Predicting the 2020 Presidential Election

Liberty Vittert^{1,2} Ryan D. Enos^{3,4} Stephen Ansolabehere³

¹Harvard Data Science Initiative, Harvard University, Cambridge, Massachusetts, United States of America,

²John M. Olin Business School, Washington University in St. Louis, St. Louis, Missouri, United States of America,

³Department of Government, Faculty of Arts and Sciences, Harvard University, Cambridge, Massachusetts, United States of America,

⁴Institute for Quantitative Social Science, Faculty of Arts and Sciences, Harvard University, Cambridge, Massachusetts, United States of America

Published on: Oct 28, 2020

DOI: <https://doi.org/10.1162/99608f92.fed3dc89>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

As co-editors of the [Special Theme on Elections](#) for *Harvard Data Science Review*, we decided that it was important to incorporate both an academic and industry look at multiple aspects of the 2020 election, from a high-level discussion of different statistical and theoretical models by the most popular and well-known pollsters, to a study on how to estimate the magnitude of the lost votes by mail, and finally a different take all together on how pollsters might predict our elections in the future. We hope this issue will illuminate some of the pitfalls and possibilities that our political system holds.

At the very beginning of our editing process in January 2020, there were multiple pollsters taking part in this special theme. However, with COVID-19, and a due date of June 1, 2020 for an initial draft of their prediction model, we had a surge of dropouts—those unwilling, unable, or very understandably too busy with COVID-related work, to make a prediction with the uncertainty surrounding COVID-19. The world changed and so did our special theme. We do want to say though, that those who chose to submit their models and predictions clearly have an undeniable belief in their methodology and put their predictions on the line despite the unpredictably. We commend them for their dedication and decisiveness in their predictions. While that does limit the breadth of predictions, we believe that this issue certainly has range—with the goal of the predictions being as evidence-based, inclusive and diverse as possible.

Now down to the brass tacks.

Before a discussion on what the pollsters are predicting, we felt the need to investigate one very important variable that may affect the accuracy of predictions. Mail-in ballots are in the news this year: with images of [discarded mail ballots](#) peppering a ditch in Pennsylvania and the subsequent outcry and questions about the abilities of the Postal Service, we wanted to ask, what are the best scientifically-based estimates of lost votes by mail? Given COVID-19, the number of mail-in-ballots is only likely to increase, so how should pollsters think about this variable in their 2020 predictions, potentially unlike any previous election? Now, to be clear, a ‘lost vote’ occurs when a voter does all that is asked, but yet that vote is uncounted in the final tally. We turned to Professor Charles Stewart III of MIT, who has been following this issue closely and believes it could be far from the initial estimates of the Caltech/MIT Voting Technology Project, which estimated the magnitude of lost votes in the 2000 presidential election as between 4 and 6 million out of the 107 million cast that year.

From this discussion of how we can or cannot estimate lost votes, and the shocking limitations of data collected which greatly increase the uncertainty around such estimates, we begin to look at the larger pictures that this one variable fits in to: who will win the presidential election?

Øptimus, in collaboration with Decision Desk HQ, presents both a presidential and congressional modeling methodology from a data set of 200+ base features spanning everything from economic indicators to candidate traits, and finance reports. However, this is not a static model, the different indicators are paired for different races, and many models are combined for a final result. Their specialty lies in feature engineering various

political variables based upon their own political knowledge—variables that are impossible to replicate precisely by others. It can almost be best described from our reading as an ever-moving beast, not just from election to election, but from day to day.

As of writing this, Optimus is forecasting a Democratic win with a probability of 82%.

From an entirely different perspective, we approached an industry-academic collaboration of *The Economist* Magazine, represented by G. Elliott Morris, and Columbia University, represented by Andrew Gelman and Merlin Heidemanns. Their approach builds in fundamental predictors (US economic growth factors, presidential approval, polls, etc.), but they spend considerable amount of time to improve specific features within the model, such as state-level trends and non-response bias—aspects that many felt the 2016 pollster predictions did not take into account.

As of writing this, Heidemanns, Gelman, and Morris are forecasting a Democratic win with a probability of 86%.

From yet another perspective, a more qualitative approach than quantitative, we turned to Professor Allan Lichtman of American University. Allan has correctly predicted the popular vote outcome of the US Presidential Election since 1984, using his own “Keys to the White House” historical-based index system. The 13 Keys form a true/false criteria based on historical correlations with presidential elections from 1860-1980, using statistical methods adapted from the work of geophysicist Vladimir Keilis-Borok for prediction earthquakes. While Lichtman himself states that the Keys do not perform a regression equation or use “horse-race polls” or “presidential approval ratings,” the statistician among your co-editors would disagree. While there are no Excel spreadsheets being loaded into R, we perform regression all the time in our everyday lives—we measure how variables move in relation to each other, from everyday decisions to the White House. In effect, this is what Lichtman’s Keys are doing—and with seven negative Keys lined up against Donald Trump.

As of writing this, Lichtman is forecasting a Democratic win.

Finally, we turn to a piece that could give us insight into this election as well as food for thought in how we will view the predictions of future elections. We turned to Michael Isakov and Shiro Kuriwaki, respectively from Harvard College and Harvard University’s Department of Government¹, for an exploration of how the pollsters got it so wrong in 2016, and specifically what statistical methodology could be worked on, not only for this election but the prediction of future elections. They look at non-response bias and the pollster’s ability to correct systematic errors in weighting, the two factors that most believe are the reason that the pollsters got it

so wrong in 2016. They translate the 2016 errors into quantifiable measures, and then use these measures to remap recent state polls about the 2020 elections, under the assumptions that these factors have not changed much since 2016. Not surprisingly, under such assumptions, their analysis of the vote share in 12 battleground states narrowed the polling gaps between Biden and Trump, and more importantly, to the data scientists among us, increases twofold the margins of error in vote share estimates.

Since they focused on the 12 battleground states, their analysis does not make an overall prediction, but your guest editors' reading of this paper believe that their analysis effectively suggests a much closer race than anyone would think (we did our best to even get this level of prediction out of them...oh those statisticians and uncertainty!).

And with that, we bring you to our own personal thoughts. As statisticians and political scientists, we have been both watching the polls right along with you all and have had the benefit/strained eyes of many nights reading over pollster's models. For what it is worth, and to have in all perpetuity, we have decided to give you our (off-the-cuff, never-said-it-if-it-didn't-turn-out-the-way-we-think) opinions (the internet isn't forever, of course...).

Professor Ryan Enos, Department of Government at Harvard University, had his political analytics class make their own predictions. As of this writing, both he and his students have similar predictions in the popular vote with 46% to 54% Democratic win, but, interestingly, Enos thinks that the Democrats will sweep the floor with a major electoral college scoop up.

As of writing this, Ryan's 'seat of the pants' prediction is a Democratic win.

Professor Stephen Ansolabehere, also in the Department of Government at Harvard University, hearkens back to 2016, showing that while both Clinton's and Trump's polling numbers were behind their actual vote shares, this was because of the undecided voters in the last few weeks of the election not making up their minds until the last minute. Inevitably, Trump made bigger gains than Clinton did, winning almost two-thirds of these undecided voters. However, right now, Trump is again polling behind his 2016 vote share, and much further behind Biden than he was behind Clinton. Moreover, the number of undecided voters in this election is significantly smaller, and Trump leads far less on those voters than he did in 2016.

As of writing this, Stephen's 'cuff of his shirt' prediction is a Democratic win.

Our final guest editor, Professor Liberty Vittert, Washington University in St. Louis, Olin Business School, takes a rather unusual viewpoint, and backs up her original prediction of Trump in 2016 with a similar one. She agrees with Stephen’s analysis, but simply doesn’t think that pollsters have changed their ways enough, nor has the non-response bias (shy voters). Hence the undecided voters at the end will push it over, even while Trump had the highest unfavorability rating in history and lowest average approval ratings of any president.

As of writing this, Liberty’s ‘collar of her jacket’ prediction is that the most unpopular winner ever may very likely win again.

We hope you all enjoy reading the nooks and crannies of this special theme as much as we do, and we will keep you updated on social media with the changing of our pollsters—and our predictions—in the coming days leading up to the election.

Disclosure Statement

Liberty Vittert, Ryan D. Enos, and Stephen Ansolabehere have no financial or non-financial disclosures to share for this article.

©2020 Liberty Vittert, Ryan D. Enos, and Stephen Ansolabehere. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](#), except where otherwise indicated with respect to particular material included in the article.

Footnotes

1. The Editor-in-Chief of HDSR, Xiao-Li Meng, was involved in an advisory capacity on Isakov and Kuriwaki’s article, “Principled Unskewing: Viewing 2020 Election Polls Through a Corrective Lens from 2016.” Due to this, the peer review process for this paper was entirely independent of the Editor-in-Chief. [↩](#)