Using the British Election Study to Understand the Great Polling Miss

Jonathan Mellon
Nuffield College, University of Oxford
British Election Study
KNIFE-EDGE ON ELECTION DAY

WELL HUNG

EXCLUSIVE

Labour

POLL

Tories

34

PREDICTION:

263 SEATS

QUEEN TO TAKE CONTROL OF ELECTION AFTERMATH

Placement U-turn over monarch's role in closest vote for decades
Polling problems

• The polls missed badly in 2015:
The British Election Study Data

• **Online Campaign wave**: ~1000 respondents each day to track opinion across the campaign

• **Online Post-election wave**: re-contacted all respondents from the campaign wave

• **Face-to-face post-election probability sample**: gold standard survey to compare with other polls and our online survey waves
Hypotheses

1. Late swing
2. Differential Turnout/registration
3. Biased samples/weighting
4. Differential don’t knows
5. “Shy Tories”
Late swing I

- Some indications in campaign wave

Vote intention in BES campaign wave by day
Late swing II

• But the post-election wave doesn’t provide much evidence for this:

• Either there is a different bias in the post-election wave, or there is no late swing
Differential turnout I

• 2010 Turnout: 65%

• Ipsos-MORI expected turnout: 72-74%

• Claimed turnout in the BES post-election wave: 91%

• Actual 2015 turnout in the UK: 66.1%

• Could be people over-reporting turnout or over-sampling more politically interested people
Evidence of turnout misreporting

• 20% of respondents in areas without local elections claim to have voted in them
• 46% of respondents who weren’t registered for EU elections claim to have voted
• 3-6% of respondents claimed to have voted by post before ballots were issued
Labour won unlikely voters by a landslide.
Differential turnout II

• What else we will do:
  – Matching face-to-face survey respondents to marked register data that will validate their turnout
  – Matching all respondents to their voter registration records
Differential turnout III

- Although we don’t have validated turnout yet, we have built a predictive model of validated turnout from 2010.

- If these figures are right, differential turnout would account for half the polling gap.
Biased samples/weighting

• Many ways that this could affect the polls
• Heterogeneity within groups?
• What we will do:
  – Compare the online panel results to the face-to-face probability sample results
  – Systematically look for heterogeneity within weighting groups
Age problems in the BES
Differences in sample

Some demographic differences could underestimate Cons:

• >70 underrepresented in our >65 age group
• >70 are significantly more Conservative leaning and have high turnout
• Too many party members (most are Lab supporters)
• Too few homeowners

But others are more likely to underestimate Lab:

• Too white
• Not routine/semi-routine
• Too many married people
• Much too educated
• Too many private school
Differential don’t knows

• In the campaign wave, don’t knows weren’t seem obviously closer to one party or the other
• Con 2.1 pt lead among don’t knows but a 0.3 lead among the rest of the sample
• But, don’t knows were only 6% of the campaign sample
• So the difference accounts for just 0.1 percentage points of the difference
Shy Tories I

• Some Labour supporters prefer Conservatives on “most important issue” or like David Cameron

• However, likely shy Tories (Labour intenders but like Cameron) have lower social desirability scores than other Labour intenders
Shy Tories II

• Several pollsters have suggested that question ordering might affect shy Tories
• We had a larger block of questions before vote intention in the campaign than in the post-election wave
• But no evidence of an ordering effect:
Shy Tories III

- Missing Tories aren’t in the right places:
<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Evidence</th>
<th>Opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Late swing</td>
<td>• Initial indications in campaign</td>
<td>Unlikely unless a new source of error appears in post wave that was absent in campaign</td>
</tr>
<tr>
<td></td>
<td>• No evidence of swing in post-wave</td>
<td></td>
</tr>
<tr>
<td>Differential turnout</td>
<td>• Large over-reporting of turnout</td>
<td>Strong evidence</td>
</tr>
<tr>
<td></td>
<td>• Predictive model suggests reweighting for liars would reduce gap substantially</td>
<td>We estimate that it might explain 3.3-4.3 percentage points of the gap</td>
</tr>
<tr>
<td>Biased samples/weighting</td>
<td>• Evidence of heterogeneity within weighting groups</td>
<td>Possible but insufficient evidence</td>
</tr>
<tr>
<td></td>
<td>• Many small biases</td>
<td></td>
</tr>
<tr>
<td>Differential don’t knows</td>
<td>• No systematic difference in don’t know shifts</td>
<td>Weak evidence but likely small effect: 0.1 percentage points</td>
</tr>
<tr>
<td>Shy Tories</td>
<td>• Lower social desirability scores among likely “shy Tories”</td>
<td>Still possible but evidence so far points against</td>
</tr>
<tr>
<td></td>
<td>• Missing Tories in wrong places</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• No ordering effect</td>
<td></td>
</tr>
</tbody>
</table>