# POL S 205 Final Project

Andrew Liang

Do campaign donation patterns for parties and corporations affect a candidate's outcome in primary elections? I hypothesize that in primary elections, increased party support for one of its candidates will result in that candidate receiving higher shares of votes in their primary election. If a party's primary election involves several candidates, the party can easily donate to candidates that align closest to the party's platform, granting them financial support. However, just as parties may end up offering more financial and political support to their preferred candidates, support from outside sources, such as corporations, can confound the impact that party support may have on a candidate's chance of advancing to the general election. If parties are likely to support a candidate that does not align with the interests of corporations, those corporations can choose to contribute financial support to their own preferred candidates. This extra support may either compound existing support to a candidate from their own party, or serve to counteract the party's influence if corporations prefer a rival candidate.

Based on these implications, my hypotheses are as follows:

*H1: Increased campaign contribution amounts to candidates from their parties are associated with an increase in the percent vote share the candidate receives in their primary election.*

*H2: Increased total campaign contribution amounts to candidates are associated with an increase in the percent vote share the candidate receives in their primary election.*

## Data and Methods

To investigate the relationship between party and corporate support for a candidate's campaign and their share of votes in the primary election, I used data from the Stanford Database on Ideology, Money in Politics, and Elections (DIME), which contains information on candidates for federal and state offices, contributions received for their campaigns, and their results in both the primary and general elections (if applicable) between 1979 and 2022 (Bonica 2023). Each observa-

tion in DIME's candidate dataset describes a specific candidate in a given election cycle (candidate-cycle); because not all election candidates are fully included, this dataset could be viewed as an unbalanced panel dataset with different candidates included or omitted based on whether they chose to run for office in a given cycle. The data in the candidate dataset was provided by the Federal Election Commission and other state election offices *after* elections concluded; as a result, it is an observational dataset where contributions were recorded without any direct manipulation by the principal investigator, based on election results across the country. Furthermore, the candidates in the dataset theoretically represent the population of all individuals who ran for office; this means that no sampling was conducted on the original dataset.

Because I wanted to investigate the effect of corporate and party support on a candidate's outcome in the primary election, I chose to look only at cases where results in the primaries were recorded for a candidate. Primary results in the candidate dataset were obtained solely from the Federal Election Commission, so data on candidates for state office are inherently omitted. I also excluded cases where the candidate ran unopposed in the primary election, regardless of whether they chose to spend money for their campaign or not. This ensured that the focus of this project was on *contested* primary elections, where each candidate is guaranteed to *not* receive every vote in the primary election. To operationalize my dependent variable – percentage of votes a candidate receives in their primary – I proceeded with a narrower definition of contested primaries by simply omitting elections where candidates received either 100 percent or 0 percent of the vote. While this potentially eliminated candidates in a multi-candidate election who simply failed to receive any votes, merely omitting candidates who received all of the votes in an election would leave the possibility that candidates in such uncontested elections remain in the dataset. This solution may slightly affect the construct validity of my dependent variable, but it also appears to be the most effective in satisfying the measure's criterion for "contested" primary elections.

I operationalize my independent variables – party and corporate support – as the total amount of donations received from each category of contributor for a candidate during a specific election cycle, using data from DIME. While DIME's candidate database contains a variable for the total amount of contributions from a candidate's party, it does not differentiate amounts received during the candidate's primary election and their general election, if applicable. Furthermore, the dataset does not contain a variable on the total amount of corporate donations a candidate receives during their primary. To create my independent variables, I instead aggregated contributions for each candidate per cycle through DIME's

*contributor* database, which includes information on every contribution reported to the Federal Election Commission between 1979 and 2022. With each observation corresponding to a single contribution, either from an individual or an organization, I could easily aggregate contributions for each candidate-cycle based on selected criteria.

To create the total party contributions variable, I specifically chose to keep only donations from the candidate's own party. Political parties often donate money to candidates through specific committees or funding accounts, so I omitted contributions from individuals. To ensure that I only analyzed donations from the candidate's own party, I filtered for contributions where the contributor's name included the name of the candidate's party. I then omitted contributions where the source was clearly not a party committee or funding account but whose names nevertheless included the party's name.[1] Next, I aggregated the total amount donated to each candidate for each cycle, identifying candidates by their state and unique DIME identification number in the dataset.

Creating the total corporate contributions variable was somewhat more difficult, though the process ultimately mirrored that of the total party contributions variable. Like the party contributions variable, I decided that I would analyze donations only from corporate organizations, omitting contributions from individuals. While this inherently excludes a considerable number of contributions from individuals who may be involved in corporations, such as CEOs, differentiating contributions from non-corporate individuals with executives and other employees would simply be an impossible task given that there were over 800,000 contributions remaining. I first filtered for contributions where the contributor's name included typical keywords such as "Incorporated" or "Corporation," as well as their abbreviated forms. Because this initial filtering failed to eliminate some individual contributions, I then kept contributions where the contributor type was "C," which denotes either a political action committee or an organization (Bonica 2023), and similarly aggregated the total amount of contributions by candidate-cycle.

After collapsing the contribution variables to the candidate-cycle level, I then attempted to filter both variables for contributions that were recorded prior to the candidate's primary election date before merging the two datasets. This, however, resulted in multiple issues: first, the Federal Election Commission only posted dates for congressional primary elections starting with the 1992 cycle; in order to obtain the dates of congressional primaries between 1980 and 1992, I instead

1. An example of this is the Majority Initiative To Keep Electing Republicans (MikeR).

relied on election statistics from several volumes of the *America Votes* handbooks.[2]

Second, the contributions datasets for the 2020 and 2022 cycles were too large to handle on any accessible computer, even those with larger memory sizes of up to 64 gigabytes. As a result, I was forced to drop the 2020 and 2022 cycles from my analysis.

Third, after aggregating the party contributions for each remaining candidate, I noticed that the amounts did not correspond to those found in the main candidate dataset. These four major issues, along with the inherent issue of attrition among units,[3] undoubtedly have impacted the construct validity of my independent variables. However, given that other avenues of trying to calculate contributions manually have been unsuccessful, this was the best method to have a comprehensive dataset of aggregated primary contributions at the candidate level. Notably, there are no other sources that provide a similar dataset on primary-election contributions from political parties and corporations, even from the Federal Election Commission. As a result, it is impossible to consult other sources to determine the construct validity of my independent variables,[4] and this project was a good-faith attempt at constructing a dataset that measures the total amount of contributions from political parties and corporations to federal Congressional candidates at the primary election stage.

Ultimately, this meant that I ran an observational panel study with an unbalanced panel dataset with 5,629 units and 19 time periods, comprising 38 years in total (Johnson, Reynolds, and Mycoff 2008). However, I was only able to obtain data on corporation contributions for 4,215 candidates.

**Variables and Descriptive Statistics**

Table 1 displays the summary statistics for our key dependent and independent variables. As noted above, while the final dataset contains values for primary vote share and party contributions for all candidates, only 4,215 candidates had data on corporate contributions.

2. In addition to searching online sources for primary dates prior to the 1992 cycle, I also attempted to reach out to the maintainer of the DIME dataset, but was unable to receive a response prior to the submission of this project. I was also unable to reach out to him while trying to identify a way to calculate the amount donated to a candidate during their primary cycle from the candidate dataset.

3. Given that not all candidates receive donations or run for election each year, some units will inevitably drop out or join each year.

4. I attempted to use other campaign-finance websites such as OpenSecrets, but found that even trying to verify the accuracy of reported donations from DIME was difficult.

**Table 1.** Summary Statistics of Key Variables

| Statistic | N | Mean | St. Dev. | Min | Median | Max |
|---|---|---|---|---|---|---|
| % Primary Vote Share | 5,629 | 56.45 | 29.51 | 0.02 | 59.84 | 99.99 |
| Total Contributions, Party | 5,629 | 22,787.74 | 128,077.50 | 0.00 | 3,150.00 | 7,421,110.00 |
| Total Contributions, Corporations | 4,215 | 124,770.60 | 314,904.20 | 0.00 | 31,100.00 | 8,000,671.00 |

Contribution amounts in U.S. Dollars.

The key dependent variable is the percentage of votes a candidate receives in their primary election, a continuous ratio variable with values between 0 and 100.[5] This variable is based on the results of the elections obtained directly from the Federal Election Commission; therefore, I assume that the variable does not have issues with construct validity, except those mentioned above. The mean percentage of candidates' vote share is about 56.5 percent, indicating that the majority of candidates present in the dataset won their primary election. Although this project focuses on contested primary elections, Figure 1 shows that the majority of candidates within the dataset receive almost all or none of the votes.

The key independent variable is the total amount contributed to a candidate in an individual cycle by that candidate's party, a continuous ratio variable. Figure 1 indicates that the variable is positively skewed, and that most candidates in our data set did not receive a significant amount of money from their own political party. Table 1 also shows that the mean total party contribution, approximately $22,788, is significantly greater than the median party contribution, $3,150. In fact, the largest amount donated to a candidate by that candidate's party was approximately $7,421,110, a stark contrast to the relatively low mean and median party contribution amounts.[6]

The confounding variable in my analysis is the total amount contributed to a candidate in an individual cycle by corporations, also a continuous ratio variable. Like the total amount of party contributions, the distribution of the total corporate contributions is positively skewed, indicating that most candidates also received insignificant amounts of money from corporations. However, those who receive contributions from corporations receive much higher amounts compared to the contributions they receive from their parties; the mean amount of money con-

5. The variable in DIME originally represented the *proportion* of votes a candidate received using values between 0 and 1. However, I decided to transform the proportions so that they represented the percentages for simpler interpretation.

6. After consulting the data, the aggregated sum was traced to multiple donations for John Boehner's 2012 reelection campaign to the House of Representatives.

tributed to a candidate by corporations is about $124,771, over five times the mean amount of money contributed to a candidate by their own party, while the median amount is nearly ten times that of total party contributions, at about $31,100. Furthermore, the largest amount ever donated to a candidate's campaign by corporations is $8,000,671.

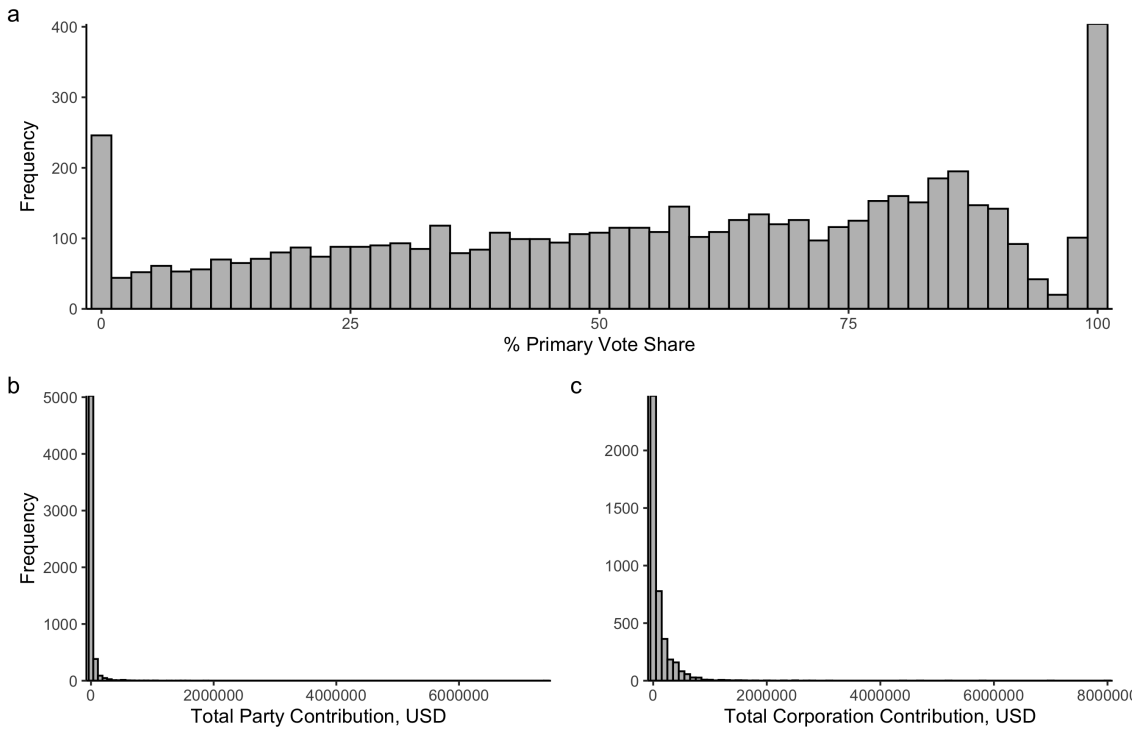


**Figure 1.** Histograms of Key Dependent and Independent Variables. Subfigure A displays histogram of vote share percentages received by candidates in primary races. Subfigure B displays histogram of total party contributions to candidates, in U.S. Dollars, and subfigure C displays histogram of total campaign contributions, in U.S. Dollars.

Because our independent variables are positively skewed, as seen in Figure 1, interpreting results with these variables is much more difficult. As a result, I logarithmically transformed the variables; Figure 2 displays histograms of these transformed variables. One immediately apparent feature is that the logarithmic transformation has resulted in both variables having a relatively normal distribution, although the distribution of the party contributions can be also described as bi-

modal. The distribution of the log-transformed corporation contribution variable also appears to have become slightly negatively skewed, though it still appears to approximate the normal distribution.

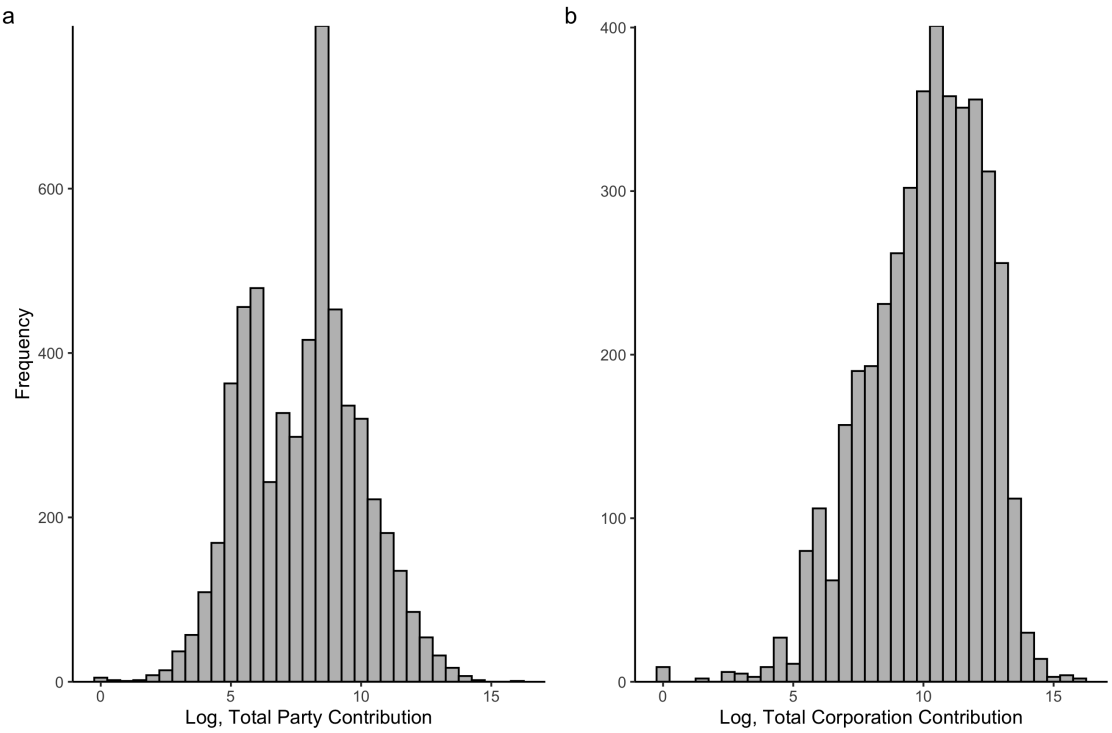


**Figure 2.** Histograms of log-transformed independent variables. Subfigure A displays histogram of log-transformed total party contributions to candidates, in U.S. Dollars, and subfigure B displays histogram of log-transformed total corporate campaign contributions, in U.S. Dollars.

Log-transforming the independent variables also provides us with a method to corroborate our findings from Table 1. Figure 3 displays boxplots of our log-transformed independent variables; the boxplots clearly show that the median amount contributed to a candidate by corporations is greater than the median amount contributed by a candidate's own party. In fact, most campaign contributions appear to be significantly greater than party contributions, which may be a result of campaign finance regulations in the United States.

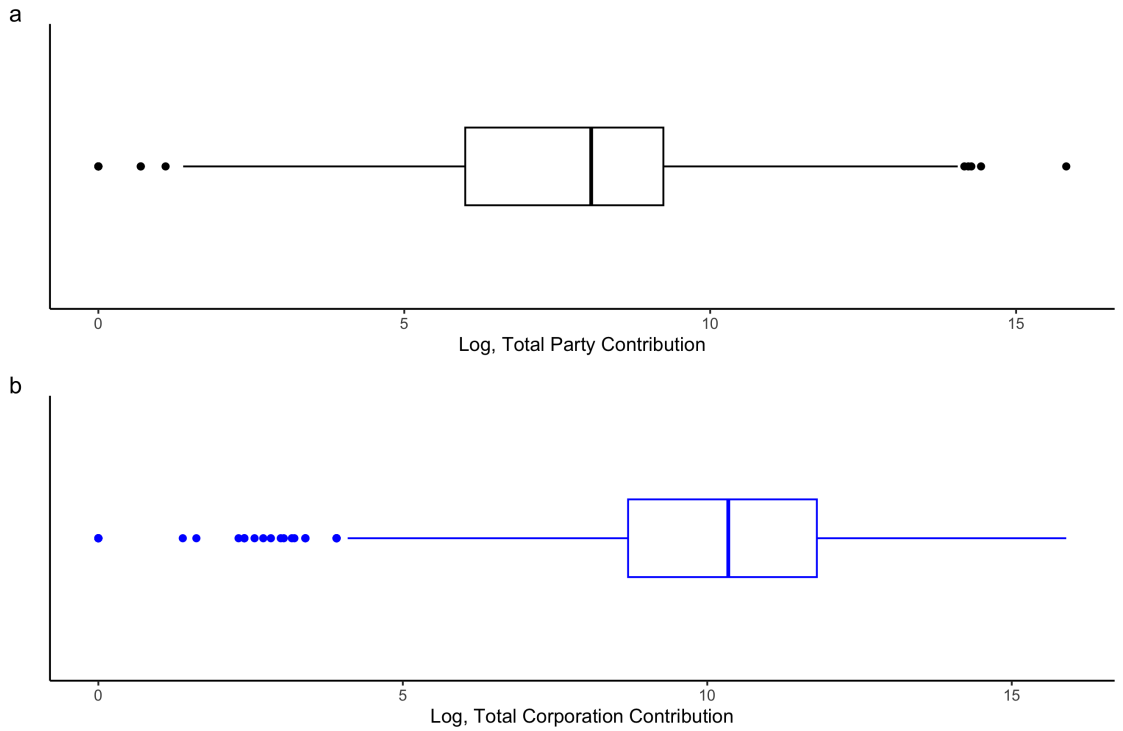


**Figure 3.** Boxplots of log-transformed independent variables. Subfigure A displays boxplot of log-transformed total party contributions to candidates, in U.S. Dollars, and subfigure B displays log-transformed boxplot of total corporate campaign contributions, in U.S. Dollars.

## Model

To estimate the effect of party and corporate support on a candidate's vote share in the primary, I estimate several ordinary least squares models regressing my dependent variable, the percentage of votes received by a candidate in their primary election, on my independent variables, total party and corporate contributions. Based on my two hypotheses, I first estimated bivariate ordinary least squares models for each of my independent variables, before estimating the full multivariate model shown below:

$$\begin{aligned}\text{Primary Vote Share}_i = \beta_0 + \beta_1 \log \text{Total Party Contribution}_i + \\ \beta_2 \log \text{Total Corporate Contribution}_i + \epsilon_i\end{aligned}$$

The full model should allow us to measure the approximate effect of party

contributions on candidates' outcomes in primary elections, while controlling for our confounding variable (the total amount of corporate contributions).

## Results

The results of all three regression models are printed in Table 2. Model 1 regresses a candidate's percentage of the votes in their primary election on the log-transformed total amount of contributions received from their party. The coefficient on the log-transformed party contributions is positively signed and statistically significant, indicating substantial support for Hypothesis 1. The effect of party contributions also appears to be substantively significant: the results suggest that a one-percent increase in the total amount of party contributions to a candidate is associated with an approximately 0.042 percentage-point increase in that candidate's vote share, on average.

**Table 2.** Primary Vote Shares and Party and Corporate Contributions.

| | *Dependent variable:* | | |
|---|---|---|---|
| | Percentage of Votes Received in Primary | | |
| | (1) | (2) | (3) |
| Log, Total Contributions, Party | 4.194*** | | 2.096*** |
| | (0.169) | | (0.207) |
| Log, Total Contributions, Corporations | | 3.286*** | 2.403*** |
| | | (0.189) | (0.206) |
| Intercept | 23.785*** | 29.258*** | 20.991*** |
| | (1.372) | (1.956) | (2.098) |
| Observations | 5,629 | 4,215 | 4,215 |
| Adjusted $R^2$ | 0.098 | 0.067 | 0.089 |
| Residual Std. Error | 28.024 | 27.078 | 26.758 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 | | |
| | OLS standard errors in parentheses. | | |
| | Significance levels reported from two-sided tests. | | |

Model 2 tells a similar story for the total amount of corporate contributions; it regresses the candidate's percentage of votes received in the primary on the

log-transformed total amount of corporate contributions. The coefficient is also positively signed and statistically significant, indicating substantial support for H2. However, the effect of corporate donations appears to be slightly lower than those of party donations; a one-percent increase in the total amount of corporate contributions is associated with an approximately 0.035 percentage-point increase in the candidate's vote share.

The full model is displayed as Model 3 in Table 2. Here, all of the coefficients on the log-transformed independent variables are statistically significant and positive, indicating support for both H1 and H2. However, while the bivariate models indicated that party support may have a larger effect on a candidate's vote share in the primaries, our full model suggests a different story. First, the effect of both variables is decreased somewhat compared to the bivariate models. A one-percent increase in the total amount of party contributions received is associated with an approximately 0.021 percentage-point increase in the candidate's vote share, holding corporate contributions constant. Holding party contributions constant, however, a one-percent increase in the total amount of corporate contributions is associated with an approximately 0.024 percentage-point increase in the candidate's vote share in their primary, on average.

Understanding the actual effect of these variables, especially in light of our logarithmic transformation, requires a more complex interpretation. As Figure 4 illustrates, exponentiating $e$ by the reciprocal of our coefficients allow us to find that candidates who seek a one percentage-point increase in their vote share should expect parties to increase their contributions by a predicted 61.2-percent of their original contribution amount, when holding corporate contributions at one dollar. This amount is substantial when the total amount that parties donate to their candidates are already significantly large, indicating that parties may be more incentivized to donate more money to ensure greater returns in primaries.

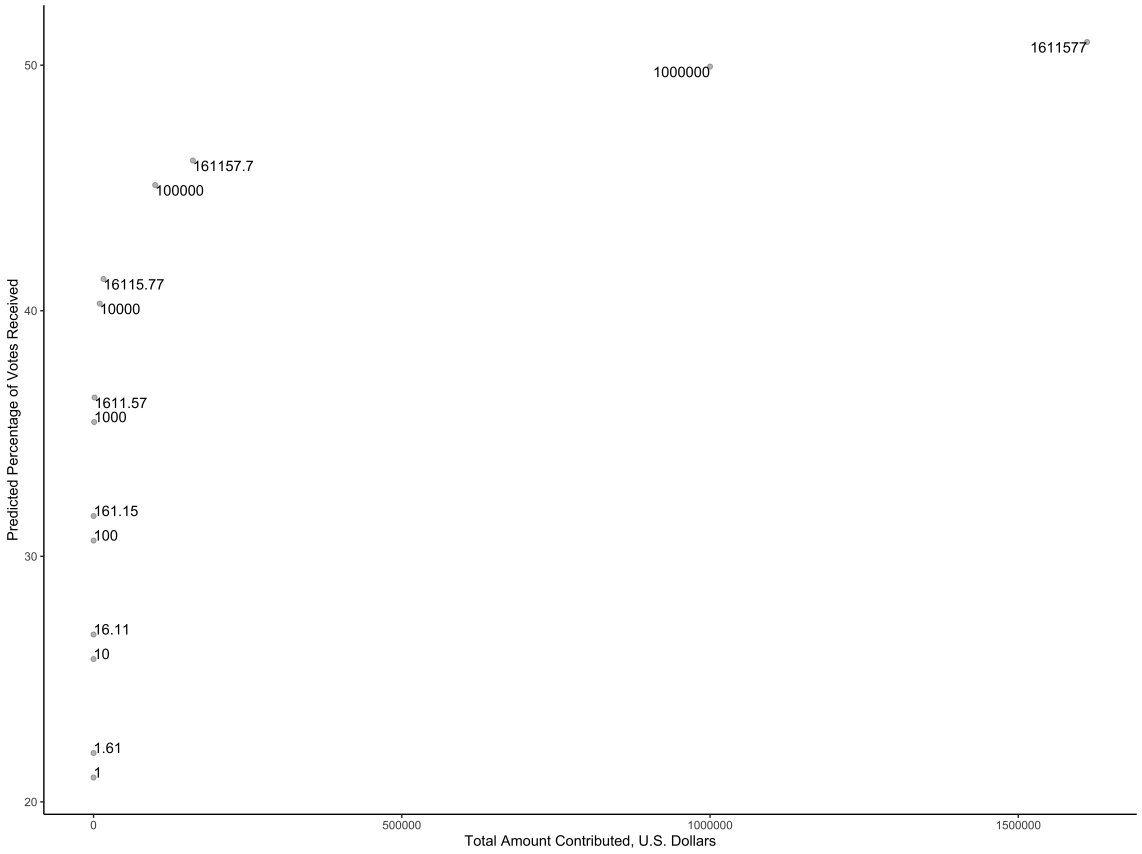


**Figure 4.** Scatterplot of Predicted Vote Shares Against Hypothetical Party Contribution Amounts.

Similarly, Figure 5 shows how, after holding total party contributions at one dollar, a candidate that seeks a one percentage-point increase in their vote share should expect corporations to increase their contributions by a predicted 51.6-percent of the current amount. These findings may explain why larger contributions from certain organizations can be instrumental in how candidates perform in their primary elections - while corporations need to contributed half their original amount more to see improvement based on our model, that burden is somewhat lessened when compared to party contributions.

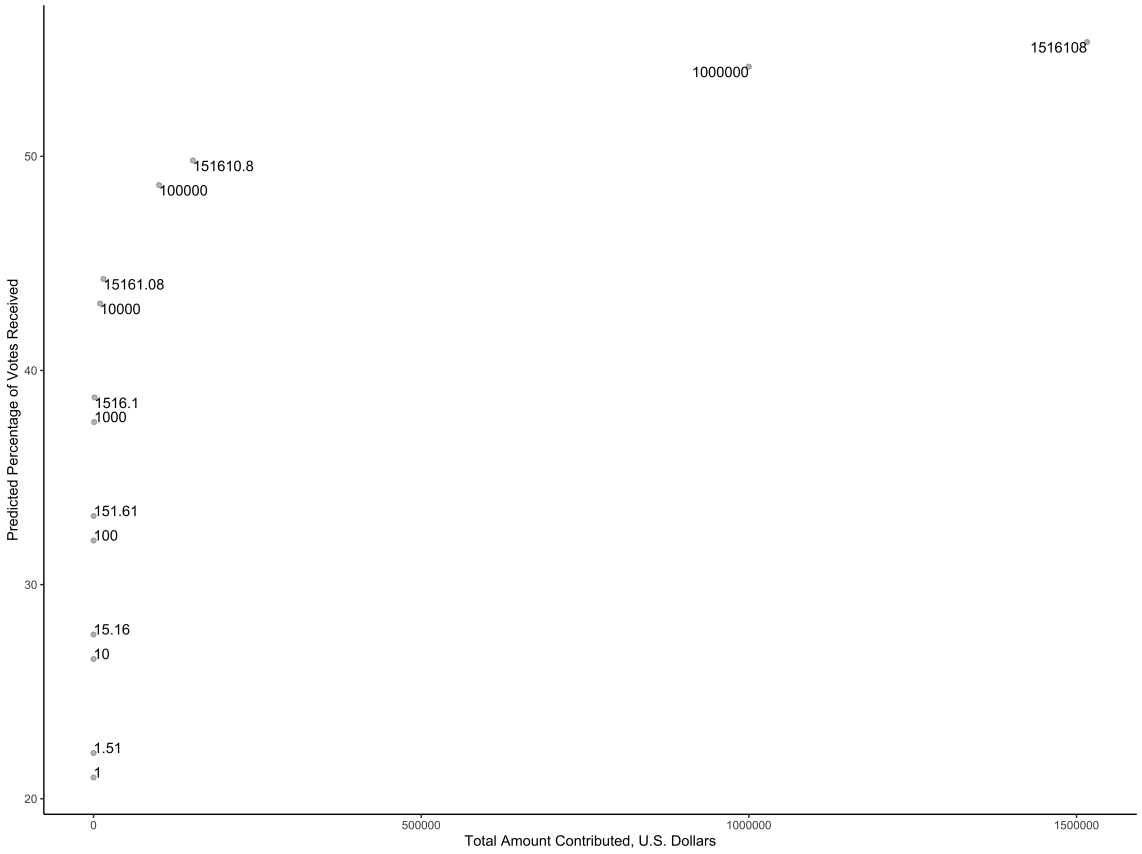


**Figure 5.** Scatterplot of Predicted Vote Shares Against Hypothetical Corporate Contribution Amounts.

## Conclusion

In this project, I attempted to investigate the relationship between party support and candidate outcomes in primary elections. By looking at the total amount of money parties and corporations donated to candidates in 19 election cycles, I found that party and corporate supports appeared to have positive effects on the percentage of votes that candidates received in their primary, with corporate support appearing to have a slightly larger effect on the candidates' vote share compared to party support. Furthermore, larger amounts of contributions appear to result in greater electoral gains for candidates. These findings supported both of my original hypotheses.

One major weakness in this project's internal validity is the fact that the data

on which I rely are observational data. Given that I have not controlled for other factors that could influence a candidate's outcomes in the primaries, it would be difficult to infer the *causal* effect of party and corporate support on candidate outcomes in primary elections. Other issues regarding reliability and validity were present in the data. For example, an inability to contact the maintainer prevented me from verifying the validity of my independent variables by comparing them to similar measures from DIME's candidate dataset, and other sources for campaign finance data appear insufficient for verification. This project was a good-faith gap in providing a look at how much candidates receive from different organizations during their primary campaigns.

To the contrary, the observational nature of DIME's datasets mean that the external validity of our findings is particularly strong. Because we relied on actual data on campaign finance in the United States, our findings are generalizable to candidates running for office in the United States House of Representatives and the Senate, who may be subjected to contested primary elections. However, the findings may not generalize as well to candidates who run unopposed in their primary, or candidates for state office, where state electoral regulations may be somewhat stricter than federal regulations. Regardless, because Congressional races are often scrutinized heavily, these findings may prove to be interesting when considering when and where organizations donate politically.

Ultimately, the results suggest a critical point for how money in politics may influence the outcomes of primary elections: while both how much a candidate spends for their campaign and contributions from parties appear to matter for the candidate's vote share in their primary election, the latter may have a higher influence in the success of candidates than how much money a candidate spends for their campaign in the primary election.

# References

Bonica, Adam. 2023. Database on Ideology, Money in Politics, and Elections: Public version 3.1 [Computer file]. https://data.stanford.edu/dime.

Johnson, Janet Buttolph, H.T. Reynolds, and Jason D. Mycoff. 2008. *Political Science Research Methods.* 6th ed. CQ Press.