Statistics and Applications {ISSN 2454-7395(online)}
Volume 19, No. 2, 2021 (New Series), pp 147-159

Modeling of Mobile Telephone Subscribers Using Piecewise Nonlinear Growth Models

P.K. Das[1] **and Anita Pal**[2]
[1]*Department of Quantitative Techniques, Indian Institute of Foreign Trade, Kolkata-700107, India*
[2]*Department of Mathematics, National Institute of Technology, Durgapur-713209, India*

Received: 04 April 2020; Revised: 04 December 2020; Accepted: 09 December 2020

Abstract

The growth trajectory of the mobile telephone subscribers is highly nonlinear. The piecewise nonlinear growth model, comprising the well-known Gompertz and Bass models, was shown to be adequately describing the underlying data generating process of mobile telephone subscribers. This study used monthly time series data from March 1997 to December 2018 of telecom Circle A, representing the industrially advanced states like Maharashtra, Gujarat, Andhra Pradesh, Karnataka, and Tamil Nadu, was applied to develop the model. We partition the monthly data into the analysis sample (March 1997 to December 2017) and the test sample (January 2018 to December 2018). The parameters of the piecewise nonlinear model were estimated using Levenberg-Marquardt and sequential quadratic programming algorithms. The piecewise nonlinear model comprising Gompertz and Bass growth models was suitable for describing the monthly mobile subscribers' data in Circle A. The developed model was statistically validated using an appropriate *coefficient of determination* for the nonlinear models and *Root Mean Squared Error* (*RMSE*). We found the *RMSE* to be comparable for both training and the test sets. The forecasting capabilities of the piecewise nonlinear model, under mild violation of residual diagnostics, are compared to exponential smoothing (Holt's) and Gompertz models. We compared the performance of the model to the best fit Gompertz and Holt's models. In the test sample, we found the *RMSE* to be lower in the piecewise nonlinear model comprising Gompertz and Bass compared to the Holt's as well as the Gompertz model. We computed the forecast values of the subscribers during April-December 2020 using the developed model. As evident from the test sample and the published data of the Telecom Regulatory Authority of India (TRAI), the prediction from the developed model is lower than the actual values. The maximum potential number of subscribers in Circle A was 421.545 million, likely to be achieved in 2027. However, as the model predicted values are marginally smaller than the actual values, the maximum potential is expected to be completed before 2027.

Key words: Piecewise nonlinear regression model; Gompertz model; Bass growth model; Sequential quadratic programming algorithm; Holt's model; Root mean squared error.

Corresponding Author: P.K. Das
Email: pkdas@iift.edu
Disclaimer: The views expressed in this paper are solely those of the authors and do not represent our employers' opinions.

1. Introduction

Indian telecommunication market is the second-largest in the world. The telecom sector showed remarkable growth and contributed substantially to creating new jobs and providing revenue to the Government. The industry will contribute ₹14.5 lakh crore to the economy and support 3 million direct jobs and 2 million indirect jobs by 2020 (GSMA, 2019). In 2015, the telecom sector accounted for 6.5% of India's Gross Domestic Product (GDP). With the leveraging of 5G technology, in 2020, it is estimated to reach 8.2% of India's GDP (ET, 2019, Wikipedia, 2020), if not getting delayed due to the COVID-19 pandemic. The bidding decision of the 5G spectrum and the telecom service providers' price in different circles would be based on the demand in that circle, among other parameters. All the stakeholders would be interested to know the future potential of the telecom markets in terms of subscriber base in the years to come. We can estimate the demand of a circle by predicting the number of mobile subscribers. There are four telecom circles in India, namely, metro, A, B, and C. In this study, we develop a model to forecast the total mobile subscribers of Circle A, which is comprising of industrially advanced states *viz*., Maharashtra, Gujarat, Andhra Pradesh, Karnataka, and Tamil Nadu. The number of subscribers (henceforth by subscribers, we would refer to customers subscribed to GSM services only) attained from 9698 in March 1997 to 3505.27 lakhs in December 2018 in Circle A – an impressive growth in subscriber base (COAI, 2018, TRAI, 2018).

In this study, our objectives are: (i) to develop an appropriate model to forecast the numbers of mobile subscribers in Circle A; (ii) to apply the developed model to forecast the numbers of mobile subscribers in Circle A. The modeling approach employed is that of piecewise nonlinear growth models. The results so obtained are compared to the exponential smoothing (Holt's) and Gompertz models. We compared the piecewise model's performance to the Holt's and Gompertz models using the test set. We divide the paper into five sections. Section 2 reviews the existing literature on piecewise nonlinear regression and exponential smoothing models and their applications. We discussed models, parameter estimation, fit statistics, and model selection metrics in Section 3. We discussed the results and salient findings in Section 4. Finally, in Section 5, we present the conclusion and the way forward.

2. Literature Review

The piecewise regression models are also known as segmented regression or broken stick regression models. In Econometrics, it is known as interrupted time series regression (Linden and Arbor, 2015). In this method, we partition the outcome (or the study) variable into regions, and a separate model is fit to each part. The piecewise regression is employed when the data is hypothesized to have been generated by more than one model (McZgee and Carleton, 1970). The piecewise linear regression models are pragmatic in the bio-physical (Vieth, 1989, Malash and El-Khaiary, 2010) and socio-economic domains (Birgit, 2006). The piecewise nonlinear regression models are also applied to biological and socio-economic studies (Oh and Kim, 2002, Maceina, 2007, Morrell *et al*., 1995, Vanli and Kozat, 2014).

The application of the growth models is the most popular approach to study the growth trajectory of mobile subscribers in different markets, namely, Central and Eastern Europe (Gruber, 2001), Asia Pacific region (Wenrong *et al*., 2006), Greece (Michalakelis *et al*., 2008) and Taiwan (Wu and Chu, 2010). Several researchers (Sridhar, 2010; Hedau and Soni, 2016) also studied India's mobile telephone market. The effect of different techno-economic variables, among other things, on the growth of mobile services in different regions in India using panel data, was studied by Sridhar (2010). It was found that competition and network were the two crucial variables to impact mobile services' growth. Though it provided significant insight into the impact of variables on mobile services' growth, it failed to make a time series forecast. The major lacuna in Hedau and Soni (2016) were in the model development and parameter estimation. The linearized model parameters were estimated; however, the original model was recommended as a forecast model. We can overcome this shortcoming by employing nonlinear estimation procedures. Though several studies were conducted to track the growth trajectory with limited success, no attempt has been made to engage piecewise nonlinear growth models to understand mobile subscribers' growth path. As growth models are mechanistic, they have advantages in understanding the data generating process and its future potential. We discuss the nonlinear growth models, piecewise nonlinear growth models, and related issues in the next section.

3. Models and Methodology

3.1. Nonlinear growth models

Let $n(t)$ and $N(t)$ denote the number of subscribers and the cumulative number of subscribers, respectively, to mobile service at time t in Circle A. If t_0 denotes the time at starting, *i.e.*, $t = 0$, then at time t, the cumulative number of subscribers can be expressed as:

$$N(t) = \int_{t_0}^{t} n(t)dt$$

where $n(t)$ is the non-cumulative number of adopters at time t. Further, let,

$$\frac{dN(t)}{dt} = \text{rate of growth at time } t\text{, and}$$

$$\frac{1}{N(t)}\frac{dN(t)}{dt} = \text{relative rate of growth at time } t.$$

Let K denote the total number of potential subscribers in Circle A. It is also known as the carrying capacity or maximum potential of the system, *i.e.*, the markets of Circle A. Let us assume that:

(i) the rate of growth is proportional to the interaction of adopters and non-adopters. This can be expressed as:

$$\frac{dN(t)}{dt} \propto N(t)\left[1 - \frac{N(t)}{K}\right]. \tag{1}$$

In differential equation form, it becomes:

$$\frac{dN(t)}{dt} = rN(t)[1 - N(t)/K]. \tag{2}$$

Here, r is the intrinsic growth rate. The solution to this equation yields the following model:

$$N(t) = \frac{K}{1 + \frac{(K - N_0)}{N_0} exp(-rt)}, \tag{3}$$

where N_0 is the number of subscribers at $t = 0$.

Reparametrizing, the model can be written as:

where $B = (K - N_0)/N_0$.

$$N(t) = \frac{K}{1 + B * exp(-rt)}, \tag{4}$$

This model is known as the logistic growth model. The inflection point of this model is at $K/2$. The model described in Eq. (3) or the reparametrized model in Eq. (4) appeared deterministically as if data never deviates from the model. It is unrealistic. To make the model realistic, independently, identically, and normally distributed error term is added to the right-hand side (RHS) of the mathematical model. The resulting nonlinear regression model is:

$$N(t) = \frac{K}{1 + B * exp(-rt)} + \varepsilon_t \tag{5}$$

(ii) The relative rate of growth is proportional to the logarithm of the ratio of carrying capacity to subscribers' number at time t. Hence, this can be represented as:

$$\frac{1}{N(t)} \frac{dN(t)}{dt} \propto \ln\left(\frac{K}{N(t)}\right). \tag{6}$$

Therefore, the model in differential equation form can be expressed as:

$$\frac{dN(t)}{dt} = rN(t) \ln\left(\frac{K}{N(t)}\right) \tag{7}$$

with the boundary condition $N(t = t_0) = N_0$ = cumulative number of adopters at time t_0.

The parameter r is known as the intrinsic rate of growth. The solution to the above differential equation results in the following model:

$$N(t) = K * exp(-B * exp(-rt)) \tag{8}$$

where $B = ln(N_0 / K)$.

This model is known as the Gompertz model. By adding an error term to the RHS, we obtain the following statistical model:

$$N(t) = K * exp(-B * exp(-rt)) + \varepsilon_t . \tag{9}$$

The model is asymmetric, and the point of inflection is at K/e.

(iii) the rate of growth is proportional to the number of non-adopters, which can be expressed as:

$$\frac{dN(t)}{dt} \propto [K - N(t)]. \tag{10}$$

In the differential equation form, it becomes:

$$\frac{dN(t)}{dt} = r[K - N(t)]. \tag{11}$$

The solution to this equation yields the following model:

$$N(t) = K - (K - B) * exp(-rt), \tag{12}$$

where $r > 0$ and $K > B > 0$. Here, K is the maximum potential, and B is the number of subscribers at $t = 0$. The statistical model can be written as:

$$N(t) = K - (K - B)*exp(-rt) + \varepsilon_t \quad (13)$$

The model in Equation 13 is known as monomolecular (MM). The models in Equations 5, 9, and 13 are S-shaped growth models. These are nonlinear models in the statistical regression sense because at least one parameter of these models appears nonlinearly. The three parameters *viz*., *K*, *B*, and *r* of the models in Equations 5, 9, and 13, are estimated. For further details on S-shaped nonlinear growth models and the Richards model, readers can refer to Seber and Wild (2003).

(iv) The rate of growth is influenced by two types of subscribers, namely, innovators and imitators. We present the differential equation below:

$$\frac{dN(t)}{dt} = p\,[m - N(t)] + \left(\frac{q}{m}\right) N(t)[m - N(t)] \quad (14)$$

Here, *m* is the market potential, *p* and *q* are the coefficients of innovation and imitation. The solution of the above differential equation results in the following model:

$$N(t) = m\frac{1 - exp^{-(p+q)t}}{1 + (\frac{q}{p}) * exp^{-(p+q)t}} \quad (15)$$

By adding an error term to the RHS, we obtain the following statistical model:

$$N(t) = m\frac{1 - exp^{-(p+q)t}}{1 + (\frac{q}{p}) * exp^{-(p+q)t}} + \varepsilon_t \quad (16)$$

In this model, $m > 0$, $p > 0$, and $q > 0$. It is also a nonlinear model in the regression sense. The parameters of the model *viz*., *m*, *p*, and *q* are to be estimated. For further details on the Bass model, the readers can refer to Bass (1969) and Rogers (2003).

3.2. Piecewise nonlinear growth models

The piecewise nonlinear model of the following type is considered in this study:

$$(t < T^*) * N1(t) + (t >= T^*) * N2(t). \quad (17)$$

In Equation (17), T^* is the value of *t* at which the growth trajectory is found to be changing from one model to another model. It is also known as the knot, breaking point, change point, or joining point. The two nonlinear functions are denoted by $N1(t)$ and $N2(t)$. If $(t < T^*)$ is true, it returns one else zero. Similarly, if $(t >= T^*)$ is true, it returns 1 else zero. Here, $N1(t)$ can be any growth model *viz*., logistic, Gompertz, monomolecular, Bass, Richards, or any other growth model. The function, $N2(t)$, can also be any growth model *viz*., logistic, Gompertz, monomolecular, Bass, Richards, or any other growth model. As an example, let us consider the following combination:

$$N(t) = (t < T^*) * (K * exp(-B * exp^{(-rt)})) \quad (18)$$

$$+ (t \geq T^{*}) * m \frac{1 - exp^{-(p+q)t}}{1 + \left(\frac{q}{p}\right) * exp^{-(p+q)t}} + \varepsilon_t$$

In this segmented (piecewise) study, the first segment of the sample data is hypothesized to have been generated by a Gompertz model. The second segment of the sample data is hypothesized to have been caused by a Bass model. In general, when we consider 'n' number of models for modeling a data set having two segments, there can be n^2 number of piecewise models.

3.3. Estimation of parameters

The nonlinear model differs in their estimation properties from linear regression models. Under the assumption of an independently and identically distributed normal error term, the linear model gives rise to unbiased, normally distributed minimum variance estimators. Nonlinear regression models tend to do so as the sample size becomes very large (asymptotically) (Ratkowsky, 1990, Ross, 1990, Intriligator, 1996). Like linear regression, in nonlinear regression also normal equations are obtained. However, these normal equations are nonlinear, and no explicit solutions can be obtained. Different algorithms are available in the literature to solve nonlinear normal equations. Three main algorithms are (i) Gauss-Newton, (ii) Sequential Quadratic Programming, and (iii) Levenberg-Marquardt. The sequential quadratic programming algorithm is appropriate for nonlinear constraint models. It is a combination of Lagrangian relaxation, active set strategy, and Newton-Raphson methods. The algorithm yields stable solutions in the majority of situations. The details of these methods and their merits and demerits are available in the literature (Draper and Smith,1998; Nocidal and Wright, 2006). These algorithms are iterative and require starting values of the parameters. A good starting value can ensure global convergence and can obtain a minimum value of the loss function. The sum of squared residuals can be considered a loss function in estimating piecewise nonlinear regression models' parameters. The choice of good starting values can influence the convergence of the algorithm in locating the fitted value or between rapid and slow convergence to the solution. However, there is no standard procedure for computing the starting values of the parameters. Sometimes a combination of two or three methods results in good starting values. In this study, a combination of techniques is used to obtain the starting values of the parameters. IBM SPSS Statistics version 26 software package is used to estimate the models' parameters and computation of goodness of fit measures (IBM, 2019). The goodness of fit of the nonlinear model is assessed by the *coefficient of determination* (R^2).

However, as Kvalseth (1985) pointed out, eight different expressions for R^2 appear in the literature. One of the most frequent mistakes occurs when the fits of a linear and a nonlinear model are compared by using the same R^2 expression. Thus, a logistic or a Gompertz model may first be linearized by using a logarithmic transformation and then fitted to data by using the ordinary least squares method. The R^2-value is then often calculated using the log of observed and log of predicted data points. The R^2 is, erroneously, interpreted as a measure of goodness of fit of even the original nonlinear model. Scott and Wild (1991) have given an example where two models are identical for all practical purposes and yet have very different values of R^2 calculated on the transformed scales. Kvalseth (1985) has emphasized that the following R^2

$$R^2 = 1 - \frac{RSS}{TSS}, \tag{19}$$

where *RSS* is the residual sum of squares, and *TSS* is the total sum of squares, which is entirely appropriate even for nonlinear models. We present below the other necessary summary measures for nonlinear models:

$$Root\ Mean\ Squared\ Error\ (RMSE) = \sqrt{\frac{\sum_{t=1}^{T}(N(t) - \widehat{N(t)})^2}{T}}. \tag{20}$$

Here, T is the total number of observed values. $N(t)$, and $\widehat{N(t)}$ are the number of actual and the predicted subscribers.

Mean Absolute Percentage Error (*MAPE*) is defined as:

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{N(t)-\widehat{N(t)}}{N(t)}\right| * 100 \tag{21}$$

$N(t)$, and $\widehat{N(t)}$ are the number of actual and the predicted subscribers, respectively.

4. Results

Monthly data on mobile subscribers in the Circle A, from March 1997 to December 2018, was collated from the Cellular Operators Association of India (COAI) (www.coai.in). Currently, in the COAI repository, monthly data is available from January 2005 - December 2018. The data before 2005 was collated from the same repository in 2013 when it was available in the database. Data is currently available on the TRAI website; however, there is a difference in the subscribers' numbers reported by COAI and TRAI. For example, in January 2018, COAI reported 340.41 million subscribers, whereas TRAI reported 402.81 million subscribers in Circle A. To avoid mixing data from two sources, only data from COAI, which is available until December 2018, has been analyzed. Moreover, cumulative data is required for the estimation of parameters of the growth models. The cumulative data is essentially increasing or equal to the previous observation where $N(t+1) >= N(t)$. However, in the reported data in some months, this essential criterion has been violated. Wherever $N(t+1)$ was reported to be smaller than the $N(t)$, it has been imputed by the $N(t)$. With this simple and essential imputation, the monthly data were pre-processed to estimate the model's parameters. Further, the complete data set was partitioned into an analysis sample and the test sample. The monthly data from March 1997 to December 2017 were used as the analysis sample, and January 2018 to December 2018 was retained as the test sample. The line plot of the data is presented in Figure 1. The line plots of Figure 1 and Figure 2 are generated in R studio (R Core Team, 2016, RStudio Team, 2015). Figure 1 depicts the growth trajectory of the corrected level data. The growth trajectory appears to be S-shaped until 2016 and follows a different path (Figure 1).

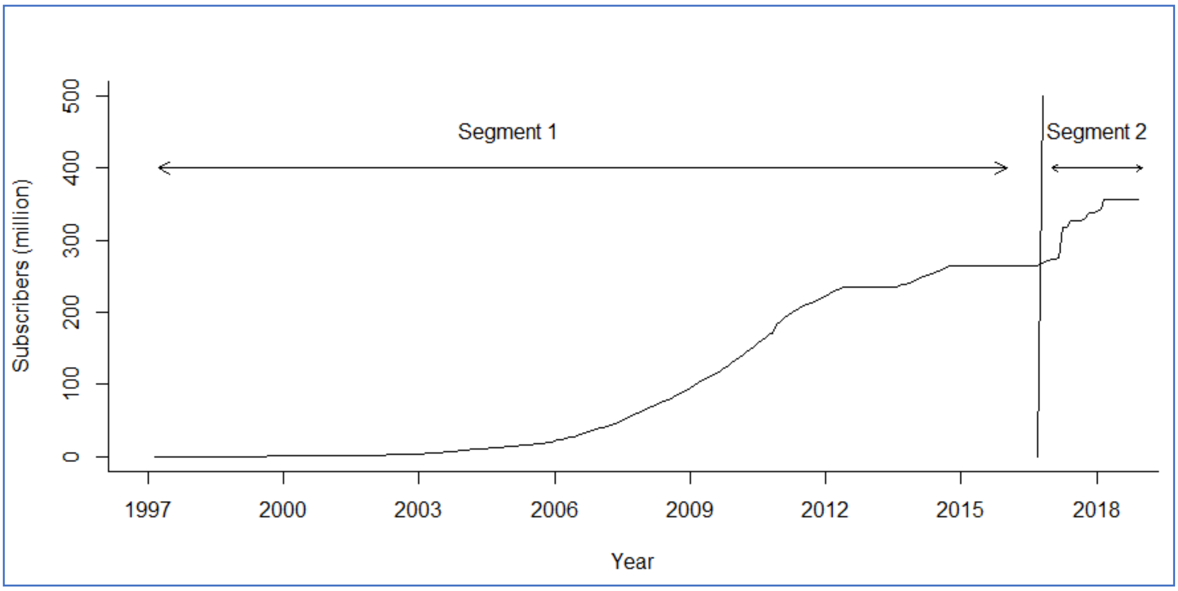

Figure 1: Partitioning of the data into Segment 1 and Segment 2 by the vertical line

It is hypothesized that the underlying data generating process follows an S-shaped model. The traditional growth models, namely, logistic, Gompertz, monomolecular, Bass, and Richards, were fitted to the data using a nonlinear estimation method. The results are presented in Table 1.

Table 1: Results of fitting nonlinear growth models to subscribers' data of Circle A

Fit Statistics/Model	**Logistic**	**Gompertz**	**MM**	**Bass**
R^2	0.98	0.99	0.952	0.98
RMSE	15.89	13.85	26.64	15.84
MAPE	1.38	0.36	49.94	0.55
K (millions)	318.79	359.29	IE	
m (millions)				319.74
IE: Inadmissible estimate				

The Gompertz model appears superior to other models in R^2, *RMSE*, and *MAPE*. The next appropriate model is the Bass model. The Richards model resulted in the non-convergence of the iterative algorithm. The maximum potential of the market is estimated using the parameter *K*. The actual maximum number of subscribers in the sample is 357.378 million. The monomolecular model resulted in an inadmissible estimate. The logistic and Bass models resulted in an estimate lower than the actual maximum value. The estimate of *K* given by the Gompertz model is only marginally higher than the actual maximum value. The Gompertz model is appropriate for describing the subscribers' data of Circle A based on the fit statistics. The estimated parameter value of *K* in the Gompertz model is admissible. We found that the results did not support the assumption of the normality of residuals in the Gompertz model using the Anderson-Darling test.

Moreover, the *RMSE* of 13.85, though minimum among all the models, is not small in an absolute sense. Hence, we look for alternative models for describing the sample data. To this end, we employ piecewise nonlinear growth models. The change point was identified visually by examining the graph (Figure 1) and scanning the values of the data during 2016 and found to be in

August 2016 ($t < 235$); t is the index of the time series representing month and year of the series. The change point divides the series into two segments, namely, segment one and segment two. We present this in Figure 1.

We hypothesize that the two segments can be modeled using one model applied to two segments separately or by two models. To evaluate this hypothesis, we fitted all 25 model combinations. To validate the piecewise nonlinear growth models' performance, we have partitioned the sample data to the training set (March 1997 to December 2017) and the test set (January 2018 to December 2018). Considering that the valuable information is present in the recent observations, we have retained only 12 observations (*i.e.*, one year's data) in the test set.

The piecewise nonlinear growth models are fitted to the training set, and the performance of the model was evaluated on the test set. Out of all 25 combinations, not all combinations converged or resulted in admissible parameter estimates. The combinations which converged and resulted in admissible parameter estimates are presented here. Two combinations, namely the Gompertz model for both the segments (let us name it Gompertz-Gompertz (*G-G*)), and the Gompertz model for Segment 1 and Bass model for Segment 2 (let us name it Gompertz-Bass (*G-B*)) were found to be comparable. We present here two sets of initial values, namely, $K = 200$, $B = 2$, $r = 0.05$, $m = 400$, $p = 0.00005$, and $q = 0.05$; and $K = 300$, $B = 2$, and $r = 0.05$, $m = 400$, $p = 0.00005$, $q = 0.05$. The first three parameters pertain to the Gompertz model, and the following three parameters pertain to the Bass model. These are obtained by combining linearization, intelligent guesses, and property of the model. We present the results in Table 2.

Table 2: Results of fitting piecewise nonlinear growth model to the analysis sample

Fit Statistics/Model	*G-G*	*G-B*
R^2	0.99	0.99
RMSE	6.94	7.14
MAPE	0.46	0.43
Parameter estimate (only maximum potential)		
K (millions)	384.07	
m (millions)		421.54

The results of the *G-G* model presented above are that of a local minimum. The algorithm failed to converge to a global minimum even when widely separated initial values were used. Therefore, it is prudent not to compare the results of *G-G* to that of *G-B*, which resulted in global convergence. However, for the sake of completeness, we presented the results here. The performance of the models (*G-G* and *G-B*) is compared using the test sample. The *RMSE* of the *G-G* model in the test sample was found to be 71.78, whereas the same value of the *G-B* model was 13.12. In the test sample, the value of the *RMSE* of the *G-B* model is much better than the *G-G* model. Furthermore, in the test sample, the value of the *RMSE* of the *G-B* model is marginally better than the only Gompertz model. Let us compare the test series results to that of the exponential smoothing model to decide the final model.

We describe below the exponential smoothing model (also known as Holt's model) (Gardner, 1985, Hanke and Wichern, 2013):

$$y_{t+1} = \alpha x_t + (1 - \alpha)\left(y_t + T_t\right) \qquad (22)$$

$$T_{t+1} = \gamma(y_{t+1} - y_t) + (1 - \gamma)\, T_t \tag{23}$$

$$H_{t+h} = y_{t+1} + hT_{t+1}, \tag{24}$$

where, y_{t+1} = smoothed value for period t+1, α = smoothing constant for the level $(0 < \alpha < 1)$, x_t = observed value in period t, T_{t+1} = trend estimate, γ = Smoothing constant for the trend estimate ($0 < \gamma < 1$), h = number of periods ahead to be forecast, and H_{t+h} = Holt's forecast value for period $t+h$.

The estimated values of alpha and gamma parameters of Holt's model were 0.933 and 0.134, respectively. The *RMSE* of Holt's model was found to be 2.883 and 14.97 for the training and the test samples, respectively. Therefore, the *RMSE* of the piecewise nonlinear growth model comprising Gompertz and Bass models was found to be superior to Holt's model. The residual diagnostics of the piecewise nonlinear model was found to deviate from the assumptions of normality and independence. However, both these deviations are mild and ignored because of the superior comparative performance of the model. Given the above, it can be concluded that the data generating process of the mobile subscribers' data of the Circle A was piecewise nonlinear, which can be modeled by Gompertz and the Bass models. We present the parameters of the final model in Table 3.

Table 3: Parameter estimates of piecewise nonlinear model

Parameters	*K*	*B*	*r*
Estimates	291.965	86.168	.031
Parameters	*m*	*p*	*q*
Estimates	421.545	7.328x10^{-9}	.07

In the final model, the estimated value of the parameter '*p*' is minimal. Such a small value, which is not exactly equal to zero but near zero, can occur and be meaningful in the present context as the data on which the models have been fitted are in millions. We present the final fitted model below:

$$N(t) = (t < T^*) * (291.965 * exp^{-86.168 * exp^{-.031*t}})$$

$$+ (t >= T^*) * 421.545 \frac{1 - exp^{-(7.328 \text{ x } 10^{-9} + .07)\,t}}{1 + \left(\frac{.07}{7.328 \text{ x } 10^{-9}}\right) * exp^{-(7.328 \text{ x } 10^{-9} + .07)\,t}}$$

The actual (dotted line) and the predicted (solid line) trajectory of the mobile subscribers are depicted in Figure 2. We present the forecasted number of subscribers using the piecewise nonlinear growth model (Gompertz-Bass) for the last three quarters of 2020 in Table 4.

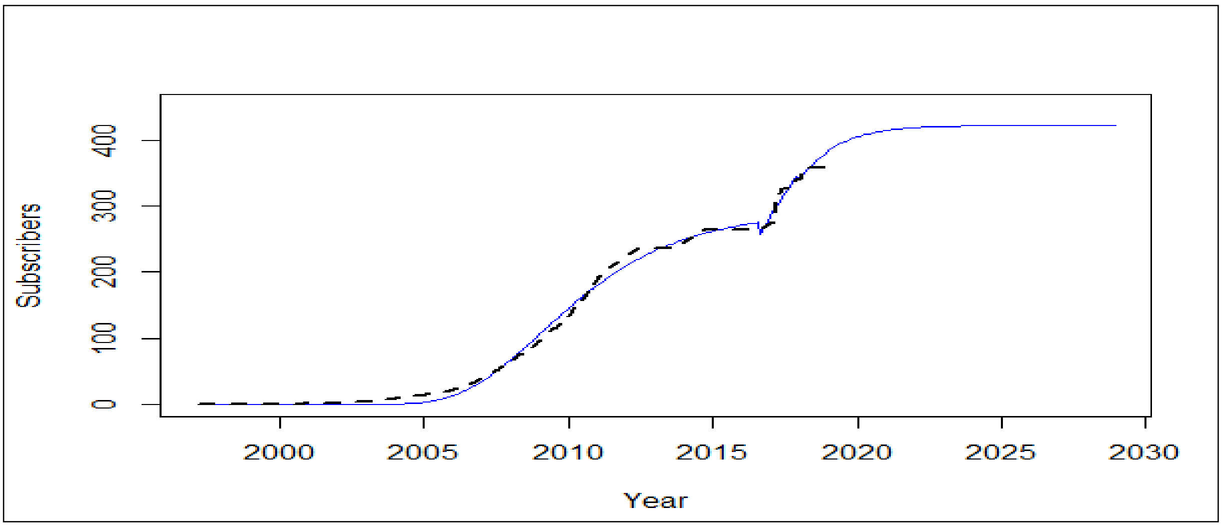

Figure 2: Actual (dotted) and predicted (solid line) mobile subscribers (in millions) in Circle A

Table 4: The forecasted subscribers (in millions) of Circle A in 2020

Month	Forecast		Month	Forecast
April	407.62		September	411.62
May	408.53		October	412.27
June	409.38		November	412.88
July	410.18		December	413.45
August	410.92			

It is evident from the predicted values and the graph (Figure 2) that the circle's maximum potential, which is 421.545 million, is predicted to be achieved in November 2027. However, it has been found earlier that the predictions from the growth models are conservative, and the maximum potential usually is attaining much before the model-predicted date (Das, 2013). Moreover, the data reported by TRAI is much higher than the data reported by COAI. As this study is based on COAI data, actual values are likely to be higher than the forecast provided in this paper.

5. Conclusions

We found the piecewise nonlinear growth model comprising Gompertz and Bass to be appropriate for describing the subscribers' data in Circle A. It confirms our assertion that the underlying data generating process can be divided into two segments, which shows strong evidence of ushering a new growth phase. Despite several issues in the telecom sector, Circle A comprises industrially advanced states like Maharashtra, Gujarat, Andhra Pradesh, Karnataka, and Tamil Nadu, which has entered into a new growth cycle. If this new cycle continues, it is likely to impact the Government in terms of revenue collection and the first- and second-degree stakeholders. As the post-COVID-19 economic scenario is different from the pre-COVID-19 economic scenario, the data policy and pricing of the 5G spectrum are required to be such that it encourages more usages so that the current growth momentum continues.

Acknowledgments

The authors are grateful to the reviewers for diligent comments.

References

Bass, F. M. (1969). A new product growth for model consumer durables. *Management Science*, **15(5)**, 215-227.

Birgit, Strikholm (2006). Determining the number of breaks in a piecewise linear regression model. *SSE/EFI Working Paper Series in Economics and Finance*, **648**, Stockholm School of Economics, The Economic Research Institute (EFI), Stockholm.

COAI (2018). *Annual Report 2018-19*. Cellular Operators Association of India. New Delhi, India.

Das, P. K. (2013). Nonlinear Statistical Model and its Applications to Diffusion of Mobile Telephony in India. *Annals of Pure and Applied Mathematics,* **5** (**1**), 64-70.

Draper, N., and Smith, H. (1998). *Applied Regression Analysis,* 3e. John Wiley, New York.

ET (2019). https://telecom.economictimes.indiatimes.com/news/telecom (Retrieved March 28, 2020).

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

Gruber, H. (2001). Competition and innovation: The diffusion of mobile telecommunications in Central and Eastern Europe. *Information Economics and Policy*, **13**, 19-34.

GSMA (2019). www.gsma.com (Retrieved March 28, 2020).

Hanke, John E., and Wichern, Dean W. (2013). *Business Forecasting*. PHI, New Delhi.

Hedau, Vandit, and Soni, Pragya, (2016). Increment of mobile subscriber in India: Gompertz curve. *Int. J. Res. Innov. Applied Science*, **1**(**1**), 1- 4.

IBM (2019). IBM Corp. Released 2019. *IBM SPSS Statistics for Windows*, Version 26.0. Armonk, NY: IBM Corp.

Intriligator, Michael D., Bodkin, Ronald, and Hsiao, Cheng (1996). *Econometric Models, Techniques, and Applications*. Prentice-Hall.

Kvalseth, T. O. (1985). Cautionary note about R^2. *Amer. Statistician*, **39**, 279-85.

Linden, Ariel, (2015). Conducting interrupted time-series analysis for single- and multiple-group comparisons. *The Stata Journal*, **15(2**), 480-500.

Maceina, M. J. (2007). Use of piecewise nonlinear models to estimate variable size-related mortality rates. *North American Journal of Fisheries Management*, **27**, 971-977.

Malash, G. F. and El-khaiary, M. I. (2010). Piecewise linear regression: A statistical method for the analysis of experimental adsorption data by the intraparticle-diffusion models. *Chemical Engineering Journal*, **163(3)**, 256-263.

McZgee, V. E. and Carleton, W. T. (1970). Piecewise regression. *Journal of the American Statistical Association*, **65**(**331**), 1109-1124.

Michalakelis, Christos, Varoutas, Dimitris and Sphicopoulos, Thomas (2008). Diffusion models of mobile telephony in Greece. *Telecommunications Policy*, **32**, 234-245.

Morrell, Christopher, H., Pearson, Jay, D., Carter Ballentine, H. and Brent, Larry, J. (1995). Estimating unknown transition times using a piecewise nonlinear mixed-effects model in men with prostate cancer. *Journal of the American Statistical Association*, **90**(**425**), 45-53.

Nocedal, J., and Wright, S. (2006). *Numerical Optimization*, 2nd *Edition*. Springer.

Oh, Kyong Joo and Kim, Kyoung-jae (2002). Analyzing stock market tick data using piecewise nonlinear model. *Expert Systems with Applications*, **22**, 249-255.

R Core Team (2016). R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ratkowsky, David A. (1990). *Handbook of Nonlinear Regression Models*. Marcel Dekker, New York.

Rogers, Everett M. (2003). *Diffusion of Innovations, 5th Edition*. Free Press, New York.

Ross, Gavin J. S. (1990). *Nonlinear Estimation*. Springer-Verlag, New York.

RStudio Team (2015). RStudio: *Integrated Development for R*. RStudio, Inc., Boston, MA.

Scott, A., and Wild, C. J. (1991). Transformations and R^2. *Amer. Statistician*, **45**, 127-28.

Seber, G. A. F. and Wild, C. J. (2003). *Nonlinear Regression*. Wiley, New York.

Sridhar, Varadharajan (2010). An econometric analysis of mobile services growth across regions of India. *Netnomics*, **11**, 205-220.

Vanli, N. Denizcan and Serdar kozat, Suleyman (2014). A comprehensive approach to universal piecewise nonlinear regression based on trees. *IEEE Transactions on Signal Processing*, **62**(**20**), 5471-5486.

Vieth, E. (1989). Fitting piecewise linear regression functions to biological responses. *Journal of Applied Physiology*, **67**(**1**), 390-396.

TRAI (2018). *Press Release No. 86/2018*. Telecom Regulatory Authority of India, New Delhi, India.

Wenrong, W., Xie, M., and Tsui, K. (2006). Forecasting of mobile subscriptions in Asia Pacific using Bass diffusion model. *IEEE International Conference on Management of Innovation and Technology*, **1**, 300-303.

Wikipedia (2020). *Telecommunications in India*. In *Wikipedia*. Retrieved April 1, 2020, from https://en.wikipedia.org/wiki/Telecommunications-in-India.

Wu, Feng-Shang, and Chu, Wen Lin (2010). Diffusion models of mobile telephony. *Journal of Business Research*, **63**(**5**), 497-501.