THE CAMPBELL COLLECTION OF PRESIDENTIAL ELECTION FORECASTS, 1984-2016. A Review[1]

Alfred G. Cuzán

8/21/2020

Abstract

For three decades, political scientists have developed and refined polimetric models of presidential elections to forecast their outcomes, usually as the incumbent's or the Democrat's share of the two-party vote. Most of these forecasts have appeared in previous symposia published in this journal. In this article, I examine the performance of the modelers, not the models, because their creators revise or replace them in light of experience. I show that across the 1984-2016 period the forecasters as a group have performed remarkably well. Moreover, their accuracy has improved over time. Also, there is some evidence of learning among the modelers.

[1] An earlier version presented at the American Political Science Association, Washington, D.C., 2019.

THE CAMPBELL COLLECTION OF PRESIDENTIAL ELECTION FORECASTS, 1984-2016. A Review

Alfred G. Cuzán

For three decades, political scientists have developed and refined polimetric models of presidential elections composed of a variety of inputs and methods to forecast the outcome of American presidential elections. Most forecasts have appeared in *PS: Political Science and Politics* in quadrennial symposia edited by James E. Campbell.[1] Thus, I have taken the liberty to name the collection after him. In this brief article, I evaluate the relative accuracy of this work, looking at the performance first of the forecasts as a group and then of the individual forecasters—the forecasters or modelers, not the models, because their authors revised or discard and replace them over time.[2] This permits the tracking of any learning that may have taken place among the forecasters.

The Campbell Collection

Thirteen scholars who designed polimetric models to forecast the incumbent party's share of the two-party vote for president appeared at least once in one of the symposia edited by James E. Campbell are included. To be considered a member of the collection, not every forecaster's model need have appeared in every symposium, as long as at least one did. Two forecasters appeared only once, and two, twice. On the other hand, nine appear in all or nearly all of them. Altogether, more than 65 forecasts are publicly available for analysis. If in any one year a forecaster posted more than one forecast or used more than one model in forecasting the election, those forecasts were averaged, unless the author indicated that one of them was the preferred one. This reduced the number of forecasts to 63. To simplify labeling the figures, where the forecast was co-authored by two or more forecasters, the name of the author that appears in all of the forecasts is used. If all names appear every time, then in alphabetical order.

Hypotheses

Findings extracted from work on the PollyVote project (PollyVote.com) supplies the rationale for the following hypotheses:

H1: The more forecasts are included, and the greater the variety of inputs and methods, the lower the expected error of the ensemble forecast.

Rationale: Since the number of modelers and the diversity of approaches has gone up with time to include state-level predictors, not just national ones (e.g., Klarner 2012, Jerome 2016), the collective forecast is expected to have become more accurate in more recent elections. The intuition is that in constructing their models forecasters rely on different, even if overlapping sets of information (Armstrong 2001, Graefe et al. 2014, 2017; Graefe, Green and Armstrong, 2019). Thus, the more forecasts, the larger the quantity of information captured in the ensemble forecast, and the higher the chance for bracketing the actual value, which makes for greater accuracy in forecasting. Ironically, errors, even large ones, may help, rather than hinder accuracy as long as they occur on opposite sides of the actual.[3]

H2: No one forecaster will emerge consistently as the most accurate across all elections.

Rationale: Again, this has to do with the different slices of information captured by various models. The circumstances in which every election takes place make it, in some sense, unique. Thus, different models will fit some elections and not others, depending on context.

Analysis

Table 1 displays the descriptive statistics—the mean, median, and standard deviation of the population, as the collection represents a universe, not a sample—of four variables: the forecasts, the election outcome (the actual incumbent share of the two-party vote), the arithmetic error, and the absolute error. Remarkably, the average ensemble forecast is within one percent point of the actual. The mean (MAE) and median (MdAE) of the absolute error are under 3 percent points, within range of the two-point error in the final pre-election poll taken in late October or early November for the same elections,[4] even though most of the forecasts were issued months in advance of the event.

. . . .

Table 1 about here

. . . .

Figure 1 shows that as the number of forecasts has increased, the average error has fallen. Observe, in table 2, that across all years the Pearson's correlation coefficient measuring the relationship between the number of forecasts and the average absolute error is weak over the entire period but strengthens substantially after the first three elections are dropped. Additionally, figure 2 reveals that in more recent years the forecasts bracket the actual value, something that goes a long way to account for the greater accuracy of ensemble forecasts. Thus, H1 receives empirical support: as the number of forecasts rose, they bracketed the actual value more frequently, making for greater accuracy.

. . . .

Table 2 and Figures 1 and 2 about here

. . . .

Turning to the modelers, figure 3 arranges the median absolute error of their forecasts from left (highest) to right (lowest). The median is used because with so few observations outliers exert a distorting effect. Along with the MdAE, the number of forecasts made by each member of the Collection is shown. Among those with five or more forecasts, Abramowitz, Campbell, Lewis-Beck, and Erikson and Wlezien have the best overall records for accuracy. However, as Figure 4 demonstrates, this does not mean any one model comes closest to the actual year after year. In fact, in some cases the forecaster's model incurred the largest error only to win the gold an election or two later (e.g., Lewis-Beck 1992-96, Campbell 1996-2000)—or vice-versa (Cuzán 2004-2012). Hypothesis 2, then, also finds support: no one modeler arrives first (or last) every time.

. . . .

Figures 3 and 4 about here

. . . .

Finally, figure 5 displays the absolute error by forecaster by year. This allows asking whether any learning has taken place, making for better forecasting performance. It appears that this is the case with Holbrook, Lewis-Beck, and Lockerbie. In all three, following an early outlier, the error of their respective forecasts has progressively shrunk, and is now among the lowest. No other forecaster shows such dramatic improvement, although Campbell, and Erikson and Wlezien did recover nicely from the 2008 and 2000 outliers, respectively. In Cuzán's case, however, the trend appears to go the other way.[5] For Abramowitz, Hibbs, and Nortpoth the error is trendless. Of course, in the case of Abramowitz's model, in which the error has consistently remained below

3.0, little learning was called for in the first place. Also, as Wink (2018-19, 206) points out, Abramowitz and Norpoth are the only ones to share the honor of having correctly called the winner in all three of the most recent elections, even as the latter's error is about half again as large as that of the former. In any case, although outside the scope of this review, studying the changes Holbrook, Lewis-Beck, and Lockerbie made in their models may well be instructive.

. . . .

Figure 5 about here

. . . .

Discussion

The foregoing analysis lends empirical support for both of the hypotheses set at the outset. As the Campbell Collection grew in numbers and diversity of inputs and methods, the ensemble forecast became more accurate, and it did so because, with more models, it became more likely that at least two members of the group would bracket the actual value. Also, no one modeler turned out to be consistently the best performer across all elections, although in estimating the vote Abramowitz, Campbell, and Lewis-Beck had the most first-place finishes, with Erikson and Wlezien almost always at their heels, and Norpoth rating highly on correctly calling the winner. The lesson is that one should not place too much confidence in any one model, but in the group mean or median forecast, a recommendation that is in keeping with the findings of the PollyVote Project. Given the number and variety of models lined up in this Symposium, however, there is one prediction I am willing to make: the error of the vote forecast will be, if not the lowest, among the lowest on record.

Conclusion

As a group, the Campbell Collection of presidential elections forecasting models performs remarkably well over the 1984-2016 period. It may not do as well as other methods in any one election (Graefe et al., 2014, 2017), but across the entire period the record is impressive. Moreover, the ensemble forecast has become more accurate. Ironically, according to one study, the experts assembled in the quadrennial PollyVote Project appear to be more influenced by polls than by the models (Graefe 2018; see, also, Jones and Cuzán 2013). Although peering into the future is always fraught with risk, the track record of the Campbell Collection's ensemble forecast demonstrates that its forecasts are not to be discounted any more, or at least not much more, than the polls.

Acknowledgments

As always, I am grateful to the University of West Florida for its generous support and benign academic environment. For their comments, suggestions, or encouragement, I thank Randall J. Jones, Jr. (who also found some of the early forecasts), Carl Klarner, Kyle Kondik, and Christopher Wlezien, as well as anonymous reviewers, and the editors of this journal and of this symposium in particular. Also, I am appreciative of the work done on the PollyVote project in collaboration with J. Scott Armstrong, Randall J. Jones, Jr., and Andreas Graefe. Finally, the usual disclaimer applies: all errors of fact or interpretation are my own.

REFERENCES

Abramowitz, Alan I. 1988. An Improved Model for Predicting Presidential Election Outcomes. *PS: Political Science and Politics*, 21(4): 843-847.

Abramowitz, Alan I. 1994. Is It Time for Them to Go? Forecasting the 1992 Presidential Election. *Political Methodologist*, 5(2): 2-3.

Armstrong, J. Scott. 2001. Combining forecasts. In J. S. Armstrong (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Norwell: Kluwer, 417-439.

Campbell, James E. 1994. Forecasts of the 1992 Presidential Election. *Political Methodologist,* 5(2): 5-9.

Campbell, James E. and Thomas E. Mann. 1992. Forecasting the 1992 Election: A User's Guide to the Models. *Brookings Review,* 10(4) 22-27.

Campbell, James E. and James C. Garand, Eds. 2000. *Before the Vote: Forecasting American National Elections*. Thousand Oaks, CA: Sage Publications.

Campbell, James E. 2001. The Referendum that Didn't Happen. The Forecasts of the 2000 Presidential Election. *PS: Political Science and Politics,* 34(1): 33-38.

Campbell, James E. 2004. Introduction—The 2004 Presidential Election Forecasts. *PS: Political Science and Politics,* 37 (4): 733-35.

Campbell, James E. 2008. Editor's Introduction: Forecasting the 2008 National Elections. *PS: Political Science and Politics,* 41(4): 679-82.

Campbell, James E. 2012. Forecasting the 2012 American National Elections. *PS: Political Science and Politics*, 45(4): 610-13.

Campbell, James E. 2016. Forecasting the 2016 American National Elections. *PS: Political Science and Politics*. 49(4): 649-654.

Cuzán, Alfred G. and Charles M. Bundrick. 1999. Fiscal Policy as a Forecasting Factor in Presidential Elections. *American Politics Quarterly,* 27(3): 338-53.

Cuzán, Alfred G. and Charles M. Bundrick. 2005. Deconstructing the 2004 Presidential Election Forecasts: The Fiscal Model and the Campbell Collection Compared. *PS: Political Science and Politics,* 35(2): 255-62.

Cuzán, Alfred G. 2019. Five Laws of Politics: A Follow-Up. *PS: Political Science and Politics,* 52(3): 457-64.

Cuzán, Alfred G. and Richard J. Heggen. 1984. A Fiscal Model of Presidential Elections in the United States: 1880-1980. *Presidential Studies Quarterly*, XIV(1): 98-108.

Fair, Ray. 2009. Presidential and Congressional Vote-Share Equations. *American Journal of Political Science*, 53(1): 55-72.

Graefe, Andreas, J. Scott Armstrong, Randall J. Jones, Jr., and Alfred G. Cuzán. 2014. Combining Forecasts: An Application to Elections. *International Journal of Forecasting,* 30(1), 43-54.

Graefe, Andreas, J. Scott Armstrong, Randall J. Jones, Jr., and Alfred G. Cuzán. 2017. Assessing the 2016 U.S. Presidential Election Popular Vote Forecasts. In A. Cavari, R Powell, and K. Mayer, Eds., *The 2016 Presidential Election: The Causes and Consequences of a Political Earthquake.* Lanham, MD: Lexington Books, 137-158.

Graefe, Andreas. 2018. Predicting Elections: Experts, Polls, and Fundamentals. *Judgment and Decision Making,* 13(4): 334-344.

Graefe, Andreas, Kesten C. Green, and J. Scott Armstrong. 2019. Accuracy Gains from Conservative Forecasting: Tests Using Variations of 19 Econometric Models to Predict 154 Elections in 10 Cuntries. PLOS One, 14(1): e0209850.

Jones, Randall J., Jr. 2008. The State of Presidential Election Forecasting: The 2004 Experience. *International Journal of Forecasting*, 24(2): 308-319.

Jones, Randall J., Jr., and Alfred G. Cuzán. 2009. Forecasting Performance of Regression Models in the 2008 Presidential Election. *Foresight*. 12(4): 41-42.

Jones, Randall J., Jr., and Alfred G. Cuzán. 2013. Expert Judgment in Forecasting American Presidential Elections: A Preliminary Evaluation. Paper presented at the 2013 meeting of the American Political Science Association, Chicago. Available at SSRN: https://ssrn.com/abstract=2303052.

Klarner, Carl E. 2012. State-Level Forecasts of the 2012 Federal and Gubernatorial Elections. *PS: Political Science and Politics*, 45(4): 655-662.

Leip, David. "United States Presidential Election Results." Atlas of U.S. Presidential Elections. https://uselectionatlas.org/RESULTS/ (August 19, 2020).

Lewis-Beck, Michael S. 1985. Elections Forecasts in 1984: How Accurate Were They? *PS: Political Science and Politics*, 18 (1): 53-62.

Lewis-Beck, Michael S. and Mary Stegmaier. 2014. Weather, Elections, Forecasts: After Richardson. *PS: Political Science and Politics*, 47(2): 322-325.

PollyVote Project. Evidence-Based Election Forecasting Since 2004. Combining Forecasts. https://pollyvote.com/en/

Wink, Kenneth A. 2018-19. Forecasting Models and the Presidential Vote. *Political Research Quarterly,* 134(2): 193-216.

Table 1. Forecasts of Incumbent Two-Party Vote, 1984-2016, Descriptive Statistics				
	Forecast	Actual	Error	Absolute Error
Mean	51.8	51.7	0.84	2.6
Median	51.6	51.2	0.80	2.5
S.D. (pop)	3.9	3.8	3.3	2.1
N=13 forecasters, 63 forecasts.				

Table 2. Absolute Error by Number of Forecasts

YEAR	N	Absolute Error	
		Mean	Median
1984	2	4.9	4.9
1988	1	2.7	2.7
1992	4	1.0	0.4
1996	8	2.3	2.5
2000	7	5.7	4.9
2004	8	2.3	2.6
2008	10	2.7	1.95
2012	13	2.2	1.5
2016	10	1.6	1.2

Correlation coefficients:
1984-2016: r=-0.27 (MAE); r=-0.38 (MdAE)
1996-2016: r=-0.54 (MAE) ; r=-0.74 (MdAE)

Figure 1. Median Absolute Error by Number of Forecasts

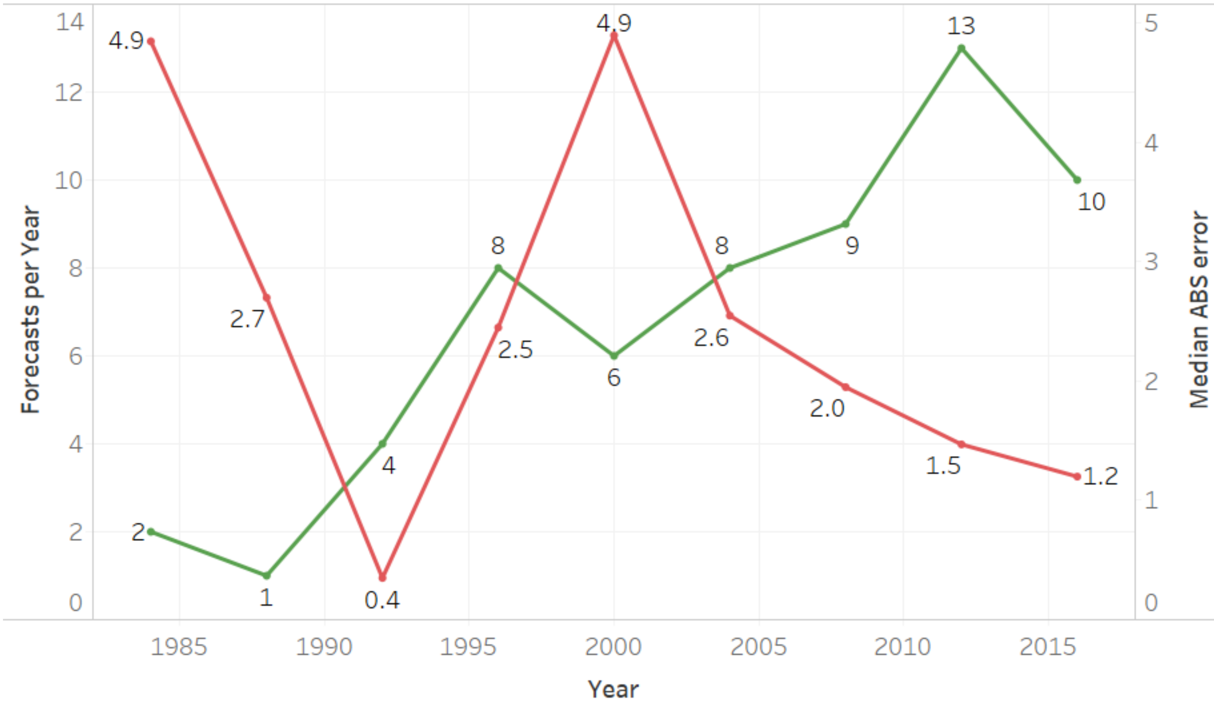
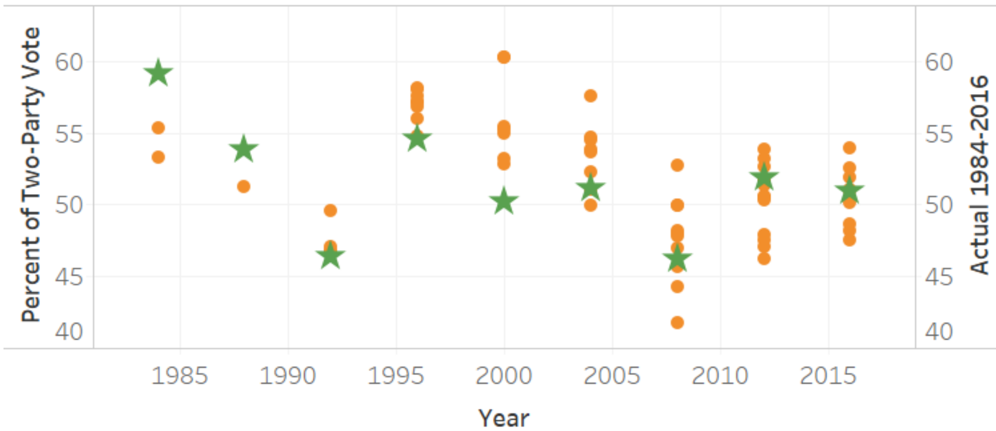
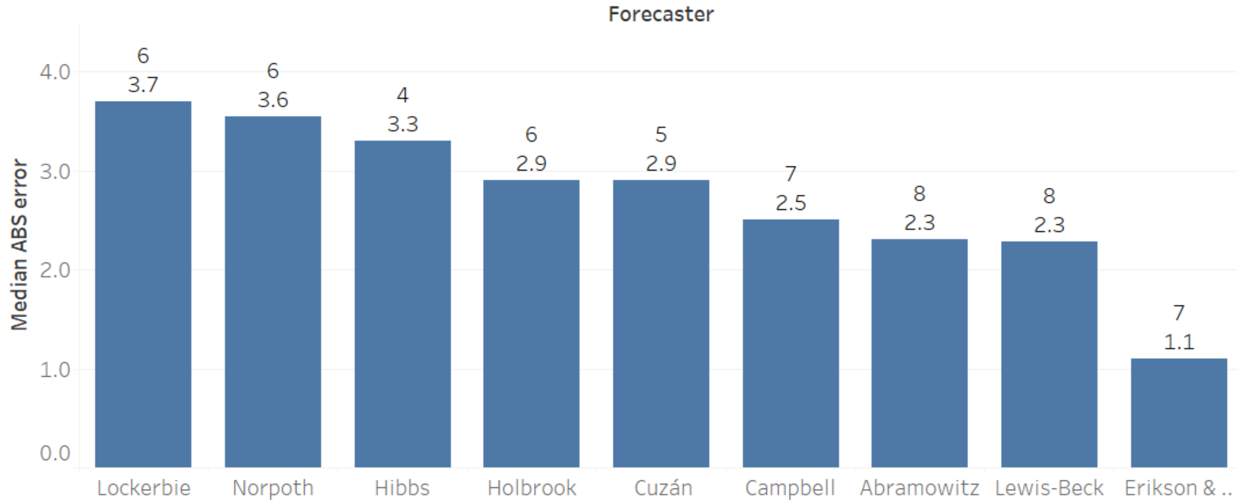


Figure 2. Incumbent Vote: Forecasts and Actual, by Year



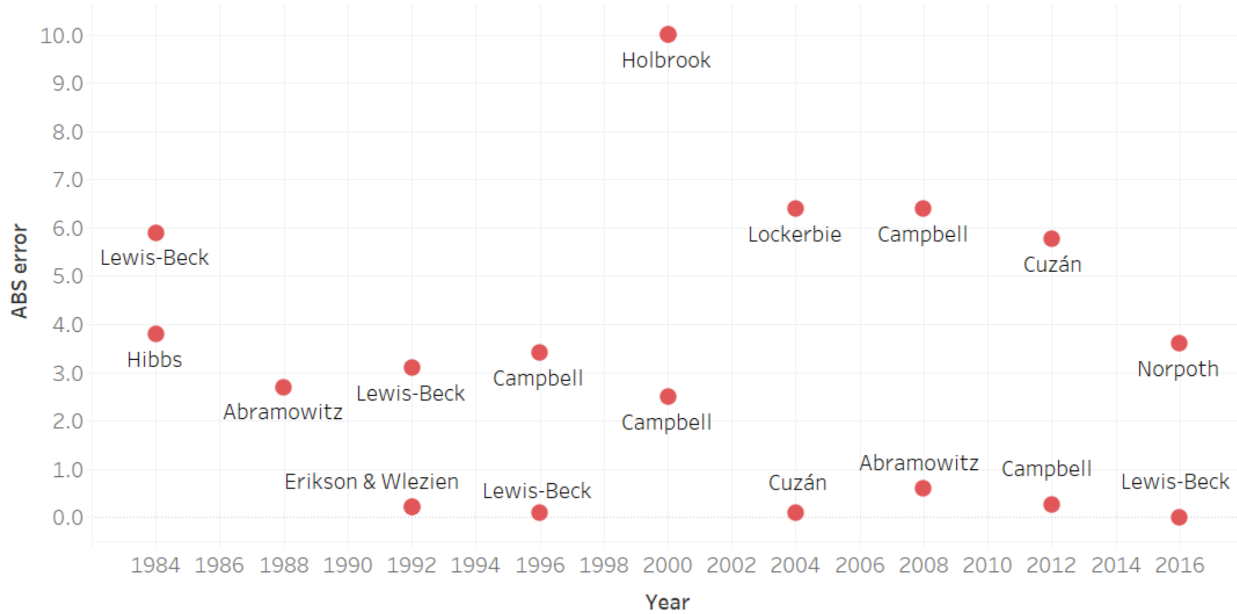
Yearly forecasts of the vote are shown by orange marks and the actual by a green star.

Figure 3. Median Absolute Error by Modeler and Number of Forecasts

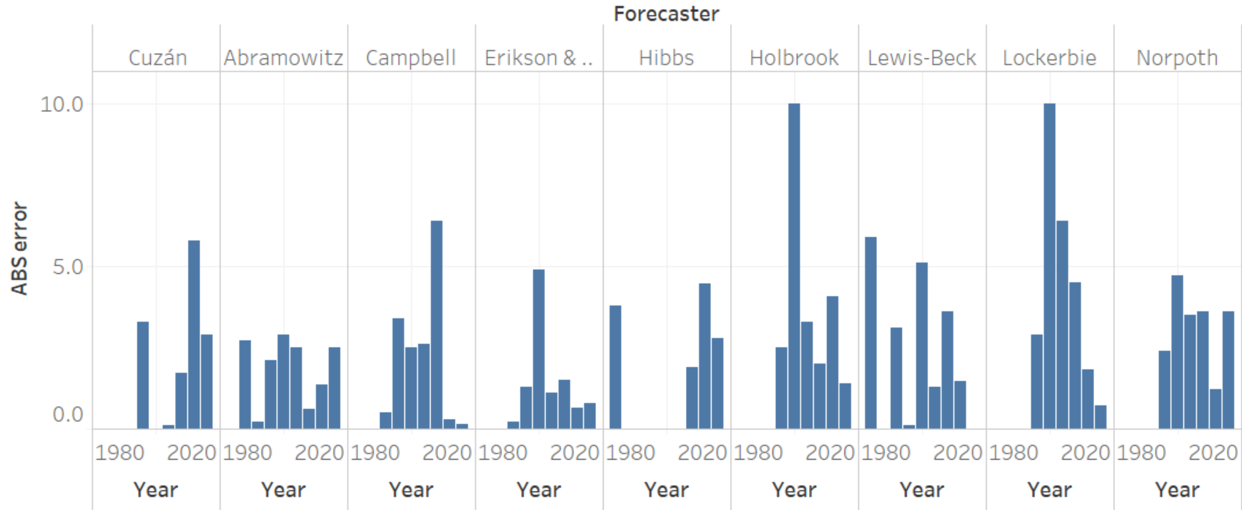


Only modelers with at least three forecasts are shown.

Figure 4. Lowest and Largest Absolute Error, by Year



Figue 5. Forecaster Error by Year, 1992-2016



Only modelers with at least three forecasts are shown.

Data Appendix					
Forecaster	Year	Actual	Forecast	Error	ABS Error
	1984	59.2			
Abramowitz	1988	53.9	51.2	-2.7	2.7
Abramowitz	1992	46.5	46.7	0.2	0.2
Abramowitz	1996	54.7	56.8	2.1	2.1
Abramowitz	2000	50.3	53.2	2.9	2.9
Abramowitz	2004	51.2	53.7	2.5	2.5
Abramowitz	2008	46.3	45.7	-0.6	0.6
Abramowitz	2012	52.0	50.6	-1.37	1.37
Abramowitz	2016	51.1	48.6	-2.5	2.5
Berrry and Bickers	2012	52.0	47.1	-4.87	4.87
Campbell	1992	46.6	47.1	0.5	0.5
Campbell	1996	54.7	58.1	3.4	3.4
Campbell	2000	50.3	52.8	2.5	2.5
Campbell	2004	51.2	53.8	2.6	2.6
Campbell	2008	46.3	52.7	6.4	6.4
Campbell	2012	52.0	51.7	-0.27	0.27
Campbell	2016	51.1	50.95	-0.15	0.15
Cuzán	1996	54.7	58	3.3	3.3[a]
Cuzán	2004	51.2	51.1	-0.1	0.1
Cuzán	2008	46.3	48	1.7	1.7
Cuzán	2012	52.0	46.2	-5.77	5.77
Cuzán	2016	51.1	48.2	-2.9	2.9
Erikson & Wlezien	1992	46.6	46.8	0.2	0.2
Erikson & Wlezien	1996	54.7	56	1.3	1.3
Erikson & Wlezien	2000	50.3	55.2	4.9	4.9
Erikson & Wlezien	2004	51.2	52.3	1.1	1.1
Erikson & Wlezien	2008	46.3	47.8	1.5	1.5
Erikson & Wlezien	2012	52.0	52.6	0.63	0.63
Erikson & Wlezien	2016	51.1	51.9	0.8	0.8
Hibbs	1984	59.1	55.3	-3.8	3.8
Hibbs	2008	46.3	48.2	1.9	1.9
Hibbs	2012	52.0	47.5	-4.47	4.47
Hibbs	2016	51.1	53.9	2.8	2.8
Holbrook	1996	54.7	57.2	2.5	2.5
Holbrook	2000	50.3	60.3	10	10
Holbrook	2004	51.2	54.5	3.3	3.3
Holbrook	2008	46.3	44.3	-2.0	2.0

Holbrook	2012	52.0	47.9	-4.07	4.07
Holbrook	2016	51.1	52.5	1.4	1.4
Jerome & Jerome	2012	52.0	51.6	-0.37	0.37
Jerome & Jerome	2016	51.1	50.1	-1.0	1.0
Klarner	2008	46.3	47	0.7	0.7
Klarner	2012	52.0	51.3	-0.67	0.67
Lewis-Beck	1984	59.2	53.3	-5.9	5.9
Lewis-Beck	1992	46.5	49.6	3.1	3.1
Lewis-Beck	1996	54.7	54.8	0.1	0.1
Lewis-Beck	2000	50.3	55.4	5.1	5.1
Lewis-Beck	2004	51.2	49.9	-1.3	1.3
Lewis-Beck	2008	46.3	49.9	3.6	3.6
Lewis-Beck	2012	52.0	50.5	-1.47	1.47
Lewis-Beck	2016	51.1	51.1	0.0	0.0
Lockerbie	1996	54.7	57.6	2.9	2.9
Lockerbie	2000	50.3	60.3	10	10
Lockerbie	2004	51.2	57.6	6.4	6.4
Lockerbie	2008	46.3	41.8	-4.5	4.5
Lockerbie	2012	52.0	53.8	1.83	1.83
Lockerbie	2016	51.1	50.4	-0.7	0.7
Montgomery, Hollenback & Ward	2012	52.0	50.3	-1.67	1.67
Norpoth	1996	54.7	57.1	2.4	2.4
Norpoth	2000	50.3	55	4.7	4.7
Norpoth	2004	51.2	54.7	3.5	3.5
Norpoth	2008	46.3	49.9	3.6	3.6
Norpoth	2012	52.0	53.2	1.23	1.23
Norpoth	2016	51.1	47.5	-3.6	3.6

[1] Campbell (1994, 2001, 2004, 2008, 2012, 2016); Campbell and Garand (2000); Campbell and Mann (1992). Some of the early forecasts appeared outside one of the Campbell symposia, e.g., Lewis-Beck (1985).

[2] For an earlier effort at such a comparison, see Jones and Cuzán 2008. Only as I was revising this article for publication did I find a similar effort by Wink (2018-19). Although we do not address very same set of questions, on those we do our findings are consistent.

[3] Although I do not think it was his intention, Wink (2018-19, 207) highlights a graphic illustration of the value of the principle. In 2008, forecasts of the McCain vote by Campbell and Lockerbie erred in opposite directions, the former overestimating it by 6.4 percent points and the latter falling short of the mark by 4.5 points. Averaging the two, however, yields an arithmetic error of 0.95.

[4] "Historical polling for United States presidential elections," *Wikipedia*, https://en.wikipedia.org/wiki/Historical_polling_for_United_States_presidential_elections.

[5] This could well be because Cuzán's "fiscal model" (Cuzán and Heggen 1984, Cuzán and Bundrick 1999, 2005), like Ray Fair's (2018), lacks one of the three elements of what Lewis-Beck and Stegmaier (2014, 323) call a "Core Political Economy Model" in election forecasting, namely, a measure of presidential popularity.