Visualization of Performance of Binary Search in Worst Case in Personal Computer Using Scatter Plot, Interpolation Lines and Loess Curve

Dipankar Das

Assistant Professor, The Heritage Academy, Kolkata, West Bengal, India
dipankr.das@gmail.com

ABSTRACT

From ancient time people have a great deal of inquisitiveness in different aspects of life. In every point of time we generate a great amount of data but crystallizing these data first into information, then from information to knowledge and ultimately into wisdom are undoubtedly one of the biggest challenges in front of us. One way to extract information hidden within the data is to identify the pattern of the data. The technique of data visualization helps us to easily identify the patterns hidden within the dataset. In the present study, the researchers have examined the performance (Data size versus Average execution time) of Binary search in the worst case by experimental means and the main objectives are to visualize the performance using scatter plot, interpolation lines and Loess curves. The scatter plot of the performance appears to be approximately random and the interpolation lines appear to be going zig-zag way. An approximately smooth Loess curve is fitted to the data points using Gaussian family with span is equal to 0.6, degree is equal to 2 and residual standard error is equal to 29.04.

Keywords—Binary search, scatter plot, interpolation lines, Loess curve

1. INTRODUCTION

Data visualization is the method of combining a number of things in a single place and presenting them more effectively or as a coherent whole in the graphical or pictorial format which in turn makes it easier for people to analyze the data, identify the hidden pattern of the data and conceptualize new ideas from the data. There are different ways of data visualization. In this present study, the researchers have employed scatter plot, interpolation lines and Loess curve to visualize the performance of Binary Search algorithm in the worst case. A scatter plot is a visualization of the relationship between two variables [1]. The interpolation line is another method of data visualization which connects the data points and can be applied to line charts, area charts, scatter plots etc [2]. The Loess curve fitting method is a nonparametric method (because in this method, the linearity assumptions of conventional regression methods have been relaxed) and is used for fitting a smooth curve between two variables [3]. We know that the Binary Search algorithm works on divide and conquer paradigm [4]. In the worst case scenario, the key will not be present in the array [5] and it will take O (log n) comparisons [6]. In the present study the performance of Binary Search has been examined by experimental means.

2. RELATED WORK

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

Literature review has revealed that different researchers had examined Binary search algorithm in their studies. Comparative analysis between Linear search and Binary search algorithm had been carried out by some researchers [7][8][9][10][11]. Three different types of searching

algorithms namely Linear search, Binary search and Interpolation search had also been compared and analyzed [12]. Linear search, Binary search and Hashing had been compared on the basis of different parameters e.g. number of comparisons, time complexity, space complexity etc. [13]. A randomized searching algorithm had been proposed and its performance had been analyzed with Binary search and Linear search [14]. The Binary search algorithm had been executed on two different computers and the performances were compared by employing curve fitting technique [15].

Visualization of the performance of Binary search had been done by using (i) polynomial of degree 22 curve identified by Bayesian information criterion and polynomial of degree 24 curve identified through Akaike information criterion [16], (ii) cubic curves identified by the use of goodness of fit statistics like R square, Adjusted R square, RMSE and residual analysis [17], (iii) spline interpolation curve fitting technique [18] and (iv) semi parametric technique with varying degrees (degree = 3, 5 and 7) [19].

3. OBJECTIVES OF THE STUDY

To visualize the performance (Data size versus Execution time) of Binary Search in worst case in a personal computer (laptop) using (i) scatter plot, (ii) interpolation lines and (iii) Loess curve

4. RESEARCH METHODOLOGY

Algorithm: Binary Search

Case: Worst case

Data size: 2500 to 20000 with an interval of 500 (total 36 data sizes)

Number of observations: 1000 execution times for each data size (data size 2500 to 20000 with an interval of 500)

Execution time (considered for the study): Average execution time for each data size

Unit of execution time: nano seconds

Implementation platform: Linux Operating System (Ubuntu 12.04.4 LTS), OpenJDK runtime environment (java version 1.6.0_36), Intel(R) Core(TM) 2 Duo CPU (T5870 @2.00 GHz) and 3 GB RAM.

Data analysis platform: R

Visualization: (a) Scatter plot, (b) Interpolation lines and (c) Loess curve

Span selection of loess curve: The objective is to find the smallest span that will give a smooth curve. This is achieved by using visual trial and error method. First, a model will be fitted and the curve obtained will be visually inspected. Based on the visual inspection, the span will be changed until most of the roughness of the curve is removed. Usually, the range of the span is from 0.3 to 0.8 [20].

Degree of Loess curve: 2

Family of Loess curve: Gaussian

Phases of the study: (i) writing the Binary Search algorithm in java programming language, (ii) executing the java code in the personal computer (laptop) under study for data size 2500 to 20000 with an interval of 500 in the worst case scenario. For each data size 1000 observations (execution time in nano seconds) are recorded, (iii) finding the average execution time for each data size, (iv) generating the Loess curves using the span from 0.3 to 0.8 with an interval of 0.05, (v) inspecting the curves visually, identifying the best smooth curve within this range which is having lowest value of span and (vi) performing the residual diagnostics of this selected model. The following residual diagnostics plots [20] will be used for the purpose: (a) residuals versus predictor, (b) quantile comparison plots of the residuals and (c) absolute residuals versus predictors, (vii) generating the scatter plot diagram (Data size versus Average execution time), (viii) visualizing the average execution time with the help of interpolation lines and (ix) visualizing the performance (Data size versus Execution time) of Binary Search in worst case in a personal computer (laptop) using selected Loess curve.

5. DATA ANALYSIS AND FINDINGS

The Loess curve of the performance of the Binary Search in worst case with span equal to 0.3 is shown below (Fig-1).

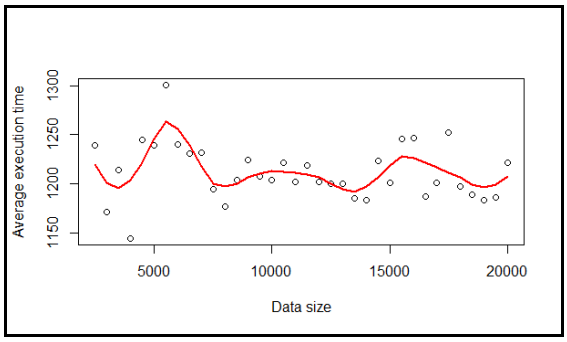

Fig-1: Loess curve with span = 0.3

The Loess curve of the performance of the Binary Search in worst case with span equal to 0.35 is shown below (Fig-2).

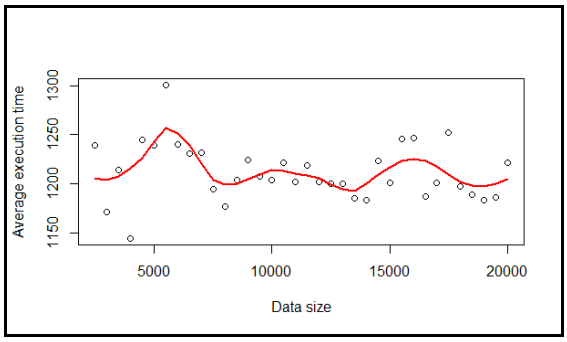

Fig-2: Loess curve with span = 0.35

The Loess curve of the performance of the Binary Search in worst case with span equal to 0.4 is shown below (Fig-3).

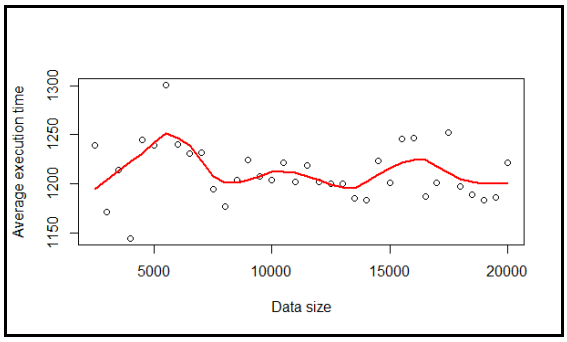

Fig-3: Loess curve with span = 0.4

The Loess curve of the performance of the Binary Search in worst case with span equal to 0.45 is shown below (Fig-4).

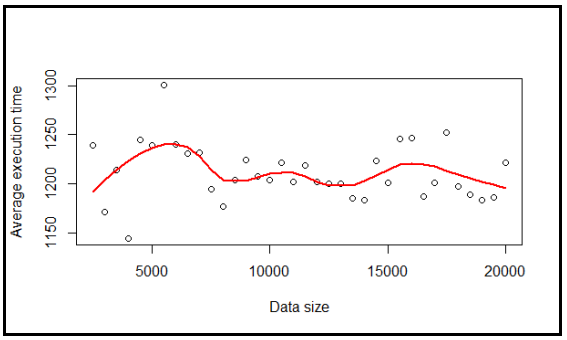

Fig-4: Loess curve with span = 0.45

The Loess curve of the performance of the Binary Search in worst case with span equal to 0.5 is shown below (Fig-5).

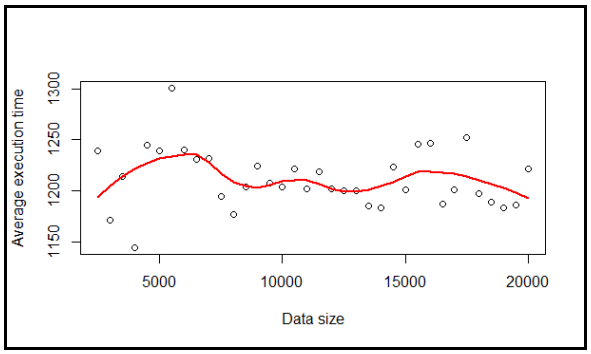

Fig-5: Loess curve with span = 0.5

The Loess curve of the performance of the Binary Search in worst case with span equal to 0.55 is shown below (Fig-6).

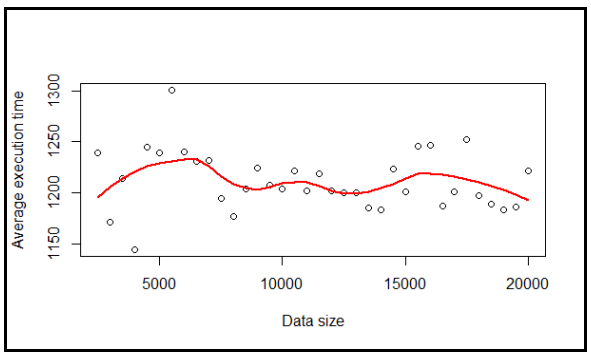

Fig-6: Loess curve with span = 0.55

The Loess curve of the performance of the Binary Search in worst case with span equal to 0.6 is shown below (Fig-7).

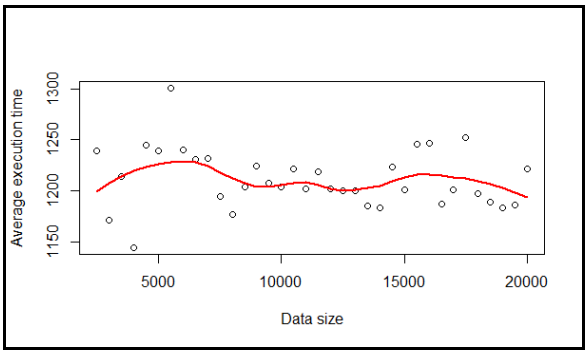

Fig-7: Loess curve with span = 0.6

The Loess curve of the performance of the Binary Search in worst case with span equal to 0.65 is shown below (Fig-8).

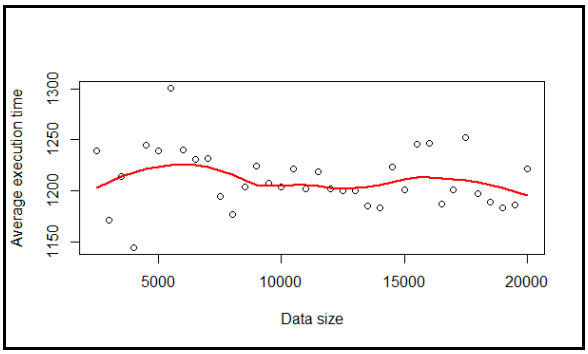

Fig-8: Loess curve with span = 0.65

The Loess curve of the performance of the Binary Search in worst case with span equal to 0.7 is shown below (Fig-9).

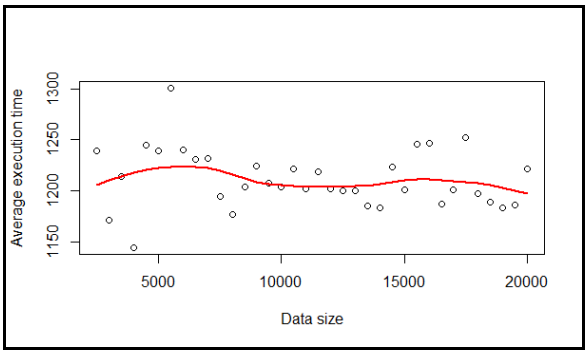

Fig-9: Loess curve with span = 0.7

The Loess curve of the performance of the Binary Search in worst case with span equal to 0.75 is shown below (Fig-10).

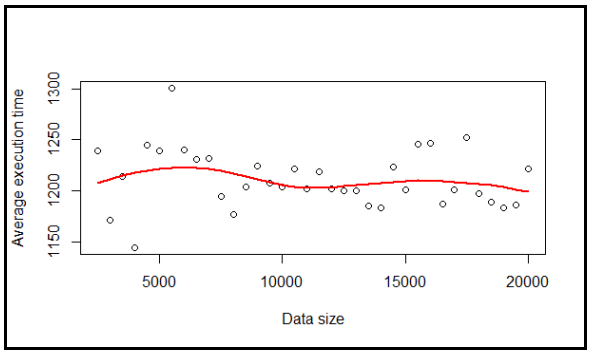

Fig-10: Loess curve with span = 0.75

The Loess curve of the performance of the Binary Search in worst case with span equal to 0.8 is shown below (Fig-11).

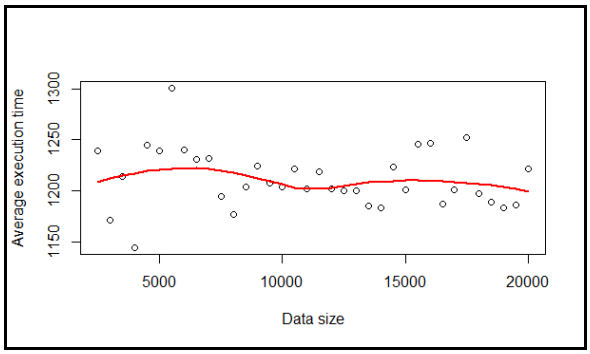

Fig-11: Loess curve with span = 0.8

It has been observed from the above Loess curves (Fig-1 to Fig-11) that for the span value equal to 0.6 (lowest value of span) the curve has became approximately smooth i.e. most of the roughness of the curve has been removed. Therefore, this span (0.6) has been selected for plotting the Loess curve for visualization purpose.

The residual diagnostics of the selected Loess model are given below:

The residual versus predictor plot of the selected model i.e. the Loess curve with span equal to 0.6 is shown in the following figure (Fig-12).

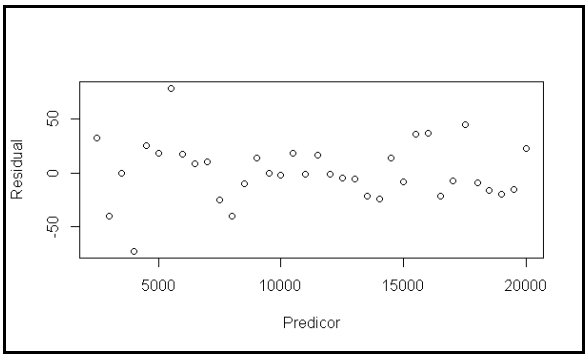

Fig-12: Residual versus Predictor plot of the selected model

Form the above figure (Fig-12) it is evident that the Residual versus Predictor plot does not show any particular pattern and therefore it may be concluded that the model fits the data well.

The Quantile comparison plot of the residuals is shown in the figure (Fig-13) below:

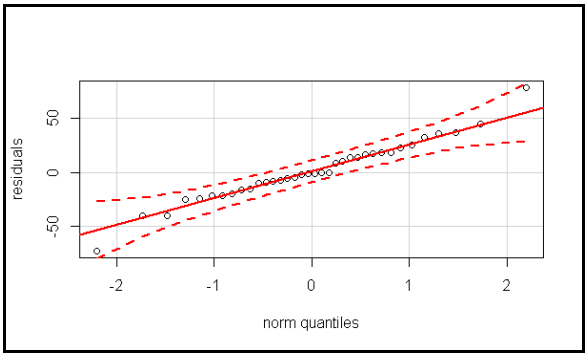

Fig-13: Quantile comparison plot of the residuals

Form the above figure (Fig-13) we observe that the errors are approximately normally distributed.

6. CONCLUSION

The main objectives of this study are to visualize the performance (Data size versus Average execution time) of the Binary search algorithm in the worst case implemented using java on Linux platform in a personal computer (laptop) with (i) scatter plot, (ii) interpolation lines and (iii) Loess curve.

The visualization of the performance (Data size versus Average execution time) of the Binary Search in worst case using scatter plot is shown in the figure (Fig-14) given below. Here, Data size has been considered as x axis and Average execution time as y axis.

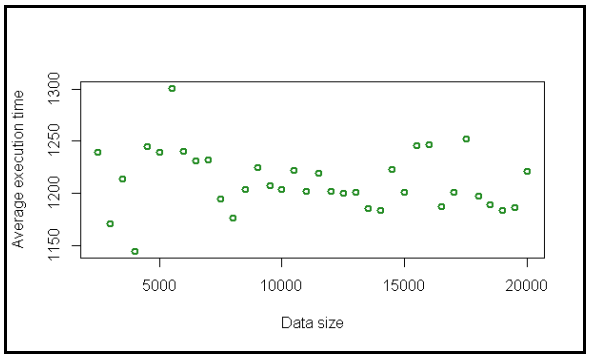

Fig-14: Scatter plot of Data size versus Average execution time

The above figure (Fig-14) shows that the data points (Data size versus Average execution time) are appearing approximately random.

The visualization of the performance (Data size versus Average execution time) of the Binary Search in worst case using interpolation lines is shown in the figure (Fig-15) given below. In this case also, Data size has been considered as x axis and Average execution time as y axis.

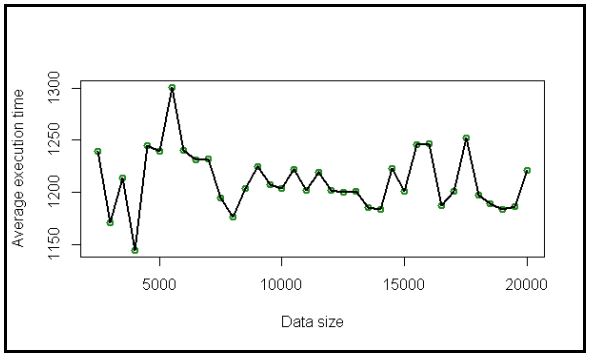

Fig-15: Interpolation line plot of Data size versus Average execution time

From the above figure (Fig-15) we observe that the performance of the Binary search in the worst case in the personal computer under study is showing a zig-zag pattern.

The visualization of the performance (Data size versus Average execution time) of the Binary Search in worst case using Loess curve (span = 0.6) is shown in the following figure (Fig-16):

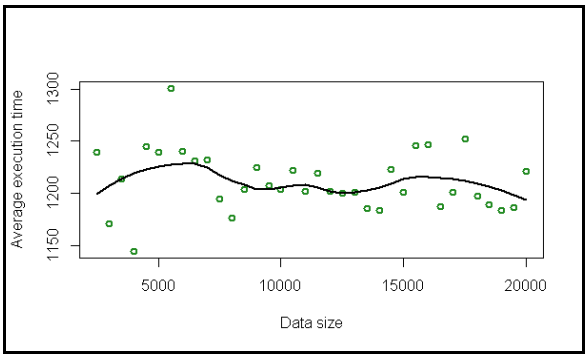

Fig-16: Loess curve of the performance of the Binary search with span = 0.6

The summary of the selected Loess curve shown above (Fig-16) is given below:

Number of Observations: 36, Equivalent Number of Parameters: 5.48, Residual Standard Error: 29.04 and Trace of smoother matrix: 6.03 (exact)

The Loess curve given above (Fig-16) provides us an easy and flexible way to represent the performance (Data size versus Average execution time) of the Binary search algorithm in the worst case implemented using java on Linux platform in a personal computer (laptop) with a smooth curve i.e. most of the roughness of the curve is removed but at the same time it is to be noted that it requires some deductive reasoning to obtain the smooth Loess curve with the lowest span and also it does not provide any mathematical equation of the curve obtained.

REFERENCES

[1] Scatter Diagram. (n.d.). Retrieved October 14, 2016, from http://mathworld.wolfram.com/ScatterDiagram.html

[2] IBM Knowledge Center. (n.d.). Retrieved October 14, 2016, from https://www.ibm.com/support/knowledgecenter/SSLVMB_20.0.0/com.ibm.spss.statistics.help/idh_webhelp_interpoltn_palette.htm

[3] LOESS Curve Fitting (Local Polynomial Regression). (n.d.). Retrieved October 14, 2016, from http://www.statsdirect.com/help/content/nonparametric_methods/loess.htm

[4] Muhammad, R. B. (n.d.). Divide-and-Conquer Algorithms. Retrieved October 14, 2016, from http://www.personal.kent.edu/~rmuhamma/Algorithms/MyAlgorithms/divide.htm

[5] *CS127 Complexity Analysis* [PDF]. (2003, September 19). Retrieved October 14, 2016, from http://cs.carleton.edu/faculty/adalal/teaching/fall03/cs127/notes/sept19.pdf

[6] Analysis of Binary Search. (n.d.). Retrieved October 14, 2016, from http://www2.hawaii.edu/~janst/demos/s97/yongsi/analysis.html

[7] Kumari, A., Tripathi, R., Pal, M., & Chakraborty, S. (2012). Linear Search versus Binary Search: A Statistical Comparison for Binomial Inputs. *International Journal of Computer Science, Engineering and Applications* (IJCSEA), 2(2).

[8] Parmar, V. P., & Kumbharana, C. (2015). Comparing Linear Search and Binary Search Algorithms to Search an Element from a Linear List Implemented through Static Array, Dynamic Array and Linked List. *International Journal of Computer Applications*, 121(3).

[9] Pathak, A. (2015). Analysis and Comparative Study of Searching Techniques. *International Journal of Engineering Sciences & Research Technology*, 4(3), 235-237.

[10] Chandrawat, Y. S., Vajpayee, A., & Pathak, A. (2015). Analysis and Comparative Study of Searching Techniques. *International Journal of Engineering Sciences & Research Technology*, *4*(10), 331-333.

[11] Sapinder, R., Singh, A., & Singh, H. L. (2012). Analysis of Linear and Binary Search Algorithms. *International Journal of Computers & Distributed Systems*, *1*(1).

[12] Roy, D., & Kundu, A. (2014). A Comparative Analysis of Three Different Types of Searching Algorithms in Data Structure. *International Journal of Advanced Research in Computer and Communication Engineering*, *3*(5), 6626-6630.

[13] Pandey, K. K., & Pradhan, N. (2014). A Comparison and Selection on Basic Type of Searching Algorithm in Data Structure. *International Journal of Computer Science and Mobile Computing*, *3*(7), 751-758.

[14] Das, P., & Khilar, P. M., Prof. (2013). A Randomized Searching Algorithm and its Performance analysis with Binary Search and Linear Search Algorithms. *The International Journal of Computer Science & Applications*, *1*(11), 11-18.

[15] Das, D., Kole, A., Mukhopadhyay, S., & Chakrabarti, P. (2015). Empirical Analysis of Binary Search Worst Case on Two Personal Computers Using Curve Estimation Technique. *International Journal of Engineering and Management Research*, 5(5), 304 –311.

[16] Das, D. (2016). Polynomial Curve Fitting of Execution Time of Binary Search in Worst Case in Personal Computer. *International Journal of Recent Trends in Engineering & Research*, 2(6), 245 – 250.

[17] Das, D., Kole, A., & Chakrabarti, P. (2015). Sample Based Visualization and Analysis of Binary Search in Worst Case Using Two-Step Clustering and Curve Estimation Techniques on Personal Computer. *International Research Journal of Engineering and Technology*, 2(8), 1508 – 1516.

[18] Das, D. (2016). Visualization of Binary Search in Worst Case Using Spline Interpolation Curve Fitting in Personal Computer. *International Journal of Recent Trends in Engineering & Research*, 2(7), 38 – 42.

[19] Das, D. (2016). Visualization of Performance of Binary Search in Worst Case in Personal Computer using Semi Parametric Technique. *International Journal for Scientific Research & Development*, 4(5), 1236 – 1239.

[20] Jacoby, B. (n.d.). *Lecture 15: Loess – Regression III: Advanced Methods* [PDF]. Retrieved October 14, 2016, from http://polisci.msu.edu/jacoby/icpsr/regress3/lectures/week4/15.Loess.pdf

