A Long-Range State-Level Forecast of the 2020 Presidential Election

Jay A. DeSart, *Utah Valley University*

In 2015, I addressed a significant shortcoming in the presidential-election forecasting literature: the lack of any longer-range state-level presidential-election forecasting models. I presented my results at the Iowa Conference on Presidential Politics in October of that year. The model used to generate this set of predictions was a simple and parsimonious one that produced forecasts for the state-level popular-vote outcome in each of the 50 states and the District of Columbia. I then extrapolated these outcomes to the national level to produce forecasts of both popular-vote and Electoral College outcomes a year in advance of the election.

In that presentation (DeSart 2015), I discussed the significant prospect of a Trump victory in 2016. The model suggested that Donald Trump would win 305 electoral votes to Hillary Clinton's 233.[1] It also projected that Trump would win a slim majority of the national two-party popular vote as well. The prediction was met with significant incredulity by those in attendance. I admit that I also was skeptical. Of course, now we have the luxury of looking back at this prediction and noting that whereas the point estimates of the national popular vote missed the mark, it at least provided an accurate forecast of who ultimately would win the election.

Using this model again to project an outcome for the 2020 election suggests a high probability that Joe Biden will win on November 3. To fully understand this forecast, we must examine its underlying inspiration and structure. The endeavor aims to address a significant challenge in presidential-election forecasting.

THE CHALLENGE OF A LONG-RANGE FORECAST

In forecasting, there is constant tension between two competing goals: *accuracy* and *lead time.* It makes sense that one of the key criteria on which we base our assessments of a forecast is how close its prediction matches the actual result it attempts to predict. However, as important as accuracy is *how far in advance* of the event of interest we can generate those predictions. A key purpose of any forecasting endeavor is to see accurately far enough into the future to reliably anticipate and prepare for future events. The key is finding the optimal balance between these two goals.

Most presidential-election forecast models have a lead time of approximately three months. The average lead time of the 11 forecasts that were featured in the 2016 *PS: Political Science & Politics* forecasting symposium was slightly more than three months (109.2 days) (Campbell 2016). The one outlier was Norpoth's Primary Model, which generated a forecast eight months in advance of the election (246 days) (Norpoth 2016).

Norpoth (2013) also has a model with an even longer lead time. His Time for Change model generates a prediction for the next election as soon as results of the current election are known, providing a four-year lead time. In 2013, Norpoth's Time for Change model projected a 61% probability that a Republican would win the 2016 election. His Primary Model predicted that Trump had an 87% certainty of defeating Clinton (Norpoth 2016).

As impressive as these lead times are, Norpoth's models have the same shortcoming as most academic presidential-election forecast models: they generate a forecast of only the national popular-vote outcome, not the Electoral College. In that sense, Norpoth's model was actually incorrect because Clinton won the national popular vote.

Only a model that produces state-level outcomes can generate reliable projections of Electoral College outcomes. A few recent models produce state-level predictions (Berry and Bickers 2012; DeSart and Holbrook 2003; Jerôme and Jerôme-Speziari 2016; and Klarner 2012). Although these models allow an extrapolation up to an Electoral College prediction, their lead times are significantly shorter than those in Norpoth's (2016) Primary Model. The longest lead time is the Jerôme and Jerôme-Speziari model, which generates a prediction roughly five months before the election—still at least three months after Norpoth's Primary Model. It was against this backdrop that I explored the possibility of a longer-range model that could generate both state- and national-level predictions of presidential-election outcomes (DeSart 2020) .

THE LONG-RANGE MODEL

Of course, one key reason why the state-level forecast models—as well as most national-level models—have lead times of only approximately three months is the availability of reliable predictor variables far in advance of the election. This creates certain structural limitations to creating longer-range forecasts of state-level outcomes. The DeSart and Holbrook (2003) state-level model used a state-level polling variable, which is not widely available for all 50 states before September. The lack of available and reliable data for all 50 states makes it challenging to build accurate longer-range models of state-level outcomes.

With the emphasis on favoring lead time, the choices of state-level indicators so far in advance of the election were relatively limited. Therefore, national-level indicators carry much of the burden in an attempt to capture the uniform shifts that may be presented from one election to the next. My efforts ultimately resulted in a fairly parsimonious model that

generates reasonably accurate forecasts a year in advance of the election. With the Democratic candidate's share of the two-party popular vote in each state, *i*, in each election, *j*, as the dependent variable, $RESULT_{ij}$, the model uses the following four predictor variables:

$RESULT_{i,j-1}$: This is the share of the two-party popular vote won by the Democratic candidate in each state, *i*, in the previous election. Pearson's r for state-level outcomes across

a one-election lag averages 0.94 for the period from 1992 to 2016, ranging from 0.91 for the 1992–1996 lag to 0.98 for the 2008–2012 lag. This establishes a baseline performance for each state.

$POLLS_j$: This is the average Democratic two-party share of national head-to-head polls taken 13 months in advance of the election, *j*. This was the real challenge and, consequently, the most important contribution to the model. I obtained data from various sources for polls conducted in October in the year before the election for each election going back to 1996. This variable is meant to capture, in part, the shift in the unique national-level context of each election year as distinct from the previous one.

Another challenge of using polling data this far in advance of the election is that the nominees for each party have not yet been determined. Therefore, polling data that pits each potential Democratic nominee against each potential Republican nominee is required as is generating a forecast for each potential matchup in a matrix. In October 2015, there were polling data for 11 possible matchups.[2] Despite coming more than a year in advance of the election, this variable was highly correlated with the national-level outcome for the six elections from 1996 to 2016 (Pearson's r=0.75).

$HOME STATE_i$: This is a dummy variable for each state indicating its status as a home state for each candidate, signed according to party (i.e., positive for a Democrat and negative for a Republican). However, because of the presence of the lagged *RESULT* variable in the model, an adjustment is made to account for the removal of the home-state advantage from the previous election. For example, the home-state "bump" that John Kerry received in Massachusetts in 2004 would no

reelection, as in 1996, 2004, and 2012, their home state receives a value of 0 to account for the fact that their advantage is already captured in the lagged *RESULT* variable.

$CONSECUTIVE TERMS_j$: This is a simple party-adjusted variable (i.e., positive for Democrats and negative for Republicans) that captures the number of terms a party has consecutively occupied the White House going into election *j*. Accounting for the now well-documented two-term penalty

(Abramowitz 1988) in this manner captures both the impact of this effect in any given election and the removal of said effect after the previous election.

IN-SAMPLE MODEL PERFORMANCE

Regressing the *RESULT* variable for each of the 50 states[3] from the six elections from 1996 to 2016 on these four predictors yields the results shown in table 1, along with various indicators of the accuracy of the model's in-sample estimates of each election outcome. All four variables achieve statistical significance at the 0.01 level, and together they explain 89% of the variance in election outcomes in each state across the six elections.

By simply examining the point estimates for each state in each year and declaring a "winner" based on which candidate wins a majority, the model correctly "predicted" the state-level outcomes 92% of the time across the six elections. This ranges from a low of 84% in 2000 to a perfect prediction of all 50 states in 2012.

The state-level predictions then can be extrapolated to the national level for both the popular vote and the Electoral College. If we ignore the uncertainty of the model's estimates by simply awarding each state's electoral votes to the model's projected popular-vote winner, we can derive a projected Electoral College outcome.

For the national popular-vote projection, the state-level predictions are weighted according to each state's contribution to the total national popular vote in the previous election and then summed. Noteworthy is that the average absolute error of this national-level extrapolation across the six elections in the analysis is approximately 1%. In terms of specific elections, the

> *Using this model again to project an outcome for the 2020 election suggests a high probability that Joe Biden will win on November 3.*

> *In forecasting, there is constant tension between two competing goals:* accuracy *and* lead time.

longer be present in 2008. Therefore, Massachusetts was coded 1 in 2008, in addition to Illinois being coded 1 to account for Barack Obama's expected home-state advantage. Likewise, in addition to Arizona being coded –1 to account for John McCain's expected boost in that state, Texas received a value of –1 to capture the loss of George W. Bush's home-state advantage in 2004. In the case of a president running for

model's in-sample projections are remarkably close to the actual popular-vote and Electoral College outcomes in 2008 and 2012.

In the critical 2000 and 2016 elections, in which there was a discrepant outcome between the national popular-vote and Electoral College outcomes, these in-sample projections were a mixed bag. For 2000, the model correctly concluded that George W. Bush would win a majority in the Electoral College

Table 1
Long-Range Forecast Model

Independent Variable	Unstandardized Regression Coefficient	Standard Error
Prior Result	1.028*	0.021
Prior October Polls	0.593*	0.056
Home-State Advantage	2.835*	0.858
Consecutive Terms	−1.066*	0.143
Constant	−31.118*	3.064
Adj. R²=0.89		
S.E. y|x=3.21		
N=300		

Model Performance over Time	1996	2000	2004	2008	2012	2016	Overall
States Correctly Predicted	88%	84%	94%	94%	100%	92%	92%
Mean Absolute Error	2.65	2.66	2.07	2.62	1.98	2.58	2.43
National-Level Predictions (Excluding DC)							
National Popular Vote	54.0	48.2	49.9	53.3	52.8	49.9	
Error	−0.7	−2.0	+1.2	−0.3	+0.9	−1.1	1.0[†]
Electoral College Vote	399	231	228	346	329	282	
Error	+36	−32	−20	−16	0	+58	27[†]

Notes: Dependent variable is Democratic share of the two-party popular vote in states for every election from 1996 to 2016; *p<0.05.
Note: [†]Mean absolute error.

—albeit by a larger margin than he actually received—but incorrectly projected that he also would win a majority of the popular vote. The model predicted that Hillary Clinton would win razor-thin majorities in both the Electoral College and popular vote.

THE 2020 PREDICTION

Utilizing the coefficients in table 1 and entering the relevant data for each predictor variable for the current election yields the prediction shown in table 2. On average, Biden held a sizeable lead of 9.7 percentage points over Trump in the national head-to-head polls in October 2019, as reported on RealClearPolitics.com.[4] This translates into a projection of Biden winning 54.8% of the national two-party popular vote to Trump's 45.2% and a majority of the Electoral College, 350 to 188.

Using these point estimates, along with the model's standard error of the estimate, I generated Biden win probabilities for each state and the District of Columbia (figure 1). Most important, I highlighted the "tipping-point" state. When all states are arranged according to their respective win probabilities, this is the state that would give either candidate the necessary 270 electoral votes to win. That state is Wisconsin, which the model suggests has a 0.857 probability of being won by Biden. With all of this information, it is possible to generate confidence intervals around the model's national-level projections by running simulated election outcomes while factoring in these probabilities. The 95% confidence intervals around both the national popular-vote and Electoral College projections in table 2 are from the distribution of 100,000 simulated elections.

Biden won a majority of the national popular vote in every one of the simulated elections, and he won an Electoral College majority 99.85% of the time. This is largely the result of the fact that the model projects many more possible paths for Biden to get to 270 than it does for Trump. Every one of Trump's Electoral College victories (i.e., 0.13% of the simulated outcomes) came as a result of an Electoral College misfire, just as it had in 2016. Time will tell if this is accurate or a foolhardy attempt to push the lead-time envelope.

Table 2
2020 Long-Range Forecast

Projected Biden share of National Two-Party Popular Vote		54.8%
95% Confidence Interval		53.6–56.1
Projected Biden Electoral College Vote Total		350
95% Confidence Interval		295–406
Chance of an Electoral College Misfire	Favoring Trump	0.13%
	Favoring Biden	<0.01%
Chance of an Outright Victory	Trump	<0.01%
	Biden	99.85%
Chance of an Electoral College Tie		0.02%
Tipping-Point State Biden Win Probability	Wisconsin	85.7%

Figure 1
Long-Range Forecast Biden State-Win Probabilities
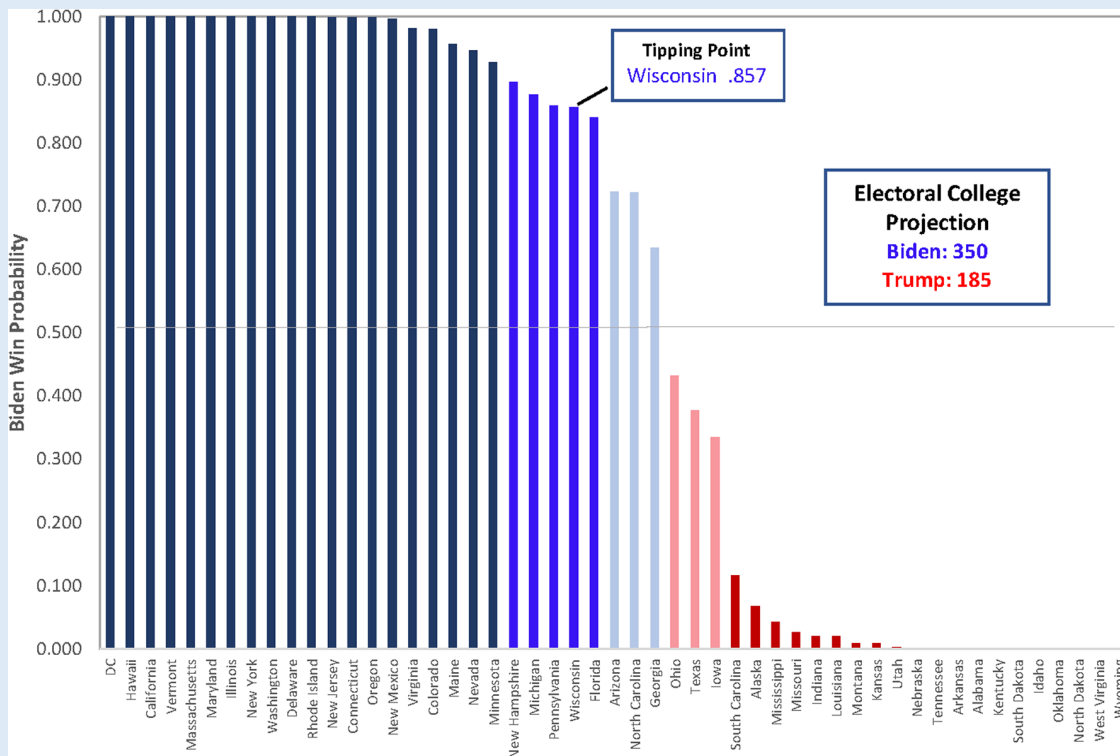
DATA AVAILABILITY STATEMENT

Replication materials are available on Dataverse at https://doi.org/10.7910/DVN/6GWHQB.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S1049096520001468. ■

NOTES

1. Because the 2015 Iowa Conference on Presidential Politics occurred during the month of October, complete polling data were not yet available; therefore, this forecast was a preliminary estimate. Subsequent October polling data obtained after the conference paper was written were slightly more favorable toward Clinton. Additionally, other data for other years were obtained after the original conference paper was written, which resulted in a slight adjustment to coefficients. These results produced an *a priori* forecast based on more complete information that showed Clinton winning a small majority in the Electoral College, 285 to Trump's 253. The four states that the model incorrectly predicted were Iowa, Michigan, Pennsylvania, and Wisconsin, with an average prediction error of 3.32.

2. In October 2015, there were polls conducted assuming that either Hillary Clinton or Bernie Sanders would be the Democratic nominee, and six potential Republican nominees: Donald Trump, Ben Carson, Carly Fiorina, Jeb Bush, Marco Rubio, and Ted Cruz. Of these 12 potential pairings, there was one for whom there was no head-to-head polling data: Sanders versus Bush. Therefore, only 11 forecasts were possible.

3. Due to its outlier status as an overwhelmingly Democratic constituency, the District of Columbia was omitted from the analysis to avoid a biasing of the estimates. However, it is included after the fact in the national-level projections of the national popular-vote and Electoral College outcomes.

4. Of course, it was not known in October 2019 that Joe Biden would be the Democratic nominee. Therefore, as in 2016, a matrix of possible matchups had to be created at that time. However, because nominees are now known, the matrix is no longer necessary.

REFERENCES

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

Berry, Michael J., and Kenneth N. Bickers. 2012. "Forecasting the 2012 Presidential Election with State-Level Economic Indicators." *PS: Political Science & Politics* 45 (4): 669–74.

Campbell, James E. 2016. "Forecasting the 2016 American National Elections: Introduction." *PS: Political Science & Politics* 49 (4): 649–54.

DeSart, Jay A. 2015. "State Electoral Histories, Regime Age, and Long-Range Presidential Election Forecasts: Predicting the 2016 Presidential Election." Presented at the 2015 Iowa Conference on Presidential Politics, Sauk Center, Iowa.

DeSart, Jay A., and Thomas M. Holbrook. 2003. "Statewide Trial Heat Polls and the 2000 Presidential Election: A Forecast Model." *Social Science Quarterly* 84 (4): 561–73.

DeSart, Jay. 2020. "Replication Data for: A Long-Range State-Level Forecast of the 2020 Presidential Election." Harvard Dataverse. doi:10.7910/DVN/6GWHQB.

Jerôme, Bruno, and Véronique Jerôme-Speziari. 2016. "State-Level Forecasts for the 2016 US Presidential Elections: Political Economy Model Predicts Hillary Clinton Victory." *PS: Political Science & Politics* 49 (4): 680–86.

Klarner, Carl. 2012. "State-Level Forecasts of the 2012 Federal and Gubernatorial Elections." *PS: Political Science & Politics* 45 (4): 680–86.

Norpoth, Helmut. 2013. "Time For Change: A Forecast of the 2016 Election." Presented at the 2013 Annual Meeting of the American Political Science Association, Chicago, August 29–September 1.

Norpoth, Helmut. 2016. "Primary Model Predicts Trump Victory." *PS: Political Science & Politics* 49 (4): 655–58.