Astronomy & Astrophysics manuscript no. output
April 14, 2020

Predicting Québec's Next Election Outcome

Use of Monte Carlo Simulation and Linear Regression in Election Forecasting

Samuel Mottard, 2607442961[1]

Rutheford Physics Building, McGill University, 3600 Rue University, Montréal
e-mail: samuel.mottard@mail.mcgill.ca

April 14, 2020

ABSTRACT

Aims. Statistical tools have historically been used to develop models of election results forecasting. I aim to develop such model using Monte Carlo simulations and linear regression while analysing past elections results.
Methods. The Monte Carlo simulation is based on obtaining the probability of winning an election based on the percentage of votes each party receives. To find the percentage of votes, polls are combined with voter history and weighted.
Results. Forecasts are compared with results from a well-known forecasting website in Québec (Qc125). The comparison shows that simple models can actually predict fairly similar results to more advanced models with many more variables.

Key words. Election Forecasting – Monte Carlo Simulation – Linear Regression

1. Introduction

Political forecasting has existed in one form or the other from the first time an election was held up until today. In the past century, many statistical tools have been developed to try to predict accurately elections. However, with the advent and development of computer science, these tools have been perfected and the data available is better and in greater quantity than ever. Political forecasting is usually done by a model which includes recent polls results as well as voter history. They also include demographics data, such as voters age, income, language of use and birth language, population density, etc. The model that I created combines the first two factors. By looking at voter history and opinion polls, one can try to determine the percentage of votes one party is likely to receive. This can then be used to determine the winner of an election. However, this process is more subtle than choosing the party with the highest percentage of votes. In Canada and many other countries, the political system in place is a representative democracy. Leaders are not determined by number of votes, but by number of won constituencies. Each constituency is itself won by number of vote for each candidate in that constituency. Considering this, a party can be elected and have a lower percentage than its adversary. Historical examples of this are numerous. In the United States, one can look no further than the elected president at this time, Donald Trump who lost the popular vote but was elected president in a system similar to that of Canada. The Canadian 2019 election saw the Conservative Party win the popular vote while still loosing the election. In order to be able to accurately give the probability of each party of winning the next Québec election, I built a simulation that takes into account our political system.

2. Linear regression from past elections

In order to quantify the next election's proportions of vote, part of he model analyses voter history. By looking at previous elections from 1998 to 2018, percentages of votes for each major party were gathered. During last elections, the four major parties were:

- Parti Libéral du Québec (PLQ)
- Parti Québécois (PQ)
- Coalition Avenir Québec (CAQ)
- Québec Solidaire (QS)

For the purposes of the model, all the parties that are not listed do not have any elected representatives and are combined in another group called "Other". Québec Solidaire participated in its first election in 2007, so for this party, only data from that year onward is used.

A few ways of using this data were considered. The first method is to simply average the percentages of votes over the whole period. A period of 21 years (from 1998 to 2018) was chosen arbitrarily, but considering that any longer period would include data that is not meaningful, as parties and voters change. However, this method is very simplistic and it is not expected that would yield good results when comparing final predictions to other algorithms.

A second method uses linear regression to observe the trends in votes across all parties over the same period of time. The regression allows us to calculate the percentage of votes for the year 2022. Because they are calculated using a linear function, percentages do not amount to 100. Therefore, results are then normalized to do so. This method gives more sensible data. The year 2022 is assumed for the next elections, because typical mandates last four years when the elected party holds the majority of seats and this is the case with the CAQ elected government. However, it does not take into account recent trends effectively, and so a third method is chosen.

The third also uses linear regression. However, each year is weighted according to how far from the year 2022 it is. The year 1998 is weighted as 1/25 or 4% of its percentage, the year 2003 as 6/25 or 24% and so on. From the weighted percentages, a

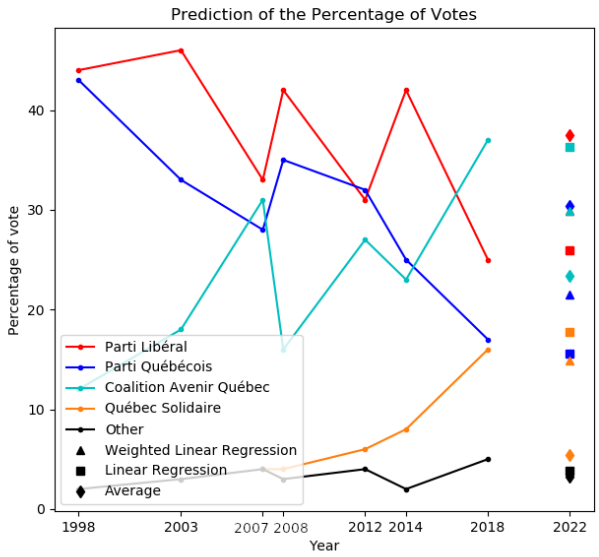

Fig. 1. Voter history for the four major parties and predictions according to the three different models.

linear regression is performed and the value of this function is calculated for the year 2022. As before, the sum of the percentages is not 100 and so they are normalized. This is expected to yield the best results out of the three models because this reflects better the recent voter trends.

Political Party	Method Used		
	Average	Linear Regression	Weighted Linear Regression
Parti Libéral	37.5%	25.9%	29.9%
Parti Québécois	30.4%	15.6%	21.5%
Coalition Avenir Québec	23.4%	36.3%	29.9%
Québec Solidaire	5.4%	17.7%	14.9%
Other	3.3%	4.4%	3.8%

Table 1. Predicted percentage of vote for every major political party assuming different methods.

3. Combining factors

The two factors used in this model are voter history and polls. The poll used in the model is the latest available by the firm EKOS Politics(1). Each factor is weighted equally and so, the final percentage used for the simulation is given as:

$$\text{Percentage} = 0.5 \times H + 0.5 \times P \quad (1)$$

Where H is the percentage obtained from analysing voter history and P is the percentage from recent polls.

4. Monte Carlo Simulation

A Monte Carlo was created to determine the probability of a political party being elected. The percentages from Table 1 are used with poll results from EKOS and combined with equation 1 to obtain the percentage of vote for each party. These percentages were input into a simulation, where for each run, 125 Gaussian distributed random numbers for each party were picked. 125 numbers are picked because this is the number of constituencies in the province of Québec. The numbers are centered around the percentage for each party. In order to determine the standard deviation of each Gaussian array, results from every constituency during the 2018 election were gathered. For each party, a list of 125 percentages was made and the standard deviation of this list was used as the standard deviation of the Gaussian random array for this party. Each of the numbers in the array was associated with a number from all four other arrays, as to simulate results of polls in a particular constituency. The highest random number in each of these represents a victory for the party to which it corresponds to in that constituency. Adding up all the victories allows us to determine which party is the winner of that run. Numerous runs are done and we can determine the probability that a party wins the elections.

5. Results

The way used to judge the accuracy of the model is to compare results with the Qc125 website(2). This organisation has developed a complex model for forecasting elections and is often cited in medias. Previous predictions of the Qc125 model demonstrate its accuracy and reliability.

Running the simulation for a 1000 steps with any of the three voter history methods discussed yields a 100% probability for the Coalition Avenir Québec to win the next elections (see Figure 2. While this may seem extreme, these results are in accordance with recent estimates according to the Qc125 website. This is because recent polls show very strong support for the current CAQ government. Obviously, opinions change and the poll used in this simulation has to be updated on a regular basis for the most accurate predictions.

A side property of the simulation is that it allows us to simulate how many constituencies are won by each party. Considering current trends, this part is actually more interesting. For instance, the averaging model (see Table 1) expects the CAQ to win 72.1 constituencies and the PLQ to win 34.6. Comparing this to the current Qc125 prediction, we see that the number for the CAQ is underestimated (91.1 ± 11.4) and the number for the PLQ is overestimated (20.7 ± 7.4). Other parties are also far off their expected values. Numbers from this model do not fall within the margin of error provided by the Qc125 website. This result is expected, because the PLQ has had strong voter support in previous elections before 2018.

The linear model on the other end overcompensates for the first model. The CAQ prediction (102.4 constituencies) is now overestimated, although at the very limit of the margin of error from Qc125 and the PLQ is underestimated (15.8 constituencies) within the error margin. The PQ obtains 5.5 constituencies against 6.4 ± 5.5 for Qc125 and QS obtains 1.3 against 6.9 ± 2.2. While it does not represent well the current trends, most numbers are within the error margins and it is overall a better model than the averaging one.

The weighted linear model shows the strongest results when comparing to Qc125 (see Figure 2). Indeed, the predicted number of constituencies is 90.9 for the CAQ, well within the error margin, 22.8 for the PLQ, also much closer than previous models, 10.0 for the PQ, also within error margins and finally 1.3 for QS, which is the same the result from the linear model. This shows that this models understands way better the current trends in politics than previous ones. In particular, it shows that many voters who voted for the PLQ are now more inclined to vote for the CAQ.

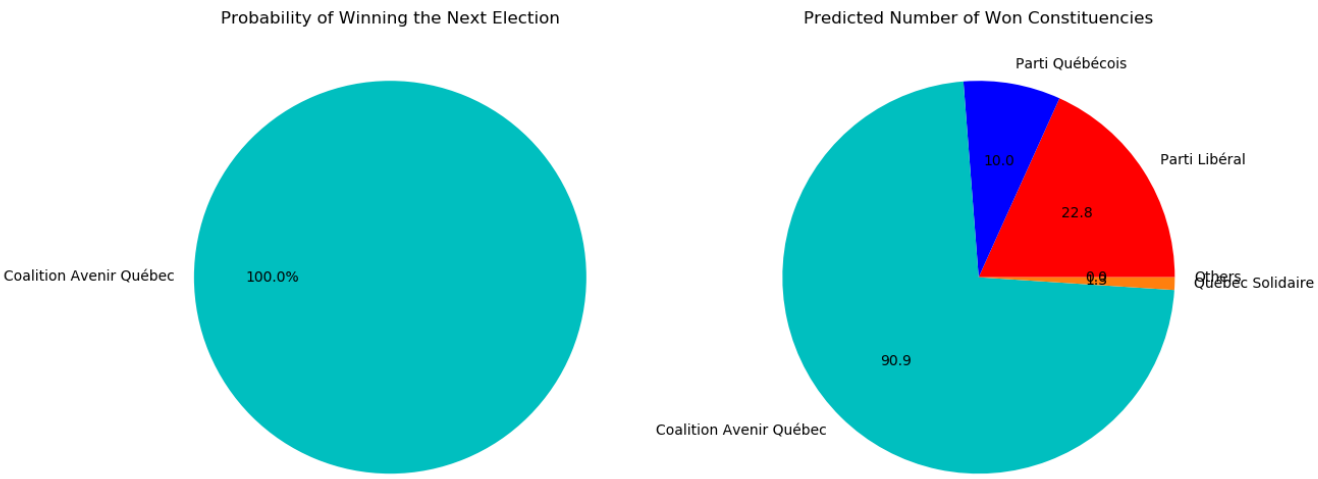

Fig. 2. Prediction of next election results. A 100.0% probability for the CAQ of winning next elections based on the current data is in accordance with what current trends show.

Finally, another way to test the model is to try to recreate the results from the 2018 elections by inputting the actual percentages obtained and see how the simulation behaves. During the 2018 elections, the CAQ won 74 seats, the PLQ won 31, the PQ won 10 and QS won 10. The model expects the CAQ should have won 80.9, the PLQ should have won 28.6, the PQ should have won 11.6 and QS should have won 3.9. While this is not a bad performance, it shows that the model needs improvement, especially considering Québec Solidaire. In all the runs performed, no matter the data, the performance of this party has been underestimated. More appropriate weighting could be a way to solve this problem, but including demographics data is suspected to be a better method to solidify the model.

6. Conclusions

1. The use of Monte Carlo in political forecasting proves to be a suitable tools with surprisingly good performances for even a simple model.
2. Averaging previous results is not a good way to quantify voting history and simple linear fitting shows little improvement over this method.
3. Weighted linear fitting is actually a better method to quantify voting history.
4. Polls can drive models but they change according to the political climate. A very strong model would have a better way of weighting their influence.
5. Including other factors such as demographics would definitely improve the model. Other factors worth considering are "star candidates" such as party leaders and high profile candidates and social circumstances such as the COVID-19 crisis, although these factors tend to affect polls.

Acknowledgements. Thanks to Professor Adrian Liu for his feedback and teaching.

References

[1] EKOS Politics, COVID-19 Produces Pervasive and Strong Anxiety, March 27, 2020, (Ottawa, Ontario), http://www.ekospolitics.com/index.php/2020/03/covid-19-produces-pervasive-and-strong-anxiety/

[2] Fournier, Philippe J., Qc125.com - Sondages, analyses et projections électorales. 2016 (Montréal, Québec), https://qc125.com/