Online Supplementary Appendix:

Forecasting the 2020 Electoral College Winner:
The State Presidential Approval/State Economy Model

Peter K. Enns
Associate Professor, Department of Government
Executive Director, Roper Center for Public Opinion Research
Co-Director, Cornell Center for Social Sciences
Cornell University
peterenns@cornell.edu

Julius Lagodny
Ph.D. Candidate, Department of Government
Cornell University
jsl364@cornell.edu

Contents

Online Appendix 1 Detailed Variable Discussion Online A-2
Online Appendix 2 Forecast Simulations Online A-4
Online Appendix 3 Forecast Error Comparisons Online A-7
Online Appendix 4 Overall Forecast Accuracy Online A-8
Online Appendix 5 Survey Data Details Online A-10
Online Appendix 1  Detailed Variable Discussion

Below we offer a detailed discussion of all model variables, which was not possible in the main text due to the word limitation.

MRP Model of State Presidential Approval

A key contribution of our approach is to estimate the percent approving of the president in each state (and Washington DC). In total we have 70 polls from June and July of each election year, with a total of almost 90,000 respondents. The number of available surveys ranges from a minimum of 3 (with 3,173 respondents) in 1988 to a maximum of 12 with 14,439 respondents in 1992. All of these surveys are national probability-based samples obtained from the Roper Center for Public Opinion Research at Cornell University (https://ropercenter.cornell.edu/) with one 2020 survey from Gallup Analytics (see Online Appendix 5 for a list of all surveys used). We use multi-level regression with post-stratification (MRP) to estimate state-level presidential approval.

MRP is a three-step approach that involves estimating a multilevel model to identify the relationship between demographic categories and the probability of survey response (in this case indicating approval of the president’s handling of the job of president), using these estimates to predict the probability of approval for each demographic-geographic “type” (e.g., African American females, age 30-44, with some college education, in Texas), and then using census data to poststratify (i.e., weight) the responses to match actual state population values. MRP has repeatedly been shown to recover valid state-level measures of public opinion from national surveys (Gelman & Little 1997, Lax & Phillips 2009, Pacheco 2014). Our model follows our earlier work (Enns & Koch 2013, Enns & Koch 2015, Enns 2016, Enns, Lagodny & Schuldt 2017) and includes age (18-29, 30-44, 45-64, 65+), education (no high school degree, high school degree, some college, college grad (and more)), race (white, black, other), and sex (male, female). We also include an indicator for each survey, state, and region (Northeast, Midwest, South, West, or DC). Post-stratification data come from the 1980-2000 census and the 2001-2018 American Community Survey (https://usa.ipums.org/usa-action/variables/group).

As noted in the main text, after estimating the percent who approve of the president in each state, we subtract a constant so that when our approval variable equals zero, it is roughly equivalent to having no incumbent advantage (Hummel & Rothschild 2014). We select a value of 48 because this maximizes model fit. After subtracting 48, we multiply the resulting number by −1 whenever the incumbent is Republican, so that lower approval of Republican incumbents corresponds with more Democratic support since the dependent variable in our models is the two-party Democratic vote share.

Coincident Economic Indicators

We use the Federal Reserve Bank of Philadelphia’s monthly index of coincident economic indicators to measure economic conditions in each state. This index, which combines multiple economic measures, is advantageous because it reflects multiple aspects of the economy. We generate our measure as follows. First, we calculate the monthly percent change in each
state’s coincident index. Presidential election outcomes reflect economic conditions prior to
the election year (Hibbs 1987, Wlezien 2015), so our measure incorporates all economic infor-
mation since the incumbent’s inauguration. We also know, however, that economic changes
closer to the election to matter more for vote choice (Hibbs 1987, Wlezien 2015), so we weight
the final quarter of data (Quarter 14: April, May, and June of election year) as 1 and each pre-
vous quarter is weighted exponentially less $0.55^{t-1}$. This weighting scheme follows Erikson
& Wlezien (2008a), Erikson & Wlezien (2008b), and Erikson & Wlezien (2016), but we select
the parameter 0.55 because it maximized model fit and minimizes forecast error. We weight
each month in a quarter the same to smooth the influence of large monthly shifts, particu-
larly during election year. We then sum the weighted values and divide by the total weight
producing a weighted cumulative average. Since the dependent variable in our models is the
two-party Democratic vote share, as with presidential approval, we multiply the resulting
number by $-1$ when the incumbent is Republican because more positive economic conditions
under a Republican president (now coded more negative) are expected to correspond with
fewer votes for the Democratic presidential candidate. Our approach mirrors that of Erikson
& Wlezien (2008a) (also see Erikson & Wlezien (2008b) and Erikson & Wlezien (2016)),
though Erikson and Wlezien use the leading economic indicators. We are unable to use lead-
ing economic indicators because the Philadelphia Fed suspended the release of these data
due to measurement complications from the COVID-19 outbreak. As of August 24, 2020,
February 2020 was the most recent available month for state leading indicators (https://

The state coincident indicator data begin in January 1979. Since we use a weighted
cumulative average, having only 6 quarters of data for 1980 (instead of 14) does not pose
a problem (the average for 1980 is based on 6 quarters instead of 14). Kansas, Oklahoma,
Texas, Virginia, Washington, and Wisconsin have missing values for January, February, and
March of 1979. We estimated these values by taking the difference between each state and the
overall US measure in April of 1979 and then subtracting this difference from the US overall
for the three prior months. The Philadelphia Fed does not produce coincident indicators
for Washington DC, so we based DC’s economic conditions on the average of neighboring
Maryland and Virginia.

**Additional Variables: Home State, 3rd Party Candidates, South**

Presidential and Vice Presidential candidates often receive a bump in their home state.
To account for this, we code the state of the Democratic candidate 1, the state of the Repub-
lican candidate $-1$, and all other states 0. If both candidates are from the same state, such
as Hillary Clinton and Donald Trump in 2016, all values are a zero. Home state was ver-
ified from the National Archives: https://www.archives.gov/electoral-college/2016.
In 2019, Trump declared his official residence to be Florida (https://www.nytimes.com/
the recency of the move, and Trump’s long-time association with New York, our 2020 model
leaves his “home state” has New York.

The model also includes the lagged value of the presidential candidates’ home state. We
expect this coefficient to be negative because it accounts for the return to typical voting levels
in that state in the subsequent election (Hummel & Rothschild 2014, Berry & Bickers 2012). We also include the home state of vice presidential candidates. Our model only includes data through July of election year. Since Biden had not yet announced his vice presidential candidate in July, our 2020 forecast codes the democratic vice presidential candidate state as 0. A lagged variable for vice presidential home state was never significant, so we do not include this lag in the model.

Similar to (Hummel & Rothschild 2014), to control for the influence of popular third party candidates, we include the percent of votes obtained in each state the year after they ran. Including third party candidates four years after they ran ensures that we are only using information available at the time of our before-the-fact forecast. For example, John Anderson won 6.6% of the national popular vote in 1980, but we code his vote share as 0 for each state in 1980, assign his actual 1980 vote share in each state in 1984, and code his vote share as 0 every subsequent year. Because Anderson’s state vote share was correlated with two-party vote share (which we confirm with a likelihood ratio test), controlling for Anderson’s vote share in each state in 1984 ensures that our estimated relationship between lagged two-party vote share deviation and current two-party vote share is not biased. Consistent with Hummel & Rothschild (2014), despite Perot’s impressive vote share in 1992, the percentage of votes he received in each state did not appear to influence two-party vote share (p=0.26), so we do not include Perot’s 1992 vote share in 1996. We do find evidence that Perot’s 1996 vote share improves model fit (p=0.076), so we do include the 1996 vote share in the 2000 model. Again, by including vote share in the subsequent election, we are only including information available before the fact in our forecasts.

During the period of analysis, southern states consistently lean Republican. Although many unmeasured factors could account for this (e.g., historical legacies of slavery and segregation, the high proportion of evangelical protestants, or greater rates of felony disenfranchisement during our period of analysis), the negative and significant coefficient in the model for South indicates that even after controlling for each state’s prior vote share, on average, southern states are typically about 1.5 percent less Democratic than we would otherwise expect. We code southern states as those in the confederacy: Alabama, Arkansas, Georgia, Florida, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Texas, and Virginia.

Online Appendix 2  Forecast Simulations

Our simulations account for two types of model uncertainty. First, we use Clarify (Tomz, Wittenberg & King 2003) to simulate 10,000 parameters for each variable in the model. These simulated parameters incorporate uncertainty based on the variance of the parameter estimates. We also need to account for the average error between our predicted values and 

\footnote{Specifically, after the 1980 election, but without adding any election results from the 1984 election, a likelihood ratio test comparing nested models with and without the vote share Anderson received in each state in 1980 shows that including Anderson’s vote share significantly improved model fit (p=0.006).}

\footnote{https://www.pewforum.org/religious-landscape-study/religious-tradition/evangelical-protestant/}

Online A-4
the actual values. To account for this error, we generate a normally distributed variable with a standard deviation equal to the root mean square error for the model. Then, for each state, we use the 10,000 simulated parameter estimates and the estimated forecast error to generate 10,000 predicted outcomes for each state. For each simulation, we assign the corresponding number of Electoral College votes if the Democrat was predicted to win (i.e., two-party vote share greater than 50 percent) and then we sum the expected Democratic Electoral College votes across each state. This gives us 10,000 estimates of the number of Electoral College votes won by the Democratic candidate.

**Accounting for Economic Conditions in 2020**

In a typical year, the steps described above would be sufficient to account for uncertainty in our forecasts. In 2020, however, recent economic shifts introduce additional uncertainty. As described above and in the text, we use the average cumulative monthly percent change in coincident economic indicators, but change is weighted the most heavily in quarter 14 (the final quarter of data prior to the election) and subsequently less each preceding quarter. We select the weighting parameter to maximize model fit (i.e., minimize the root mean square error). In 2020, most states experienced a major decline in the coincident economic indicators due to the negative economic shock caused by COVID-19 in the final quarter of data (April 2020). The timing and magnitude of this economic decline means that small changes in the weighting parameter have a major influence on our 2020 measures of economic conditions. Figure A-1 illustrates the unique sensitivity to parameter choice in 2020. The figure reports three measures of economic conditions for the four states with the most Electoral College votes for each election year. The solid black line reflects our measure with a weight parameter of 0.55, selected to minimize model error. The dashed lines show that in preceding elections, choosing a slightly different weight parameter (i.e., 0.52 or 0.58) would have almost no impact on the subsequent measure in any year before 2020. However, in 2020, these slightly different weight parameters lead to distinct conclusions about current economic conditions.

Figure A-2 plots the root mean square error (RMSE) for 25 different regressions using data from 1980 through 2016, where the only difference across models is the measure of economic conditions. Each measure of economic conditions varied the weight parameter from 0.42 to 0.66 (at increments of 0.01). The parameter 0.55 minimizes the root mean square error, which is why we select this value. However, neighboring weight parameters only alter the RMSE slightly. While a parameter of 0.55 is the best choice according to model fit, neighboring parameter values fit the data almost as well in a substantive sense. Figure A-1 shows that between 1980 and 2016, the economic measures are not sensitive to the exact weighting parameter. But 2020 is more complicated. Weight parameters immediately adjacent to 0.55 might be considered nearly equivalent form a model-fit perspective, but using an alternate parameter would lead to different conclusions about current economic conditions, and thus different forecasts. In other words, if voters in 2020 slightly change how they weight

\[^3\]The number of electoral college votes for each state were obtained from the National Archives: [https://www.archives.gov/electoral-college#2016](https://www.archives.gov/electoral-college#2016). Although Maine and Nebraska award two electoral votes to the statewide winner and one vote to the winner of each congressional district, our forecasts assign all electoral college votes to the statewide winner.
Figure A-1: The Influence of the Weighting Parameter on the 2020 Measure of State Coincident Economic Indicators

Note: Since we forecast the percent Democrat (two-party vote share), measures have been adjusted based on the incumbent presidential candidate; i.e., the economic downturn in 2020 is coded positively because it is expected to disadvantage Trump, the Republican incumbent, and benefit Biden, the Democratic challenger.

past economic performance, this could have significant consequences for our prediction. This sensitivity increases our uncertainty in 2020 compared to other years.

We build this uncertainty into our 2020 forecasts. In addition to 10,000 simulations based on our economic measure using a weight parameter for 0.55, we conduct 10,000 simulations for six additional models that include a measure of economic conditions based on a weight parameter of 0.52, 0.53, 0.54, 0.56, 0.57, and 0.58—for a total of 70,000 simulations. Incorporating these neighboring parameter values into our simulations accounts for measurement uncertainty that stems from not knowing the exact weight parameter. For context, this range of parameter values is a little over half the range of values other scholars have used for
similar measures across elections. Although Erikson and Wlezien use the leading economic indicators in their forecasts, across elections the best fitting weight parameter has varied by 0.1 (Erikson & Wlezien 2008b, Erikson & Wlezien 2016). Hibbs uses weighted average growth of per capita real disposable income personal income (Hibbs 1987, Hibbs 2012), and the weight parameter has also varied by about 0.1 (see, e.g., Wlezien 2015, footnote 7).

Note, this approach only accounts for uncertainty around the weight parameter. This does not account for uncertainty about how the economy will shift between quarter 14 and the election. If the economy does not shift more than in previous elections, the forecast error discussed above will be sufficient. Our simulations do not, however, account for atypical economic shifts after our forecast.

Online Appendix 3  Forecast Error Comparisons

The text compares our forecast error for past elections with other prominent forecasts. Although this is our first public forecast, to generate our past forecasts we only used model estimates from before the election and data available through July of each election year (economic data is through June of election year). So our 1984 forecast was based on the model from 1980 and data through July of 1984. Our 1988 forecast was based on a model using data from 1980 and 1984 and data through July 1988. The comparison data came from the following sources.

State Vote Share

The weighting parameters used by Erikson and Wlezien and Hibbs are both higher (ranging from 0.8 to 0.9), perhaps because they use a different economic measure or because they focus on the national economy instead of the state economy, but the key here is that our simulations incorporate a range of parameter a little over half the range of values they have used, suggesting that our range accounts for expected variation across elections.
• Hummel & Rothschild (2014): 2012 forecast error reported on p.136. We report estimates from their February model. Their June forecast based on the same model was slightly less accurate with a mean/median error of 2.94/2.56 points (p.136).

• Klarner (2012): 2012 state forecasts reported in Table 4.

• Berry & Bickers (2012): 2012 state forecasts reported in Table 3.

• Jerôme & Jerôme-Speziari (2012): 2012 state forecasts reported in Table 3, Main (1). Jerôme & Jerôme-Speziari (2012) forecast the incumbent percent out of the popular vote (not two-party vote), so we compare their forecasts with the popular vote (not two-party vote).

• Jerôme & Jerôme-Speziari (2016): 2016 state forecasts reported in Table 2. Again, we compare Jerôme & Jerôme-Speziari’s (2016)’s forecast of the popular vote with the actual popular vote outcome.

**National Vote Share** We select these three models for comparison because the days prior to the election of the forecast roughly align with ours (Campbell 2016, Table 2) and because they report the average forecast error of their models for previous elections.

• Erikson & Wlezien (2016): Page 669 and footnote 7 indicate an average absolute error of 1.6 percentage points for 1996–2012 and their 2016 forecast had a 0.89 error for an overall average of 1.48 percentage points.

• Abramowitz (2016): Table 2 (1988–2012) and 2016 forecast error.

• Lewis-Beck & Tien (2016): Table 1 (1984–2012) and 2016 forecast error, based on out-of-sample (Jackknife) predictions.

**Online Appendix 4  Overall Forecast Accuracy**

The text reported that our model predicts the national two-party vote and Electoral College outcomes with a high degree of accuracy. Table A-1 reports these results by year. Our mean/median Electoral College error (46/42) is less than the Electoral College votes of the swing states of Florida (29) and Pennsylvania (20) together.
<table>
<thead>
<tr>
<th>Year</th>
<th>% Dem National Forecast</th>
<th>% Dem National Actual</th>
<th>Difference</th>
<th>Correctly Predicted</th>
<th>Democratic EC Votes Forecast</th>
<th>Democratic EC Votes Actual</th>
<th>Difference</th>
<th>Correctly Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>1984</td>
<td>44.3</td>
<td>40.8</td>
<td>3.4</td>
<td>Yes</td>
<td>59</td>
<td>13</td>
<td>-46</td>
<td>Yes</td>
</tr>
<tr>
<td>1988</td>
<td>43.7</td>
<td>46.1</td>
<td>-2.4</td>
<td>Yes</td>
<td>146</td>
<td>112†</td>
<td>-34</td>
<td>Yes</td>
</tr>
<tr>
<td>1992</td>
<td>49.3</td>
<td>53.5</td>
<td>-4.1</td>
<td>No</td>
<td>257</td>
<td>370</td>
<td>113</td>
<td>No</td>
</tr>
<tr>
<td>1996</td>
<td>53.9</td>
<td>54.7</td>
<td>-0.8</td>
<td>Yes</td>
<td>389</td>
<td>379</td>
<td>-10</td>
<td>Yes</td>
</tr>
<tr>
<td>2000</td>
<td>54.3</td>
<td>50.3</td>
<td>4.0</td>
<td>Yes</td>
<td>363</td>
<td>267†</td>
<td>-96</td>
<td>No</td>
</tr>
<tr>
<td>2004</td>
<td>47.4</td>
<td>48.8</td>
<td>-1.4</td>
<td>Yes</td>
<td>238</td>
<td>252†</td>
<td>14</td>
<td>Yes</td>
</tr>
<tr>
<td>2008</td>
<td>55.6</td>
<td>53.7</td>
<td>1.9</td>
<td>Yes</td>
<td>363</td>
<td>365</td>
<td>2</td>
<td>Yes</td>
</tr>
<tr>
<td>2012</td>
<td>50.1</td>
<td>52.0</td>
<td>-1.9</td>
<td>Yes</td>
<td>290</td>
<td>332</td>
<td>42</td>
<td>Yes</td>
</tr>
<tr>
<td>2016</td>
<td>51.7</td>
<td>51.1</td>
<td>0.6</td>
<td>Yes</td>
<td>285</td>
<td>232†</td>
<td>-53</td>
<td>No</td>
</tr>
</tbody>
</table>

All forecasts are “before-the-fact.” †These totals ignore faithless electors, which cannot be predicted by the model. The actual Electoral College votes were 111 in 1988, 266 in 2000, 251 in 2004, and 227 in 2016.
Online Appendix 5  Survey Data Details

We utilized the standard presidential approval question, which asks, “Do you approve or disapprove of the way [president's name] is handling his job as President?” All surveys but one were obtained from the Roper Center for Public Opinion Research at Cornell University (https://ropercenter.cornell.edu/). One 2020 survey came from Gallup Analytics.

Surveys from the Roper Center for Public Opinion Research


Online Appendix References


