Project CMAP: Flows of voters between election rounds: What happend in the Chilean Election of 2017

Juan Vila

12/9/2019

Introduction

In the chilean second round election fo 2017, most to the analyst were expecting a very thight race because the results in the first round of the election were not definitive the current president got only 36% of the votes and the center-left and left candidates between them add up to 43% of the votes. Nevertheless, the result was astonishing the current president Sebastian Piñera won with 54% of the votes and particularlly the turnout rate between the rounds increase in a 4%. There has been different hypothesis about this result. First, the increasing in competitivity of the race. Second, how the right wing manege to use different discourses to mobilize voters.

The Chilean presidential election rules are a majoritarian one, with second round election if any of the candidates do not achieve 50%+1 of the votes. Particularly, we are interested in see how they behave between the second round and first. This is because, in the last election (2017) we see that turn up increase between rounds, and in the 2013, election was in the other way around. This imply that candidates can try to de-activate or activate particular voters for certain elections.

The objective of this paper is to analyze how the voters in the Chilean Presidential election of 2017 behave between the first and second election round, especially analyzing the how they move their votes between first and second round. With especial focus what happend with the people that decide to vote only one of the elections. Trying to identify how the votes transfer from candidate in the first round to the second round.

For do this we use electoral data and voter file demographics. We use to methods to try to untangle this questions, first we use unconstrained regressions of the three status of the second round (Vote for Sebastian Piñera, Vote for Alejandro Guille or not vote) and constrained regression where force the estimator to add up to one. We found eveidence of how the different flow of voter move between elections and we see that 630 thousands new voters that do not had vote in the first and vote in the second round. Nevertheless, was not possible to asert which of the hypothesis provoke this results.

The document have the following structure. In the first part we speak about the context of the election, describing the history of the second round rule, when have been applied and the particular context of the election of 2017. Then we develop the principal hypothesis that explain the fenomena in the 2017 election. In the next section we discuss the methodology. Later, the results of the two methodologies. Finally, the conclusions and the limitations of the approach taked.

Context

Chilean Presidential elections

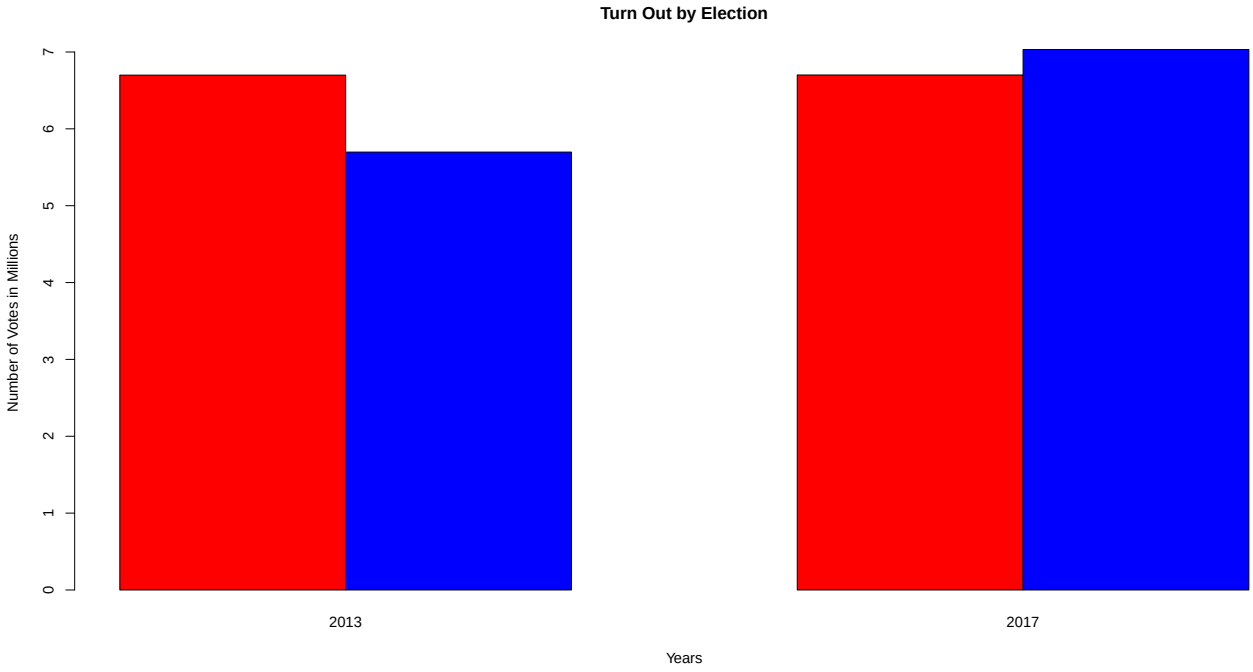
Historically, the Chilean election scheme for the presidency had been a majoritarian system but had different rules according what happened when the neither of the candidates get the absolute majority. There had been different rules for solving this issue. For example, in the 1925 constitution on that case says that in this situation the congress will elect the winner. In the other hand, in the current constitution this issue is solved with a second round.

Since 1990, these rules have generated that the presidential elections with different results related to the use of second round. From 1990 until the 2000, the two presidential elections in that period elections were

resolved in first round. From, 2000 until now is complete opposite there be always a second-round election, which is a very similar issues if we compare with the French election that have also that characteristic.

Normally, the results of turnout-rate between election was that the turn out rate diminish a little related to the round. Which is explained how the registration rules change in 2009. Before the registration in the voter register was voluntary but the vote was mandatory. After 2009, the system changes to a system were the registration was automatic and vote is voluntary. This change creates an incentive to try to mobilize groups between rounds. Seeing if some particular vote of people will be activated in one particular round, this create variance in the turn-out rate between rounds.

We can see that the results of the 2013 and 2017 elections are quite different. In the 2013 election, we can see that the people that comparing the first with the second round election we have a diminish of the 15%. One explination for this result is the lack of competition of the second round presidential election in that time, due to, in the first round election the former president Michelle Bachelet won the first round with 46% of votes, been followed by the center-right candidate with 25% of the votes. Then in the second round Bachelet won easly with a 62% of the votes.



In the 2017 election, the history is quite different we have a more competitive election that in 2013. In the first round the current president Sebastian Piñera got 36%, but the center left candidate got 22% of the votes. This is explained because there where 2 strong candidates that were outsiders in either the left and right parts of the political spectrum. In the case of the left, was created a new coalition called Frente Amplio, the presidential candidate of this coalition Beatriz Sanchez got 20% of the votes. In the other side of the political spectrum, a far right candidate Jose Antonio Kast got 10% of the votes.

The election have a total of 8 candidates, from the all political spectrum from the far right or alt-right, until far left candidates. As we discuss previously we can see that SP got 2.4 MM in the first round which is 36% of the votes. In second, place the candidate was Alejandro Guille, he was the official candidate of the Nueva Mayoria coalition that has won the election in 2013, he got 1.5MM votes. In third place, appear the candidate from a newly createad left wing coalition called Frente Amplio, Beartiz Sanchez(BS) got a total of 1.3 MM votes.We can see that the far right and center right candidates add up a total of 43.9% of of the votes, but the rest of the coalitions from the center unti the far left add up to 54.5% of the votes, then with this result was expected that would be a very tigh race in the second round. One important think to notice is that the candidates,Carolina Goic, Alejandro Guille and Alejandro Navarro cames from parties that where part of the goverment of Michelle Bachelet and the coalition Nueva Mayoria.

Table 1: Results first Round

Candidates	Political_Ideology	NVotes	Por.
Carolina Goic(CG)	Centrist	386396	5.78
Jose Antonio Kasta(JK)	Far-right	521962	7.81
Alejandro Guille(AG)	Center-Left	1490532	22.32
Sebastian Piñera(SP)	Center-Right	2409922	36.09
Beatriz Sanchez(BS)	Left	1331191	19.93
Marco Enriquez Ominami(MEO)	Center-Left	375769	5.62
Eduardo Artes(EA)	Far-Left	33468	0.50
Alejandro Navarro(AN)	Left	23880	0.35
Blank Votes		64858	0.97
Annulated Votes		39314	0.58
Total		6677292	100.00

Nevertheless, we can see th in following table the second round result, we can see that SP have an ashtonishing result getting 54% of the votes. There is a rule of thumb for the second round elections, that is if the winner of the first round get that 40%, normally it would be beatable in the second round. Whith this set up much people was thinking that the second round would be very tight. Nevertheless, was the other way around, Piñera won with a 54% of the votes. This open the question how this happend, there is two possible hypothesis about what happend, which are analyzed in the next section.

Table 2: Results Second Round

Candidates	NVotes	Por.
Alejandro Guille(AG)	3147142	44.88
Sebastian Piñera(SP)	3788302	54.03
Blank Votes	56338	0.80
Annulated Votes	19421	0.27
Total	7011203	100.00

Posible Hypothesis.

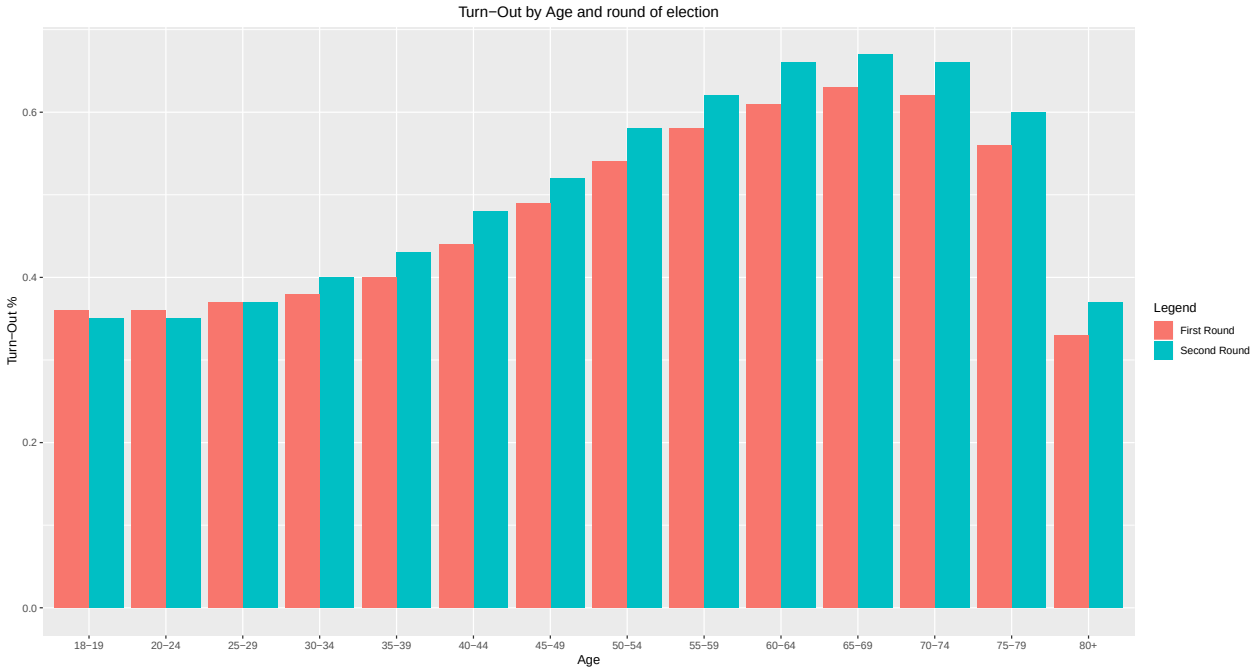
First, an important element to take into consideration in the Latin-American political context, is the polarization between the leftwing and right wing government in a continental level. Shifts in the goverments ideology influence the current political debate in other countries in the regrion. For example, the current elected president of Uruguay before the start of riots in Chile, says that Uruguay must became like Chile or in the Chilean case where the current president for the second round election use the concept of "Chilezuela" as an element to mobilize population to vote in the second round. This in terms of electoral data is related to possible shift of people that do not vote in first round and vote in second round.

Second, element is related to the competitiveness of the election, as the election is more interest more people want to participate. Impliying that this explination could help to explain the difference between what happend in 2013 with 2017. The first election was basically decided in the first round election, actually in that time was the idea that maybe Bachelet was going to win in first round.

One way to think about how to differenciate this two hypothesis is the following: if on average all people decides not to vote between rounds meaning that if the probability diminish the equal for every person independent of the age,gender, political preference, income, etc. We could argue that is related to a competetiveness issue if all people on overall decide not vote maybe they don't find interesing enough. In other hand if we start to see that certain groups start to increase the turn-out it's possible to think that there is issues that move people that allow to certain groups increase or diminish their turn-out. Even we could think in institutional

elements or during the campaing to incetivze/de-incentivize groups to vote.

One way to star to shed light to this problem is to see the distribution of the result through the demographics characteristics of between age. In the following graph we can see how the distribution of voters change between round specially for old people that they increase their participation in the election. This reenforce the idea that certain discourse targeted to people with more conservative world view could had affect the electorate.



Data

We are going to use three data sources:

1. Voter File
2. Electoral results

The voter file have 14.3 million rows corresponding to every voter in Chile. The information that it have is only the name, gender, address, comune where she lives and the ballot box that she have to vote. This don't give much information but allows to you link some demographic information of the voter file with the electoral results which is usefull.

Finally, the electoral results are at the level of ballot box, they are around 43,000 ballot boxes results for each election. Meaning that we don't know if a person vote in a particular election, only we have the result of the ballot box were she is registered to vote. Also, there is information in which part of the country is located and this mean that potentially we can link it with economic data by comune.

Methodology.

We are going to work with two different types of approach for triying to figure out the answer.

First approach is to run regression at a ballot box level, always running three equations. First, the number of people that do not vote in second round, the number of votes that the winning candidate Sebastian Piniera and the number of votes of Alejandro Guille. The basic model that we are going to run is:

$$V_c^B = \alpha^B + \sum_{j=1}^{n} \beta_j^B 1stRoundResults_j^B + \sum_{i=1}^{n} \gamma_i^B DemVars_i^B + \epsilon$$

Where 1st Round results are the results of first round of all j candidates in the ballot, Demographics variables depends in which level we are working, but in the case of the ballot level we are going to control by the average age in the ballot box and porcentage of men registered on it, the upper script reflect the level of the data.

In the second approach, what we are going to do is estimate a similar model but with out the demographic variables constrain the values of the estimator betwee 0 and 1, and estimate for two equation simultaneously and for complement estimate the third one.Such we can get a possible decomponsition of how are the transition of votes between round. Nevertheless, this is not a perfect identification of the system because we need as number of equations that we estimate the same numbers of variables to identify the system. This is a limitation of the approach in general.

$$Min(V_c^B - \sum_{j=1}^{n} \beta_j^B 1stRoundResults_j^B)^2$$

$$st. \sum_{j=1}^{n} \beta_j^B = 1$$

$$\beta_j^B > 0$$

$$\beta_j^B < 1$$

Results

Fix effect OLS models

We can see if we ran the models with fix effect estimation by columns to try to take into account elements of the comunes where the ballot box are located.

One of the main issue is sections is how to interpretate this regression. Ideally we would like to have estimators between 0 and 1, thats why we build the restricted model in the next section.

First the clasical interpretation of a OLS regression, in the case that are negative or postive that indicates the correlation, the ballot boxes where the candidate X votes increase in one vote have a z(coef. value) effect in the votes of SP, AG or the non voters of the second round.

Second, which is the complicated is how close is that over to one or not. Because as we are working with aggregate data we loss some internal correlations. For example, if a estimator is bigger than one can be explianed that people that don't vote in the first round. If they had voted would like to vote for the candidate in that estimator and they vote in second round. In other words, people that vote only in second round if they had voted in the first would had voted for candidates which estimator is way over one. Another reason thar can explain this is because we maybe we have some nonlinearities that provoke that the values where negative or over one.

We are going to start with models that regress the second round results with all the ballots votes of the first round. We can see that the parameters for the equation of non participation in the second round have the most quantity of parameters estatiscally equal to zero, for candidates GC, JK, Artes and Navarro. This could imply that most their voters went to vote into the second round. The other interesting thing is that most of the parameters sum closely to one.

Now we analyze every variables of the model into the dependent variable of the threemodels.

- First Round non Voters.

We can see that the relation between the quantity of people that not vote in the first round and the people that not vote in the second, vote for SP and AG in second round is .92, .05 and .02, is interesting because we can see that basically that in the ballot box people if they report an extra non voter, increase the non voters in .92 votes, which imply in certain way that some people that vote in second round have not vote in the first one. In the case case of the other two estimators, we can see that in case of the votes of second round voters, the candidate SP capture part of the non voters of the first round that AG, approximately 3 times more.

- Carolina Goic voters.

In the case of votes of the candidate CG, we can see that where was votes of her do not increse the abstention on second round. In the case of the votes of SP in the second round just one third go to SP and the other third went to AG. It's important to notice that she was from the same coalition of goverment of AG.

- Jose Antonio Kast voters.

Now in the case the candidate JK we can see that most of him votes went to SP, because the estimator was .95, 2% went to non voters of second round and AG respectivly.

- Sebastian Piñera voters.

We can see that in this case all estimator sum closely to one as the other ones, but in this case we have values over one and negative values. This is explained because we are running his votes in first round and the second round which are highly correlated and we can see that he also camptures votes of non voters population of first round. Additionally, some of his votes according to this regression go to AG, which is very non-likey, but the interesting element is the negative sign in the regression for second round non voters, where he reduce the amount of abstention in second round by .11 votes by each vote that he got in first round. Which is consistent with result of the in the equation of the non voters in second round. In other words, if the sign of SP in this equation was positive and the sign of non voters of first round of was also positive would not made sence.

- Beatriz Sanchez voters.

We can see that .27 of his votes went to abstention, the other .12 went to SP and finally around .58 of her votes went to AG. This idelogically make sence. Since, she is the candidate from a leftist coalition that the voters historically tend to vote by the old version of the Nueva Mayoria coalition called Concertacion which govern the country from 1990 until 2009.

- Other variables.

In case of MEO we can see that around .11 of his votes went to asbtention, .31 to SP and .57 to AG. In the case of Eduardo Artes we can see that most of the votes went to AG in second round, which is consistent with the ideological position of the candidates. In the case of AN, we can see that around most of the votes went to AG, but the effect over the abstention in the second round is negative meaning that is posible that some of first round potential voters have increase the numbers of voters. Finally, we can see the blank and annuled votes, the results in this case are strange because they do not add to one.

In the case of the model with the demographics variables we decide to use squares in the age variable because we know that the particiption is non lineal acording to the graph that we present in the first part. The result that we can see is that in general male voters tend to participate less than woman in elections, because we have a positive relation between the porcentage of men that are registered in the ballot box and the abstention in the second round election, this exactly the oposite in the case of the results of the candidates AG and SP votes in second round. In the case of age we can see that for both model of the candidates AG and SP we see that have a inverse U shape relation and with very similar estimators numbers for the square term. The contrary is true for the case of the non voters in second round that they have a U shape effect.

In the case when we merge the previous models we found that most estimators are the same. Nevertheless, we notice some differences. For the demographic estimators we can see that their value diminish in the case of the age estimators, but in the case of the porcentage of men we can see that the effect of the gender over the votes of SP in second round are now negative, meaning that controling for the votes in the ballot we can see that now men are less likly to vote for SP in second round. For the political variables we can see that biggest change that we see between the models presented in table 3 and 5 is for the votes of the candidate

	Model for non-votes	Model for Votes of SP	Model for votes of AG
novoto_pv	0.94*	0.05*	0.01*
	[0.94; 0.94]	[0.04; 0.05]	[0.01; 0.01]
goic_pv	−0.00	0.32*	0.66*
	[−0.03; 0.02]	[0.30; 0.34]	[0.64; 0.68]
kast_pv	0.02	0.95*	0.02*
	[−0.00; 0.04]	[0.94; 0.97]	[0.01; 0.04]
pinera_pv	−0.11*	1.06*	0.05*
	[−0.12; −0.10]	[1.05; 1.06]	[0.04; 0.05]
guillier_pv	−0.20*	0.05*	1.14*
	[−0.21; −0.19]	[0.05; 0.06]	[1.14; 1.15]
sanchez_pv	0.27*	0.12*	0.58*
	[0.26; 0.28]	[0.12; 0.13]	[0.57; 0.59]
meo_pv	0.11*	0.31*	0.57*
	[0.09; 0.14]	[0.29; 0.33]	[0.55; 0.59]
artes_pv	0.02	0.06	0.86*
	[−0.07; 0.12]	[−0.01; 0.13]	[0.79; 0.93]
navarro_pv	−0.06	0.24*	0.79*
	[−0.16; 0.05]	[0.15; 0.32]	[0.71; 0.87]
nulos_pv	−0.49*	0.15*	0.22*
	[−0.54; −0.44]	[0.11; 0.18]	[0.18; 0.25]
blanco_pv	−0.26*	0.04*	0.18*
	[−0.30; −0.22]	[0.01; 0.07]	[0.15; 0.21]
R^2	0.92	0.88	0.85
Adj. R^2	0.92	0.88	0.85
Num. obs.	42890	42890	42890

* 0 outside the confidence interval

Table 3: Models with political variables for every possible status in second round with electoral variables

	Model for non-votes	Model for Votes of SP	Model for votes of AG
porc_men	12.54***	−17.06***	−8.41***
	(0.75)	(0.46)	(0.43)
av_age	−11.71***	6.25***	5.32***
	(0.09)	(0.06)	(0.05)
av_age_sq	0.12***	−0.06***	−0.06***
	(0.00)	(0.00)	(0.00)
R^2	0.31	0.28	0.22
Adj. R^2	0.30	0.27	0.22
Num. obs.	42890	42890	42890

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 4: Models for every possible status in second round with demographics variables

	Model for non-votes	Model for Votes of SP	Model for votes of AG
novoto_pv	0.92^{***}	0.06^{***}	0.02^{***}
	(0.00)	(0.00)	(0.00)
goic_pv	0.01	0.31^{***}	0.65^{***}
	(0.01)	(0.01)	(0.01)
kast_pv	0.03^{**}	0.91^{***}	0.06^{***}
	(0.01)	(0.01)	(0.01)
pinera_pv	-0.07^{***}	1.05^{***}	0.01^{***}
	(0.00)	(0.00)	(0.00)
guillier_pv	-0.09^{***}	0.04^{***}	1.05^{***}
	(0.01)	(0.00)	(0.00)
sanchez_pv	0.15^{***}	0.10^{***}	0.72^{***}
	(0.01)	(0.00)	(0.00)
meo_pv	0.17^{***}	0.26^{***}	0.56^{***}
	(0.01)	(0.01)	(0.01)
artes_pv	0.06	0.14^{***}	0.75^{***}
	(0.05)	(0.04)	(0.03)
navarro_pv	0.06	0.28^{***}	0.64^{***}
	(0.05)	(0.04)	(0.04)
nulos_pv	-0.48^{***}	0.17^{***}	0.19^{***}
	(0.02)	(0.02)	(0.02)
blanco_pv	-0.23^{***}	0.05^{***}	0.13^{***}
	(0.02)	(0.02)	(0.02)
porc_men	-3.82^{***}	-3.23^{***}	6.76^{***}
	(0.26)	(0.20)	(0.20)
av_age	-2.16^{***}	0.93^{***}	1.30^{***}
	(0.04)	(0.03)	(0.03)
av_age_sq	0.02^{***}	-0.01^{***}	-0.01^{***}
	(0.00)	(0.00)	(0.00)
R^2	0.93	0.88	0.86
Adj. R^2	0.92	0.88	0.86
Num. obs.	42890	42890	42890

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 5: Models for every possible status in second round with demographics+electoral variables

BS, where we can see that the estimator are .15, .10 and .72 for the three equations in table 5. In the other hand, we can see that the estimate in table 3 are .27, .12 and .58, which tell us that the result of BS is highly related to the demography of the voters. This make sense because as we explain previously we know that

Before to advance to the Constrained OLS strategie is important to discuss that even if the number make sence as consistent approach for that we would requiere some identification strategy that we are currently missing, in the next section we are going to try to solve simultaneously equation for AG and SP and with those results for construction resolve the system for the second round abstention.

Constrained OLS

In this section we present the results of the Constrined OLS estimate, basically as we show in the previous part is try to jointly minimize the error of the

First we can see the result of the constrained optimization of the SP equation, we can see that people that not vote in first round add up to SP 5%, 35% of GC votes go to SP, Kast all the votes practically, same as himself, AG coeficitient is not significant, from BS voters the 10% go to SP, from MEO 27% of him votes

go to SP, Artes 0% and from Navarro voters 21% go to SP. This show that SP manege to get 5.9% of the persons how did not vote in first round to vote in second round for him.

Table 6: Equation for Second Round for SP

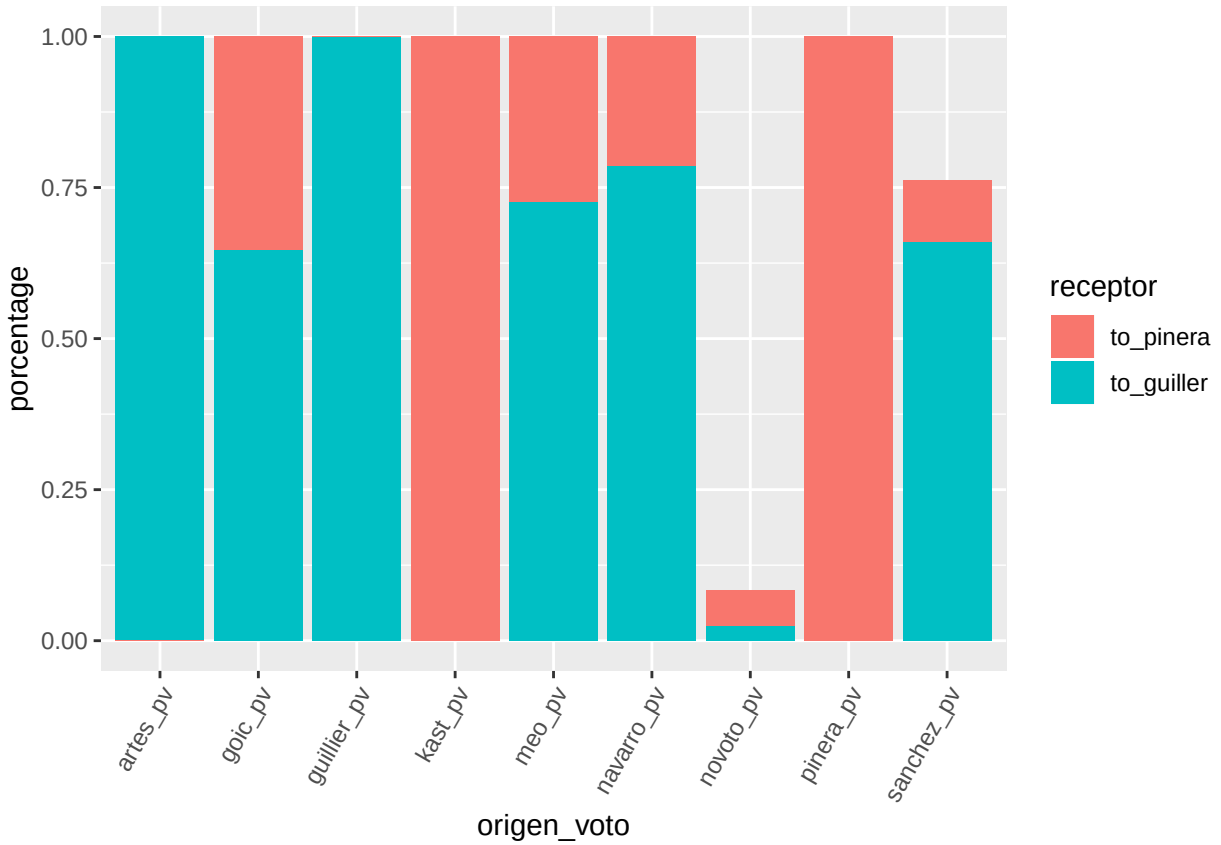
Variable	Coef	SE
novoto_pv	0.0591708	0.0000000
goic_pv	0.3527198	0.0096233
kast_pv	0.9999983	0.0065700
pinera_pv	0.9999941	0.0017718
guillier_pv	0.0000009	0.0033147
sanchez_pv	0.1010549	0.0026577
meo_pv	0.2745380	0.0108861
artes_pv	0.0000000	0.0439660
navarro_pv	0.2150549	0.0413061

Now if we see the same for AG, we can see that from non voters of the first round he manege to get 2.4% of the total of person that do not vote in first round.From GC voters 64% vote for him in second round. From Kast and SP zero. From himself and Artes basically all. From BS 2/3 of the voters went for him in second round. From Meo 72% of his voters went for AG. Finally, Navarro's voters only 78% voted for AG in the second round.

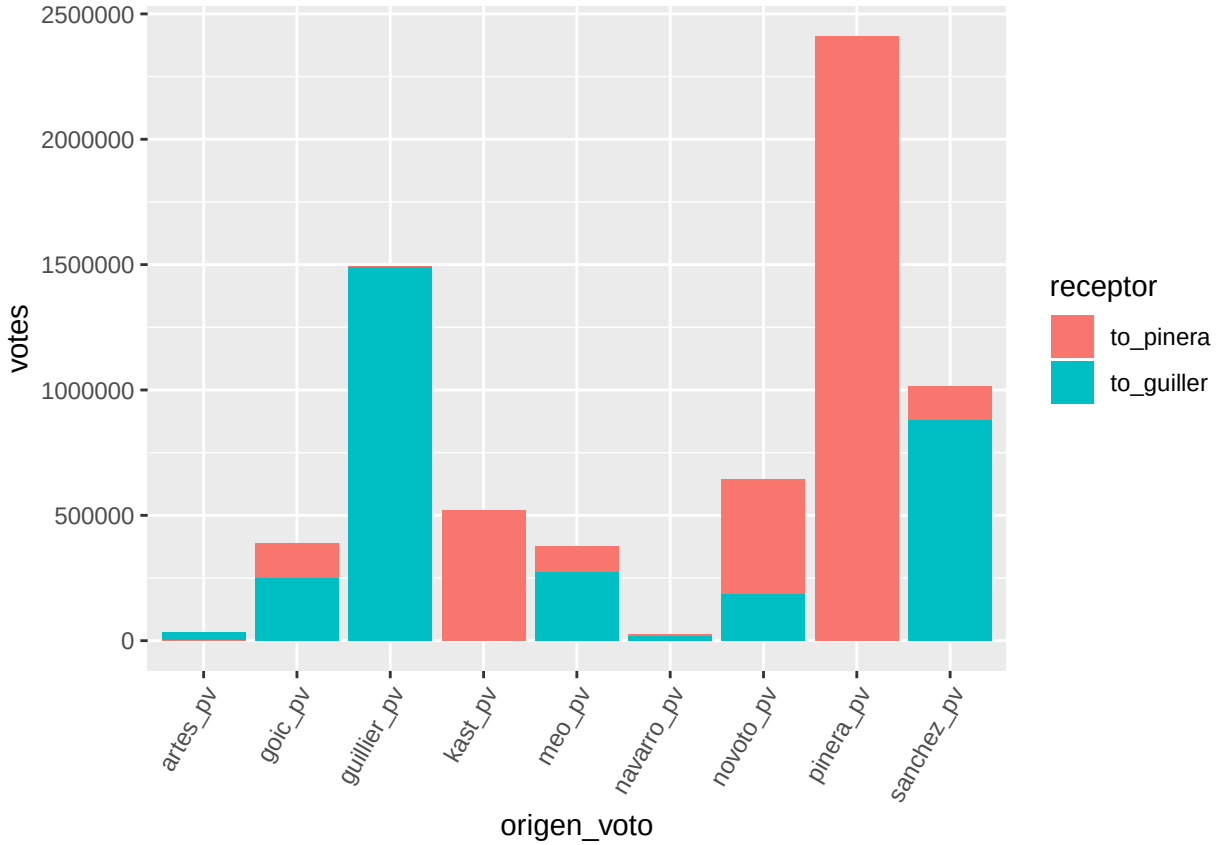
Table 7: Equation for Second Round for AG

Variable	Coef	SE
novoto_pv	0.0243149	0.0000000
goic_pv	0.6472797	0.0090934
kast_pv	0.0000008	0.0062082
pinera_pv	0.0000025	0.0016742
guillier_pv	0.9999974	0.0031322
sanchez_pv	0.6607903	0.0025114
meo_pv	0.7254615	0.0102867
artes_pv	1.0000000	0.0415450
navarro_pv	0.7849450	0.0390316

In the follwing graph we see what resume the previous tables. If they no add up to one is the persons how decide not vote in second round. In the case of BS(the last column) implies that they are voters of hers that decide not to vote in second round. Is strange this result, that probably es related a the strategy utilized to identify the decomposition of voters between rounds.



In this third graph we see the result instead of porcentages in total of votes, we can see how the vote of SP is constructed in the second round throught votes of all candidates expect Artes.



One problem with this methodology that is dificult to decompose the vote for type of demograpy. For example is difficult to know how this distribute by age. We could try to rebuild the great aggregates throught microdata and then re run this regression. Another issue is why it's concentrate all the non voters in second round that voted in first round in the candidate BS, this could explain only by bad motives, if we see that the uncontrain regression of the previous part the highest paramater value of the equation for non vote in second round is for the candidate BS, this tell us that the algorithm for constraining the data probably it's not optimizing correctly, because concentrate all the result in one single variable. One of the reason why we choose only to run jointly two of three models is to ease the convergency, but we have a trade off, that is that the optimization process decide to go in the more simple local optima.

Conclusions and discussion.

Our results tell us an history where we can see how the first round voters decide what to do in the second round election. We have see that there is some evidece that voter for a particular candidate do not garantee some kind of ideological loyalty. This tell us that is some heterogeneous preferences of voters. For example, we can see that an important part of the Beatriz Sanchez voter went to vote for Sebastian Piñera. Alse, we see that in overall election we have flow of new voters in the second round is 636.000 new voters and around 300.000 voters that vote only in the firs round election, showing that is some espace for campings in this type of election to atract or new voters.

Nevertheless, is importan to consider some limitations in the analysis. Fist, we are working with aggregate data meaning that is difficult how interpret the data and the results of this regression, we have to take into consideration the posibility of ecological fallacy. Second, the lack of free variables for identify the system of equations of the OLS part, ideally we would like to estimate jointly estimate the three equations, but for doing that we need three variables that help to identify the equation system correctly. But as we know that variable are difficult to come bay and we need more elemtens that probably at this level of aggregation are difficult to get. Third, in the case of the constrained regression we have see that is very strange that from

all the voters that do not vote in the second round that voted in the first one are concentrated only in her voters. What is proof or evidence that the algorithm choose the candidate which voters have the highest estimator in the unconstrained version, this open a discution that maybe the constrain force to take higher values that should be.

In other hand, some of this problems could be solutionated at least parcially. In the case of looking for free variables that can help to identify the effect we could exploit elements as the expenditure of the campaing in the geografic area or number of visits to that comune of the candidates or if the speeches point to certain demographics throughr social network, the importan point is to look what differenciate the strategy of the candidate and how to build the variable that reflects that. In the case of the constrained regression problem, maybe could be solved if is implemented throught a ridge regression.

Even with this consideration we can see from the unconstrined analysis that there are patron and behaviour of the voters of the different candidates that at least even taking into consideration the limitations in the interpretation with agrgragte data over individual beahaviour. We see certain mobilization of voters to different candidates and this flow of voters from one to another. Nevertheless, we think that we do not have enough infromation to see if which hypothesis prevail For doing this requiere go to analyze the variance of this result to Comune and demographic level that this kind of information do not allow to do.