International Journal of Engineering & Technology, 7 (2.29) (2018) 912-916

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Novel S-Regression Model on an Auto Price

Fadzilah Salim[1]* , Nur Azman Abu[2]

[1]Faculty Of Engineering Technology, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka.
[2] INSFORNET, Faculty Of Information Technology And Communication, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka.
**Corresponding Author E-Mail: Fadzilah@Utem.Edu.My*

Abstract

A simple linear regression model is useful in a prediction model. A general linear regression beyond a single independent variable is still not popular. A nonlinear regression can be easily produced a better predictive model but it is difficult to construct. The objective of this paper is to propose a technique for predicting the price of used cars in Malaysia using S-shaped curve model. In this paper, the S-shaped Membership Function [SMF] is used as the basis to develop a novel S-Regression model. Comparisons between linear regression, cubic regression and S-Regression have been made on the used car prices. The mean squared error of S-Regression model is found to be closer to cubic regression than the linear regression. S-Regression model is found to be quite suitable to represent the relationship between the price of a used car and the make year of a car. The result demonstrates that the S-Regression model gives better and practical estimate of the price of a used car in Malaysia.

Keywords: *S-Regression model; S-shaped curve; Prediction on used car price*

1. Introduction

Regression analysis is a form of a statistical model. Typically, a regression analysis is used for one or more of three reasons: modeling the relationship between x and y; prediction of the target variable [forecasting]; and testing of hypotheses [1].

Predicting the price of used car is very important prior to a car sale in Malaysia. Due to the economic and financial circumstances and in addition to the cost of car insurance, road tax, fuel and maintenance, purchasing a new car is a luxury thing to do. Thus, many people would prefer to opt for cheaper alternatives, that is, buying a good second hand car or used car.

It is useful to note that prediction and forecasting have been very essential in the business world for a long time. Predicting a price of used car also is a very interesting problem. There are techniques or models used by researchers for predictive analysis. Pudaruth [2] predicted the price of used car by using machine learning techniques. Aimin et al. [3] on the other hand, looked into prediction on the developing trend of Global Electric automobile based on the logistic model. Sharma et al. [4] did a sales forecast of an automobile industry using multiple linear regression model.

Finding the best tool to give a better estimate of the price of used cars can be quite challenging. Predictive modeling can be done by using statistical regressions or machine learning techniques. Aydin [5] found that a spline S-regression model gives the best prediction performance as compared to the other six regression models, namely, linear, logarithmic, power, exponential, inverse, and growth regression models when forecasting natural gas production. Alternatively, this paper concentrates on the prediction model of used car price in Malaysia using S-shaped curve method or S-Regression model.

2. Literature Review

2.1. S-Shaped Curve as a Tool for Prediction

An S-shaped curve is a well-established function. It has been identified throughout modern computational sciences. It has many applications especially in the area of neural network, fuzzy logic, artificial intelligence, and so on. One of the functions that has an S-shaped curve is known as sigmoid function. It is widely used in artificial neural network to introduce the nonlinearity model [6–8] and backpropagation learning [9]. It has also been used to predict and make some decision making on how much money can be saved based on the amount available in the current account balance [10].

An S-shaped curve often refers to the special case of the continuous logistic function which has a mathematical equation as follows [11]:

$$f(x) = \frac{1}{1+e^{-x}} \tag{1}$$

A logistic function is frequently used to model a population growth of some set. It is also commonly used in Statistics to describe a cumulative distribution function. It is a continuous function and an increasing function for all x. Its gradient is always nonnegative for all real value of x. Its second derivative changes from concave upward to concave downward. The graph of equation [1] is shown in Figure 1.

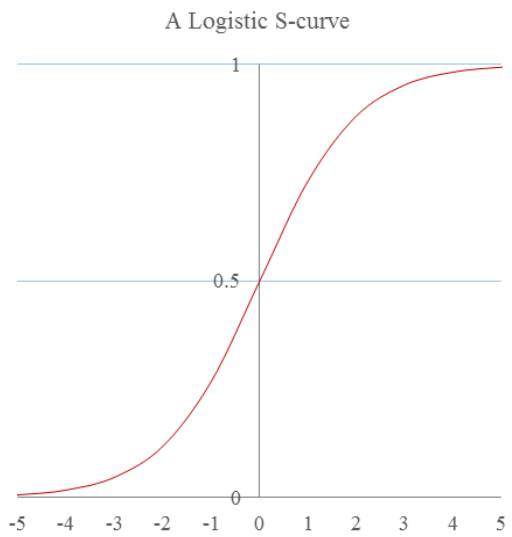

Fig. 1: The Logistic "S" curve

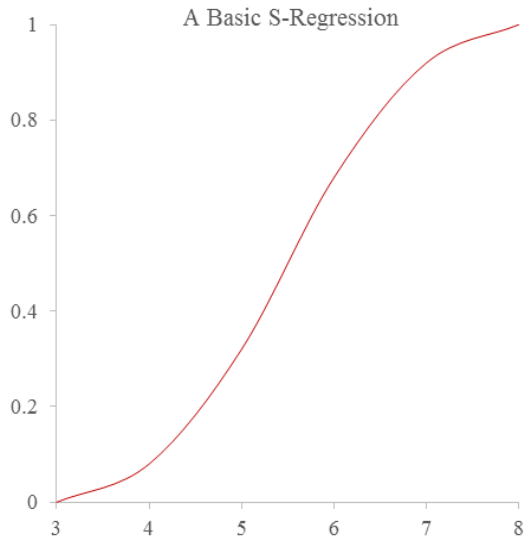

Fig. 2: A Basic S-Regression graph based on SMF

The S-shaped curve can be divided into three stages. The first stage is where the graph starts slowly, then it accelerates in the second stage before slowing down again to reach its saturation level at the final stage. Such curves can be used to capture a diffusion process [12]. The logistic S-curve of natural growth is a basic model of the Volterra-Lotka equations, which are reliable for describing and forecasting different forms of competition and technology substitution [13].

There are many research studies that have been carried out using S-shaped curve in forecasting and predicting analysis. One of the leading forecasting journal which is the International Journal of Technological Forecasting and Social Change published hundreds of articles which mentioned a lot on S-shaped curve on their pages. Further reading on the types of forecasting methods can be referred to [14] which covers all types of forecasting methods from various experts in which some mentioned about S-shaped curve in their chapters.

2.2. A Novel S-Regression Model Based on S-Shaped Membership Function

S-function is becoming popular. An S comparative index function is also used in Fuzzy Logics. The shape of membership function is important for a particular problem since they effect on a fuzzy inference system. An example of S-shaped membership function [SMF] is a spline-based curve mapping on the input vector ***x*** to a membership value [or degree of membership] between 0 and 1. A typical example of SMF is as follows.

$$f(x)=\begin{cases}0, & x\le a\\ 2\left(\dfrac{x-a}{b-a}\right)^2, & a<x\le\dfrac{a+b}{2}\\ 1-2\left(\dfrac{x-b}{b-a}\right)^2, & \dfrac{a+b}{2}<x<b\\ 1, & b\le x\end{cases} \quad (2)$$

An S-shaped curve has also been found in technical analysis of comparative index by the second author [15]. **Figure 2** shows the graph of equation [2] which follows the shape of an "S" with extreme parameters $a = 3$ and $b = 8$.

Although there are numerous research using S-curve shaped as a form of prediction, none has come up with S-Regression model for better estimation. Thus, it is proposed here in this paper that the novel S-Regression model would be developed as a mean to forecast value of a particular time-series data.

3. Methodology/Materials

Data from www.carlist.my website, a famous website of used car price in Malaysia have been gathered. Real data have been collected on the price of used cars and their make years for two established car models on the market. Other features which may affect the price of used cars such as mileage, colors, and engine capacity are not being measured. The two car models selected are a BMW 325i 3 series which is a sports car and a VOLVO S60 Sedan car. The data range in interest is within the year of 2000 until 2016. However, the data available in the website for BMW 325i is only from 2004 to 2012, while for VOLVO S60, the data available is from 2002 until 2014. The samples of the collected data are shown in Table 1 and Table 2.

Table1: Sample data collection of 65 used car price of BMW 325*i* from year 2004 to 2012

Year	Price	Year	Price
2004	36800	2010	91989
2004	37800	2010	108000
2004	35800	2010	110000
2005	62800	2010	96000
2005	45800	2010	81800
2005	37800	2010	89900
2005	53800	2010	102800
2005	38800	2010	115000
2005	58888	2010	98800
2005	28800	2010	106800
2008	63800	2010	103888
2008	79888	2010	105000
2008	56800	2010	93800
2008	70000	2010	78888
2008	59800	2010	115000
2008	73800	2010	103800
2009	53800	2010	97800
2009	53988	2011	115000
2009	106988	2011	112000
2009	88000	2011	118000
2009	98800	2011	93800
2009	97000	2011	101800
2009	84800	2011	99888
2009	78800	2011	108500
2009	88800	2011	103800
2009	95800	2012	89800
2009	79800	2012	87800

2009	83800	2012	93999
2009	83888	2012	125000
2009	109900	2012	109880
2009	63900	2012	111888
2009	78000	2012	108888
2009	146888		

Table2: Sample data collection of 45 used car price of VOLVO S60 from year 2002 to 2014

Year	Price	Year	Price
2002	19000	2011	85000
2002	22800	2011	86800
2002	21800	2011	159000
2002	18300	2011	115000
2002	22500	2011	95000
2002	24799	2011	107888
2003	19800	2011	97500
2003	17800	2011	105888
2003	23800	2011	88800
2004	28500	2011	89800
2004	28700	2012	105800
2004	29800	2012	91800
2004	29900	2012	99800
2005	32900	2012	92800
2005	23222	2012	98,800
2005	28800	2012	102800
2005	26800	2012	88500
2005	29799	2012	112000
2006	32800	2012	88858
2006	47500	2013	93800
2007	31800	2013	105800
2008	29800	2014	129800
2008	35800		

4. Results and Findings

Based on the data collected, three different regression graphs have been produced for both car models, which are linear regression, cubic regression and S-regression as shown in Figures 3 to 8.

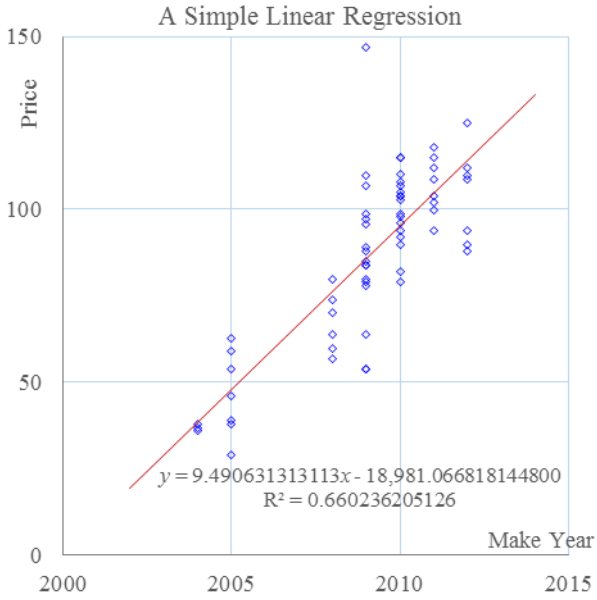

Fig.3: A linear regression on BMW 325i used car price

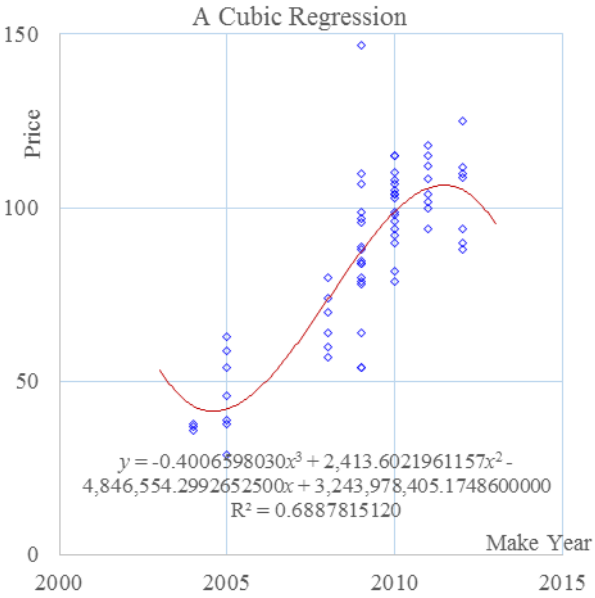

Fig.4: A cubic regression on BMW 325i used car price

Fig.5: An S-Regression on BMW 325i used car price

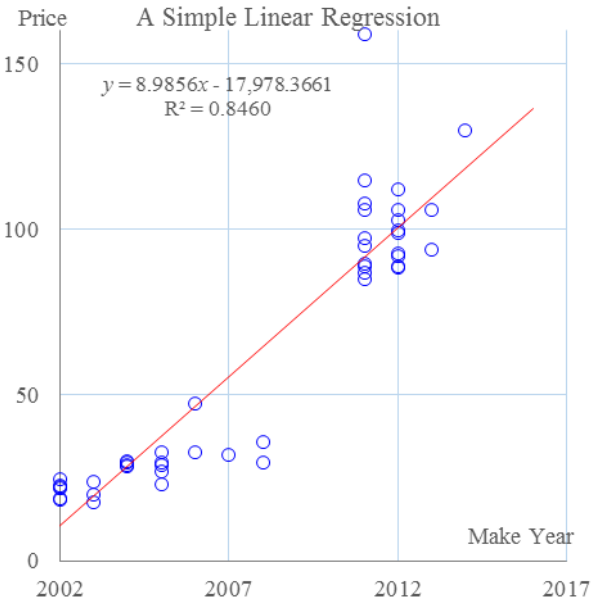

Fig.6: A linear regression on VOLVO S60 used car price

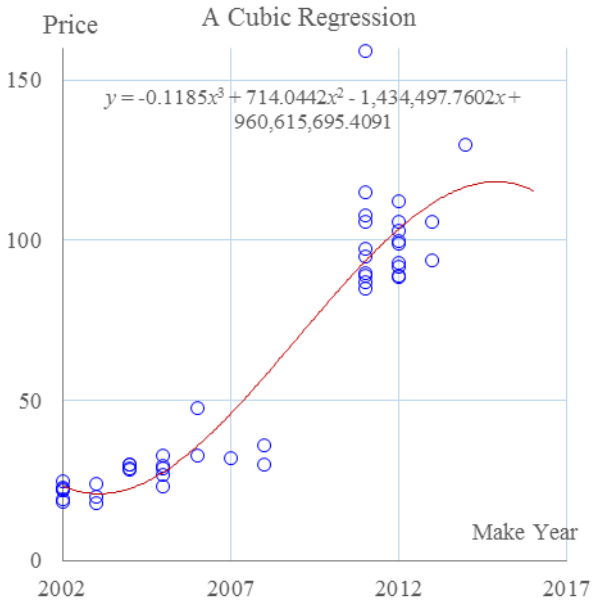

Fig.7: A cubic regression on VOLVO S60 used car price

Fig.8: An S-Regressieon on VOLVO S60 used car price

In this research, the linear and cubic regression models have been produced using Microsoft Excel, while S-Regression model has been produced using the S-shaped Membership Function [SMF] as in equation [2]. Using equation [2], the minimum and maximum values of the data have to be set in order for the equation to produce correct calculation. Based on the cubic model, the assumption of S-Regression model for BMW 325i would be as follows:

$$f(x)=\begin{cases}0, & x\le 2003\\ 2\left(\dfrac{x-2003}{10}\right)^2, & 2003<x\le 2008\\ 1-2\left(\dfrac{x-2013}{10}\right)^2, & 2008<x<2013\\ 1, & x\ge 2013\end{cases} \quad (3)$$

Thus, $\hat{y}=40000+70000\,f(x)$.

In this case, the minimum price for S-regression model is set to be RM40,000 while the maximum price is RM110,000.

The assumption for S-Regression model for VOLVO S60 would be as follows:

$$f(x)=\begin{cases}0, & x\le 2001\\ 2\left(\dfrac{x-2001}{7}\right)^2, & 2001<x\le 2008\\ 1-2\left(\dfrac{x-2015}{7}\right)^2, & 2008<x<2015\\ 1, & x\ge 2015\end{cases} \quad (4)$$

Thus, $\hat{y}=20000+90000\,f(x)$.

For VOLVO S60, the minimum price for S-regression model is set to be RM20,000 while the maximum price is RM110,000. All predicted values, the least square errors, and also the mean squared errors were calculated. The values of the mean squared errors for all the three types of regressions for both car models are shown in Table 3 below.

Table3: Mean Squared Error [MSE] for BMW 325i and VOLVO S60 used car price

Regression	Mean Squared Error [MSE]	
	BMW 325i	VOLVO S60
Linear	217.22	237.99
Cubic	198.97	188.93
S	201.99	204.84

From Table 3, the mean squared errors for S-Regression is found to be closer toward cubic regression quantitatively. The smaller the mean squared error, the closer we are to finding the line of best fit. Depending on the data, it may be inconceivable to get a very small value for the mean squared error due to the quantitative nature of the data.

This research is a preliminary study toward the developing a novel S-Regression model. A used car is expected to rise over the make years. Even though cubic regression produces smaller MSE, the curves on the far ends have already gone into negative derivative. These phenomena negate the principle of rising price over time. Even though the S-Regression takes a slightly higher MSE, it gives a closer price prediction in real practical life.

4.1. S-Regression Model on Used Car Price

It is very challenging and difficult task to make a quick decision with minimal knowledge or minimal information. Theoretically, a statistical model is typically very ideal but realistic. In this research we want to emulate the real scenario in regression model. The basic prediction model is a linear regression. However, in reality, a valuable item does not depreciate in linear value over time. Linear regression is not suitable and best practice for predictive analysis as it has sensitivity to outliers and subject to overfitting and also very rare cases when ordinary least square assumptions perfectly exists [16].

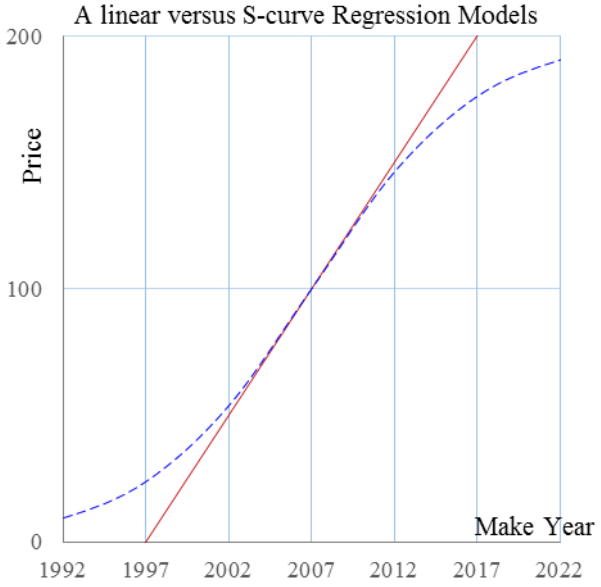

Fig. 9: A linear versus S-curve regression model

Figure 9 shows the relationship between the current prices of used car and the make year of the car. A straight line is a linear regression line, while the dotted lines represents an "S" shaped curve or S-Regression line. Assuming that the range of the car make year is between 1992 and 2022. Under the policy the car older than 12 years is not roadworthy. The life span of a car is between five to twelve years in most of the countries. It is clearly shown that if the best fit line is a straight line, then the price of used car in 1997 is 0. In reality, this is not possible since the car must have some steel value even if the car is already too old to be on the roads. The value of the car made in 1997 should be set [fixed] as a junk steel price. In 2017, the price of the second hand car logically cannot be too high as compared to 2012 because of the depreciation value. It has to be near to the price of 2012 used car. Therefore, the price of used car cannot follow the linear model, instead, it is better or practical to follow the S-shaped curve.

In general, a regression model will start from the center mean point $(\bar{x}, \bar{y})$ which is estimated by the sample mean point. At this center point, the regression model is expected to have the lowest error. The error grows larger as the point estimate moves away from the center point. A practical objective of S-Regression is to better fit the real world situation. In this case, the S-Regression model is meant to give better price prediction on the used cars in Malaysia, instead of linear regression which projects the price to go down at the same pace per year throughout a life span of a car.

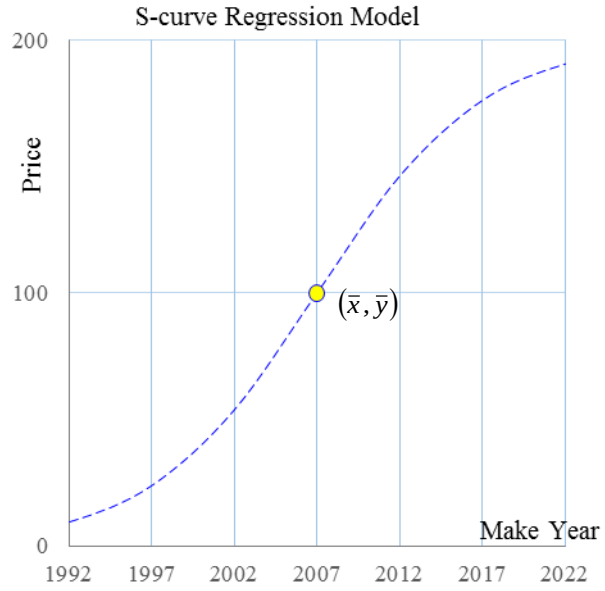

Fig.10: S-Regression model on used car price

Based on Figure 10, the strategy is to remove the points on the top right corner above the tradition linear regression line and the points on the bottom left corner below the tradition linear prior to the minimizing the prediction error during curve fitting. Under the current linear regression model the variance is wider as the year is far away from the center median year. The accuracy of the model is lower as the make year is moving away from the center. However, the value of an accurate prediction model is higher on both sides. There is a challenging problem to come out with a better regression model which give a more accurate prediction on both sides, further away from the center point.

5. Conclusion

The S-shaped curve pattern gives a significant meaning on what is happening in the real situation. S-shaped regression model has the potential to represent a better model in the world of prediction or forecasting the real scenario. The intention of this research is to give a fair price of used cars. Thus, S-regression model can be predicted to give a more realistic forecast value on the price of used cars in Malaysia.

Acknowledgement

The authors wish to thank Universiti Teknikal Malaysia Melaka, Malaysia for their encouragement and support in carrying out this research.

References

[1] Chatterjee S, Simonoff JS. Handbook of Regression Analysis. 1st ed. Wiley; 2012. 252 p.
[2] Pudaruth S. Predicting the Price of Used Cars using Machine Learning Techniques. Int J Inf Comput Technol. 2014;4[7]:753–764.
[3] Aimin W, Shunxi L. Prediction on the Developing Trend of Global Electric Automobile Based on the Logistic Model. In: 2011 International Conference on Business Management and Electronic Information [BMEI]. IEEE; 2011. p. 31–33.
[4] Sharma R, K. Sinha A. Sales Forecast of an Automobile Industry. Int J Comput Appl. 2012;53[12]:25–28.
[5] Aydin G. Forecasting Natural Gas Production Using Various Regression Models. Pet Sci Technol. 2015;33[15–16]:1486–1492.
[6] Faiedh H, Gafsi Z, Besbes K, Torki K. Digital hardware implementation of sigmoid function and its derivative for artificial neural networks. Proc Int Conf Microelectron ICM. 2001;2001–Janua:189–192.
[7] Jamel TM, Khammas BM. Implementation of a Sigmoid Activation Function For Neural Network Using FPGA. 13th Sci Conf Al-Ma'moon Univ Coll. 2012;[April].
[8] Tsai CH, Chih YT, Wong WH, Lee CY. A Hardware-Efficient Sigmoid Function with Adjustable Precision for a Neural Network System. IEEE Trans Circuits Syst II Express Briefs. 2015;62[11]:1073–1077.
[9] Han J, Moraga C. The influence of the sigmoid function parameters on the speed of backpropagation learning. In: Mira J, Sandoval F [eds] From Natural to Artificial Neural Computation IWANN 1995 Lecture Notes in Computer Science. Springer, Berlin, Heidelberg; 1995. p. 195–201.
[10] Mahalingam PR, Vivek S. Predicting Financial Savings Decisions Using Sigmoid Function and Information Gain Ratio. Procedia Comput Sci 93. 2016;93[September]:19–25.
[11] Weisstein EW. "Sigmoid Function." From MathWorld--A Wolfram Web Resource. [Internet]. [cited 2017 Jan 15]. Available from: http://mathworld.wolfram.com/SigmoidFunction.html
[12] Meade N, Islam T. Forecasting the diffusion of innovations: Implications for time-series extrapolation. Princ Forecast. 2002;[Exhibit I]:577–595.
[13] Kucharavy D, De Guio R. Application of S-shaped curves. In: Procedia Engineering. 2011. p. 559–572.
[14] Armstrong JS. Principles of Forecasting: A Handbook for Researchers and Practitioners. Kluwer Academic Publishers; 2001.
[15] Abu NA, Salim F, Kamel Ariffin MR. Introducing S-Index into Factoring RSA Modulus via Lucas Sequences An Overview on S - index Function. Malaysian J Math Sci [Internet]. 2017;11[Special Issue]:72–89. Available from: http://einspem.upm.edu.my/journal/
[16] Mishra S. 10 Types of regression techniques and their application [Internet]. 2015 [cited 2016 Dec 20]. Available from: http://hashanalytics.co.in/10-types-of-regression-techniques-and-their-application/