A Look at the 2012 Presidential Election

Eric Hare & Andee Kaplan

Dec. 12, 2012

Abstract

Motivated by the 2010 Citizens United ruling and the subsequent birth of "Super PACs", this paper uses independent expenditures data from the Federal Elections Commission, in conjunction with presidential polling data to analyze the 2012 presidential campaign. Using R, and the packages XML, ggplot2, plyr, lubridate, knitr, we scrape data from these sources and analyze them in order to highlight interesting trends in campaign spending. Furthermore, we correlate these trends in spending over time to the changes in the polls. Ultimately, we did not find a clear and direct relationship between increases in spending and changes in public support. However, our analysis does provide evidence for some commonly held views of the candidate's spending habits and geographical areas of strength and weakness.

1 Introduction

In 2010, the United States Supreme Court released its decision on Citizens United v. Federal Election Commission. The decision found that it is a violation of the constitutional right to free speech for limits to be placed on money spent by corporations and labor unions to support or oppose political candidates. For the first time since the passage of the Tillman Act in 1907, corporations and unions could spend unlimited amounts of their own money on the presidential election. Such power led to the coining of the term Super PAC, as well as a cloud of uncertainty as to what effect such money could have on the election outcome.

In this report, we analyze the independent expenditures data set from the Federal Election Commission in an attempt to determine the effect of this spending on the 2012 election. We also retrieve data from the NationalPolls.com database of polling results to quantify the effect of this spending. The results of our analysis suggest that the spending may have had a measurable impact on public opinion, but external factors, such as the presidential debates, had much stronger impacts.

```
## Error:  cannot open the connection
```

2 Data

We are primarily working with two data sources to complete our analysis: independent expenditures (PAC) data and polling data. The PAC spending data comes from the Federal Election Commision and the polling data comes from `http://nationalpolls.com/`.

2.1 Variable Definitions

The variables present in the PAC spending data are defined as follows:

Tag	Field Name	Data Type	Description
can_id	Candidate ID	Character	unique ID of whom the ex
spe_id	Spender ID	Character	Unique ID o group makin
spe_nam	Spender Name	Character	Name of com making expe
ele_typ	Election Type	Character	code for spec penditure wa
can_off_sta	Candidate State	Character	Postal state a date
can_off_dis	Candidate District	Number	District num
can_off	Office	Character	Office Sought
can_par_aff	Party	Character	Party abbrev
exp_amo	Expenditure Amount	Currency	Dollar amou

Error in format(min(spend.data$exp_amo, na.rm = TRUE), nsmall = 2, scientific = FALSE) :
object 'spend.data' not found Max: $
Error in format(max(spend.data$exp_amo, na.rm = TRUE), nsmall = 2, scientific = FALSE) :
object 'spend.data' not found Mean: $
Error in mean(spend.data$exp_amo, na.rm = TRUE) :
object 'spend.data' not found Median: $
Error in median(spend.data$exp_amo, na.rm = TRUE) :

object 'spend.data' not found

exp_dat	Expenditure date	Date	Date of spe D/YYYY
agg_amo	Aggregate amount	Currency	Total amoun endar year, sought

Error in format(min(spend.data$agg_amo, na.rm = TRUE), nsmall = 2, scientific = FALSE) :
object 'spend.data' not found Max: $
Error in format(max(spend.data$agg_amo, na.rm = TRUE), nsmall = 2, scientific = FALSE) :
object 'spend.data' not found Mean: $
Error in mean(spend.data$agg_amo, na.rm = TRUE) :
object 'spend.data' not found Median: $
Error in median(spend.data$agg_amo, na.rm = TRUE) :
object 'spend.data' not found

sup_opp	Support or Oppose	Character	S=Support,
pur	Purpose of expenditure	Character	description o vision or radi
pay	name of payee	Character	Name of the entity receivi
file_num	Filing number	Number	Unique iden (which may ments)
amn_ind	Amendment Indicator	Character	New report c
tra_id	Transaction ID	Character	Unique iden (unique with
ima_num	Image number	Number	Image locatio action appea
rec_dt	Filing receipt date	Date	Date on whic ted to FEC M

prev_file_num	Previous filing number	Number	Reference to

The variables present in the polling data are defined as follows:

Tag	Field Name	Data Type	Description
Pollster	Polling Company	Character	Company that conducted the poll.
State/US	State	Character	State poll was conducted of. If national poll, then value is "National".
Date	Poll Date	Date	Range of dates that the poll was being conducted.
Obama	Support for Mr. Obama	Number	Integer rounded percent of support in the poll.
Romney	Support for Mr. Romney	Number	Integer rounded percent of support in the poll.

2.2 Cleaning

Most of the data cleaning had to be performed on the PAC spending data. We made sure all the columns were formatted correctly in regards to data type (dates are date, numbers are numeric, etc.) before cleaning and reformatting the data more extensively.

One challenge we faced was that the purpose of independent expenditures column is a free text field on the FEC reporting form. For example, the Conservative Super PAC Americans for Prosperity tended to use verbose descriptions of the spending purpose. An entry from mid-August read "oppose advertising-tv production (voted for)." This description lists whether it was an add in support or opposition, that it was a television ad, and the name of the ad. In contrast, the anti-abortion Women Speak Out PAC prefered short descriptions, such as "ads" for their early-October expense. The result was that when trying to explore what PACs spent the majority of their money on, we were unable to group expenditures together. To solve this issue we searched for matches to patterns in the purpose field. We chose these patters by looking at the expenditure purposes and manually finding common threads among the purposes. From these patterns we were able to create buckets that each expense fell in to, as well as high level buckets that more generally classified the expenditures. The buckets we chose are:

- Ads - Advertisement spending, including television, radio, and online
- Direct Contact - Direct voter contact, such as canvassing
- Salary - Payments made to staff of the organization
- Swag - Clothing, signs, and other promotional material
- Transport - Transportation costs, such as taxis or van rentals

- Other - All expenses that do not fit into the above categories

An additional complication we faced was with the support/oppose column. This column in conjunction with the candidate column are used to indicate which candidate benefits from the expenditure. An example would be if the support/oppose column is oppose and the candidate name is Romney, then Mr. Obama benefits because the money is being spent to "oppose Romney". Likewise, if the support/oppose column equals support and the candidate name is Romney, then Mr. Romney benefits. We solved this by adding a new field which simply stores the candidate that 'benefits' from the particular expense. That is, an opposition advertisement expense entered with a candidate name of Obama will benefit Romney.

The polling data did not require very much data cleanup. We needed to split the date range and format the second date as a date for our use. We also removed some differences in state naming by the different pollsters. One complication we faced is that a fairly significant number of polls were excluded from the nationalpolls.com database. Some new national polling outfits, such as the RAND Corporation, were excluded, as were a large number of online polls, including Google Consumer Surveys. We were unable to find a methodology for the inclusion/exclusion of certain polls, but we suspect that polls without a track record in previous presidential elections were strong candidates for exclusion. As it turns out however, online surveys performed more strongly in the 2012 election than did automated phone surveys or live interviewer surveys according to Nate Silver. [1]

3 Findings

Our key findings are summarized in the sections below.

3.1 Type of Spending by Week

This plot displays the spending in each of the six categories by independent organizations benefiting Barack Obama and Mitt Romney for each week since April 25, 2012. The spending amounts are split by whether the goal was to support or oppose the candidate. There are three markers which signify major events in the campaign: the selection of Paul Ryan as Mitt Romney's vice presidential nominee, the Republican National Convention, the Democratic National Convention, the 47% video and the debates. Interesting trends can be observed in advertisement spending, and direct contact. Ads in support of both candidates have increased since the end of July. However, ads in opposition of the candidates were airing well before. You can, however, see an increase in negative ads benefiting Romney. There has also been a noticeable increase in spending on direct contact for Obama, while the Romney organizations have maintained consistent spending in negative direct contact for quite some time.

[1] http://fivethirtyeight.blogs.nytimes.com/2012/11/10/which-polls-fared-best-and-worst-in-the-2012-presidential-race/ff

```
## Error:  object 'num.weeks' not found
```

3.2 Spending Categories

These plots shows the comparison of amount spent for each candidate by categories of independent expenditure, since April 25, 2012. The categories are mined from the self-reported text description of the expense. The left plot shows the spending to benefit Mr. Obama against the spending to benefit Mr. Romney by category, with values to the right of the reflection line showing PACs supporting Mr. Romney are outspending PACs supporting Mr. Obama on these categories, and vice verse for those to the left. The right plot shows bar chart comparisons of the spending in main categories, on a log scale. Organizations supporting Romney are greatly outspent those supporting Obama, and particularly in direct contact and advertisements. On the other hand, organizations supporting Obama only outspent those supporting Romney in salary.

```
## Error:  object 'sum_exp2' not found

## Error:  object 'sum_exp2_p' not found
```

3.3 Spending by Independent Organizations

This chart displays the total spending by the top independent organizations split by candidate. The cumulative amounts spent is displayed vertically, by the benefiting candidate. The organizations supporting Romney have spent significantly more than those for Obama. In fact, two Romney Super PACs (Restore our Future, Inc and Americans for Prosperity) have spent more than all the organizations supporting Obama combined.

```
## Error:  object 'sum_exp.spep' not found
```

3.4 Swing State Trends

This plot displays the change in polling support for both Romney and Obama over time, in 12 swing states. The data includes all polls in the National Polls database from April 25th to now. Vertical lines indicate difference between the two candidates, and color shows who is ahead. There are five markers which signify major events in the campaign: the selection of Paul Ryan as Mitt Romney's vice presidential nominee, the Republican National Convention, the Democratic National Convention, the 47% video and the first presidential debate. The plots suggest that the reaction to these major campaign events was not consistent state to state. Some states, such as Wisconsin, have produced drastic changes in polling support over time ("bounces"), while others, such as North Carolina, have maintained a consistent margin of support for the two candidates.

```
## Error:  object 'polls.data' not found
```

3.5 Spending By Week

Looking at weekly spending by PACs benefiting Romney, we can see a sharp increase in spending that occurred during the week of July 18th. We look at the national and swing-state polls to determine if there is any noticeable effect due to these changes in spending.

```
## Error:  object 'num.weeks.sum' not found
```

3.6 Effect of Spending on Polls

This plot shows the polling margin (Obama - Romney) over time, colored by swing states versus the national polls. It can be seen that Mr. Obama consistently maintained an advantage in swing states relative to his national numbers. A marker is placed on July 18th, corresponding to the sharp increase in spending by Romney-supporting PACs. It does not seem that this spending increase had a measurable effect on the overall trend in the polls at this time.

```
## Error:  object 'polls.data' not found
```

This plot compares the difference in spending by week by the PACs to the difference in polling between Mitt Romney and Obama with a one week lag. The points are colored by week and certain weeks are labeled to indicated important events. The goal is to see if there is a relationship between PAC spending and poll results. In both Obama-supporting and Romney-supporting PACs we can see a weak positive relationship between spending and polling.

```
## Error:  object 'final.df3' not found

## Error:  object 'final.df3' not found
```

4 Conclusions/Future Work

Ultimately, our findings suggest that while election spending itself was a very significant part of the 2012 elections, it is not clear whether the existence of Super PACs had a measurable impact on the outcome. Instead, the events that garnered large amounts of media coverage, such as the 47% video and the first presidential debate, seemed to have the most noticeable effect on the polling.Furthermore, our analysis did not take into account the spending done by the candidates themselves. It may be that such spending muted the effect of Super PAC spending.

Our analysis also makes clear that Mr. Obama's polling held up more strongly in the swing states than in the national polling throughout the campaign. This advantage was evident in the election results, as well. According to the current AP vote tally as of November 16th, Mr. Obama won the overall popular vote by 2.8%. But he won Colorado, the state that put him over the 270 electoral votes needed for victory, by 4.7%. This suggests that Mr. Obama could have lost the popular vote by nearly 2% and still have been elected president of the United States.

An extension of this project should take into account spending done by the candidates themselves, in addition to independent expenditures. This would allow for a more informed look at the relative spending difference, and how it may or may no correlate with changes in the polling averages. Another area of exploration would be the purpose field of the independent expenditures data set. Our analysis attempted to categorize spending entries into broad categories, but many of the entries in the database have more specific information. For instance, an ad buy may list the name of the ad, as well as the state the ad is running in. It may also be possible to incorporate another data source in order to link the expenditure entries with ad buys in particular states.