Explore the relationship between the total amount of donations received by a candidate and the frequency of donations

Zixi Song,Yiyi Yao,Pu Yuan

2024-03-11

Abstract

This study looks at political campaign finance donations and their value in mayoral elections, specifically exploring the correlation between the total amount and frequency of donations to candidates. Historical data show that financial donations are closely linked to election success, while small and frequent donations reflect broad grassroots support. By analyzing comprehensive donation data from the 2014 Toronto mayoral election, this study aims to uncover the relationship between donation frequency and the total amount of donations received, and explore the potential impact of this relationship on election results. The generalized linear model analysis confirmed that the number of donations has a significant positive impact on the total amount of donations, indicating that the increase in the frequency of donations may lead to a higher total amount of funds raised. In addition, the model suggests that the high variability of donation data may need to be further explored in future studies. The findings underscore the importance of valuing the relationship between the number of donations and the total amount when evaluating campaign strategies, while also providing new perspectives on understanding the role of money in the democratic process. Our findings provide valuable insights into the dynamics of political money and the optimization of election financing strategies.[1]

Introduction

The dynamics of political campaign finance have long been the subject of scrutiny and analysis, where the correlation between financial contributions and election success is a persistent research interest [1]. In the field of local elections, especially mayoral races, patterns of donations can reveal strategic alliances in community engagement and political campaigning [2]. This study looked at the 2014 mayoral election to explore the complex relationship between the candidates who received the highest total donations and those who received the highest frequency of donations. While past research has highlighted the importance of large donations to campaign viability [3], more recent discourse has also considered the importance of small but frequent donations as a measure of grassroots support and as a predictor of election success. By analyzing a comprehensive data set detailing the amount and frequency of donations, the study contributes to a broader conversation about campaign finance, the democratization of money, and election outcomes. Our survey does not only refer to the existing literature on political financing [4], but also introduces a new perspective on the strategic significance of the mayoral campaign finance model to explore the positive relationship between the between the total amount of donations received by a candidate and the frequency of donations.

Method

In the methodological framework of the study,we analyzed campaign finance data from the open Toronto platform, including the name, address, zip code and donation details of each donor in the 2014 mayoral race. Detailed data sets are provided in the section of Experiments and Analyses section .

We take a hybrid approach to dissect patterns in this data. In terms of quantity, we apply statistical technology to count the frequency of donations received by candidates and the total amount of donations

[1]The project can be access via https://github.com/KristySzx/TermPaper

received, and carry out data cleaning on missing values, outliers and outliers. Then, we conduct regression analysis on the frequency and total finance, and explore the relationship between candidates, total donations and the number of donations received through the regression model information. The deep relationship between the two is further inferred from the conclusion of the model.

Data and Analysis

Introduction to Data Sets

This investigation draws on a comprehensive data set of 10,201 entries that intricately document financial contributions made during the 2014 mayoral campaign. Each entry is rich in detail, spanning several information features:

- **Donor name:** The identity of the donor.
- **Donor address:** Record the address of the donor and provide geographical context.
- **Contributors' ZIP codes:** Zip codes are provided to allow for finer spatial analysis.
- **Contribution amount:** Indicates the monetary value of the contribution, which is crucial to assessing the total amount of the contribution.
- **Contribution Type description:** Describes the nature of the contribution.
- **Description of goods or services:** If applicable, specify the goods or services offered.
- **Description of Donor types:** Categorize the types of donors.
- **Relationship with the candidate:** Note any specific relationship between the contributor and the candidate.
- **President/Business Manager:** List any senior executives or business managers involved that may indicate greater organizational support.
- **Authorized Representative:** Designate any authorized representative in connection with the donation.
- **Candidate:** Name of the beneficiary candidate of the donation.
- **Office:** The office for which a candidate is designated to run.
- **Constituency:** Indicates the electoral constituency to which the contribution relates.

This dataset, which examines not only the number and total amount of contributions, but also the frequency of contributions by each candidate, is a key asset. By dissecting these data, this study aims to shed light on the relationship between the size of financial support and the breadth with which donations occur, providing insight into the dynamics of campaign finance in municipal elections. The depth and breadth of the dataset allows us to conduct a multifaceted analysis that is both refined and comprehensive, providing a substantive basis for subsequent statistical and thematic studies.

Data Pre-processing

Missing value processing When studying the relationship between the total amount of donations received by candidates and the frequency of donations, data integrity is critical to ensure the accuracy and reliability of the analysis. Therefore, in the data preparation phase, we paid close attention to the lack of donation amounts and candidate information, and took steps to supplement the missing data to strengthen the research foundation.

Considering the purpose of the study and the nature of the data, we decided to adopt the mean interpolation method to deal with these missing values [5]. Mean interpolation is a simple and effective technique that preserves the overall statistical properties of a dataset by replacing the missing values with the mean values of the corresponding variables. However, after a preliminary review of the dataset, we found no missing

values, particularly in terms of donation amounts and candidate information. Therefore, this part mainly implements data screening but does not process the data.

Outlier detection processing After the detection of missing data is completed, the outlier value is detected[6]. The specific processing process is as follows:

1. Boxplot visualization: Create a box diagram. This visual representation quickly identifies potential outliers, data points that appear outside the whisker of the boxplot. This is typically 1.5 times the quartile range below the first quartile or above the third quartile.
2. Calculate the IQR: The IQR is a measure of statistical dispersion and is the difference between the 75th and 25th percentiles of the contribution amount. It is used as a scale to define the "normal" range.
3. Establish boundaries: Using IQR, calculate the lower and upper limits. Any data point that lies outside these boundaries is considered an outlier. The lower limit is the 25th percentile minus 1.5 times IQR, and the upper limit is the 75th percentile plus 1.5 times IQR.
4. Filter outliers: The data is then filtered to exclude any contributions beyond the lower and upper limits. This results in a subset of the data that excludes outliers, which can distort the analysis of donation patterns.

Histogram of filtered data: After the outliers are removed, a histogram of the cleaned data is generated. The histogram provides a visual representation of the distribution of the frequency of donations, providing insights into the amount of joint contributions and the general pattern of contributions after accounting for outliers.

By conducting outlier detection, this study ensures that the analysis of the relationship between total donation and donation frequency is not unduly influenced by extreme values. This step is critical to maintaining the integrity of statistical analysis, especially when trying to understand the typical financial support a candidate receives, as outliers may represent atypical large contributions and not necessarily reflect the broader base of supporters. The clean, nontrivial data set paves the way for more accurate examination of potential patterns and relationships in campaign contribution data.

The boxplot created is shown in Figure 1, and the histogram obtained is shown in Figure 2. As can be seen from Figure 1 and Figure 2:

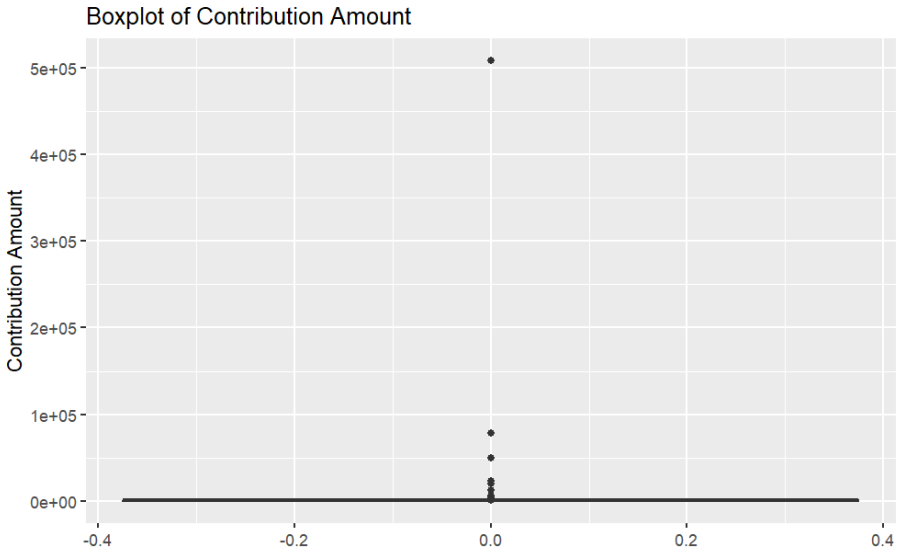

Figure 1: The boxplot of contributions amount shows an abnormal distribution of large contributions

1. Boxplot: Several points above the top whisker, indicating the presence of outliers. In a boxplot diagram, any data point that lies outside the whisker line is considered an outlier. The presence of these points indicates that there is indeed an outlier in the donation amount. However, the outlier of this data is in

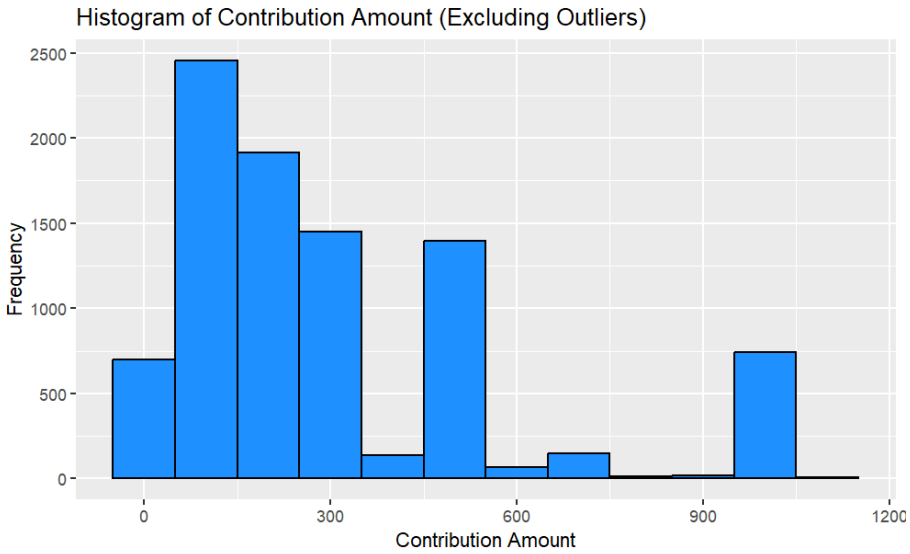

Figure 2: The histogram of total donations shows the right distribution of donations

line with the right-sided distribution in the donation process: that is, the majority of small donations, and the proportion of large donations is not high.

2. Histogram: Most donation amounts are concentrated in the lower range, the higher the amount, the frequency gradually decreases. There are still bars in the histogram at the higher donation amounts, indicating that although the number of such donations is small, they do exist.

Given these observations, it is clear that while outliers have been identified, the overall trend in donation amounts favors smaller values, which is a common pattern in real-world financial giving data.

Regression result analysis

After data preprocessing, reasonable data are integrated and analyzed. The specific operations are as follows[7]:

1. Integrate the frequency of donations received by different candidates as an independent variable
2. Aggregate the total amount of donations received by different candidates as the dependent variable
3. Perform GLM-based regression analysis on the independent and dependent variables mentioned above

A Generalized Linear Model (GLM) is an extension of traditional linear regression that allows the probability distribution of the response variable (dependent variable) to belong to an exponential family of distributions rather than just a normal distribution. This allows GLM to be applied to more types of data, such as count data, binomial distribution data.

In GLM, the response variable. The mean of Y is considered an independent variable The result of a linear combination of X transformed by a link function. It can be mathematically expressed as:

$$g(E(Y)) = X\beta \tag{1}$$

where $g()$ is the link function,$E(Y)$ is the expected value or mean of the response variable,X is the design matrix of the independent variables, and β is the regression coefficient [8].

In our study, Exploring the Relationship between the Total Amount of Donations Received by Candidates and the frequency of donations, we used generalized linear regression to model the relationship between the total amount of donations received by candidates and the frequency of donations. Given that total contributions are a positive continuous variable, it is reasonable to assume that their distribution is close to a

normal distribution. In this case, we choose the identity link function, which makes our GLM similar to a traditional linear regression model. Regression analysis showed that donation frequency was a significant positive predictor, indicating that candidates who donated more frequently tended to have higher total donations.

Table 1: Regression model key parameter results

term	estimate	std.error	statistic	p.value
(Intercept)	72955.45	77788.13	0.937	3.57e-01
candidate_frequency	414.82	63.76	6.505	8.18e-07

Table 2: The overall test of the model

null.deviance	df.null	logLik	AIC	BIC	deviance
9.94e+12	26	-384.47	774.94	778.83	3.69e+12

From Table 1 and Table 2, we can draw the following conclusions:

- **Intercept:** The intercept of the regression model is 72955.45. The intercept represents the predicted amount_sum when candidate_frequency is 0. However, the p-value of the intercept is 0.357, which means that the intercept is not statistically significant, and we do not have enough evidence to show that the intercept term is significantly non-zero.
- **Candidate_frequency variable:** The coefficient of candidate_frequency is 414.82, indicating that each increase in the number of donations, the total amount of donations received by the candidate increases by 414.82 units on average. The T-value of this variable is 6.505, and the P-value is very small, far less than the common significance level of 0.05, which indicates that candidate_frequency has a significant positive impact on amount_sum.
- **Model Deviance:** Null deviance and Residual deviance show how well the model fits the data. The original deviation is the difference between the null model with only the intercept term and the full model, and the residual deviation is the difference between our model and the data [9]. There is a significant decrease in the residual deviation relative to the original deviation, suggesting that the model containing candidate_frequency provides a better fit.
- **Model validity**: Given the statistical significance of candidate_frequency, we can conclude that there is a significant positive correlation between the number of donations (donation frequency) and the total amount of donations received by a candidate. However, due to the size of the dispersion parameters and the AIC value, we may need to further evaluate whether the model is the best representation of the data or explore other potential variables.
- **Number of Fisher Scoring iterations:** The model converges after only two Fisher Scoring iterations, which generally indicates that the model converges well.

In summary, this model shows that the number of donations is an important factor affecting the total amount of donations received by a candidate.

Discussion

In our study Exploring the Relationship between the Total Amount of Donations Received by a Candidate and the frequency of donations, we analyzed in detail the statistical association between the number of donations received by a candidate donation frequency and the total amount of donations received by the

candidate by constructing a generalized linear model (GLM). The results show that there is a significant positive correlation between the frequency of donation and the total amount of donation. In other words, as the number of donations received by a candidate increased, the total amount of contributions accumulated increased significantly.

The results reinforce the importance of frequency in political fundraising: Frequent small donations can be just as important as a few large donations, since an increase in the number of small donations can significantly boost a candidate's total fundraising. In addition, the increase in the frequency of donations may also reflect the activity and breadth of a candidate's base of supporters, which is a positive sign for the campaign.

However, we need to be careful with our model choices and assumptions. While the frequency of giving is a key indicator, we cannot rule out that other unobobserved variables may also have an impact on total giving. Future research may therefore need to explore additional explanatory variables, such as the socioeconomic status of the donor, the political climate at the time of the donation, and the phased nature of the campaign, to fully understand the complex dynamics behind giving behavior. In addition, because the assumption of a Gaussian distribution may not fully apply to highly right-skewed donation data, further research may need to consider more sophisticated statistical models to more accurately capture the characteristics of the data.

References

[1] Maria Petrova, Ananya Sen, and Pinar Yildirim. Social media and political contributions: The impact of new technology on political competition. *Management Science*, 67(5):2997–3021, 2021.

[2] Alexei V Ovtchinnikov and Eva Pantaleoni. Individual political contributions and firm performance. *Journal of Financial Economics*, 105(2):367–392, 2012.

[3] Ank Michels. Innovations in democratic governance: how does citizen participation contribute to a better democracy? *International Review of Administrative Sciences*, 77(2):275–293, 2011.

[4] Daniella Acker, Ayan Orujov, and Helen Simpson. Political donations and political risk in the uk: Evidence from a closely-fought election. *Journal of Banking & Finance*, 92:146–167, 2018.

[5] Norazian Mohamed Noor, Mohd Mustafa Al Bakri Abdullah, Ahmad Shukri Yahaya, and Nor Azam Ramli. Comparison of linear interpolation method and mean method to replace the missing values in environmental data set. In *Materials science forum*, volume 803, pages 278–281. Trans Tech Publ, 2015.

[6] Karanjit Singh and Shuchita Upadhyaya. Outlier detection: applications and techniques. *International Journal of Computer Science Issues (IJCSI)*, 9(1):307, 2012.

[7] Richard B Darlington and Andrew F Hayes. Regression analysis and linear models. *New York, NY: Guilford*, pages 603–611, 2017.

[8] John Ashworth Nelder and Robert WM Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 135(3):370–384, 1972.

[9] Annette J Dobson and Adrian G Barnett. *An introduction to generalized linear models*. Chapman and Hall/CRC, 2018.