LIBERAL PARTY LOSING VOTES DUE TO COVID-19'S ECOMIC 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, with particular interest on the opinion of the Liberal Party. The study is conducted by random selection of participants from 13 pools of candidates corresponding to the different Canadian provinces. An invitation is first sent to a selected candidate and upon registration administration of the survey is done by email. The survey is coupled with a small monetary incentive ($5 coupon) which is rewarded upon completion of the survey. The survey is designed using the tool www.typeform.com. No personal data is collected to protect the privacy of respondents.

Research has long suspected on the correlation between the length of a survey and response rate. As a result, we decided to limit the survey to five questions:

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?

Responses where simulated based on data collected from a list of news videos in which a representative of the Liberal party either speaks about the economic situation due to the pandemic, or addresses the pandemic as a whole. The videos used where from September and first week of October. Data collected includes: *number of views*, *number of likes*, *number of dislikes* and *number of comments*.

We focus our study on quantifying the number of lost/gained votes in favor of the Liberal Party. Figure 1 displays the raw data collected for questions two and four. This figure helps to visualize the actual findings of the study.

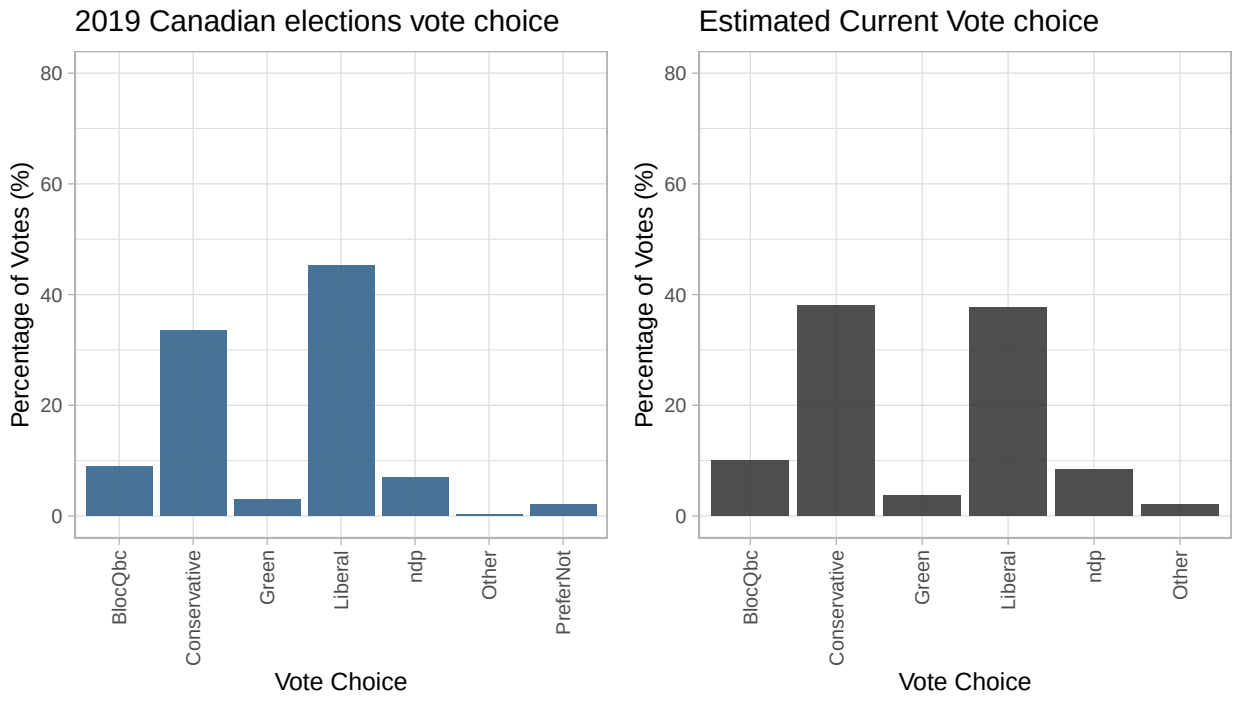

Figure 1: Results for question 'For whom did you vote in the 2019 Canadian federal elections?'

After statistical analysis of the data we suspect an approximate loss of **6.78%** of votes in favour of the Liberal party, and estimate **56%** of Canadians feel our economy is moving in the wrong direction or is at a standstill.

Some major weakness to our approach lie in the simplicity of the survey. We believe a longer study should be carried out, as future work, to gather more information on the social aspects that society thinks are the most impacted by COVID-19. We particularly chose economic impact as the focus given empirical evidence of the negative effects of the pandemic in Canada's and the world economy.

Introduction

There is enough evidence to make the claim that the COVID-19 pandemic has nearly created a state of economic collapse in Canada. What is worse is that this second wave promises to be more severe than the first one. Strict measures where taken by Canadian leadership as a countermeasure to the spread of the virus. These measures produced an overwhelming amount of jobs being lost and businesses that are now in a state of bankruptcy. With this as primary motivation, we claim there has been a shift in the acceptance of the Canadian Liberal party as a result of these measures. This experiment attempts to quantify such claim and evaluate its implications.

Following, there is thorough discussion on the methodology of the study conducted, in this section we discuss logistics, statistical basis and costs, to name a few components. This is followed by the results obtained, which quantify and provide evidence to the original claim.

Finally there is a short discussion on the weaknesses of our methods and areas of future work. References to the survey, code and data modeling methods can be found in the Appendix.

Survey Methodology

In this section we discuss 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 costs.

We conduct the experiment as a web survey created using Typeform as our survey design platform. 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 registration steps. If the candidate accepts, we send an email with a link to the survey and upon competition, the survey redirects users to a $5 CAD coupon 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 province h is given by

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

where n_h is as above (Wu and Thompson 2020). A complete table of 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 they suggests that the optimal survey should take less than thirteen minutes to complete. Our survey takes approximately two to four minutes with the following questions:

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 given there is no empirical result that either proves or disproves the efficiency of randomization. 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 of establishing a more conversational type interaction and 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.

However, avoiding the question gives respondents an alternative, a way out, to the cognitive burden of the question which as a result makes them 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 four and five. Typeform fills missing values with *NA* and will not be included in the analysis, since we are purposefully targeting eligible voters. All other questions are mandatory and we will discard any surveys that are left half answered. Finally, we clarify that we do not ask for any personal information. This is with the purpose to ensure anonymity for participants. We understand these questions consist of very sensitive topics. Screen shots of the survey can be found in the Appendix and we advise the reader to look at these pictures before moving to the "Results and Discussion" section.

The logic Typeform feature is only available under the professional plan. We used Typeform's 3 month free offering for anyone working on COVID-19 related projects that are strictly not-for-profit. Therefore, there is no cost attached with this tool. Hence, 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. An estimated 20,000 invitations with a price of $0.92 CAD per invitation sent and estimated total of $18,400 CAD, together with the cost for printing of $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, followed by a discussion of the data and statistical inferences. The last section is reserved for a weaknesses and future work discussion. 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 three and six. Recall question three asked participants which party did they vote for in the 2019 Federal elections, and question six asks to choose a party if general elections where held tomorrow. Similarly we grouped question four and five, which asked about the general economic direction of our country and the impact of COVID-19 in the participants response to question four. Figures 2 and 3 display the raw data simulated for this report.

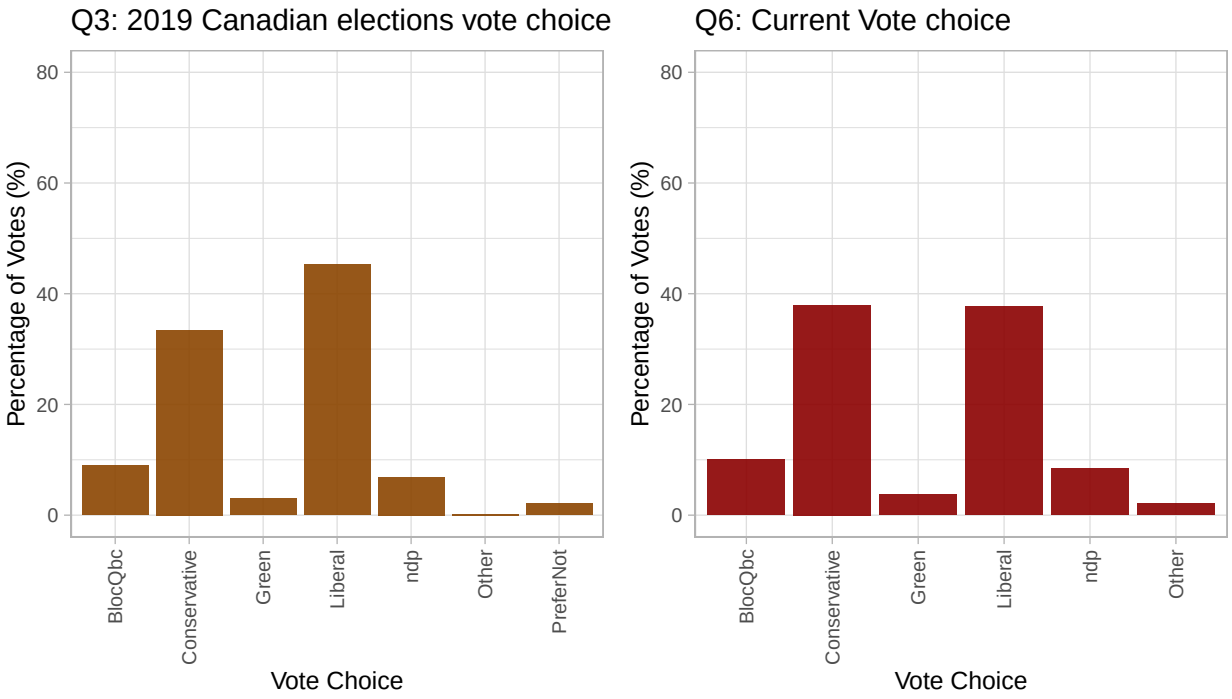

Figure 2: Answers for questions three and six.

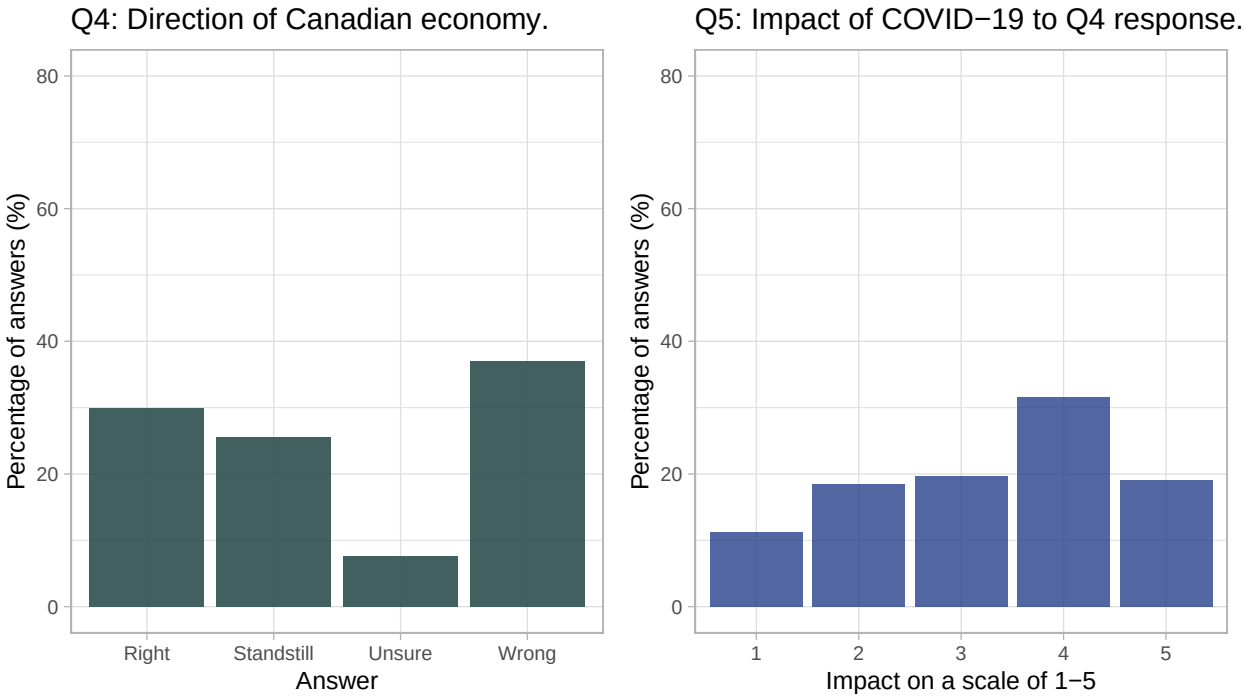

Figure 3: Answers for questions four and five.

Raw data shows that there is an important percentage of respondents that are choosing to vote for another party which is not the Liberal party. Similarly more than half of respondents feel like the Canadian economy is moving in the wrong direction or that is at a standstill. Moreover, approximately half of respondents also claim that COVID-19 had a high impact on their view of Canadian economy. However, this are just the raw results and we cannot make any empirical conclusions based just on these.

Further analysis is required to support what the raw data shows at a glance. To properly estimate such population parameters we first require to calculate an unbiased estimator for the population mean. Consider the stratified sample mean, 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. Sample mean of 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. This is important because, 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

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

here N_h is the total stratum population size. A proof of this statement can be found in Theorem 3.1 in Wu and Thompson (2020) pg. 36.

As per the introduction, the variable of interest is the percentage of votes the Liberal party has with respect to the population. Then compare this number to what the current percentage of votes is estimated to be. Using the methods 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 approximate loss of 6.78% of votes.

Using the same method, we estimate that **33.31%** of Canadians feel the general economy is 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 $[19.84, 47.72]$.

With these results, which support our original claim, we now discuss some of the weaknesses of our approach. The most significant come from the design of the survey. Although it has

been justified the reason for the shortness of our survey, this is also a weakness because it does not provide nearly enough information to study and discuss. Due to various factors we had to limit this study to two main results, the loss of votes and the percentage of people that believe the Canadian economy is moving in the wrong direction. This greatly limited our ability to create stronger conclusions because we are only considering one social aspect. 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 surveys 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 and attempts to draw stronger conclusions. More complex theoretical tools are required for the statistical analysis of such survey.

Appendix

Links

All code used in this project can be found at:

- **https://github.com/cesar-yoab/survey_modelling**

The survey can be found at:

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

We present a set of screen shots of the survey created in Typeform

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.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

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.