**Modelling Federal Campaign Fundraising Trends**

**1: Introduction**

I took AP US History last year, and I’m taking AP US Government and Politics this year. As an avid fan of social sciences, I have always been intrigued by the intricacies of the US government. Unfortunately, the US government has not always been portrayed in the most positive light throughout much of history; it has often been deemed as “corrupt”. This is understandable, as seen through evident scandals in the Grant administration and other incidents such as Watergate. In contemporary times, outright scandals are less common, but the public still charges the campaign financing system with corruption (Burke).

The US government is elected by the people, so politicians must campaign to obtain or maintain positions in office. Because elections are hugely publicity-based, whichever candidate spends more money on their campaign often happens to win. Recognizing this, big corporations and other special interests donate huge amounts to support campaigns, in exchange for promotion of their concerns if the supported candidate wins. In recent years, following events such as the 2010 Citizens United case, critics have held more firmly than ever that the campaign finance system is corrupt (Sandler).

Math and government did not always go hand in hand for me. At first, they seemed like irreconcilably different disciplines. However, in government class, we recently learned about the efficiency gap formula – a math equation measuring the extent of gerrymandering. This manifestation of math in politics fascinated me and let me realize that math could be used to approach political trends.

**2: Aim and Approach**

In this paper, I aim to interpret statistics regarding outside money in federal campaign financing. I will be modeling the trend of campaign financing in Senate elections since the year 2008. The year 2008 was chosen because it is the earliest year on which I can find data regarding campaign fundraising. I will be tracking Senate races because it is where most federal campaign money goes – Senate races are far more expensive than House of Representatives races, and they also happen every 2 years, which is twice as frequent as presidential elections. As is seen in the news, the amount of money poured into campaign financing has peaked in 2020 and is projected to continue increasing in the future. I will be trying to find a model that can best describe this increase, and which can possibly project how much more money will be spent on campaigns in the future.

To model Senate campaign financing, I will first obtain the amount raised by the 10 most expensive Senate campaigns for every election cycle since 2008 (OpenSecrets). Senate races are staggered and happen every two years, so these years would be 2008, 2010, 2012, 2014, 2016, 2018, and 2020. I will solely be looking at the top 10 campaigns in terms of money raised, because 10 data points per year is not too many to make the graph cluttered, and not too little for a few abnormally expensive campaigns to skew the data.

Next, I will plot these points on a graph and test out different trendlines to model the trend in recent years. Specifically, I will be using 3 different fits – linear, polynomial, and exponential. I will also attempt to explain and make sense of the different features of each graph. Through this investigation, I hope to learn more about the current trend in American federal campaign financing and what implications it may have for the future.

**3: Data and Basic Processing**

3.1: Raw Data

The raw data with the top 10 amounts raised every Senate campaign cycle is in the appendix of this paper. When this raw data is plotted, the following chart is attained:

At first sight, this data seems interesting. The amount of money raised by candidates seems to be increasing over time. However, there is a dip in the 2014 and 2016 cycles, but the amount raised picks up immediately afterwards and continues to skyrocket after that.

3.2: Outlier Handling

Because I am looking at 10 of the most expensive Senate campaigns every cycle, I am bound to hit outliers; some candidates may raise an extraordinarily large amount not characteristic of the rest of the campaigns that year. For example, in 2010, Connecticut candidate Linda McMahon raised $50 million for her campaign, but no other candidate raised more than $30 million that cycle. I will remove outliers to prevent a few individuals from skewing the entire trend of the data.

To detect and remove outliers, I will be using the interquartile range method. For each year, any value that lies more than 1.5 interquartile ranges above the 3rd quartile or below the 1st quartile out of all values that year will be considered an outlier.

I will discover an outlier in the 2010 data as an example. In 2010, the top 10 amounts of money raised for Senate campaigns in millions of dollars was 50.29, 28.28, 23.78, 21.74, 21.41, 19.38, 18.41, 18.27, 17.76, and 17.14.

The interquartile range is the difference between the first and third quartiles. In a dataset of 10 numbers, the median would be the average between the 5th and 6th numbers, meaning the third quartile is the 3rd largest number and the first quartile is the 3rd smallest number. Hence, the interquartile range here is:

Next, any value that is more than 1.5 IQRs above the 3rd quartile or below the first quartile is considered an outlier. Therefore, if a value satisfies either of the following inequalities, it is an outlier:

In 2010, the 3rd quartile is 23.78 and the interquartile range is 5.51. Hence, the right-hand side of the first inequality evaluates to:

Thus, any value from 2010 which is greater than 32.045 million dollars would be an outlier. Because the greatest value in 2010, 50.29 million dollars, is greater than 32.045 million dollars, it is considered an outlier and will be removed from the set of 2010 points.

3.3: Factoring in Inflation

It must be recognized that this data was collected over more than a decade – in this time, the value of the US dollar has fluctuated, due to global trade and inflation. Thus, I will standardize all dollar values raised by Senate campaigns to 2020 value.

An online inflation calculator will be used to adjust each year’s campaign finance amounts to match the dollar value in 2020. A list of conversion values from each year’s dollar value to the 2020 dollar value is included in this paper’s appendix.

After outliers were removed and all campaign fundraising amounts were standardized in terms of the dollar’s 2020 value, the following plot is obtained:

**4: Mathematical Modelling**

4.1: Linear Model

The first model I will try to fit is a linear model, both because it is the simplest model and because it is the most seen model in the world around us. Following are steps to find the linear equation that best fits these points (Haese).

Let there be data points. Then, each data point could be referred to as , where . A linear line on a 2D Cartesian plane has the equation . Hence, let the y-value of the best fit line at the x-value of the -th data point be defined as:

Here, would be the y-intercept and would be the slope of the line of best fit. This line will try to minimize the cost function , which will be defined as the sum of all distances of the points to the line squared. Mathematically, this function would be:

The previous definition of   can be substituted in:

The expression can then be manipulated and expanded:

It follows that the summation on the right-hand side can be split up and simplified:

Now, the expression for is written as a quadratic of the variable . Because , or the number of data points, is always positive, this quadratic expression in terms of faces upward, thus having a minimum at its vertex. A quadratic equation in the form has a vertex at . Thus, the cost function has a vertex at and is minimized when:

The 2’s and the negative signs in the numerator can be canceled out to obtain:

The sum of all values divided by the number of values is just the average of the values. Note that the average is denoted by the bar symbol. Thus, as and :

Substituting this expression for back into the expression for the cost function , an equation can be obtained which directly relates to (note that all and are constant):

Grouping terms inside the parentheses, the following is obtained:

This expression can then be factored into a quadratic in terms of and the summation can then be split apart, like what was done with :

Because the coefficient of , or , is the sum of squared values, it is always positive. Thus, this quadratic terms of is upwards facing and has a minimum at its vertex too. Thus is minimized when:

In conclusion, to find a linear line of best fit , the values and could be calculated as such:

In my data, let the x-value be the year since 2000 and the y-value be the amount raised in the campaign that year. When all my campaign finance data is substituted into the best fit formula, and are obtained.

From those values of and , the following best fit line can be plotted:

Unfortunately, it appears that few of the data points are close to the line. In all years except for 2020, the line either passes underneath all the points that year or passes above all the points that year, suggesting severe underestimating and overestimating. This supported by the value of this graph, which was calculated by Excel as well. is the coefficient of determination, a number from 0 to 1 which tells us how many values fit the model (Statistics How To). The low value of in this regression is a sign that the line is underfitting.

Furthermore, with this model, there is an x-intercept at ; this would imply that in the 2006 cycle and before, candidates raised zero or negative sums of money! This does not make logical sense, so a linear fit is suboptimal, and I will need to use other models.

4.2: Polynomial Model

At next sight, I want to fit a polynomial model. The graph that I have looks like a cubic function with a positive leading coefficient because it starts out increasing, turns to decreasing, but ends off increasing again.

I will use Microsoft Excel to regress a cubic function through these data points. The graph Excel plotted is shown below:

To begin, the cubic function seems to fit the data much better than the linear model. It passes through cluster of data points every cycle and captures the main shifts of the data from increasing to decreasing and vice versa. It has a higher value than the linear model as well, suggesting better fitting of the line.

However, upon further analysis, it is unlikely that the trend of campaign fundraising dollars follows a polynomial trend. Firstly, while the model seems to fit the years 2008 to 2020 fine, if the data is extrapolated to the left, there is an x-intercept at and the function is negative for any x-value . This hence creates the same problem as there was in the linear model – a cubic model would mean that in cycles prior to 2008, candidates raised negative amounts of money. This does not make logical sense.

Additionally, we can analyze the derivative of the regressed cubic:

The derivative of a function describes the rate at which the function’s value changes. This derivative has two zeros, at a bit past and a bit past . At these two points, the amount of money raised by candidates went from increasing to decreasing and from decreasing to increasing, respectively. These reversals in the trend are likely due to events that happened around these time periods. The amount of money donated to campaigns can be interpreted as having a direct relationship with the public’s interest in politics – the more the public cares about who is in power, the more money people will pour into elections.

Thus, the derivative’s zero around could be explained by the aftermath of the Great Recession in 2008, which devastated the United States economy. Political interest understandably spiked after that – the American people wanted the federal government to act promptly. However, as the economy sailed to recovery in 2010, political interest then dropped back down. However, the derivative’s second zero around is harder to explain, as there were no major events throughout the early to mid-2010s. Thus, the rapid increase in political interest from 2014 to 2018 is unexplained. In this respect, a cubic model for the data does not make a lot of sense either.

4.3: Exponential Model

Because of the shortfalls of both the linear model and the polynomial model, I will use one last model – the exponential model – to fit the data and see if it models the trend well. I think an exponential fit might work best because exponents and logarithms are applicable to a variety of other situations in human science. For example, in economics, ideas such as interest or inflation are based on exponential functions. Other concepts such as population growth are also often exponential. In an exponential model, the value increases at an increasing rate. This looks fitting for this data because the amounts raised in 2020, the last year on which I have data, is higher than the amounts raised for all the other years – such increases are a characteristic of exponential functions.

Excel was used to regress the data exponentially, and the following chart is obtained:

This plot looks more legitimate than the linear and polynomial fits. To begin, this function fits the data better than the linear plot – the value for this model is greater, suggesting that the data better follows this exponential model. Unfortunately, this model’s is lower than that of the polynomial model. Nevertheless, social science cannot be quantified by pure math, so deviation from the trendline to some extent is expected.

More importantly, the physical features of this exponential model are more reasonable. There is a horizontal asymptote at as , which makes sense: it means that at elections further back in the past, less money was raised for campaigns.

Furthermore, the derivative of this function makes logical sense as well. When differentiated using the chain rule, the following is obtained:

The first derivative is another exponential function with all positive coefficients, meaning that it is always positive. This means that campaign fundraising dollars are expected to keep increasing as time passes. It is also interesting to note that the first derivative is just 0.1028 times the original function – this means that the first derivative is increasing as well, meaning that the amount of money being raised by Senate candidates is increasing at an increasing rate year after year. This is consistent with what many news sources say about federal elections becoming more and more expensive as time goes on, so such an exponential model seems to make sense.

However, one deficiency with this model is the fact that it severely underestimates the set of values in 2020, by a margin as wide as the range of all values for other years such as 2014 and 2016. However, this and other deviations can be explained by special circumstances in those years. For example, in 2020, amid the coronavirus and other social issues that have risen to the forefront of the political agenda, the public has become more attentive than ever to politics. Candidates have hence been able to raise at all-time highs because there is great public interest. Similarly, the trendline underestimating the amount raised in Senate campaigns in 2008 and 2010 can be possibly attributed the Great Recession. On the other hand, the trendline overestimating the amount raised in 2014, 2016, and 2018 can be explained by the relative lack of major political or socioeconomic events in those years.

In any case, social science is predictably irrational, and fundraising trends cannot be characterized perfectly by mathematical models. Overestimating and underestimating from time to time is thus expected. With everything that was mentioned in consideration, an exponential model seems to best model the trend of the money raised by Senate race candidates over time, both from a data-fitting and a graph feature standpoint.

**5: Conclusion**

5.1: Conclusion and Implications

In conclusion, the amount of money that Senate candidates are raising for their campaigns is roughly increasing every cycle. Specifically, the amount raised increases in an approximately exponential manner as time passes. An exponential relationship was established because it was the best fit of the 3 models that I tried fitting – linear, polynomial, and exponential.

It was optimal out of the three mainly for the following reasons:

* It captured the overall increase of the data, and its value was better than that of the linear graph
* The physical features of the graph (intercepts, asymptotes, etc.) made logical sense
* The derivative of the function makes sense and is consistent with what is often said about federal campaign costs
* Deviations from the trendline can be explained

Because Senate races are becoming more and more expensive, it can be reasonably well inferenced that House and presidential races are likely costing more and more money as well. If this trend continues, it means that much more money will be poured into politics as time goes on, which is detrimental to the US economy and consistent with what critics would deem as “corruption” in the government. If the nation was aware of this, would we perhaps be led to reevaluate our system and our choices?

5.2: Evaluation and Extensions

This paper had both strengths and weaknesses. One strength was that I tested out different models and justified why each model would fit. I also only used data that was likely representative – I took out outliers and also adjusted values to deflation.

However, a weakness of this paper is that I only tried to fit 3 models. I attempted to choose models that would make sense, but the linear and polynomial models did not capture the essence of the data, while the exponential model may have seemed underfitting due to deviation, especially in 2020. Furthermore, my original objective was to track trends in federal campaign financing. Tracking the top 10 amounts of money raised for Senate races for each 2-year cycle can help demonstrate the general trend of campaign financing, but it is not the whole story and can arguably be misrepresentative of federal campaign financing as a whole. It may be the case that my data is drastically different from trends in campaigns for the House of Representatives, or even from Senators who raised lesser amounts of money each year.

If I had more time, I would collect and process greater amounts of data to get a more complete picture of campaign financing. For sake of clarity and avoiding cluttered data, I only chose the top 10 sums of money raised by Senate candidates each cycle, but I could use data from all Senators next time. I could also analyze campaigns in the House of Representatives or the presidency to see if those trends are consistent with my findings here. Furthermore, I could try out some other models, such as power function, to see if it fits my data better. Lastly, I may even specifically compare campaign financing before and after monumental events such as 911, the 2002 McCain-Feingold Act, or the Great Recession, to see how they shift the amount of money injected into the intricate system of campaign financing.

**6: Works Cited**

“Bivariate Statistics.” Mathematics: Applications and Interpretation HL 2, by Michael Haese et al., Haese Mathematics, 2019.

Burke, Thomas F., "The Concept of Corruption in Campaign Finance Law." (1997). Constitutional Commentary. 1089. https://scholarship.law.umn.edu/concomm/1089

“Excel Regression Analysis Output Explained.” Statistics How To, www.statisticshowto.com/probability-and-statistics/excel-statistics/excel-regression-analysis-output-explained/.

“Fundraising Totals: Who Raised the Most?” OpenSecrets, www.opensecrets.org/elections-overview/fundraising-totals.

Sandler, Joseph E. Federal Election Campaign Act of 1971, www.mtsu.edu/first-amendment/article/1078/federal-election-campaign-act-of-1971.

**7: Appendix**

7.1: Raw Data for Senate Campaign Financing

|  |  |  |
| --- | --- | --- |
| **Year** | **Rank** | **Amount Raised ($)** |
| 2008 | 1 | 22502124 |
| 2 | 19298843 |
| 3 | 16733486 |
| 4 | 16634310 |
| 5 | 13967029 |
| 6 | 13727473 |
| 7 | 13663049 |
| 8 | 11667048 |
| 9 | 10883172 |
| 10 | 9249611 |
| 2010 | 1 | 50285122 |
| 2 | 28276674 |
| 3 | 23777064 |
| 4 | 21741330 |
| 5 | 21408417 |
| 6 | 19378713 |
| 7 | 18414247 |
| 8 | 18272033 |
| 9 | 17764831 |
| 10 | 17141810 |
| 2012 | 1 | 50302456 |
| 2 | 42506349 |
| 3 | 28576174 |
| 4 | 28159602 |
| 5 | 21229108 |
| 6 | 20944796 |
| 7 | 18912557 |
| 8 | 18836531 |
| 9 | 18045722 |
| 10 | 15735457 |
| 2014 | 1 | 22520800 |
| 2 | 21129747 |
| 3 | 20420380 |
| 4 | 18980575 |
| 5 | 18624052 |
| 6 | 17718139 |
| 7 | 17412237 |
| 8 | 16159074 |
| 9 | 16080504 |
| 10 | 15548343 |
| 2016 | 1 | 24544070 |
| 2 | 21571013 |
| 3 | 19566922 |
| 4 | 19562175 |
| 5 | 18575732 |
| 6 | 18561299 |
| 7 | 17183006 |
| 8 | 16935850 |
| 9 | 16494739 |
| 10 | 16364966 |
| 2018 | 1 | 85000301 |
| 2 | 78979726 |
| 3 | 39158866 |
| 4 | 37111560 |
| 5 | 34962008 |
| 6 | 32016390 |
| 7 | 29038046 |
| 8 | 27225359 |
| 9 | 25635466 |
| 10 | 25627244 |
| 2020 | 1 | 138257050 |
| 2 | 130608856 |
| 3 | 124278473 |
| 4 | 103133119 |
| 5 | 99042619 |
| 6 | 94133955 |
| 7 | 92135745 |
| 8 | 85899409 |
| 9 | 74495369 |
| 10 | 71483697 |

7.2: Dollar Conversions due to Inflation

The value of one dollar in 2020 has the value of:

* $0.83 in 2008
* $0.84 in 2010
* $0.89 in 2012
* $0.91 in 2014
* $0.93 in 2016
* $0.97 in 2018
* $1.00 in 2020