Forecasting US Voter Turnout

Michael Bednarczuk, *Austin Peay State University, USA*

ABSTRACT Voter turnout is a crucial indicator of democratic health, yet forecasting turnout remains an understudied area in political science. This article presents two pioneering models for predicting US presidential election turnout: the national and the state model. The national one, using data from 1868 from 2020, employs lagged turnout as its sole predictor. The state model, covering 1984 to 2020, incorporates demographic and institutional variables to forecast state-level participation. The national predicts 65.3% turnout for 2024, whereas the state model forecasts increased turnout in 41 states compared to 2020. The models' ability to generate early predictions offers valuable lead time for planning and resource allocation, which has implications for election administrators and political campaigns, as well as for the vibrancy of civic engagement in America.

WHY TURNOUT?

Predicting elections has a long history in *PS: Political Science & Politics*. Although early predictions focused on forecasting the winner of the popular vote for the US presidency, more recent models have expanded to predicting the winner of the Electoral College or the distribution of congressional seats.

However, predicting election *turnout* is a far less-studied phenomenon. The most recent article on the topic is now more than a decade old. Evans and Ivaldi (2012) built a model to predict turnout for European Union elections. They found that lagged turnout was the most important predictor, but that electoral circumstances and structures—such as the absence or presence of a compulsory voting system, the time since the last election, and the number of parties—also mattered. In sharp contrast to forecasts of elections, no economic variables were significant. Furthermore, no studies to date have focused on forecasting turnout in the United States.

Voter turnout is a fundamental pillar of democracy and a key indicator of the health of a country. Participation in elections has profound implications for the proper functioning of government. Turnout rates are linked to important aspects of democratic governance, ranging from the accuracy of representation to the perceived legitimacy of those who are elected.

The importance of turnout can be seen in examining its role across different political contexts. In authoritarian regimes, for instance, reported turnout figures are frequently viewed as manipulated or coerced and are often met with skepticism by international observers. Conversely, in established democracies, low turnout rates can be a source of considerable concern and frustration, especially for elections that receive less attention, such as midterm congressional races or special elections. In these cases, diminished voter participation may raise questions about the mandate of those elected and the overall health of the democratic process.

Many factors encompassing a range of social, political, and institutional variables help shape turnout. Some of the rules and structures governing the electoral process may inadvertently create barriers to voting, and others may encourage or facilitate greater participation. Differences in social or demographic factors can influence turnout and may reflect varying levels of political engagement, access to information, or feelings of efficacy. Additionally, the perceived importance of a particular election, the competitiveness of races, and the broader political climate may affect how individuals approach each election.

Even though we understand the importance of turnout, as well as the forces that can change it, there has not been work in forecasting turnout in the US context. This article is thus an attempt to lay a foundation for this field.

The exercise of forecasting turnout has several real-world implications. For practical reasons, forecasting may shape how election administrators perform their work. It could assist with preparation for elections to ensure that there are an adequate number of polling places or poll workers. It may also serve as a check against anomalies: forecasting could check against unusually high or low turnout. Given an atmosphere in which many worry about the fair counting of ballots, forecasting turnout may serve as a balm to guard against politicized claims of fraud.

Michael Bednarczuk *is an assistant professor of political science at Austin Peay State University. He can be reached at bednarczukm@apsu.edu.*

© The Author(s), 2024. Published by Cambridge University Press on behalf of American Political Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

doi:10.1017/S1049096524000829

https://doi.org/10.1017/S1049096524000829 Published online by Cambridge University Press

There are also strategic reasons to forecast turnout. Campaigns may use forecasting to better understand the structural forces in elections. Modeling turnout may also help with the distribution of financial resources, because campaigns may wish to prioritize mobilization under some circumstances and persuasion in others.

Two models are introduced in this article: the national and the state model. As their names imply, the former model attempts to predict national turnout, whereas the latter model is a forecast at the state level. The national model predicts that turnout will be 65.3% for 2024, whereas the state model finds that turnout will generally increase across the country from 2020.

CREATING THE MODELS

Data comes from Bednarczuk (2024). In deciding what to model, several considerations were top of mind. Variables should be logical; that is, there should be prima facie support for their inclusion. Ideally, the variables should be available far in advance of the election so that the forecast can be made early; however, this consideration may have to be weighed against the increased precision that may come from the use of data gathered shortly before the election.

Inspired by the literature on voter turnout and on the earlier model of forecasting turnout, I decided to classify the independent variables as either institutional, campaign, economic, or demographic. Institutional variables capture the rules and structures that may shape turnout. These rules include laws such as same-day voter registration (Grumbach and Hill 2022) and automatic voter registration (Kim 2023), both of which have been shown to increase turnout.

Campaign variables focus on features in the political environment that can affect turnout. For example, the presence of other high-profile races, such as senatorial or gubernatorial contests (Jackson 2002; Springer 2012), and of additional candidates (Bol and Ivandic 2022) can increase turnout.

Economic variables include measures of the economic forces that may affect turnout, ranging from the global (Park 2023) to the national (Frank and Martinez i Coma 2023) to the individual; for example, Aytac et al. (2020) recently found that turnout is lower among the unemployed.

Demographic variables include group identities that may shape turnout. For example, Barber and Holbein (2022) recently documented how turnout is much lower among minority citizens and young people.

As with many models, some ideas worked; others did not. Given the novelty of these models, the following sections describe what worked and what did not work when formalizing them.

THE NATIONAL MODEL

The national model is designed to forecast turnout for the country as a whole. It includes the years 1868 through 2020 and is represented by the following formula:

$$Y = \beta_0 + \beta_1(\text{lagged turnout}) + \varepsilon, \text{ where}$$

Y = voter turnout, β_0 = intercept,

β_1 = coefficient for lagged voter turnout, and ε = error term

Data for the initial model came from elections held from 1960 to 2020. The values for the dependent variable came from the United States Election Project and reflect turnout for the voting-eligible population.

Table 1

The National Model

	Years: 1868–2020	
Variable	**Coefficient**	**SE**
Lagged turnout	**0.835**	**0.083**
Constant	0.101	0.053
	N = 39	
	R^2 = 0.7312	

* *Note*: Variables that are significant at the 0.05 level are in **bold.**

Campaign variables included the presence of an elected incumbent, the presence of a sizable third-party challenge, and whether a poll showed a minor party earning more than 10% of the potential vote during an election year. Economic variables included the national unemployment rate for the year and the change in GDP from the prior presidential election. This data came from the Bureau of Labor Statistics and the Federal Reserve Bank of St. Louis (FRED), respectively. Demographic variables included the percent of the country over 65 and the percent of the country that was nonwhite, both of which came from US Census data. Lagged turnout was also included.

Using a time-series cross-section regression model and various combinations of independent variables, the only variable that was significant was prior turnout (table 1). Additionally, the results showed poor model fit. Altogether, this is an unsatisfactory result for forecasting.

What might help explain these results? Two possible explanations center around the sample size and the variation in the dependent variable. Fewer than 20 observations were included, and turnout varied by fewer than 20 percentage points. Given either complicating factor, creating a useful forecast may have been challenging, but the presence of both may have been too much to overcome.

Therefore, I considered additional models. Based on the results of the earlier model, the only independent variable was lagged turnout. Given data availability, I was able to extend the series to the election of 1792 to increase the sample size to 58. Although doing so greatly improved the model fit, there were sizable residuals, which were largely driven by the low levels of turnout in antebellum elections. Thus, another model was considered that only included post–Civil War turnout. This became the national model.

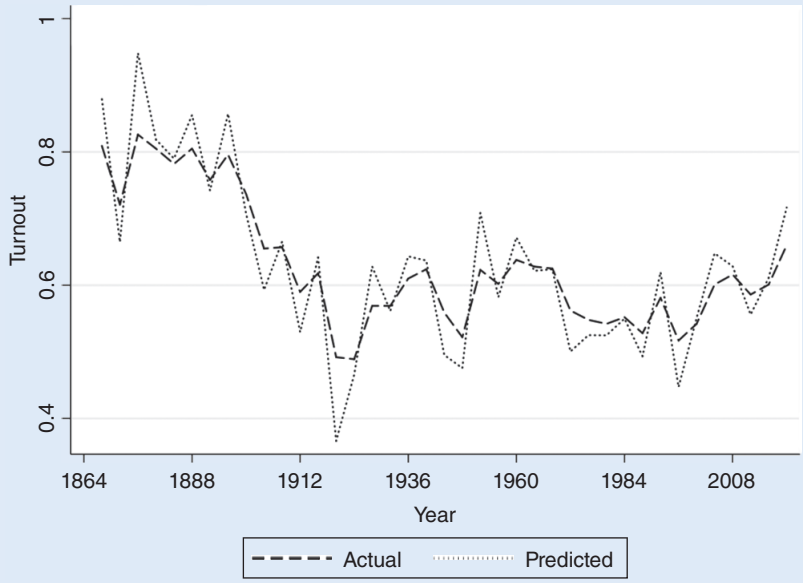
A plot of the predicted results and the actual turnout can be seen in figure 1. Some of the largest misses seem to be driven by times of great change in the composition of the electorate. For example, the predicted turnout of 37% in 1920 was far lower than the actual turnout of 49%; this may be because it was the first presidential election following the ratification of the Nineteenth Amendment.

An additional check analyzed the out-of-sample performance of the model. This was done by dropping an election, reestimating the model, and checking the difference between the prediction for this missing election and the actual turnout rate. Doing this showed that the model fit improved for recent elections; the average miss since 1956 has been less than 3%, with the overall

https://doi.org/10.1017/S1049096524000829 Published online by Cambridge University Press

Figure 1

Plot of Predicted versus Actual National Turnout, 1868–2020



average miss at 4%. For example, the prediction for 2016 was 59%, whereas actual turnout was 60%.

The national model predicts a national turnout for 2024 of 65.3%, with a confidence interval that ranges from 63.6% to 67.0%.

THE STATE MODEL

The state model attempts to forecast turnout state by state. Again, data come from the United States Election Project, which only has state-level data from 1980 through 2020. Given that lagged turnout was included, this meant that only data from 1984 to 2020 elections would be used. Therefore, this state-level model had a sample size of 500. Turnout varies across states from a low of 36.5 (Nevada in 1996) to a high of 74.8 (Minnesota in 2020), with a mean turnout across years of 56.1.

The state model is represented by the following formula:

$$Y = \beta_0 + \beta_1(\text{lagged turnout}) + \beta_2(\text{same-day registration}) + \beta_3(\text{percent white}) + \beta_4(\text{percent college}) + \varepsilon,$$

where Y = voter turnout, β_0 = intercept β_1 = coefficient for lagged voter turnout, β_2 = coefficient for same-day voter registration, β_3 = coefficient for percent of population that is white, β_4 = coefficient for percent of population with a four-year degree, and ε = error term

Several other variables were tested before formalization of this model. First, I introduced a new category of independent variables —institutional—because there are varying laws across states that may shape turnout; for instance, whether states allow same-day registration through Election Day. These data were collected from the National Conference of State Legislatures. Campaign variables included the presence of significant statewide contests (gubernatorial or senatorial) and whether those races featured an incumbent; these data were gathered from a review of elections results from each state. Economic variables included the unemployment rate in the state for the first three months of the election year, as well as the median household income per state. I gathered unemployment data from FRED and derived median household income from census data. Demographic variables included the percent of the state that is nonwhite, the percent that has at least a bachelor's degree, and the percent of the state in poverty, all of which came from census and American Community Survey data. Lagged turnout was also included.

I again used a time-series cross-section regression model. After testing various combinations of the proposed independent variables, the best-fitting model included lagged turnout, same-day registration, race, and education. More college-educated voters and a smaller minority population suggest higher turnout in a state, as does the presence of same-day registration. However, the variable that has the largest effect was lagged turnout (table 2).[1]

To check model fit, I created a plot of the predicted turnout and the actual turnout, as shown in figure 2. The predicted turnout for each state was weighed by its voting-eligible population, aggregated

Table 2

The State Model, 1984–2020

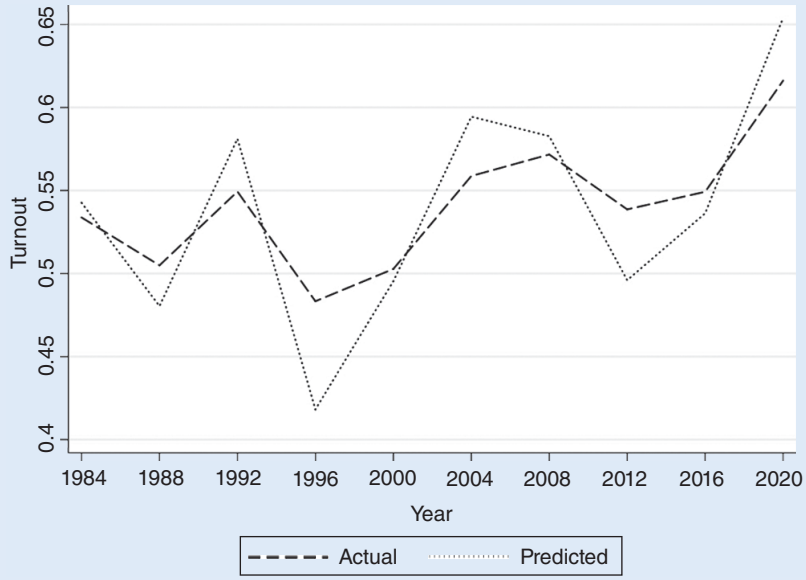
Variable	Coefficient	SE
Lagged turnout	**0.687**	**0.027**
Same-day registration	**2.018**	**0.561**
Percent white	**0.063**	**0.021**
Percent college	**0.237**	**0.025**
Constant	7.243	1.703
N = 500		
R^2 = 0.7092		

* *Note*: Variables that are significant at the 0.05 level are in **bold.**

https://doi.org/10.1017/S1049096524000829 Published online by Cambridge University Press

Figure 2

Plot of Predicted versus Actual Aggregated State Turnout, 1984–2020



by year, and then compared to the actual national turnout. The forecast tracks closely to actual turnout: the average result misses by less than 3%. As an additional check, the out-of-sample performance of this model was also analyzed. Much like in the national, the average miss per state was less than 4%.

This model was then used to forecast turnout for each state for the 2024 election. Data for unemployment came from FRED, whereas college graduate data and data on the racial makeup of states came from the most recent American Community Survey.

Turning to predictions for the upcoming election, 41 states are forecasted to see an increase in their turnout from 2020; the average shift from 2020 turnout is 1.5 percentage points. States seeing the largest prospective gains in turnout tend to be those with same-day voter registration. Their turnout shifts by an average of 2.3 percentage points; those without it move around 0.8 percentage points. Nine states are forecasted to have a turnout higher than 70%, whereas thirteen are expected to see turnout of less than 60% (table 3).[2]

What do these results suggest for the potential swing states in 2024? First, let us look at the Democratic "blue wall" of the Midwest, which consists of Wisconsin, Michigan, and Pennsylvania. All three are high-turnout states, but they have some important differences. Wisconsin's turnout could be higher than it was in 2016, driven by its largely homogeneous population and same-day registration. Michigan should be right behind, brought down slightly by its more diverse population. Lagging these states is Pennsylvania, which is broadly similar in demographics to Michigan but lacks same-day registration. Given the high turnout in these states, motivating people to go to the polls may not be as large a problem as in other states, but Democrats would have to focus more on Pennsylvania than on the other two.

There are other states to consider as well. Potential key states in the Sun Belt include Arizona, Nevada, and Georgia. These are lower-turnout states than the Midwestern states; all three cluster around 60% turnout. This suggests there may be more room for increasing the size of the electorate. Arizona and Nevada both have same-day registration, which may make them easier states to mobilize given their late deadlines. Georgia, in contrast, would require an earlier get-out-the-vote effort.

DISCUSSION

Aside from lagged turnout, Why are there no independent variables for forecasting turnout with the national model? This could be based on its lack of campaign variables. Perhaps the inclusion of measures such as campaign interest and spending or the number of battlegrounds may improve fit; however, those variables would restrict the model to the most recent elections, and such a reduction in sample size may not be worth the trade-off.

A similar critique can be made of the state model. Although related variables, such as down-ballot campaigns, did not have a demonstrable effect on turnout, other electoral variables, such as whether a state is a battleground or swing state, should be considered in additional models. Relatedly, this model does not include any candidate-specific variables. Measures of interest or popularity may be relevant.

Future models may also wish to consider the effect of other demographic variables, especially if the current variables become less predictive over time. Considerations may need to be given to the dependent variable as well; there may be factors that shape the size of the voting-eligible population that should be modeled. Perhaps changes over time in state laws regarding the restoration of felon voting rights should be included.

The lack of economic indicators is also striking. Although these indicators are normally linked to turnout, no combination of economic measures in either model had an effect. That was also found in Evans and Ivaldi (2012)'s forecast of EU turnout. Perhaps these indicators do more to shape vote choice than the choice to vote.

https://doi.org/10.1017/S1049096524000829 Published online by Cambridge University Press

Table 3

The State Model Predicted Turnout, 2024

State	Predicted Turnout	Lower Bound	Upper Bound
Alabama	60.68	59.87	61.49
Alaska	63.82	62.86	64.78
Arizona	59.96	59.22	60.7
Arkansas	54.15	53.7	54.6
California	60.58	59.36	61.8
Colorado	74.11	72.71	75.51
Connecticut	67.99	66.8	69.18
Delaware	64.46	63.33	65.59
Florida	63.34	62.4	64.27
Georgia	60.69	59.59	61.79
Hawaii	55.76	54.32	57.2
Idaho	65.7	64.75	66.65
Illinois	64.83	63.76	65.9
Indiana	59.5	58.88	60.11
Iowa	70.2	69.21	71.17
Kansas	62.26	61.35	63.18
Kentucky	61.58	60.94	62.23
Louisiana	57.64	56.85	58.43
Maine	74.79	73.62	75.96
Maryland	67.2	65.73	68.67
Massachusetts	67.93	66.44	69.43
Michigan	70.27	69.2	71.34
Minnesota	75.13	73.84	76.42
Mississippi	56.42	55.54	57.31
Missouri	63.51	62.7	64.32
Montana	71.9	70.84	72.95
Nebraska	65.34	64.38	66.29
Nevada	59.1	58.05	60.14
New Hampshire	74.52	73.24	75.79
New Jersey	66.52	65.09	67.96
New Mexico	59.73	58.76	60.69
New York	59.08	57.98	60.18
North Carolina	67.65	66.51	68.78
North Dakota	62.82	62.03	63.6
Ohio	64.54	63.67	65.4
Oklahoma	54.29	53.72	54.86
Oregon	69.09	67.91	70.26
Pennsylvania	67.26	66.17	68.35
Rhode Island	63.13	62.04	64.21
South Carolina	61.28	60.41	62.15
South Dakota	63.16	62.37	63.94
Tennessee	58.46	57.83	59.1
Texas	54.93	54.15	55.71
Utah	67.35	66.28	68.42
Vermont	75.07	73.68	76.47
Virginia	67.07	65.7	68.45
Washington	69.71	68.57	70.85
West Virginia	57.46	56.81	58.11
Wisconsin	72.02	70.96	73.09
Wyoming	64.52	63.6	65.44

However, the lack of economic indicators does point to a general strength of both models: they are able to generate a prediction well in advance of an election. The national model can issue its prediction for 2028 once the official turnout records for 2024 are tabulated, whereas the state model will be ready at the beginning of an election year. Given the sometimes competing interests of lead time and accuracy in forecasting, these models decidedly tilt toward the former. This could be useful for both election administrators and political campaigns as they prepare for the fall election cycles.

CONCLUSION

I set out to create a model to forecast turnout for US presidential elections and ended up creating two: the national and the state model. The national model relies strictly on lagged turnout, whereas the state model includes demographic and institutional variables.

This forecast suggests a more positive reading of the health of democracy across the country in 2024 that might be garnered from other sources. Turnout is forecasted to be higher across most of the country for 2024, which is impressive given the high turnout in 2020. Furthermore, although some forces may work against each other—increasing levels of education contrasting with increasing racial diversity—with respect to turnout, there are some reforms that do appear to have a liberalizing effect on turnout, such as same-day voter registration. May this model be one of many to analyze turnout, and may there be a robust debate on this topic in *PS* in the years to come.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit http://doi.org/10.1017/S1049096524000829.

ACKNOWLEDGEMENTS

I am grateful to the anonymous reviewers and to the editors of this symposium for their feedback and guidance.

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 on publication of the final article.

CONFLICTS OF INTEREST

The author declares no ethical issues or conflicts of interest in this research. ■

NOTES

1. A model that mirrored the National Model was also tested, but model fit improved with additional covariates, so the more parsimonious model is omitted.
2. Forecasting this fall's national turnout using the aggregate of state turnout for 2024 would require knowing the voting-eligible population in each state for this year to accurately weight the forecast, so such a prediction is omitted.

REFERENCES

Aytaç, S. Erdem, Eli G. Rau, and Susan Stokes. 2020. "Beyond Opportunity Costs: Campaign Messages, Anger and Turnout among the Unemployed." *British Journal of Political Science* 50 (4): 1325–39. doi:10.1017/S0007123418000248.

Barber, Michael, and John B. Holbein. 2022. "400 Million Voting Records Show Profound Racial and Geographic Disparities in Voter Turnout in the United States." *PLOS ONE* 17 (6): e0268134. https://doi.org/10.1371/journal.pone.0268134.

Bednarczuk, Michael. 2024. "Replication Data for: Forecasting US Voter Turnout." PS: Political Science & Politics. https://doi.org/10.7910/DVN/E4TVRo.

https://doi.org/10.1017/S1049096524000829 Published online by Cambridge University Press

Bol, Damien, and Ria Ivandic. 2022. "Does the Number of Candidates Increase Turnout? Causal Evidence from Two-Round Elections." *Political Behavior* 44 (4): 2005–26. https://doi.org/10.1007/s11109-022-09810-5.

Evans, Jocelyn, and Gilles Ivaldi. 2012. "Deriving a Forecast Model for European Election Turnout." *Political Research Quarterly* 65 (4): 855–67. https://doi.org/10.1177/1065912911421016.

Frank, Richard W., and Ferran Martínez i Coma. 2023. "Correlates of Voter Turnout." *Political Behavior* 45 (2): 607–33. https://doi.org/10.1007/s11109-021-09720-y.

Grumbach, Jacob M., and Charlotte Hill. 2022. "Rock the Registration: Same Day Registration Increases Turnout of Young Voters." *Journal of Politics* 84 (1): 405–17. https://www.journals.uchicago.edu/doi/abs/10.1086/714776.

Jackson, Robert A. 2002. "Gubernatorial and Senatorial Campaign Mobilization of Voters." *Political Research Quarterly* 55 (4): 825–44. https://doi.org/10.1177/106591290205500405.

Kim, Seo-young Silvia. 2023. "Automatic Voter Reregistration as a Housewarming Gift: Quantifying Causal Effects on Turnout Using Movers." *American Political Science Review* 117 (3): 1137–44. doi:10.1017/S0003055422000983.

Park, Brandon B. 2023. "How Does a Relative Economy Affect Voter Turnout?" *Political Behavior* 45: 855–875. https://doi.org/10.1007/s11109-021-09736-4.

Springer, Melanie J. 2012. "State Electoral Institutions and Voter Turnout in Presidential Elections, 1920–2000." *State Politics & Policy Quarterly* 12 (3): 252–83.

https://doi.org/10.1017/S1049096524000829 Published online by Cambridge University Press