Forecasting the 2012 Presidential Election from History and the Polls

Drew Linzer

Assistant Professor
Emory University
Department of Political Science

Visiting Assistant Professor, 2012-13
Stanford University
Center on Democracy, Development,
and the Rule of Law

votamatic.org

Bay Area R Users Group
February 12, 2013

The 2012 Presidential Election: *Obama 332–Romney 206*

The 2012 Presidential Election: *Obama 332–Romney 206*

But also: *Nerds 1–Pundits 0*

The 2012 Presidential Election: *Obama 332–Romney 206*

But also: *Nerds 1–Pundits 0*

Analyst forecasts based on history and the polls

Drew Linzer, Emory University	332-206
Simon Jackman, Stanford University	332-206
Josh Putnam, Davidson College	332-206
Nate Silver, New York Times	332-206
Sam Wang, Princeton University	303-235

Pundit forecasts based on intuition and gut instinct

Karl Rove, Fox News	259-279
Newt Gingrich, Republican politician	223-315
Michael Barone, Washington Examiner	223-315
George Will, Washington Post	217-321
Steve Forbes, Forbes Magazine	217-321

What we want: Accurate forecasts as early as possible

The problem:

- The data that are available early aren't accurate:
  Fundamental variables (economy, approval, incumbency)
- The data that are accurate aren't available early:
  Late-campaign state-level public opinion polls
- The polls contain sampling error, house effects, and most states aren't even polled on most days

What we want: Accurate forecasts as early as possible

The problem:

- The data that are available early aren't accurate:
  Fundamental variables (economy, approval, incumbency)
- The data that are accurate aren't available early:
  Late-campaign state-level public opinion polls
- The polls contain sampling error, house effects, and most states aren't even polled on most days

The solution:

- A statistical model that uses what we know about presidential campaigns to update forecasts from the polls in real time

What we want: Accurate forecasts as early as possible

The problem:

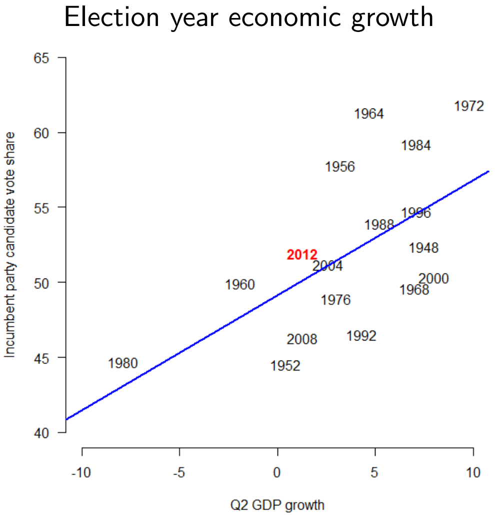
- The data that are available early aren't accurate:
  Fundamental variables (economy, approval, incumbency)
- The data that are accurate aren't available early:
  Late-campaign state-level public opinion polls
- The polls contain sampling error, house effects, and most states aren't even polled on most days

The solution:

- A statistical model that uses what we know about presidential campaigns to update forecasts from the polls in real time

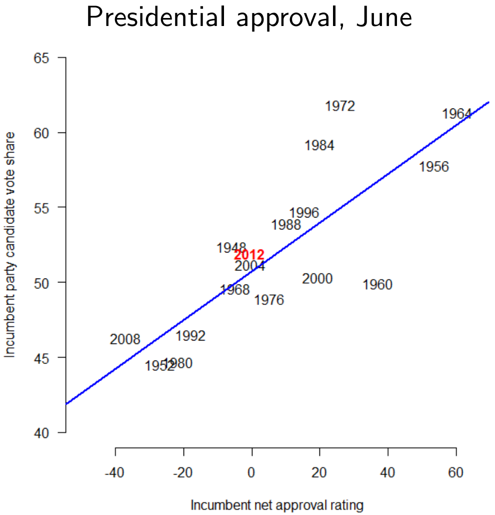
What do we know?

1. The *fundamentals* predict national outcomes, noisily



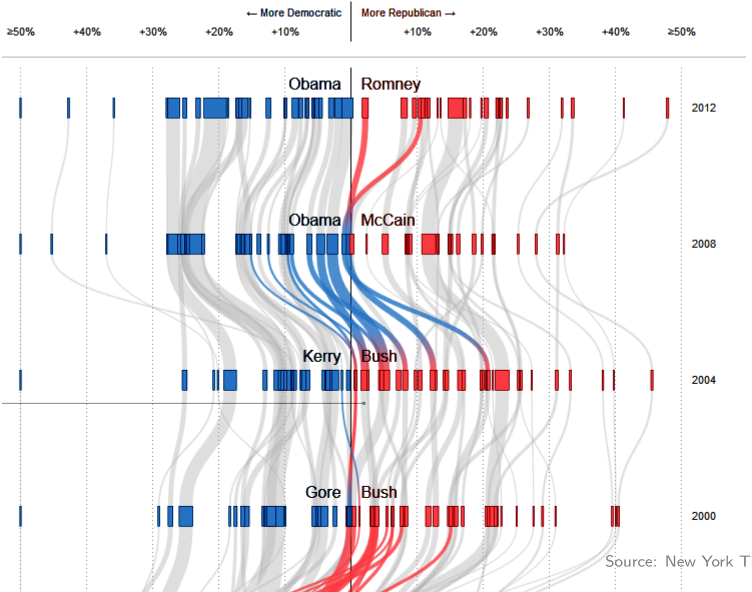
Source: U.S. Bureau of Economic Analysis

1. The *fundamentals* predict national outcomes, noisily



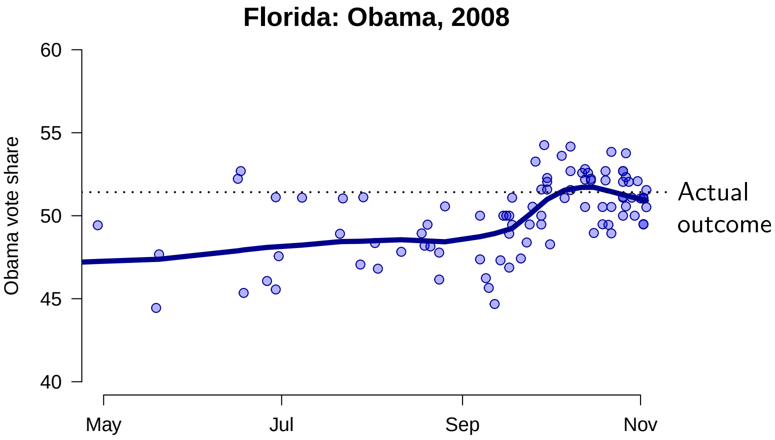
Source: Gallup

2. States vote outcomes swing (mostly) in tandem



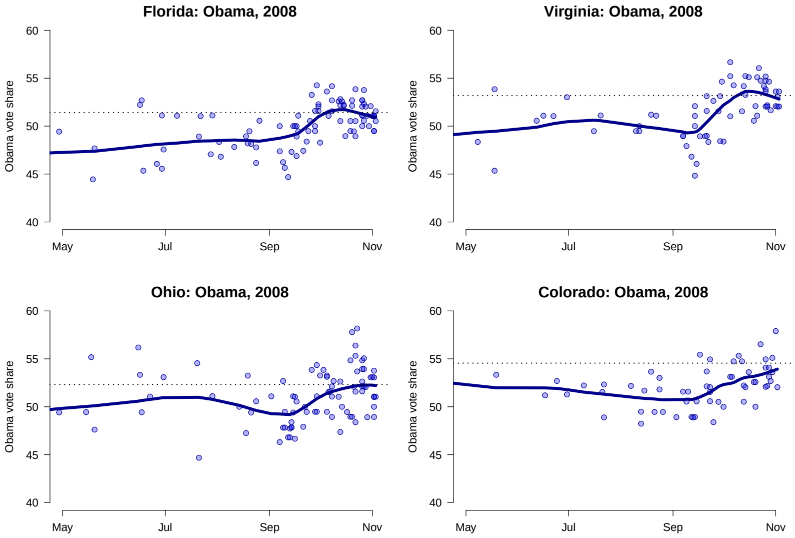
Source: New York Times

3. Polls are accurate on Election Day; maybe not before



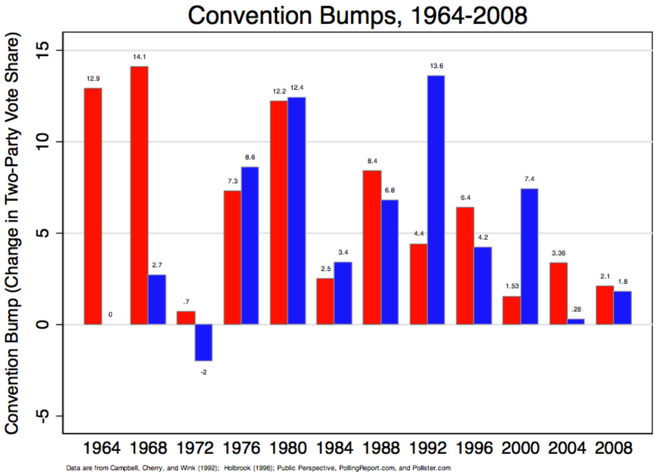
Source: HuffPost-Pollster

4. Voter preferences evolve in similar ways across states



Source: HuffPost-Pollster

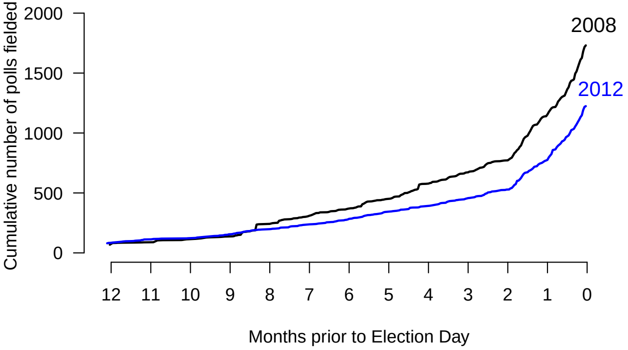
5. Voters have short term reactions to big campaign events



Source: Tom Holbrook, UW-Milwaukee

All together: A forecasting model that learns from the polls

Publicly available state polls during the campaign



Forecasts weight fundamentals $\longleftrightarrow$ Forecasts weight polls

Source: HuffPost-Pollster

First, create a baseline forecast of each state outcome

Abramowitz *Time-for-Change* regression makes a *national* forecast:

$$\begin{aligned} \text{Incumbent vote share} &= 51.5 + 0.6\ \text{Q2 GDP growth} \\ &\quad + 0.1\ \text{June net approval} \\ &\quad - 4.3\ \text{In office two+ terms} \end{aligned}$$

First, create a baseline forecast of each state outcome

Abramowitz *Time-for-Change* regression makes a *national* forecast:

$$\begin{aligned}\text{Incumbent vote share} &= 51.5 + 0.6\ \text{Q2 GDP growth} \\ &\quad + 0.1\ \text{June net approval} \\ &\quad - 4.3\ \text{In office two+ terms}\end{aligned}$$

$$\begin{aligned}\text{Predicted Obama 2012 vote} &= 51.5 + 0.6\ (1.3) \\ &\quad + 0.1\ (\text{-}0.8) \\ &\quad - 4.3\ (0)\end{aligned}$$

First, create a baseline forecast of each state outcome

Abramowitz *Time-for-Change* regression makes a *national* forecast:

$$\begin{aligned}\text{Incumbent vote share} &= 51.5 + 0.6\ \text{Q2 GDP growth} \\ &\quad + 0.1\ \text{June net approval} \\ &\quad - 4.3\ \text{In office two+ terms}\end{aligned}$$

$$\begin{aligned}\text{Predicted Obama 2012 vote} &= 51.5 + 0.6\ (1.3) \\ &\quad + 0.1\ (\text{-}0.8) \\ &\quad - 4.3\ (0)\end{aligned}$$

$$\text{Predicted Obama 2012 vote} = \mathbf{52.2\%}$$

First, create a baseline forecast of each state outcome

Abramowitz *Time-for-Change* regression makes a *national* forecast:

$$\begin{aligned} \text{Incumbent vote share} &= 51.5 + 0.6\ \text{Q2 GDP growth} \\ &\quad + 0.1\ \text{June net approval} \\ &\quad - 4.3\ \text{In office two+ terms} \end{aligned}$$

$$\begin{aligned} \text{Predicted Obama 2012 vote} &= 51.5 + 0.6\ (1.3) \\ &\quad + 0.1\ (\text{-}0.8) \\ &\quad - 4.3\ (0) \end{aligned}$$

$$\text{Predicted Obama 2012 vote} = \mathbf{52.2\%}$$

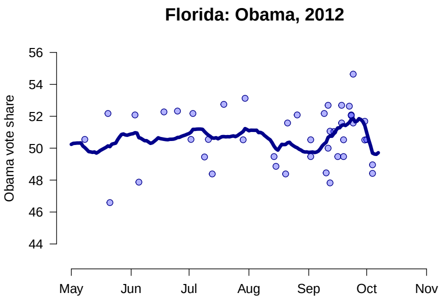
Use *uniform swing* assumption to translate to the state level:

Subtract 1.5% for Obama from his 2008 state vote shares

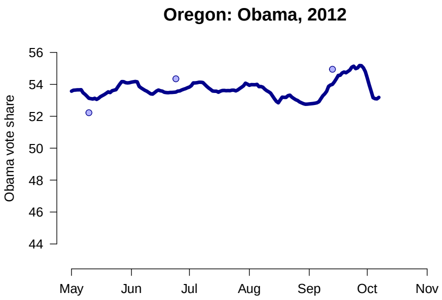
Make this a Bayesian prior over the final state outcomes

Combine polls across days *and* states to estimate trends

States with many polls

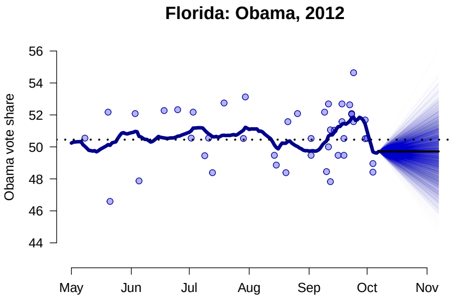


States with fewer polls



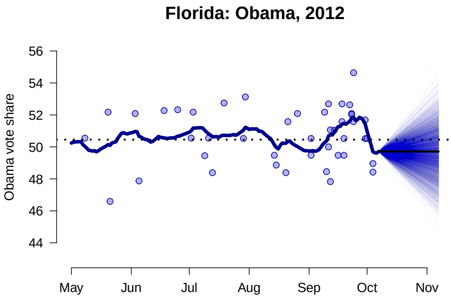
Combine with baseline forecasts to guide future projections

Random walk (no)

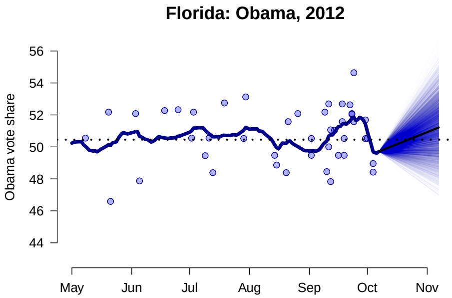


Combine with baseline forecasts to guide future projections

Random walk (no)



Mean reversion



Forecasts compromise between history and the polls

Model specification

$y_k \sim Binomial(\pi_{i[k]j[k]}, n_k)$	Number of people preferring Democrat in survey k, in state i, on day j
$\pi_{ij} = logit^{-1}(\beta_{ij} + \delta_j)$	Proportion reporting support for the Democrat in state i on day j

National effects: δ_j
State components: β_{ij}
Election forecasts: $\hat{\pi}_{iJ}$

Priors

$\beta_{iJ} \sim N(logit(h_i), \tau_i)$	Informative prior on Election Day, using historical predictions h_i, precisions τ_i
$\delta_J \equiv 0$	Polls assumed accurate, on average
$\beta_{ij} \sim N(\beta_{i(j+1)}, \sigma^2_\beta)$	Reverse random walk, states
$\delta_j \sim N(\delta_{(j+1)}, \sigma^2_\delta)$	Reverse random walk, national

Model specification

$y_k \sim Binomial(\pi_{i[k]j[k]}, n_k)$ — Number of people preferring Democrat in survey k, in state i, on day j

$\pi_{ij} = logit^{-1}(\beta_{ij} + \delta_j)$ — Proportion reporting support for the Democrat in state i on day j

National effects: δ_j

State components: β_{ij}

Election forecasts: $\hat{\pi}_{iJ}$

Estimated for all states simultaneously

Priors

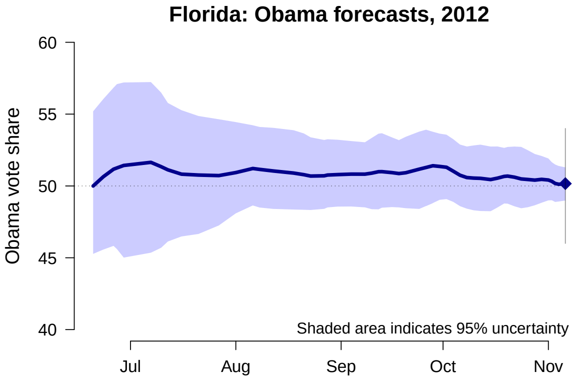
$\beta_{iJ} \sim N(logit(h_i), \tau_i)$ — Informative prior on Election Day, using historical predictions h_i, precisions τ_i

$\delta_J \equiv 0$ — Polls assumed accurate, on average

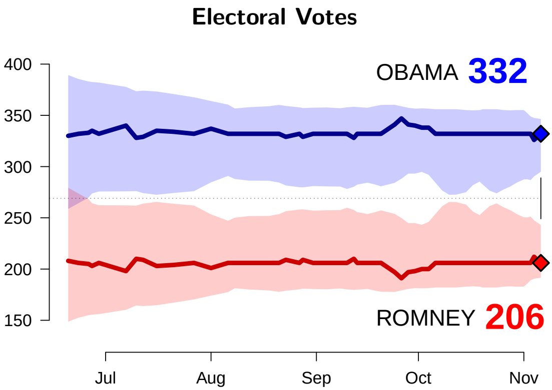
$\beta_{ij} \sim N(\beta_{i(j+1)}, \sigma_\beta^2)$ — Reverse random walk, states

$\delta_j \sim N(\delta_{(j+1)}, \sigma_\delta^2)$ — Reverse random walk, national

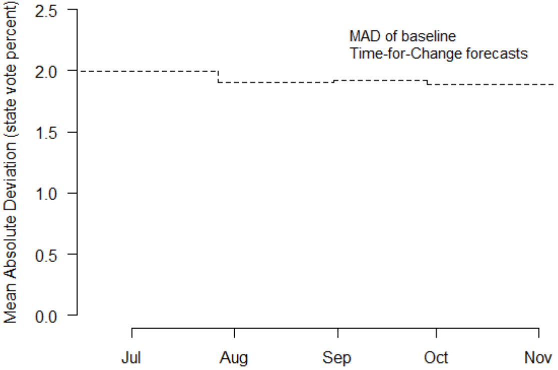
Results: Anchoring to the fundamentals stabilizes forecasts



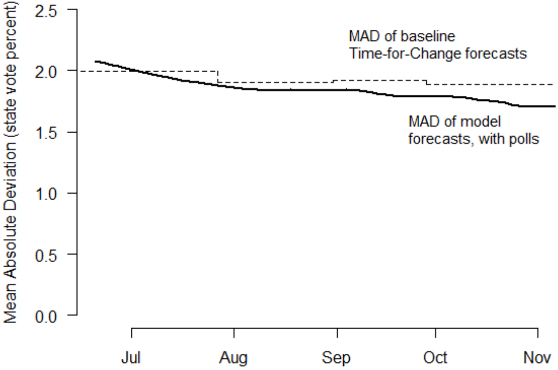
Results: Anchoring to the fundamentals stabilizes forecasts



There were almost no surprises in 2012

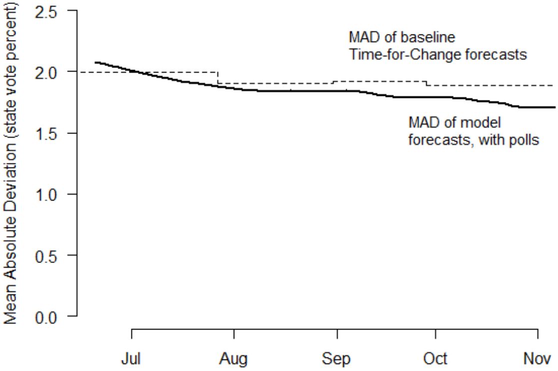


There were almost no surprises in 2012



On Election Day, average error = 1.7%

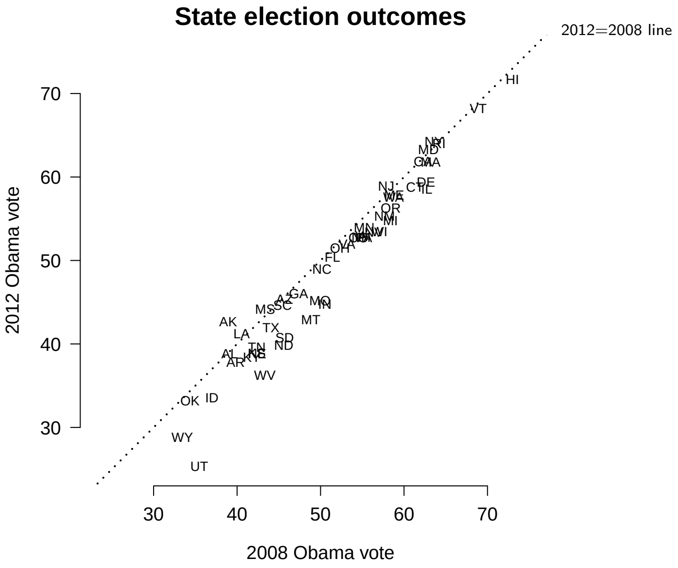
There were almost no surprises in 2012



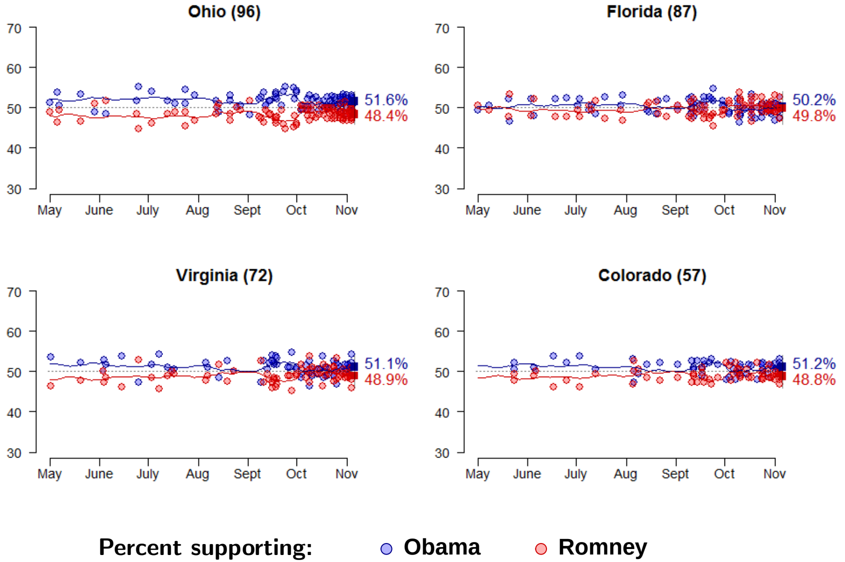
On Election Day, average error = 1.7%

Why didn't the model improve forecasts by more?

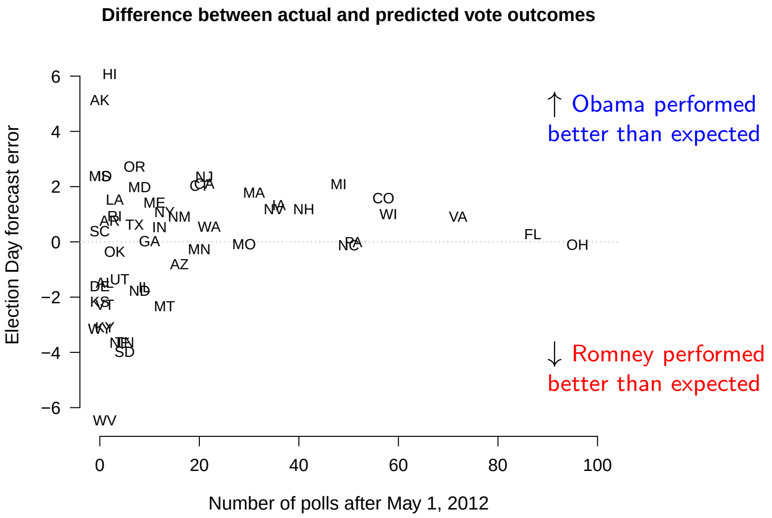
The fundamentals and uniform swing were right on target



Aggregate preferences were *very* stable



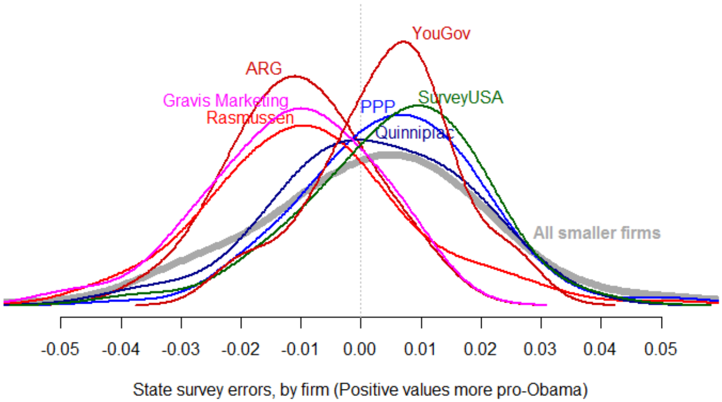
Could the model have done better? *Yes*



Forecasting is only one of many applications for the model

1. Who's going to win?
2. Which states are going to be competitive?
3. What are current voter preferences in each state?
4. How much does opinion fluctuate during a campaign?
5. What effect does campaign news/activity have on opinion?
6. Are changes in preferences primarily national or local?
7. How useful are historical factors vs. polls for forecasting?
8. How early can accurate forecasts be made?
9. Were some survey firms biased in one direction or the other?

House effects (biases) were evident during the campaign



Much more at votamatic.org

✉ drew@votamatic.org
@DrewLinzer