of Indian incidence tuberculosis(per 100,000 people) till year 2020.

Year	Holt's Linear Smoothing model		
	LCL	Prediction	UCL
2015	161	163	164
2016	155	158	162
2017	149	154	159
2018	143	150	157
2019	136	146	155
2020	129	142	154

UCL =Upper Confidence Limit and LCL= Lower Confidence Limit

CONCLUSIONS

Among the different exponential smoothing model the data set contains the single exponential smoothing technique was not found suitable then other model are suitable. Holt's model was found to be an appropriate model and selected based on model selection criteria. These remaining three models were found to be appropriate model to forecast the tuberculosis data set. Because of the trend pattern of smear positive tuberculosis incidence, time series models with trend Holt's linear model are suitable for prediction. The results of this prediction shows degreasing trend in smear positive tuberculosis cases in India.

References

1. Raviglione M, Uplekar M. WHO's new Stop TB strategy. The Lancet,2006;367:952–5.
2. Uplekar M, Weil D, Lonnroth K, Jaramillo E, Lienhardt C, Dias HM, et al. WHOs new End TB Strategy. The Lancet. 2015; 385:1799–801.
3. (2015) WHO TB Report 2014. *http://www.who.int/tb/publications/global-report/en/*

4. Makridakis, S., Wheelwright, S. C. and Hyndman, R. J. (1998). Forecasting - Methods and Applications. John Wiley and Sons, New York.
5. Holt C. C. (2004). Seasonal and trends by exponentially weighted moving averages, International Journal of Forecasting, Vol. 20(1), pp. 5–10.
6. Brown, R. G and R. F. Meyer and D. A. D'Esopo (1961). The fundamental theorem of exponential smoothing, Operations Research, Vol. 9(5), pp. 673–685.
7. Hyndman R., A. B. Koehler, J. K, Ord and R. D, Snyder (2008). Forecasting with Exponential Smoothing: The State space Approach, Berlin: Springer–Verlag.

Assistant Professor cum Statistician, Dept. of Community medicine, Sri Venkateshwaraa Medical College Hospital & Research Centre, Puducherry, India.

E-mail address: thirustat10@gmail.com