EAESP
EESP
CENTRO DE POLÍTICA
E ECONOMIA DO SETOR
PÚBLICO

Forecasting Runoff Elections: A Parsimonious Model Using First-round Results

We present a simple model of runoff elections forecasting based on the distribution of votes in the first round. Using a collection of 128 presidential elections from 44 countries, our model indicates a strong positive relationship between the top two candidates' first-round vote share and the second-round outcome. We confirm these results with an analysis of 287 elections for executive office in Brazil and evaluate our parsimonious approach next to more sophisticated specifications, but find limited improvement from including additional predictors. We argue the model performs well because (1) first-round results incorporate candidate and campaign-specific characteristics, (2) abstention rates increase slightly across rounds, (3) trailing candidates often need a large share of the remaining vote, and (4) voters of first-round losing candidates rarely have homogenous-enough preferences to turn elections around. As the model requires only readily available information, it can be easily applied to different contexts.

Working
PAPER

Forecasting Runoff Elections: A Parsimonious Model Using First-round Results

Abstract

We present a simple model of runoff elections forecasting based on the distribution of votes in the first round. Using a collection of 128 presidential elections from 44 countries, our model indicates a strong positive relationship between the top two candidates' first-round vote share and the second-round outcome. We confirm these results with an analysis of 287 elections for executive office in Brazil and evaluate our parsimonious approach next to more sophisticated specifications, but find limited improvement from including additional predictors. We argue the model performs well because (1) first-round results incorporate candidate and campaign-specific characteristics, (2) abstention rates increase slightly across rounds, (3) trailing candidates often need a large share of the remaining vote, and (4) voters of first-round losing candidates rarely have homogenous-enough preferences to turn elections around. As the model requires only readily available information, it can be easily applied to different contexts.

Recent inaccurate predictions based on polling data such as the 2016 U.S. presidential election (Chalabi 2016; Siegel 2016), the Brexit referendum (Duncan 2016; Hanretty 2016) and the Colombian peace agreement vote (Dube 2016) have raised criticisms toward the effectiveness of forecasting efforts from statistical modeling (Cassino 2016; Berger 2016; Zukin 2015). An explanation for these mis-predictions was the use of polls as the single data input. Forecasts based on electoral surveys tend to assume each individual poll as a random draw of the population; however, if a number of polls suffer from systematic bias, predictions with polling data alone will likely be inaccurate.[1] One strategy to overcome this limitation is to include other data inputs in forecasting analysis, such as results from past elections.[2]

This paper contributes to the study of election forecasting by presenting a simple model of second-round races that uses only the first-round vote share of the top two candidates as predictors. The two-round voting system is the most common method for the direct election of presidents around the world, currently in place in 86 countries (IDEA 2008).[3] In the majority of these countries, a second ballot is held between the two Largest vote-grabbers if no candidate receives an absolute majority of first-round votes.[4] We test our model with a sample of 128 presidential runoff elections from 44 countries. Our results indicate that the vote share received by the first-round winner as well the distance to the

[1]Techniques to account for differences in sampling, interviewing and other procedures across polling companies such as house effects, weighting or Bayesian updating may improve a model's estimates (Jackman 2005; Linzer 2013), but they often require decisions that are based on the same polling data, which could actually exacerbate bias, or on external information, which means the analysis does not rely on polls alone. Relying on the accumulation of polls to generate high-confidence predictions also limits the application of this approach to contexts where numerous polls are conducted.

[2]Other potential inputs include economic and social indicators, observations from focus groups and narrative accounts, and predictions from betting markets. These additional pieces of information may improve accuracy by reducing potential biases from individual sources (Young and Bernd 2014).

[3]This list represents 76% of all presidential democracies and includes France and most of its former colonies, the majority of Latin American countries, a number of former-Soviet republics, among others. For more information, see IDEA's electoral system database

[4]In Argentina, Costa Rica and Bolivia, for instance, a candidate is elected with 45% of the vote or 40% and a 10-point lead over the second. Systems may also vary in the number of candidates who move through to the second round (e.g., in French legislative elections, any candidate who reaches 12.5% of the electorate goes on to the second round). For a discussion of the trade-offs and optimization of the threshold for first-round victory, see Green-Armytage and Tideman (2019), O'Neill (2007), Pérez-Liñán (2006), and Shugart and Taagepera (1994).

trailing candidate are important predictors of the probability of win.

We then examine a second set of 287 executive elections in Brazil and find equivalent results. The analysis of the Brazilian case offers additional information on the candidates and campaigns that allows us to evaluate our parsimonious model next to more sophisticated specifications that include incumbency, demographic characteristics, information on the candidates' political offices, as well as the distribution of votes among first-round losing contestants. Yet, the comparison across specifications indicates that the inclusion of extra variables does not improve our ability to predict second round results. We argue the model performs well because (1) first round results largely incorporate important candidate and campaign-specific characteristics, (2) rates of abstention tend to increase across election rounds, (3) trailing candidates often need a large share of the remaining vote, and (4) voters of first-round losing candidates rarely have homogenous enough preferences in order to turn most elections around.

Finally, we test the prediction power of our model by comparing the predicted forecasts for the set of international elections with the actual results. Overall, the results indicate an accuracy rate of roughly 70%, which improves to 80% for elections of narrower confidence intervals. As a second test, we use the first-round results of the as inputs to correctly predict the winner of 13 out of 15 runoff races in the 2018 Brazilian elections, the same performance by polls conducted in the day before the second round vote. In contrast, polling data estimates from before the first-round vote anticipated the winner in only 7 elections.

The parsimony of our approach provides a few advantages. First, the model inputs are public and readily available. Second, it does not require the collection of external data such as exit polls (Selb et al. 2013). Third, the model can be applied to national and sub-national runoff elections around the world. Fourth, the model provides not only predicted scores and predicted probabilities for a wide range of first-round scenarios, but also the level of confidence of each prediction. Fifth, the model seems to outperform polls from before the first round of voting and is available in advance of polls conducted during the second-round campaign period.

We conclude by arguing that even analysts who rely on polling data may benefit from using the model as an independent test. Estimates of uncertainty generated by our electoral model can serve as a criterion for the decision of including or not, and with what weight, the model prediction in more elaborate forecasting analyses.

International presidential elections

Runoff elections are different from most electoral disputes in the sense they are preceded by a first ballot in which voters express their preferences by selecting two candidates to move through to the runoff. The nature of a second ballot, then, offers a simple method of predicting the final winner based on the first round's results. We formalize this approach by developing a model to predict the first-round front-runner's 1) vote percentage in the second round and 2) probability of winning.

To this end, we collected information on 128 runoff presidential elections from 44 countries available on Adam Carr's website of international elections.[5] A table with the list of countries and elections is presented in appendix A.[6]

The model is composed of three predictors: the front-runner's percentage of valid votes in the first round (mean=37.7 and sd=7.6), the difference in percentage points between the front-runner and the trailing candidate (mean=10.2 and sd=7.3), and an interaction

[5]The total number of two-round elections available on the website was 136 races, but 8 are dropped from the analysis because of judicial decisions and alleged irregularities that interfered with the selection of candidates or the second round's vote count (Afghanistan 2014, Benin 2001, Haiti 2010, Ukraine 2004), instances in which the front-runner received more than 50% of the valid votes but was not outright elected due to low turnout (Bulgaria 2006, Macedonia 2014), and outlier second-round results (Liberia 2011, Niger 2016, Benin 2001). Still, the inclusion of these 8 cases does not have any meaningful impact on the results

[6]The structure of the dataset is equivalent to analyses of American presidential elections in which the observations analyzed only include one of the two party candidates (e.g., Abramowitz 1988; Lewis-Beck and Rice 1982). In our database, all observations are the elections' front-runner with the total second round vote always adding to 100.

between the two variables.[7] That is, our specification models the performance in the runoff election as a product of the candidate's first-round vote percentage and magnitude of the difference to the opponent.

We evaluate our specification with an OLS regression to explain variation in the percentage of valid vote share in the second round (mean=`54.4` and sd=`9.9`) and a logistic regression to predict the probability the first-round front-runner wins the runoff. For an easier interpretation of the constant, the independent variables were centered at their means. The results are presented in Table 1.

Table 1: Parsimonious Model of Runoff Elections, International Elections

	Pct. Valid Votes in Runoff	Victory (dummy)
Pct. Valid Votes in 1st Round	0.141	0.095*
	(0.153)	(0.047)
Pct.-Point Difference 1st and 2nd Candidates	0.441**	0.084*
	(0.156)	(0.042)
Pct. Valid Votes * Pct.-Point Difference	−0.00001	0.005
	(0.019)	(0.006)
Constant	54.741**	1.049**
	(1.008)	(0.262)
N	128	128
Adjusted R^2	0.136	

* $p<0.05$; ** $p<0.01$

The coefficients of the first regression present the estimated change in the front-runner's second round vote percentage associated with a one-unit increase in the predictor variable. But because the model includes the interactive term, the first two coefficients indicate the estimated change in the second-round ballot of a one percentage point increase in the variable, holding the other variable at its mean.

The first coefficient indicates that an increase of one percentage point in the first-round voting is associated with a small and non-significant increase in the second-round vote when the difference to the trailing candidate is at its mean (`10.2`). In other words, when

[7] The descriptive statistics of the dependent and independent variables are presented in the appendix B.

the front-runner is ahead by about 10 percentage points- e.g., 49-39, 45-35 or 35-25, the change in second-round vote share associated with a higher vote percentage in the first round is small and not distinguishable from 0.

On the other hand, an increase of one percentage point in the difference between the two candidates is associated with a statistically significant increase of 0.44 percentage points in the second round when the front-runner received 37.7% in the first round. Considering the coefficient for the interactive term is roughly zero, the model predicts a meaningful improvement of 0.44 percentage points in favor of the front-runner for every one-point increase in the lead against the first-round runner-up, e.g. 42-30 compared to 41-30 or 41-30 compared to 41-31. These results imply that the distance in the first round is more important than the front-runner's percentage of votes per se. Finally, the constant indicates that the predicted vote score for the front-runner in a scenario with the mean first-round percentage (37.7) and mean difference to the trailing candidate (10.2) is 54.7.

The coefficients of the logistic regressions also indicate that the gap between first-round winner and runner-up is an important predictor of the probability of victory. Yet, in this model, the vote share of the front-runner is associated with a statistically significant increase in the probability of win. The substantive interpretation of the coefficients is not as easily interpretable, however. An easier way to interpret their expected impact is to graph the predicted vote share and predicted probabilities of victory as a function of varying first-round scenarios. To do this, we first plot the expected second round vote percentage for the front-runner across all existing first-round vote percentages and the 25th, 50th and 75th percentile of the distribution in the difference between leading and trailing candidate: roughly a gap of 3, 8 and 14 percentage points, respectively. The results are presented in Figure 1.

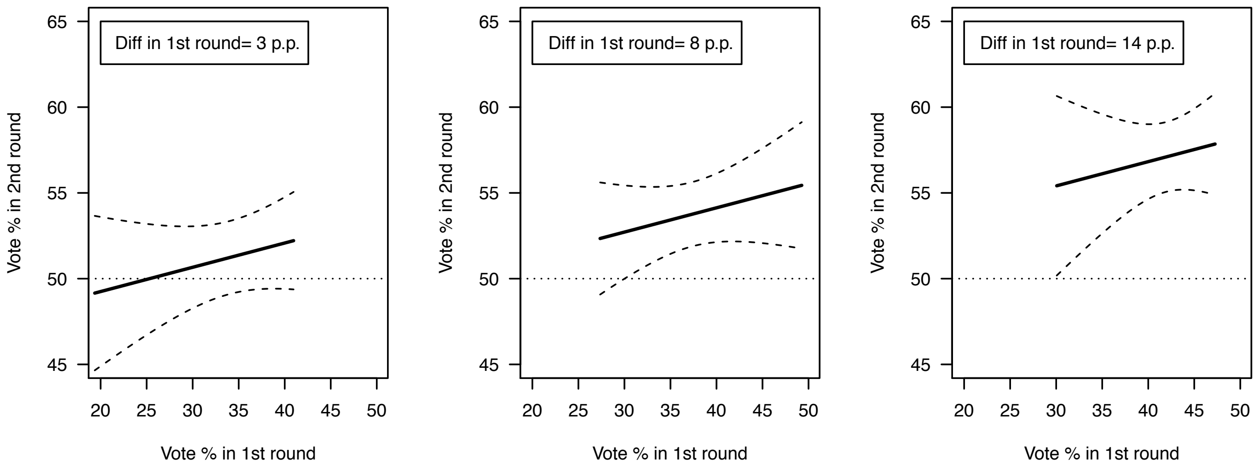

Figure 1: Predicted Vote Percentage of Front-Runner, International Elections

The low steepness of the lines indicates that the percentage of valid votes in the first round (x-axis) is not strongly associated with substantive differences in the predicted vote share in the second round (y-axis) when the difference is fixed. In contrast, as the difference between the front-runner and trailing candidate increases (moving from the graph on the left to the right), the predicted second-round scores are significantly higher. That is, while in scenarios in which the gap is approximately 3 percentage points, the model predicts a final share of between about 49 and 52%. In scenarios of an 8-p.p. gap, the model predicts a vote share between 52 and 55%, with higher levels of confidence as indicated by the dashed lines that portray the 95% confidence interval around these estimates. And when the gap jumps to 14 p.p., the predicted scores range around 57%.[8]

When we analyze the second dependent variable, the predictions corroborate the scenarios for the first-round winner. Figure 2 shows the predicted probability of victory for the same set of scenarios.

[8] It is important to note that while the graphs indicate that front-runners would be expected to win, the model predicts scenarios in which the front-runner is no longer the favorite. For example, in an election in which the front-runner receives 24% of valid votes in the first round and the competitor receives 21%, the model predicts the second-placed candidate as favorite, albeit with a large confidence interval.

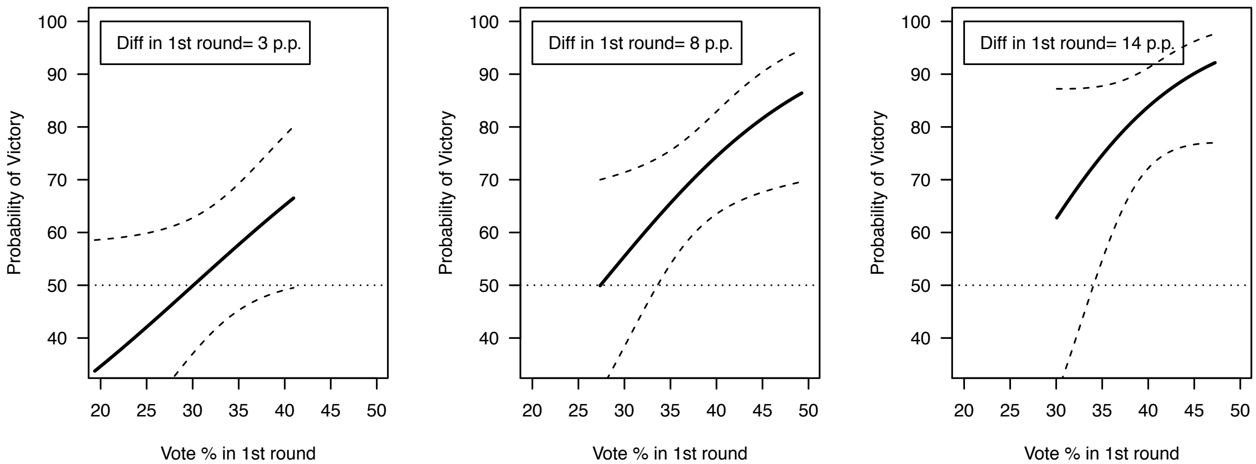

Figure 2: Predicted Probability of Front-Runner Winning, International Elections

An increase in the vote percentage of the first round always improves a candidate's chances of winning. In scenarios of a 3-p.p. gap, the probability of a front-runner victory improves from roughly 41% in a 25-21 first-round to more than 60% in a 41-38 race. However, these predictions are generated with large confidence intervals that always include the 50% margin. In scenarios with an 8-p.p. gap, the probability of winning increases from roughly 52% in a 27-19 scenario to more than 80% in a 49-41 scenario. Moreover, the level of confidence around the latter estimate is substantially narrower than in scenarios in which the front-runner receives smaller vote percentages. Lastly, in scenarios where the difference between the two candidates is about 14 points, the predicted probability of winning is never smaller than 60% and continues to increase when the vote percentage in the first round is larger. In an election in which the front-runner received 47% of the first-round vote and the opponent 33%, the model suggests a probability of victory of about 85%, with a high degree of confidence.

This set of international elections shows we can reasonably predict the second-vote outcome with the vote share of the top two candidates. We turn to a second group of two-round elections to test if these results hold when we look at a larger and more similar set of races. That is, we use the case of Brazilian subnational elections for executive office as a complementary analysis that offers more observations and information about the candidates.

Brazilian subnational elections

The majority runoff two-round system is used in Brazil not only for the election of the president, but also of governors of the 27 states and mayors of municipalities with 200 thousand registered voters or more. Since 1998, a total of 287 runoff elections have taken place.[9] This collection of elections provide an opportunity to assess the extent to which the results from the set of international elections are consistent with a larger and naturally more similar group of subnational runoff elections (Snyder 2001). Using the same predictors and dependent variables, we repeat the estimation of the OLS and logistic regressions as the table above.[10]

Table 2: Parsimonious Model of Runoff Elections, Brazil 1998-2018

	Pct. Valid Votes in Runoff	Victory (dummy)
Pct. Valid Votes in 1st Round	-0.024	0.067^{*}
	(0.084)	(0.031)
Pct.-Point Difference 1st and 2nd Candidates	0.433^{**}	0.063^{*}
	(0.076)	(0.027)
Pct. Valid Votes * Pct.-Point Difference	-0.006	0.005
	(0.011)	(0.004)
Constant	54.446^{**}	1.148^{**}
	(0.481)	(0.162)
N	287	287
Adjusted R^2	0.139	

* p<0.05; ** p<0.01

Overall, the results are remarkably consistent. The first coefficient indicates that an increase of one percentage point in the first-round is again associated with a non-statistically significant change in the second-round vote. Also consistent with the previous table, an

[9]These include 5 presidential, 74 gubernatorial and 208 mayoral.

[10]The mean vote percentage of the first-round front runner is 54.3 and its standard deviation,9.9.

increase of one percentage point in the difference between the two candidates is associated with a significant increase of 0.43 percentage points in the second round when the front-runner received 40.7% in the first round.

The results from the second regression are also analogous. The size of first-round difference and vote share of the front-runner are again associated with a larger probability of victory for the front-runner, and the moderating variable is not distinguishable from 0. When plotting these regression results using the same scenarios as above, the predicted vote share does not vary as much as in the previous analysis but rather ranges between 51 and 56%, with narrower confidence intervals. The 14-p.p. scenario is also slighly different as it shows a negative although rather monotous relationship.

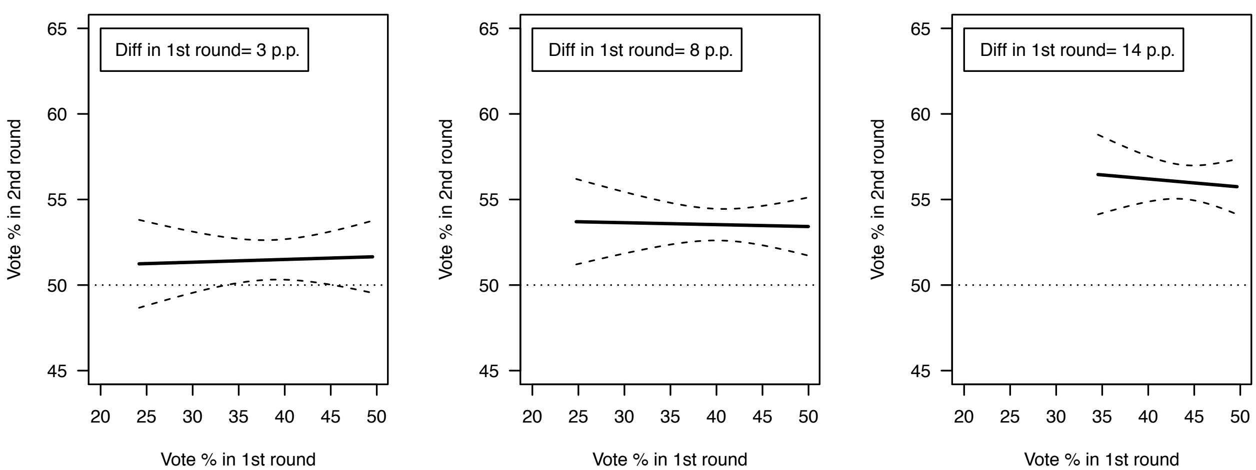

Figure 3: Predicted Vote Percentage of Front-Runner, Brazil 1998-2018

The predicted probability of victory also follows the parameters from the group of international elections, but again, there is more precision around the point estimates. Figure 4 indicates that an increase in the vote percentage of the first round always improves a candidate's chances of winning. In scenarios of a 4-p.p. gap, the probability of a front-runner victory improves from 55% in a 25-21 first-round to more than 70% in a 49-45 race. Similarly, the probability of winning increases from roughly 56% in a 27-19 scenario to about 80% in a 49-41 scenario.[11]

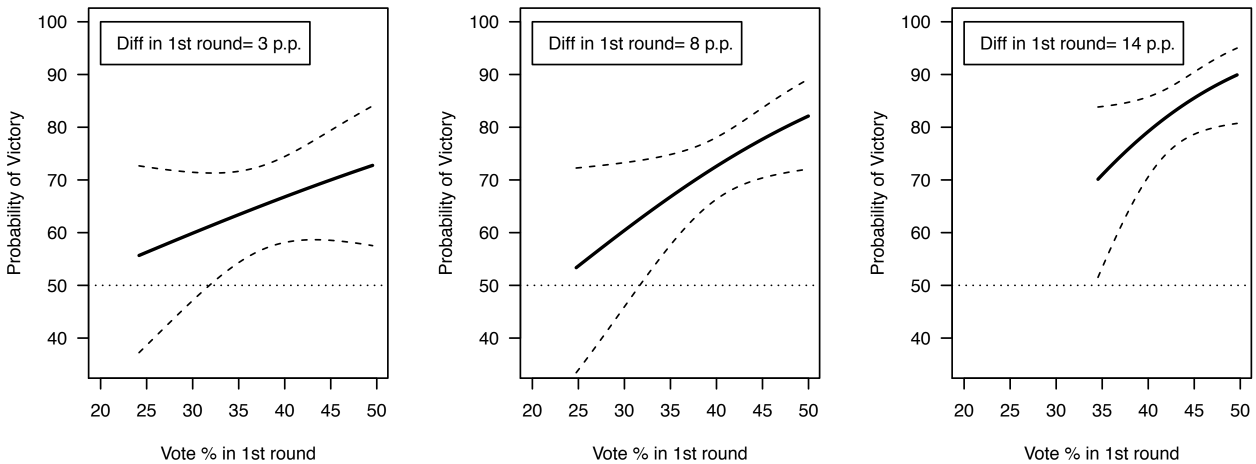

Figure 4: Predicted Probability of Front-Runner Winning, Brazil 1998-2018

Alternative specifications

We believe the parsimony is an important advantage of the model. Yet, one may wonder whether its simplicity leaves out key information about the candidates, parties, and campaign strategies. We believe the results of the first round incorporate a great deal of the value associated with factors featured in traditional models of voting behavior.

To substantiate this position, we make use of additional electoral data from Brazil to test a number of alternative specifications that consider differences across years and elected offices, characteristics of the candidates such as sex, age, incumbency, political office at the time of the election (if any), as well as the number and vote percentages of the first-round losing candidates. Yet, the differences in second-round vote share and probability of winning associated with these variables are predominantly non-significant and only affected the predictions marginally. That is, after testing numerous different specifications with these variables, we find very little change in the percent of elections

[11] In appendix G, we present an alternative version of the confidence intervals when calculating with bootstrapping. The results are roughly identical.

Table 3: Model Specification, PCP and Betas

Model	# of IVs	β % Votes 1st round	β Pct.-Point Difference	PCP
Original	3	0.07	0.06	75.3
+ Type of office	5	0.06	0.06	75.3
+ Number of candidates	6	0.08	0.05	75.3
+ % Abstention	7	0.07	0.06	76.3
+ Sex of Candidates	10	0.07	0.06	76
+ Age of Candidates	11	0.07	0.06	75.5
+ Education of Candidates	14	0.06	0.07	76.9
+ Largest expenditure	15	0.06	0.07	76.5
+ Incumbency of candidates	18	0.06	0.08	76.5

correctly predicted and magnitude of our coefficients of interest.

Table 3 summarizes results from these models by identifying the variables included in each specification, number of independent variables, the variation of the two principal coefficients of the model - the percentage of votes and the point-difference between the first two candidates, and the percentage of elections correctly predicted (PCP) from the logistic regression model.

As suggested, there is very little variation in the ability to predict elections across specifications.[12] The first row shows the percent of elections correctly predicted and first two coefficients of the second regression model presented in Table 2 (i.e. Original). The subsequent rows indicate the stepwise addition of other variables to the model specification. For instance, the second row reports the percent of elections correctly predicted and coefficients of interest when we added dummies for the type of electoral office (presidency, governorship or mayorship). Overall, the results show that accounting for different averages across offices, incumbency and personal characteristics does very little to the model. Burdening the model with more independent variables that may require data collection efforts does not seem to be compensated by a substantive improvement in our predictions. With the best-performing model, we observe an increase of only 1.6 percentage points in

[12]The actual results from these regressions are presented in appendix C, along with a version of the table that presents the number of observations and adjusted-R^2 of the models.

PCP in spite of adding eleven independent variables.

Why the model works

We theorize the power of the model derives from three simple axioms about the electoral behavior of voters across rounds.

Voters are consistent across rounds

Voters of the two leading candidates rarely flip from change votes between the first and second round. In fact, data from the 2002, 2006, 2010 and 2014 rounds of the Brazilian Electoral Study indicate that only 6% of voters (in presidential and gubernatorial elections) who chose a candidate who moved through to a second round flipped their vote to the opposing candidate. And because these changes tend to occur in both directions with roughly the same frequency, we should not expect an impact on the final results from this group of voters.[13] Hence, the majority of votes in runoff races are set after the first round: on average, 71.7 of valid votes in the first round are cast for the two candidates who move through to the second round (sd=11.5).

Rates of abstention are higher in the runoff

The share of the electorate "available" to turn elections around is composed of voters of the first-round losing candidates and those who have not participated in the first ballot. Yet, we know from the first round that in the presence of alternatives, these two groups preferred to not vote for the candidates who moved through to the runoff. In the second round, they can choose one of the candidates left in the race or not vote by abstaining or voting invalidly.

[13] Given the nature of measurement errors associated with survey data (Groves and Lyberg 2010; Holbrook and Krosnick 2010), this is likely an overestimation of the phenomenon.

The same electoral data from the international elections indicate the number of votes cast for candidates is, on average, slightly smaller in the second round. This drop is driven by an average increase in rates of abstention of 1.65 percentage points, but the number of invalid votes actually tends to drop slightly, on average, by 0.35 points.[14] The abstention rate is also, on average, slightly higher (2.5 p.p.) in the second round among Brazilian executive elections, in spite of voting being mandatory.[15] But again, the number of invalid votes actually tends to decrease across rounds, which almost offsets the drop in turnout. Overall, this suggests that, on average, there is little voter mobilization across rounds, or mobilization efforts are not very effective, and the final group of voters available to turn elections around tends to be a minority of the electorate.

Trailing candidates need the lion's share of the remaining electorate

For most elections to be flipped, the group of voters of first-round losing candidates have to predominantly choose the first-round runner-up. For instance, in the presidential elections' average first-round scenario 37.7-27.5, the trailing candidate would need roughly 64% (22.5 out of 34.8) of the remaining vote just to even the race- a distribution more one-sided than the vote of the candidates in the first-round. In scenarios of 41-31 or a closer 45-41, the trailing candidate would need 68 and 64% of the "unclaimed" electorate to equalize. And even in a race not as close to the majority threshold such as a 35-27 dispute, the trailing candidate would still require about 61%.

One may argue that scenarios in which the first-round third candidate has a large share of votes should lead to more turnaround results than similar elections with a more fragmented group of first-round losing-candidate voters. That is, the concentration of remaining voters around the third-place candidate may render it easier to coordinate, therefore we should expect more inversions of the first-round result. Yet, this hypothesis

[14]Due to missing data, these estimates are based on only 94 presidential elections.

[15]This result is also aligned with primary elections in the United States where turnout rates was smaller in 77% of second-round races when compared to the initial vote (Wright 1989.)

assumes that the group of voters who can flip the electoral outcome will coordinate or at least choose the second candidate more frequently than the first-round winner. Recent works on elections under majority runoff rule indicate that voters tend to engage in coordination and strategic voting in the first round (Blais et al. 2011; Bouton 2013; Bouton et al. 2019; Bordignon, Nannicini, and Tabellini 2016), which suggests that we should not expect the remaining electorate to have favorable views toward the trailing candidate more often than toward the front-runner.

In accordance with this strand of the literature, we do not find evidence that the percentage of votes for the third candidate in the first round is associated with differences in the percentage of votes for the leading candidate in the second round, nor with the probability of winning.[16]

To demonstrate the prediction power of this model, we compare the forecasts generated by the model for two sets of elections: the 128 second-round presidential races described in the first section and the 2018 Brazilian runoff elections- with their actual results.

Application: International Presidential Elections

We assess the extent to which the model parameters generated with the Brazilian elections' dataset produces accurate estimates of the 128 runoff presidential elections from around the world. Using the first-round vote percentages as inputs, we generate predictions of the second-round vote and compare with actual results in Figure 6. The filled circles indicate races in which the model correctly predicts the winner.

[16]This hypothesis is tested with the regression model and presented in appendix E. We do find, nonetheless, that the variance around the percentage of votes for the leading candidate in the Brazilian runoffs is larger when the third candidate receives a higher percentage. In other words, as the percentage of votes for the third-placed candidate increases, the variation in electoral results tends to be larger, thus indicating more homogeneity among the remaining electorate when the third-placed candidate received more votes. To show this, in the same, we plot the relationship between the vote percentage of the third-placed candidate and the residuals from the regression model predicting the front-runner second-round vote share after controlling for the first-round of the candidates who moved through to the runoff.

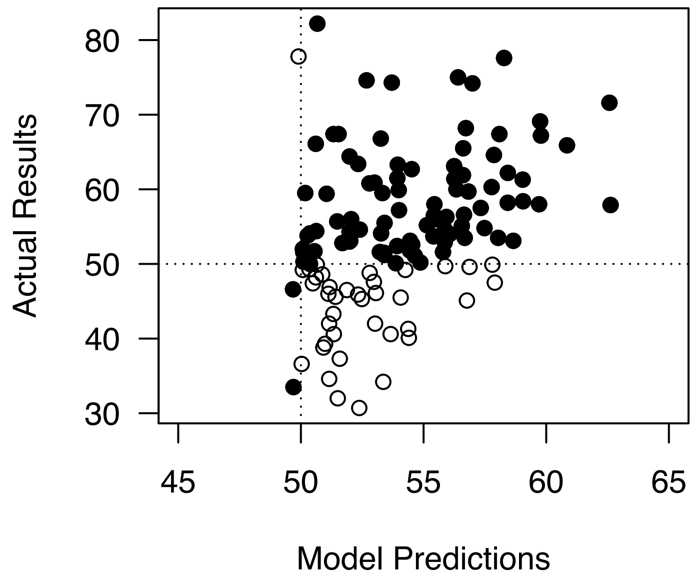

Figure 5: Scatterplot of Model Predictions and Actual Results, International Elections

Given the model tends to cluster prediction estimates close to the mean, its predictions yield favorable forecasts to the front-runner. As a whole, the model correctly predicted the winner 70.3% of the time, with a mean absolute error of 6.7. When we restrict the sample to instances in which the lower bound of the prediction interval was above 50% (i.e., when we have more confidence on the prediction), the percent correct predicted improves to 79% and the mean absolute error drops to 6.1.[17]

It is important to note that there is variation in the performance of the model across elections. For instance, when we compare across continents, the model performs better in elections in Europe and Latin America than Africa or Asia. More precisely, the mean absolute error for elections in Europe and Latin America were about 6 percentage points as opposed to 8 and 13 in Africa and Asia, respectively. The percent of elections correctly predicted was also higher in the prior two, while not largely different in Africa, but substantially lower for the 5 elections in Asia.

[17] As a whole, the model's accuracy tends to improve as our measures of confidence increases.

Table 4: Predictions of International Elections, per continent

Continent	Observations	Mean Absolute Error	PCP
Africa	31	8.2	67.7
Americas	36	5.7	72.2
Asia	5	12.9	40.0
Europe	56	6.1	73.2

Application: 2018 Brazilian Runoff Elections

As a second test, we use a version of the model based on the Brazilian executive elections from 1998 and 2016, and the 2018 first-round vote share of the candidates who move through as inputs to generate forecasts for each of the 15 runoff elections in Brazil.[18] In addition to comparing the predictions to the actual results, we also evaluate the model's performance next to estimates from polls conducted the day before the first and second round votes. That is, we compare the model's estimates with poll-based predictions from IBOPE's (Brazilian Institute of Public Opinion and Statistics) last survey prior to election day.[19]

The first three columns of Table 3 present the district of the race (state name abbreviations for gubernatorial and BR for presidential), names and parties of the candidates, and the valid vote share of the *first-round front-runner*, who is listed first in column 2. The winner of the race is indicated by the asterisk next to the name. Columns four and five present the expected second-round vote share for the front-runner according to the last polls before the first and second round, as well as their margin of error interval in parentheses. The last two columns present the expected vote share based on our OLS model and the front-runner's probability of win, with the 99% prediction interval in parentheses.

[18]The model coefficients with data up to 2016 vary only marginally compared to the results presented above when we include the 2018 data. See appendix F for the comparison across regression models.

[19]IBOPE is perhaps the most reputable polling company in Brazil. It was founded in 1942, and since then has been at the forefront of media and public opinion research in the country. All public polls produced by IBOPE during the 2018 campaign period are available online: https://especiais.gazetadopovo.com.br/eleicoes/2018/pesquisas-eleitorais/governador/ (Accessed April 22, 2019).

Table 5: Predictions for the 2018 Runoff Elections

Race	Candidates	Result	Poll 1	Poll 2	Model	Win Prob.
AM	Lima* (PSC) v. Mendes (PDT)	59	52 (49,55)	64 (61,67)	51 (49,53)	62 (45,76)
AP	Waldez* (PDT) v. Capi (PSB)	52	43 (40,46)	47 (44,50)	52 (50,54)	63 (50,74)
BR	Bolsonaro* (PSL) v. Haddad (PT)	55	52 (50,54)	54 (52,56)	57 (55,59)	89 (77,95)
DF	Ibaneis* (MDB) v. Rollemberg (PSB)	70	75 (72,78)	75 (72,78)	62 (58,65)	92 (73,98)
MG	Zema* (Novo) v. Anastasia (PSDB)	72	NA	68 (65,71)	56 (55,57)	82 (72,89)
MS	Azambuja* (PSDB) v. Odillon (PDT)	52	55 (52,58)	51 (48,54)	56 (54,57)	83 (74,90)
PA	Barbalho* (MDB) v. Miranda (DEM)	55	63 (60,66)	57 (54,60)	57 (55,59)	91 (78,97)
RJ	Witzel* (PSC) v. Paes (DEM)	60	NA	54 (51,57)	59 (56,62)	86 (68,95)
RN	Bezerra* (PT) v. Eduardo (PDT)	58	55 (52,58)	55 (52,58)	56 (54,57)	86 (75,92)
RO	Junior (PSDB) v. Rocha* (PSL)	34	NA	37 (34,40)	54 (51,56)	60 (42,76)
RR	Denarium* (PSL) v. Anchieta (PSDB)	53	44 (41,47)	54 (51,57)	52 (50,53)	69 (56,79)
RS	Leite* (PSDB) v. Sartori (MDB)	54	60 (57,63)	57 (54,60)	53 (51,54)	65 (55,74)
SC	Merísio (PSD) v. Moisés* (PSL)	29	NA	41 (38,44)	51 (49,54)	61 (42,77)
SE	Belivaldo* (PSD) v. Valadares (PSB)	65	46 (43,49)	61 (58,64)	58 (56,61)	84 (68,93)
SP	Doria* (PSDB) v. França (PSB)	52	50 (48,52)	50 (47,53)	55 (52,58)	62 (38,81)
	Mean Abs. Error (# Correct Predict)		6.5 (7)	3.8 (13)	6.4 (13)	

The last row presents the mean absolute error and number of elections correctly predicted according to each method.

The results from the last polls conducted prior to the first round of voting (column 4) were limited in predicting the winner of the runoff races. Out of 15 elections, responses to the hypothetical second-round scenario only indicated the correct winner 7 times. In 4 of the 14 gubernatorial elections, respondents were not asked about a potential final dispute between the candidates who moved through due to oversight of the pollsters. Among the elections in which the final contest was asked about, the mean absolute error was 6.5 percentage points.

The polls from the day prior to the second round of voting performed substantially better. Not only the mean absolute error was smaller (3.8), but the polls correctly indicated the winner 13 out of the 15 runoff elections.[20] The poll conducted in Amapá (AP) underestimated the elected-governor Waldez who received roughly 52% of the valid vote

[20]The magnitude of the error by polls near election day is comparable to the average root mean square error reported by Jennings and Wlezien (2016).

as opposed to the expected 47% from the final IBOPE electoral survey. The poll from São Paulo (SP) suggested the race was even, whereas then-candidate Doria received 52% of the second ballot.[21]

The predictions from our parsimonious model were also correct 13 out of 15 times, although the mean absolute error was larger than the polls.[22] As discussed in the previous section, because the model clusters its predictions around the mean, it tends to miss elections with low and high vote shares, as well as turnaround results such as the elections in Roraima (RO) and Santa Catarina (SC) where the candidates from the same party of the president-to-be, Jair Bolsonaro, received much higher shares of the final vote than the model predicted.

Finally, the last column indicates the predicted probability of victory based on the logistic regression model. Although all forecasts give the edge to the front-runner, the probability and confidence around the result vary substantially. While high probabilities tend to be connected with small intervals, elections with lower probabilities are associated with higher degrees of uncertainty. Consequently, predictions of the latter type should be seen with more caution than for the former. The two races the electoral model missed (RO and SC) are among the four (along AM and SP where the polls mispredicted) the model indicated the most uncertainty, with the lower bound of the prediction interval not surpassing 45%.

When the two approaches pointed in the same direction, all 11 elections were correctly predicted. Considering the chronological order of when these numbers are available: poll 1, first round results and poll 2- elaborate forecasting approaches can continually update their estimates with all three pieces of information. Interested political parties and poll houses can also use the predictions from the electoral model proposed here to make informed decisions about how to best allocate resources during the second-round

[21] If we consider only elections in which the prediction interval does not include 50%, the polls would not have got any incorrect, but would also drop an accurate one (Azambuja's election in the state of Mato Grosso do Sul- MS).

[22] And 10 out of 11 correct if we dispense those in which the prediction interval includes the midpoint.

campaign period. We interpret the similarity in results between the application of our model to presidential elections around the world and the 2018 Brazilian elections as an indication that the model can be applied to subnational races around the world.

Conclusion

Predictions based on polls tend to rely on important assumptions about the unbiasedness of its estimates. We suggest that forecasting analysts can (and should) test the reliability of their polling results with independent data sources, such as past elections

We propose a simple model for runoff elections constructed from 128 international presidential elections and 287 races for executive office in Brazil. Our tests indicate that the first-round vote share of the two candidates who move through to the runoff can indicate the size of the front-runner's advantage, and more importantly, which elections analysts may want to generate more information in order to produce high-confidence estimates. The positive and significant relationship between the first and second round results also provide evidence in favor of arguments that adjustments to the minimum threshold for first-round victory may reduce costs (Blais et al. 2011; Bouton 2013; Shugart and Taagepera 1994; O'Neill 2007), polarization (Bordignon, Nannicini, and Tabellini 2016), and likelihood of crisis of governance (Pérez-Liñán 2006) without having an impact on the great majority of runoff races.[23]

We recognize the degree of validity of the estimates produced here might be conditional on a couple of characteristics of our sample. First, the elections analyzed here are always disputed by the two first vote-grabbers when no candidate reaches an absolute majority in the first round. Some countries use modified versions of the two-round system in which the number of candidates who move through can be larger than the two, and in which the threshold for first-round victory is lower. Second, the temporal distance between first

[23] Obviously, the counterargument is that lower thresholds may lead to a winner who is not preferred by the majority and do not offer voters an opportunity for a head-to-head comparison.

and second ballot may interfere with the model's performance. The length of the second round campaign of all Brazilian elections analyzed here were either 3 or 4 weeks. It is also important to note that there may be a high degree of heterogeneity in the relationship between first and second round vote as product of contextual characteristics such as party system institutionalization (Pérez-Liñán 2006).

The simplicity of the model, nonetheless, offers a quick and free-of-charge tool to not only create predictions about the final electoral results in the dozens of countries where two-round elections take place for president and lower-level offices, but also to use the confidence around these estimates as information for other decisions such as which races to allocate resources or the introduction of weights in more sophisticated analyses.

We believe the model presented here may be particularly helpful for elections where not many polls conducted, such as those in poorer nations and two-round subnational disputes like the gubernatorial and mayoral elections in countries around the world.

Appendix A- List of International Presidential Runoff Elections

References

Abramowitz, Alan I. 1988. "An Improved Model for Predicting Presidential Election Outcomes." *PS: Political Science and Politics* 21(4): 843–47.

Berger, James. 2016. "Polls Are Bad for Democracy." *US News*, November. https://www.usnews.com/opinion/articles/2016-11-22/polls-are-bad-for-democracy.

Blais, André, Simon Labbé-St-Vincent, Jean-François Laslier, Nicolas Sauger, and Karine Van der Straeten. 2011. "Strategic Vote Choice in One-Round and Two-Round Elections: An Experimental Study." *Political Research Quarterly* 64(3): 637–45.

Bordignon, Massimo, Tommaso Nannicini, and Guido Tabellini. 2016. "Moderating Political Extremism: Single Round Versus Runoff Elections Under Plurality Rule." *American Economic Review* 106(8): 2349–70.

Bouton, Laurent. 2013. "A Theory of Strategic Voting in Run-Off Elections." *American Economic Review* 103(4): 1248–88.

Bouton, Laurent, Jorge Gallege, Aniol Llorente-Saguer, and Rebecca Morton. 2019. "Runoff Elections in the Laboratory." *National Bureau of Economic Research* 25949.

Cassino, Dan. 2016. "Why Pollsters Were Completely and Utterly Wrong." *Harvard Business Review*, September. https://hbr.org/2016/11/why-pollsters-were-completely-and-utterly-wrong.

Chalabi, Mona. 2016. "Yes, the Election Polls Were Wrong. Here's Why." *The Guardian*, September. https://www.theguardian.com/commentisfree/2016/nov/09/polls-wrong-donald-trump-election.

Dube, Ryan. 2016. "On Colombian Peace Vote, Pollsters Miss Again." *Wall Street Journal*, October. https://www.wsj.com/articles/on-colombia-peace-vote-pollsters-miss-again-1475524316.

Duncan, Pamela. 2016. "How the Pollsters Got It Wrong on the Eu Referendum." *The Guardian*, June. https://www.theguardian.com/politics/2016/jun/24/how-eu-referendum-pollsters-wrong-opinion-predict-close.

Green-Armytage, James, and T. Nicolaus Tideman. 2019. "Selecting the Runoff Pair." *Public Choice* Forthcoming: 1–19.

Groves, Robert M., and Lars Lyberg. 2010. “Total Survey Error: Past, Present, and Future.” *Public Opinion Quarterly* 74(5): 849–79.

Hanretty, Chris. 2016. “Here’s Why Pollsters and Pundits Got Brexit Wrong.” *Washington Post: The Monkey Cage*, June. https://www.washingtonpost.com/news/monkey-cage/wp/2016/06/24/heres-why-pollsters-and-pundits-got-brexit-wrong/?utm_term=.b8f6f0b7cda0.

Holbrook, Allyson L., and Jon A. Krosnick. 2010. “Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Technique.” *Public Opinion Quarterly* 74(1): 37–67.

IDEA, International Institute for Democracy and Electoral Assistance. 2008. “Electoral System Design: The New International Idea Handbook.” https://www.idea.int/sites/default/files/publications/electoral-system-design-the-new-international-idea-handbook.pdf.

Jackman, Simon. 2005. “Polling the Polls over an Election Campaign.” *Australian Journal of Political Science* 40(4): 499–517.

Jennings, Will, and Christopher Wlezien. 2016. “The Timeline of Elections: A Comparative Perspective.” *American Journal of Political Science* 60(1): 219–33.

Lewis-Beck, Michael S., and Tom W. Rice. 1982. “Presidential Popularity and Presidential Vote.” *Public Opinion Quarterly* 46(4): 534–37.

Linzer, Drew A. 2013. “Dynamic Bayesian Forecasting of Presidential Elections in the States.” *Journal of the American Statistical Association* 108(501): 124–34.

O’Neill, Jeffrey C. 2007. “Choosing a Runoff Election Threshold.” *Public Choice* 131(3-4): 351–64.

Pérez-Liñán, AnÍbal. 2006. “Evaluating Presidential Runoff Elections.” *Electoral Studies* 25(1): 129–46.

Selb, Peter, Michael Herrmann, Simon Munzert, Thomas Schübel, and Susumu Shikano. 2013. “Forecasting Runoff Elections Using Candidate Evaluations from First Round Exit

Polls." *International Journal of Forecasting* 29(4): 541–47.

Shugart, Matthew Soberg, and Rein Taagepera. 1994. "Plurality Versus Majority Election of Presidents: A Proposal for a 'Double Complement Rule'." *Comparative Political Studies* 27(3): 323–48.

Siegel, Ethan. 2016. "The Science of Error: How Polling Botched the 2016 Election." *Forbes*, September. https://www.forbes.com/sites/startswithabang/2016/11/09/the-science-of-error-how-polling-botched-the-2016-election/#44ab81033795.

Snyder, Richard. 2001. "Scaling down: The Subnational Comparative Method." *Studies in Comparative International Development* 36(1): 93–110.

Wright, Stephen G. 1989. "Voter Turnout in Runoff Elections." *Journal of Politics* 51(2): 385–96.

Young, Clifford, and Elisa Bernd. 2014. "The Rise of the Forecaster-Pundit: Big Data, Data Aggregation and the 2012 Us Presidential Election." *IPSOS MORI: Understanding Society*, July, 22–26. https://www.ipsos.com/sites/default/files/publication/9211-03/sri-understanding-society-july-2014.pdf.

Zukin, Cliff. 2015. "What's the Matter with Polling?" *The New York Times*, June. https://www.nytimes.com/2015/06/21/opinion/sunday/whats-the-matter-with-polling.html.

Supplementary Information

Forecasting Runoff Elections: A Parsimonious Model Using First-round Results

Contents

Appendix A- List of International Presidential Runoff Elections

Table 1: Two-Round Elections in Database

Country	Elections
Argentina	2015
Armenia	1999, 2003
Austria	2016
Benin	2006
Brazil	2002, 2006, 2010, 2014, 2018
Bulgaria	1996, 2001, 2011, 2016
Cape Verde	2001, 2011
Central African Republic	1993, 2005, 2015
Chad	1996
Chile	1999, 2005, 2010, 2013, 2017
Colombia	2010, 2014, 2018
Costa Rica	2002, 2014, 2018
Croatia	2002, 2005, 2009, 2014
Cyprus	1998, 2008, 2013, 2018
DR Congo	2006
East Timor	2007, 2012
Ecuador	1998, 2002, 2006, 2017
El Salvador	2014
France	1965, 1969, 1974, 1981, 1988, 1995, 2002, 2007, 2012, 2017
Georgia	2018
Ghana	2008
Guatemala	1999, 2003, 2007, 2011, 2015, 2019
Guinea	2016, 2010
Guinea-Bissau	2005, 2009, 2014
Indonesia	2004
Iran	2005
Ivory Coast	2010
Liberia	2005, 2017
Lithuania	1997, 2002, 2004, 2014, 2019
Macedonia	2004, 2009, 2019
Madagascar	2001, 2013, 2018
Mali	2002, 2013, 2018
Niger	1999, 2004, 2011
Peru	2000, 2001, 2006, 2011, 2016
Poland	1990, 1995, 2005, 2010, 2015
Romenia	1994, 1996, 2000, 2004, 2009, 2014
São Tomé and Príncipe	2011
Senegal	2000, 2012
Seychelles	2015
Sierra Leone	2007
Slovakia	1999, 2004, 2009, 2014, 2019
Slovenia	2002, 2007, 2012, 2017
Ukraine	1999, 2010, 2019
Uruguay	1999, 2009, 2014

Appendix B- Descriptive Statistics

Table 2: Descriptive statistics, International Elections

Statistic	N	Mean	St. Dev.	Min	Max
Percent Vote in Runoff	128	54.4	9.9	30.7	82.2
Victory (dummy)	128	0.7	0.5	0	1
Percent Vote in 1st Round	128	37.7	7.6	19.9	49.8
P.p. diff between 1st and 2nd in 1st round	128	10.2	7.3	0.2	32.3
Percent Vote of 3rd candidate in 1st Round	128	16.4	5.7	1.3	31.8

Table 3: Descriptive statistics, Brazil 1998-2018

Statistic	N	Mean	St. Dev.	Min	Max
Year	287	2,009.0	6.1	1,998	2,018
Percent Vote in Runoff	287	54.3	7.7	27.0	85.4
Victory (dummy)	287	0.8	0.4	0	1
Percent Vote in 1st Round	287	40.7	6.4	20.5	50.0
P.p. diff between 1st and 2nd in 1st round	287	9.7	7.1	0.1	33.6
Number of Candidates in 1st round	287	7.1	2.2	3	16
Percent Vote of 3rd candidate in 1st Round	287	16.5	6.3	0.7	30.0
Abstention Rate in 1st Round	287	19.9	5.3	6.1	49.9
Total Percent of Losing Candidates in 1st round	287	28.3	11.5	1.0	59.3
Age of Front-runner	286	51.8	10.2	27.0	83.0
Difference of Age between 1st and 2nd	286	0.05	14.3	−41.0	36.0
Incumbent (dummy)	245	0.1	0.2	0.0	1.0
In executive office (dummy)	245	0.1	0.3	0.0	1.0

Appendix C- Alternative Specifications of Regression Models

Table 4: Model Specification and Goodness-of-fit

Model	# of IVs	# of observations	Adusted-R2	PCP
Original	3	287	0.139	75.3
+ Type of office	5	287	0.138	75.3
+ Number of candidates	6	287	0.135	75.3
% Vote 3rd candidate	6	287	0.136	75.3
+ % Abstention	7	287	0.143	76.3
+ Sex of Candidates	10	287	0.14	76
+ Age of Candidates	11	286	0.147	75.5
+ Education of Candidates	14	286	0.148	76.9
+ Largest expenditure	15	285	0.145	76.5
+ Incumbency of candidates	18	285	0.14	76.5

Table 5: Alternative Specifications 1, Brazil 1998-2018

	Pct. Valid Votes in Runoff	Victory (dummy)
Pct. Valid Votes in 1st Round	-0.011	0.065^{*}
	(0.086)	(0.031)
Pct.-Point Difference 1st and 2nd Candidates	0.420^{**}	0.061^{*}
	(0.077)	(0.027)
Race- Mayoral	1.212	0.096
	(1.000)	(0.331)
Race- Presidential	1.905	14.889
	(3.310)	(1,042.620)
Pct. Valid Vote * Pct.-Point Difference	-0.007	0.005
	(0.011)	(0.004)
Constant	53.567^{**}	1.062^{**}
	(0.862)	(0.285)
N	287	287
Adjusted R^2	0.138	

* p<0.05; ** p<0.01

Table 6: Alternative Specifications 2, Brazil 1998-2018

	Pct. Valid Votes in Runoff	Victory (dummy)
Pct. Valid Votes in 1st Round	−0.024	0.076*
	(0.096)	(0.034)
Pct.-Point Difference 1st and 2nd Candidates	0.429**	0.054
	(0.082)	(0.028)
Race- Mayoral	1.165	0.129
	(1.011)	(0.334)
Race- Presidential	2.075	14.724
	(3.355)	(1,047.108)
Number of Candidates in 1st Round	−0.070	0.058
	(0.213)	(0.073)
Pct. Valid Vote * Pct.-Point Difference	−0.008	0.006
	(0.012)	(0.004)
Constant	54.101**	0.628
	(1.840)	(0.617)
N	287	287
Adjusted R^2	0.135	

* p<0.05; ** p<0.01

Table 7: Alternative Specifications 3, Brazil 1998-2018

	Pct. Valid Votes in Runoff	Victory (dummy)
Pct. Valid Votes in 1st Round	−0.039	0.055
	(0.106)	(0.037)
Pct.-Point Difference 1st and 2nd Candidates	0.435**	0.067*
	(0.083)	(0.029)
Race- Mayoral	1.168	0.078
	(1.006)	(0.333)
Race- Presidential	1.921	14.895
	(3.315)	(1,039.725)
Pct. of 3rd Candidate in 1st Round	−0.042	−0.015
	(0.091)	(0.030)
Pct. Valid Vote * Pct.-Point Difference	−0.007	0.005
	(0.011)	(0.004)
Constant	53.586**	1.072**
	(0.864)	(0.286)
N	287	287
Adjusted R^2	0.136	

* p<0.05; ** p<0.01

Table 8: Alternative Specifications 3, Brazil 1998-2018

	Pct. Valid Votes in Runoff	Victory (dummy)
Pct. Valid Votes in 1st Round	−0.002	0.069
	(0.107)	(0.038)
Pct.-Point Difference 1st and 2nd Candidates	0.423**	0.063*
	(0.083)	(0.030)
Race- Mayoral	1.855	0.320
	(1.068)	(0.357)
Race- Presidential	1.853	14.846
	(3.301)	(1,040.941)
Pct. of 3rd Candidate in 1st Round	−0.019	−0.005
	(0.091)	(0.031)
Pct. Rate of Absention in 1st Round	0.160	0.058
	(0.086)	(0.031)
Pct. Valid Vote * Pct.-Point Difference	−0.007	0.005
	(0.011)	(0.005)
Constant	53.088**	0.920**
	(0.901)	(0.296)
N	287	287
Adjusted R^2	0.143	

* p<0.05; ** p<0.01

Table 9: Alternative Specifications 4, Brazil 1998-2018

	Pct. Valid Votes in Runoff	Victory (dummy)
Pct. Valid Votes in 1st Round	0.007	0.072
	(0.108)	(0.038)
Pct.-Point Difference 1st and 2nd Candidates	0.413**	0.061*
	(0.083)	(0.030)
Race- Mayoral	1.898	0.310
	(1.073)	(0.363)
Race- Presidential	2.336	15.261
	(3.356)	(1,007.433)
Pct. of 3rd Candidate in 1st Round	−0.013	−0.005
	(0.092)	(0.031)
Pct. Rate of Absention in 1st Round	0.159	0.053
	(0.087)	(0.032)
(Front-runner) Female v. Female	−5.036	−17.237
	(7.135)	(2,399.545)
(Front-runner) Female v. Male	−0.900	−0.817
	(1.745)	(0.559)
(Front-runner) Male v. Female	1.547	0.638
	(1.429)	(0.580)
Pct. Valid Vote * Pct.-Point Difference	−0.007	0.005
	(0.012)	(0.005)
Constant	52.975**	0.949**
	(0.939)	(0.313)
N	287	287
Adjusted R^2	0.140	

* $p<0.05$; ** $p<0.01$

Table 10: Alternative Specifications 5, Brazil 1998-2018

	Pct. Valid Votes in Runoff	Victory (dummy)
Pct. Valid Votes in 1st Round	−0.010	0.067
	(0.108)	(0.038)
Pct.-Point Difference 1st and 2nd Candidates	0.422**	0.064*
	(0.083)	(0.031)
Race- Mayoral	2.050	0.288
	(1.077)	(0.370)
Race- Presidential	2.802	15.404
	(3.355)	(992.012)
Pct. of 3rd Candidate in 1st Round	−0.020	−0.010
	(0.091)	(0.031)
Pct. Rate of Absention in 1st Round	0.156	0.048
	(0.087)	(0.031)
(Front-runner) Female v. Female	−3.998	−16.952
	(7.133)	(2,399.545)
(Front-runner) Female v. Male	−1.072	−0.910
	(1.743)	(0.567)
(Front-runner) Male v. Female	1.629	0.662
	(1.425)	(0.584)
Age difference	−0.057	−0.016
	(0.030)	(0.010)
Pct. Valid Vote * Pct.-Point Difference	−0.007	0.005
	(0.011)	(0.005)
Constant	52.869**	0.995**
	(0.945)	(0.321)
N	286	286
Adjusted R^2	0.147	

* $p<0.05$; ** $p<0.01$

Table 11: Alternative Specifications 6, Brazil 1998-2018

	Pct. Valid Votes in Runoff	Victory (dummy)
Pct. Valid Votes in 1st Round	−0.036	0.062
	(0.109)	(0.039)
Pct.-Point Difference 1st and 2nd Candidates	0.432**	0.071*
	(0.083)	(0.031)
Race- Mayoral	2.026	0.295
	(1.078)	(0.373)
Race- Presidential	3.490	15.551
	(3.387)	(993.790)
Pct. of 3rd Candidate in 1st Round	−0.028	−0.013
	(0.092)	(0.031)
Pct. Rate of Absention in 1st Round	0.156	0.049
	(0.087)	(0.032)
(Front-runner) Female v. Female	−3.660	−16.719
	(7.138)	(2,399.545)
(Front-runner) Female v. Male	−0.904	−0.864
	(1.745)	(0.573)
(Front-runner) Male v. Female	1.819	0.786
	(1.430)	(0.597)
Age difference	−0.069*	−0.021
	(0.031)	(0.011)
(Front-runner) College v. Less than College	1.139	0.552
	(1.253)	(0.460)
(Front-runner) Less than College v. College	−1.217	−0.046
	(1.301)	(0.443)
(Front-runner) Less than College v. Less than College	2.644	1.631
	(2.243)	(1.127)
Pct. Valid Vote * Pct.-Point Difference	−0.010	0.004
	(0.012)	(0.005)
Constant	52.798**	0.891**
	(0.971)	(0.329)
N	286	286
Adjusted R^2	0.148	

* $p<0.05$; ** $p<0.01$

Table 12: Alternative Specifications 7, Brazil 1998-2018

	Pct. Valid Votes in Runoff	Victory (dummy)
Pct. Valid Votes in 1st Round	−0.035	0.062
	(0.109)	(0.039)
Pct.-Point Difference 1st and 2nd Candidates	0.432**	0.072*
	(0.084)	(0.031)
Race- Mayoral	2.001	0.254
	(1.098)	(0.376)
Race- Presidential	3.482	15.577
	(3.400)	(980.627)
Pct. of 3rd Candidate in 1st Round	−0.028	−0.013
	(0.092)	(0.032)
Pct. Rate of Absention in 1st Round	0.155	0.047
	(0.088)	(0.032)
(Front-runner) Female v. Female	−3.660	−16.717
	(7.165)	(2,399.545)
(Front-runner) Female v. Male	−0.906	−0.857
	(1.752)	(0.571)
(Front-runner) Male v. Female	1.808	0.754
	(1.438)	(0.600)
Age difference	−0.069*	−0.020
	(0.031)	(0.011)
(Front-runner) College v. Less than College	1.151	0.559
	(1.261)	(0.460)
(Front-runner) Less than College v. College	−1.210	−0.028
	(1.307)	(0.443)
(Front-runner) Less than College v. Less than College	2.674	1.684
	(2.261)	(1.134)
Difference in Largest Campaign Expense	−0.00000	−0.00000
	(0.00000)	(0.00000)
Pct. Valid Vote * Pct.-Point Difference	−0.010	0.004
	(0.012)	(0.005)
Constant	52.815**	0.915**
	(0.982)	(0.332)
N	285	285
Adjusted R^2	0.145	

* p<0.05; ** p<0.01

Table 13: Alternative Specifications 8, Brazil 1998-2018

	Pct. Valid Votes in Runoff	Victory (dummy)
Pct. Valid Votes in 1st Round	−0.018	0.059
	(0.112)	(0.040)
Pct.-Point Difference 1st and 2nd Candidates	0.419**	0.076*
	(0.086)	(0.032)
Race- Mayoral	2.436	0.489
	(1.252)	(0.414)
Race- Presidential	3.808	15.887
	(3.463)	(974.875)
Pct. of 3rd Candidate in 1st Round	−0.012	−0.011
	(0.094)	(0.033)
Pct. Rate of Absention in 1st Round	0.161	0.051
	(0.089)	(0.032)
(Front-runner) Female v. Female	−3.829	−16.860
	(7.257)	(2,399.545)
(Front-runner) Female v. Male	−0.748	−0.870
	(1.764)	(0.581)
(Front-runner) Male v. Female	1.835	0.788
	(1.444)	(0.601)
Age difference	−0.066*	−0.021
	(0.031)	(0.011)
(Front-runner) College v. Less than College	1.209	0.565
	(1.269)	(0.462)
(Front-runner) Less than College v. College	−1.180	−0.010
	(1.312)	(0.447)
(Front-runner) Less than College v. Less than College	2.681	1.597
	(2.272)	(1.151)
Difference in Largest Campaign Expense	−0.000	−0.00000
	(0.00000)	(0.00000)
(Front-runner) Not Incumbent v. Not Incumbent	−0.159	−1.817
	(2.260)	(1.132)
(Front-runner) Not Incumbent v. Incumbent	3.090	−1.748
	(3.263)	(1.391)
(Front-runner) NA Incumbent v. NA Incumbent	−0.058	−1.645
	(2.426)	(1.175)
Pct. Valid Vote * Pct.-Point Difference	−0.009	0.004
	(0.012)	(0.005)
Constant	52.496**	2.479*
	(2.063)	(1.096)
N	285	285
Adjusted R^2	0.140	

* $p<0.05$; ** $p<0.01$

Table 14: Alternative Specifications without 1st Round Results

Model	# of IVs	# of observations	Adusted-R2
Original	3	287	0.139
Type of office	2	287	0.007
+ Number of candidates	3	287	0.004
% Vote 3rd candidate	3	287	0.003
+ % Rate of Abstention	6	287	0.005
+ Sex of Candidates	6	287	0.005
+ Age of Candidates	7	286	0.011
+ Education of Candidates	10	286	0.014
+ Party of front-runners	33	286	0.006
+ Largest expenditure	11	285	0.011
+ Incumbency of candidates	14	285	0.008
In executive office	15	285	0

Appendix D- Goodness-of-fit without First Round Results

Appendix E- Predictions with 3rd-placed Candidate's Vote Share

Table 15: Models with Vote Share of 3rd Candidate, International Elections

	Pct. Valid Votes in Runoff	Victory (dummy)
Pct. Valid Votes in 1st Round	0.085	0.090
	(0.155)	(0.048)
Pct.-Point Difference 1st and 2nd Candidates	0.451**	0.086*
	(0.155)	(0.043)
Pct. of 3rd Candidate in 1st Round	−0.261	−0.029
	(0.150)	(0.039)
Pct. Valid Vote * Pct.-Point Difference	0.004	0.005
	(0.019)	(0.007)
Constant	54.648**	1.050**
	(1.001)	(0.263)
N	128	128
Adjusted R^2	0.150	

* p<0.05; ** p<0.01

Table 16: Models with Vote Share of 3rd Candidate, Brazil 1998-2018

	Pct. Valid Votes in Runoff	Victory (dummy)
Pct. Valid Votes in 1st Round	−0.058	0.058
	(0.104)	(0.036)
Pct.-Point Difference 1st and 2nd Candidates	0.450**	0.069*
	(0.082)	(0.029)
Pct. of 3rd Candidate in 1st Round	−0.051	−0.015
	(0.090)	(0.030)
Pct. Valid Vote * Pct.-Point Difference	−0.005	0.006
	(0.011)	(0.004)
Constant	54.432**	1.145**
	(0.482)	(0.162)
N	287	287
Adjusted R^2	0.137	

* p<0.05; ** p<0.01

As indicated in the third section, we find that the variance around the percentage of votes for the leading candidate in the Brazilian runoffs is larger when the third candidate receives a higher percentage. In other words, as the percentage of votes for the third-placed candidate increases, the variation in electoral results tends to be larger, thus indicating more homogeneity among the remaining electorate when the third-placed candidate received more votes. To show this, we plot the relationship between the vote percentage of the third-placed candidate and the residuals from the regression model predicting the front-runner second-round vote share after controlling for the first-round of the candidates who moved through to the runoff. The dashed line represents two standard deviations of the mean residual per decile.

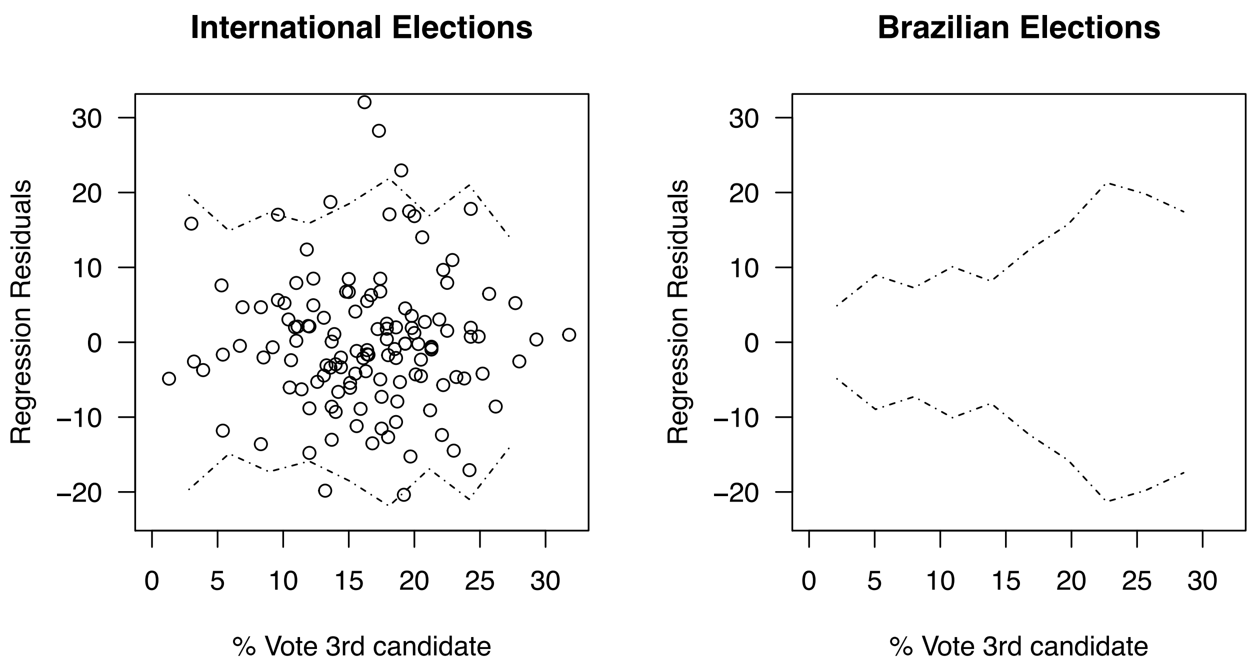

Figure 1: Scatterplot of Regression Residuals and Vote Share of Third-placed Candidate

The results indicate that “undecided” voters do not choose the trailing candidate more often, even in less fragmented scenarios. But the graph on the right suggests that a larger vote share for the third-placed candidate is associated with a more diversified set of results in the second round- i.e. larger wins and losses. The analysis of the international elections does not show the same pattern, however.

Appendix F- Regressions Models without the 2018 Elections

Table 17: Original Models in the Text, Brazil 1998-2018

	Pct. Valid Votes in Runoff	Victory (dummy)
Pct. Valid Votes in 1st Round	-0.024	0.067^{*}
	(0.084)	(0.031)
Pct.-Point Difference 1st and 2nd Candidates	0.433^{**}	0.063^{*}
	(0.076)	(0.027)
Pct. Valid Vote * Pct.-Point Difference	-0.006	0.005
	(0.011)	(0.004)
Constant	54.446^{**}	1.148^{**}
	(0.481)	(0.162)
N	287	287
Adjusted R^2	0.139	

* p<0.05; ** p<0.01

Table 18: Models Used to Predict the 2018 Elections, Brazil 1998-2016

	Pct. Valid Votes in Runoff	Victory (dummy)
Pct. Valid Votes in 1st Round	-0.039	0.065^{*}
	(0.085)	(0.031)
Pct.-Point Difference	0.407^{**}	0.059^{*}
	(0.078)	(0.027)
Pct. Valid Vote * Pct.-Point Difference	-0.003	0.006
	(0.011)	(0.004)
Constant	54.414^{**}	1.100^{**}
	(0.486)	(0.165)
N	272	272
Adjusted R^2	0.125	

* p<0.05; ** p<0.01

Appendix G- Intervals Calculated with Bootstrapping

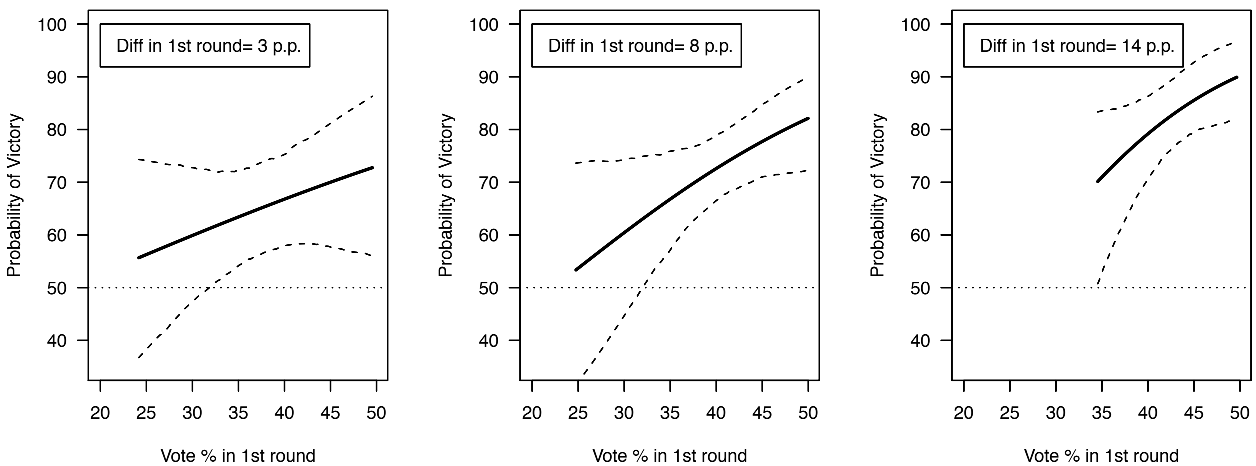

Figure 2: Predicted Probability of Winning, Brazil 1998-2018