3 Reasons Election Forecasts Made False Projections Favoring Joe Biden

Election Day showed us all Joe Biden's lead had been wildly exaggerated.

Casey B. Mulligan

November 6, 2020

Forecasts of this presidential election were missing economics. As I wrote the week before the election, the missing economics fit into three categories: information aggregation, voter incentives to tell the truth, and voter incentives to participate in the general election.

As of the week before the election, the RealClearPolitics poll average was Biden +7.9 (51.4-43.5) percent nationally and +3.1 (48.9-45.9) in the top battleground states (Florida, Pennsylvania, Michigan, Wisconsin, North Carolina, and Arizona). Based on a very similar data set, FiveThirtyEight put Biden's chance of winning at 90 percent while The Economist put it at 96 percent. FiveThirtyEight and The Economist forecasted that Biden would win with a margin of 156 and 162 electoral votes, respectively.

The election turned out to be much closer than indicated by the forecasts and RCP averages (hereafter, "the forecasters"). Two days *after* the election, Donald Trump still had a path to victory in the Electoral College, albeit unlikely. Biden's national popular vote lead is only 2.5 percent. As of Thursday of election week, Biden actually had 0.9 percent fewer combined votes than Trump did in RCP's top battleground states. This departure between real-world results and the forecasts is due at least in part to a failure to use economic reasoning and methods.

Information Aggregation

The information aggregation methods forecasters used are inferior to alternatives that have been well-studied and readily available, and that better reflect results on the economics of information. The methods forecasters used in this election cycle were often more than 10 percentage points more favorable to Biden than the alternatives. I have not seen any explanation from the forecasters why the alternatives are not part of their forecasts.

U.S. voters are heterogeneous and scattered over millions of square miles. One of the challenges for an election forecaster is to assemble or reweight a polling sample that is representative of those who actually vote. For this purpose, the polls used by the forecasters do some reweighting of their samples by demographics.

Most famously, less-educated voters are often (but not always) upweighted, given that their relative propensity to vote in recent years has exceeded their relative propensity to participate in polls. Other demographic variables sometimes used for these purposes are voter registration and voting history.

The polls used by the forecasters contact potential voters by landline, cell phone, and internet and ask them, "Who do you think *you will vote for* in the election for president?" (emphasis

added). These polls ask other questions about the election, but the answers to these questions are not part of the aforementioned averages or forecasts.

As recently as a few years ago, it was well known that another poll question has a much better track record at predicting election outcomes: "Who do *you think will be elected president* in November?" This "expectation" question is included in some of the polls used by the forecasters but, as far as I can tell, is not any part of the forecast.

David Rothschild and Justin Wolfers found in 2012 that the optimal forecasting weight on the expectations question is about nine times the weight on the own-vote question. Furthermore, the optimal weight on voter expectations is, they said, even greater when: (1) "voters are embedded in heterogeneous … social networks," (2) "they don't rely too much on common information," and (3) "small samples are involved." Two of the three of their criteria describe 2020 voters even better than they described voters in the past.

A couple of pollsters expressed interest in voter expectations and found them to deviate significantly from the own-vote intentions that are the foundation of election forecasts. The USC Dornsife poll (seven-day window ending Oct. 29) found Biden leading by 11.6 in own-vote intentions but only 2.1 points in election expectations — a significant gap.

When respondents were asked about how they expected their social contacts to vote, Biden led by a lot less (5.3 points). The Fox News poll found Biden winning by 10 points on the own-vote question but losing by 9 points on "do you think more of your neighbors are voting for Joe Biden or Donald Trump?"

In other words, the foundational own-vote question differed from the other questions by 6 to 19 percentage points, which is several multiples of the RCP battleground average Biden-Trump gap of 3 percentage points.

The Economist forecaster Andrew Gelman, not an economist but an eminent Bayesian statistician, is now rather disingenuously shifting all the blame onto forecasters for assembling skewed samples. Arguably most of his forecast error came instead from his seemingly arbitrary choice of which questions to use from the polls. Gelman has claimed that own-vote questions are better forecasters than expectation questions, which is a respectable conclusion but no reason to completely ignore the expectation questions instead of assigning them somewhat less weight.

Gallup, which does not even ask the own-vote intention question, found that 56 percent of Americans expect Trump to win. During the days before the election, betting markets were putting Trump's chances at about 34 percent. Both of these are far different from the election forecasts built on own-vote polls.

Betting markets are rich enough to allow bets on the electoral margin of victory. The most expensive contracts were Biden 150-209 and Trump 60-99. One of these coincides with the election forecasts, but the other is wildly different. In other words, betting markets seemed to put significant probability on the event that the polls used by the forecasters were way off in Biden's favor.

Voter Incentive: Social Desirability

The incentives to truthfully participate in polls are different from those to participate in the general election, and these differences correlate with political affiliation.

One of the potential incentives is "social desirability." A Cato poll found that 62 percent of Americans "say the political climate these days prevents them from saying things they believe because others might find them offensive."

Trump is the Bad Orange Man to many. Biden is the overwhelming choice among the rich and famous, many of whom blamed Trump's 2016 victory on his allegedly deplorable racist supporters. More recently, looting has become more common and might be related to antiracism in the minds of some voters. This suggests that some fraction of Trump supporters would not acknowledge their support for him — the "shy Trump voter" — especially in Democratic communities.

The quantitative question, of course, is the magnitude of the effect of shy-Trump and shy-Biden voters on poll results relative to election results. I have been amazed how little effort the forecasters put into assessing this magnitude and applying a correction.

Nate Silver dismissed the shy-Trump voter theory by transforming it into a straw man that a huge red wave is coming, which he says is contradicted by early-voting data. In the past, he did nothing more than take the difference between the results of phone polls and internet polls (note that Biden's edge is 2 percentage points more in phone surveys, which is not negligible and is in the expected direction).

Social desirability bias is one of the reasons pollsters have fielded the voter-expectation and voter-neighbors questions. Silver says this approach is "stupid" because it deviates from "the shy-Trump voter narrative." In other words, he does not have a real argument.

Another step, taken by the renegade pollsters Democracy Institute and Trafalgar, is to repeatedly assure respondents that their responses are fully confidential. They assert that the assurances affect the poll results, although I have not seen estimates of the magnitude. In any case, the renegades can be proud of the accuracy of their much-maligned forecasts of the 2020 election.

Voter Incentive: Turnout

Voting in person or by mail is different than picking up the phone or opening an email to begin an online survey. The difference is particularly stark during a pandemic. I do not see that forecasters were accounting for any relationship between the difference and political affiliation.

Republicans are doing much less to withdraw from normal activities in an effort to protect themselves from COVID-19. Gallup found that more than 77 percent of Democrats worried about getting the coronavirus whereas less than 29 percent of Republicans worried about that. More than 70 percent of Democrats were avoiding going to public places as compared to less than 38 percent of Republicans. Meanwhile, confirmed cases were surging in the Midwest as the election neared. A widely televised expert even went so far as to say that Midwesterners would not be adequately protected by a cloth mask and that going out in public required an N-95 mask instead.

In short, COVID-19 surges create a perceived cost of in-person voting that is likely greater for Democrats than Republicans. Some will switch to voting by mail, but perhaps others will not vote at all. The magnitude of this effect cannot be estimated.

Overall, economics by itself suggested, before Election Day, that the widely cited polls were exaggerating Biden's electoral advantage. I am no polling expert and could not assess the

vastness of these polling biases. Betting markets and other methods of information aggregation did not show as much optimism for Biden as did the election forecasters.

These markets appeared to put significant probability mass on the event that Biden's election results would be far worse than his polling results. Indeed, Election Day showed us all that Biden's lead had been wildly exaggerated.