Debia$ed Aggregated Poll$ and Prediction Market Price$

David Rothschild

D o your hopes and fears over an upcoming election shift with each new poll released? If so, consider Figure 1 the antidote to your daily roller coaster. Other than being an enjoyable pastime for many people, accurately documenting the underlying probability of victory at any given day before an election is critical for enabling academics to determine the impact of shocks to campaigns. Further, the more accurate and real-time the probabilities, the more efficient the public can be in choosing how to invest their time and money (in terms of having the greatest marginal impact for them and/or their cause) and the more efficient political organizations and campaigns can be in spending those resources.

Shifts in the underlying probability of victory cannot justify the volatility illustrated by the daily poll-based probability in Figure 1, which is standard in the previous literature. Aggregated poll-based probabilities provide a more realistic progression of probabilities near the trend of the daily movement. As a practi-

cal implication, daily polls are so volatile it is hard to grasp anything else on a chart that includes their probabilities.

The aggregated poll-based probability is derived with coefficients from a linear regression of results on the daily poll snapshot in past election cycles. The poll snapshot is the linear trend of all the polls released up to that day. It is projected onto Election Day by regressing the final vote share on the poll snapshot for each day before the election in previous election cycles (2000 and 2004 data are used for the presidential races): $V_{yr} = \alpha + \beta P_{yr} + e_{yr}$, where y is a given year and r is a given race. The daily projections for 2008 are created using the unique alpha and beta derived for each day before the election (T): $\hat{V}_{2008.T} = \alpha_T + \beta_T P_{2008.T}$. The alpha corrects for anti-incumbency bias (i.e., the depression of incumbent party numbers in the late summer/early fall), and the beta corrects for reversion to the mean (i.e., elections narrow as Election Day approaches). Both α_T and β_T are statistically significant, confirming that debiasing polls does make more accurate forecasts. A probability of victory is calculated by assuming the projections are the mean of a normal distribution. After assigning a variance based on the historic accuracy of the projections on a given day, one can determine the percentage of possible outcomes in which each candidate has the higher number of votes.

Do you debate (or bet) with your friends and family the probability of a candidate winning an upcoming election and how that probability shifts after a controversial vice presidential choice or televised debate? The chart in Figure 2 is the same as Figure 1, but excludes the daily poll-based probability and adds prediction marked–based forecasts and

annotations of the major events from the election cycle.

The prediction market prices are from Intrade (*www.Intrade.com*). Intrade trades binary options that pay, for example, $10 if the chosen candidate wins and $0 otherwise. Thus, if there are no transaction costs, an investor who pays $6 for a "democrat to win" stock and holds the stock through Election Day, earns $4 if the Democrat wins and loses $6 if the Democrat loses. In that scenario, the investor should be willing to pay up to the price equaling the estimated probability of the Democrat winning the election.

Figure 2 illustrates why prediction market prices need to be debiased to correct for the favorite-longshot bias (i.e., the restriction of market prices as they approach 0 and 1 due to transaction costs, liquidation concerns, and nonrisk neutral investing), where the previous literature only debiased polls. This bias is a major problem for the accuracy of prediction markets, as 44 of 51 Electoral College races had consensus predictions over 90% for one of the candidates on the eve of the 2008 election.

The cost of not debiasing is seen in the difference between the prediction market price and the debiased prediction market-based probability as Obama's victory becomes increasingly certain. The favorite-longshot bias is ameliorated with the transformation suggested by Andrew Leigh in "Is There a Favorite-Longshot Bias in Election Markets?" presented at the 2007 University of California, Riverside Conference on Prediction Markets, which is estimated (and suggested) prior to my sample using data from presidential predication markets from 1880 to 2004: $Pr = \Phi (1.64 * \Phi^{-1}(price))$, where Φ is the cumulative normal distribution

function. The transformation converts price to a normal Z-score (standardized to mean 0 and variance 1), multiples by 1.64 (thereby inflating the value), and then computes a probability.

Figure 2 also demonstrates why extending forecast research to state-level races is essential to gathering the data necessary to determine some causality or, at minimum, a fuller description of correlations between events and electoral outcomes.

First, the figure shows that while there is strong correlation between the polling and prediction market-based forecasts, there is still considerable variation at points during the cycle. Both sets of probabilities have Republican candidate John McCain moving up after the Republican National Convention and the announcement of Sarah Palin as his running mate, but only the market has him crossing the 50% threshold (i.e., predicting he wins the election).

Second, even if there were a consensus on the underlying national values, it is impossible to determine causality of events on outcomes using national data calibrated daily; there are too many overlapping events in too few races.

Finally, there is evidence that the national popular vote prediction markets may suffer from manipulation by people motivated to gain publicity for their chosen candidate, but this evidence does not extend to the state-level markets. Moreover, the national popular vote does not determine the winner of the U.S. presidential election, since the election outcome hinges on the results in 51 individual sovereignties through the Electoral College. **C**

Further Reading

Erikson, Robert S., and Christopher Wlezien. 2008. Are political markets really superior to polls as election predictors? *Public Opinion Quarterly* 72:190-21.

Leigh, Andrew, Justin Wolfers, and Eric Zitzewitz. 2007. Is there a favorite-longshot bias in election markets? Paper presented at the 2007 UC Riverside Conference on Prediction Markets, Riverside, California.

Rothschild, David. 2009. Forecasting elections: Comparing prediction markets, polls, and their biases. *Public Opinion Quarterly* 73:895-916.
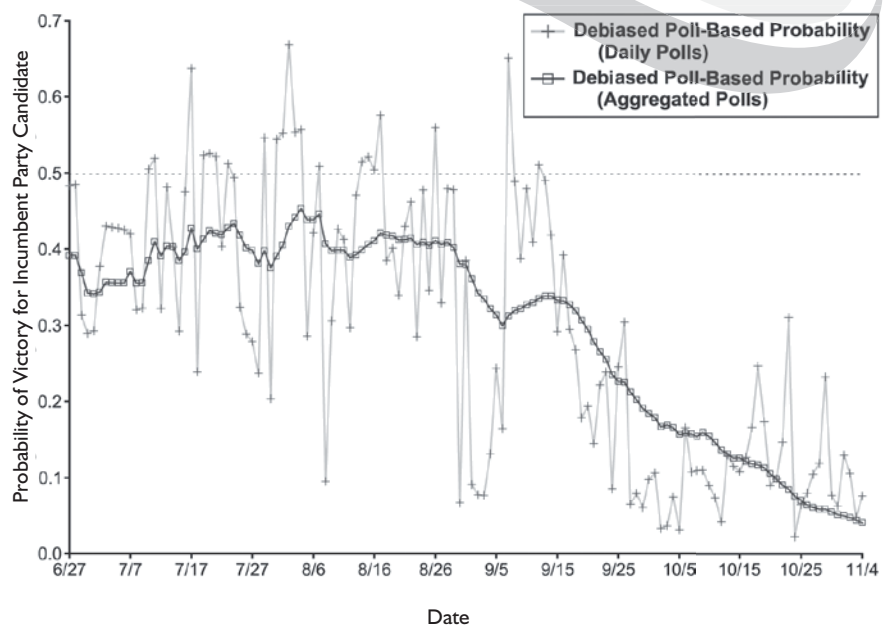

Figure 1. Probability of victory in the national popular vote for the incumbent party candidate (John McCain) in the 2008 presidential election based on debiased daily and aggregated polls
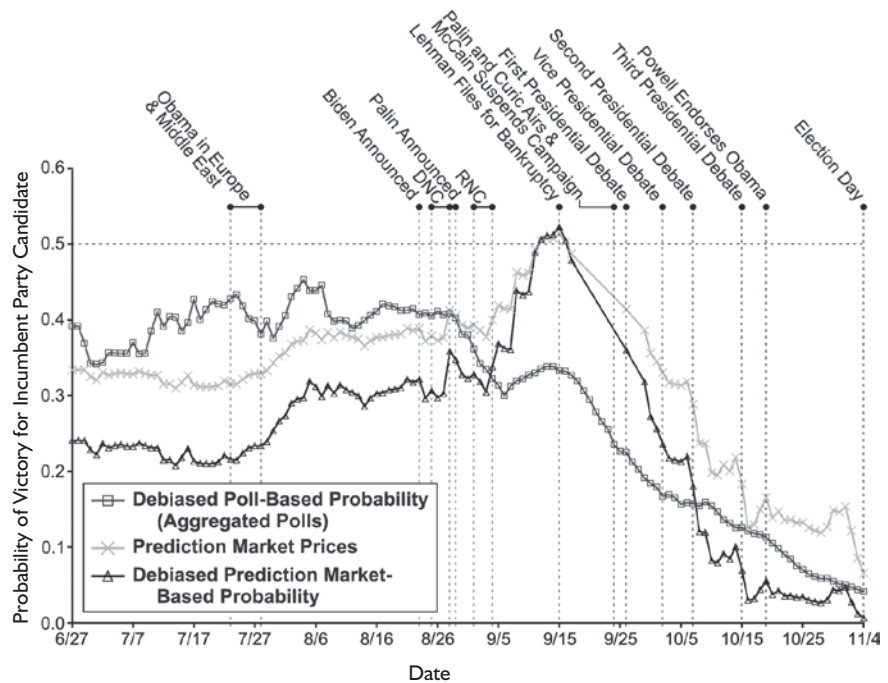

Figure 2. Probability of victory in the national popular vote for the incumbent party candidate (John McCain) in the 2008 presidential election based on debiased aggregated polls, prediction market prices, and debiased prediction market prices