SUPPLEMENTARY MATERIAL: MODELING VOTING DYNAMICS IN A TWO-PARTY SYSTEM

MATERIAL SUPLEMENTARIO: MODELAR LA DINÁMICA DE VOTACIÓN EN UN SISTEMA DE DOS PARTIDOS

T. Phan, M.G. Navas-Zuloaga, C. Ignace. C.M. Kribs,
C.W. Castillo-Garsow & V.M. Moreno

5 Data

We use two different data sets (polling data) that contain information on voting preferences from the people who self-identify as either Democrat or Republican during the 2012 and 2016 presidential elections for parameter fitting. We further cross reference the credibility of each data source with the rating published by Nate Silver [76]. All polls have a rating of $A-$ or higher, with the exception of Politico/George Washington University with a B rating.

Specifically, the first data set is compiled of weekly[1] data sets from multiple sources spanning 14 weeks from the beginning of August of 2012 to right before the election day, November 6, 2012, from [52, 54, 55, 56, 57, 59, 60, 61, 62, 63, 66, 67, 68]. All polls are nationwide polls with the exception of one statewide poll, which still fits the trend of the data. The second set is collected through a single source. In this case, the data are taken weekly[2] between May 18 to July 12 of 2016 from [69]. Descriptive statistics of the polling data are provided in Tables 5 and 6.

Table 5: Descriptive statistics for polling data of the 2012 presidential election.

	Mean	SD	Min	Q1	Med	Q3	Max
V_1	0.458	0.0263	0.412	0.441	0.464	0.481	0.498
M_1	0.0180	0.0111	0.00476	0.00906	0.0165	0.0254	0.0436
V_2	0.503	0.0217	0.458	0.491	0.504	0.521	0.540
M_2	0.0205	0.0144	0.00518	0.00948	0.0168	0.0294	0.0545

[1]We use Monday as a reference. If the data set is taken within that week, then we categorize it as data for that week.

[2]Wednesday is used as the reference similarly.

Table 6: Descriptive statistics for polling data of the 2016 presidential election.

	Mean	SD	Min	Q1	Med	Q3	Max
V_1	0.351	0.0128	0.330	0.343	0.352	0.363	0.370
M_1	0.087	0.0174	0.0579	0.0732	0.0906	0.0994	0.113
V_2	0.461	0.0189	0.437	0.448	0.456	0.474	0.497
M_2	0.101	0.00747	0.0922	0.0938	0.100	0.108	0.112

We also collected data on the number of times a candidate was mentioned on TV and headlines either favorably or unfavorably. For the 2012 election, data were taken from the studies in [65]. For 2016, the period from May to June is covered by the report in [64]. Data on the month of July are taken from [77]. (Analysis was performed in July 2016, at the end of the primary season).

6 Sensitivity

For the peer interaction model, we calculate the closed forms of the sensitivity of $V \equiv V_1(t+1) - V_2(t+1)$ with respect to each parameter between two consecutive time points [53]. Recall that V_1, V_2, M_1, M_2 are taken to be evaluated at time t.

$$\frac{a_1}{V}\frac{\partial V}{\partial a_1} = \frac{V_1^2 a_1}{V}, \qquad \frac{a_2}{V}\frac{\partial V}{\partial a_2} = \frac{V_1 M_1(1-2b_1)a_2}{V},$$

$$\frac{a_3}{V}\frac{\partial V}{\partial a_3} = \frac{V_1 M_2(2b_2 + c_1 - b_2 c_1 - 1)a_3}{V}, \qquad \frac{a_4}{V}\frac{\partial V}{\partial a_4} = \frac{V_2 V_1(b_3 - 2)a_4}{V},$$

$$\frac{a_5}{V}\frac{\partial V}{\partial a_5} = \frac{V_1 M_1 a_5}{V}, \qquad \frac{a_6}{V}\frac{\partial V}{\partial a_6} = 0,$$

$$\frac{a_7}{V}\frac{\partial V}{\partial a_7} = \frac{V_1 V_2(2 - b_4)a_7}{V}, \qquad \frac{a_8}{V}\frac{\partial V}{\partial a_8} = \frac{V_2 M_1(b_5 c_2 - 2b_5 - c_2 + 1)a_8}{V},$$

$$\frac{a_9}{V}\frac{\partial V}{\partial a_9} = \frac{V_2 M_2(2b_6 - 1)a_9}{V}, \qquad \frac{a_{10}}{V}\frac{\partial V}{\partial a_{10}} = -\frac{V_2^2 a_{10}}{V},$$

$$\frac{a_{11}}{V}\frac{\partial V}{\partial a_{11}} = 0, \qquad \frac{a_{12}}{V}\frac{\partial V}{\partial a_{12}} = -\frac{V_2 M_2 a_{12}}{V},$$

$$\frac{b_1}{V}\frac{\partial V}{\partial b_1} = \frac{2V_1 M_1(1-a_2)b_1}{V}, \qquad \frac{b_2}{V}\frac{\partial V}{\partial b_2} = -\frac{M_2 V_1(1-a_3)(2-c_1)b_2}{V},$$

$$\frac{b_3}{V}\frac{\partial V}{\partial b_3} = \frac{V_2 V_1(a_4 - 1)b_3}{V}, \qquad \frac{b_4}{V}\frac{\partial V}{\partial b_4} = \frac{V_1 V_2(1-a_7)b_4}{V},$$

$$\frac{b_5}{V}\frac{\partial V}{\partial b_5} = \frac{M_1 V_2(1-a_8)(2-c_2)b_5}{V}, \qquad \frac{b_6}{V}\frac{\partial V}{\partial b_6} = \frac{2M_2 V_2(a_9 - 1)b_6}{V},$$

$$\frac{c_1}{V}\frac{\partial V}{\partial c_1} = -\frac{M_2 V_1(1-a_3)(1-b_2)c_1}{V}, \qquad \frac{c_2}{V}\frac{\partial V}{\partial c_2} = \frac{M_1 V_2(1-a_8)(1-b_5)c_2}{V}.$$

Thus the normalized sensitivity of $V(t_E)$ is obtained by evaluating V in the last time-step.

7 A note on the sensitivity indices

As a side note, we would like to know what actions can effectively change the parameter with the highest sensitivity value. Unfortunately, this information is not directly obtainable with our models. This is due to each parameter being the net effective cause of the corresponding transition, so there is not a direct way to figure out the separate mechanisms that determine the value of the parameter. In order to suggest a method to obtain the mechanism behind each parameter, we make the following observation. If we define an extreme parameter to be parameter with value either above 0.75 or below 0.25, then the person-to-person interaction model has 20 extreme parameters in 2012 and 15 extreme parameters in 2016. We propose the following explanation for this observation.

Notice that our data for the 2012 election were first taken approximately three months before the election day and span 14 weeks, whereas our data for the 2016 election were first taken approximately six months before the election day and span 8 weeks. This difference in time periods in which the two data sets are taken may indicate the different level of uncertainty in the decisions people of the two parties make. Suppose this is not caused by the inherent differences between the two elections—which we do not know for certain. Then since there are significant difference between the two time periods, e.g., the choice of vice president, the resulting difference in the number of extreme parameters can alternatively be explained by the difference in the stages of the election. If so, this suggests that a more appropriate way to model the dynamics behind voting behaviors is to represent the parameters using functions which take intermediate values initially; then, as the election draws closer, these values become more extreme, e.g. greater than 0.75 or smaller than 0.25. This follows the assumption that people are less likely to change their mind as the election draw closer. A similar assumption has been examined before by Halu in his agent-based model on social effects in political elections [58]. Furthermore, since we are interested in knowing what mechanism drives change in parameters, we can structure these functions to incorporate different assumptions and use data fitting as a way to justify which assumption is appropriate. This should be carried out after determining which parameters matter most in the election under consideration.

8 Additional notes on the limitation of our approach

We notice that the error for fitting the data collected from a single source is significantly smaller than the one fitting data collected using multiple sources. This is understandable due to the inconsistency between differences in methods for data collection and data samples.A better collection of data would be to use the average from multiple weekly sets of data, since it is a trend that the average of pollsters tends to be a good predictor for the outcome of the election. This, however, is difficult to achieve because polling data is generally collected through private agencies. Moreover, as mentioned in the previous section, data sets taken at different time periods with respect to the election day could potentially cause ambiguity in the meaning of the parameters. Thus, for the purpose of comparing voting behaviors between different elections, data sets from the same time period (relative to the election day) should be used. On the other hand, if the purpose is to find the mechanisms that determine the value of the parameters, then a data set of significant length, e.g., a data set that spans six months before the election day, should be used in parallel with function forms for the parameters.

Additionally, there is a significant proportion of eligible voters who are not aligned with either major party; thus, they are excluded from the data used to fit our models. This is important in two aspects. First, our models cannot measure the general trend of the voting behavior. Secondly, without considering this population, the two groups that do not wish to vote for either candidate, M_1 and M_2, are much smaller than in reality. This may limit the effectiveness of the sensitivity analysis and render any conclusion drawn from it meaningless.

Recall that our fitting does not guarantee a global minimum for the fitting error. However, since we do not have a complete understanding of the distribution of the parameter values with respect to the fitting error, or the distribution of local minima, we cannot say for sure what effects it could have on our system. Furthermore, obtaining the global minimum for the fitting error does not mean we have the most realistic set of parameters. We do not disregard the potential benefits of having a global minimum fitting error, we simply acknowledge the possibility that it can be an artifact of our method. In fact, since we vary our initial guesses randomly each time we run the program for parameter fitting and still obtain the same fitting error, we may as well assume that our fitting error is the global minimum on a large region in the parameter space.

Another limitation associated with the data comes from our use of the data to estimate the parameters for the models. These polling data take a sizable amount of time to collect. But when fitting, we assume them to be taken exactly at one point and two consecutive polls are always exactly seven days apart.

This assumption is necessary for our discrete-time Markov model, but it is another potential significant source of error. Not only quantitatively but philosophically, it is a strong assumption to consider that the least amount of time in which people can change their political positions is a week. Furthermore, to ease the complexity of the model, we choose to approximate the media coverage to be constant between any two data points. In addition to that, since our models, especially the person-to-person interaction model, have a large set of parameters, it is easy to over-fit. This can be addressed using a more thorough sensitivity analysis to find the least sensitive parameters for each election and eliminate them from the model. This can be carried out using the Latin Hypercube Sampling (LHS) method to vary all parameters within a certain range simultaneously to find their effects on the system relative to one another.

References

[52] ABC News, The Washington Post, *ABC News/Washington post poll: 2012 presidential election*, [Data set]. USABCWASH2012-1140, Version 2. Langer Research Associates/Abt SRBI, Inc. [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [distributor], September 2012, consulted on 15-Jul-2016.

[53] L. Arriola, J. M. Hyman, *Sensitivity analysis for uncertainty quantification in mathematical models* (G. Chowell, J. M. Hyman, L. M. A. Bettencourt and C. Castillo-Chavez, editors), Mathematical and Statistical Estimation Approaches in Epidemiology, Springer, 195–247, 2009.

[54] Cable News Network, *CNN/ORC poll: 2012 presidential election/gun control/Syria*, [Data set]. USORCCNN2012-007, Version 2. Opinion Research Corporation [Producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], August 2012, consulted on 15-Jul-2016.

[55] Cable News Network, *CNN/ORC poll: Presidential election*, [Data set]. USORCCNN2012-WI1E, Version 2. Opinion Research Corporation [Producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], August 2012, consulted on 15-Jul-2016.

[56] Cable News Network, *CNN/ORC poll: Wisconsin presidential election poll # 1*, [Data set]. USORCCNN2012-009, Version 2. Opinion Research Corporation [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], August 2012, consulted on 15-Jul-2016.

[57] CBS News, The New York Times, *CBS News/New York Times poll: 2012 presidential election/issues impacting the election*, [Data set]. USORCCNN2012-09A, Version 2. Opinion Research Corporation [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], September 2012, consulted on 15-Jul-2016.

[58] A. Halu, K. Zhao, A. Baronchelli, G. Bianconi, *Connect and win: the role of social networks in political elections*, EPL (Europhysics Letters) **102** (2013), no. 1, 16002.

[59] Monmouth University Poll, *National: all tied up in Monmouth poll, Obama 48%; Romney 48%*, `http://www.monmouth.edu/polling-institute/reports/`, consulted on 05-Jul-2016.

[60] Monmouth University Poll, *National: Obama gains over Romney, incumbent has edge on handling foreign policy in wake of Middle East violence*, `http://www.monmouth.edu/polling-institute/reports/`, consulted on 05-Jul-2016.

[61] Monmouth University Poll, *National: Romney Edges Obama in Monmouth Poll, post-debate: more voters like Romney, give him edge on key issues*, `http://www.monmouth.edu/polling-institute/reports/`, consulted on 05-Jul-2016.

[62] Monmouth University Poll, *National: Romney leads in Monmouth poll, Republican holds momentum after second debate*, `http://www.monmouth.edu/polling-institute/reports/`, consulted on 05-Jul-2016.

[63] Monmouth University Poll, *National: Ryan's initial impact, presidential horse race remains tight*, `http://www.monmouth.edu/polling-institute/reports/`, consulted on 05-Jul-2016.

[64] T. E. Patterson, *News coverage of the 2016 presidential primaries: horse race reporting has consequences*, `http://shorensteincenter.org/news-coverage-2016-presidential-primaries/`.

[65] Pew Research Center: Journalism & Media Staff, *Winning the media campaign 2012*, Pew Research Center, Technical Report, (2012).

[66] Politico, *George Washington battleground brushfire–week 1 #13560: weighted tables, September 24–27, 2012*, http://images.politico.com/global/2012/10/bg_brushfire_tables_oct12012.html, consulted on 05-Jul-2016.

[67] Politico, *George Washington battleground brushfire–week 2 #13697: weighted tables, October 15–18, 2012*, http://images.politico.com/global/2012/10/politico_gwbgp_oct22_weightedtables.html, consulted on 05-Jul-2016.

[68] Politico, *George Washington battleground brushfire–week 2 #13697: weighted tables, October 29–November 1, 2012*, http://images.politico.com/global/2012/11/politico_gwbgptracking_nov4_weighted_tables.html, consulted on 05-Jul-2016.

[69] Reuters News, *Core political data, May 14–18, 2016*, http://big.assets.huffingtonpost.com/2016ReutersTrackingCorePolitical5182016.pdf, consulted on 15-Jul-2016.

[70] Reuters News, *Core political data, May 28–June 1, 2016*, http://big.assets.huffingtonpost.com/2016ReutersTrackingCorePolitical60116.pdf, consulted on 15-Jul-2016.

[71] Reuters News, *Core political data, June 4–8, 2016*, http://big.assets.huffingtonpost.com/2016ReutersTrackingCorePolitical6082016.pdf, consulted on 15-Jul-2016.

[72] Reuters News, *Core political data, June 11–15, 2016*, http://big.assets.huffingtonpost.com/2016ReutersTrackingCorePolitical615.pdf, consulted on 15-Jul-2016.

[73] Reuters News, *Core political data, June 18–22, 2016*, http://big.assets.huffingtonpost.com/2016ReutersTrackingCorePolitical6222016.pdf, consulted on 15-Jul-2016.

[74] Reuters News, *Core political data, June 25–29, 2016*, http://big.assets.huffingtonpost.com/2016ReutersTrackingCorePolitical.pdf, consulted on 15-Jul-2016.

[75] Reuters News, *Core political data, July 2–6, 2016*, http://big.assets.huffingtonpost.com/2016ReutersTrackingCorePolitical7062016.pdf, consulted on 15-Jul-2016.

[76] N. Silver, *Fivethirtyeight's pollster ratings*, http://projects.fivethirtyeight.com/pollster-ratings/, consulted on 05-Jul-2016.

[77] TV News Archive, *Presidential campaign 2016: candidate television tracker*, http://television.gdeltproject.org/cgi-bin/iatv_campaign2016/iatv_campaign2016, consulted on 11-Jul-2016.