



Origins of Presidential poll aggregation: A perspective from 2004 to 2012



Samuel S.-H. Wang*

Princeton Neuroscience Institute and Department of Molecular Biology, Princeton University, Princeton, NJ 08544, United States

ARTICLE INFO

Keywords:

Poll aggregation
Bayesian estimation
Opinion polls
Election forecasting
Election snapshots

ABSTRACT

US political reporting has become extraordinarily rich in polling data. However, this increase in information availability has not been matched by an improvement in the accuracy of poll-based news stories, which usually examine a single survey at a time, rather than providing an aggregated, more accurate view. In 2004, I developed a meta-analysis that reduced the polling noise for the Presidential race by reducing all available state polls to a snapshot at a single time, known as the Electoral Vote estimator. Assuming that Presidential pollsters are accurate in the aggregate, the snapshot has an accuracy equivalent to less than $\pm 0.5\%$ in the national popular-vote margin. The estimator outperforms both the aggregator FiveThirtyEight and the betting market InTrade. Complex models, which adjust individual polls and employ pre-campaign “fundamental” variables, improve the accuracy in individual states but provide little or no advantage in overall performance, while at the same time reducing transparency. A polls-only snapshot can also identify shifts in the race, with a time resolution of a single day, thus assisting in the identification of discrete events that influence a race. Finally, starting at around Memorial Day, variations in the polling snapshot over time are sufficient to enable the production of a high-quality, random-drift-based prediction without a need for the fundamentals that are traditionally used by political science models. In summary, the use of polls by themselves can capture the detailed dynamics of Presidential races and make predictions. Taken together, these qualities make the meta-analysis a sensitive indicator of the ups and downs of a national campaign—in short, a precise electoral thermometer.

© 2015 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

In 2012, polling aggregation entered the public spotlight as never before. Typically, political horserace commentaries in the US are dominated by pundits who are motivated by pressure, not to be accurate, but to attract readers and viewers. For example, one day before the

2012 U.S. presidential election, former Reagan speechwriter Noonan (2012) wrote that “nobody knows anything” about who would win, asserting that Republican candidate Mitt Romney’s supporters had the greater passion and enthusiasm, while columnist George Will predicted a Romney electoral landslide (Poor, 2012).

In the end, the aggregators were correct. The pundits largely failed to report the fact that, according to public opinion polls with collectively excellent track records, President Obama had an advantage of three to four percentage points for nearly the entire campaign season. Ignoring the data, many commentators expressed confidence—and were wrong.

* Correspondence to: Princeton Neuroscience Institute, Washington Road, Princeton University, Princeton, NJ 08544, United States. Tel.: +1 609 258 0388; fax: +1 609 258 1028.

E-mail address: sswang@princeton.edu.

<http://dx.doi.org/10.1016/j.ijforecast.2015.01.003>

0169-2070/© 2015 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

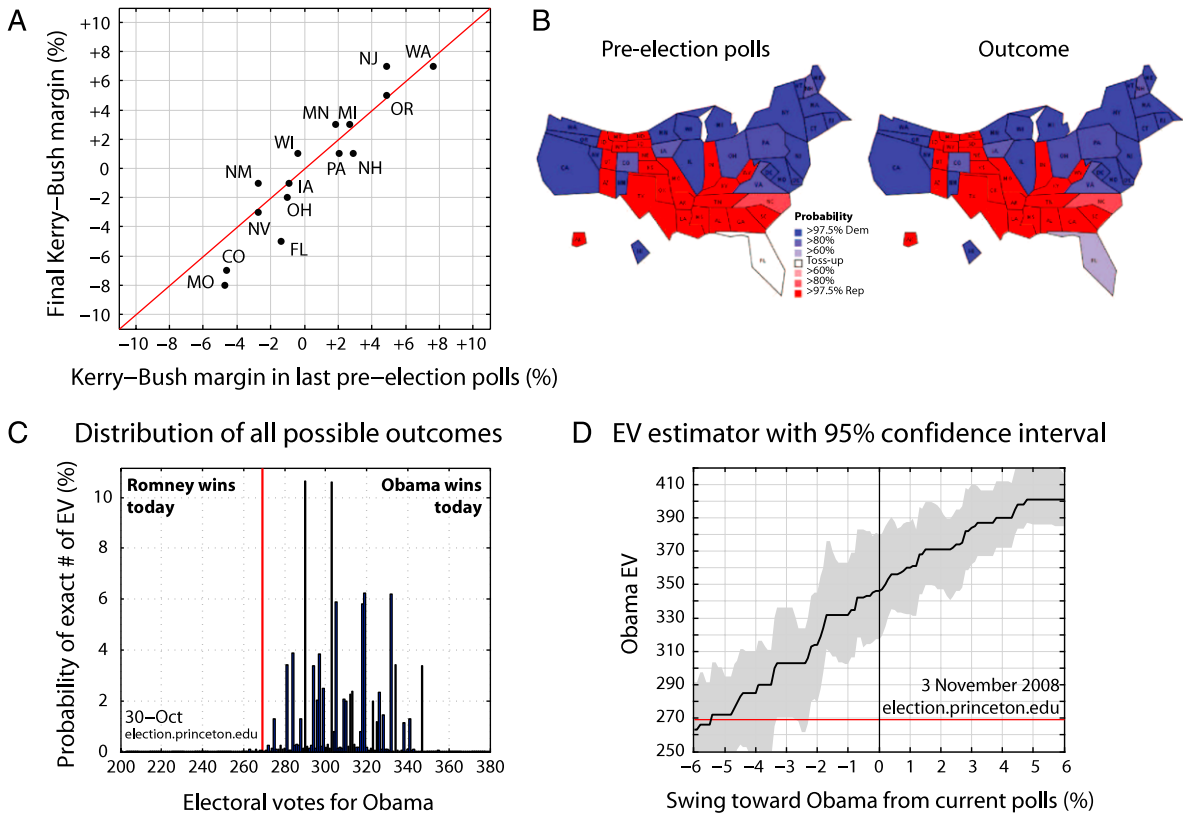


Fig. 1. Foundations of the Presidential meta-analysis. (a) State-by-state election margins as a function of final pre-election polls in the 2004 Kerry vs. Bush race. (b) Pre-election win probabilities and actual outcomes in the 2012 Obama vs. Romney race. (c) A snapshot of the exact distribution of all $2^{51} = 2.3$ quadrillion outcomes calculated from the win probabilities in (b). The electoral vote estimator is defined as the median of the distribution. (d) Electoral effect of a uniform shift in state polls through a constant swing. The gray band indicates a nominal 95% confidence interval, including uncorrected pollster-to-pollster variation.

In this article, I describe an early approach to the aggregation of Presidential state polls, the meta-analytic method, which has been being used at the Princeton Election Consortium (PEC; <http://election.princeton.edu>) since 2004. PEC’s approach uses Electoral College mechanisms and can be updated on a daily basis. Its only input is publicly available data, and it runs on open-source software, thus providing a high level of transparency. I will describe this method, and give both public and academic perspectives (see also Jones, 2008, for a review). I provide both an academic account and a history, under the assumption that the evolution of the meta-analysis may interest some readers.

Polling aggregators have been outperforming pundits since at least 2004, when a number of websites began to collect and report polls on a state-by-state basis in Presidential, Senate, and House races. State polls are of particular interest for the Presidency, for three reasons. First, the Presidency is determined via the Electoral College, which is driven by state election win-lose outcomes. Second, state polls have the advantage of being accurate predictors of state election outcomes, on average (Fig. 1(a)), though national polls can have significant inaccuracies. For example, in 2000, Al Gore won the popular vote over George W. Bush by 0.5%, yet election-eve national polls favored Bush by an average of 2.5%, a 3.0% error that got the sign

of the outcome wrong. State polls may owe their superior accuracy levels to the fact that local populations are less complex demographically, and therefore easier to sample, than the nation as a whole. Third and last, state presidential polls are also remarkably abundant: Electoral-vote.com contains the results of 879 polls from 2004, 1189 from 2008, and 924 from 2012.

Early sites—RealClearPolitics in 2002, followed in 2004 by Andrew Tanenbaum’s Electoral-vote.com, the Princeton Election Consortium, and several others (Forelle, 2004a)—reported average or median polling margins (i.e., the percentage difference in support between the two leading candidates) for individual races. An additional step was taken by PEC (then titled “Electoral college meta-analysis”, <http://synapse.princeton.edu/~sam/pollcalc.html>), which calculated the electoral vote (EV) distribution of all possible outcomes, using polls to provide a simple tracking index, the EV estimator. The calculation, an estimate of the EV outcome for the Kerry vs. Bush race, was updated in a low-graphics, hand-coded HTML webpage, together with a publicly posted MATLAB script. PEC gained a following among natural scientists, political and social scientists, and financial analysts. Over the course of the 2004 campaign, PEC attracted over a million visits, and the median decided-voter calculation on election eve captured the exact final outcome (Forelle, 2004b).

Download English Version:

<https://daneshyari.com/en/article/7408404>

Download Persian Version:

<https://daneshyari.com/article/7408404>

[Daneshyari.com](https://daneshyari.com)