MODELING THE 2012 PRESIDENTIAL ELECTION

By
Christopher P. Alexander
Ethan Krohn
Selman Kaldiroglu
Vanessa Moreno

1. Introduction:

This research project aims at predicting the November, 2012 Presidential Election outcomes in Ohio, Florida, Georgia, Pennsylvania, and North Carolina. In analyzing the progression of the race and eventually predicting the winner, certain demographics were conjectured to be key factors. Peoples ideologies and choices are tremendously influenced by their demographics. This model conjectures that by recording game changing demographics such as, race, age, and gender, the outcome of the election in the state level could be predicted. Additionally, this project aims to show how certain segments of the population alter the outcome of the election.

By looking at surveys starting from 2010, we recorded these demographics opinions towards the candidates (Mitt Romney and Barack Obama). After observing these demographics with respect to time, we obtained a trend for each candidate and for those remaining undecided.

For the final step, regressions were run to model the trends and assemble a prediction regression for the 2012 November Presidential Election between Barack Obama and Mitt Romney. We hoped to find a consistent and accurate method that can be generalized across states and despite data inaccuracies. We found that the 3rd degree consistently predicts the election results.

1.1. **Discussion of Literature.**

1.1.1. Bread and Peace

The article, *"Obama's Re-election Prospects Under Bread and Peace Voting in the 2012 Presidential Election,"* uses a different mathematical model to estimate the percent of the vote that Obama will receive in the election (Hibbs). The model presented in this article is based on just two factors, the weighted average growth of disposable personal income and the cumulative fatalities of in the U. S. military in unprovoked, hostile, foreign wars. The author refers to these variables as bread and peace respectively, and it is from these labels that the model gets its name. Because it is restricted to two variables this model cannot take into account transient variables that are specific to a certain election, such as Obama's race or Romney's religion. The advantage of this model is that it can be computed for any U. S. presidential election with two easily found statistics. In the 2012 election this model predicted that Obama would lose with 47.5 percent of the vote. This model has been proved incorrect by

the election results. The article was provided to us by our sponsor, Dr. Suarez. Our model varies tremendously from this one. It relies on polling information to find how voting preferences for different demographics change over time. The advantage that our approach has compared to this article is that ours takes into account the transient issues which the bread and peace model ignores. This model is useful as a good example of how others are modeling the coming election, even if its method does not have much in common with ours.

1.1.2. A Mathematical Model of Political Affiliation

Our original model was based on an unpublished model written in 2007, named *A Mathematical Model of Political Affiliation*. This article was beneficial in building our former model for the election. The model constructs five basic groups in categorizing people: fanatical republicans, fanatical democrats, moderate republicans, moderate democrats, and the susceptible group. The article uses various interpersonal factors and external factors that affect ones political affiliation. We were interested in developing a similar model that would have considered four key demographics as starting points and then integrating interactions with the media and other groups as a rate of change of ones political beliefs. However, the article has its drawbacks that we wanted to avoid. It assumes that everyone in a certain group is homogeneous and that peoples personal influences do not change over the course of the surveyed period. Another key element missing from the article is the fact that it fails the mass action law which creates an imbalance in the differential equations. Thirdly, the model does not account for the probability of actually voting. This is crucial to the model since even people who will not vote in the election might be responding to the surveys, therefore to increase the accuracy of our model, we wanted to integrate whether someone is likely to vote in our analysis. Since it is hard to measure someones degree of political affiliation as a slight change from this paper, we were only going to assume that there are three categories: republicans, democrats and susceptible. The base model of the article was useful in our project however, due to its main disadvantages, we planned on building our own model with significant changes.

2. Discussion of Data:

The states that our model focused on were Ohio, North Carolina, Florida, Georgia, and Pennsylvania. Ohio was considered to be the most important swing state in the election. North Carolina was important as it was the only swing state to go Republican. We chose Florida because the race was extremely tight. Georgia and Pennsylvania were selected in order to test our model on states that were not swing states. The polling data that our model used came from the `http://www.realclearpolitics.com/epolls/latest_polls/president/` which archives current and past polling data from a variety of polling companies. This site provided numerous free polls that reported how different demographics were leaning in the election. Unfortunately, polls for states other than the swing states were limited and many of the polling companies that we had to rely on were biased. For instance, one of the most available polls, Public Policy Polling (PPP) is consistently considered a democrat leaning poll. Additionally, the weights for the different demographics were taken from the United States Census Bureaus report on how demographics voted in the 2008 presidential

election. It was felt that the 2008 data would be more accurate than the more recent 2010 election simply because so many more people vote in the presidential elections.

2.1. **Issues.** Several issues regarding the regressions were encountered throughout the process. One of the main problems was the bias of surveys that were collected. Some polling companies tend to be more Democratic whereas others tend to be more Republican-leaning. In order to reduce this issue, numerous polling companies were included. We hoped that the idiosyncratic biases of each polling company would offset. Secondly, some of the states that were chosen to be modeled do not have enough number of surveys to have a confident result. A low number of surveys increases the variance of the model for those states. The methodology of the model resulted in difficulties when trying to predict far into the future. Higher level polynomial regressions tend to shoot to infinity in cases where the election date was far away from the last survey collected.

One might think that Income or Education Level would also be key factors in determining the race between Barack Obama and Mitt Romney. These demographics were considered however due to unavailability of data and a problem associated with the cutoffs of the categories prevented them from being used to be accurately modeled into the regression. In particular for income, the polling data we found for income normally divided income groups differently: one poll would cutoff at 50,000 another at 35,000 and 75,000. The demographic Age was not affected as much by this issue due to the consensus among the polling companies to have similar cutoff points.

The last issue with the model is the weighting of each demographic. In our model, the regressions for each demographic is multiplied by the percentage of that demographic in the voting population. These weights were taken from the 2008 election results which created an issue of population and cultural change. The values that were used do not accurately portray the number of people who voted due to the insufficient information to determine the changes within the population in those states from 2008 to 2012.

3. Methodology

Dr. Suarez provided articles that discussed voters, the relevant issues in the 2012 presidential election, and previous models used to predict elections. *A Mathematical Modeling of Political Affiliations* was part of the set that Dr. Suarez recommended. The paper, presented by California State Polytechnic University, Pomona and Loyola Marymount University, explored the factors that play a role in a voters political affiliation. The model presented in this paper separated voters into five categories- Susceptible, Moderate Democrat, Moderate Republican, Fanatical Democrats, and Fanatical Republicans- and tracked the movement of voters within these groups using a dynamical system. Based on the research we did, we decided to create a dynamical system that would compute the movement of voters, whose affiliation was dependent on their demographics, between parties over the time. We condensed our political affiliations to a group of three: Democrat, Republican, and Susceptible.

For our model, we needed information of potential voters that was broken down to political affiliations and demographics. The site `http://www.realclearpolitics.com/epolls/latest_polls/president/` collected polls from several polling companies that questioned potential voters across the country. After taking a look at the

information available for free, we realized our preliminary model would not be compatible with the data we could collect. The data from polls was given in percentages of potential voters surveyed at that moment and focused on a potential voters choice of candidate. For this reason, a dynamical system was not a good candidate to model our data. We decided to change our groups to Pro-Obama, Pro-Romney, and Undecided supporters to reflect our data. The new model would have to focus on the changes of each group independently over time, rather than the flow within groups. Ohio was considered a swing state, and thus a critical state for the 2012 presidential election. As a first step, we collected data from Ohio in order to predict the outcome of the election. We collected polls dating back to March 2011 up to days prior to the election. Time was calculated starting from the first poll. Change in time was decided according to the number of days since the first poll. The demographics we focused included gender, race, and age and we tracked how the percentage of each breakdown of demographic would vote: that being either Pro-Obama, Pro-Romney, or Undecided. The data we collected was placed in an excel file. The excel file visually separated our data. We took number of days since the first poll to be the independent variable and the percentage of voters in a given sub-group at that particular time who voted for a particular candidate as the dependent variable. For example, 93 days after the first poll, a new poll was published for Ohio in which 51% of females claimed they would vote in favor of Obama if the elections were held on that day. The sub-group in the example would then be, females Pro-Obama. We input this information into a Matlab file to run regressions on each sub-group. We ran the file to compute the regressions, and then we multiplied our answer by the weight of that sub-group. The weight of each subgroup was decided based on the 2008 Census Data for Ohio. Our next step was to add up each subgroup in favor of the same candidate and divide by three, since we had a total of three demographic groups. Using this method, we predicted the values for the time of the election.

4. Results

Once the regressions were run the results were compared to what happened in the actual election. The election results only reported the percent of the vote that each candidate won, while the model predicted the percentage of likely voters that would vote for each candidate as well as the percentage that would remain undecided and either ended up not voting or voted for a third party candidate. In order to obtain results that could be compared, the percentage that we predicted each candidate would win by was divided by the combined percentage that we predicted either Obama or Romney would win. Then the same was done to the actual election results in order to remove any percentage that went to a third party candidate. This gave the percentage of the votes for the two main candidates th at went to either Obama or Romney. At this point, the model had results for how each regression predicted the race would go in each of the five states. Next, each regression was ranked in order of how well they performed for each state. The regression that predicted a result closest to the actual outcome of the election would be ranked number one in that state and the worst regression would be ranked number ten. The mean and median values for the rankings that each regression had received across the five states were found. The mean difference between that the predicted results from each regression

FIGURE 4.1. The Ranking of the Polynomials

Order	Avg. Rank	Median Rank	Avg. Error
1	4	4	1.01%
2	3.8	3	0.99%
3	2.2	2	0.95%
4	4	5	1.32%
5	6	6	1.84%
6	5.2	4	2.09%
7	7.2	7	2.11%
8	7.4	8	2.67%
9	6	9	1.28%
10	8.8	10	13.60%

and the actual results were also calculated. The third level polynomial had a mean ranking of 2.2, a median ranking of 2, and an average error of 0.95 percent. This tells us that over the five states the third level polynomial consistently performed well.

5. Conclusion:

This method of predicting the 2012 Presidential Election aims at taking freely available data and turning them into accurate election results. Additionally, this method considers the effects of certain sections of the voting population on the results of an election.

We hoped to find a method that could be generalized to future elections: And in some sense, we found a good starting spot. The third degree polynomial accurately predicted the percentage outcome in each state, and was consistently a better predictor than any of the other polynomial fits. Moving forward, we hope to show that this method would have predicted well in previous Presidential Elections, and for states beyond our limited sample.

By focusing on race, gender, and age we concentrate on some of the most significant explanatory variables in any election. Each of these groups have had different experiences and that causes each of these groups to have very different expectations for the President. Racism, misogyny, and ageism all have real political ramifications in regards to policy development: affirmative-action, welfare reform, and unemployment concerns. This information on population composition by demographics is also commonly available, allowing ourselves an opportunity to use this method across time periods, as long as we account for changes in population make-up.

This study will also serve as a good platform into further research in the association between population composition and the results of elections. The development

of a metric about the *blueness* or *redness* of a state, as it pertains to its demographics, is one such possibility. This coupled with census projections about population composition mean, *ceteris paribus*, we can find the point Texas turns Democrat.

The simplicity of our method is certainly one of its advantages. The method works (for our limited sample) despite all of the imperfections inherent to the data. This is ideal. We will never be able to have perfect information about the electorate, nor will we ever have perfectly unbiased data. But despite all of that our third degree polynomial provided surprisingly accurate predictions. Most of the other models we examined, relied exclusively on issues. However, they over-rely on economic issues, as those are the only data that are available in numeric form and freely available. A lot of these models, thus, projected Romney to win, as the economy has struggled under Obama. Our method, on the other hand, implicitly accounts for economic and social views, as well as personality characteristics of each candidate.

6. Bibliography

References

[1] Hibbs, Douglas A, Obamas Re-election Prospects Under Bread and Peace Voting in the 2012 US Presidential Election. PS: Political Science and Politics 31 July 2012. Online.
[2] Rasmussen Reports, LLC. Pulse Opinion Research. Survey. May, 2011 - November, 2012.
[3] Gravis Marketing, Inc. Florida. Survey. June, 2011 - November, 2012
[4] Public Policy Polling. Raleigh, North Carolina. Survey. March, 2011 - November, 2012
[5] University of Cincinnati. The Ohio Poll. Survey. January, 2012 - November, 2012.
[6] SurveyUSA. Survey. October, 2011 - November, 2012.
[7] Fox News Poll. Anderson Robbins Research. Survey. October, 2011 - November, 2012.
[8] he Huffington Post. Huffpost Politics: Election Resulst. September, 2012- November, 2012.
[9] The Purple Strategies. PurplePoll. September, 2012- November, 2012
[10] American Research Group. Survey. September, 2012- November, 2012
[11] Cable News Network. CNN/ORC Poll. October, 2012- November 2012
[12] Carol Ambrose, Jennifer Jones, Kurt Larson, Lucy Orozco,David Uminsky, Steve A. Wirkus, California State Polytechnic University, Pamona and Loyola Marymount University. A Mathematical Model of Political Affiliation. August 2007.
[13] United States Census Bureau. Voting and Registration in the Election of November 2008−Detailed Tables. October, 2012. Online.

FIGURE 7.1. Ohio With Power 3

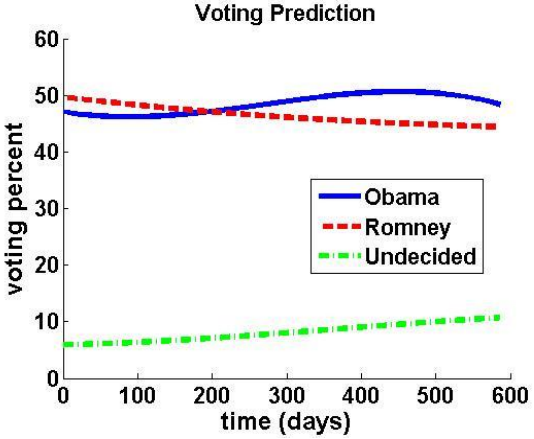


FIGURE 7.2. Florida With Power 3

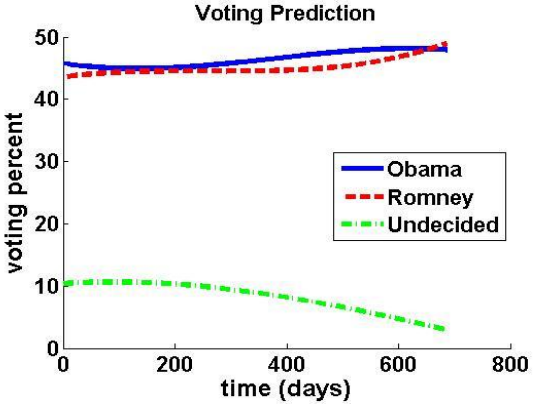
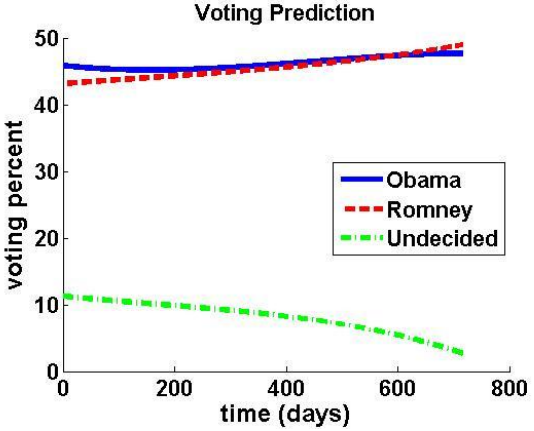


FIGURE 7.3. North Carolina With Power 3



7. ADDENDUM

FIGURE 7.4. Georgia With Power 3

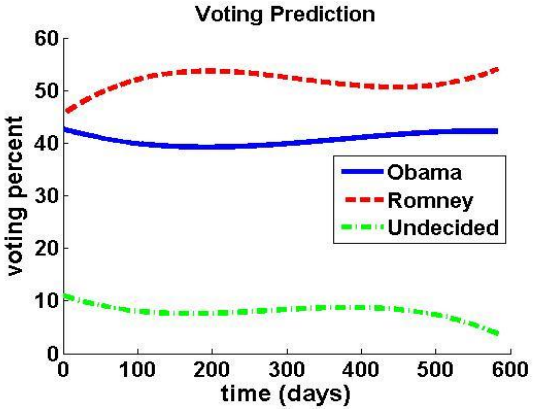


FIGURE 7.5. Pennsylvania With Power 3

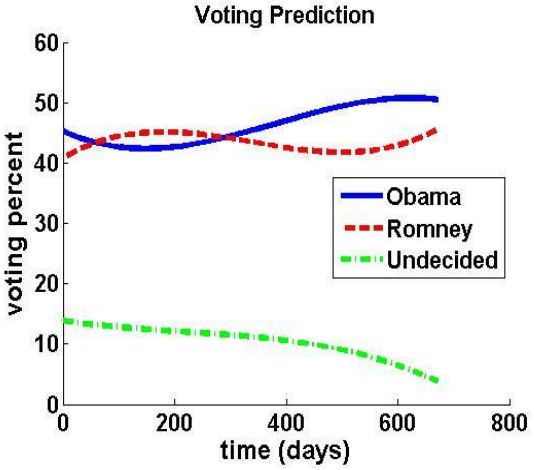


FIGURE 7.6. Ohio With Power 9

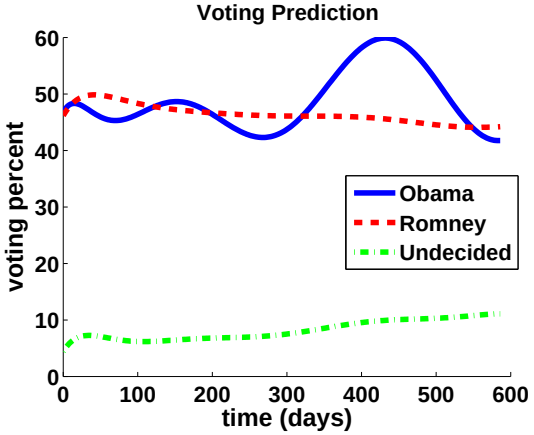


FIGURE 7.7. Florida With Power 9

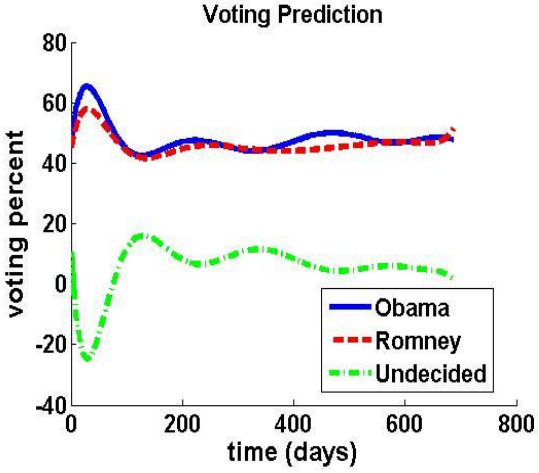


FIGURE 7.8. North Carolina With Power 9

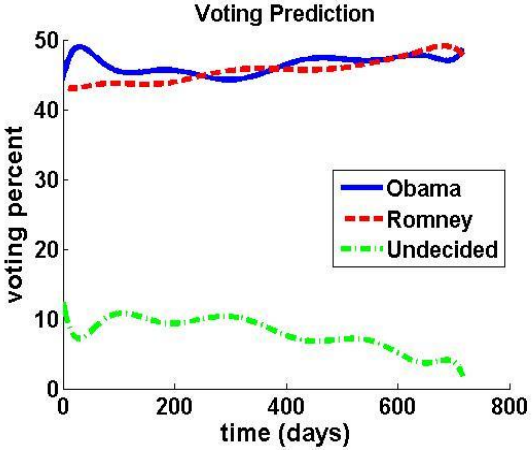


FIGURE 7.9. Georgia With Power 9

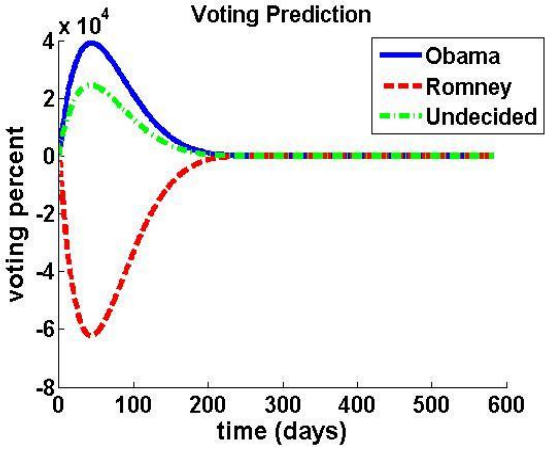


FIGURE 7.10. Pennsylvania With Power 9