A bootstrap method for estimating the sampling variation in point estimates from quota samples

Jouni Kuha
London School of Economics and Political Science

Patrick Sturgis
University of Southampton

Conference on Inference for Non-Probability Samples
Sciences Po, Paris
16.3.2017

UK General Election 2015: Vote shares (%) from final polls

Pollster	Survey mode†	Days of fieldwork	Sample size	Party‡					
				Con	Lab	Lib	UKIP	Green	Other
Populus	O	5–6 May	3917	34	34	9	13	5	6
Ipsos-MORI	P	5–6 May	1186	36	35	8	11	5	5
YouGov	O	4–6 May	10307	34	34	10	12	4	6
ComRes	P	5–6 May	1007	35	34	9	12	4	6
Survation	O	4–6 May	4088	33	34	9	16	4	4
ICM	P	3–6 May	2023	34	35	9	11	4	7
Panelbase	O	1–6 May	3019	31	33	8	16	5	7
Opinium	O	4–5 May	2960	35	34	8	12	6	5
TNS UK	O	30/4–4/5	1185	33	32	8	14	6	6
Ashcroft*	P	5–6 May	3028	33	33	10	11	6	8
BMG*	O	3–5 May	1009	34	34	10	12	4	6
SurveyMonkey*	O	30/4-6/5	18131	34	28	7	13	8	9
Election result				37.7	31.2	8.1	12.9	3.8	6.4
Mean absolute error				*3.9*	*2.7*	*0.9*	*1.4*	*1.3*	*1.1*

2015 polling inquiry

- Commissioned by the British Polling Council (BPC) and the Market Research Society
- Inquiry team: Patrick Sturgis (chair), Nick Baker, Mario Callegaro, Stephen Fisher, Jane Green, Will Jennings, Jouni Kuha, Benjamin Lauderdale, and Patten Smith
- Thank you to BPC polling companies for the data: ComRes, ICM, Ipsos-MORI, Opinium, Panelbase, Populus, Survation, TNS UK, and YouGov

Research questions

For the polling inquiry:

- Why were the polls wrong?
  - google "*2015 polling inquiry report*" for the report of the findings and recommendations (Sturgis et al. 2016)

Research questions

For the polling inquiry:

- Why were the polls wrong?
  - google "*2015 polling inquiry report*" for the report of the findings and recommendations (Sturgis et al. 2016)

For today's talk:

- How to quantify sampling variability in the poll estimates?
  - i.e. sample-to-sample variance around whatever the polls are estimating
  - i.e. also MSE/confidence intervals for the true vote shares if the polls were actually estimating them

Methodology of the election polls

Methodology of the election polls

(i) Collect a quota sample of individuals, with quota targets for some marginal distributions of some variables $\mathbf{X}^*$

Methodology of the election polls

(i) Collect a quota sample of individuals, with quota targets for some marginal distributions of some variables $\mathbf{X}^*$

(ii) Create poststratification weights w_i^* for respondents i, so that weighted marginal distributions of elements of $\mathbf{X} = (\mathbf{X}^*, \mathbf{X}^\dagger) = (\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(p)})$ match population targets

Methodology of the election polls

(i) Collect a quota sample of individuals, with quota targets for some marginal distributions of some variables $\mathbf{X}^*$

(ii) Create poststratification weights w_i^* for respondents i, so that weighted marginal distributions of elements of $\mathbf{X} = (\mathbf{X}^*, \mathbf{X}^\dagger) = (\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(p)})$ match population targets

(iii) Assign for each respondent a probability p_{Ti} that the respondent will turn out to vote ($T_i = 1$) in the election

- conditional on self-reported likelihood to vote (L_i), and possibly other variables

Methodology of the election polls

(i) Collect a quota sample of individuals, with quota targets for some marginal distributions of some variables $\mathbf{X}^*$

(ii) Create poststratification weights w_i^* for respondents i, so that weighted marginal distributions of elements of $\mathbf{X} = (\mathbf{X}^*, \mathbf{X}^\dagger) = (\mathbf{X}_{(1)}, \ldots, \mathbf{X}_{(p)})$ match population targets

(iii) Assign for each respondent a probability p_{Ti} that the respondent will turn out to vote $(T_i = 1)$ in the election
- conditional on self-reported likelihood to vote (L_i), and possibly other variables

(iv) Estimate (predict) shares of vote in the election (P_i) as weighted proportions of self-reported intended vote (V_i), with weights $w_i = w_i^* p_{Ti}$

Assumptions of the election polling methodology

This delivers approximately unbiased estimates of the election results if the following assumptions are satisfied:

Assumptions of the election polling methodology

This delivers approximately unbiased estimates of the election results if the following assumptions are satisfied:

1. *Representative sampling*: $p(V, L|\mathbf{X})$ is the same in the sample as in the population

Assumptions of the election polling methodology

This delivers approximately unbiased estimates of the election results if the following assumptions are satisfied:

1. *Representative sampling*: $p(V, L|\mathbf{X})$ is the same in the sample as in the population
2. *Correct model for turnout probabilities*: Assigned turnout weights p_{Ti} are equal to $p(T_i = 1|V_i, L_i, \mathbf{X}_i)$ in the population

Assumptions of the election polling methodology

This delivers approximately unbiased estimates of the election results if the following assumptions are satisfied:

1. *Representative sampling*: $p(V, L|\mathbf{X})$ is the same in the sample as in the population
2. *Correct model for turnout probabilities*: Assigned turnout weights p_{Ti} are equal to $p(T_i = 1|V_i, L_i, \mathbf{X}_i)$ in the population
3. *Agreement between stated vote intention and actual vote*: $p(V|T = 1) = p(P|T = 1)$

Sampling variability: "Margin of error"

This — if anything — is commonly reported with poll estimates.

For an estimated proportion $\hat{\pi}$, the margin of error is the half-width of a 95% confidence interval under the assumption of simple random sampling

- i.e. $1.96\sqrt{\pi(1-\pi)/n}$

If this is calculated under $n = 1000$ and taking $\pi = 0.5$ to get an upper bound, we obtain the commonly reported margin of error of "$\pm$3%"

Sampling variability: "Margin of error"

This — if anything — is commonly reported with poll estimates.

For an estimated proportion $\hat{\pi}$, the margin of error is the half-width of a 95% confidence interval under the assumption of simple random sampling

- i.e. $1.96\sqrt{\pi(1-\pi)/n}$

If this is calculated under $n = 1000$ and taking $\pi = 0.5$ to get an upper bound, we obtain the commonly reported margin of error of "$\pm$3%"

Note: For the election polls, even the "n" is not entirely clear, because the number of actual voters is not known.

- Should it be the number of poll respondents, or the predicted number of voters ($\sum_i p_{Ti}$)?

Sampling variability: A bootstrap approach

We propose this to better reflect the actual sampling design.

Should give a good estimate of the sampling variability if the observed sample is representative of what would be observed in repeated samples using the same design.

Sampling variability: A bootstrap approach

We propose this to better reflect the actual sampling design.

Should give a good estimate of the sampling variability if the observed sample is representative of what would be observed in repeated samples using the same design.

Bootstrap procedure:

Sampling variability: A bootstrap approach

We propose this to better reflect the actual sampling design.

Should give a good estimate of the sampling variability if the observed sample is representative of what would be observed in repeated samples using the same design.

Bootstrap procedure:

1. Draw bootstrap resamples from the observed sample, in a way which mimics the quota sampling.

Sampling variability: A bootstrap approach

We propose this to better reflect the actual sampling design.

Should give a good estimate of the sampling variability if the observed sample is representative of what would be observed in repeated samples using the same design.

Bootstrap procedure:

1. Draw bootstrap resamples from the observed sample, in a way which mimics the quota sampling.
2. Estimate vote shares from each bootstrap sample, using the same estimation procedure as for the real sample

Sampling variability: A bootstrap approach

We propose this to better reflect the actual sampling design.

Should give a good estimate of the sampling variability if the observed sample is representative of what would be observed in repeated samples using the same design.

Bootstrap procedure:

1. Draw bootstrap resamples from the observed sample, in a way which mimics the quota sampling.
2. Estimate vote shares from each bootstrap sample, using the same estimation procedure as for the real sample
3. Use variation across estimates from the boostrap samples to estimate the sampling variation

Creating the bootstrap resamples

0a. Quota targets: Distributions of the quota variables in the observed sample

0b. Initial settings:

- Retained sample so far: Empty
- Resampling pool, of size $m(= n)$: The observed sample

Creating the bootstrap resamples

0a. Quota targets: Distributions of the quota variables in the observed sample

0b. Initial settings:

- Retained sample so far: Empty
- Resampling pool, of size $m(= n)$: The observed sample

1. Sample m observations with replacement from the pool

Creating the bootstrap resamples

0a. Quota targets: Distributions of the quota variables in the observed sample

0b. Initial settings:

- Retained sample so far: Empty
- Resampling pool, of size $m(= n)$: The observed sample

1. Sample m observations with replacement from the pool
2. Retain all the sampled observations which (when combined with previously retained sample) do not go over any quota

Creating the bootstrap resamples

0a. Quota targets: Distributions of the quota variables in the observed sample

0b. Initial settings:

- Retained sample so far: Empty
- Resampling pool, of size $m(=n)$: The observed sample

1. Sample m observations with replacement from the pool
2. Retain all the sampled observations which (when combined with previously retained sample) do not go over any quota
3. Reduce the pool to include only observations where no variable values correspond to a full quota

Creating the bootstrap resamples

0a. Quota targets: Distributions of the quota variables in the observed sample

0b. Initial settings:

- Retained sample so far: Empty
- Resampling pool, of size $m(= n)$: The observed sample

1. Sample m observations with replacement from the pool
2. Retain all the sampled observations which (when combined with previously retained sample) do not go over any quota
3. Reduce the pool to include only observations where no variable values correspond to a full quota
4. Repeat 1.-3. until all targets are full or pool is empty

Uncertainty intervals for Con-Lab difference (%)

Pollster	Con-Lab (%) (election result = +6.5%)		
	Estimate	95% interval*	N
Populus	-0.1	(-2.5; +2.0)	3695
Ipsos-MORI	-0.3	(-6.6; +6.1)	928
YouGov	+0.4	(-1.1; +1.8)	9064
ComRes	+0.8	(-4.6; +6.3)	852
Survation	+0.1	(-2.2; +2.5)	3412
ICM	+0.0	(-2.8; +3.1)	1681
Panelbase	-2.7	(-5.6; +0.2)	3019
Opinium	+0.4	(-1.8; +2.5)	2498
TNS UK	+0.8	(-3.6; +5.2)	889

* Adjusted percentile intervals, calculated using 10,000 bootstrap samples.

- None of the intervals includes the election result
- These results are not consistent with the claim that the error in the polls was due to sampling variation only

Bootstrap standard errors of party vote shares

	Con(%)	Lab(%)	N
Populus	0.7	0.7	3695
Ipsos-MORI	1.8	1.9	928
YouGov	0.4	0.4	9064
ComRes	1.5	1.9	852
Survation	0.7	0.7	3412
ICM	0.9	1.0	1681
Panelbase	0.9	0.9	3019
Opinium	0.6	0.7	2498
TNS UK	1.4	1.3	889

- Yellow ones are those where the standard error is ≥ 1.5, i.e. where the "$\pm 3\%$" margin of error would be too low. But these also have sample size of less than 1000.
- More generally, is the SRS assumption behind margins of error adequate? Is the sampling actually more or less efficient than SRS?

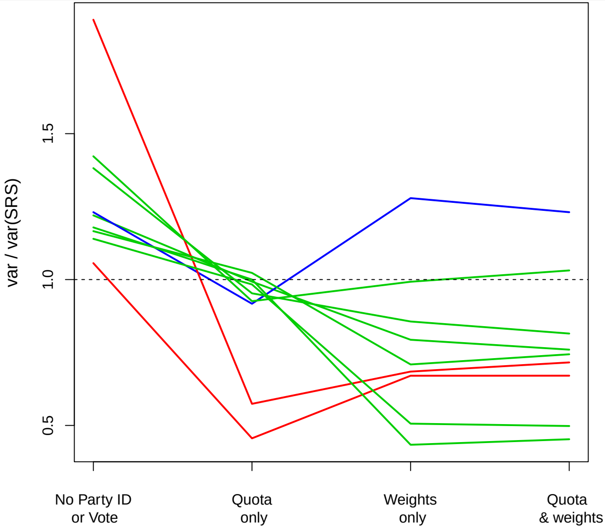
Design effects of Con-Lab difference

We have also generated bootstrap replicates under the following scenarios for the sampling:

1. "SRS": Using neither quota targets nor poststratification weights
2. Using quotas but not weights
3. Using weights but not quotas
4. Published estimates: Using both quotas and weights
5. Using both quotas and weights, but omitting from both variables related to party identification and past vote (which are expected to be best predictors of upcoming vote)

On the next slide: Variance of Con-Lab difference from 5,2,3,4, divided by variance from 1 (i.e. the "design effects"), separately for each pollster

Variance of Con-Lab (vs. "SRS")



Red: Party ID/Vote used as quota; Blue: No Party ID/Vote used at all

Observations

- In general, the variances behave as we would expect
  - Similar patterns as for design variances in probability sampling, with quota in the role of strata and poststratification weights as weights
- Specifics for the quota and weights used in these election polls:
  - Estimates are less variable than "SRS" only if party ID or past vote is used for quota and/or weights
  - Quota reduce variability, roughly to the level of "SRS" if quota do not include party ID/vote
  - Weights increase variability, except when they introduce party ID/vote