How do Internet Polls Fare in Predicting Election Results?

Presented at the World Social Science Forum, Montréal October 13-15, 2013

> By Claire Durand, Département de sociologie, Université de Montréal

> > © C. Durand, 2013



Outline

- What is an internet poll?
 - Why would an internet poll be less reliable?
- How to assess empirically whether internet polls differ?
- Four cases:
 - Canada 2011, as a whole, in Ontario, in Quebec
 - Alberta 2012
 - Quebec 2012
 - BC 2013
- Compared to U.S. presidential 2012.
- Synthesis & conclusion.



What is an internet poll?

Diverse methods

- Usually opt-in panel, but... how are panel members recruited, contacted, etc.?
 - It varies...
 - Recruitment via web sites, social media, telephone polls, river sampling.
 - The goal: Establish a sampling base.
- Invitations: sample of the sampling base.
 - Members of underrepresented groups receive more invitations.
- A convenience sample BUT sometimes close to Probability sampling with quotas (PSQ).

Why would they differ?

- Differences between internet users and nonusers (Bigot et al. 2010, etc.)
- Differences between convenience vs probability sample (Stephenson & Crete, 2011; Krosnick & coll., etc.).
- Use of quotas & bad sample management.
- Attempts at adjusting samples using weighting, propensity scores,did not give good results (Tourangeau et al., Loosveldt, etc...
- Main differences: religious practice, marital status, activity status, values,...
 - Related to political preferences & voting behaviour.

How to assess empirically...

The impact of web polls in estimates during electoral campaigns

- First, establish the likely evolution of preferences during the campaign.
- Second, establish the direction of bias, when present, i.e. the difference between prediction and results.
- Third, establish whether web polls estimates differ from polls using other methodologies all things being equal, i.e., conducted at the same period.



Four recent cases in Canada

Did Internet polls contribute to bias in estimates?

- Canada 2011:
 - As a whole
 - In Ontario
 - In Quebec
- Alberta 2012
- Quebec 2012
- BC 2013



Canada 2011 - total



+ Bias against the Conservative party.
 + Methods (Web and IVR) contribute to bias in a similar direction.



Canada 2011 - Ontario



+Cons: -4.1% +Nanos: + +2.0 Cons + +3.2 Libs

+Bias against the Conservative party.

+Only Nanos, differs significantly from the other pollsters/methods.



Canada 2011 - Quebec



+Though there was huge movement, prediction is almost perfect.

+IVR underestimates the Bloc (contrary to usual).

Alberta 2012



+Wild Rose: +4.9; Cons: -11.8 = 16.7
+Web: WR:-2.6; PC: +2.1: reduces bias.
+IVR: WR: +2.1; PC: -2.9: contributes to bias.

Quebec 2012



- +Underestimation of the Quebec Liberal Party though no substantial movement.
- +Web differs but has no impact on bias per se.
- +IVR reduces bias.



British Columbia 2013



+There is movement towards the Liberal party.

- +Liberals: -7.5; NDP:+2.9, Total: **10.4 pts**
- +WEB (68% of polls): NPD: +4.3. contributes to bias.

U.S. 2012 Presidential election



- Web polls=15% of all polls.
 No difference between web polls and others.
- Impact of likely voter models.



Synthesis

In Canada,

- Web polls differed from other polls in all elections since 2011.
- They underestimated the right-wing vote in Canada as a whole and in Alberta
 - Contributing to bias in Canada and reducing it in Alberta.
- They overestimated the left-wing vote in BC; their contribution to the catastrophe is substantial since they constitute 68% of the polls.
- They overestimated CAQ in Quebec but this had no effect on the overall bias in the polls (underestimation of the Quebec Liberal Party).
- In the U.S. 2012 presidential campaign, Web polls are less used, no systematic difference with other polls.



Limits

- The main limit in these analyses is the fact that methods and firms go together.
 - This is even more the case in the Quebec 2012 election where the 3 firms used 3 different methods.
 - We can't conclude for sure that it is a question of methods and not mainly a question of firm.
- Often, there are not many polls. The analyses will have to be corroborated in other elections.



Conclusion - 1

- There is much to be done if we want to arrive at reliable internet polls.
 - The coverage problem → it will be solved rapidly and easily.
 - Constitution of a sampling base→Much to be done and... not much work seems to have been done yet except for very expensive bases (Knowledge Networks, etc.).
 - Management of sampling base and sample→ much to be done to rely on more classical methods and tackle non response problems.
 - ► Focus of research→ should be away from adjustment, towards improvement of sampling base & management.



Conclusion - 2

- It is likely that more and more polls will be conducted using the Web,
 - Especially in less populated geographical areas where media cannot pay for more expensive polls.
- The method used by most pollsters now is akin to quota sampling
 - Sometimes with a probabilistic sampling frame
 - Most of the time with a convenience sample.
- Representing the population demographically does not mean that it is well represented sociopolitically.
- Results from Web polls may be misleading, especially when they are conducted on valuerelated topics and there is a split around 50/50.