Fundamentals-Based State-Level Forecasts of the 2020 US Presidential Election

Clemens Nollenberger and Gina-Maria Unger

*University of Konstanz, Germany,
Department of Politics and Public Administration*

This version: 20 October 2020

Abstract:

Forecasts of US presidential elections have gained considerable attention in recent years. However, as became evident in 2016 with the victory of Donald Trump, most of them consider presidential elections only at the national level, neglecting that these are ultimately decided by the Electoral College. In order to improve accuracy, we believe that forecasts should instead address outcomes at the state-level to determine the eventual Electoral College winner. We develop a political economy model of the incumbent vote share across states based on different short- and long-term predictors, referring up to the end of the second quarter of election years. Testing it against election outcomes since 1980, our model correctly predicts the eventual election winner in 9 out of 10 cases – including 2016 –, with the 2000 election being the exception. For the 2020 election, it expects Trump to lose the Electoral College, as only 6.2 percent of simulated outcomes cross the required threshold of 270 Electoral Votes, with a mean prediction of 106 Electoral Votes.

This work is based on a manuscript that was submitted to the 2020 Election Forecasting Seminar at the University of Konstanz on 15 September 2020. We thank its participants for their comments, which benefitted the development of this paper, and would particularly like to express our gratitude to Arndt Leininger for his invaluable feedback and support.

Contact: clemens.nollenberger[at]uni-konstanz.de, gina-maria.unger[at]uni-konstanz.de

1. Introduction

In a heretofore considerably eventful year 2020, the US were caught by a global pandemic within a highly polarized political environment. In midst of this sensitive setting, the presidential election will take place on November 3, pitting Republican President Donald Trump against his Democratic challenger Joe Biden. A comparison with the preceding 2016 election, which among other things was also characterized by an intensely polarized environment, might raise the question whether the 2020 election could be equally difficult to forecast. In 2016, most election forecasts favored the Democratic candidate Hillary Clinton, but Donald Trump eventually gained a majority in the Electoral College despite losing the Popular Vote (Campbell et al. 2017).

Although fundamentals-based forecast models performed relatively well in 2016 in terms of the Popular Vote (Campbell et al. 2017), they generally did not expect that the majorities of the Popular Vote and Electoral College might differ. While this had not happened throughout the 20th century, it then occurred already for the second time since 2000. As this divergence may arise particularly in close elections, a focus on the national Popular Vote alone might not be the best approach to forecast the winner in upcoming elections. We argue that instead of relying on the national level only, the distribution of Electoral Votes should be emphasized and forecasters should thus concentrate on the state-level.

Consequently, we aim to forecast the 2020 election with a state-by-state political economy model relying on local as well as national data for presidential elections from 1980 to 2016. We develop a parsimonious model capturing information about local economic development and general public satisfaction with the incumbent president in election years, institutional (dis-)advantages for candidates across states in given elections as well as states' long-term partisan domination. On that basis, we aim to quantify the underlying uncertainty of our forecast, regarding both the uncertainty in estimating the relationship of our independent variables with the incumbent vote share as well as the fundamental uncertainty associated with predicting single events (King et al. 2000). To this end, we simulate a range of potential state-level election outcomes, accounting for their correlation across states, from which we derive our eventual forecast. When conducting out-of-sample tests on historical elections, our model correctly predicts the Electoral College winner in 9 out of the 10 elections since 1980, including

Trump's victory in 2016. We hope that our undertakings contribute to a rising literature of state-level forecasts.

Our forecast was completed in early September, giving it a lead time of approximately two months. Taken together, it indicates 106 Electoral Votes and a 39.8% share of the national two-party Popular Vote for incumbent President Trump. Incorporating the uncertainty associated with this estimate, it marks Trump's chance of winning the Electoral College at 6.2 percent. In essence, our forecast suggests that the Democratic candidate Biden is on a trajectory towards winning the Presidency, but these conclusions come with a considerable extent of uncertainty.

The paper is organized as follows: After providing a brief overview of the general logic of state-by-state political economy models, we describe our data and embed our choice of variables in a broader theory of voting. Afterwards, we outline the methodology of our forecasting model, examine its predictive accuracy on historical elections, and depict the output of the underlying regression model on our full dataset. The last paragraph presents and discusses our forecast.

2. State-by-state Models for US Elections

The first political economy forecasting models were utilized in the late 1970s to early 1980s by Sigelman (1979), Lewis-Beck and Rice (1982, 1984) and Rosenstone (1983) to forecast US presidential elections. Rosenstone's model particularly relied on the states as unit of study instead of the entire nation. Although more state-based models were implemented afterwards by Holbrook (1991) and Campbell (1992) following Rosenstone's example, they are still outnumbered by other approaches. Notable recent applications for US presidential elections that provide the foundation of our own model include Berry and Bickers (2012), Klarner (2012), as well as Jérôme and Jérôme-Speziari (2012, 2016) for the elections in 2012 and 2016, respectively. For the upcoming 2020 election, state-level forecasts were conducted by DeSart (2020), Enns and Lagodny (2020) and Jérôme et al. (2020).

Political economy models for state-level forecasts are mainly built on a vote function, which often provides a model of voters' choice to vote for an incumbent (or the party of the incumbent) based on the political and economic context. These models thus follow a structural approach. The used data commonly have a lead time of one quarter, allowing these models to make predictions several months in advance.

The main logic of state-level models is to employ disaggregated data to obtain predictions across states, which are then aggregated on the national level to get an overall prediction. Including state-level data can improve predictive accuracy on the national level as it allows to incorporate information about potentially heterogeneous local economic and socio-political circumstances. This is an important feature as the US presidential election highly depends on local voting behavior, with the Electoral College determining the winner and not the national Popular Vote. Furthermore, pooling data across states increases the number of observations and the degrees of freedom, what should reduce uncertainty as well as biases due to aggregation effects. In addition, as they rely on panel data of states across elections, these models allow to discern effects that are difficult to identify with longitudinal datasets only at the national level (Jerôme and Jérôme-Speziari 2012, 2016).

3. Data

We compiled a dataset with the variables described below comprising information about presidential elections within the 50 states and the District of Columbia. State-level election results were retrieved from the MIT Election Data And Science Lab (2017), which provides information about the total number of votes every candidate received across states for presidential elections since 1976. From these quantities, we calculate the two-party-vote share for the respective candidates of the Democratic or Republican party as the number of votes cast for them individually divided by the sum of votes cast for either the Republican or the Democratic candidate. We adopt the two-party-vote share of the respective incumbent party, i.e. the party of the incumbent president, in a given election as our dependent variable.¹

Our approach to modeling election outcomes is based on the notion that voters hold incumbents accountable by either “rewarding” or “punishing” them for their performance in office (Key 1966; Ferejohn 1986). Accordingly, good performance can be expected to translate into more votes for incumbents at subsequent elections, whereas voters may punish them for bad performance by voting for oppositional candidates. Comparative studies demonstrate that the economy plays a pivotal part in voters’ evaluation of government performance (Lewis-Beck

¹ Note that these data do not allow to discern between sub-state levels allocating their Electoral Votes independently in Maine and Nebraska. Following other state-level forecasts (e.g. DeSart 2020; Enns and Lagodny 2020; Jérôme et al. 2020), these states are instead assumed to form a unit.

and Stegmaier 2000; Duch 2007). However, Healy and Lenz (2014) establish that voters tend not to consider economic development over the entire period between elections in equal measure, but to rely primarily on the current situation in the year of the election. As an indicator for local economic circumstances, we thus include state-level unemployment in our model and use the change in the unemployment rate within a state in the election year compared to the previous year. To allow sufficient lead time, this variable refers to the 12-month change at the end of the second quarter of the election year, approximately four months before the election takes place. These data are retrieved from the US Bureau of Labor Statistics (2020).

Moreover, we include presidential job approval in our model. By indicating voters' overall satisfaction with the incumbent president, this measure allows to capture both how voters evaluate the government's handling of the given economic circumstances (measured by unemployment) as well as it reflects a broad range of other determinants of their evaluation of government performance. We draw on the percentage of respondents indicating approval with the way the president is handling his job in Gallup opinion polls at the end of the second quarter in the election year (American Presidency Project 2020). This variable refers to the national level, allowing to incorporate general national-level swings into our model that affect election outcomes independently from local circumstances. Primarily, we considered polls with fieldwork being conducted between June 23 and July 7 in election years, and calculated the arithmetic mean of the corresponding shares of respondents across polls. In 2000, 2004, 2008 and in 2020, there were no polls conducted exclusively within this time frame. For these years, we thus averaged over the two next available polls before and after the time frame.

Against the backdrop that it might affect a challenger's ability to mobilize support whether large parts of the public are actually discontent with an incumbent president or simply neutral, focusing on approval alone is not sufficient. We thus also include net approval in our model. For this variable, we subtracted the share of approving respondents from the share of respondents indicating disapproval with the president (calculated analogously to approval as described above).

As a further variable affecting vote choice, we regard an incumbent party's tenure and count its number of terms in office. This variable intends to reflect the "cost of ruling" for incumbents in the form of losing votes at subsequent elections (Nannestad and Paldam 2002), where the vote loss tends to increase with their length in office. In the case of the US, Abramowitz (2008)

demonstrates that incumbent parties are less likely to win an election if they were in power for two or more terms than after their first term, receiving less votes on average.

Furthermore, we include a variable indicating the home states of the candidates of the incumbent and challenging party, aiming to account for the increased vote share they tend to achieve in these (Mixon and Tyrone 2004; Powell 2004). Early scholarship related this effect to voters' increased identification with a candidate from their local area as well as a higher perceived probability that this candidate would favor their state in resource distribution (Lewis-Beck and Rice 1983). Some newer reasoning refers to disproportionate partisan mobilization of non-voters (Devine and Kopko 2013). Based on biographical information about the candidates, we regard their home state as the state of their primary political affiliation, expressed through their activity or highest political position before candidacy. The variable either takes the value of 1, if the respective state is the home state of the incumbent party's candidate, -1 if it is the home state of the challenging party's candidate, or 0 otherwise.

Apart from these rather short-term factors affecting vote choice at the aggregate level, we include two variables capturing long-term partisan domination of states. We first classify states that were always won by a single party with a margin of more than 5 percentage points of the two-party-vote over the entire period covered in our data as Republican- or Democrat-dominated states, respectively.² We encode two variables, *repstate* and *demstate*, which both take on the value of 0 for states that are not classified as partisan-dominated. When a state is dominated by one party, we assign the value of 1 to the corresponding variable when the respective party is the incumbent party and -1 when the party is not the incumbent party.

In order to address long-term differences in the magnitude of electoral support for different parties across states, we include the two-party-vote share that the respective incumbent party received at the preceding election as additional variable in our model. Note that this is not simply the lagged value of our dependent variable, as the incumbent party might differ between consecutive elections. Instead, we adopt the previous vote share of the party which is the incumbent party at the current election.

² We classify Alaska, Idaho, Kansas, Nebraska, North Dakota, Oklahoma, Utah, and Wyoming as Republican-dominated, and the District of Columbia as Democrat-dominated.

Additional variables used to generate our forecast as described in section 4 are the number of Electoral Votes across states for each election. Furthermore, in addition to the lagged share of the two-party vote, we also count the lagged absolute number of votes for these parties within a state in the previous election.

As we have no local unemployment data available for the 1976 election, we consider elections from 1980. Our first final dataset, used to fit the model, thus contains 510 observations at the state-level for the 10 presidential elections between 1980 and 2016. The second dataset, used to generate the forecast, comprises values for the independent variables across states in 2020.

4. Forecasting Model

This section describes the methodology with which we derive our forecast. In a first step, we run a linear regression of the incumbent two-party-vote share across elections and states on the change in unemployment, presidential job approval, net job approval, the number of terms of the incumbent party in office, variables indicating home state advantages and partisan-dominated states, and the respective party's lagged two-party-vote share:

$$incumbent2pv = \beta X + \epsilon \quad (1)$$

where X denotes the matrix of covariates, β the coefficient vector to be estimated, and ϵ the vector of residuals, which are assumed to follow a normal distribution with mean 0 and a standard deviation to be estimated by the model (residual standard error). In order to address the potential problem that election outcomes across states might not be independent from each other but correlated (Lauderdale and Linzer 2015), we cluster standard errors by state and election.

In a second step, the estimated coefficients $\hat{\beta}$ are used to compute the forecasted incumbent two-party-vote share for the 2020 election with the corresponding values of the predictor variables. In doing so, we follow the approach proposed by King et al. (2000) to simulate random draws from the respective distributions of coefficient estimates and residuals in order to incorporate both the uncertainty in the estimation of coefficients as well as the underlying fundamental uncertainty about the outcome in our predictions. Hence, we first generate 100,000 random draws from a multivariate normal distribution with $\hat{\beta}$ as mean vector and the cluster-robust variance-covariance matrix retrieved from the regression model specifying the

covariance between estimates in $\hat{\beta}$. This yields a matrix where the first column contains simulated intercepts and the remaining 8 columns simulated coefficients for our predictor variables. We then multiply these columns with the respective predictors for the 2020 election in each state and add the resulting quantities, together with the intercept, to obtain a matrix of expected values across states.

From these, we derive predicted values of the incumbent two-party-vote share that reflect the underlying fundamental uncertainty as we add to each the value of a random draw from a normal distribution with mean 0 and the residual standard error estimated by the regression model as standard deviation. Resulting out-of-range forecasts greater than 1 or less than 0 are corrected by setting them to 1 or 0, respectively. From the resulting matrix with 51 columns depicting the states, we calculate our forecast for the incumbent two-party-vote share in each state as the arithmetic mean of each respective column. The lower and upper bound of a 95% prediction interval are extracted by the 2.5 and 97.5 percentiles of each column, respectively. Moreover, we calculate a win probability for each state as the percentage of predicted values greater than 0.5.

In a final step, these state-level forecasts are aggregated to the national-level through the Electoral College. We count the number of Electoral Votes from all states in which the forecasted vote share exceeds 0.5. Similarly, we calculate the lower and upper bound for a 95% prediction interval with the sum of Electoral Votes from states where the lower or upper bound of the state-level forecast is larger than 0.5, respectively. Furthermore, we obtain the distribution of simulated Electoral Votes by first sorting the matrix of simulated state-level predictions for each state and replacing these predictions with the state-specific number of Electoral Votes when they exceed 0.5, and with 0 otherwise. We then calculate the sum of Electoral Votes for each row of simulated outcomes across states. Based on this, we determine the proportion of instances where this sum equals at least 270 in order to derive a probability that the incumbent candidate will win the Electoral Vote.

Although the winner of the election is only determined by the number of Electoral Votes, we also calculate a national-level Popular Vote forecast in order to facilitate comparability with other forecasts that tend to focus on this quantity. Thus, we adopt the approach by Berry and Bickers (2012), who use turnout in the previous election as predictor for turnout in the following one. As our interest lies in the two-party-vote share, we use the total number of votes cast for

these parties for this purpose, exploiting the high correlation between the absolute two-party-vote and its lagged value across states in our dataset ($r = 0.99$). Consequently, we sum over the product of our forecasted state-level two-party-vote shares with the lagged absolute two-party-votes across states, and then divide this by the sum of the lagged absolute two-party-vote across states. We proceed similarly with the lower and upper bounds of the prediction intervals for the state-level two-party-vote shares in order to obtain a corresponding interval at the national level.

5. Empirical Results

Table 1 presents the results of the regression model as outlined in the previous chapter using our dataset of ten presidential elections between 1980 and 2016 within the 50 states and the District of Columbia.

Overall, our model is able to account for 87.6% of the variance in the dependent variable, with the residual standard error amounting to 3.95 percentage points. Except for the indicating home state advantages and Democrat-dominated states, the coefficient estimates for our predictors reach conventional levels of significance. Accordingly, each increase of 1 percentage point in the unemployment rate in the June of the election year, compared to 12 months earlier, is associated with a decrease in the incumbent two-party-vote share by 1.2 percentage points on average. Each percentage point increase in presidential job approval at the end of the second quarter tends to be followed by an additional incumbent two-party-vote share of 0.6 percentage points, while this effect is counteracted by circa -0.3 percentage points when net approval also increases. For each additional term that an incumbent party was in office before the election, one would expect a decrease in the vote share by 3 percentage points. Incumbent candidates and challenger candidates may each expect to fare ca. 1.7 percentage points better in their respective home state. Incumbent vote shares are on average 2.7 percentage points higher (lower, respectively) in Republican-leaning states than in other states when the incumbent president is a Republican (Democrat). Likewise, Democrat-leaning states exhibit a difference of ca. 4.2 percentage points compared to other states. Finally, each percentage point of an incumbent party's vote share at the previous election is on average associated with 0.89 percentage points of its vote share in the current election.

Table 1: State-Level Regression Analysis

	<i>Dependent variable:</i>
	Incumbent Two-Party- Vote Share
Constant	-0.2030 (0.1505)
Δ Unemployment	-0.0124** (0.0055)
Approval	0.0061** (0.0027)
Net Approval	-0.0026** (0.0012)
Tenure	-0.0303*** (0.0067)
Home State	0.0169* (0.0096)
Rep. State	0.0272** (0.0138)
Dem. State	0.0417 (0.0279)
Previous Vote Share	0.8938*** (0.0686)
Observations	510
R ²	0.8764
Adjusted R ²	0.8744
Residual Std. Error	0.0395 (df = 501)
F Statistic	443.8710*** (df = 8; 501)

Note: Standard errors clustered by election and state
(in parentheses)

*p<0.1; **p<0.05; ***p<0.01

6. Model Diagnostics

Before generating a forecast for 2020, we evaluate our model considering its performance on previous elections. As a more conservative test, we examine its ability to predict election outcomes that were not used in estimating it. To this end, we run different rounds to generate out-of-sample predictions, in which we partition our dataset of ten elections between 1980 and 2016 each into a training set of nine elections and a test set of one election. We then fit a model on the training data and predict the dependent variable in the test data, employing the same

Table 2: Model Diagnostics (State-Level)

Election	RMSE	Missed States
1980	0.0588	9
1984	0.0701	14
1988	0.0449	7
1992	0.0458	10
1996	0.0406	7
2000	0.0558	10
2004	0.0243	3
2008	0.0412	5
2012	0.0403	4
2016	0.0392	3
Mean	0.0461	7.2

procedure as described in section 4. Each of the ten elections was used as test data once, so that we run this algorithm ten times in total. In every round, we also generate the variables indicating Republican- and Democrat-dominated states newly using only the respective training data, as they would otherwise be based on the outcome of the test election as well. After retrieving predictions for the test data in terms of the state-level vote shares as well as the national Popular Vote and Electoral Votes, we compare these with the actual results of the respective election.

Table 2 exhibits the Root Mean Squared Error (RMSE) for the predicted state-level incumbent two-party-vote shares as well as the number of states where the wrong winner was predicted. It illustrates that on average across all elections, the state-level predictions differed by 4.6 percentage points from the actual two-party-vote shares. Remarkably, the accuracy of the model seems to improve over time with regard to missed states. The relatively high mean of 7.2 missed states per election seems to be driven by the six elections until 2000, while this number varied only between 3 and 5 in the four more recent ones since 2004. Despite these inaccuracies at the state-level, as tables 3 and 4 show, the model is nevertheless able to correctly predict the winner of both the Electoral College and the Popular Vote in nine out of ten elections.

Table 3 summarizes model diagnostics with regard to Electoral Votes. The mean difference between predicted and actual Electoral Votes of incumbent candidates is 1.2, indicating that our model does not tend to over- or underestimate incumbent Electoral Votes to a large extent in the long run. In absolute terms, the individual predictions for the respective elections deviate from the actual outcome by 60.2 Electoral Votes on average. However, the model predicted nine out of ten Electoral College winner correctly, with the 2000 election being the exception. Likewise, in nine out of ten elections the 95% prediction interval comprised the eventual number of Electoral Votes, with exception of the 1984 election.

Table 3: Model Diagnostics (Electoral Votes)

Election	Actual EV	Predicted EV	Pred. EV (Lower)	Pred. EV (Upper)	Diff. Pred. - Actual EV	EV Winner Correct
1980	90	180	21	485	90	Yes
1984	525	387	153	523	-138	Yes
1988	426	462	126	535	36	Yes
1992	168	72	0	322	-96	Yes
1996	379	345	61	518	-34	Yes
2000	267	375	119	512	108	No
2004	286	307	82	488	21	Yes
2008	174	147	30	355	-27	Yes
2012	332	371	194	510	39	Yes
2016	233	246	119	410	13	Yes
Mean					1.2	
Mean Absolute Error					60.2	

Table 4: Model Diagnostics (National Popular Vote)

Election	Actual PV	Predicted PV	Pred. PV (Lower)	Pred. PV (Upper)	Diff. Pred. - Actual PV	PV Winner Correct
1980	0.4484	0.4787	0.3955	0.5621	0.0303	Yes
1984	0.5912	0.5418	0.4666	0.6170	-0.0494	Yes
1988	0.5383	0.5546	0.4673	0.6418	0.0163	Yes
1992	0.4638	0.4354	0.3483	0.5225	-0.0284	Yes
1996	0.5480	0.5243	0.4439	0.6047	-0.0237	Yes
2000	0.5026	0.5366	0.4590	0.6142	0.0341	Yes
2004	0.5123	0.5094	0.4246	0.5943	-0.0029	Yes
2008	0.4623	0.4485	0.3605	0.5364	-0.0138	Yes
2012	0.5201	0.5449	0.4637	0.6260	0.0248	Yes
2016	0.5116	0.4949	0.4121	0.5777	-0.0167	No
Mean					-0.003	
Mean Absolute Error					0.024	

Table 4 provides similar information regarding the Popular Vote. The mean difference between the predicted and actual national two-party-vote share of incumbent candidates is only 0.3 percentage points, indicating that our model does not tend to over- or underestimate the incumbent Popular Vote to a large extent in the long run. In absolute terms, the individual predictions for the respective elections deviate from the actual outcome by 2.4 percentage points

on average. As with Electoral Votes, the model predicted nine out of ten Popular Vote winner correctly. The exception here is the 2016 election, for which our model incorrectly predicted that incumbent candidate Hillary Clinton would lose the Popular Vote. For all elections, the 95% prediction interval comprised the eventual two-party-vote share.

7. Forecast for 2020 and Discussion

Following the procedure described in section 4, we obtain forecasts of the two-party-vote share of Republican candidate and incumbent President Donald Trump in the upcoming 2020 election across states, the corresponding number of Electoral Votes as well as the national two-party Popular Vote. *Table 5* lists our state-level forecasts of the Republican two-party-vote share with a 95% prediction interval and a probability of Trump winning this particular state. As it can be seen from this table, our model forecasts Trump to win only 18 states, whereas it forecasts 33 states to be garnered by his challenger Biden. Perhaps unsurprisingly, given electoral history, our model predicts Trump to receive his lowest state-level vote share in the District of Columbia (ca. 2.1%, 95% prediction interval from 0% to ca. 9.2%) and his highest vote share in Wyoming (ca. 70.5%, prediction interval from 60.9% to 80.1%). Generally, the states that we predict to be won by Trump are among the states where the increase in unemployment was less than the average increase across all states, including those which we classified as Republican-dominated on a long-term basis. However, the prediction intervals for individual states reflect a substantial extent of uncertainty in our forecast. We provide a visualization of the distribution of forecasted two-party-vote shares for each state across our simulations in the Appendix.

The two closest states according to our forecast are Indiana and South Carolina (vote share forecast: 49.1% and 49.2%, respectively), where the outcomes are essentially toss-ups but in slight favor of Biden. If we look at the probability of a Trump victory across states, there are 20 states for which this is less than even 5%, with these states alone accounting for 251 Electoral Votes. Complementarily, there are only 9 states accounting for 44 Electoral Votes where this probability exceeds 95%.

When comparing our forecasted two-party-vote shares with the ones Trump achieved in 2016, our forecast does not expect him to increase his vote share in any state if we only look at the point estimate of the vote share. Out of the 30 states that Trump won in 2016, we predict 12 to change sides; among them, perhaps most remarkably given its electoral history and number of

Electoral Votes, Texas. However, the relatively wide range of the prediction intervals should be brought to mind again and caution the reader.

Table 5: State-Level Forecasts

State	EV	Two-Party-Vote	Winner	Lower Bound	Upper Bound	Probability of Trump Win
Alabama	9	0.5676	Trump	0.468	0.6671	90.9 %
Alaska	3	0.5225	Trump	0.4186	0.6251	66.7 %
Arizona	11	0.45	Biden	0.3548	0.5454	15.3 %
Arkansas	6	0.5693	Trump	0.4718	0.6672	91.7 %
California	55	0.2176	Biden	0.0844	0.3502	0 %
Colorado	9	0.3753	Biden	0.2605	0.4888	1.6 %
Connecticut	7	0.3541	Biden	0.2543	0.4533	0.2 %
Delaware	3	0.3164	Biden	0.1932	0.4388	0.1 %
D.C.	3	0.0208	Biden	0	0.0917	0 %
Florida	29	0.4148	Biden	0.3052	0.524	6.3 %
Georgia	16	0.4701	Biden	0.3807	0.5597	25.7 %
Hawaii	4	0.2087	Biden	0.0775	0.3384	0 %
Idaho	4	0.6526	Trump	0.5664	0.7393	100 %
Illinois	20	0.2864	Biden	0.1533	0.4191	0.1 %
Indiana	11	0.4909	Biden	0.3705	0.6111	44 %
Iowa	6	0.4723	Biden	0.3718	0.5725	29.4 %
Kansas	6	0.5693	Trump	0.477	0.6612	93 %
Kentucky	8	0.6363	Trump	0.5542	0.718	99.9 %
Louisiana	8	0.5276	Trump	0.4296	0.6261	71 %
Maine	4	0.436	Biden	0.3498	0.5229	7.4 %
Maryland	10	0.3137	Biden	0.2262	0.402	0 %
Massachusetts	11	0.1831	Biden	0.0099	0.3528	0 %
Michigan	16	0.3656	Biden	0.2267	0.5055	2.9 %
Minnesota	10	0.423	Biden	0.3273	0.5188	5.9 %
Mississippi	6	0.5366	Trump	0.4477	0.6249	79.2 %
Missouri	10	0.5271	Trump	0.4302	0.6236	70.8 %
Montana	3	0.5509	Trump	0.4593	0.6423	86.3 %
Nebraska	5	0.616	Trump	0.5328	0.6992	99.7 %
Nevada	6	0.3452	Biden	0.2003	0.4892	1.8 %
New Hampshire	4	0.4125	Biden	0.3068	0.5179	5.1 %
New Jersey	14	0.2645	Biden	0.101	0.4264	0.2 %
New Mexico	5	0.4124	Biden	0.3284	0.4971	2.1 %
New York	29	0.249	Biden	0.1072	0.3899	0 %
North Carolina	15	0.4711	Biden	0.3848	0.5576	25.5 %
North Dakota	3	0.6398	Trump	0.5412	0.7385	99.7 %

Table 5: State-Level Forecasts

State	EV	Two-Party-Vote	Winner	Lower Bound	Upper Bound	Probability of Trump Win
Ohio	18	0.4498	Biden	0.3401	0.5592	18.4 %
Oklahoma	7	0.6575	Trump	0.5698	0.7455	100 %
Oregon	7	0.3454	Biden	0.2346	0.4558	0.3 %
Pennsylvania	20	0.39	Biden	0.2662	0.5128	4 %
Rhode Island	4	0.3112	Biden	0.1912	0.4306	0.1 %
South Carolina	9	0.4922	Biden	0.389	0.595	44 %
South Dakota	3	0.5917	Trump	0.4962	0.6871	97 %
Tennessee	11	0.5422	Trump	0.4319	0.6519	77.4 %
Texas	38	0.4773	Biden	0.3814	0.5728	32.1 %
Utah	6	0.6018	Trump	0.5177	0.6863	99.1 %
Vermont	3	0.2735	Biden	0.1712	0.3757	0 %
Virginia	13	0.4064	Biden	0.3119	0.5	2.5 %
Washington	12	0.348	Biden	0.2524	0.4426	0.1 %
West Virginia	5	0.6234	Trump	0.5101	0.7366	98.4 %
Wisconsin	10	0.4365	Biden	0.341	0.5311	9.5 %
Wyoming	3	0.7048	Trump	0.6094	0.801	100 %

Table 6: National-Level Forecasts**6a: Electoral Votes**

EV (Forecast)	EV (Lower Bound)	EV (Upper Bound)	Probability of Trump Win
106	41	336	6.2 %

6b: Popular Vote

PV (Forecast)	PV (Lower Bound)	PV (Upper Bound)
0.3975	0.2851	0.5098

Figure 1: Histogram of Electoral College outcomes based on 100,000 simulated state-level outcomes. The dashed red line indicates the majority threshold of 270 Electoral Votes.

Aggregating these state-level forecasts on the national level, we obtain a forecast of 106 Electoral Votes in favor of Trump. Looking at the distribution of Electoral Votes throughout our simulations, as visualized in *Figure 1*, one can see that relatively many simulated outcomes seem to cluster in this area. However, the range of simulated Electoral Votes is relatively large. *Table 6* summarizes our national-level forecasts, indicating that the 95% prediction interval includes 41 as lower bound and 336 as upper bound. In 6.2% of our simulated Electoral College outcomes, Trump is able to gain 270 or more Electoral Votes, exceeding the threshold of Electoral Votes needed for a majority. This renders it rather unlikely, but not impossible for him to secure his re-election in November. Regarding the Popular Vote, our model forecasts a two-party-vote share of approximately 39.8% with a 95% prediction interval ranging from 28.5% to 51%.

An essential limitation of our forecast is, firstly, that our model is not able to predict to what extent voters will hold an incumbent president accountable for economic circumstances largely determined by a global pandemic, and whether voters will consider the situation as an exogenous shock beyond the incumbent's accountability or not (cf. Duch and Stevenson 2010). Secondly, the strength of the economic vote may also be unclear due to the unprecedented increase in unemployment in 2020. Whereas our model assumed a constant effect of

unemployment change on the incumbent vote, Garz and Martin (2020) recently found evidence of potential non-linear effects. Whether one attributes them to a left-digit bias in reports about the economy like Garz and Martin or perhaps to decreasing returns to scale, we find it conceivable that voters could perceive it less important, for example, if an increase in unemployment amounts to 11 or 14 percentage points than if it totals 1 or 4 percentage points. Thirdly, the effects of the pandemic on turnout and whether they differ across partisan groups are unclear, notwithstanding additional reports about fears of active efforts to asymmetrically decrease turnout for the benefit of Republicans (e.g. Smith 2020).

As all these factors might contribute to a closer election outcome than our model suggests, they pose strong and important caveats to our forecast. It should therefore be kept in mind that our forecast comes with a relatively large range of uncertainty. Taken together, this suggests that, although our model projects Democrat Joe Biden to win the election with a very high probability, one should be prudent in its interpretation as there is also a significant amount of scenarios in which Trump might win a second term.

8. References

- Abramowitz, Alan I. (2008): Forecasting the 2008 Presidential Election with the Time-for-Change Model. In: *PS: Political Science & Politics* 41 (4), pp. 691–695.
- American Presidency Project (2020): Presidential Job Approval. Available online at <https://www.presidency.ucsb.edu/statistics/data/presidential-job-approval>, checked on 9/12/2020.
- Berry, Michael J.; Bickers, Kenneth N. (2012): Forecasting the 2012 Presidential Election with State-Level Economic Indicators. In: *PS: Political Science & Politics* 45 (4), pp. 669–674.
- Campbell, James E. (1992): Forecasting the Presidential Vote in the States. In: *American Journal of Political Science* 36 (2), pp. 386–407.
- Campbell, James E.; Norpoth, Helmut; Abramowitz, Alan I.; Lewis-Beck, Michael S.; Tien, Charles; Erikson, Robert S. et al. (2017): A Recap of the 2016 Election Forecasts. In: *PS: Political Science & Politics* 50 (2), pp. 331–338.
- DeSart, Jay A. (2020): A Long-Range State-Level Forecast of the 2020 Presidential Election. In: *PS: Political Science & Politics*.
- Devine, Christopher J.; Kopko, Kyle C. (2013): Presidential Versus Vice Presidential Home State Advantage: A Comparative Analysis of Electoral Significance, Causes, and Processes, 1884-2008. In: *Presidential Studies Quarterly* 43 (4), pp. 814–838.
- Duch, Raymond M. (2007): Comparative Studies of the Economy and the Vote. In: Carles Boix, Susan C. Stokes (Eds.): *The Oxford Handbook of Comparative Politics*. Oxford: Oxford University Press, pp. 805–844.
- Duch, Raymond M.; Stevenson, Randy (2010): The Global Economy, Competency, and the Economic Vote. In: *The Journal of Politics* 72 (1), pp. 105–123.
- Enns, Peter K.; Lagodny, Julius (2020): Forecasting the 2020 Electoral College Winner: The State Presidential Approval/State Economy Model. In: *PS: Political Science & Politics*.
- Ferejohn, John (1986): Incumbent performance and electoral control. In: *Public Choice* 50 (1), pp. 5–25.
- Garz, Marcel; Martin, Gregory J. (2020): Media Influence on Vote Choices: Unemployment News and Incumbents' Electoral Prospects. In: *American Journal of Political Science*.
- Healy, Andrew; Lenz, Gabriel S. (2014): Substituting the End for the Whole. Why Voters Respond Primarily to the Election-Year Economy. In: *American Journal of Political Science* 58 (1), pp. 31–47.
- Holbrook, Thomas M. (1991): Presidential Elections in Space and Time. In: *American Journal of Political Science* 35 (1), pp. 91–109.
- Jerôme, Bruno; Jérôme-Speziari, Véronique (2012): Forecasting the 2012 US Presidential Election: Lessons from a State-by-State Political Economy Model. In: *PS: Political Science & Politics* 45 (4), pp. 663–668.

- Jerôme, Bruno; Jérôme-Speziari, Véronique (2016): State-Level Forecasts for the 2016 US Presidential Elections: Political Economy Model Predicts Hillary Clinton Victory. In: *PS: Political Science & Politics* 49 (4), pp. 680–686.
- Jérôme, Bruno; Jérôme, Véronique; Mongrain, Philippe; Nadeau, Richard (2020): State-Level Forecasts for the 2020 US Presidential Election: Tough Victory Ahead for Biden. In: *PS: Political Science & Politics*.
- Key, Valdimer O. (1966): *The Responsible Electorate. Rationality in Presidential Voting 1936-1960*. Cambridge: Belknap Press of Harvard University Press.
- King, Gary; Tomz, Michael; Wittenberg, Jason (2000): Making the Most of Statistical Analyses: Improving Interpretation and Presentation. In: *American Journal of Political Science* 44 (2), pp. 347–361.
- Klarner, Carl E. (2012): State-Level Forecasts of the 2012 Federal and Gubernatorial Elections. In: *PS: Political Science & Politics* 45 (4), pp. 655–662.
- Lauderdale, Benjamin E.; Linzer, Drew (2015): Under-performing, over-performing, or just performing? The limitations of fundamentals-based presidential election forecasting. In: *International Journal of Forecasting* 31 (3), pp. 965–979.
- Lewis-Beck, Michael S.; Rice, Tom W. (1982): Presidential Popularity and Presidential Vote. In: *Public Opinion Quarterly* 46 (4), pp. 534–537.
- Lewis-Beck, Michael S.; Rice, Tom W. (1983): Localism in Presidential Elections: The Home State Advantage. In: *American Journal of Political Science* 27 (3), pp. 548–556.
- Lewis-Beck, Michael S.; Rice, Tom W. (1984): Forecasting Presidential Elections: A Comparison of Naïve Models. In: *Political Behavior* 6 (1), pp. 9–21.
- Lewis-Beck, Michael S.; Stegmaier, Mary (2000): Economic Determinants of Electoral Outcomes. In: *Annual Review of Political Science* 3, pp. 183–219.
- MIT Election Data And Science Lab (2017): U.S. President 1976–2016. Harvard Dataverse. Available online at <https://doi.org/10.7910/DVN/42MVDX>, checked on 9/12/2020.
- Mixon, Franklin G.; Tyrone, J. Matthew (2004): The ‘Home Grown’ Presidency: empirical evidence on localism in presidential voting, 1972–2000. In: *Applied Economics* 36 (16), pp. 1745–1749.
- Nannestad, Peter; Paldam, Martin (2002): The cost of ruling. A foundation stone for two theories. In: Han Dorussen, Michael Taylor (Eds.): *Economic Voting*. London, New York: Routledge, pp. 17–44.
- Powell, Richard J. (2004): The Strategic Importance of State-Level Factors in Presidential Elections. In: *Publius* 34 (3), pp. 115–130.
- Rosenstone, Steven J. (1983): *Forecasting Presidential Elections*. Yale University Press.
- Sigelman, Lee (1979): Presidential Popularity and Presidential Elections. In: *Public Opinion Quarterly* 43 (4), pp. 532–534.

Smith, David (2020): Trump v American democracy: the real battle on the ballot this November. The Guardian. Available online at <https://www.theguardian.com/us-news/2020/aug/23/donald-trump-v-democracy-us-election>, checked on 9/15/2020.

US Bureau of Labor Statistics (2020): Local Area Unemployment Statistics. Available online at <https://www.bls.gov/lau/data.htm>, checked on 9/12/2020.

9. Appendix

Figure A.1a: Distribution of forecasts for the incumbent two-party-vote share across states. The red dashed line indicates a vote share of 0.5.

Figure A.1b: Distribution of forecasts for the incumbent two-party-vote share across states. The red dashed line indicates a vote share of 0.5.

Figure A.1c: Distribution of forecasts for the incumbent two-party-vote share across states. The red dashed line indicates a vote share of 0.5.

Figure A.1d: Distribution of forecasts for the incumbent two-party-vote share across states. The red dashed line indicates a vote share of 0.5.