

Online Appendix to

**A Long-Range State-Level Forecast of the
2020 Presidential Election**

Detailed Explanation on the derivation of the 2020 Long-Range Forecast

The prediction for each state, i , is derived from the coefficients found in Table 1 of the Symposium article:

FORECAST $_i$ =

$$(1.029 \times \text{PRIOR}_i) + (0.592 \times \text{POLLS}) + (2.693 \times \text{HOME}_i) - (1.070 \times \text{TERMS}) - 31.161$$

Where:

PRIOR $_i$ = The share of the 2PPV won in the state by the Democratic candidate in the previous election,

POLLS = Joe Biden's average two-party share of all national head-to-head polls pitting him against Donald Trump during the month of October, 2019. [Value for 2020: 54.88]

HOME $_i$ = A party-adjusted (positive for Democrat, negative for Republican) indicating the home state of the candidate, adjusted for the one election lag of effect in the previous election. (Explained in detail below)

TERMS = A party-adjusted (positive for Democrat, negative for Republican) indicating the number of consecutive terms the current incumbent party has occupied the White House. [Value for 2020: -1]

The HOME variable captures two effects:

- 1) The boost in support a candidate receives in their home state, and
- 2) The loss of support their party's candidate in the previous election received in their own home state

For 2020, Joe Biden's home state of Delaware is coded 1, and Donald Trump's home state of New York is coded -1. Even though Trump officially changed his State of Residence from

Florida to New York in September of 2019, I determined that this change would not fundamentally change his appeal to either New Yorkers or Floridians. Complicating matters is the fact that in 2016, both candidates were from New York. Therefore, in 2016, New York was coded 0 on HOME, to indicate that the home state effect was effectively cancelled out for both candidates. For all practical purposes, then, New York's value of -1 for 2020 is representing *the loss of Hillary Clinton's home state effect from 2016 to 2020*.

Out-of-Sample Model Performance

A more stringent test of the model's accuracy in making forecasts is its *post-casting* performance when an election is excluded from the analysis to generate coefficients with the data from *all other elections* in the analysis. These coefficients are then used to "forecast" outcomes in the excluded election. Understandably, these out-of-sample forecasts typically generate slightly less accurate "predictions" than those found in in-sample predictions presented in the Symposium article. Appendix Table 1, presented below, replicates the bottom half of the Article's Table 1 showing the model's year-to-year performance based on these out-of-sample estimates.

Not surprisingly the model performs less well than with the in-sample estimates. Even so, these out-of-sample results are largely comparable to those from the in-sample results. Only one additional state outcome is incorrectly assigned in the Out-of-Sample analysis: Florida in 2004.

Overall, 2000 stands out as the election for which the national-level extrapolations of the Out-of-Sample estimates perform much less well than those from the In-Sample estimates. While the overall percentage of cases correctly predicted remained the same, that masks a shift that

ultimately made the national level extrapolations significantly under-predict Gore's performance in both the popular vote and electoral college. Three states that the In-Sample estimates incorrectly predicted would be won by Gore in 2000 (Arkansas, Louisiana, and West Virginia) were *correctly* assigned in the Out-of-Sample analysis. On the other hand, three states shifted to from *correct* in the In-Sample analysis to *incorrect* predictions in the Out-of-Sample analysis. California, Michigan, and Washington. Clearly, this suggests that the 2000 election represents an important data point that definitely needs to be included in future forecasting endeavors.

Appendix Table 1: Long-Range Model Out-of-Sample Model Performance

	<u>1996</u>	<u>2000</u>	<u>2004</u>	<u>2008</u>	<u>2012</u>	<u>2016</u>	<u>OVERALL</u>
States Correctly Predicted	88%	84%	92%	94%	100%	92%	91.7%
Mean Absolute Error	2.65	3.80	2.34	2.91	2.21	2.60	2.75

National-Level Predictions (Excluding DC)

	<u>1996</u>	<u>2000</u>	<u>2004</u>	<u>2008</u>	<u>2012</u>	<u>2016</u>	<u>OVERALL</u>
National Popular Vote Error	54.1 -0.6	46.1 -4.1	50.7 +2.0	52.3 -1.3	53.1 +1.2	50.1 -0.9	1.7*
Electoral College Vote Error	399 +36	128 -105	284 +35	322 -40	329 0	282 +58	45.7*

* Mean Absolute Error