POLITICS

Special Issue on Forecasting the 2024 US Elections

A Dynamic Forecast: An Evolving Prediction of the 2024 Presidential Election

Spencer Lindsay, *University of Arizona, United States of America*

Levi Allen, *Indiana State University, United States of America*

Abstract: In this article, we build a model to predict the state-level results of the 2024 election. We do so by using polling from similar points in past election cycles and by using the results of the previous election. Notably, we update our model over time and the coefficients of the two variables change- the model puts more weight on polling as the election gets closer. As of September 1, 2024, we find that Kamala Harris is a narrow favorite to win the 2024 election, with a 57 percent chance of doing so. Currently the model predicts she will win 289 electoral votes to Trump's 249. However, there remains significant uncertainty and the model will continue to update as the election nears.

This is a "preproof" accepted article for *PS: Political Science & Politics*. This version may be subject to change during the production process.

DOI: 10.1017/S1049096524000878

The 2024 presidential election has many unique characteristics. While it at first appeared to be the first rematch for the Presidency since Dwight Eisenhower defeated Adlai Stevenson for a second term in 1956, it will instead be a race between a former President and the first woman of color to lead a major party's presidential ticket. Not since 1940 and Herbert Hoover's ill-fated attempt to claim the Republican Party's nomination for President has a former occupant of the White House attempted to reclaim the nation's highest office. The current matchup between Donald Trump and Kamala Harris results from the sitting President, and winner of his party's primary process, stepping aside from immense pressure from within, and outside, his own political party. We have also seen an attempt on the former President's life, which, to many, seems like a distant memory. This creates a complex landscape that challenges election forecasters in unique ways. Despite these challenges, elections tend to follow predictable patterns, and we rely on this knowledge to predict the 2024 election. To that end, we have constructed a model that is responsive to the everyday fluctuations that exist in a campaign environment while also taking into context states' electoral history.

How our Model Works

Our model has two variables: the previous election results of a given state, as measured by the margin of the winning candidate, and the average of current polls in that state. Consequently, our model provides a balance of contemporary, present-day horserace polling paired with an accounting of the partisanship and vote history of a state in the most recent presidential election. Notably, our model's coefficients change over time; as the election approaches, we recalculate our coefficients, and the model tends to put more weight on polls and less on the previous election result to account for polling becoming a more accurate predictor as the election cycle goes on. Below, we begin by discussing how we calibrate our model, provide a brief description of these two variables, and how we collect and measure them for the purposes of constructing our forecasting model.

Nuts and Bolts: Calibration

To collect data for the model, we use data sources that documented both general election polling by state and previous election results. To gather previous polling averages, we leverage the data collected by two data sources: *Real Clear Politics (RCP)* and *FiveThirtyEight (538)*. Simple polling averages from 2004 through 2012 are constructed using polls documented in the *RCP* databases and the 2016 and 2020 polling averages are constructed from polling in 538's (more comprehensive) database. To gather election results, we collect data from the Federal Election Commission (FEC) reports from the years 2000 through 2020.

We create simple averages based on the polls in said databases that were available at the date to which the model is calibrated, in this case, September 1. We pre-selected 19 states[1] in March which we viewed as having the potential to be the most competitive. We analyze how the polling – and the errors associated with that polling – changed over time across five election cycles, re-estimating the model at six dates between April 15 and Election Day (see Table 1). This approach emphasizes correctly estimating the outcomes and win probabilities of the states most likely to decide the election and doing so only using data that was available at similar points in the election cycle.

Our reasoning for this is that polling averages become more accurate as the election nears. For example, on April 15, the earliest calibration of our current model, the model puts more weight on past results than on poling. But, as we near election day, we find that the model will typically

[1] Those states are: Alaska, Arizona, Colorado, Florida, Georgia, Iowa, Maine, Michigan, Minnesota, Nevada, New Hampshire, New Mexico, North Carolina, Ohio, Pennsylvania, Texas, Utah, Virginia, and Wisconsin.

continue to put more weight on the polls, providing us with something of akin to a weighted average between the polls and the prior results that changes as the election approaches. Thus, the weight of each factor changes as we near election day. Our model is estimated using an OLS regression and is estimated without an intercept, for reasons we discuss below.

Polling Average

To create a polling average for each state, whenever possible, we only include polls from pollsters that are ranked in the Top 100 as determined by the reputable outlet, *FiveThirtyEight,* whenever such polls are available. This equates to just over one-third of tabulated pollsters and includes polling giants such as The New York Times/Siena College, Monmouth's University Polling Institute, YouGov, Selzer, Marist College and Quinnipiac. Furthermore, whenever possible, we only utilize these pollsters if they have a documented history of polling presidential general or primary elections or other statewide general election races in the past and only if the poll in question does not have a partisan sponsor.[2] These selection criteria were chosen to avoid the inclusion of low-quality pollsters who, despite much media attention, may skew the averages with questionable data. When no such polls are available, we give preference to higher ranked (top 150) polls. We base our averages on the margin between the Democratic and Republican candidates. When possible, we use likely voter samples, but will include registered voter samples if the former is unavailable. Additionally, given that there are no third-party candidates currently in the race consistently polling above 1 percent, we include head-to-head polls whenever possible.

[2] This excludes Abt Associates, Innovative Research Group, Kaplan Strategies, Ward Research and CVoter, who, as far as we can tell, have not polled a presidential primary or general election, nor any statewide general election race prior to this cycle.

Polls are included in the model if they are recent (within a month)[3] and we follow the following process for maintaining the averages: when new polls are added to each state's average, old polls are deleted if (1) they are more than one month old *and* (2) deleting the poll would not cause there to be less than three polls in the state's average. To compute our averages, we weight by sample size and the age of the polls relative to others in the average.[4]

Previous Result

We include in our model the result from the previous presidential election. We view this as an important component of the model for two reasons. First, it represents real, actual data that is reflective of how voters cast their ballots in the past; it is, in some ways, the ultimate horse-race poll. Second, this variable allows us to account for state-level partisanship as well as the elastic nature – by which we mean how much states tend to fluctuate between presidential elections – in our model. Through this, the model may capture factors that polls may not and help to account partially for some potential biases within polling that arise from methodological challenges, such as partisan non-response bias. We measure this variable by taking the margin between Joe Biden and Donald Trump from the previous presidential election. For example, Donald Trump garnered 57.02% of the vote in Indiana in the 2020 presidential election compared to Joe Biden's 41.96%. Thus, the margin is rounded and entered into the model as 15.1 percentage points.

[3] Notably, we include only polls in which Harris is the candidate and after Biden had withdrawn from the race.
[4] To weight for the age of the poll, the newest polls are generally weighted at an 8 and the weight reduced by 1 for each week older the poll is until it ages out of the model. If the average spans more than eight weeks, the weight of the newest poll will be 16 and the same process will apply to older polls.

Other Modeling Considerations and Strategies

Our model is notably "simple," which we argue is a strength; parsimony, after all, is something desirable in political science, and given that we are relying on past electoral performance and the best performing polls in recent cycles, we are confident in the integrity of the model. However, there are many other variables that have been used – to varying degrees of success – in modeling presidential outcomes, such as incumbency (Abramowitz 2016), varying measures of economic growth or sentiment (Erikson and Wlezien 2021; Lewis-Beck and Tien 2021; Lockerbie 2020), war deaths (Hibbs 2000), presidential approval (Abramowitz 2016), primary support (Norpoth 2021), perceived competency and leadership capability (Graefe 2021), candidate's home state (DeSart 2021) and the number of consecutive terms served by the incumbent party (Abramowitz 1988, DeSart 2021).

We take seriously the possibility that, in our pursuit for an accessible and parsimonious model, we have left out variables that may matter. We put forward a three-pronged justification of our model. First, we are skeptical that adding additional variables will further refine the model and not create a problem of overfitting. For instance, overfitting can lead to overconfidence in the output of our models, accentuating some "idiosyncratic features" that may be specifically powerful in explaining some presidential elections, but not others (Dowding and Miller 2019).

Second, many models incorporate a state-level approach with a minimum number of variables to great success. For example, Campbell and Wink (1990) build a model that was comprised of the Gallup Poll trial-heat question and real GNP growth, borrowing that variable from Lewis-Beck and Rice (1984). Others also leverage the economic variables within the state context, utilizing unemployment change, presidential job approval, and local partisan domination (Jerôme and Jerôme-Speziari 2012). Others still incorporate limited variables in the statewide context, such as asking respondents who they believe will win in the upcoming election (Murr and Lewis-Beck 2020). Consequently, we view ourselves on sure ground for positing a model with a small number of variables at the state level.

Finally, we believe that horse-race polling may account for much of these potentially confounding variables. For example, if one is dissatisfied with the current economic condition of the United States, one is likely incorporating that into your evaluation of who to vote for in the election, which is then reflected in the preference that you provide to a pollster. Furthermore, the unit of analysis for many of these variables is this election cycle at the state level. Thus, our observations are drawn from just five presidential elections, and we want to ensure that cycle-level factors do not lead us to incorrectly estimate the coefficients by placing undue certainty on our results due to temporal or idiosyncratic factors.

Notably, we are not the first to incorporate polling and past election results in a model forecasting presidential elections at the state level, though our approach in doing so differs from DeSart's 2020 model. DeSart (2021) attempted to forecast the election using a variable aggregating several past election results and polling from a year prior to the election in addition to the candidate's home state and the number of terms the incumbent party had been in power. We, instead, use just the previous election results and calibrate our model to changes in polling over time. Rather than predicting the election from a single date, our model changes, and becomes less error prone, over time.

Data Collection and Building the Model

The coefficients of the independent variables change over time as the model is recalibrated due to the election date becoming closer and closer. As it nears, we find that the model will generally put more weight on polling and less on previous election results given that polling becomes more predictive of election outcomes as we near the actual date of the election. [5] How much weight the model places on each variable at various points in the cycle can be seen in Table 1. We expect the final update to come the night before the election. The model can be viewed throughout the election cycle at the attached link, which we provide to readers to see how the model continues to evolve over time post-publication and to increase the transparency of how we reach the forecast below.[6]

We estimate outcomes using an Ordinarily Least Squares (OLS) regression model and elect to do so without an intercept. We do this because the intercept is likely to be the result of aggregate cycle-level errors, of which our concern with which is noted in the discourse above. It is unclear that the intercept provides any new information beyond which party has benefitted from cycle-level polling error in the past five elections. Notably, polling "bias tends to shift unpredictably from election to election" (Silver 2018). Therefore, we believe estimating the model without the intercept is the best choice. We are assuming, instead, that if polls and previous election results both indicate a tie, we should assume that the election will come close to being a tie. We then create a point estimate for each observation based on the model and create an estimated error term using the model's residuals based on the square root of the mean of squared errors.

To create uncertainty estimates, we gather a z-score using the expected error and take the p-value, assuming a normal distribution. For the overall outcome, we take the percentage chance a

[5] For states in which no polling data is available, we estimate the result using an OLS model with only the previous result as a predictor.
[6] Model Link: https://docs.google.com/spreadsheets/d/1lY2K4ZINICHXUbFjrcVFcNv4PL-e1OHIk1SoxO3yu0M/edit?gid=705324711#gid=705324711

given candidate has of winning in the tipping point state as their percentage change of winning the race. In doing so, we are assuming uniform errors in model's estimates. While this is not the case in reality, we do know that polling error is correlated between states (Silver 2016) and that the assumption of uniform error is better than the assumption of independent error (ibid). The model additionally provides an estimated popular vote outcome by taking an average of states' predicted outcomes, weighted by the total number of votes in the previous election. The model also gives us a mean electoral vote outcome based on the sum of the candidate's expected electoral votes from each state.

Our forecast relies on three simple values: (1) we predict election based only on information that was available at the same time point in previous election cycles, (2) we predict state election outcomes because they are determinative of the outcome of the election and, (3) we do not make assumptions about the direction of error and avoid modeling choices that may lead us to do so unduly.

The Model

Table 1 shows the model calibrated at six different points spanning from April 15 through election day. It shows that the amount of weight the model puts on each variable shifts overtime. We use the following equation to predict the outcome of the election on September 1.

Dem. Margin = (.732* Polled) +(.316*Previous)

However, on election day, the model puts more weight on polls and less on the previous election margin, though both remain significant at $p<.001$. The November equation is shown below.

Dem. Margin = (.887* Polled) +(.186*Previous)

Notably, as election day near, each model has a higher R^2, indicating that as election day approaches, each model explains a higher portion of the variance (as expected). Also, the expected error of the model's prediction changes over time. The expected error of the April 15 model is 6.9 points, whereas the election day model has an expected error of just 3.4 points. As the election grows closer, the model's predictions become more precise.

Table 1: Regression Models Predicting Election Outcomes Throughout an Election Year

Variable	April 15	June 1	July 15	Sept. 1	Oct. 1	Election Day
Polled Dem. Margin	.333* (.110)	.679* (.096)	.746* (.082)	.732* (.078)	.778* (.073)	.887* (.053)
Previous Dem. Margin	.522* (.072)	.287* (.081)	.309* (.064)	.316* (.062)	.238* (.062)	.186* (.044)
N	87	92	94	94	94	94
R^2	.65	.77	.82	.83	.85	.91
Expected Error of Model Output (1 SD)	+/- 6.94	+/- 5.57	+/- 5.01	+/- 4.91	+/- 4.61	+/- 3.41

Note: *- p<.01

To test how the model would have performed in previous election cycles, we test the election-day model error against the error of polling alone. Additionally, we test this using both coefficients from the full sample and coefficients in which the election being tested is taken out of the sample to provide an out-of-sample test. The error of the model excluding the election being predicted from the sample was within 0.2

points on average, between 2016 and 2020, of the error of the model when using the full sample. We find that regardless of model specification, our model beat the error of election day polling in both 2020 and 2016. Additionally, we find that the errors of the out-of-sample test are fairly similar to the full-sample errors.

Table 2: Model Error vs. Polling Error in 2016 and 2020

	Full Sample Coefficients	Coefficients Excluding Tested Election from Sample	Election Day Polling
2016	4.87	4.90	5.20
2020	3.73	4.14	4.65

Table 3 shows our model's outputs from key states as of September 1, 2024, while Figure 1 shows the model's predicted winner in each state and Washington DC. We find that Kamala Harris is currently a narrow favorite, and assuming uniform errors, that she has a 57% chance of winning the Presidency. The average outcome is Harris winning the electoral college 289-249. The median outcome is Harris winning 292 electoral votes to Trump's 246. Our model also currently estimates Harris will win the popular vote by 3.8 percentage points. The model currently gives Harris a 95 percent chance of receiving between 162 and 447 electoral votes.

Table 3: Model Outputs in States of Interest[7]

State	Predicted Outcome	Leader Win Probability
Arizona	Trump +0.2	52%

[7] The full predictions for all 50 states on September 1 are available in the Appendix

Florida	Trump +4.4	82%
Georgia	Harris +0.4	53%
Michigan	Harris +2.6	70%
Minnesota	Harris +6.1	89%
Nevada	Harris +1.2	60%
New Hampshire	Harris +6.4	91%
New Mexico	Harris +8.5	96%
North Carolina	Trump +0.9	57%
Ohio	Trump +9.8	98%
Pennsylvania	Harris +1.8	65%
Texas	Trump +5.4	87%
Virginia	Harris +5.4	86%
Wisconsin	Harris +0.8	57%

Figure 1: September 1 Prediction of 2024 Presidential Race

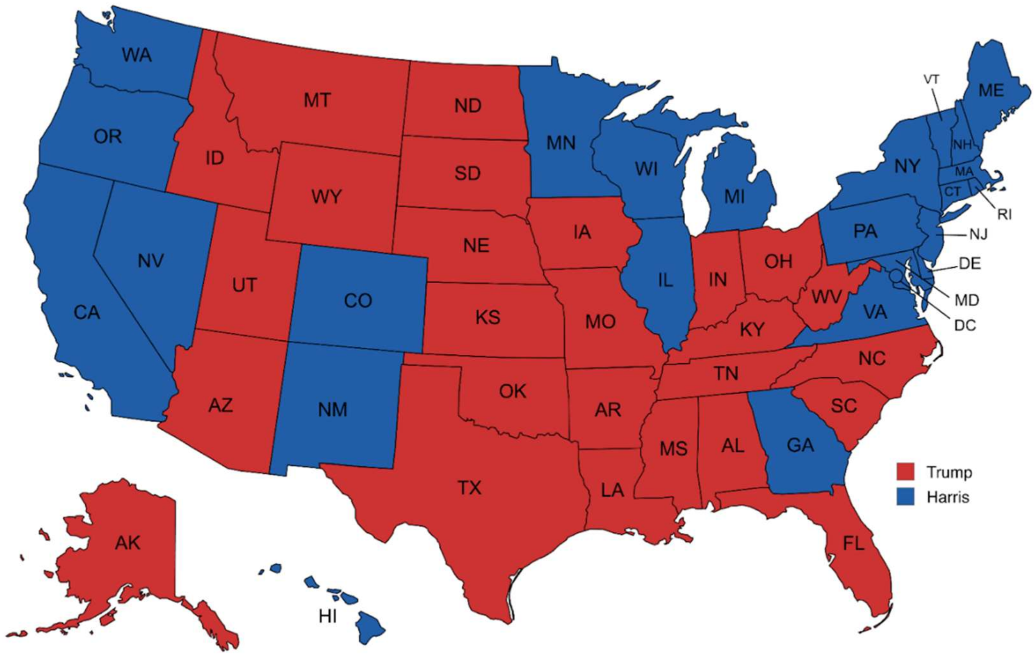


Figure 2: Democrat's Projected Odds of Winning 2024 Election Over Time

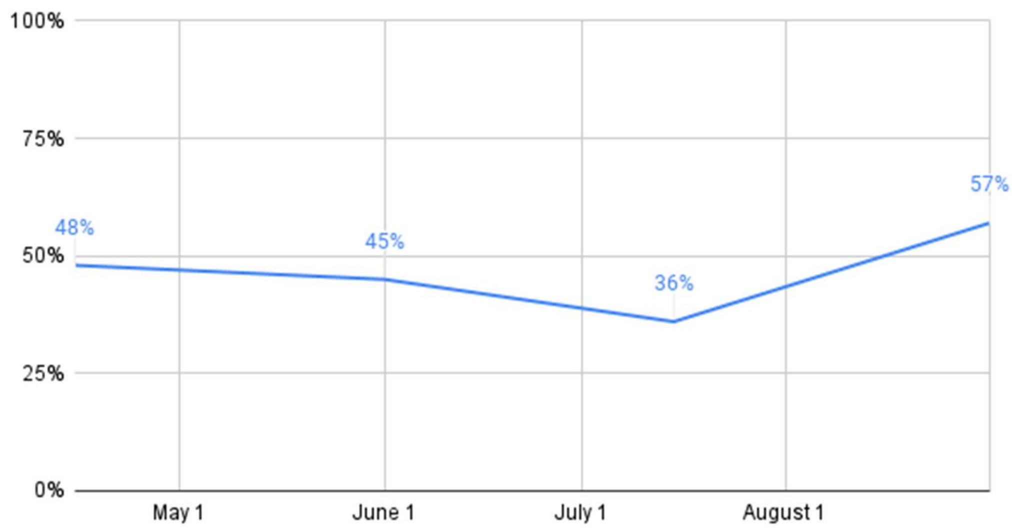
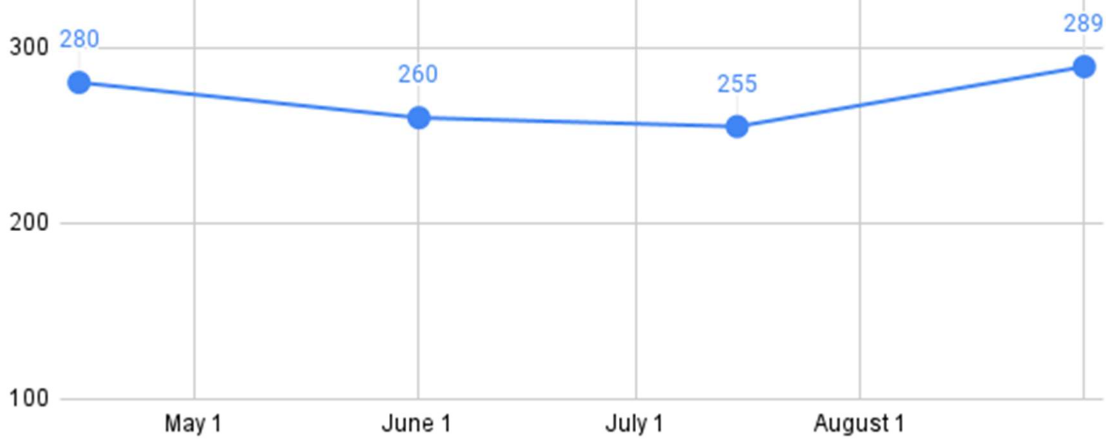


Figure 3: Democrat's Projected Average Electoral Votes in 2024 Election Over Time



Figures 2 shows the Democrats' probability of winning the election over time. Currently, Kamala Harris is favored and has a 57% chance of winning the Presidency. On July 15, shortly before Biden dropped out of the race, his odds of winning were just 36 percent. This indicates that Democrats' chances of winning the presidency are 21 percentage points higher now than they were on July 15. In a separate analysis available in the appendix, we find that the change in candidates led to a 3.3-point increase in the Democrats' standing in national polling. In light of this data, it appears that the choice of Biden to withdraw from the race greatly improved the Democrats' chances of holding the Whitehouse.

It is important to note that there is significant uncertainty associated with predicting an election at this stage, and while uncertainty does decrease as the election approaches, forecasters are never fully free from it. Our goal is not to downplay that uncertainty, but rather to embrace it and do the best we can with the data available to us. As the election gets closer, we will have a better idea of what will happen.

In this spirit of transparency, it is also important to acknowledge two important limitations of this model. First, it does not have a way to recognize discrete campaign events that may have a fleeting effect on the polls. If, for example, the Democratic National Convention had an effect on Harris' polling that was only temporary, rather than durable, as the effects of conventions tend to be (Erikson and Wlezien 2012), our model may currently be underestimating Trump's chances of winning. Second, in places where there is less polling available or where the polling is less current, the model's estimates may be less accurate. Thankfully, states that are more likely to decide the election are polled more often, and the volume of polling tends to increase as the election approaches, so the estimates are likely to become more accurate as the election approaches and are likely more accurate in decisive states.

Conclusion

Our model is parsimonious, balances two important yet simple criteria, and is accessible to the general public. We have also attempted to incorporate Victor's (2021) recommendations for election forecasters in this piece. For instance, we place a heavy emphasis on the precise parameters – in this case, horserace polling and previous election results – that are predicting the outcome. We have also endeavored to be transparent, providing access to our data and recognizing the uncertainty associated with our predictions by discussing the amount of confidence we have in the results.

Currently Harris is a narrow favorite to win the election, but Trump remains in strong position for an upset. Given the stark policy difference between these two candidates, the stakes are tremendously high. A Harris Presidency is likely to bring about new pushes for abortion rights protections at the federal level, attempts to enshrine protections on voting rights into federal law, continued pressure on Russia to withdraw from Ukraine and an expansion on economic regulations and social spending. A Trump presidency, on the other hand, would likely bring about a remaking of the federal bureaucracy in Trump's image, harsher immigration policy and an embodiment of authoritarian leaders across the world. Our model shows that the former is slightly more likely, but the latter remains a strong possibility. Only on November 5, when over 125 million voters will decide the fate of the nation, will we begin to understand what is in store for the future of the United States.

Acknowledgements: *We would like to thank Jeffery Harden (University of Notre Dame) for his informal advice during the review process*

Data Availability Statement: *Research documentation and data that support the findings of this study have not yet been verified by PS's replication team. Data will be openly available at the Harvard Dataverse upon publication of the final article.*

Conflict of Interest Statement: *The authors declare no ethical issues or conflicts of interest in this research.*

References

Abramowitz, A. I. (1988). An improved model for predicting presidential election outcomes. *PS: Political Science & Politics*, *21*(4), 843-847.

Abramowitz, A. I. (2016). Will time for change mean time for Trump?. *PS: Political Science & Politics*, *49*(4), 659-660.

Campbell, J. E. and K.A. Wink (1990). Trial-Heat forecasts of the presidential vote. *American Politics Research*, 18(3), 251-269.

DeSart, J. A. (2021). A long-range state-level forecast of the 2020 presidential election. *PS: Political Science & Politics*, *54*(1), 73-76.

Dowding, K. and C. Miller. (2019). On prediction in political science." *European Journal of Political Science* 58(3): 1001-1018.

Erikson, R. S., & Wlezien, C. (2012). *The timeline of presidential elections: How campaigns do (and do not) matter*. University of Chicago Press.

Erikson, R. S., & Wlezien, C. (2021). Forecasting the 2020 presidential election: Leading economic indicators, polls, and the vote. *PS: Political Science & Politics*, *54*(1), 55-58.

Graefe, A. (2021). Of Issues and Leaders: Forecasting the 2020 US Presidential Election. *PS: Political Science & Politics*, *54*(1), 70-72.

Hibbs Jr, D. A. (2000). Bread and peace voting in US presidential elections. *Public Choice*, *104*(1), 149-180.

Jerôme, Bruno and V. Jerôme-Speziari. 2012. Forecasting the 2012 US presidential election: lessons from a state-by-state political economy model. *PS: Political Science and Politics*.

Lewis-Beck, M. S., & Tien, C. (2021). The political economy model: A blue wave forecast for 2020. *PS: Political Science & Politics*, *54*(1), 59-62.

Lindsay, S. (2024, January 16). Polls overestimate support for third-party candidates. Political Wire. https://politicalwire.com/2024/01/16/polls-are-overestimating-support-for-third-party-candidates/

Lewis-Beck, M.S. and T. W. Rice. (1984). Forecasting presidential elections: a comparison of naïve models. *Political Behavior* 6(1): 9-21.

Lockerbie, B. (2021). Economic Pessimism and Political Punishment in 2020. *PS: Political Science & Politics*, *54*(1), 67-69.

Murr, A. E. and Lewis-Beck, M. S. (2020). Citizen forecasting 2020: a state-by-state experiment. *PS: Political Science and Politics* 54(1): 91-95.

Norpoth, H. (2021). Primary model predicts trump reelection. *PS: Political Science & Politics*, *54*(1), 63-66.

Silver, N. (2016, Nov. 11). Why *FiveThirtyEight* gave trump a better chance than almost anyone else. https://fivethirtyeight.com/features/why-fivethirtyeight-gave-trump-a-better-chance-than-almost-anyone-else/

Silver, N. (2018, May 30). The polls are all right. *FiveThirtyEight*. https://fivethirtyeight.com/features/the-polls-are-all-right/

Victor, J. N. (2021). Let's Be Honest about Election Forecasting. *PS: Political Science & Politics*, *54*(1), 107-110.

All replication material available at: *Lindsay, Spencer and Levi Allen. 2024. "Replication Data for: A Dynamic Forecast: An Evolving Prediction of the 2024 Presidential Election." PS: Political Science & Politics. DOI: XXXXXXXXX.*