Journal of Mountain Area Research
Vol. 2, 2017
http://journal.kiu.edu.pk/index.php/JMAR

Full length article

AN EFFICIENT AND COST-EFFECTIVE MATHEMATICAL MODEL TO ANALYZE BIG DATA

Ubaidullah, W. Akram, I. A. Memon*
Sukkur Institute of Business Administration, Sindh Pakistan.

ABSTRACT

An efficient and cost-effective piecewise mathematical model is presented to represent a descriptive huge data mathematically. The techniques of function lines as decision boundaries are applied to incorporate the big data of the organization into slope intercept form. Which may be very helpful for a better understanding of discrete data to obtain sustainable and accurate results. Based on the boundaries limitation results of the collected data of the Federal Board of Revenue, the income tax against the income is studied. And finally the reliability of piecewise function to optimize the role of strategic management in any organization is investigated. The results showed that, the slope rate measured in the boundaries of income in percentage or increased slope rate is in good agreement with that predicted by the organization in descriptive form.

KEYWORDS: Big data; Mathematical Model; Boundaries; Tools; Computational technique

* Corresponding author: (E-mail: ubaidullah@iba-suk.edu.pk)

1. INTRODUCTION

The software engineering and computer science characterized big data as the large data set that become hard to do work. Due to the size and complexity of big data it is difficult to obtain the required results using on-hand database management devices or traditional data processing techniques. Scientist, business executives and technocrats hypothesize that the phenomena of big data is difficult to explain due to the unequal growth rate and huge volume. The term big is not big in volume if we go back ten years ago a hard derive of 30 MB was big but nowadays a 2TB derives are common. One historically important distinction between the collected data from dawn of civilization to 2003 is 5 exabytes, now we are creating 5 exabytes every two days. Today we are not only creating the size that is volume of data but we are also creating variety of data with a faster rate, this is called three Vs of Big Data, Volume, Variety and Velocity. In the era 1440s and by 1500s, the printed data were 10

million texts including 2 million books, while in the 6th and 7th centuries only 120 books were produced annually in Western Europe. The International Data Corporation stated that discovery and analysis of data economically with high velocity from very huge volumes of data by using new technologies and architecture designs is big data.

Viktor Mayer-Schönberger and Kenneth Cukier described that big data is a way to take out new insights or make new forms of value in traditions that change markets, organizations etc.

To collect, store and analyze the data sets from unstructured data we have need advance techniques, software's and systems. In advance laboratories and advance engineering research centers big data provided significant advances. Whether we are in technology or business the term big data is absolutely different than the past things. Martin Hilbert and Priscilla Lopez stated that in 1996 digital data was one percent and by 2007 almost ninety four percent data was digital. Around 2000 each things became digital while in the 20th century the digital data was only texts and numbers.

Andrew Whit stated that big data has the potential to open up stirring new opportunities in social research, it while it is difficult to access. Hall [1] stated that curiosity, litheness and motivation to learn by doing assortment and job experience. Lorenz [2] described that data is like an assembly of facts, but it is not necessarily the facts always truth. The weakly interpretation causes the wrong conclusions so we have needed to understand the big data.

Jeanne Harris described that the importance to understand the mathematical reasoning and statistical models is not only need for technical experts but it is also need for managers to meet the challenges of big data. Harris [4] stated that sixty percent of respondent on a survey feel the need to develop new skills for their employees to translate big data into insight and business value. To reduce large raw data sets into small dimensions the topology technique is a flexible technique for different systems. Big Data has the potential not only to update research, but it has the potential also transform education [8]. Hamann [9] described the technique of data discretization for resembling the curve by line segments. Discretization technique is a first step making the data suitable for numerical assessment and execution on digital computers. McMaster in 1987 provided a scheme of data reduction of piecewise linear curves.

Gar in 2011 stated that the industry analysis companies are not facing challenges only in volume but also in velocity and Variety. Big data is a data which is recorded from a data generating source. One challenge is the collection of required data from these sources without losing the exact required information. Another challenge is to automatically collect the right data from the data source. We have need also an information extraction method to take out the associated data from the data source and express it in an ordered form for analysis. Data analysis is also a challenge, so to overcome this challenge we need domain analysis scientist to create effective data base design. The piecewise defined function is a well-defined mathematical technique to formulate and interpret the big data. The Federal Board of Revenue is a supreme federal agency of Pakistan for auditing, enforcing and collecting revenue for the government of Pakistan. The data of collection federal taxes is a big data. In our paper we have used the piecewise function to formulate and interpret the collected data. We formulated the income tax slabs for salaried class in Pakistan for Financial Year 2014-15 into

piecewise defined form that is limitations. Our developed mathematical model is a cost effective and time efficient.

2. MATHEMATICAL AND GRAPHICAL REPRESENTATION OF DATA

2.1 Mathematical model

The general mathematical form for n dimensional piecewise continuous and convex linear functions is $f : R^n \to R$.

Like $\Pi \in P(P(R^{n+1}))$

So that: $f\left(\vec{x}\right) = \min_{\sum \in \Pi} \max_{(\vec{a},b)\in \sum} \vec{a}.\vec{x}+b$

If the function is convex and continuous then,

$$\sum \in P(R^{n+1})$$

So that: $$f\left(\vec{x}\right) = \max_{(\vec{a},b)\in \sum} \vec{a}.\vec{x}+b$$

Here $\vec{a}.\vec{x}+b$ is a linear polynomial such that $a \neq 0 \ \ and \ \ a, b \in R$.

The piecewise linear function effectively reduced the problem size and enhanced the computational efficiency.

2.2 Data Collection and Processing

According to the Finance Act passed by the government of Pakistan, these below mentioned income tax rates will be followed for salaries in the year 2014-2015. Suppose x represents the income and $T(x)$ represents the income tax. The tax slabs are as follows:

S#	Taxable Income	Rate of Tax
1	Where the taxable income does not exceed Rs.400,000	0%
2	Where the taxable income exceed Rs.400,000 but does not exceed Rs.750,000	5% of the amount exceeding Rs.400,000
3	Where the taxable income exceed Rs.750,000 but does not exceed Rs.1,400,000	Rs.17,500+10% of the amount exceeding Rs.750,000
4	Where the taxable income exceed Rs.1,400,000 but does not exceed Rs.1,500,000	Rs.82,500 +12.5% of the amount exceeding Rs.1,400,000
5	Where the taxable income exceed Rs.1,500,000 but does not exceed Rs.1,800,000	Rs.95,000+15% of the amount exceeding Rs.1,500,000
7	Where the taxable income exceed Rs.1,800,000 but does not exceed Rs.2,500,000	Rs.140,000+17.5% of the amount exceeding Rs.1,800,000
8	Where the taxable income exceed Rs.2,500,000 but does not exceed Rs.3,000,000	Rs.262,000+20% of the amount exceeding Rs.2,500,000

9	Where the taxable income exceed Rs.3,000,000 but does not exceed Rs.3,500,000	Rs.362,500+22.5% of the amount exceeding Rs.2,500,000
10	Where the taxable income exceed Rs.3,500,000 but does not exceed Rs.4,000,000	Rs.475,000+25% of the amount exceeding Rs.3,500,000
11	Where the taxable income exceed Rs.4,000,000 but does not exceed Rs.7,000,000	Rs.600,000+27.5% of the amount exceeding Rs.4,000,000
12	Where the taxable income exceed Rs.7,000,000	Rs.1,425,000+30% of the amount exceeding Rs.7,000,000

- The rate of income tax is zero 0% if the taxable salary income does not exceed Rs. 400,000 i.e. $T(x) = 0$.
- The rate of income tax is 5% if the taxable salary income exceed Rs. 400,000 but does not exceed Rs 750,000 i.e.

$$T(x) = 0.05(x - 4000{,}000)$$
$$T(x) = 0.05x - 20000$$

.

- The rate of income tax is 10% if the taxable salary income exceed Rs. 750,000 but does not exceed Rs. 1,400,000

$$T(x) = 0.05(750{,}000) - 20000$$

- i.e. $T(x) = 17500 + 0.10(x - 750{,}000)$
$$T(x) = 0.10x - 57500$$

- The rate of income tax is 12.5% if the taxable salary income exceed Rs. 1,400,000 but does not exceed Rs. 1,500,000
- i.e. $T(x) = 82500 + 0.125(x - 1{,}400{,}000)$
$$T(x) = 0.125x - 92500$$

- The rate of income tax is 15% if the taxable salary income exceed Rs. 1,500,000 but does not exceed Rs. 1,800,000
- i.e. $T(x) = 95000 + 0.15(x - 1{,}500{,}000)$
$$T(x) = 0.15x - 130000$$

- The rate of income tax is 17.5% if the taxable salary income exceed Rs. 1,800,000 but does not exceed Rs. 2,500,000
-
- i.e. $T(x) = 140000 + 0.175(x - 1{,}800{,}000)$
$$T(x) = 0.175x - 175{,}000$$

- The rate of income tax is 20% if the taxable salary income exceed Rs. 2,500,000 but does not exceed Rs. 3,000,000
- i.e. $T(x) = 262500 + 0.2(x - 2{,}500{,}000)$
$$T(x) = 0.2x - 237{,}500$$

- The rate of income tax is 22.5% if the taxable salary income exceed Rs. 3,000,000 but does not exceed Rs. 3,500,000
- i.e. $T(x) = 362500 + 0.225(x - 3{,}000{,}000)$
$$T(x) = 0.225x - 312{,}500$$

- The rate of income tax is 25% if the taxable salary income exceed Rs. 3,500,000 but does not exceed Rs. 4,000,000
- i.e. $T(x) = 475000 + 0.25(x - 3{,}500{,}000)$
$$T(x) = 0.25x - 400{,}000$$

- The rate of income tax is 27.5% if the taxable salary income exceed Rs. 4,000,000 but does not exceed Rs. 7,000,000

- i.e. $T(x) = 600{,}000 + 0.275(x - 4{,}000{,}000)$
  $T(x) = 0.275x - 500{,}000$
- The rate of income tax is 30% if the taxable salary income exceed Rs. 7,000,000
- i.e. $T(x) = 1{,}425{,}000 + 0.3(x - 7{,}000{,}000)$
  $T(x) = 0.3x - 675{,}000$

$$T(x) = \begin{cases} 0 & 0 \le x \le 400{,}000 \\ 0.05x - 20{,}000 & 400{,}000 < x \le 750{,}000 \\ 0.1x - 57{,}500 & 750{,}000 < x \le 1{,}400{,}000 \\ 0.125x - 92{,}500 & 1{,}400{,}000 < x \le 1{,}500{,}000 \\ 0.15x - 130{,}000 & 1{,}500{,}000 < x \le 1{,}800{,}000 \\ 0.175x - 175{,}000 & 1{,}800{,}000 < x \le 2{,}500{,}000 \\ 0.2x - 237{,}500 & 2{,}500{,}000 < x \le 3{,}000{,}000 \\ 0.225x - 312{,}500 & 3{,}000{,}000 < x \le 3{,}500{,}000 \\ 0.25x - 400{,}000 & 3{,}500{,}000 < x \le 4{,}000{,}000 \\ 0.275x - 500{,}000 & 4{,}000{,}000 < x \le 7{,}000{,}000 \\ 0.3x - 675{,}000 & x > 7{,}000{,}000 \end{cases}$$

2.3 Graphical Representation

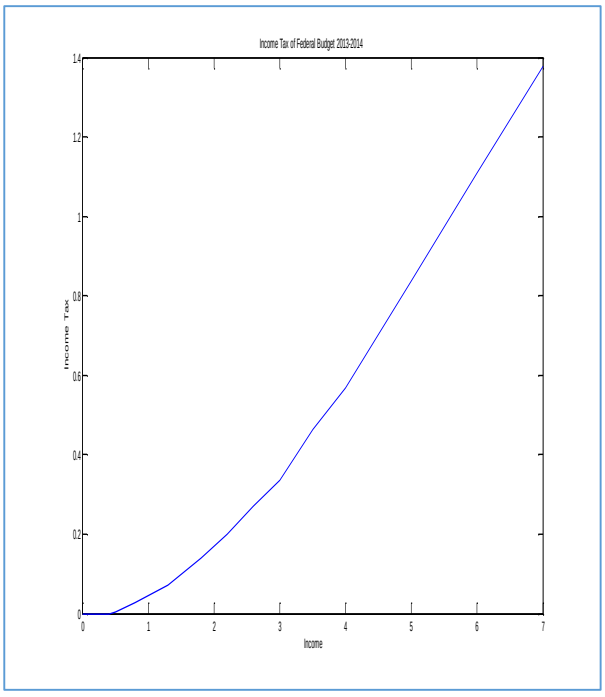

Figure 1. Represents the Income in million on x-axis and Income tax in million on y-axis in standard form.

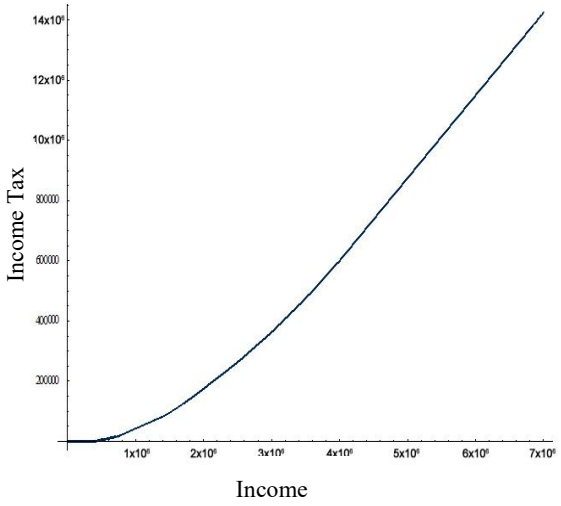

Figure 2. Represents the Income in million on x-axis and Income tax in million on y-axis in scientific notation.

3. RESULTS AND DISCUSSION

From the derived mathematical model and the graphical representation we can conclude that the data we have collected of the fiscal year 2014-2015 of federal budget of Pakistan is big in term of volume. To manage, share, analyze and visualize the data in a timeframe it is difficult without advanced tools, software, and systems. The used mathematical model summarized the big data into a small form such that we can calculate with a faster rate easily and efficiently. The income tax depends on the income so we have taken income on x-axis and income tax on y-axis. Scaling on the axis is as; on x-axis income is in millions and on y-axis income tax is in hundred thousand. The graph shows as income increase the income tax is also increase. Figure 1and Figure 2 indicate the increase of income tax due to increase of income. If the income of a pair is 750,000 the income tax is 17500. This showed the reliability of the piecewise linear mathematical model. From the above Model of T(x), the slope of the intervals are as:

0, 0.05, 0.10, 0.125, 0.15, 0.175, 0.2, 0.225, 0.25, 0.27 and 0.3. These are the mathematical indicators which are efficient and cost effective to analyze and interpret the big data into small one. These indicator indicates that as income increase the income tax is also increase.

4. CONCLUSION

We presented a piecewise mathematical model which converts a descriptive data into a single model based on the linear coefficients, assigned variables and tax slab's percentage into a single model. We used the high level language software 'MATLAB' that is able to reliably detect and sharply the tax slab of the tax payer. This software also accurately calculate the exact amount of the individual taxpayer. The problem here is to find the slab percentages that appears in the acquired data. Finally we did optimize our collected data.

References

[1] D. Lazer, A. Pentland, L. Adamic, S. Aral, A-L. Barabási, D. Brewer. Computational Social Science". *Science*: 323 (2009), 721-723.

[2] S. Shvetank, H. Andrew, C. Jaime. "Good Data Won't Guarantee Good Decisions". Harvard Business Review, HBR.org. Retrieved (2012).

[3] V. Mayer- Schönberger & K. Cukier. Big Data: A Revolution that Will Transform How We Live, Work, and Think", (2013) New York, Houghton Mifflin Harcourt Publishing Company.

[4] J. Harris. Data is useless without the skills to analyze it, (2012) HBR Blog Network.

[5] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, & A. H. Byers (2011).

[6] "Big data: The next frontier for innovation, competition, and productivity". McKinsey Global Institute.

[7] D. Raywood. Big data analyst shortage is a challenge for the UK. SC Magazine, (2012).

[8] CCC, Advancing Personalized Education. Computing Community Consortium". Spring 2011.

[9] B. Hamann, J. L. Chen. "Data point selection for piecewise linear curve approximation". *Computer Aided Geometric Design* 11 (1994).

[10] M. H. Lin, J. G. Carlsson, D. Ge, J. Shi and J. F. Tsai, A Review of Piecewise Linearization Method. Mathematical Problems in Engineering (2013).

[11] K. Holmberg. Solving the Staircase Cost Facility Location Problem with Decomposition and Piecewise Linearization. European Journal of Operational Research, 75(1994) 41-61.

[12] A. B. Keha, I. R. De Farias, and G. L. Nemhauser, Models for Representing Piecewise Linear Cost Function". Operation Rsearch Letters, 32 (2004) 44-48.

[13] V. Ford and A. Siraj, Clustering of Smart Meter Data for Disaggregation, In Proc. IEEE Global Conference on Signal and Information Processing (Global SIP),Austin, TX (2013).

[14] www.fbr.gov.pk

[15] W. Huang, P. Eades, S. H. Hong, C. C. Lin. Improving multiple aesthetics produces better graph drawings. J Vis Lang Comput 24 (2013) 262-272.

[16] M. J. Baker, S. G. Eick. Space-filling Software Visualization. Journal of Visual Languages & Computing 6(1995)119-133.