CITIZEN FORECASTING 2020: A STATE-BY-STATE EXPERIMENT

Andreas E. Murr, University of Warwick
Michael S. Lewis-Beck, University of Iowa

Accepted for publication at *PS: Political Science and Politics*.

Andreas E. Murr is Associate Professor in Quantitative Political Science in the Department of Politics and International Studies at the University of Warwick. His research interests are election forecasting, party leadership elections, and political methodology. He may be reached at a.murr@warwick.ac.uk.

Michael S. Lewis-Beck is F. Wendell Miller Distinguished Professor of Political Science at the University of Iowa. He has authored or co-authored over 300 articles and books, including *Economics and Elections*, *The American Voter Revisited*, *French Presidential Elections*, *Forecasting Elections*, *The Austrian Voter* and *Applied Regression*. He may be reached at michael-lewis-beck@uiowa.edu.

= pull quotes

The leading approaches to scientific election forecasting in the United States consist of structural models, prediction markets and opinion polling. With respect to the last, by far the dominant mode relies on vote intention polling, e.g., "If the election were held tomorrow, who would you vote for?" However, there exists an abiding opinion polling strategy that shows a good deal of promise—citizen forecasting. That is, rather than query on vote intention, query on vote expectation, e.g., "Who do you think will win the upcoming election?" This approach has been pursued most extensively in the United Kingdom (Murr 2016) and the United States (Lewis-Beck and Tien 1999). Recent performance evaluations have shown that in the United Kingdom vote expectations clearly offer more predictive accuracy than vote intentions (Murr et al. forthcoming) and that in the United States vote expectations appear to be superior to an array of rival forecasting tools (Graefe 2014). However, the timing of the data collection has forced most of the studies using citizen forecasts to forecast elections *ex post*, i.e., after they occurred. Indeed, to date, there are only two *ex ante* citizen forecasting papers to have appeared before a

national election (Lewis-Beck and Stegmaier 2011; Murr 2016). Both these efforts forecasted British General Elections, with Murr (2016) relatively most accurate among 12 academic forecasts (Fisher and Lewis-Beck 2016).

With respect to the United States, the case at hand, none of the work has been *ex ante* and all studies have focused on the national level, with the exception of a lone study carried out at the state level (Murr, 2015). The latter point seems critical, since the final selection of the president takes place in the Electoral College. The citizen forecasting research here stands unique, being *ex ante* and focusing on the states. Utilizing survey questions on Amazon.com's Mechanical Turk (MTurk), administered in July, we render forecasts for the November 2020 presidential contest. This experiment, which has been conducted before-the-fact and looks at the states, provides a strong test of the quality of citizen forecasting in this American election.

ANES EXAMPLES

A standard dependent variable in American citizen forecasting studies is the two-party popular vote share received by the president's party, in the nation as a whole. The American National Election Study (ANES) survey question on voter expectations asks, "Who do you think will win in the November election?" Posed in each ANES pre-election survey, 1952 to 2016, it serves as a standard independent variable tapping voter expectations. We estimate a simple regression equation on the seventeen elections across this period to show how voter expectations predict the popular vote. (See table 1 and columns 2 to 4, table 2, below.) [The data and code to replicate all analyses in this article are available on the *PS: Political Science & Politics* Dataverse within the Harvard Dataverse Network (Murr and Lewis-Beck 2020).]

[Place table 1 here]

[Place table 2 here]

We observe that the model provides a snug fit to the data and close out-of-sample predictions. As a benchmark, say that a Democratic "win" consists of a two-party popular vote of 50.1 percent or more. What voter expectation value will generate such a prediction? Answer: when 54 percent of the respondents say the Democratic candidate will win. [That is, 39.3 + .2(54.0) = 50.1.]

We can explore a similar equation at the state level, since for twelve of these ANES surveys it is possible to match the respondent's state to his or her expectation about "who will win in the state?" (See table 3). We test how well it predicts both the popular vote in states (see table 3) and the electoral vote in the nation as a whole (see columns 8 to 10, table 2).

[Place table 3 here]

Again, the model both fits the data well and predicts the popular vote well in out-of-sample tests, though less so than at the national-level. In terms of forecasting, we see that to achieve a Democratic "win-the-state" popular vote prediction of 50.1 percent, we need an expectations value of 57 percent. [That is, 33.0 + .3 (57.0) = 50.1.]

These popular vote predictions also imply a prediction of who wins the state, which in turn implies a prediction for the electoral vote in the nation as a whole. Table 2 shows that this model-based forecast of the electoral vote worked quite will in the last three elections, getting two of them right. [Instead of using the regression model, we can also predict the winner to be the candidate who most citizens say will win. This data-based forecast performs similarly, and is explored in the middle columns of table 2.]

These ANES findings, at both the state and national levels, offer strong evidence that citizen forecasting can "work." Why should vote expectations predict well, perhaps even better

than vote intentions? Their predictive power comes from their utilization of more information, drawing on the citizens' social networks (Leiter et al. 2018; Murr 2017). Put simply, two (or more) heads forecast better than one. This increase in information depends for its utility on the "wisdom of the crowds," as formally expressed in the Condorcet Jury Theorem, e.g., citizens are not all forecasting randomly and their accuracy increases with size (Murr 2015).

The foregoing results provide guidelines for our MTurk experiment in the states, which we unfold below.

THE MTURK EXAMPLE

Utilizing Amazon.com's Mechanical Turk, we gathered *ex ante* election expectations data on a sample of citizens in each American state during the month of July. We obtained responses to the national question, "Who do you think will be elected President in November?" We also obtained responses to a state question, "Which candidate for President do you think will win in this state?" The total N = 2483. (Within each state, the aim was to poll at least N = 30.)

Of immediate interest is the percentage of respondents, in the nation as a whole, who believe the Democratic candidate, Joe Biden, will win. (See column 2, table 4.) That number is 51.35 percent, a value encouraging to Biden supporters. However, it is not enough, taking into account what that expectation score can deliver, according to our national level model in table 4 (columns 4 and 5), which predicts Biden will lose. That is, he will win 39.3 + .2 (51.35) = 49.60 percent of the national popular vote.

[Place table 4 here]

What is the prediction for the Electoral College? Suppose we look directly at the prediction for each state, simply declaring a "state win" for Biden if 50 percent-plus of the

respondents in that state forecast Biden. The expectation percentages, state-by-state, are given in columns 3 and 7 of table 5. One observes results with face validity, e.g., the red state of Oklahoma shows 37% for Trump, the blue state of Massachusetts shows 59% for Biden. However, curious classifications exist, foremostly California at 46% for Biden. We have double-checked our calculations for this case, as well as other surprises, and can find no errors. They may not hold for the nation on election day. But if they do, Biden would receive 204 electoral votes and Trump 334. The ANES-derived forecast model (see table 3) suggests an even stronger Trump Electoral College victory, with 357 electoral votes. (See the electoral votes summary in table 6 and the map in figure 1.) [We report the uncertainty of our forecasts in the Supplemental Material.]

[Place table 5 here]

[Place table 6 here]

[Place figure 1 here]

DISCUSSION

The results of the ANES analysis, especially at the state level, clearly indicate that citizen forecasting can produce accurate state-by-state results. How much confidence should we have in the MTurk results? Consider the quality of the sample. Take ANES as a case in point; while it uses probability sampling at the national level, it does not follow probability sampling at the state level. But those state level results make for convincing forecasts. Can we say the same about the MTurk convenience samples?

Here are some things we know. First, essentially because of Condorcet's Jury Theorem and the "wisdom of the crowd," random samples are not required for citizen forecasting (Murr 2011). What is needed is an independent sample that predicts better than chance. The state

forecasts are different from chance ($p<0.001$). Whether they are better or worse than chance can be tested. One data credibility check is how many respondents forecast a candidate winning both the state and the nation. In our MTurk sample the number is 77%, close to the historical ANES average of 73%. Further, the MTurk national forecast receives concurrent validity, since it is essentially the same as the current national citizen forecast from *The Economist/YouGov Polls* (where Biden = 39%, Trump = 40%, as of July 12–14, 2020).

The MTurk workers appear to be doing as well as some more known national samples. Further, their workers have some characteristics that are shown to improve individual forecasting accuracy. For one, our MTurk workers are, on average, quite interested in the election campaign (scoring 2.5 on a 3.0 scale), a characteristic positively correlated with accuracy in both state forecasts (Murr 2015) and national forecasts (Lewis-Beck and Tien 1999). Moreover, Amazon monitors the quality and identity of the workers. For example, they have to provide official forms of identification and Amazon validates they are working where they claim. Also, the workers we employed had approval ratings of 90% or more.

In sum, the charge of a fatally flawed sample is harder to make than might be supposed. However, another issue exists, besides that of sample quality— the issue of sample timing. Recall that these are July data. Is that date too far away from the election to yield accurate forecasts? Extended work on the UK case has shown 93 percent of citizen forecasts were accurate two quarters before the election (Murr et al. forthcoming). For the US, a current paper exploring the accuracy-lead time trade-off for the 2016 presidential contest gave an optimal lead (before T =1) of 48 days before, even 86 days before if one sacrifices a half-point (Jennings et al. 2020, table 1, 954). These numbers imply the July estimates, grosso modo, may hold. The election results themselves will provide a sharp answer to this question.

REFERENCES

Fisher, Stephen D. and Michael S. Lewis-Beck. 2016. "Forecasting the 2015 British General Election: the 1992 Debacle All Over Again?" *Electoral Studies* 41 (1): 225–229.

Graefe Andreas. 2014. "Accuracy of Vote Expectation Surveys in Forecasting Elections." *Public Opinion Quarterly* 78 (S1): 204–232.

Jennings, Will, Michael S. Lewis-Beck, and Christopher Wlezien. 2020. "Election Forecasting: Too Far Out?" *International Journal of Forecasting* 36 (3): 949–962.

Leiter, Debra, Andreas E. Murr, Ericka Rascón Ramírez, and Mary Stegmaier. 2018. "Social networks and citizen election forecasting: The more friends the better." *International Journal of Forecasting* 34(2), 235–248.

Lewis-Beck, Michael S. and Mary Stegmaier. 2011. "Citizen Forecasting: Can UK Voters See the Future?" *Electoral Studies* 30 (2): 264–268.

Lewis-Beck, Michael S. and Charles Tien. 1999. "Voters as Forecasters: a Micromodel of Election Prediction." *International Journal of Forecasting* 15 (2): 175–184.

Murr, Andreas E. 2011. "'Wisdom of Crowds'? A Decentralised Election Forecasting Model That Uses Citizens' Local Expectations" *Electoral Studies* 30 (4): 771–783.

— — —. 2015. "The Wisdom of Crowds: Applying Condorcet's Jury Theorem to Forecasting US Presidential Elections." *International Journal of Forecasting* 31 (3): 916–929.

— — —. 2016. "The Wisdom of Crowds: What Do Citizens Forecast for the 2015 British General Election?" *Electoral Studies* 41 (1): 283–288.

— — —. 2017. "Wisdom of Crowds." In: *The Sage Handbook of Electoral Behaviour*, ed. Kai Arzheimer, Jocelyn Evans, and Michael S. Lewis-Beck, 835–860. London: SAGE.

Murr, Andreas E. and Michael S. Lewis-Beck. 2020. "Replication Data for: Citizen Forecasting 2020: A State-by-State Experiment." Harvard Dataverse. doi: 10.7910/DVN/S1594S.

Murr, Andreas E., Mary Stegmaier, and Michael S. Lewis-Beck. Forthcoming. "Vote Expectations Versus Vote Intentions: Rival Forecasting Strategies." *British Journal of Political Science*.

Figure 1. Map of model-based election forecast by state. The numbers indicate the state's electoral votes.

Table 1. Linear regression model of national two-party popular vote for Democratic candidate on Democratic national forecasts, 1952–2016.

	Coefficient	Standard error
Intercept	39.3	1.5
Percentage of Democratic national forecasts	0.2	0.0
N	17	
R-squared	0.77	
RMSE (leave-one-national-election-out cross-validation)	3.07	

Table 2. Out-of-sample one-step ahead forecasting of Democratic national two-party popular vote and electoral vote in the last three elections.

	Popular vote			Electoral vote (Data)			Electoral vote (Model)		
Year	Actual	Prediction	Error	Actual	Prediction	Error	Actual	Prediction	Error
2008	53.7	53.9	–0.3	364	328	+36	364	305	+59
2012	52.0	53.1	–1.1	332	343	–11	332	301	+31

2016	51.1	52.3	–1.3	233	279	–46	233	275	–42

Note: In the predictions for 2008 the 88 electoral votes of 17 states without ANES respondents are split equally between Democrats and Republicans.

Table 3. Linear regression model of state two-party popular vote for Democratic candidate on Democratic state forecasts, 1952, 1972–1996, 2004–2016. Data are weighted by the number of observations in a state.

	Coefficient	Standard error
Intercept	33.0	0.5
Percentage of Democratic state forecasts	0.3	0.0
N	450	
R-squared	0.75	
RMSE (leave-one-national-election-out cross-validation)	6.44	

Table 4. Forecast of national popular vote.

	Data		Model	
Number of observations	Democratic forecasts	Forecast winner	Forecast Biden vote share	Forecast winner
2483	51.3%	Biden	49.6%	Trump

Table 5. Forecast of state popular vote and winner.

	N	Democratic Forecasts (Data)	Forecast Biden vote share (Model)		N	Democratic Forecasts (Data)	Forecast Biden vote share (Model)
AL	60	16.7	38.0	MT	16	25.0	40.6
AK	17	52.9	49.1	NE	34	8.8	35.7
AZ	60	31.7	42.6	NV	60	46.7	47.2
AR	55	16.4	38.0	NH	40	72.5	55.1
CA	60	41.7	45.7	NJ	60	71.7	54.8
CO	60	61.7	51.8	NM	34	88.2	59.9
CT	60	38.3	44.7	NY	60	73.3	55.3
DE	35	77.1	56.5	NC	60	41.7	45.7
DC	18	83.3	58.4	ND	19	21.1	39.4
FL	60	41.7	45.7	OH	60	48.3	47.7
GA	60	35.0	43.6	OK	60	13.3	37.0
HI	21	100.0	63.5	OR	60	83.3	58.4
ID	32	3.1	33.9	PA	60	53.3	49.2
IL	60	61.7	51.8	RI	30	80.0	57.4
IN	60	16.7	38.0	SC	60	15.0	37.5
IA	58	32.8	43.0	SD	26	19.2	38.8
KS	53	22.6	39.9	TN	60	18.3	38.6
KY	59	18.6	38.7	TX	60	35.0	43.6
LA	60	25.0	40.6	UT	56	12.5	36.8
ME	30	83.3	58.4	VT	14	78.6	56.9
MD	60	70.0	54.3	VA	60	56.7	50.2
MA	60	85.0	58.9	WA	60	71.7	54.8
MI	60	60.0	51.3	WV	40	27.5	41.4
MN	60	90.0	60.4	WI	60	45.0	46.7
MS	46	13.0	36.9	WY	10	30.0	42.1
MO	60	20.0	39.1				

Table 6. Forecast of electoral votes.

	Forecast (Data)	Forecast (Model)
Biden	204	181
Trump	334	357