Liberal Party losing votes due to COVID-19 economic impact

Cesar Y. Villarreal Guzman

October 8, 2020

Executive Summary

In this experiment we conduct a simulation based survey study to investigate the impact of the second wave of the COVID-19 pandemic on the political views of Canadian's. The study follows the premise of random selection of participants from a pool of candidates corresponding to the different Canadian provinces. The number of participants from each province was determined by following the *stratified simple random sampling* method (Wu and Thompson 2020). The survey was designed using the Typeform platform. The plan of administration is by email , coupled with a small monetary incentive ($5 coupon) upon completion.

Research has long suspected on the correlation between the length of a survey and response rate. One exampleis from a recent experiment which showed a sizable negative effect of length on completion in web surveys (Marcus et al. 2007). As a result, we decided to limit the survey to four questions:

1. For whom did you vote in the 2019 Canadian federal elections?
2. How do you feel about the general direction of our country at the moment?
3. How much did the COVID-19 pandemic influence your response to the last question?
4. If there was a general election tomorrow, which party would you vote for?

Responses where simulated based on data collected from a list news videos covering the address of Prime Minister Trudeau on the COVID-19 pandemic. The videos used where from September and first week of October. Overall we collected: *number of views, number of likes*, *number of dislikes* and *number of comments*. This was achieved by calling the official YouTube API provided by Google. Data was then processed and used to generate responses. Figure 1 displays important results from this study.
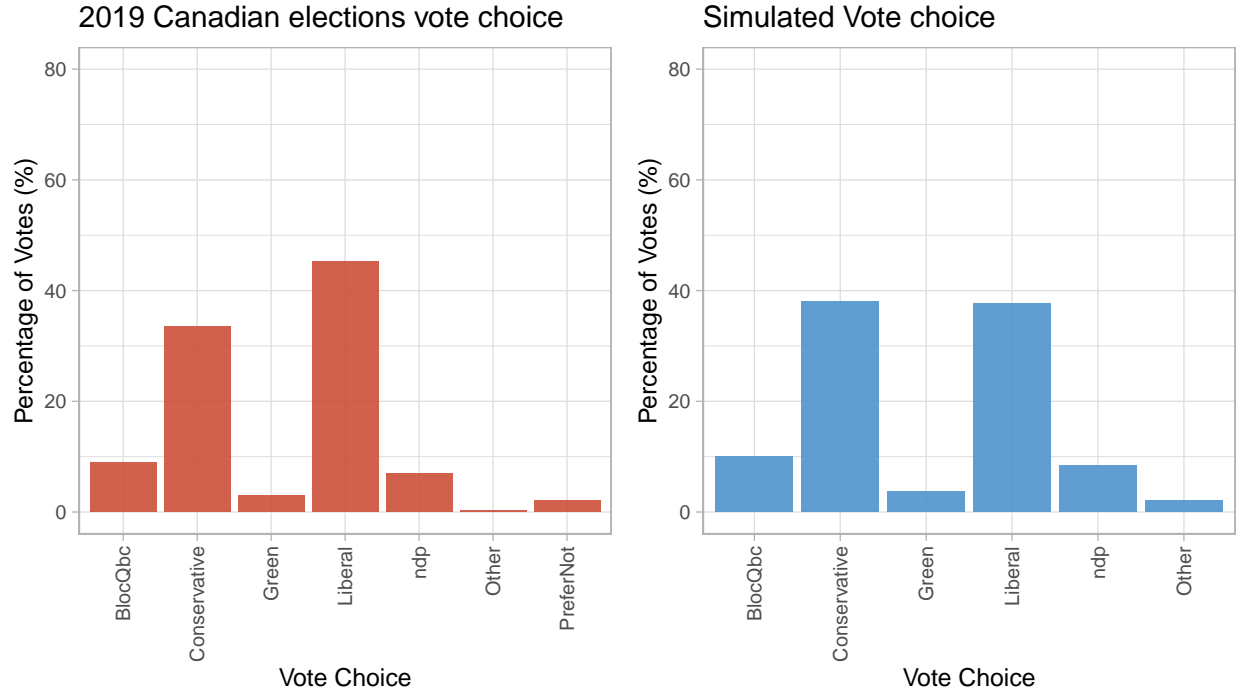
Figure 1: Results for question 'For whom did you vote in the 2019 Canadian federal elections?'

As seen there is a loss of approximately 4% in the number of Liberal voters. The conclusion reached in this investigation, is that there is some correlation with the COVID-19 pandemic. Further investigation is required to achieve a stronger conclusion and we advise the Liberal political team to take these results with a grain of salt.

Introduction

The COVID-19 pandemic has nearly created a state of economic collapse in Canada, and this second wave promises to be more severe than the first one encountered in April and May of this year. Strict measures where taken by the Canadian leadership as a countermeasure to the pandemic. These measures produced an overwhelming amount of lost jobs and businesses to go bankrupt. We believe this has shifted the mentality and overall perspective of the Canadian population which as a result may impact the overall acceptance of the Canadian Liberal party. This experiment attempts to quantify such claims and evaluates the implications of the validity of such claims.

Survey Methodology

In this section we thoroughly discuss the all aspects of the design of the experiment. First we talk about logistics, statistical basis, frame, sampling methodology and methods for non-responses this is then followed by content of the survey and budgeting

We conduct the experiment as a web survey created using Typeform as a survey design framework. A complete list of Canadian addresses has been provided and is used as frame for

the survey. We follow stratified simple random sampling as our design, with each province as strata. The selection of candidates is by simple random sampling without replacement (SRSWOR). The selected candidate is sent an invitation to register to complete the survey. If the candidate accepts the invitation, we send an email with a link to the survey. Upon competition, the survey redirects respondents to a \$5 CAD reward. In the case the candidate does not accept, we draw another candidate following the same SRSWOR method, this is repeated until the desired sample size of $n = 10,000$ is reached.

The number of candidates we select from each province is calculated using proportional allocation. Say the population has size N and the population of province h is N_h, then the number of samples to draw from province h is

$$n_h = \frac{n}{N} N_h$$

where n is the total sample size. Similarly, we compute the stratum weight, which is the relative size of the stratum within the overall population, a needed quantity for population parameter estimation. The stratum weight of strata h is given by

$$W_h = \frac{n_h}{n}$$

where n_h is as above (Wu and Thompson 2020). A complete table of sample computed strata sample sizes and weights can be found in Table 1 in the Appendix.

To ensure a high response rate we selected a small set of questions. This is because, research has long showed some correlation between the length of a survey and response rate, for example, an experiment evaluating redesigns of the U.S. Census found that shortening the questionnaire increased response rate (Dillman, Sinclair, and Clark 1993), a more recent experiment showed a sizable negative effect of length on completion in web surveys (Marcus et al. 2007).

In constructing the survey, we followed (Fan and Yan 2010) very closely. In particular Fan and Yan (2010) suggests that the optimal survey should take less than thirteen minutes to complete. Our survey takes approximately two to four minutes. The questions are:

1. Are you eligible to vote?
2. Which Canadian province do you live in?
3. For whom did you vote in the 2019 Canadian federal elections?
4. How do you feel about the economic direction of our country at the moment?
5. How much did the COVID-19 pandemic influence your response to the last question?
6. If there was a general election tomorrow, which party would you vote for?

Questions are presented in this order and not randomized since there are no empirical results that either prove or disprove the efficiency of this method. The actual wording of the questions is different than what is presented above, however the language used is significantly less formal, done with the intention to yield higher response rates. The third question has the option to not provide an answer since we understand this might be a sensitive topic and could negatively impact our response rate.

It is important to mention the drawbacks of providing this option. Avoiding the question gives respondents an alternative to the cognitive burden of the question and they are more likely to take this mental shortcut referred in the literature as satisficing (Krosnick 1991). Consequently, question two is the only question where this is allowed.

Logic is implemented to the survey in question 1, if the respondent is not an eligible voter then we only display questions two, three and four, Typeform fills all other values with *NA* and will not be included in the analysis, since we are purposefully targeting Canadian adults. All other questions are mandatory and we will discard any surveys that are left half answered. This Typeform feature is only available under the professional plan. We obtained this plan for free because, Typeform is offering 3 months free for anyone working on COVID-19 related projects that are strictly not-for-profit.

The expected cost for this project is broken down as follows. First the $5 CAD monetary reward delivered upon completion of the survey, with a total of approximately $50,000 CAD, for 10,000 respondents. An estimated 20,000 invitations with a price of $0.92 CAD per invitation and estimated total of $18,400 CAD. Finally, the cost for printing 20,000 invitations is $400 CAD. **Therefore, the expected cost of this experiment is of approximately $68,800 CAD**.

Results and Discussion

In this section we begin by displaying the modeled data in the form of frequency plots .We then follow it by a discussion of the data and statistical inferences, we then finish with a weaknesses and future work section. Data was pre-cleaned by removing all NA values corresponding to all non-eligible respondents. The resulting sample size is $n = 8961$.For convenience we group question two and five, question two asked respondents which party did they vote for in the 2019 Federal elections, and question five asks to choose a party if general elections where held tomorrow. Similarly we grouped question three and four, which asked about the general economic direction of our country and the impact of COVID-19 in the response of question three. Figure 2 and 3 display the raw data simulated in this experiment.
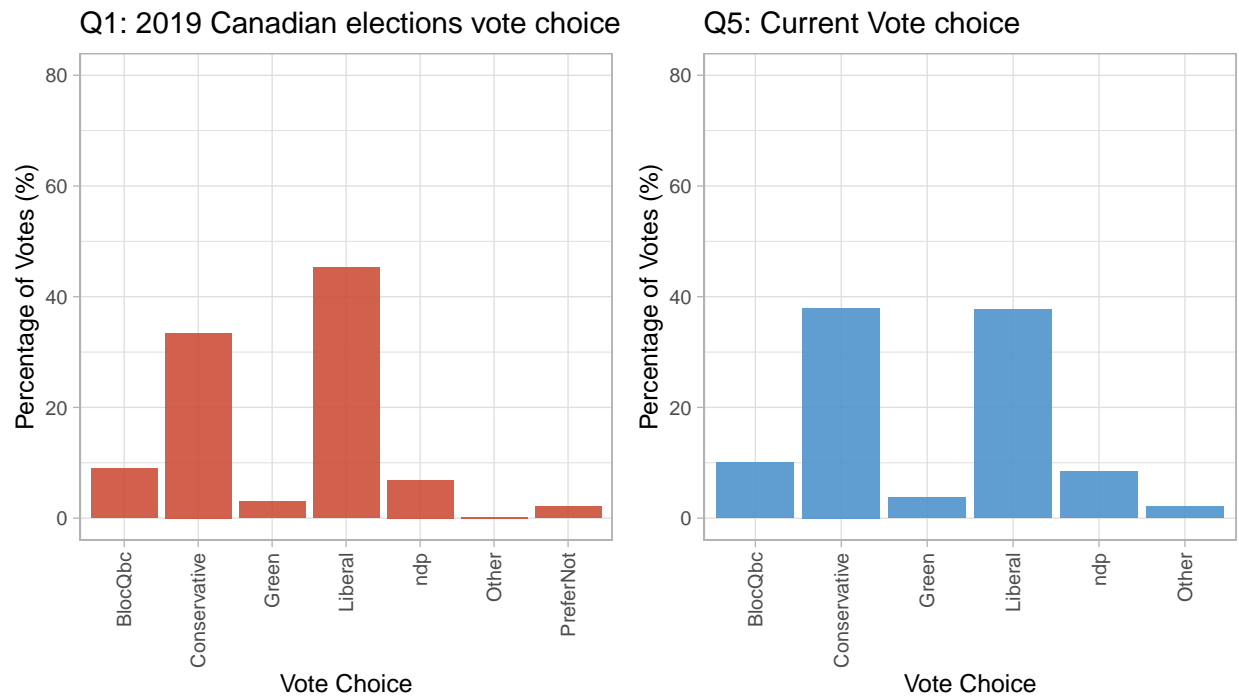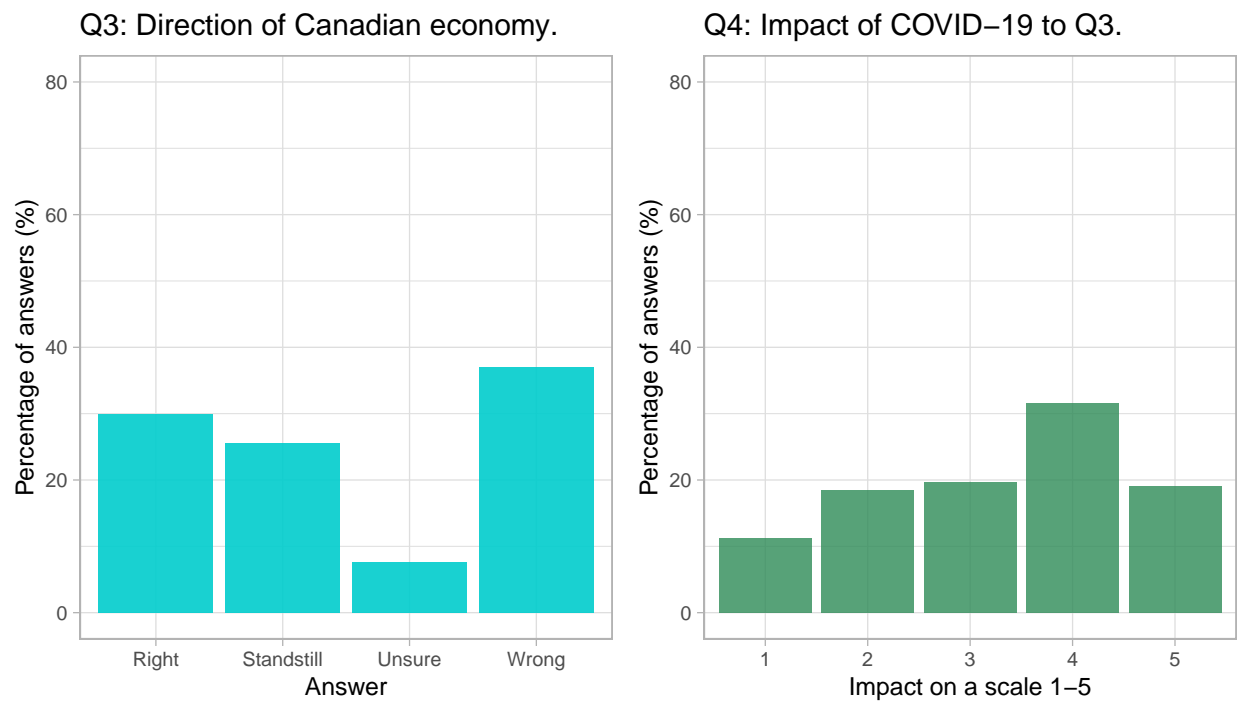
Figure 2: Answers for questions two and five.

Figure 3: Answers for questions three and four.

Raw data shows that there is a small percentage of respondents that are choosing to vote for another party which is not the Liberal party. Similarly a majority feel like the Canadian economy is moving in the wrong direction with a significant portion of respondents claiming that COVID-19 has a high impact in their response to the previous question. However, this are just the raw results and we cannot make any empirical conclusions based just on these.

To properly estimate population parameters we seek to calculate the stratified sample mean. The stratified sample mean is given by

$$\overline{y}_{st} = \sum_{h=1}^{H} W_h \overline{y}_h$$

where W_h is as in the previous section, the stratum weight, and $\overline{y}_h$ the sample mean for stratum h is given by:

$$\overline{y}_h = \frac{1}{n_h} \sum_{i \in S_h} y_i$$

Here S_h is the sample corresponding to stratum h. Moreover, the stratified sample mean is an unbiased estimator of the population mean μ_y, i.e.

$$E(\overline{y}_{st}) = \mu_y$$

similarly an unbiased variance estimator for $\overline{y}_{st}$ is given by

$$v(\overline{y}_{st}) = \sum_{h=1}^{H} W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{s_{yh}^2}{n_h}$$

as per Theorem 3.1 in Wu and Thompson (2020).

As per the introduction, the first thing to compute is the percentage of votes the Liberal party has with respect to the population and compare this number to what the current percentage of votes is estimated to be. Using the method described above, the estimated percentage of votes obtained in the 2019 election is of **40.56%**, and as a point of comparison the actual percentage of votes with which the Liberal party won the 2019 Canadian elections is 39.47%. Similarly, we estimate that the current percentage of votes is **33.78%** with a variance of 2.72% and a 90% confidence interval [12.64, 54.92]. An estimated loss of 6.78% of votes.

Similarly we find that most people that claim Canadian economy is moving in the wrong direction, feel this is partially caused by the COVID-19 pandemic. Following the same methodology we compute an estimate that **33.31%** of Canadians feel the economy moving in the wrong direction, with a computed variance of 2.62% and a 90% confidence interval [13.03, 54.53]. Similarly **22.62%** of Canadians feel the economy is at a standstill, with a computed variance of 1.18% and a 90% confidence interval [19.84, 47.72].

Finally we discuss some of the weaknesses of our approach. The first comes from the design of the survey. It has been justified the reason for such a short survey however, this is also a weakness because it does not provide nearly enough information to study and discuss. We had to limit this study to two main results, the loss of votes and the percentage of people that

are feeling a negative direction in our economy. Moreover, this survey comes with a significant cost, nearly $70,000 CAD. A non-probability, opt-in web survey offers the promise of being significantly cheaper, and possibly generate a bigger sample size at the cost of theoretical complications.

Future work could include a slightly longer survey, in which there is a large set of questions and with the use of active matrix factorization, we can automate the design of shorter by a variance-minimizing active learning criterion which chooses the most informative questions per respondent (Zhang et al. 2020). This type of survey can be a non-probability based survey that attempts to gather at least $50,000$ samples. More complex theoretical tools are required for the statistical analysis of such survey.

Appendix

Table 1: Stratum Sizes and Weights

Province	Stratum Weight	Stratum Sample Size
Newfoundland and Labrador	0.014	138
Prince Edward Island	0.004	43
Nova Scotia	0.026	258
New Brunswick	0.021	206
Quebec	0.226	2255
Ontario	0.388	3874
Manitoba	0.036	363
Saskatchewan	0.031	311
Alberta	0.116	1162
British Columbia	0.135	1355
Yukon	0.001	12
Northwest Territories	0.001	12
Nunavut	0.001	11

References

Bennett, Stephen T, and Mark Steyvers. 2020. "Estimating Covid-19 Antibody Seroprevalence in Santa Clara County, California. A Re-Analysis of Bendavid et Al." https://doi.org/10.1101/2020.04.24.20078824.

Couture-Beil, Alex. 2018. *Rjson: JSON for R.* https://CRAN.R-project.org/package=rjson.

Dillman, Don A., Michael D. Sinclair, and Jon R. Clark. 1993. "Effects of Questionnaire Length, Respondent-Friendly Design, and a Difficult Question on Response Rates for Occupant-Addressed Census Mail Surveys." *Public Opinion Quarterly* 57 (3): 289. https://doi.org/10.1086/269376.

Fan, Weimiao, and Zheng Yan. 2010. "Factors Affecting Response Rates of the Web Survey: A Systematic Review." *Computers in Human Behavior* 26 (2): 132–39. https://doi.org/10.1016/j.chb.2009.10.015.

Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5 (3): 213–36. https://doi.org/10.1002/acp.2350050305.

Marcus, Bernd, Michael Bosnjak, Steffen Lindner, Stanislav Pilischenko, and Astrid Schütz. 2007. "Compensating for Low Topic Interest and Long Surveys." *Social Science Computer Review* 25 (3): 372–83. https://doi.org/10.1177/0894439307297606.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wu, Changbao, and Mary E. Thompson. 2020. *Sampling Theory and Practice.* Springer International Publishing.

Zhang, Chelsea, Sean J. Taylor, Curtiss Cobb, and Jasjeet Sekhon. 2020. "Active Matrix Factorization for Surveys." *Annals of Applied Statistics* 14 (3): 1182–1206. https://doi.org/10.1214/20-aoas1322.