

The 2016 Presidential Election

(He isn't going to win, right?)

Benjamin E Lauderdale

London School of Economics and Political Science

7 November 2016

Introduction

Three things I am going to talk about today:

1. Methodology for analysing pre-election polling
2. If you only look at the microdata, this is a pretty boring US presidential election
 - Nearly all Romney voters are voting for Trump and nearly all Obama voters are voting for Clinton.
 - There has been essentially no movement since the first debate.
 - Versus 2012, there are increasing associations of voting with gender and education, and some slight realignments of the electoral geography.
3. Trump probably will not win tomorrow, but it is not *that* unlikely.
 - The national vote lead for Clinton looks to be similar to the 2012 election margin, which was Obama by 3.9%.
 - However, polls can easily be off by that much, and unlike Romney in 2012, Trump has an electoral college advantage.

Collaboration

- This talk is based on a collaboration with Doug Rivers (YouGov/Stanford).
 - Delia Bailey (YouGov) and Jack Blumenau (LSE) have also worked extensively on the project.
- To see the estimates updated daily (today and tomorrow), visit <https://today.yougov.com/us-election/>

A Useful Decomposition

- A 'voter type' is a set of measurable characteristics for an eligible voter X_i (age, gender, education, who they voted in 2012, where they live, etc). For each voter type, there are three important quantities that we would like to know. . .
 1. $p(V_i|X_i, T_i = 1)$ What proportion of each type will vote for Clinton, Trump, Johnson, Stein, etc, among those who do vote?
 2. $p(T_i = 1|X_i)$ What proportion of each voter type will turn out to vote?
 3. $p(X_i)$ What proportion of the electorate is of each type?
- The proportion of the electorate turning out to vote is:

$$\sum_i p(T_i = 1|X_i)p(X_i)$$

- The proportion of the electorate voting for each candidate is:

$$\sum_i p(V_i|X_i, T_i = 1)p(T_i = 1|X_i)p(X_i)$$

Estimating $p(V_i|X_i, T_i = 1)$

What proportion of each type will vote for Clinton, Trump, Johnson, Stein, etc, among those who do vote?

- Key assumption: we can model voter preferences in a large, online, opt-in panel so as to get the conditional vote choices of different types (close enough to) correct.
- Data
 - YouGov US panel, 5-10k responses, per day, nationally. 14 day data window, with time trends modeled.
 - Key variables: congressional district, state, race, age, education, marital status, gender, and 2012 vote.
- Model
 - Multilevel (district-level within state-level) multinomial logistic regression with many interactions.
- Potential problems:
 - If the YouGov panelists have different preferences than the public, conditional on everything in the model.
 - If the 2012 vote measures for the panelists are biased, we will have big problems, but we have validated extensively, see:
<https://today.yougov.com/news/2016/11/01/beware-phantom-swings-why-dramatic-swings-in-the-p/>

Estimating $p(T_i = 1|X_i)$

What proportion of each voter type will turn out to vote?

- Key assumption: we cannot reliably predict changes in relative turnout of different groups. Given historical stability of turnout, it is better to have a high quality estimate of 2012 turnout patterns than a low quality estimate of 2016 turnout patterns.
- Data
 - 2012 Current Population Survey (CPS) is the largest (~100k) available high-quality probability sample with 2012 turnout as a function of demographics.
 - We impute 2008 turnout onto the 2012 CPS using state voter files.
- Model
 - Multilevel (state-level) logistic regression with many interactions.
- Potential problems:
 - If CPS self-reported turnout is more problematic than we think it is.
 - If turnout patterns are very different than in 2012.

Estimating $p(X_i)$

What proportion of the electorate is of each type?

- Data
 - 2010 US Census 1% sample provides ~2m obs with joint distribution of *most* variables.
 - We update the marginal distributions using the 2015 American Community Survey (ACS) via raking.
 - We impute registration onto this using current voter files.
 - We impute 2012 vote onto this using YouGov's 2012 polling data subject to constraint of actual election results.
- Potential problems:
 - Some potential for imputations to cause problems.
 - If YouGov 2012 vote data is biased conditional on demographics (it was not nationally, but could be in sub-groups).

Putting it all together with multilevel regression and poststratification

- We take a multilevel regression and poststratification (MRP) approach to estimating each of the three components and combining them.
 - Bayesian implementation implemented in Stan (<http://mc-stan.org>)
 - 37 parallel MCMC chains hosted on AWS
 - About 8 hours estimation time
 - Data pulled each day around 8pm Eastern time in US, estimates posted by 8am following morning

Should you trust this kind of analysis?

- We used the same setup for the 2016 EU referendum in the UK
 - We never estimated Remain ahead.
 - Final estimate was Leave 50.6% versus the 51.9% actual result.
 - Local authority (reporting area) estimates were correlated with results at 0.92, RMSE of 4.1 percentage points.
 - See <http://www.benjaminlauderdale.net/wp/?p=210>
- We retroactively predicted the 2012 US presidential election
 - We did not do this until the model was done (we did not cheat by training the model on the data).
 - Final wave of YouGov data from early November 2012 yielded estimate of 4.1% Obama national margin (versus 3.9% actual) and every state correctly predicted except NC.
 - We then estimated the degree of under-coverage of state-level results and feed that into the 2016 model as additional non-sampling uncertainty.

2012-2016 Transition Matrix

2012 Vote	Clinton	Trump	Johnson	Stein	Other	% of Voters
Obama	88	7	2	2	1	45
Romney	5	87	5	0	3	40
Other	17	34	28	13	8	2
None	41	48	6	3	2	9
Under 18	64	20	10	4	3	5

- Nearly 90% of Obama voters will vote for Clinton and nearly 90% of Romney voters will vote for Trump.
 - About 1.5x as many switchers as in 2012 vs 2008; still not many though.
- Clinton loses about as much among 2012 voters and non-voters as she gains because of new 18-21s.

Transitions by Education

2012 Vote	Education	Clinton	Trump	Johnson	Stein	Other
Obama	High School	85	10	2	2	1
Obama	Some College	86	8	2	3	1
Obama	College	91	5	2	2	1
Obama	Post Graduate	93	3	1	2	1
Romney	High School	4	92	3	0	1
Romney	Some College	4	89	4	0	2
Romney	College	6	83	7	0	4
Romney	Post Graduate	8	81	7	0	4

- Clinton is retaining the high education Obama vote better than the low education Obama vote.
- Trump is retaining the low education Romney vote better than the high education Romney vote.

Transitions by Race

2012 Vote	Race	Clinton	Trump	Johnson	Stein	Other
Obama	White	86	8	2	3	1
Obama	Black	94	3	1	2	1
Obama	Hispanic	87	7	2	4	1
Obama	Other	84	8	2	4	2
Romney	White	4	88	5	0	3
Romney	Black	18	75	3	1	3
Romney	Hispanic	12	79	6	1	2
Romney	Other	6	84	6	1	4

- Less stark than you might think, but remember that Black and Hispanic Romney voters were pretty committed to being Republicans.

Transitions by Gender

2012 Vote	Gender	Clinton	Trump	Johnson	Stein	Other
Obama	Male	86	8	2	3	1
Obama	Female	90	6	1	2	1
Romney	Male	4	88	5	0	3
Romney	Female	6	87	5	0	3

- Clinton is retaining the female Obama vote better than the male Obama vote.
- Trump is retaining the male Romney vote better than the female Romney vote.

Worst states for Democratic retention

2012 Vote	State	Clinton	Trump	Johnson	Stein	Other
Obama	WV	80	14	2	3	1
Obama	SD	82	11	4	2	1
Obama	MT	82	9	4	4	1
Obama	HI	83	9	2	4	2
Obama	IN	84	10	3	1	2
Obama	ME	84	9	2	4	1
Obama	NE	85	9	3	2	1
Obama	WY	85	9	2	3	1
Obama	DE	86	9	2	3	1
Obama	KY	86	9	2	3	1
Obama	NV	86	9	2	2	1
Obama	OH	86	9	2	2	1
Obama	OK	85	8	3	2	1
Obama	PA	87	10	1	2	1
Obama	RI	85	8	3	3	1

Best states for Democratic retention

2012 Vote	State	Clinton	Trump	Johnson	Stein	Other
Obama	IL	89	6	2	2	1
Obama	MN	89	6	2	2	1
Obama	NH	89	6	2	3	1
Obama	SC	89	6	1	2	1
Obama	TX	89	6	2	2	1
Obama	UT	87	4	3	3	3
Obama	WA	88	5	2	3	1
Obama	VA	90	6	2	1	1
Obama	MD	90	5	1	2	1
Obama	MA	90	5	2	3	1
Obama	NC	91	6	1	1	1
Obama	LA	91	5	2	2	1
Obama	GA	92	5	1	1	1
Obama	MS	93	4	1	2	1
Obama	DC	95	2	1	1	1

Worst states for Republican retention

2012 Vote	State	Clinton	Trump	Johnson	Stein	Other
Romney	DC	15	65	14	1	5
Romney	UT	6	67	8	1	19
Romney	MA	9	79	8	1	3
Romney	NH	7	82	6	1	4
Romney	NM	5	80	10	1	4
Romney	RI	5	83	8	1	4
Romney	VT	6	84	5	1	4
Romney	CT	6	85	7	0	2
Romney	HI	5	84	7	0	4
Romney	ID	3	82	6	1	9
Romney	IL	7	86	5	1	2
Romney	MD	7	86	5	1	2
Romney	WA	5	84	7	1	4
Romney	AK	4	84	7	0	4
Romney	CA	5	85	6	1	3

Best states for Republican retention

2012 Vote	State	Clinton	Trump	Johnson	Stein	Other
Romney	FL	5	90	4	0	1
Romney	IA	3	88	5	0	3
Romney	LA	4	89	5	0	2
Romney	OH	5	90	3	0	2
Romney	SC	4	89	4	0	3
Romney	WY	3	88	5	1	3
Romney	MS	4	90	3	0	2
Romney	MO	3	89	4	1	2
Romney	NE	3	89	5	0	3
Romney	NV	3	89	5	0	2
Romney	SD	3	89	6	0	2
Romney	TN	3	89	4	0	2
Romney	KY	3	91	3	0	2
Romney	WV	3	91	4	0	2
Romney	AL	3	92	3	0	2

State-level Changes vs 2012

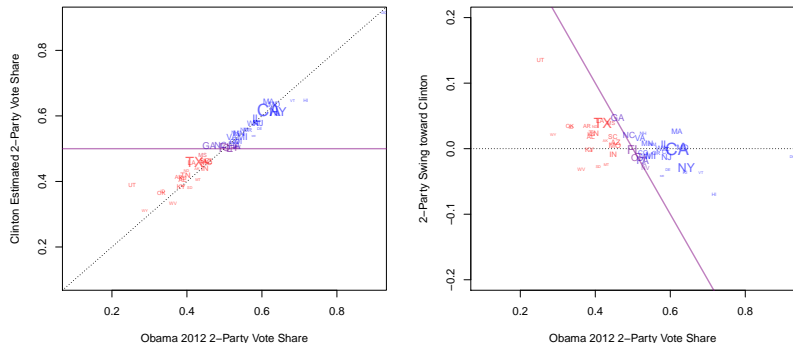


Figure 1: Left: 2016 estimates vs 2012 results. Right: 2016 swings vs 2012 results.

Current Estimated Clinton Margins

Predicted Clinton - Trump Margin

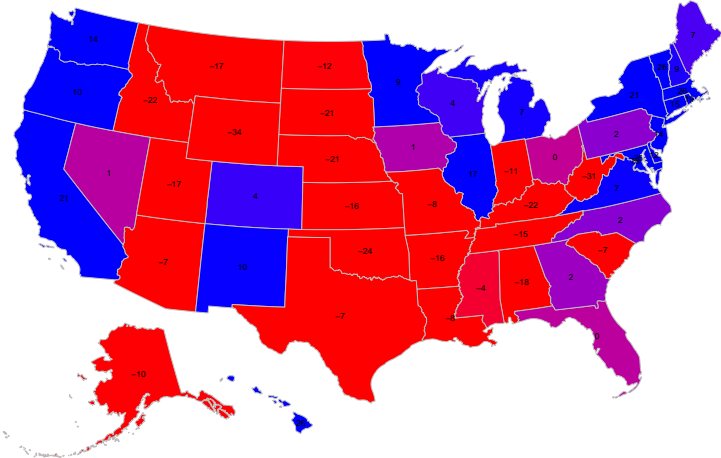


Figure 2: Clinton - Trump margin (% points)

Current Estimated Clinton Probability of Victory

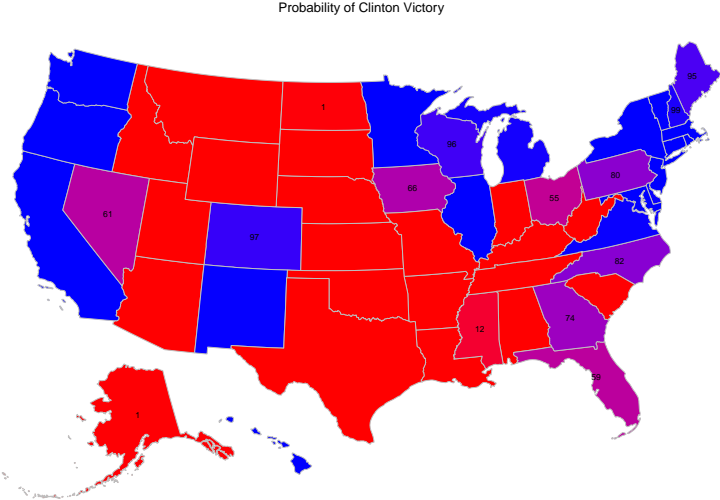


Figure 3: Clinton probability of victory (%) given our best estimate of sampling and non-sampling uncertainty.

Electoral College

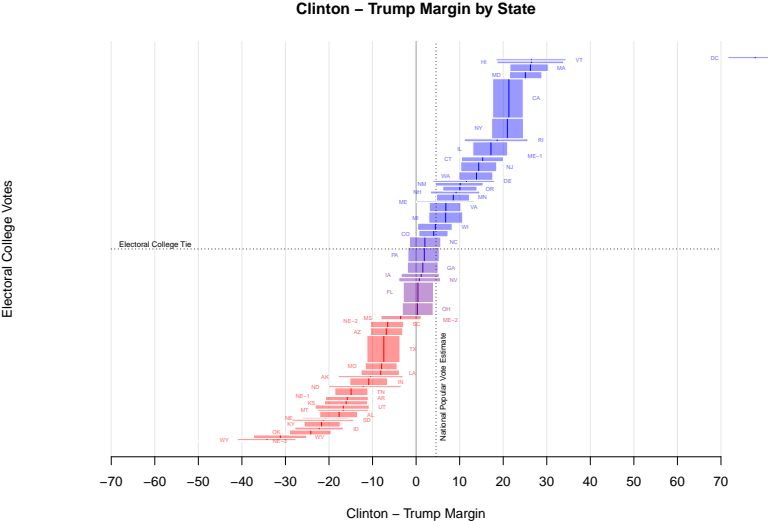


Figure 4: The height of each state is proportional to number of electoral votes.

Electoral College for Trump

- In 2012, Obama had an electoral college advantage
 - Obama's margin in the pivotal state (CO, 5.4%) was larger than his national margin (3.9%).
- In 2016, it looks like Trump has a 1.5% - 2.0% electoral college advantage.
 - California and New York “waste” a lot of votes, no comparably large, heavily Republican states.
 - Texas used to counterbalance more than it does now.
- Pivotal states:
 - Our estimates suggest NC, PA, WI or GA are particularly likely to be pivotal in this sense (note, several of these are not what others are expecting)
 - Note that these are pivotal in a world where Clinton loses FL, OH, IA and NV, and needs to win two of the above.

Pennsylvania

- If you decide to stay up to watch the votes come in, PA is a key state you should watch.
 - 20 electoral votes
 - Obama won by 5.5% in 2012
 - Has been drifting Republican relative to national vote slowly in recent elections.
 - Demographically strong for Trump
 - ▶ Low education White population has traditionally been relatively Dem-leaning (room for gains)
 - ▶ Minority population is primarily Black rather than Hispanic (room for turnout letdown vs Obama in 2012)
 - Very difficult for Trump to win without PA; he would need to win all of NC, WI, GA, FL, OH, IA and NV.
 - Is the most likely tipping point state in our analysis.

Popular and Electoral Vote Uncertainty

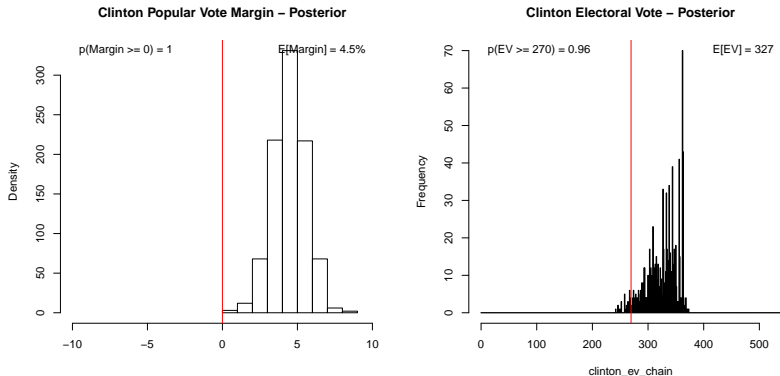


Figure 5: Sampling plus non-sampling uncertainty in national popular and electoral vote.

Time Trend

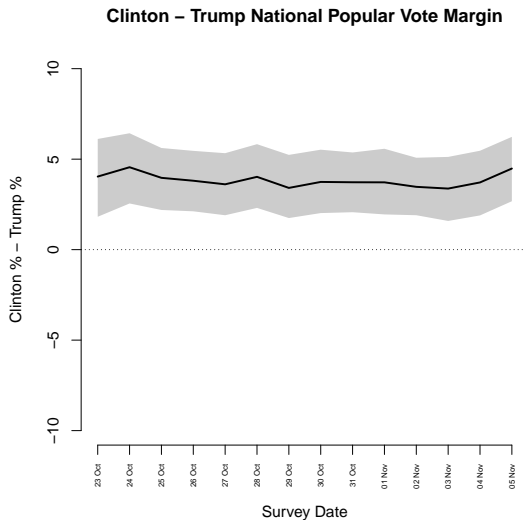


Figure 6: Clinton - Trump margin (% points)

Plausible mechanisms for bias

- Generic reasons we could be wrong
 - Turnout patterns might be consequentially different than 2012
 - Voting intention of YouGov panel might be conditionally unrepresentative of the electorate's voting behaviour
- More specific reasons we might understate Clinton
 - Depressed turnout among (high education?) Republicans
 - Increased turnout among Hispanics
 - Poor online panel coverage of Hispanics who primarily speak Spanish
 - Out of date information about composition of the electorate
- More specific reasons we might understate Trump
 - Increased turnout among disaffected 2012 non-voters
 - Decreased turnout among Black voters versus 2012
 - Johnson supporters "coming home"
 - Social desirability bias (online mitigates, does not eliminate risk)

Table of Contents

Intro

Methods

This is a boring US presidential election

So, Trump is not going to win, right?