UNIVERSITY OF CAMBRIDGE
Faculty of Economics

CAMBRIDGE WORKING PAPERS IN ECONOMICS

Forecasting 2024 US Presidential Election by States Using County Level Data: Too Close to Call

M. Hashem Pesaran
(University of Southern California, and Trinity College, Cambridge)

Hayun Song
(Korea Institute for International Economic Policy)

Abstract

This document is a follow up to the paper by Ahmed and Pesaran (2020, AP) and reports state-level forecasts for the 2024 US presidential election. It updates the 3,107 county level data used by AP and uses the same machine learning techniques as before to select the variables used in forecasting voter turnout and the Republican vote shares by states for 2024. The models forecast the non-swing states correctly but give mixed results for the swing states (Nevada, Arizona, Wisconsin, Michigan, Pennsylvania, North Carolina, and Georgia). Our forecasts for the swing states do not make use of any polling data but confirm the very close nature of the 2024 election, much closer than AP's predictions for 2020. The forecasts are too close to call.

Reference Details

CWPE	2464
Published	10 October 2024
Key Words	Voter Turnout, Popular and Electoral College Votes, Simultaneity and Recursive Identification, High Dimensional Forecasting Models, Lasso, OCMT
JEL Codes	I23, I24, I28
Website	www.econ.cam.ac.uk/cwpe

Forecasting 2024 US Presidential Election by States Using County Level Data: Too Close to Call *

M. Hashem Pesaran
University of Southern California, and Trinity College, Cambridge

Hayun Song
Korea Institute for International Economic Policy

October 19, 2024

Abstract

This document is a follow up to the paper by Ahmed and Pesaran (2020, AP) and reports state-level forecasts for the 2024 US presidential election. It updates the 3,107 county level data used by AP and uses the same machine learning techniques as before to select the variables used in forecasting voter turnout and the Republican vote shares by states for 2024. The models forecast the non-swing states correctly but give mixed results for the swing states (Nevada, Arizona, Wisconsin, Michigan, Pennsylvania, North Carolina, and Georgia). Our forecasts for the swing states do not make use of any polling data but confirm the very close nature of the 2024 election, much closer than AP's predictions for 2020. The forecasts are too close to call.

JEL-Codes: C530, C550, D720

Key Words: voter turnout, popular and electoral college votes, simultaneity and recursive identification, high dimensional forecasting models, Lasso, OCMT.

*We would like to thank Rashad Ahmed for his valuable insights and contributions to this research. We have benefited greatly from his expertise and feedback. We are also thankful to Ron Smith for his comments.

1 Introduction

There are many polls that provide snapshots of the likely outcome of the upcoming November 2024 US presidential election. The national polls predict Kamala Harris will win the popular vote, but given the US system of electoral voting, it is agreed that the outcome of the presidential election is most likely determined by seven so-called swing states: Nevada, Arizona, Wisconsin, Michigan, Pennsylvania, North Carolina, and Georgia, with the results for the remaining 44 states generally well predicted historically. The polls for the swing states are often very close and tend to be unstable from one week to the next. In this paper we follow Ahmed and Pesaran (2020) and produce state level forecasts of voter turnout and the Republication vote shares using a panel econometric model covering 3,107 counties over five election cycles (2004, 2008, 2012, 2016 and 2020).[1]

We confine our analysis to the 48 US mainland states plus the District of Columbia. As with the 2020 forecast exercise we do not include Hawaii and Alaska in our forecasting models, and assume the electoral outcomes in these two states follow a random walk and will be the same as in 2020.

We do not use poll data, but consider a large number of socioeconomic variables at mixed frequencies (totaling 15,535 data points), and use machine learning techniques to select the variables used in forecasting voter turnout and the Republican vote shares for 2024. In particular, we use Lasso (Least Absolute Shrinkage and Selection Operator) due to Tibshirani (1996), and OCMT (One Covariate at a time Multiple Testing) by Chudik, Kapetanios, and Pesaran (2018).

Both variable selection procedures are applied to pooled and regional models as detailed in Ahmed and Pesaran (2020, 2022). As before, we also produce a combined forecast where Lasso and OCMT forecasts are averaged. We ended up with six sets of forecasts: pooled-Lasso, regional-Lasso, pooled-OCMT, regional-OCMT, pooled-LassoOCMT, regional-LassoOCMT.

Before discussing the 2024 forecasts, it is interesting to note that the variables selected by Lasso and OCMT for the extended sample (2000-2020) are very similar to those reported by AP using the sample 2000-2016. The estimates are also very close across the two samples for the voter turnout panel regressions, although there are some notable differences when we consider the parameter estimates for the voting outcome panel regressions. Out of

[1] Ahmed and Pesaran released their 2020 forecasts on October 19, 2020 in the CESifo working paper Ahmed, Rashad and Pesaran, M. Hashem, Regional Heterogeneity and U.S. Presidential Elections (2020). CESifo Working Paper No. 8615, Available at SSRN: https://ssrn.com/abstract=3711891 or http://dx.doi.org/10.2139/ssrn.3711891. The 2020 forecasts were then subsequently evaluated and published in the International Journal of Forecasting in 2022. See Ahmed and Pesaran (2022).

the demographic variables (migration, education, population, income, poverty, rural score), the migration variable saw, by far, the largest change in its estimated coefficient between 2000-2016 and 2000-2020 samples. The coefficients (across both Lasso and OCMT) were previously large and negative (migration was negatively associated with the Republican vote share), but the coefficients have attenuated substantially toward zero in the 2000-2020 sample. This seems consistent with the more recent data reflecting a shift in voter attitudes toward immigration policies. By contrast, the coefficients on the other demographic variables have broadly remained stable.

Turning to the forecasts, all six models predict substantial drops in voter turnout across all states. The results at the national level are summarized in Table 1 with detailed state-level forecasts provided in Tables 6, 7, 8, 9 for voter turnout and the Republican vote share, with state level electoral forecasts summarized in Table 10. At the national level, turnout is predicted to fall from around 66 per cent in 2020 to somewhere between 55 and 57 per cent in 2024. It is worth noting the turnout forecasts reported by AP for 2020 turn out to be very close to the realized value of 66.23 per cent. If true, this could be an important consideration for election outcomes favouring the Republican candidate in the forthcoming election.

Table 1: National Republican Voter Turnout and Vote Share: Realized and Forecast Values for the 2020 and 2024 U.S. Presidential Elections

		Pooled Forecasts				Regional Forecasts			
	Realized	Lasso		OCMT		Lasso		OCMT	
	2020	2020	2024	2020	2024	2020	2024	2020	2024
Voter Turnout	0.6623	0.6512	0.5497	0.6619	0.5500	0.6256	0.5745	0.6197	0.5642
Vote Share	0.4776	0.4529	0.5065	0.4763	0.4725	0.4742	0.4910	0.4935	0.4730
EC Votes	232	188	312	236	261	249	335	270	215

Notes: The realized national Republican vote refers to the 2020 Republican share of the two-party vote across the U.S. mainland states and Washington, D.C. Electoral College (EC) votes represent both realized and predicted Republican electoral votes, assuming Hawaii casts its votes for the Democratic candidate and Alaska and Florida cast their votes for the Republican candidate.

All state-level forecast results are summarized in Table 10. There is a clear divide between Lasso and OCMT forecasts, with pooled and regional Lasso both predicting a Republican win with 312 and 335 electoral votes (well above the 270 votes needed), respectively. In contrast, pooled and regional OCMT models predict Democratic wins with 261 and 215 electoral votes, respectively. A similar divide can also be observed when we consider pooled and regional combined Lasso-OCMT forecasts. The pooled combined procedure, predicting a Republican win with 282 electoral votes, and a Democratic win with 274 electoral votes.

A graphic representation of the forecasts is also given in Figures 7, 8, 9 and 10. The results for the swing states are given in Table 2.

Table 2: US Presidential Election 2024 Forecasts of Swing States Using County Level Data

		Lasso				OCMT			
		Pooled		Regional		Pooled		Regional	
State	Total EC	$\hat{V}_S$	EC Votes	$\hat{V}_S$	EC Votes	$\hat{V}_S$	EC Votes	$\hat{V}_S$	EC Votes
AZ	11	0.5142	11	0.5572	11	0.4805	0	0.5351	11
GA	16	0.5206	16	0.5046	16	0.5132	16	0.4936	0
MI	15	0.5179	15	0.5705	15	0.4843	0	0.5017	15
NC	16	0.5274	16	0.5098	16	0.5122	16	0.4834	0
NV	6	0.5168	6	0.4274	0	0.4898	0	0.4879	0
PA	19	0.5150	19	0.5384	19	0.4748	0	0.4952	0
WI	10	0.5381	10	0.6061	10	0.5097	10	0.5512	10
All Votes	93		93		87		42		36

Notes: Total EC refers to the number of electoral votes per state. EC votes are the predicted Republican electoral votes, excluding Hawaii (Democratic) and Alaska (Republican). Regional forecasts are based on eight panel regressions, one for each BEA region.

The forecast uncertainty remains even if we decided to eliminate those forecasts that did not correctly predict the outcomes for all the 41 mainland non-swing states and the District of Columbia. The situation does not change if we follow Ahmed and Pesaran (2022, p. 673) and focus on pooled forecasts, on the grounds that the time dimension of the panel is very short (5 election cycles) and regional forecasts that allow for parameter heterogeneity could lead to poor estimates and less reliable forecasts. Using pooled Lasso we have a Republican win with 312 electoral votes, and if we use pooled OCMT we have a marginal Democratic win with 261 electoral votes. No clear pattern emerges. The election is forecast to be too close to call.

The rest of this document provides the details of model specifications, the predictors and the data sources, and how the original data set used by Ahmed and Pesaran (2020, 2022) was updated, and the forecasts were generated. A detailed explanation of the variables used and the data cleaning process conducted is also provided in the Data Appendix.

Table 2 (cont.) 2024 Lasso-OCMT Average

		Lasso-OCMT Average			
		Pooled		Regional	
State	Total EC	$\hat{V}_S$	EC Votes	$\hat{V}_S$	EC Votes
AZ	11	0.4973	0	0.5461	11
GA	16	0.5169	16	0.4991	0
MI	15	0.5011	15	0.5361	15
NC	16	0.5198	16	0.4966	0
NV	6	0.5033	6	0.4577	0
PA	19	0.4949	0	0.5168	19
WI	10	0.5239	10	0.5787	10
All Votes	93		63		55

Notes: Total EC refers to the number of electoral votes per state. EC votes are the predicted Republican electoral votes, excluding Hawaii (Democratic) and Alaska (Republican). Regional forecasts are based on eight panel regressions, one for each BEA region.

2 Model specifications for voter turnout and voting outcomes

We consider the same model specifications as in Ahmed and Pesaran (2020, 2022). For county c in region r and election years $t = 2004, 2008, 2012, 2016, 2020, 2024$ we use a panel across counties to model the log-odds ratio of Republican to Democrat votes given by

$$\mathrm{LRO}_{cr,t} = \ln\left(\frac{R_{cr,t}}{D_{cr,t}}\right) = \ln\left(\frac{V_{cr,t}}{1 - V_{cr,t}}\right),$$

where $R_{cr,t}$ and $D_{cr,t}$ represent Republican and Democratic votes, respectively, and $V_{cr,t} = R_{cr,t}/(R_{cr,t} + D_{cr,t})$ is the Republican vote share in year t. We follow the same assumption as in Ahmed and Pesaran (2022), which states that third-party candidates have a negligible effect on the two-party outcomes. The BEA regional classification divides the 48 contiguous states and the District of Columbia into eight regions: New England, Mideast, Southeast, Great Lakes, Plains, Rocky Mountain, Southwest, and Far West.

We use the following panel regression

$$\mathrm{DLRO}_{cr,t+4} = a_{DLRO,r} + \boldsymbol{\phi}_r' \mathbf{z}_{DLRO,cr} + \beta_r \mathrm{VT}_{cr,t+4} + \gamma_r' \mathbf{x}_{DLRO,cr,t+3} + \varepsilon_{cr,t+4},$$

where $\mathrm{DLRO}_{cr,t+4} = \mathrm{LRO}_{cr,t+4} - \mathrm{LRO}_{cr,t}$ represents the change in the log-odds ratio between two election cycles for county c in region r. $a_{DLRO,r}$ are region-specific fixed effects, $\mathbf{z}_{DLRO,cr}$ is a set of county-specific variables that do not vary over time, and $\mathbf{x}_{DLRO,cr,t+3}$ includes the

covariates that vary across counties and over time, measured before the election date (as of October 14, 2024, to be specific).

The voter turnout computed as

$$\mathrm{VT}_{cr,t} = \frac{R_{cr,t} + D_{cr,t}}{\mathrm{VAP}_{cr,t}},$$

is the proportion of total two-party votes relative to the voting-age population (VAP) in county c, region r, during the election year t.

In an election year, voter turnout is influenced by various demographic and socioeconomic factors. We model turnout as:

$$\mathrm{VT}_{cr,t+4} = a_{VT,r} + \boldsymbol{\psi}_r' \mathbf{z}_{VT,cr} + \lambda_r \mathrm{VT}_{cr,t} + \delta_r \mathrm{DLRO}_{cr,t+4} + \boldsymbol{\theta}_r' \mathbf{x}_{VT,cr,t+3} + v_{cr,t+4},$$

where turnout is a function of both time-invariant and time-varying variables, including the turnout in the preceding election cycle, and $\mathrm{DLRO}_{cr,t+4}$, which acknowledges the the simultaneous nature of voters' decisions, whether to vote and how to cast their vote.

Time-invariant factors include stable demographic and socioeconomic characteristics such as education, migration patterns, religiosity, and the urban-rural divide. Time-varying factors include variables like unemployment rates, poverty levels, household income, housing prices, and employment statistics.

In this paper, we impose a triangular restriction by setting $\delta_r = 0$, implying that an individual's voting decision is unaffected by their expectations of the overall election outcomes. Under this restriction, the parameters for both $\mathrm{VT}_{cr,t+4}$ and $\mathrm{DLRO}_{cr,t+4}$ can be estimated recursively using a two-stage process. The estimating equations are:

$$\hat{\mathrm{VT}}_{cr,t+4} = \hat{a}_{VT,r} + \hat{\boldsymbol{\psi}}_r' \mathbf{z}_{\mathrm{VT},cr} + \hat{\lambda}_r \mathrm{VT}_{cr,t} + \hat{\boldsymbol{\theta}}_r' \mathbf{x}_{VT,cr,t+3},$$

and

$$\widehat{\mathrm{DLRO}}_{cr,t+4} = \hat{a}_{DLRO,r} + \hat{\boldsymbol{\phi}}_r' \mathbf{z}_{DLRO,cr} + \hat{\beta}_r \hat{\mathrm{VT}}_{cr,t+4} + \hat{\boldsymbol{\gamma}}_r' \mathbf{x}_{DLRO,cr,t+3}.$$

We estimate two models: a pooled model and a heterogeneous model. The heterogeneous model accounts for regional differences by allowing all coefficients to vary across regions, estimating eight separate regional panel regressions. The pooled model is a restricted version where coefficients are assumed to be the same across all regions. Both models are estimated using least squares after variable selection, applying techniques such as Lasso and OCMT.

2.1 Forecasting equations

To predict the state level electoral votes and the national outcome, we need to calculate the Republican and Democratic vote shares at the state level. The Republican votes for county c in region r at time $t+4$ are given by:

$$\begin{aligned} R_{cr,t+4} &= \frac{\mathrm{VAP}_{cr,t+4}\mathrm{VT}_{cr,t+4}}{1+\exp(-\mathrm{LRO}_{cr,t+4})} \\ &= \mathrm{VAP}_{cr,t+4}\mathrm{VT}_{cr,t+4}\left(\frac{\exp(\mathrm{LRO}_{cr,t+4})}{1+\exp\left(\mathrm{LRO}_{cr,t+4}\right)}\right), \end{aligned}$$

where $\mathrm{VAP}_{cr,t+4}$ represents the voting age population, $\mathrm{VT}_{cr,t+4}$ is the voter turnout, and $\mathrm{LRO}_{cr,t+4}$ is the log-odds ratio of Republican to Democratic votes in the county. Similarly, the Democratic votes are:

$$D_{cr,t+4} = \mathrm{VAP}_{cr,t+4}\mathrm{VT}_{cr,t+4}\left(\frac{1}{1+\exp(\mathrm{LRO}_{cr,t+4})}\right).$$

These county-level votes are then aggregated to the state level. Let $\mathcal{C}_s$ denote the set of counties in state s. The total Republican and Democratic votes for the state are:

$$R_{s,t+4} = \sum_{cr\in\mathcal{C}_s} R_{cr,t+4} \quad \text{and} \quad D_{s,t+4} = \sum_{cr\in\mathcal{C}_s} D_{cr,t+4}.$$

The Republican vote share for state s is calculated as:

$$\begin{aligned} V_{s,t+4} &= \frac{\sum_{cr\in\mathcal{C}_s} R_{cr,t+4}}{\sum_{cr\in\mathcal{C}_s}(R_{cr,t+4}+D_{cr,t+4})} \\ &= \frac{\sum_{cr\in\mathcal{C}_s}\mathrm{VAP}_{cr,t+4}\mathrm{VT}_{cr,t+4}\left(\frac{\exp(\mathrm{LRO}_{cr,t+4})}{1+\exp(\mathrm{LRO}_{cr,t+4})}\right)}{\sum_{cr\in\mathcal{C}_s}\mathrm{VAP}_{cr,t+4}\mathrm{VT}_{cr,t+4}}. \end{aligned}$$

Following the U.S. Electoral College rules, the Republican candidate is predictor as the winner if:

$$\sum_{s=1}^{51} w(d_s)\mathbf{1}(V_{st}-0.5) > 0.5,$$

where $\mathbf{1}(a) = 1$ if $a > 0$, and 0 otherwise, $w(d_s) = d_s/d$, d_s is the number of delegates for state s, and d is the total number of delegates (538).

3 Model validation

Before forecasting the 2024 presidential election, we first validate our model by applying it to data from 2004 to 2016, including updated information from the 2020 election. This validation process tests whether the model can accurately predict electoral outcomes for Republican and Democratic states, the so called "non-swing states". Republican states are those that have consistently supported Republican candidates over the past 5 election cycles. In our analysis, these states include Alabama (AL), Alaska (AK), Arkansas (AR), Florida (FL), Idaho (ID), Indiana (IN), Kansas (KS), Kentucky (KY), Louisiana (LA), Mississippi (MS), Missouri (MO), Montana (MT), Nebraska (NE), North Dakota (ND), Oklahoma (OK), South Carolina (SC), South Dakota (SD), Tennessee (TN), Utah (UT), West Virginia (WV), Wyoming (WY), Iowa (IA), Ohio (OH), and Texas (TX).

Democratic states, which have consistently favored Democratic presidential candidates, include California (CA), New York (NY), Illinois (IL), Massachusetts (MA), Maryland (MD), Vermont (VT), Hawaii (HI), Rhode Island (RI), Connecticut (CT), New Jersey (NJ), Delaware (DE), Washington (WA), Oregon (OR), New Mexico (NM), Maine (ME), Minnesota (MN), Colorado (CO), Virginia (VA), the District of Columbia (DC), and New Hampshire (NH). The remaining states, referred to as the swing states (where support is nearly evenly split between the two major parties) are critical battlegrounds in determining the election outcomes. These states in our analysis include Arizona (AZ), Georgia (GA), Michigan (MI), Pennsylvania (PA), Wisconsin (WI), North Carolina (NC), and Nevada (NV). Although Texas is becoming more competitive due to demographic shifts, it remains classified as Republican-leaning based on recent elections, where the Republican margin, though narrowing, was still decisive enough to maintain this designation.

If our model can reliably predict electoral outcomes for the Republican and Democratic states, excluding swing states, it is more likely that model predictions can be relied on for predicting the election outcomes of the 7 swing states.

Maine (ME) and Nebraska (NE) use a unique electoral method called the congressional district system. Under this approach, two electoral votes are awarded to the statewide popular vote winner, while the remaining votes are distributed according to the popular vote in each congressional district (two districts in ME, three in NE). This allows for the possibility of split electoral votes, which has happened twice in each state: Nebraska in 2008, Maine in 2016, and both in 2020. However, for the purposes of our forecasting model, we apply the winner-take-all method used by most U.S. states, disregarding the possibility of vote splits in ME and NE.

Table 3 presents state-level forecasts for Republican vote shares (V_s) and electoral college

Table 3: US Presidential Election 2020 Forecasts by States Using County Level Data

		Lasso				OCMT			
		Pooled		Regional		Pooled		Regional	
State	Total EC	$\hat{V}_S$	EC Votes	$\hat{V}_S$	EC Votes	$\hat{V}_S$	EC Votes	$\hat{V}_S$	EC Votes
AK	3	N/A	3	N/A	3	N/A	3	N/A	3
AL	9	0.6284	9	0.6364	9	0.6412	9	0.6541	9
AR	6	0.6291	6	0.6502	6	0.6465	6	0.6652	6
AZ	11	0.4889	0	0.5495	11	0.5208	11	0.5695	11
CA	55	0.3024	0	0.3130	0	0.3371	0	0.3399	0
CO	9	0.4050	0	0.4127	0	0.4214	0	0.4167	0
CT	7	0.3689	0	0.4980	0	0.3984	0	0.4787	0
DC	3	0.0317	0	0.0355	0	0.0328	0	0.0403	0
DE	3	0.4005	0	0.4249	0	0.4152	0	0.4689	0
FL	29	0.4640	0	0.4620	0	0.4880	0	0.4823	0
GA	16	0.4914	0	0.5062	16	0.5109	16	0.5246	16
HI	4	N/A	0	N/A	0	N/A	0	N/A	0
IA	6	0.5240	6	0.5510	6	0.5351	6	0.5781	6
ID	4	0.6654	4	0.6484	4	0.6772	4	0.6550	4
IL	20	0.3799	0	0.4171	0	0.3981	0	0.4189	0
IN	11	0.5810	11	0.5761	11	0.6038	11	0.5818	11
KS	6	0.5766	6	0.5764	6	0.5848	6	0.5916	6
KY	8	0.6381	8	0.6538	8	0.6550	8	0.6691	8
LA	8	0.5771	8	0.5815	8	0.6003	8	0.6091	8
MA	11	0.2915	0	0.4084	0	0.3177	0	0.3911	0
MD	10	0.3133	0	0.3460	0	0.3329	0	0.3732	0
ME	4	0.4540	0	0.5107	4	0.4650	0	0.4788	0
MI	16	0.4740	0	0.5076	16	0.4972	0	0.4999	0
MN	10	0.4494	0	0.4661	0	0.4653	0	0.5015	10
MO	10	0.5928	10	0.6230	10	0.6082	10	0.6512	10
MS	6	0.5809	6	0.5821	6	0.6019	6	0.6098	6
MT	3	0.5730	3	0.5705	3	0.5935	3	0.5728	3
NC	15	0.4856	0	0.4884	0	0.5036	15	0.5082	15
ND	3	0.6606	3	0.6931	3	0.6881	3	0.7340	3
NE	5	0.6016	5	0.6413	5	0.6111	5	0.6934	5
NH	4	0.4432	0	0.5265	4	0.4695	0	0.4961	0
NJ	14	0.3711	0	0.4140	0	0.4066	0	0.4454	0
NM	5	0.4101	0	0.4650	0	0.4416	0	0.4790	0
NV	6	0.4678	0	0.4606	0	0.5001	6	0.4634	0
NY	29	0.3404	0	0.3450	0	0.3685	0	0.3421	0
OH	18	0.5198	18	0.5423	18	0.5398	18	0.5413	18
OK	7	0.6729	7	0.6796	7	0.6870	7	0.6995	7
OR	7	0.4034	0	0.4070	0	0.4229	0	0.4235	0
PA	20	0.4678	0	0.4973	0	0.4993	0	0.5547	20
RI	4	0.3777	0	0.4944	0	0.3888	0	0.4629	0
SC	9	0.5601	9	0.5586	9	0.5714	9	0.5812	9
SD	3	0.6408	3	0.6316	3	0.6523	3	0.6404	3
TN	11	0.6202	11	0.6375	11	0.6419	11	0.6548	11
TX	38	0.5032	38	0.5123	38	0.5335	38	0.5379	38
UT	6	0.6027	6	0.5983	6	0.6150	6	0.6062	6
VA	13	0.4237	0	0.4139	0	0.4396	0	0.4405	0
VT	3	0.3176	0	0.3770	0	0.3326	0	0.3610	0
WA	12	0.3701	0	0.3844	0	0.3991	0	0.4111	0
WI	10	0.4743	0	0.5070	10	0.4951	0	0.5051	10
WV	5	0.7121	5	0.7201	5	0.7363	5	0.7399	5
WY	3	0.7257	3	0.7284	3	0.7498	3	0.7365	3
All Votes	538		188		249		236		270

Notes: Total EC refers to the number of electoral votes per state. EC votes are the predicted Republican electoral votes, excluding Hawaii (Democratic) and Alaska (Republican). Regional forecasts are based on eight region-specific panel regressions, one for each BEA region. Red indicates historically Republican states, blue indicates historically Democratic states, and green represents swing states.

Table 3 (cont.) 2020 Lasso-OCMT Average

		Lasso-OCMT Average			
		Pooled		Regional	
State	Total EC	$\hat{V}_S$	EC Votes	$\hat{V}_S$	EC Votes
AK	3	N/A	3	N/A	3
AL	9	0.6348	9	0.6453	9
AR	6	0.6378	6	0.6577	6
AZ	11	0.5049	11	0.5595	11
CA	55	0.3198	0	0.3264	0
CO	9	0.4132	0	0.4147	0
CT	7	0.3837	0	0.4883	0
DC	3	0.0322	0	0.0379	0
DE	3	0.4078	0	0.4469	0
FL	29	0.4760	0	0.4721	0
GA	16	0.5011	16	0.5154	16
HI	4	N/A	0	N/A	0
IA	6	0.5295	6	0.5646	6
ID	4	0.6713	4	0.6517	4
IL	20	0.3890	0	0.4180	0
IN	11	0.5924	11	0.5789	11
KS	6	0.5807	6	0.5840	6
KY	8	0.6465	8	0.6615	8
LA	8	0.5887	8	0.5953	8
MA	11	0.3046	0	0.3997	0
MD	10	0.3231	0	0.3596	0
ME	4	0.4595	0	0.4947	0
MI	16	0.4856	0	0.5037	16
MN	10	0.4573	0	0.4838	0
MO	10	0.6005	10	0.6371	10
MS	6	0.5914	6	0.5959	6
MT	3	0.5832	3	0.5716	3
NC	15	0.4946	0	0.4983	0
ND	3	0.6744	3	0.7136	3
NE	5	0.6064	5	0.6674	5
NH	4	0.4564	0	0.5113	4
NJ	14	0.3888	0	0.4297	0
NM	5	0.4259	0	0.4720	0
NV	6	0.4840	0	0.4620	0
NY	29	0.3545	0	0.3435	0
OH	18	0.5298	18	0.5418	18
OK	7	0.6799	7	0.6896	7
OR	7	0.4131	0	0.4152	0
PA	20	0.4836	0	0.5260	20
RI	4	0.3832	0	0.4786	0
SC	9	0.5658	9	0.5699	9
SD	3	0.6465	3	0.6360	3
TN	11	0.6311	11	0.6461	11
TX	38	0.5184	38	0.5251	38
UT	6	0.6089	6	0.6022	6
VA	13	0.4316	0	0.4272	0
VT	3	0.3251	0	0.3690	0
WA	12	0.3846	0	0.3977	0
WI	10	0.4847	0	0.5060	10
WV	5	0.7242	5	0.7300	5
WY	3	0.7378	3	0.7325	3
All Votes	538		215		265

Notes: Total EC refers to the number of electoral votes per state. EC votes are the predicted Republican electoral votes, excluding Hawaii (Democratic) and Alaska (Republican). Regional forecasts are based on eight panel regressions, one for each BEA region. Red indicates Republican states, blue indicates Democrat states, and green represents swing states.

(EC) votes for the 2020 election, using data from Ahmed and Pesaran (2020). The column labeled "Total EC" represents the total number of electoral votes per state. The forecasts cover the U.S. mainland, with Hawaii (HI) assumed to vote Democratic and Alaska (AK) expected to vote Republican. Regional predictions are generated through eight separate panel regressions, corresponding to the eight Bureau of Economic Analysis (BEA) regions: Southeast, Southwest, Far West, Rocky Mountain, New England, Mideast, Great Lakes, and Plains. The forecasts are based on data from 2000 to 2016 as the training set, with 2020 data used for predictions. In all tables, historically Republican states are shown in red, historically Democratic states in blue, and swing states in orange.

As can be seen from the forecast results summarized in Table 3, for 2020 all models accurately predict the outcomes for all Republican states, except for Florida. However, the regional forecast using the Lasso model mistakenly predicts Republican victories in two Democratic states: Maine (ME) and New Hampshire (NH). Similarly, the OCMT model incorrectly forecasts a Republican win in Minnesota (MN), a state that has historically voted Democratic. Consequently, the combined Lasso-OCMT model also erroneously predicts a Republican victory in New Hampshire. Aside from these three errors, all discrepancies in the total vote counts ("All Votes") are due to variations in the projected outcomes for the Swing states, rather than misclassifications of Republican or Democratic states.

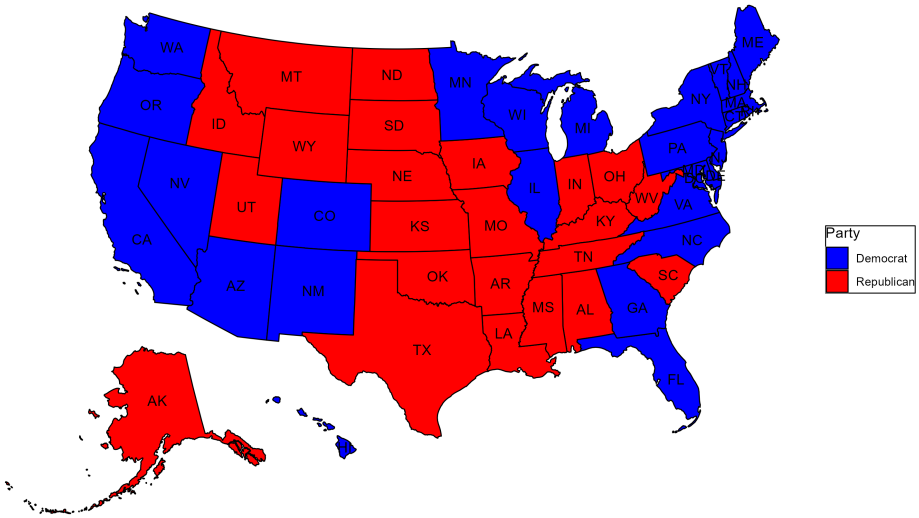

Figure 1: Pooled Forecasts of Lasso in 2020

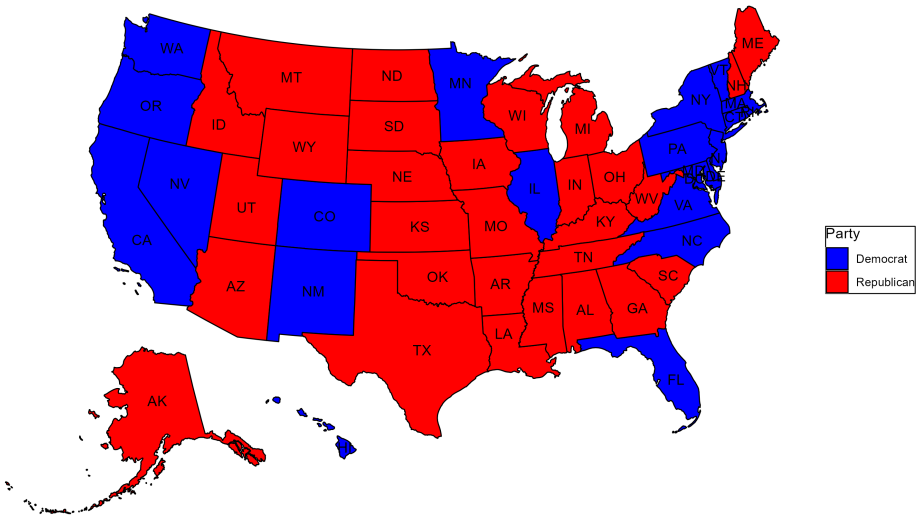

Figure 2: Regional Forecasts of Lasso in 2020

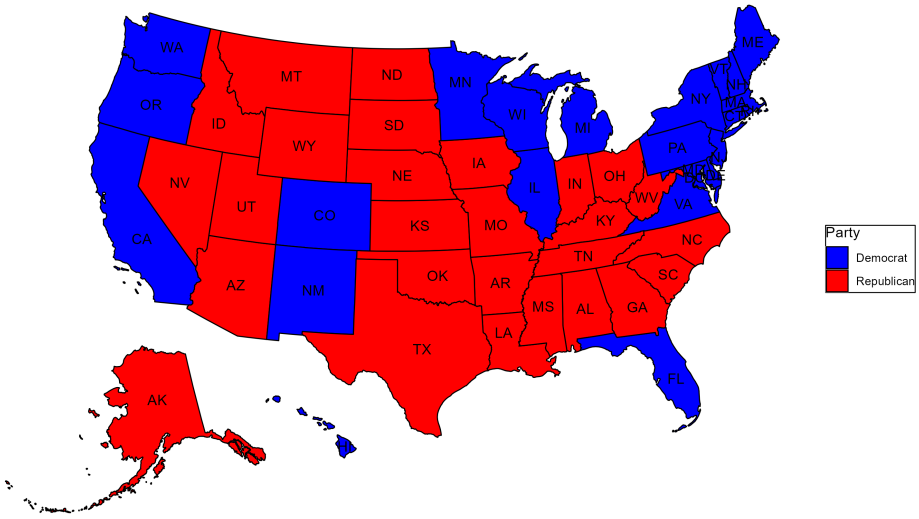

Figure 3: Pooled Forecasts of OCMT in 2020

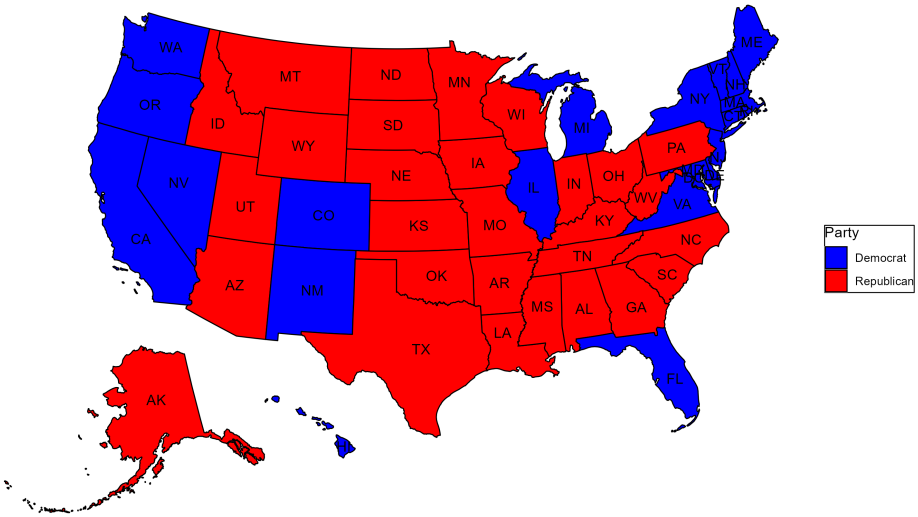

Figure 4: Regional Forecasts of OCMT in 2020

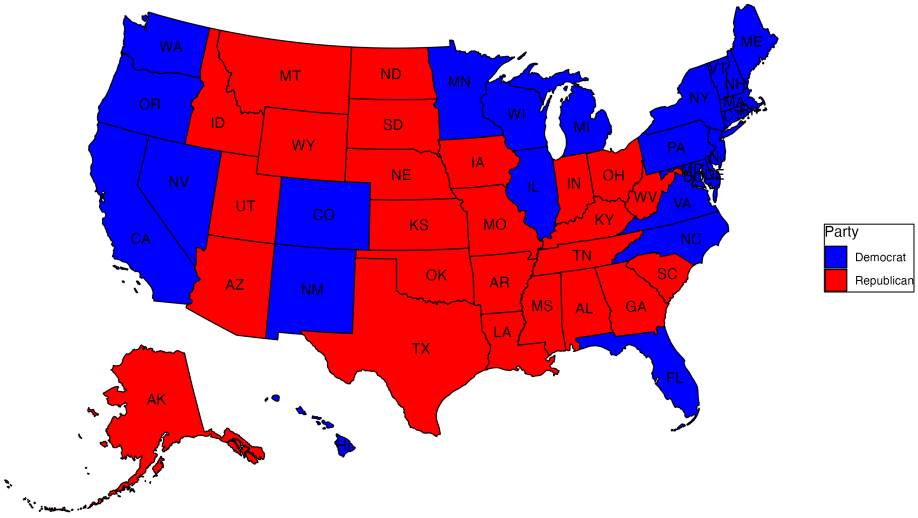

Figure 5: Pooled Forecasts of Laaso+OCMT Average in 2020

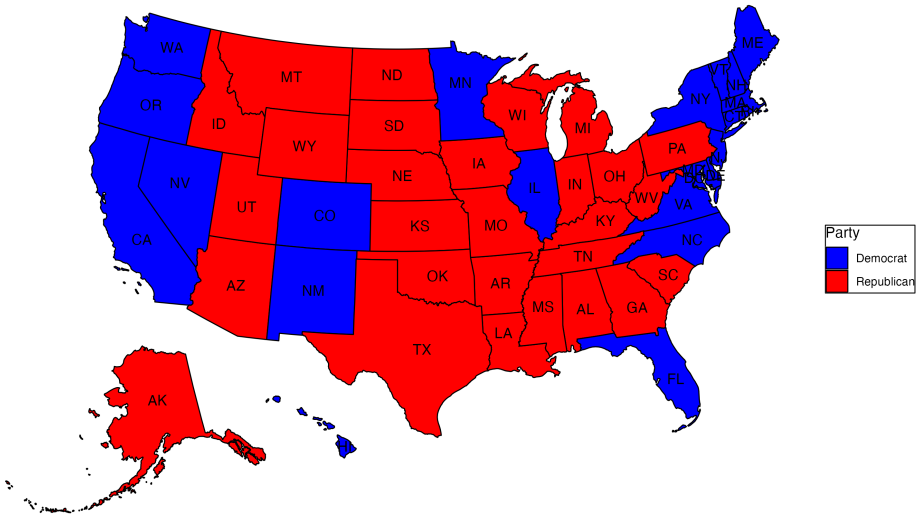

Figure 6: Regional Forecasts of Lasso+OCMT Average in 2020

4 2024 updates of variable selection and regression results

Tables 4 and 5 display the results for the pooled panel regressions with variable selection, where voter turnout (VT) and changes in the log Republican odds (DLRO) serve as the dependent variables, respectively. The regressions are conducted for two sample periods: 2000–2016 and 2000–2020. All models are estimated with an intercept, but to save space intercept estimates are not reported.

Tables 6, 7, 8, and 9 present the state-level Republican vote share and voter turnout for both 2020 and 2024. Tables 6 and 7 give the results for pooled and regional forecasts for 2020 and 2024, estimated using the Lasso method, alongside the realized values for 2020 . Tables 8, and 9 give the forecasts generated using OCMT. As can be seen, the voter turnout in 2024 is forecast to be lower than in 2020, in all states with one or two exceptions. This prediction is reasonably robust to whether we use pooled or regional forecasting models or whether the variables are selected by Lasso or OCMT. This is an issue that requires further investigation.

Table 4: Pooled panel regression with variable selection for voter turnout (VT) as the dependent variable estimated over the 2000-2016 and 2000-2020 election cycles

Covariate	OCMT		SE-OCMT		Lasso	
Sample Period	2000-16	2000-20	2000-16	2000-20	2000-16	2000-20
Incumbent party	-0.2179	-0.3351	(0.1085)	(0.0804)		
Incumbent president	0.0314	0.0349	(0.0053)	(0.0050)	0.0289	0.0323
VT(L1)	0.7977	0.8089	(0.0172)	(0.0165)	0.7789	0.7886
VT(L1) x incumbent party	-0.0461	-0.0338	(0.0166)	(0.0160)		
Healthcare costs (L1)					-0.1294	-0.1958
Government employment (L1)	0.1557	0.1399	(0.1870)	(0.1698)	0.1449	0.0859
Unemployment (L1)	0.4034	0.4465	(0.1044)	(0.0967)	0.2964	0.2676
House price (L1)						
Rent price (L1)					-0.0576	0.0771
Religiosity					-0.0050	-0.0074
Religiosity x incumbent party	-0.0116	-0.0152	(0.0066)	(0.0063)		
Migration	-0.3548	-0.3632	(0.1023)	(0.0823)	-0.2490	-0.3594
Migration x incumbent party	-0.2233	-0.2383	(0.1025)	(0.0812)		-0.0048
Education	0.0978	0.0798	(0.0163)	(0.0140)	0.0955	0.0694
Education x incumbent party	0.0907	0.0724	(0.0158)	(0.0137)	0.0835	0.0939
ln(Median income)	0.0026	0.0062	(0.0109)	(0.0079)		0.0182
ln(Median income) x incumbent party	0.0231	0.0338	(0.0096)	(0.0072)		
Poverty	-0.1517	-0.1616	(0.0495)	(0.0357)	-0.1596	-0.1220
Poverty x incumbent party	0.0277	0.0345	(0.0472)	(0.0334)		
Rural	-0.0004		(0.0004)		-0.0004	
Rural x incumbent party	-0.0008		(0.0004)		-0.0014	-0.0006
Mail-in voting	0.0065	0.0091	(0.0032)	(0.0029)	0.0063	0.0085
Observations	12438	15535			12438	15535
Covariates Selected	18	16			14	15
Adj.R2	0.8058	0.8265			0.8034	0.8233
Reg SE	0.0397	0.0393			0.0399	0.0397

Notes: This table presents estimates from the recursive voter turnout (VT) and voting outcome model. First, the VT equation is estimated, and the resulting estimated voter turnout ($\hat{VT}$) is then incorporated into the active set for estimating the log Republican odds (DLRO) equation. The estimates provided here are for the VT equation. Standard errors are clustered at the state-year level to account for potential within-cluster correlation. For the Lasso regression, model selection is conducted using 10-fold cross-validation with a fixed seed of 123 to ensure reproducibility. The selected model is the one whose cross-validated mean squared error (CV-MSE) is within one standard deviation (1-SD) of the minimum MSE, following the one-standard-error rule.

Table 5: Pooled panel regression with variable selection for changes in log Republican odds ($DLRO$) as the dependent variable estimated over the 2000-2016 and 2000-2020 election cycles

Covariate	OCMT		SE-OCMT		Lasso	
Sample Period	2000-16	2000-20	2000-16	2000-20	2000-16	2000-20
Incumbent party	-0.8364	-0.7914	(0.3566)	(0.3294)	-0.1478	-0.2052
House DLRO	0.0250		(0.0281)		0.0186	0.0035
VT	-0.3735	-0.3036	(0.1069)	(0.1000)	-0.2839	-0.2035
VT x incumbent party	-0.1786	-0.1343	(0.0938)	(0.0875)	-0.1094	-0.0457
Left-behind (L1)					0.0051	0.0185
Left-behind (L1) x incumbent party					-0.1118	-0.0969
Healthcare costs (L1)						0.0653
Government employment (L1)					2.4752	2.2079
USD REER (L1)	-0.0198	0.0915	(0.8802)	(0.4914)		0.3495
USD REER (L1) x incumbent party	0.0737		(0.8872)		-0.5045	0.2691
USD REER (M3)	-0.0389	-0.0300	(0.2468)	(0.1591)		
USD REER (M3) x incumbent party	-0.7329	-0.7392	(0.2311)	(0.1245)		-0.4217
Unemployment (L1)					-0.9088	-2.4790
Unemployment (L1) x incumbent party	-2.6836	-2.3867	(1.5928)	(0.8097)		-0.0002
Unemployment (M3)						1.7287
Unemployment (M3) x incumbent party	4.8527	4.3289	(1.6454)	(0.7058)	0.9323	1.4309
House price (L1)	-0.3884	-0.3005	(0.4050)	(0.3958)		-0.0320
House price (L1) x incumbent party						0.0854
House price (M3)	0.7047	0.6223	(0.3133)	(0.3081)	0.3722	0.4086
House price (M3) x incumbent party						
Rent price (L1)					-0.8429	-0.3035
Inflation (L1)					1.0148	1.1577
Migration	-1.7827	-0.2636	(0.4813)	(0.4519)	-1.4525	-0.1714
MigrationL	0.7900		(1.1803)		0.2324	0.6078
Education	-0.6296	-0.6022	(0.0962)	(0.0734)	-0.6864	-0.6699
EducationL	-0.7883	-0.6515	(0.2045)	(0.1265)	-0.7032	-0.6370
ln(Population density)	-0.0010	-0.0020	(0.0056)	(0.0044)	0.0004	0.0003
ln(Median income)		-0.0325		(0.0417)		0.0048
ln(Median income) x incumbent party	0.0649	0.0593	(0.0336)	(0.0318)		
Poverty	-0.6909	-0.6755	(0.1486)	(0.1822)	-0.5167	-0.4095
Rural	0.0089	0.0081	(0.0021)	(0.0019)	0.0050	0.0061
Observations	12438	15535			12438	15535
Covariates Selected	20	18			20	28
Adj.R2	0.5071	0.5056			0.5185	0.5172
Reg SE	0.1788	0.1689			0.1768	0.1669

Notes: This table presents estimates from the recursive voter turnout (VT) and voting outcome model. First, the VT equation is estimated, and the resulting estimated voter turnout ($\hat{VT}$) is then incorporated into the active set for estimating the log Republican odds (DLRO) equation. The estimates provided here are for the log Republican odds equation. Standard errors are clustered at the state-year level to account for potential within-cluster correlation. For the Lasso regression, model selection is conducted using 10-fold cross-validation with a fixed seed of 123 to ensure reproducibility. The selected model is the one whose cross-validated mean squared error (CV-MSE) is within one standard deviation (1-SD) of the minimum MSE, following the one-standard-error rule.

Table 6: State Republican Vote Share and Voter Turnout: Realized and Pooled Forecast Values for the 2020 and 2024 U.S. Presidential Elections using Lasso

	Realized		Pooled Forecasts			
	Vote Share	Voter Turnout	Vote Share		Voter Turnout	
State	2020	2020	2020	2024	2020	2024
AK	0.5526	0.6388	N/A	N/A	N/A	N/A
AL	0.6291	0.6151	0.6284	0.6761	0.6197	0.5322
AR	0.6421	0.5305	0.6291	0.6898	0.5571	0.4790
AZ	0.4984	0.6578	0.4889	0.5142	0.6147	0.5156
CA	0.3509	0.6614	0.3024	0.3704	0.6364	0.5294
CO	0.4306	0.7583	0.4050	0.4316	0.7472	0.6394
CT	0.3980	0.6837	0.3689	0.4124	0.6939	0.5714
DC	0.0553	0.6335	0.0317	0.0507	0.6883	0.5610
DE	0.4037	0.6847	0.4005	0.4336	0.6805	0.5809
FL	0.5169	0.7132	0.4640	0.5474	0.6956	0.5830
GA	0.4987	0.6551	0.4914	0.5206	0.6340	0.5346
HI	0.3497	0.5486	N/A	N/A	N/A	N/A
IA	0.5430	0.7078	0.5240	0.6055	0.6914	0.6078
ID	0.6588	0.6646	0.6654	0.6957	0.6005	0.5165
IL	0.4134	0.6507	0.3799	0.4451	0.6675	0.5656
IN	0.5820	0.5995	0.5810	0.6279	0.6084	0.5148
KS	0.5749	0.6352	0.5766	0.6256	0.6182	0.5191
KY	0.6320	0.6209	0.6381	0.6750	0.6074	0.5172
LA	0.5946	0.6061	0.5771	0.6372	0.6276	0.5281
MA	0.3288	0.7027	0.2915	0.3336	0.7135	0.5820
MD	0.3297	0.6836	0.3133	0.3444	0.7055	0.5966
ME	0.4554	0.7357	0.4540	0.4860	0.7164	0.6269
MI	0.4859	0.7223	0.4740	0.5179	0.6850	0.5888
MN	0.4636	0.7749	0.4494	0.4982	0.7390	0.6430
MO	0.5964	0.6104	0.5928	0.6360	0.6309	0.5425
MS	0.5838	0.5759	0.5809	0.6314	0.5938	0.5075
MT	0.5840	0.7092	0.5730	0.6271	0.6410	0.5565
NC	0.5068	0.7091	0.4856	0.5274	0.6901	0.5848
ND	0.6722	0.6144	0.6606	0.7496	0.6342	0.5364
NE	0.5978	0.6727	0.6016	0.6552	0.6503	0.5604
NH	0.4625	0.7396	0.4432	0.4812	0.7418	0.6328
NJ	0.4193	0.7259	0.3711	0.4193	0.7055	0.5900
NM	0.4448	0.5925	0.4101	0.4716	0.5695	0.4679
NV	0.4877	0.6597	0.4678	0.5168	0.6322	0.5385
NY	0.3828	0.6135	0.3404	0.3939	0.6268	0.5189
OH	0.5408	0.6570	0.5198	0.5810	0.6718	0.5787
OK	0.6694	0.5311	0.6729	0.7189	0.5742	0.4852
OR	0.4169	0.7357	0.4034	0.4462	0.6944	0.5982
PA	0.4941	0.6970	0.4678	0.5150	0.6830	0.5862
RI	0.3949	0.6334	0.3777	0.3990	0.6490	0.5408
SC	0.5593	0.6405	0.5601	0.6018	0.6242	0.5239
SD	0.6343	0.6314	0.6408	0.6950	0.6143	0.5370
TN	0.6183	0.5870	0.6202	0.6551	0.5703	0.4800
TX	0.5283	0.5958	0.5032	0.5635	0.5790	0.4784
UT	0.6069	0.6809	0.6027	0.6469	0.5526	0.4610
VA	0.4485	0.7011	0.4237	0.4748	0.6998	0.5931
VT	0.3170	0.7114	0.3176	0.3509	0.6521	0.5451
WA	0.4007	0.7364	0.3701	0.4225	0.6826	0.5767
WI	0.4968	0.7352	0.4743	0.5381	0.7112	0.6218
WV	0.6980	0.5445	0.7121	0.7610	0.5410	0.4642
WY	0.7248	0.6076	0.7257	0.7653	0.6111	0.5270

Notes: Vote share refers to the proportion of total votes cast for a specific candidate or party, expressed as a percentage of the total valid votes in a given election. In this context, it represents the percentage of the two-party vote received by the Republican candidate. Voter turnout refers to the percentage of eligible voters who participated in the election, calculated as the number of votes cast divided by the total eligible voting-age population.

Table 7: State Republican Vote Share and Voter Turnout: Realized and Regional Forecast Values for the 2020 and 2024 U.S. Presidential Elections using Lasso

	Realized		Regional Forecasts			
	Vote Share	Voter Turnout	Vote Share		Voter Turnout	
State	2020	2020	2020	2024	2020	2024
AK	0.5526	0.6388	N/A	N/A	N/A	N/A
AL	0.6291	0.6151	0.6364	0.6538	0.6217	0.5570
AR	0.6421	0.5305	0.6502	0.6771	0.5581	0.4979
AZ	0.4984	0.6578	0.5495	0.5572	0.5396	0.5025
CA	0.3509	0.6614	0.3130	0.2382	0.6015	0.5455
CO	0.4306	0.7583	0.4127	0.4735	0.7176	0.6640
CT	0.3980	0.6837	0.4980	0.3647	0.6547	0.6227
DC	0.0553	0.6335	0.0355	0.0599	0.6229	0.6280
DE	0.4037	0.6847	0.4249	0.4477	0.6057	0.6363
FL	0.5169	0.7132	0.4620	0.5199	0.6946	0.6238
GA	0.4987	0.6551	0.5062	0.5046	0.6311	0.5548
HI	0.3497	0.5486	N/A	N/A	N/A	N/A
IA	0.5430	0.7078	0.5510	0.6332	0.6249	0.5744
ID	0.6588	0.6646	0.6484	0.7166	0.6303	0.5929
IL	0.4134	0.6507	0.4171	0.5130	0.5916	0.5285
IN	0.5820	0.5995	0.5761	0.6555	0.6632	0.6018
KS	0.5749	0.6352	0.5764	0.6288	0.6058	0.5416
KY	0.6320	0.6209	0.6538	0.6501	0.6098	0.5499
LA	0.5946	0.6061	0.5815	0.6173	0.6242	0.5593
MA	0.3288	0.7027	0.4084	0.3065	0.6705	0.6411
MD	0.3297	0.6836	0.3460	0.3633	0.6188	0.6383
ME	0.4554	0.7357	0.5107	0.4921	0.6616	0.6318
MI	0.4859	0.7223	0.5076	0.5705	0.6866	0.5961
MN	0.4636	0.7749	0.4661	0.5297	0.7169	0.6531
MO	0.5964	0.6104	0.6230	0.6807	0.6099	0.5531
MS	0.5838	0.5759	0.5821	0.6177	0.5899	0.5292
MT	0.5840	0.7092	0.5705	0.6361	0.6380	0.5989
NC	0.5068	0.7091	0.4884	0.5098	0.6863	0.6150
ND	0.6722	0.6144	0.6931	0.7497	0.6267	0.5527
NE	0.5978	0.6727	0.6413	0.7092	0.6235	0.5477
NH	0.4625	0.7396	0.5265	0.4638	0.6690	0.6328
NJ	0.4193	0.7259	0.4140	0.4669	0.6055	0.6220
NM	0.4448	0.5925	0.4650	0.4746	0.6411	0.5921
NV	0.4877	0.6597	0.4606	0.4274	0.6108	0.5700
NY	0.3828	0.6135	0.3450	0.3977	0.6253	0.6215
OH	0.5408	0.6570	0.5423	0.6163	0.6681	0.5811
OK	0.6694	0.5311	0.6796	0.6785	0.5552	0.4979
OR	0.4169	0.7357	0.4070	0.3445	0.6621	0.6075
PA	0.4941	0.6970	0.4973	0.5384	0.5937	0.6285
RI	0.3949	0.6334	0.4944	0.3741	0.6167	0.5813
SC	0.5593	0.6405	0.5586	0.5751	0.6204	0.5511
SD	0.6343	0.6314	0.6316	0.6709	0.6065	0.5508
TN	0.6183	0.5870	0.6375	0.6422	0.5677	0.4994
TX	0.5283	0.5958	0.5123	0.5288	0.5678	0.5054
UT	0.6069	0.6809	0.5983	0.6613	0.5396	0.4990
VA	0.4485	0.7011	0.4139	0.4327	0.6965	0.6119
VT	0.3170	0.7114	0.3770	0.3257	0.6081	0.5752
WA	0.4007	0.7364	0.3844	0.3234	0.6540	0.5860
WI	0.4968	0.7352	0.5070	0.6061	0.5743	0.5015
WV	0.6980	0.5445	0.7201	0.7304	0.5363	0.4794
WY	0.7248	0.6076	0.7284	0.7732	0.6033	0.5684

Notes: Vote share refers to the proportion of total votes cast for a specific candidate or party, expressed as a percentage of the total valid votes in a given election. In this context, it represents the percentage of the two-party vote received by the Republican candidate. Voter turnout refers to the percentage of eligible voters who participated in the election, calculated as the number of votes cast divided by the total eligible voting-age population.

Table 8: State Republican Vote Share and Voter Turnout: Realized and Pooled Forecast Values for the 2020 and 2024 U.S. Presidential Elections using OCMT

	Realized		Pooled Forecasts			
	Vote Share	Voter Turnout	Vote Share		Voter Turnout	
State	2020	2020	2020	2024	2020	2024
AK	0.5526	0.6388	N/A	N/A	N/A	N/A
AL	0.6291	0.6151	0.6412	0.6511	0.6275	0.5348
AR	0.6421	0.5305	0.6465	0.6727	0.5655	0.4756
AZ	0.4984	0.6578	0.5208	0.4805	0.6282	0.5116
CA	0.3509	0.6614	0.3371	0.3114	0.6566	0.5232
CO	0.4306	0.7583	0.4214	0.4129	0.7595	0.6343
CT	0.3980	0.6837	0.3984	0.3453	0.7074	0.5705
DC	0.0553	0.6335	0.0328	0.0403	0.6967	0.5582
DE	0.4037	0.6847	0.4152	0.3821	0.6922	0.5758
FL	0.5169	0.7132	0.4880	0.4990	0.6969	0.6016
GA	0.4987	0.6551	0.5109	0.5132	0.6414	0.5356
HI	0.3497	0.5486	N/A	N/A	N/A	N/A
IA	0.5430	0.7078	0.5351	0.5859	0.6965	0.6083
ID	0.6588	0.6646	0.6772	0.6963	0.6095	0.5129
IL	0.4134	0.6507	0.3981	0.4135	0.6774	0.5632
IN	0.5820	0.5995	0.6038	0.6094	0.6191	0.5137
KS	0.5749	0.6352	0.5848	0.5993	0.6276	0.5201
KY	0.6320	0.6209	0.6550	0.6729	0.6145	0.5275
LA	0.5946	0.6061	0.6003	0.6029	0.6368	0.5386
MA	0.3288	0.7027	0.3177	0.2799	0.7207	0.5848
MD	0.3297	0.6836	0.3329	0.3116	0.7170	0.5879
ME	0.4554	0.7357	0.4650	0.4774	0.7221	0.6217
MI	0.4859	0.7223	0.4972	0.4843	0.6943	0.5894
MN	0.4636	0.7749	0.4653	0.4778	0.7458	0.6435
MO	0.5964	0.6104	0.6082	0.6184	0.6409	0.5429
MS	0.5838	0.5759	0.6019	0.5988	0.6032	0.5091
MT	0.5840	0.7092	0.5935	0.6174	0.6531	0.5501
NC	0.5068	0.7091	0.5036	0.5122	0.6974	0.6011
ND	0.6722	0.6144	0.6881	0.7115	0.6425	0.5453
NE	0.5978	0.6727	0.6111	0.6426	0.6554	0.5626
NH	0.4625	0.7396	0.4695	0.4359	0.7552	0.6262
NJ	0.4193	0.7259	0.4066	0.3603	0.7233	0.5853
NM	0.4448	0.5925	0.4416	0.4310	0.5813	0.4716
NV	0.4877	0.6597	0.5001	0.4898	0.6430	0.5342
NY	0.3828	0.6135	0.3685	0.3439	0.6428	0.5114
OH	0.5408	0.6570	0.5398	0.5519	0.6801	0.5793
OK	0.6694	0.5311	0.6870	0.6970	0.5822	0.4865
OR	0.4169	0.7357	0.4229	0.4111	0.7094	0.5991
PA	0.4941	0.6970	0.4993	0.4748	0.6921	0.5831
RI	0.3949	0.6334	0.3888	0.3442	0.6585	0.5327
SC	0.5593	0.6405	0.5714	0.5854	0.6302	0.5292
SD	0.6343	0.6314	0.6523	0.6838	0.6274	0.5343
TN	0.6183	0.5870	0.6419	0.6433	0.5783	0.4762
TX	0.5283	0.5958	0.5335	0.5343	0.5872	0.4794
UT	0.6069	0.6809	0.6150	0.6382	0.5648	0.4498
VA	0.4485	0.7011	0.4396	0.4369	0.7090	0.5949
VT	0.3170	0.7114	0.3326	0.3215	0.6638	0.5396
WA	0.4007	0.7364	0.3991	0.3882	0.6986	0.5787
WI	0.4968	0.7352	0.4951	0.5097	0.7173	0.6206
WV	0.6980	0.5445	0.7363	0.7271	0.5564	0.4599
WY	0.7248	0.6076	0.7498	0.7587	0.6258	0.5170

Notes: Vote share refers to the proportion of total votes cast for a specific candidate or party, expressed as a percentage of the total valid votes in a given election. In this context, it represents the percentage of the two-party vote received by the Republican candidate. Voter turnout refers to the percentage of eligible voters who participated in the election, calculated as the number of votes cast divided by the total eligible voting-age population.

Table 9: State Republican Vote Share and Voter Turnout: Realized and Pooled Forecast Values for the 2020 and 2024 U.S. Presidential Elections using OCMT

	Realized		Regional Forecasts			
	Vote Share	Voter Turnout	Vote Share		Voter Turnout	
State	2020	2020	2020	2024	2020	2024
AK	0.5526	0.6388	N/A	N/A	N/A	N/A
AL	0.6291	0.6151	0.6541	0.6339	0.6524	0.5514
AR	0.6421	0.5305	0.6652	0.6653	0.5827	0.4867
AZ	0.4984	0.6578	0.5695	0.5351	0.5611	0.4980
CA	0.3509	0.6614	0.3399	0.2832	0.5635	0.5245
CO	0.4306	0.7583	0.4167	0.4180	0.7023	0.6375
CT	0.3980	0.6837	0.4787	0.4152	0.6388	0.6039
DC	0.0553	0.6335	0.0403	0.0531	0.5741	0.6236
DE	0.4037	0.6847	0.4689	0.4120	0.5491	0.6346
FL	0.5169	0.7132	0.4823	0.4781	0.7297	0.6335
GA	0.4987	0.6551	0.5246	0.4936	0.6658	0.5502
HI	0.3497	0.5486	N/A	N/A	N/A	N/A
IA	0.5430	0.7078	0.5781	0.5625	0.6291	0.5522
ID	0.6588	0.6646	0.6550	0.6749	0.6074	0.5638
IL	0.4134	0.6507	0.4189	0.4746	0.5689	0.5136
IN	0.5820	0.5995	0.5818	0.6253	0.6485	0.5897
KS	0.5749	0.6352	0.5916	0.5698	0.6248	0.5133
KY	0.6320	0.6209	0.6691	0.6439	0.6396	0.5415
LA	0.5946	0.6061	0.6091	0.5919	0.6570	0.5550
MA	0.3288	0.7027	0.3911	0.3333	0.6588	0.6224
MD	0.3297	0.6836	0.3732	0.3434	0.5717	0.6330
ME	0.4554	0.7357	0.4788	0.4960	0.6639	0.6301
MI	0.4859	0.7223	0.4999	0.5017	0.6390	0.5863
MN	0.4636	0.7749	0.5015	0.4603	0.7199	0.6381
MO	0.5964	0.6104	0.6512	0.6132	0.6062	0.5433
MS	0.5838	0.5759	0.6098	0.5939	0.6193	0.5220
MT	0.5840	0.7092	0.5728	0.6080	0.6244	0.5675
NC	0.5068	0.7091	0.5082	0.4834	0.7212	0.6125
ND	0.6722	0.6144	0.7340	0.6636	0.6216	0.5262
NE	0.5978	0.6727	0.6934	0.5688	0.6157	0.5106
NH	0.4625	0.7396	0.4961	0.4956	0.6715	0.6261
NJ	0.4193	0.7259	0.4454	0.4430	0.5589	0.6232
NM	0.4448	0.5925	0.4790	0.4394	0.6835	0.5779
NV	0.4877	0.6597	0.4634	0.4879	0.5746	0.5773
NY	0.3828	0.6135	0.3421	0.4207	0.5958	0.6182
OH	0.5408	0.6570	0.5413	0.5601	0.6292	0.5737
OK	0.6694	0.5311	0.6995	0.6582	0.5777	0.4827
OR	0.4169	0.7357	0.4235	0.3807	0.6343	0.5874
PA	0.4941	0.6970	0.5547	0.4952	0.5315	0.6341
RI	0.3949	0.6334	0.4629	0.4250	0.5943	0.5499
SC	0.5593	0.6405	0.5812	0.5577	0.6529	0.5433
SD	0.6343	0.6314	0.6404	0.6456	0.6260	0.5333
TN	0.6183	0.5870	0.6548	0.6308	0.5983	0.4884
TX	0.5283	0.5958	0.5379	0.4903	0.5995	0.4858
UT	0.6069	0.6809	0.6062	0.6128	0.5374	0.4387
VA	0.4485	0.7011	0.4405	0.4106	0.7406	0.6092
VT	0.3170	0.7114	0.3610	0.3652	0.5903	0.5519
WA	0.4007	0.7364	0.4111	0.3468	0.6186	0.5832
WI	0.4968	0.7352	0.5051	0.5512	0.5446	0.4923
WV	0.6980	0.5445	0.7399	0.7148	0.5619	0.4630
WY	0.7248	0.6076	0.7365	0.7514	0.5856	0.5195

Notes: Vote share refers to the proportion of total votes cast for a specific candidate or party, expressed as a percentage of the total valid votes in a given election. In this context, it represents the percentage of the two-party vote received by the Republican candidate. Voter turnout refers to the percentage of eligible voters who participated in the election, calculated as the number of votes cast divided by the total eligible voting-age population.

5 2024 Forecasts

Table 10 presents state-level forecasts of Republican vote shares and electoral votes for the 2024 U.S. presidential election. Similar to the analyses in Tables 2 and 3, the training sample covers the years 2000 to 2020. It is important to highlight that the 2024 Electoral College allocations were updated based on the 2020 census. The key changes in electoral votes occurred in the following states: California (-1), Colorado (+1), Florida (+1), Illinois (-1), Michigan (-1), Montana (+1), North Carolina (+1), New York (-1), Ohio (-1), Oregon (+1), Pennsylvania (-1), Texas (+2), and West Virginia (+1). The numbers in parentheses reflect the changes in electoral votes compared to the 2020 election. The election outcomes for Hawaii (Democratic) and Alaska (Republican) were pre-set. OCMT and Lasso-OCMT forecasts for Florida turned out to be borderline, and in view of the historical trends we decided to over-rule our forecasts and classify as a Republican win.

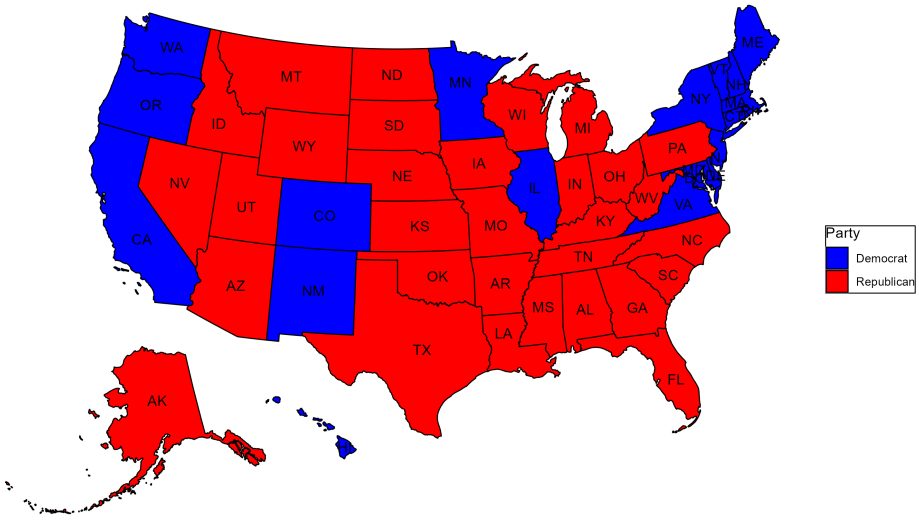

Figure 7: Pooled Forecasts of Lasso in 2024

Table 10: US Presidential Election 2024 Forecasts by States Using County Level Data

		Lasso				OCMT			
		Pooled		Regional		Pooled		Regional	
State	**Total EC**	$\hat{V}_S$	**EC Votes**	$\hat{V}_S$	**EC Votes**	$\hat{V}_S$	**EC Votes**	$\hat{V}_S$	**EC Votes**
AK	3	N/A	3	N/A	3	N/A	3	N/A	3
AL	9	0.6761	9	0.6538	9	0.6511	9	0.6339	9
AR	6	0.6898	6	0.6771	6	0.6727	6	0.6653	6
AZ	11	0.5142	11	0.5572	11	0.4805	0	0.5351	11
CA	54	0.3704	0	0.2382	0	0.3114	0	0.2832	0
CO	10	0.4316	0	0.4735	0	0.4129	0	0.4180	0
CT	7	0.4124	0	0.3647	0	0.3453	0	0.4152	0
DC	3	0.0507	0	0.0599	0	0.0403	0	0.0531	0
DE	3	0.4336	0	0.4477	0	0.3821	0	0.4120	0
FL	30	0.5474	30	0.5199	30	0.4990	30	0.4781	30
GA	16	0.5206	16	0.5046	16	0.5132	16	0.4936	0
HI	4	N/A	0	N/A	0	N/A	0	N/A	0
IA	6	0.6055	6	0.6332	6	0.5859	6	0.5625	6
ID	4	0.6957	4	0.7166	4	0.6963	4	0.6749	4
IL	19	0.4451	0	0.5130	19	0.4135	0	0.4746	0
IN	11	0.6279	11	0.6555	11	0.6094	11	0.6253	11
KS	6	0.6256	6	0.6288	6	0.5993	6	0.5698	6
KY	8	0.6750	8	0.6501	8	0.6729	8	0.6439	8
LA	8	0.6372	8	0.6173	8	0.6029	8	0.5919	8
MA	11	0.3336	0	0.3065	0	0.2799	0	0.3333	0
MD	10	0.3444	0	0.3633	0	0.3116	0	0.3434	0
ME	4	0.4860	0	0.4921	0	0.4774	0	0.4960	0
MI	15	0.5179	15	0.5705	15	0.4843	0	0.5017	15
MN	10	0.4982	0	0.5297	10	0.4778	0	0.4603	0
MO	10	0.6360	10	0.6807	10	0.6184	10	0.6132	10
MS	6	0.6314	6	0.6177	6	0.5988	6	0.5939	6
MT	4	0.6271	4	0.6361	4	0.6174	4	0.6080	4
NC	16	0.5274	16	0.5098	16	0.5122	16	0.4834	0
ND	3	0.7496	3	0.7497	3	0.7115	3	0.6636	3
NE	5	0.6552	5	0.7092	5	0.6426	5	0.5688	5
NH	4	0.4812	0	0.4638	0	0.4359	0	0.4956	0
NJ	14	0.4193	0	0.4669	0	0.3603	0	0.4430	0
NM	5	0.4716	0	0.4746	0	0.4310	0	0.4394	0
NV	6	0.5168	6	0.4274	0	0.4898	0	0.4879	0
NY	28	0.3939	0	0.3977	0	0.3439	0	0.4207	0
OH	17	0.5810	17	0.6163	17	0.5519	17	0.5601	17
OK	7	0.7189	7	0.6785	7	0.6970	7	0.6582	7
OR	8	0.4462	0	0.3445	0	0.4111	0	0.3807	0
PA	19	0.5150	19	0.5384	19	0.4748	0	0.4952	0
RI	4	0.3990	0	0.3741	0	0.3442	0	0.4250	0
SC	9	0.6018	9	0.5751	9	0.5854	9	0.5577	9
SD	3	0.6950	3	0.6709	3	0.6838	3	0.6456	3
TN	11	0.6551	11	0.6422	11	0.6433	11	0.6308	11
TX	40	0.5635	40	0.5288	40	0.5343	40	0.4903	0
UT	6	0.6469	6	0.6613	6	0.6382	6	0.6128	6
VA	13	0.4748	0	0.4327	0	0.4369	0	0.4106	0
VT	3	0.3509	0	0.3257	0	0.3215	0	0.3652	0
WA	12	0.4225	0	0.3234	0	0.3882	0	0.3468	0
WI	10	0.5381	10	0.6061	10	0.5097	10	0.5512	10
WV	4	0.7610	4	0.7304	4	0.7271	4	0.7148	4
WY	3	0.7653	3	0.7732	3	0.7587	3	0.7514	3
All Votes	538		312		335		261		215

Notes: Total EC refers to the number of electoral votes per state. EC votes are the predicted Republican electoral votes, excluding Hawaii (Democratic), Alaska (Republican) and Florida (Republican). Regional forecasts are based on eight panel regressions, one for each BEA region. Red indicates Republican states, blue indicates Democrat states, and green represents swing states.

Table 10 (cont.) 2024 Lasso-OCMT Average

		Lasso-OCMT Average			
		Pooled		Regional	
State	**Total EC**	$\hat{V}_S$	**EC Votes**	$\hat{V}_S$	**EC Votes**
AK	3	N/A	3	N/A	3
AL	9	0.6636	9	0.6438	9
AR	6	0.6813	6	0.6712	6
AZ	11	0.4973	0	0.5461	11
CA	55	0.3409	0	0.2607	0
CO	9	0.4222	0	0.4458	0
CT	7	0.3789	0	0.3899	0
DC	3	0.0455	0	0.0565	0
DE	3	0.4079	0	0.4298	0
FL	30	0.5232	30	0.4990	30
GA	16	0.5169	16	0.4991	0
HI	4	N/A	0	N/A	0
IA	6	0.5957	6	0.5979	6
ID	4	0.6960	4	0.6958	4
IL	20	0.4293	0	0.4938	0
IN	11	0.6186	11	0.6404	11
KS	6	0.6124	6	0.5993	6
KY	8	0.6740	8	0.6470	8
LA	8	0.6201	8	0.6046	8
MA	11	0.3068	0	0.3199	0
MD	10	0.3280	0	0.3533	0
ME	4	0.4817	0	0.4940	0
MI	15	0.5011	15	0.5361	15
MN	10	0.4880	0	0.4950	0
MO	10	0.6272	10	0.6469	10
MS	6	0.6151	6	0.6058	6
MT	3	0.6223	4	0.6220	4
NC	16	0.5198	16	0.4966	0
ND	3	0.7305	3	0.7067	3
NE	5	0.6489	5	0.6390	5
NH	4	0.4586	0	0.4797	0
NJ	14	0.3898	0	0.4550	0
NM	5	0.4513	0	0.4570	0
NV	6	0.5033	6	0.4577	0
NY	29	0.3689	0	0.4092	0
OH	17	0.5664	17	0.5882	17
OK	7	0.7080	7	0.6683	7
OR	7	0.4287	0	0.3626	0
PA	19	0.4949	0	0.5168	19
RI	4	0.3716	0	0.3995	0
SC	9	0.5936	9	0.5664	9
SD	3	0.6894	3	0.6582	3
TN	11	0.6492	11	0.6365	11
TX	38	0.5489	40	0.5095	40
UT	6	0.6426	6	0.6370	6
VA	13	0.4558	0	0.4216	0
VT	3	0.3362	0	0.3455	0
WA	12	0.4053	0	0.3351	0
WI	10	0.5239	10	0.5787	10
WV	5	0.7441	4	0.7226	4
WY	3	0.7620	3	0.7623	3
All Votes	538		282		274

Notes: Total EC refers to the number of electoral votes per state. EC votes are the predicted Republican electoral votes, excluding Hawaii (Democratic), Alaska (Republican) and Florida (Republican). Regional forecasts are based on eight panel regressions, one for each BEA region. Red indicates Republican states, blue indicates Democrat states, and green represents swing states.

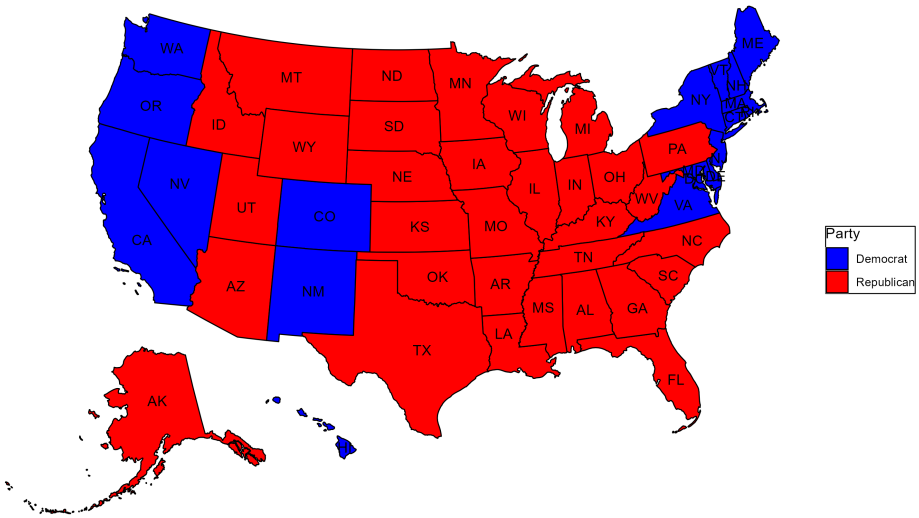

Figure 8: Regional Forecasts of Lasso in 2024

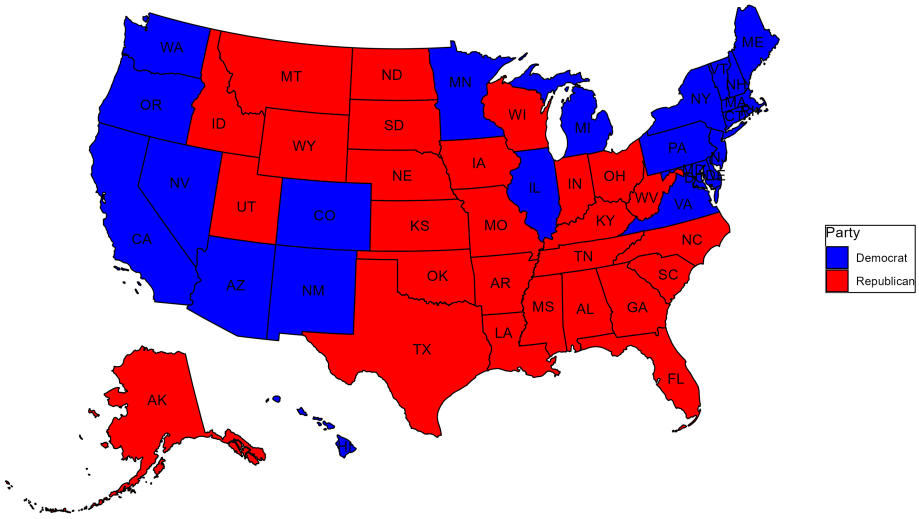

Figure 9: Pooled Forecasts of OCMT in 2024

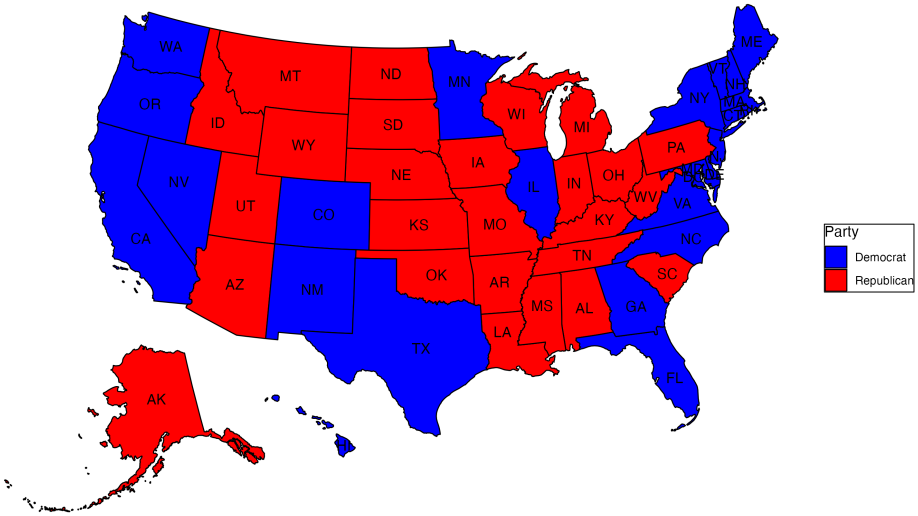

Figure 10: Regional Forecasts of OCMT in 2024

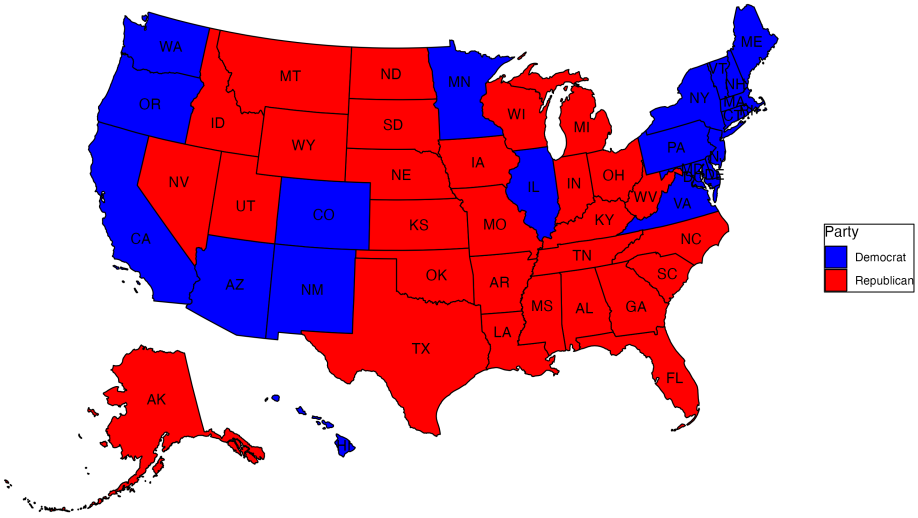

Figure 11: Pooled Forecasts of Lasso+OCMT Average in 2024

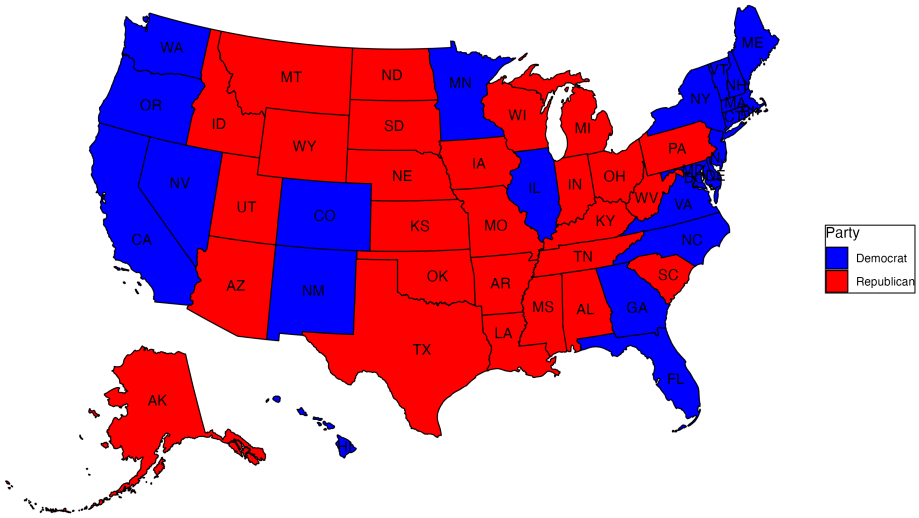

Figure 12: Regional Forecasts of Lasso+OCMT Average in 2024

6 Concluding Remarks

This paper has followed Ahmed and Pesaran (2020) and produced forecasts of voter turnout and election outcomes for the US 2024 Presidential election, using county level data and a novel validation approach whereby the validity of the panel regressions are judged by their ability to produce accurate forecasts for the non-Swing states (those states that have been consistently won by the same party over the past five presidential election cycles). The updated panel regressions are not that different from those reported by AP for 2020 election, with the notable exception that migration trends have become more favourable to the Republican party. Another important feature of the updated results is the sharp fall predicted in voter turnout across all states, from a national average of 66.23 per cent in 2020 to a figure of around 52-54 per cent predicted for 2024. This result is particularly important noting that AP's models predicted the turnout for 2020 very accurately, ranging from 61.97 to 66.19 per cent, as compared to the realized value of 66.23 per cent. See the published version of AP paper (Ahmed and Pesaran (2022) where 2020 predictions and realized values are compared). Both of these shifts tend to favour the Republicans, and the electoral vote forecasts seem to reflect this, with the Republican electoral votes predicted to be stronger in 2024 as compared to the 2020 election.

References

Ahmed, R., and M. H. Pesaran (2020): "Regional Heterogeneity and U.S. Presidential Elections," CESifo Working Paper No. 8615.

Ahmed, R., and M. H. Pesaran (2022): "Regional heterogeneity and US presidential elections: Real-time 2020 forecasts and evaluation," *International Journal of Forecasting*, 38(2), 662–687.

Chudik, A., G. Kapetanios, and M. H. Pesaran (2018): "A one covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models," *Econometrica*, 86, 1479–1512.

Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 58, 267–288.

Data Appendix: Descriptions and Sources

A Introduction

This appendix provides a description of the covaraites used in the forecasting exercise and the provides the data sources used. For further details see Ahmed and Pesaran (2020, 2022).

A.1 Incumbent Party and Presidency

incumbent party (r_incu_pa) The variable 'incumbent party' provides a numerical representation of the political party of the sitting U.S. president during a presidential election. It is defined as:

$$\text{r_incu_pa} = 1(\text{current party is Republican}) - 1(\text{current party is Democratic}).$$

where the expressions 'current party is Republican' and 'Current party is Democratic' are are indicator functions that return 1 if true and 0 if false. For a detailed derivation, refer to Table 5, which outlines this calculation based on U.S. presidential election results.

incumbent president (r_incu_pr) The variable 'incumbent president' quantifies whether the current president (or vice president) is running for re-election, as well as their political party affiliation. It is defined by the following equation:

$$\begin{aligned}\text{r_incu_pr} = {} & 1(\text{current incumbent is Republican and running for re-election}) \\ & - 1(\text{current incumbent is Democratic and running for re-election}) \\ & + 0(\text{no re-election candidates}),\end{aligned}$$

where each term is an indicator function, equal to 1 if the condition is met and 0 if not. For further details, see Table A.1, which shows the derivation based on U.S. presidential election results.

A.2 Massachusetts Institute of Technology (MIT) Election Data and Science Lab

A.2.1 County US Presidential Votes

The following variables are derived from the dataset "County Presidential Election Returns 2000-2020":

Table A.1: U.S. Presidential Election Results (2000–2024)

Election Year	President at Election Year	Party at Election Year	Major Candidates		r_incu_pa	r_incu_pr
			Republican	Democratic		
2000	Bill Clinton	Democratic	George W. Bush	Al Gore	-1	0
2004	George W. Bush	Republican	George W. Bush	John Kerry	1	1
2008	George W. Bush	Republican	John McCain	Barack Obama	1	0
2012	Barack Obama	Democratic	Mitt Romney	Barack Obama	-1	-1
2016	Barack Obama	Democratic	Donald Trump	Hillary Clinton	-1	0
2020	Donald Trump	Republican	Donald Trump	Joe Biden	1	1
2024	Joe Biden	Democratic	Donald Trump	Kamala Harris	-1	0

Notes: r_incu_pa = 1(current party is Republican) – 1(current party is Democratic). When the incumbent president belongs to the Republican Party, the indicator current party is Republican equals 1, and current party is Democratic equals 0. r_incu_pr = 1(current incumbent is Republican and running for re-election) – 1(current incumbent is Democratic and running for re-election) + 0(no re-election candidates). When the incumbent president is a Democrat running for re-election, the first indicator equals 1 and the second equals 0. If there is no incumbent running for re-election, both indicators equal 0.

change in logged republican votes odds ratio (dLRO) This variable measures the change in the logged odds ratio of Republican to Democrat votes at the county level during U.S. presidential elections. It is calculated as: $\text{dLRO}_{cr,t+4} = \text{LRO}_{cr,t+4} - \text{LRO}_{cr,t}$, where $\text{LRO}_{cr,t}$ represents the log-odds ratio of Republican to Democratic votes for county c in region r for election year t.

change in Republican vote share (dV) This variable captures the change in Republican vote share by county and is computed as: $\text{dV}_{cr,t+4} = V_{cr,t+4} - V_{cr,t}$, where $V_{cr,t}$ is the Republican vote share for county c in region r for the election year t.

change in logged Republican vote share (dLV) This variable measures the change in the log of the Republican vote share by county. It is computed as: $\text{dLV}_{cr,t+4} = \log(V_{cr,t+4}) - \log(V_{cr,t})$, where $V_{cr,t}$ is the Republican vote share for county c in region r for election year t.

Republican votes (repV) This variable represents the total number of Republican votes cast in a U.S. presidential election by county.

Democratic votes (demV) This variable represents the total number of Democratic votes cast in a U.S. presidential election by county.

A.2.2 Biennial State U.S. Midterm Votes

The following variables are derived from the dataset "U.S. House 1976–2022." Prior to generating these variables, it is necessary to update party affiliations. For instance, Bernard Sanders (or Bernie Sanders) should be classified as a Democrat, and similar parties such as "Democratic-Farmer-Labor," "Democratic-Nonpartisan League," and "Democratic-NPL" should also be relabeled as "Democrat."

House midterm changes in log Republican votes odds ratio (dLRO_house) This variable tracks changes in the logged odds ratio of Republican to Democrat votes during biennial U.S. House elections by state. It is computed as: $\text{dLRO_house}_{s,t+4} = \text{LRO_house}_{s,t+4} - \text{LRO_house}_{s,t}$, where $\text{LRO_house}_{s,t}$ is the log-odds ratio of Republican to Democratic votes for state s for election year t.

House midterm changes in Republican votes (dV_house) This variable represents the change in Republican vote share during biennial U.S. House elections by state. It is computed as: $\text{dV_house}_{s,t+4} = \text{V_house}_{s,t+4} - \text{V_house}_{s,t}$, where $\text{V_house}_{s,t}$ is the Republican vote share for state s in election year t.

House midterm changes in log Republican votes (dLV_house) This variable tracks changes in the logged Republican vote share during biennial U.S. House elections by state. It is calculated as: $\text{dLV_house}_{s,t+4} = \log(\text{V_house}_{s,t+4}) - \log(\text{V_house}_{s,t})$, where $\text{V_house}_{s,t}$ represents the Republican vote share for state s in election year t.

A.3 Bureau of Economic Analysis (BEA)

We retrieved four datasets from the BEA website for our analysis. The first is the CAGDP9 dataset, which provides real GDP in chained dollars by county and MSA from the 2021 Local Area Gross Domestic Product report. The second dataset is the quarterly GDP by State and Industry from the 4th Quarter and Annual 2022 Gross Domestic Product by State report, including historical SIC tables. The third dataset, labeled 'SAPCE' contains personal consumption expenditures (PCE) by state. The fourth dataset also labeled 'SAPCE,' provides annual personal income and employment by state.

Let $Y_{cr,t}$ represent real GDP for county c in region r during year t, and let $Y^*_{cr,t}$ denote its "local" or "regional" counterpart. Local real GDP is defined as:

$$Y^*_{cr} = \sum_{c'=1}^{N_R} w_{c,c'} Y_{c',r,t},$$

where N_R is the number of counties within 100 miles of c, including county c itself. The weight $w_{c,c'} = 1/N_R \cdot \mathbf{1}\{c' \text{ is within 100 miles of } c\}$, ensuring that $\sum_{c'=1}^{N} w_{c,c'} = 1$.

individual left-behind measure of the county-level real GDP growth (LBCG) The variable LBCG measures whether county c is economically lagging behind relative to the national level. It is computed as:

$$G_{cr,t}(h) = \frac{\ln(Y_{cr,t}) - \ln(Y_{cr,t-h})}{h} - \frac{\ln(Y_t) - \ln(Y_{t-h})}{h},$$

where h represents the time horizon and Y_t is the national real GDP. If $G_{cr,t}(h) < 0$, county c is considered left behind. If $G_{cr,t}(h) > 0$, county c is not left behind. For the forecast for the 2024 election, we updated the dataset to 2021.

local left-behind measure of the county-level real GDP growth (LBLG) The variable LBLG assesses whether county c is economically lagging behind relative to other counties within 100 miles. It is computed as:

$$G^*_{cr,t}(h) = \frac{\ln(Y^*_{cr,t}) - \ln(Y^*_{cr,t-h})}{h} - \frac{\ln(Y_t) - \ln(Y_{t-h})}{h}.$$

County c is locally left behind if $G^*_{cr,t}(h) < 0$, and not left behind if $G^*_{cr,t}(h) > 0$. For the forecast for the 2024 election, we updated the dataset to 2021.

inflation rate (inf) The variable 'inf' represents the U.S. inflation rate, which does not vary across counties or states. We calculate the inflation rate as the logged change in each state's GDP deflator from the previous quarter. The GDP deflator is defined as (Nominal GDP/Real GDP) $\times$ 100. For election forecasting, we use the inflation rate for Q3 of the year preceding the election to Q2 of the election year (e.g., for the 2004 election, the inflation rate from Q3 2003 to Q2 2004).

log change of the health care expenditure (hlt) The variable 'hlt' measures the annual growth rate of health care expenditures by state. We calculate the logged annual growth rate, with the last available data for the 2024 forecast being from 2023.

log change of the government employee share (gov) The variable 'gov' represents the growth rate in the share of government employees by state. We compute the annual logged growth rate, with the most recent data for the 2024 forecast available from 2022.

log change of the rent price of the housing (rp) The variable 'rp' tracks the log change in housing rent prices by state. We calculate the annual logged growth rate, with the most recent data for the 2024 forecast available from 2023.

A.4 United States Department of Agriculture (USDA)

We derived the following variables from two USDA datasets: "Educational Attainment for Adults Age 25 and Older for the United States, States, and Counties, 1970–2022," and the "Atlas of Rural and Small-Town America".

proportion of the people with bachelors degree or higher (edu2000) This variable represents the proportion of adults aged 25 and older who hold a bachelor's degree or higher. For the years 2000 to 2016, the 2000 values were used. For 2020, the average proportion from 2017 to 2021 was used.

local proportion of the people with bachelors degree or higher (edu2000L) This variable calculates the local proportion of individuals with a bachelor's degree or higher by averaging the values from counties within 100 miles of county c. The same time periods and cumulative values used for the edu2000 variable are applied here. The local measure follows the method described in the 'left-behind' calculation.

total net international migration rate (migrate10) This variable measures the total net international migration. For the years 2000 to 2016, cumulative values from 2010 to 2015 were used, and for 2020, cumulative values from 2010 to 2019 were applied.

local total net international migration rate (migrateL) This variable calculates the local net international migration by averaging the net migration rates of counties located within 100 miles of county c. The same time periods and cumulative values used for migrate10 are applied. The local measure follows the same method as the 'left-behind' measure.

total population (pop) This variable represents the total population of county c.

local population (popL) This variable calculates the average population of counties located within 100 miles of county c.

population density (popdens) This variable measures the population density (people per square mile) by county. The 2020 values have been updated.

degree of urbanization (rural) This variable represents the degree of urbanization by county, with lower values indicating more urbanized areas and higher values corresponding to more rural areas.

A.5 U.S. Census

The variables were sourced from the U.S. Census "Small Area Income and Poverty Estimates (SAIPE) School District Estimates."

poverty estimates (povr) This variable represents the percentage of individuals living at or below the poverty level within each county.

median income (m_inc) This variable captures the median household income for each county.

A.6 Association of Religion Data Archives (ARDA)

The following variables are derived from the "U.S. Religion Census - Religious Congregations and Membership Study, 2020 (County File)" provided by the Association of Religion Data Archives.

religiosity (religion) This variable represents the proportion of individuals in county c who are affiliated with a religion.

local religiosity (religionL) This variable measures the local religiosity of county c by calculating the average religiosity of counties located within 100 miles of county c.

A.7 American Community 5-year Surveys (ACS)

The following variable is obtained from the "Citizen Voting Age Population (CVAP) Special Tabulation from the 2018-2022 5-Year American Community Survey" in the American Community 5-Year Surveys. The ACS dataset spans from 2000 to the 2018-2022 survey.

voting age population (VAP) This variable represents the voting age population for each county. Due to data availability, the estimates used for each election year are as follows: 2016-2020 estimates for 2020, 2012-2016 estimates for 2016, 2008-2012 estimates for 2012, and 2005-2009 estimates for the 2004 and 2008 elections.

A.8 Bureau of Labor Statistics (BLS)

The following variables are obtained from the BLS download website. However, before use, the data must be cleaned and formatted into a panel data structure. The dataset is available up to August 2024.

unemployment rate (ump) This variable represents the unemployment rate for each county. For election year t, the annual average unemployment rate is calculated using data from July of the previous year $t-1$ to June of year t. Additionally, a short-term average for the 3-month period from July to September of each election year is also computed.

change in unemployment rate (umpd) This variable represents the change in the unemployment rate for each county.

A.9 Federal Reserve Bank of Dallas

The following variables are sourced from the dataset "Real Trade-Weighted Value of the Dollar by U.S. State" provided by the Federal Reserve Bank of Dallas. The dataset is available up to June 2023.

real effective exchange rates (rusd) This variable represents the state-level real effective exchange rate (REER) appreciation rate of the U.S. dollar. For each election year t, the annualized change in the log of REER is calculated using data from July of the previous year $t-1$ to June of year t. Additionally, short-term averages for the 3-month period from July to September of each election year are also computed.

A.10 Zillow

The following variable is obtained from the Zillow Home Value Index (ZHVI) All Homes (SFR, Condo/Co-op) Time Series, which provides smoothed, seasonally adjusted data on typical home values and market trends across regions and housing types. ZHVI represents the typical home value in the 35th to 65th percentile range. The dataset is available through August 2024.

house prices (hpret) This variable measures the log changes in house prices at the county level. Since not all counties or year-month observations are available, local house price returns are estimated by averaging the prices of counties within 100 miles of county c, including county c itself. For each election year t, the logged annual change in house prices

is calculated as the average monthly change from July of the previous year $t-1$ to June of year t, then annualized. Additionally, short-term averages over the 3-month period from July to September of each election year are also computed.

A.11 National Bureau of Economic Research (NBER): County Distance Database (2010)

The dataset containing distances between counties, used in calculating local variables, is sourced from the National Bureau of Economic Research (NBER) "County Distance Database." The data is from 2010, and any counties need to be updated according to the FIPS code revisions that occurred after 2010.

A.12 Voter Turnout (TO)

voter turnout (TO) This variable represents voter turnout, calculated as a percentage of the voting age population (VAP). Table A.2 presents the realized voter turnout and Republican vote share for U.S. presidential elections from 2004 to 2020.

Table A.2: National Voter Turnout and Republican Vote Share of the US President Election from 2004 to 2020

	Realized				
	2004	2008	2012	2016	2020
Voter Turnout	0.5770	0.6178	0.5843	0.5705	0.6623
Vote Share	0.5128	0.4635	0.4802	0.4892	0.4776
EC Votes	286	173	206	304	232

Notes: The realized national Republican vote refers to the 2004-2020 Republican share of the two-party vote across the U.S. mainland states and Washington, D.C. Electoral College (EC) votes represent both realized and predicted Republican electoral votes, assuming Hawaii casts its votes for the Democratic candidate and Alaska casts its votes for the Republican candidate.

A.13 Other Sources

County land area in squared miles (land_mi) The variable land_mi represents the land area of U.S. counties in square miles. We use the same values for 2020 and 2024 as those provided in Ahmed and Pesaran (2022).

Mail-in voting (vmail_d) The variable vmail_d is a state-level indicator that takes one of three values: +1 if mail-in voting is the default method, 0 if it is optional but available

to everyone, and -1 if an excuse is required to vote by mail.

References

AHMED, R., AND M. H. PESARAN (2020): "Regional heterogeneity and US presidential elections: Real-time 2020 forecasts and evaluation," CESifo Working Paper No. 8615.

AHMED, R., AND M. H. PESARAN (2022): "Regional heterogeneity and US presidential elections: Real-time 2020 forecasts and evaluation," *International Journal of Forecasting*, 38(2), 662-687.