

Predicting 2006 General Election Outcomes Using Campaign Finance Data

Abstract: Statistical analysis can provide a valuable tool for discerning trends and outcomes in politics. In this study, we built a model that attempted to quantify the impact of a political campaign's finances as a baseline before the 2010 Citizens United ruling, which removed caps for corporate donations. Using 2005-2006 campaign finance data from the Federal Election Commission, we began our investigation with seven initial predictor variables and a subset of 935 cases (each being an individual congressional candidate). Exploratory data analysis revealed a promising relationship between incumbency and election results, along with a severe right skew of all quantitative variables that informed us to proceed with caution. A multiple logistic regression was most appropriate given the binary nature of the response variable. Through automated stepwise selection, we obtained the optimal final model and concluded that the most relevant predictors of election results were incumbency status, political party, and contributions from political committees.

I. Background and Introduction

Statistical analysis provides helps us predict important developments and outcomes in politics, from polling to demographic shifts. We were specifically interested in using statistics to understand the influence of money on Congressional elections. Campaign donations and expenditures have consistently attracted significant attention since the early 20th century in response to concerns about election integrity and the responsiveness of candidates to their financial supporters at the expense of their constituency at large (Mann 2003). Campaign finance reform has remained a contentious issue, particularly in response to the impact of the 2010 Supreme Court *Citizens United v. FEC* ruling, which removed caps on corporate donations to political campaigns as enacted under the 2002 Bipartisan Campaign Reform Act (Oyez).

In order to determine a baseline sample method for quantifying the impact of hard money donations to congressional candidates prior to the Citizens United ruling, we conducted a statistical analysis of campaign finances in the 2006 general election. We wanted to analyze whether the amount or type of financial resources a campaign receives can be used to predict whether it will be successful, or if other factors are more significant. In doing so, we hope to provide information about the impact of money on political campaigns, which can help inform voters on candidates' different sources of power and their relative importance. This kind of research is vital to maintaining election integrity and transparency as well as exposing the influence of special interests in our political system.

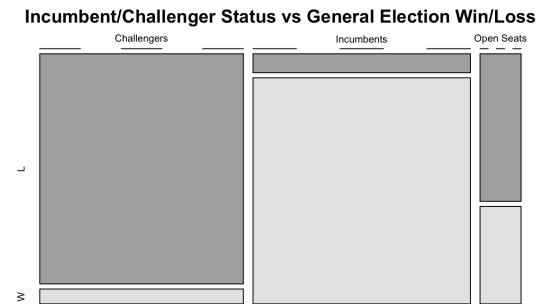
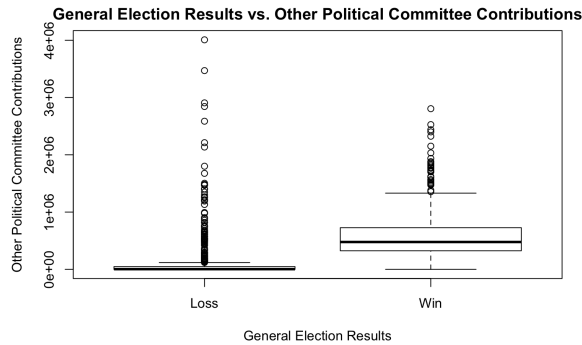
II. Data and Exploratory Analysis

a. Data and Variables

For our analysis, we used a Federal Election Commission dataset for campaign finance by candidate from 2005 to 2006, which includes 2410 cases, each case representing a candidate who raised or spent money during the period (regardless of whether they ran during the election). After sub-setting to only include candidates who ran in a general election that year, 935 cases remained. Some limitations include that the data are dependent on summary information reported by the campaign itself to the Federal Election Commission. The timeliness of the data may vary, as the period of time covered may be different for candidates in different states, depending on the dates of their primaries. We used data from 2005-2006 because it was the most recent set publicly on file from the FEC that included key variables relevant to our research question. While elections and voting trends may have changed slightly in the years since, it is still valuable to consider how various factors affected past election results. The seven predictors that we initially investigated were incumbent/challenger status, political party, total receipts, total disbursements, total individual contributions, total candidate contributions, and contributions from other political committees. The first two are categorical, while the other five are quantitative variables.

b. Exploratory Data Analysis

In the exploratory data analysis, our most notable findings were a strong relationship between incumbent/challenger status (CAND_ICI) and general election results, and the strong right skew of all quantitative predictors. In regard to the former, this informs us that CAND_ICI is likely to remain in our model. The quantitative variable that displayed the strongest relationship with election results was contributions from other political committees—it seems the more money a candidate received from PACs, they more likely they are to win. The right skew of the quantitative variables suggests that we may have to transform the predictors to satisfy all model assumptions (see figures below).



III. Model and Results

a. Analytic Methods

In order to predict the outcomes of the binary predictor—general election win or loss—we used a multiple logistic regression model. After exploring the linearity assumption using empirical logit plots for all the quantitative predictors, we noticed that all the plots displayed a concave downward curve; to ameliorate this pattern, we transformed all the x variables progressively down Tukey’s ladder of powers, finding that $x^{1/3}$ was the optimal transformation for straightening the plots. This was the only transformation we applied, as all other assumptions were reasonably satisfied. Next, we applied automated selection processes (forward, backward, and stepwise selection) and found all three methods eliminated the candidate contributions variable. Since the additional variation explained by this predictor did not outweigh the complexity it added to the model in terms of minimizing the AIC, we saw fit to eliminate it.

b. Final Model and Results

Our final model is as follows:

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = 0.06448 + 0.7114(CAND.ICII) + 0.2642(CAND.ICIO) + 0.00292(OTHER.POL.CMTE.CONTRIB.TRANS) - 0.0996(PTY.CD2) - 0.0907(PTY.CD3) - 0.00837(TTL.DISB.TRANS) + 0.007125(TTL.RECEIPTS.TRANS)$$

Predictor/Coefficient	Slope Estimate	P-value	CI for slope (e^{bj})
Incumbent (as opposed to challenger)	0.7144	<2e-16	1.914191 to 2.148543
Open seat (as opposed to challenger)	0.2642	3.53e-14	1.218590 to 1.394152
Other political committee contributions (transformed)	0.00292	5.30e-7	1.001469 to 1.003856
Republican (as opposed to Democrat)	- 0.0996	1.21e-7	-0.137161 to -0.063818
Independent (as opposed to Democrat)	- 0.0907	0.00658	-0.157376 to -0.026669
Total disbursements (transformed)	- 0.00837	3.90e-7	0.988495 to 0.994867
Total receipts (transformed)	0.007125	6.77e-5	1.003917 to 1.011121

The final model from the automated selection included the transformed predictor total individual contributions. However, given the p-value of 0.15018 in a Wald Test for the slope of this

predictor, we fail to reject that total individual contributions is a useful predictor for general election win given all other predictors in the model.

The model is appropriate because the logistic regression assumptions are satisfied: the result of each election, a candidate's win or loss, is not predetermined and therefore random; the success of each campaign is not reliant on the success of other campaigns (except for ones in the same race); and linearity was improved after the transformation of the quantitative predictors. When we made pairwise plots of the predictors, we discovered some correlations, for instance between total receipts and total disbursements, so some predictors could be less effective given others in the model due to multicollinearity.

The model is also appropriate because when we conducted a likelihood ratio test, we obtained a very large G-value of 167.8 that in return produced a p-value of almost 0, allowing us to reject the null hypothesis and conclude that at least one predictor is useful in the model. Although we found that categorical predictors such as incumbent challenger status and political party were generally more effective at predicting odds of winning than financial variables, contributions from other political committees, mainly political action committees (PACs), was the most significant predictor related to campaign finance. Though the confidence interval for the slope of this variable suggests a 1.001469-fold to 1.003856-fold increase in the odds of a candidate winning the 2006 general election given all other predictors, this change is associated with donations in the dollar^{1/3} unit, which is less than a typical PAC donation (usually in thousands of dollars). For instance, we used our model to predict that the chance of winning for a hypothetical candidate with no PAC donations was about 4%. When we kept all other information consistent and added 300K (not an unreasonable value) of PAC donations, the chances of our hypothetical candidate's win increased to about 24%.

IV. Conclusion

The objective of our research question was to investigate the existence of a relationship between the financial resources of a political campaign and its likelihood of winning the 2006 election, given the presence of other factors such as political party and incumbency. The results of the logistic regression model showed that there is indeed such a relationship—specifically, a candidate's odds of winning in the 2006 general election increased most significantly if they were an incumbent or running for an open seat (as opposed to being a challenger), if they were a Democrat, or if they were financially supported by political action committees. There are several reasons for this: incumbents are more often reelected because they have name recognition among their constituents, which aids in fundraising and contributes to their electability overall. In 2006, there was a Democratic gain in Congress, which can be explained by the presence of President Bush in the White House, given that during midterms the party not represented by the president tends to push back.

Some limitations of this study include issues of dependence on a small scale—within an individual election, only one candidate can win, and the others must all lose. However, considering the scale of all elections across the nation, none of whom influence each other, this can mostly be ignored. Additionally, our data only captured one year of funding for the campaign and only general elections, and since political trends change year to year, it may not be reliable to extrapolate by predicting results for other years or primaries. Finally, immeasurable quantities such as a candidate's personality, qualifications, and likability were not accounted for by our model.

For future study, since our data was from 2005-2006, it would be interesting to consider whether campaign finances have become more statistically significant in determining congressional election results after the 2010 Citizens United ruling.

References

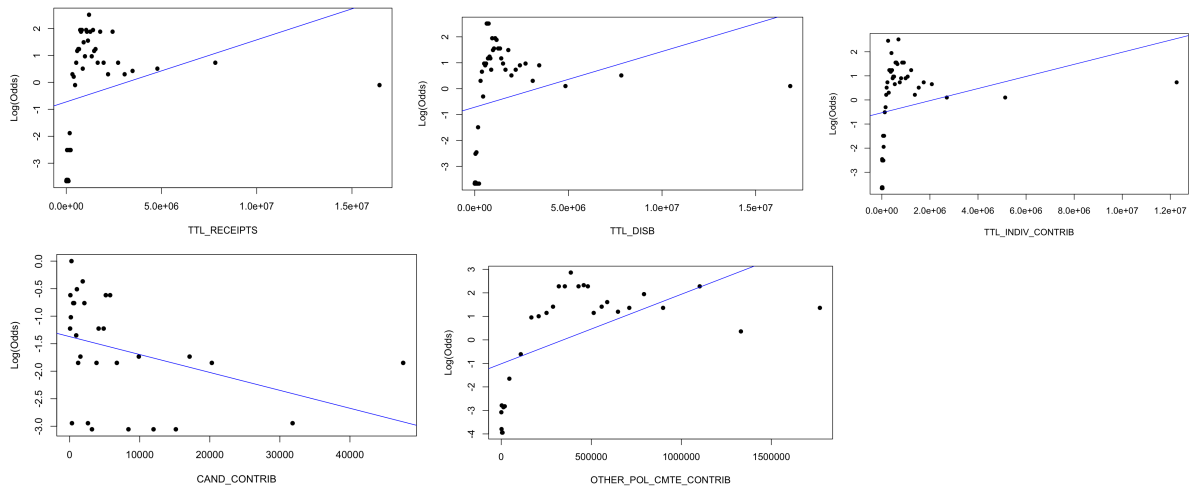
1. "Campaign Finance Data, All Candidates, 2005-2006" Federal Election Commission, <https://www.fec.gov/data/browse-data/?tab=bulk-data>.
2. Mann, Thomas E. "[Linking Knowledge and Action: Political Science and Campaign Finance Reform](#)." *Perspectives on Politics* Vol. 1, No. 1. American Political Science Association, 2003, pp. 69-83.
3. "Citizens United v. Federal Election Commission." Oyez, www.oyez.org/cases/2008/08-205. Accessed 4 Dec. 2019.

Appendix

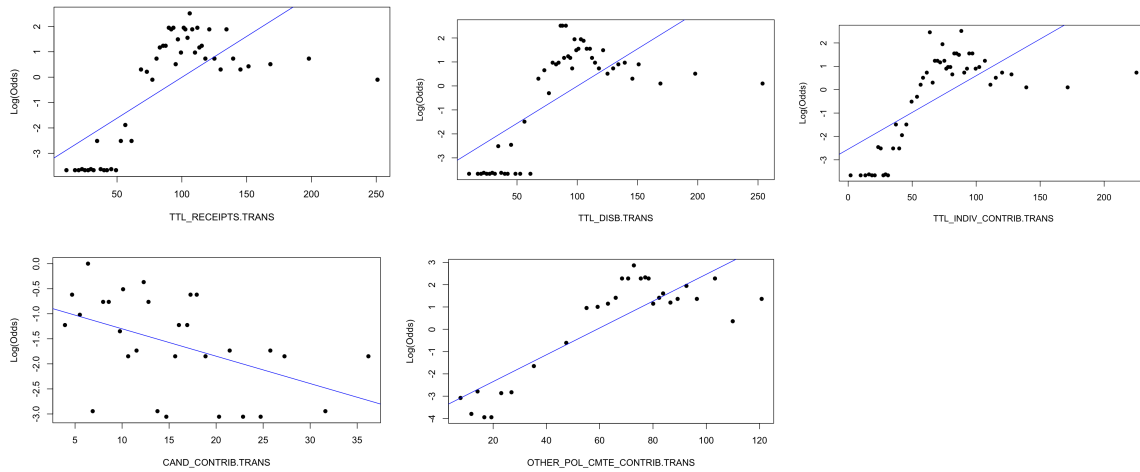
Full description of variables:

Variable Name	Description
CAND_ICI	Incumbent challenger status (categorical): I (for incumbent), C (for challenger), or O (open seat where no incumbent sought reelection)
PTY_CD	Party code (categorical): 1 (Democrats), 2 (Republicans), or 3 (Independent)
TTL_RECEPTS	Total receipts (quantitative, measured in \$): total amount of money raised by the campaign
TTL_DISB	Total disbursements (quantitative, measured in \$): total expenditures by the campaign committee (expenses such as advertising, travel, fundraising, etc.)
TTL_INDIV_CONTRIB	Total individual contributions (quantitative, measured in \$): sum of the contributions from all individual donors (not committees)
CAND_CONTRIB	Contributions from candidate (quantitative, measured in \$): contributions from the candidates to their own campaign
OTHER_POL_CMT E_CONTRIB	Contributions from other political committees (quantitative, measured in \$): mostly contributions from PACs, also includes contributions from other candidates
GEN_ELECTION	General election status (categorical): W (win) or L (loss)

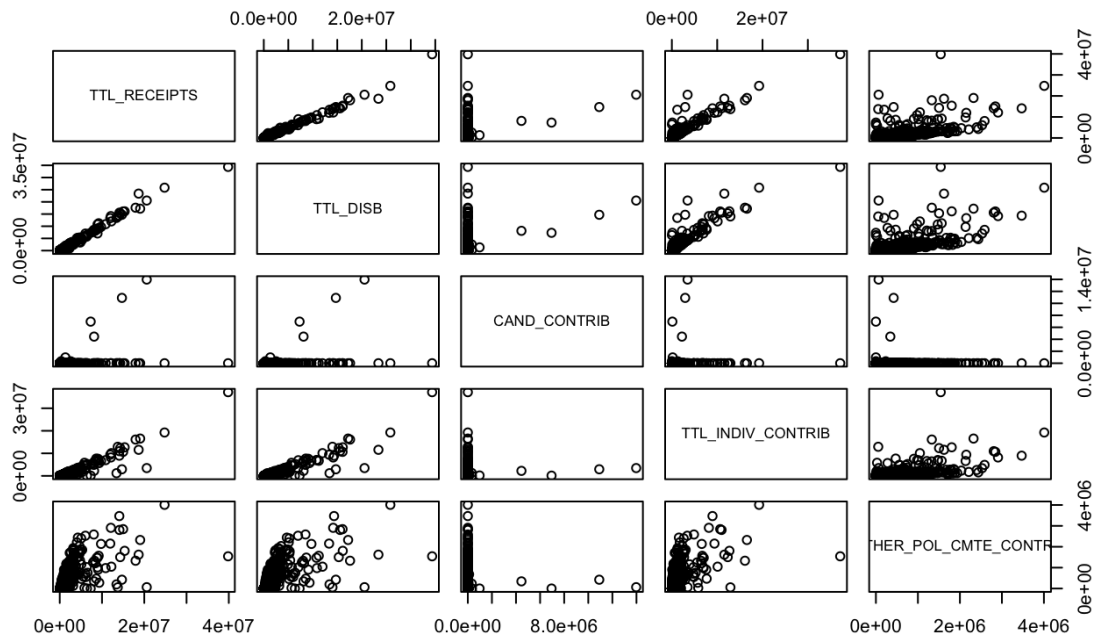
Initial Empirical Logit Plots:



Transformed Empirical Logit Plots:



Pairwise scatterplots:



Initial Model with all predictors:

Call:

```
glm(formula = GEN_ELECTION_WIN ~ CAND_ICI + PTY_CD + TTL_RECEIPTS.TRANS +
  TTL_DISB.TRANS + TTL_INDIV_CONTRIB.TRANS + CAND_CONTRIB.TRANS +
  OTHER_POL_CMTE_CONTRIB.TRANS, data = campaign.finance.new)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.01526	-0.09181	0.02334	0.09219	0.81274

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.0664651	0.0220484	3.015	0.00264 **
CAND_ICII	0.7070425	0.0294634	23.997	< 2e-16 ***
CAND_ICIO	0.2649902	0.0343354	7.718	3.06e-14 ***
PTY_CD2	-0.1004895	0.0187103	-5.371	9.92e-08 ***
PTY_CD3	-0.0920229	0.0333444	-2.760	0.00590 **
TTL_RECEIPTS.TRANS	0.0074841	0.0018241	4.103	4.44e-05 ***
TTL_DISB.TRANS	-0.0083586	0.0016391	-5.099	4.13e-07 ***
TTL_INDIV_CONTRIB.TRANS	0.0007119	0.0007588	0.938	0.34836
CAND_CONTRIB.TRANS	-0.0005989	0.0006620	-0.905	0.36590
OTHER_POL_CMTE_CONTRIB.TRANS	0.0027562	0.0006070	4.541	6.35e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.07070668)

Null deviance: 233.737 on 934 degrees of freedom
 Residual deviance: 65.404 on 925 degrees of freedom
 AIC: 188.34

Final Model Summary:

```
Call:
glm(formula = GEN_ELECTION ~ CAND_ICI + OTHER_POL_CMTE_CONTRIB.trans +
    PTY_CD + TTL_DISB.trans + TTL_RECEIPTS.trans + TTL_INDIV_CONTRIB.trans,
    data = campaign.finance.new)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.01170	-0.09064	0.02502	0.09009	0.81926

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.0644824	0.0219370	2.939	0.00337	**
CAND_ICII	0.7114171	0.0290610	24.480	< 2e-16	***
CAND_ICIO	0.2642229	0.0343215	7.698	3.53e-14	***
OTHER_POL_CMTE_CONTRIB.trans	0.0029221	0.0005786	5.051	5.30e-07	***
PTY_CD2	-0.0996585	0.0186859	-5.333	1.21e-07	***
PTY_CD3	-0.0907253	0.0333102	-2.724	0.00658	**
TTL_DISB.trans	-0.0083755	0.0016389	-5.111	3.90e-07	***
TTL_RECEIPTS.trans	0.0071254	0.0017803	4.002	6.77e-05	***
TTL_INDIV_CONTRIB.trans	0.0009952	0.0006911	1.440	0.15018	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.0706928)

Null deviance: 233.737 on 934 degrees of freedom
Residual deviance: 65.462 on 926 degrees of freedom
AIC: 187.17

Number of Fisher Scoring iterations: 2

Sample slope interpretations:

Predictor	Slope	Interpretation (in 2006 Election)
CAND_ICII	0.7144	On average, we predict a 2.0429-fold increase in the odds of winning the election associated with being an incumbent as opposed to being a challenger, given all other predictors.
CAND_ICIO	0.2642	On average, we predict a 1.3024-fold increase in the odds of winning the election associated with running for an open seat as opposed to being a challenger, given all other predictors.
OTHER_POL_CMTE_CONT_RIB.TRANS	0.00282	On average, we predict a 1.0028-fold increase in the odds of winning the election associated with each additional dollar ^{1/3} contribution from political committees to a candidate's campaign, given all other predictors.