

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

*Juan Vila*

*12/9/2019*

## Introduction

In the Chilean second-round election of 2017, most of the analysts were expecting a very tight 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 added up to 43% of the votes. Nevertheless, the result was astonishing: the current president Sebastian Piñera won with 54% of the votes, and particularly the turnout rate between the rounds increased by 4%. There has been a different hypothesis about this result. First, the increase in competitiveness of the race. Second, how the right-wing managed to use different discourses to mobilize voters.

The Chilean presidential election rules are a majoritarian one, with the second-round election if any of the candidates do not achieve 50%+1 of the votes. Mainly, we are interested in seeing how they behave between the second round and first. Because, in the last election (2017), we know that turn-up increases between rounds, and in 2013, voting was the other way around. Implying that candidates can try to de-activate or activate particular voters for specific elections.

The objective of this paper analyzes how the voters in the Chilean Presidential election of 2017 behave between the first and second election round — especially analyzing how they move their votes between first and second round. With particular focus on what happened with the people that decide to vote only one of the elections. Trying to identify the transfer of votes from candidates to another.

For doing this, we use electoral data and voter file demographics. We use methods to try to untangle these questions. First, we use unconstrained regressions of the three statuses of the second round (Vote for Sebastian Piñera, Vote for Alejandro Guille or not vote) and constrained regression where we force the estimator to add up to one. We found evidence of how the different flow of voter move between elections, and we see that 630 thousand new voters that do not have a vote in the first and vote in the second round. Nevertheless, it was not possible to assert which of the hypotheses provoked these results.

The document has the following structure. In the first part, we speak about the context of the election. Describing the history of the second-round elections of the sXX. When have been applied and the particular context of the election of 2017. Then we develop the principal hypothesis that explains the phenomena in the 2017 election. In the next section, we discuss the methodology. Later, the results of the two methods. Finally, the conclusions and limitations of the approach taken.

## Context

### Chilean Presidential elections

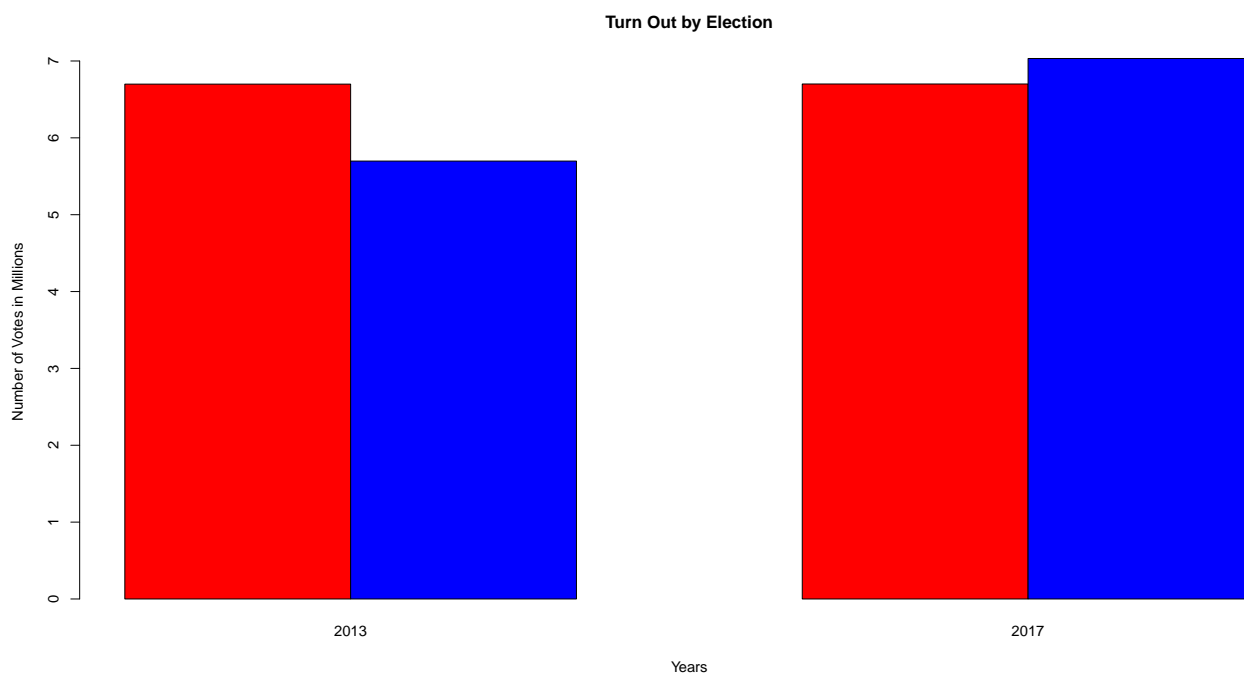
Historically, the Chilean election scheme for the presidency had been a majoritarian system but had different rules according to what happened when neither of the candidates gets the absolute majority. There had been different rules for solving this issue. For example, in the 1925 constitution on that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

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

first-round elections. From 2000 until now, it is complete opposite there be always a second-round election, which is a very similar issue if we compare with the French election that also has that characteristic.

Usually, the result of the turnout-rate between election was that the turn out rate diminish a little related to the round, due to how the registration rules changed in 2009. Before the registration in the voter register was voluntary, but the vote was mandatory. After 2009, the system changes to a system where the registration was automatic, and the vote is voluntary. This change creates an incentive to try to mobilize groups between rounds. Seeing if some particular group of voters can be activated in a specific round, this creates a 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 are comparing the first with the second-round election, we have a diminishing of the 15%. One explanation 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, in second place the center-right candidate with 25% of the votes. In the second round, Bachelet won easily with 62% of the votes.



In the 2017 election, the history is quite different. It is a more competitive election than the one of 2013. In the first round, the current president Sebastian Piñera got 36%, but the center-left candidate got 22% of the votes. Because that election has two 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. On the other side of the political spectrum, a far-right candidate Jose Antonio Kast got 10% of the votes.

The election has a total of 8 candidates, from the all political spectrum from the far-right or alt-right, until far-left candidates. As we discussed 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 created left-wing coalition called Frente Amplio, Beatriz Sanchez(BS) reached 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 the votes, but the rest of the coalitions from the center until the far left add up to 54.5% of the votes, then with this result was expected that would be a very tight race in the second round. One important thing to notice is that the candidates, Carolina Goic, Alejandro Guille, and Alejandro Navarro, came from parties that were part of the government 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 that in the following table, the second round result, we can see that SP has an astonishing 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 gets that 40%, it usually would be beatable in the second round. With this setup, many people thought that the second round would be very tight. Nevertheless, was the other way around, Piñera won with 54% of the votes, this opens the question of how this happened, there are two possible hypotheses about what happened, which will be 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

## Possible Hypothesis.

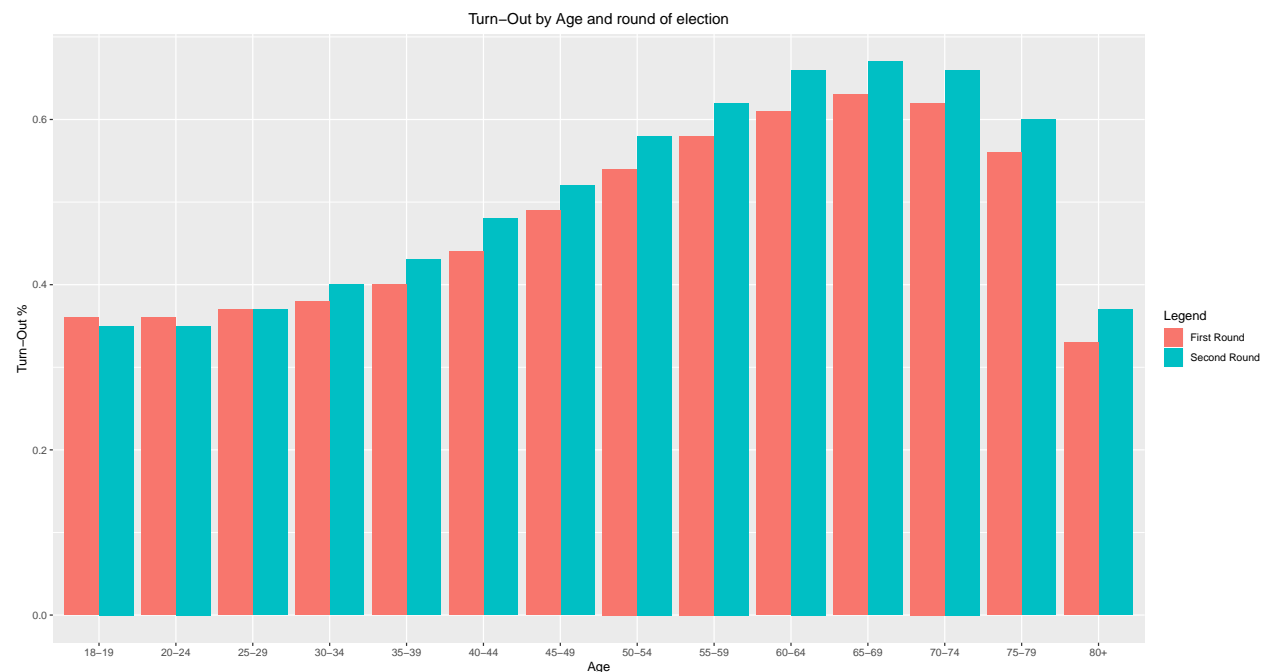
First, an essential element to take into consideration in the Latin-American political context is the polarization between the leftwing and right-wing government at a continental level. Shifts in the government’s ideology influence the current political debate in other countries in the region. For example, the current elected president of Uruguay before the start of riots in Chile, says that Uruguay must become like Chile or in the Chilean case where the current president for the second-round election uses the concept of “Chilezuela” as an element to mobilize the population to vote in the second round. In other words, can be such that voters do not vote in the first round, but vote in the second round.

Second, an element is related to the competitiveness of the election, as the election is more interest more people want to participate. I was implying that this explanation could help to explain the difference between what happened in 2013 with 2017. The first election was decided in the first round election, and even there was debate if Bachelet was going to win in the first round.

One way to think about how to differentiate these two hypotheses is the following: if on average all people decide not to vote between rounds meaning that if the probability diminishes the equal for every person independent of the age, gender, political preference, income, etc. We could argue that it is related to a competitiveness issue if all people on overall decide not to vote; maybe they don’t find exciting enough. On the other hand, if we start to see that certain groups begin to increase the turn-out, it’s possible to think that there are issues that move people that allow to certain groups increase or diminish their turn-out. Even we

could believe in institutional elements or during the campaign to incentivize/de-incentivize groups to vote.

One way to start to shed light on this problem is to see the distribution of the result through the demographics characteristics. In the following graph, we can see how the distribution of voters changes between round, especially for older adults, that they increase their participation in the election. Which reinforces the idea that specific discourse targeted to people with a more conservative world view could have affected the electorate.



## Data

We are going to use three data sources:

1. Voter File
2. Electoral results

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

Finally, the electoral results are at the level of the ballot box. They are around 43,000 ballot boxes results for each election, which means that we don't know if a person votes in a particular election. Only we have the result of the ballot box where she is registered to vote. Also, there is information in which part of the country lives, and this means that potentially, we can link it with economic data by commune.

## Methodology.

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

The 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 the second round, the number of votes that the winning candidate Sebastian Pinera 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 the first round of all j candidates in the ballot, Demographics variables depend on which level we are working. Still, in the case of the ballot level, we are going to control by the average age in the ballot box, and the percentage of men registered on it, the upper script reflects the level of the data.

In the second approach, what we are going to do is estimate a similar model, but without the demographic variables constrain the values of the estimator between 0 and 1, and estimate for two-equation simultaneously and for complement estimate the third one. Such we can get a possible decomposition of how are the transition of votes between round. Nevertheless, this is not a complete identification of the system because we need as number of equations that we estimate the same numbers of variables to identify the system — provoking a limitation of the approach in general.

$$\begin{aligned} &Min(V_c^B - \sum_{j=1}^n \beta_j^B 1stRoundResults_j^B)^2 \\ &st. \sum_{j=1}^n \beta_j^B = 1 \\ &\beta_j^B > 0 \\ &\beta_j^B < 1 \end{aligned}$$

## 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 communes where the ballot box are.

One of the main issues is sections is how to interpret this regression. Ideally, we would like to have estimators between 0 and 1, that's why we build the restricted model in the next section.

First the classical interpretation of an OLS regression, in the case that is negative or positive 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 nonvoters of the second round.

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

We are going to start with models that regress second-round results with all options in the ballot on the first round. We can see that the parameters for the equation of nonparticipation in the second round have the most quantity of parameters statistically equal to zero for candidates GC, JK, Artes, and Navarro, which implies that most of their voters went to vote into the second round. The other exciting thing is that most of the parameters sum closely to one.

Now we analyze every variable of the model into the dependent variable of the three models.

- First Round nonvoters.

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

- Carolina Goic voters.

We can see that where were votes of her do not increase the abstention on the 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 the government of AG.

- Jose Antonio Kast voters.

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

- Sebastian Piñera voters.

We can see that, in this case, all estimators sum closely to one as the other ones, but in this case, we have values over one and negative values. Due we are running his votes in the first round and the second round, which is highly correlated, and we can see that he also captures votes of the nonvoters population of the first round. Additionally, some of his votes, according to this regression, go to AG, which is very non-likely. Still, the interesting element is the negative sign in the regression for second-round nonvoters. Where he reduced the amount of abstention in the second round by .11 votes by each vote that he got. Which is consistent with the result on the equation for nonvoters in the second round. In other words, if the sign of SP in this equation was positive and the sign of nonvoters of the first round was also positive would not make sense.

- 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. Ideologically makes sense. 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 the case of MEO, we can see that around .11 of his votes went to abstention, .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 the second round, which is consistent with the ideological position of the candidates. In the case of AN, we can see that most votes went to AG. Nevertheless, the effect in the abstention of the second round is negative. Implying that some of the first-round potential voters may have increased the numbers of voters. Finally, we can see the blank and annulled 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 participation is nonlinear according to the graph that we present in the first part. The result that we can see is that in ballot boxes where there are more men than woman, tend to participate more in elections. Also, we have a positive relationship between the percentage of men registered in the ballot box and the abstention in the second-round election. The other equations are precisely the opposite. In the case of age, we can see that for both models of the candidate's AG and SP; we see that have an inverse U shape relation and with very similar estimators numbers for the square term. The contrary is valid for the case of the nonvoters 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 diminishes in the case of the age estimators. Still, in the case of the percentage of men, we can see the effect of gender over

	Model for non-votes	Model for Votes of SP	Model for votes of AG
novoto_pv	0.94* [0.94; 0.94]	0.05* [0.04; 0.05]	0.01* [0.01; 0.01]
goic_pv	-0.00 [-0.03; 0.02]	0.32* [0.30; 0.34]	0.66* [0.64; 0.68]
kast_pv	0.02 [-0.00; 0.04]	0.95* [0.94; 0.97]	0.02* [0.01; 0.04]
pinera_pv	-0.11* [-0.12; -0.10]	1.06* [1.05; 1.06]	0.05* [0.04; 0.05]
guillier_pv	-0.20* [-0.21; -0.19]	0.05* [0.05; 0.06]	1.14* [1.14; 1.15]
sanchez_pv	0.27* [0.26; 0.28]	0.12* [0.12; 0.13]	0.58* [0.57; 0.59]
meo_pv	0.11* [0.09; 0.14]	0.31* [0.29; 0.33]	0.57* [0.55; 0.59]
artes_pv	0.02 [-0.07; 0.12]	0.06 [-0.01; 0.13]	0.86* [0.79; 0.93]
navarro_pv	-0.06 [-0.16; 0.05]	0.24* [0.15; 0.32]	0.79* [0.71; 0.87]
nulos_pv	-0.49* [-0.54; -0.44]	0.15* [0.11; 0.18]	0.22* [0.18; 0.25]
blanco_pv	-0.26* [-0.30; -0.22]	0.04* [0.01; 0.07]	0.18* [0.15; 0.21]
R <sup>2</sup>	0.92	0.88	0.85
Adj. R <sup>2</sup>	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*** (0.75)	-17.06*** (0.46)	-8.41*** (0.43)
av_age	-11.71*** (0.09)	6.25*** (0.06)	5.32*** (0.05)
av_age_sq	0.12*** (0.00)	-0.06*** (0.00)	-0.06*** (0.00)
R <sup>2</sup>	0.31	0.28	0.22
Adj. R <sup>2</sup>	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*** (0.00)	0.06*** (0.00)	0.02*** (0.00)
goic_pv	0.01 (0.01)	0.31*** (0.01)	0.65*** (0.01)
kast_pv	0.03** (0.01)	0.91*** (0.01)	0.06*** (0.01)
pinera_pv	-0.07*** (0.00)	1.05*** (0.00)	0.01*** (0.00)
guillier_pv	-0.09*** (0.01)	0.04*** (0.00)	1.05*** (0.00)
sanchez_pv	0.15*** (0.01)	0.10*** (0.00)	0.72*** (0.00)
meo_pv	0.17*** (0.01)	0.26*** (0.01)	0.56*** (0.01)
artés_pv	0.06 (0.05)	0.14*** (0.04)	0.75*** (0.03)
navarro_pv	0.06 (0.05)	0.28*** (0.04)	0.64*** (0.04)
nulos_pv	-0.48*** (0.02)	0.17*** (0.02)	0.19*** (0.02)
blanco_pv	-0.23*** (0.02)	0.05*** (0.02)	0.13*** (0.02)
porc_men	-3.82*** (0.26)	-3.23*** (0.20)	6.76*** (0.20)
av_age	-2.16*** (0.04)	0.93*** (0.03)	1.30*** (0.03)
av_age_sq	0.02*** (0.00)	-0.01*** (0.00)	-0.01*** (0.00)
R <sup>2</sup>	0.93	0.88	0.86
Adj. R <sup>2</sup>	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

votes of SP in second-round are now negative, meaning that controlling for the votes in the ballot, we can see that now men are less likely to vote for SP in the second-round. For the political variables, we can see that most significant change that we see between the models presented in table 3 and 5 is for the votes of the candidate BS, where we can see that the estimator are .15, .10 and .72 for the three equations in table 5. On the other hand, we can see that the estimate in table 3 is .27, .12, and .58, which tell us that the result of BS is highly related to the demography of the voters. Making sense because as we explain previously, we know that is a new coalition that attracts young people

Before to advance to the Constrained OLS strategy, it is important to discuss that even if the number makes sense. Still, we don't have a consistent identification strategy that we are currently missing; in the next section, we are going to try to solve a simultaneous equation system 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 Constrained OLS estimate, basically as we show in the previous part is try to jointly minimize the error of the two-equation restricting the value that the parameters can take.

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. For Kast and himself, all the votes



practically. In the case of the AG, the coefficient 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. The major highlight of this regression is 5.9% of the persons how not vote in the first round but vote in the second round for him.

Table 6: Equation for Second Round for SP

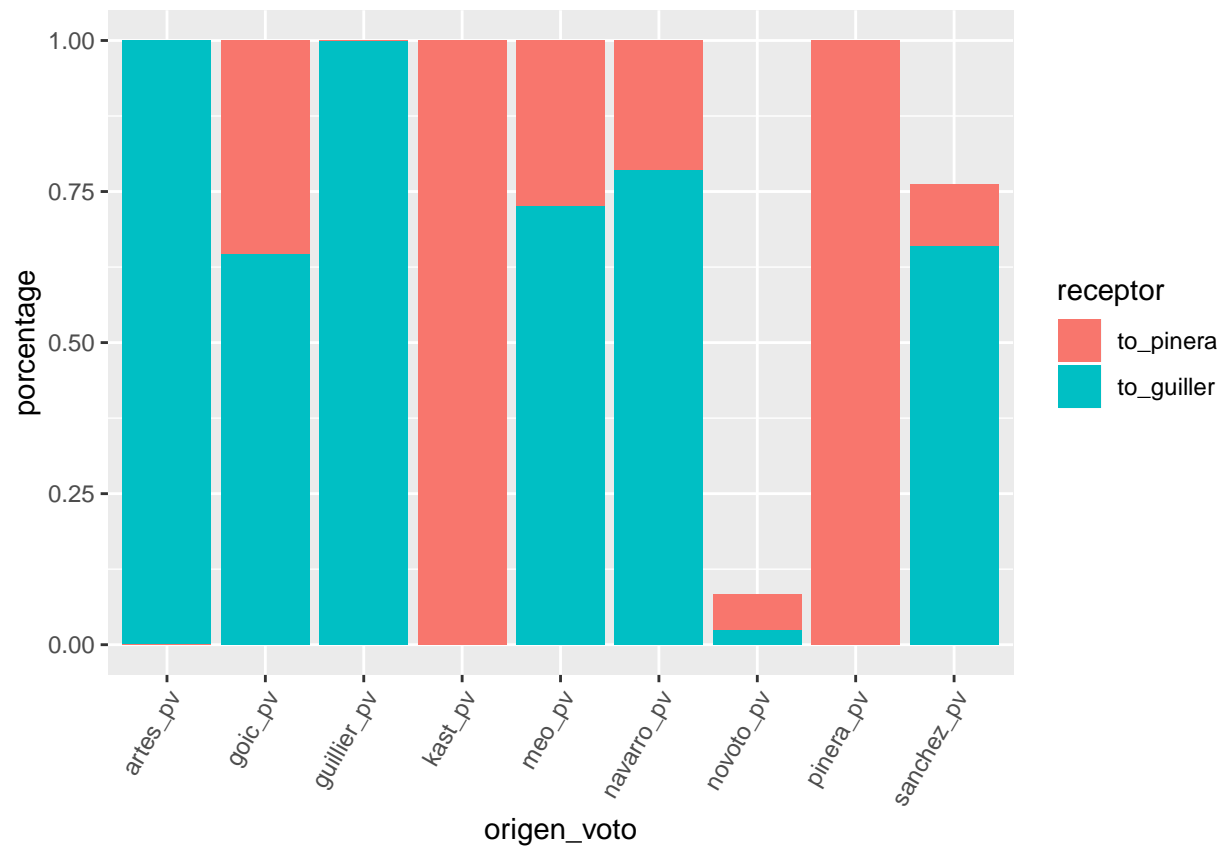
Variable	Coef	SE
novoto_pv	0.0591708	0.0000000
goic_pv	0.3527198	0.0096233
kast_pv	0.9999983	0.0065700
pinera_pv	0.9999941	0.0017718
guillier_pv	0.0000009	0.0033147
sanchez_pv	0.1010549	0.0026577
meo_pv	0.2745380	0.0108861
artes_pv	0.0000000	0.0439660
navarro_pv	0.2150549	0.0413061

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

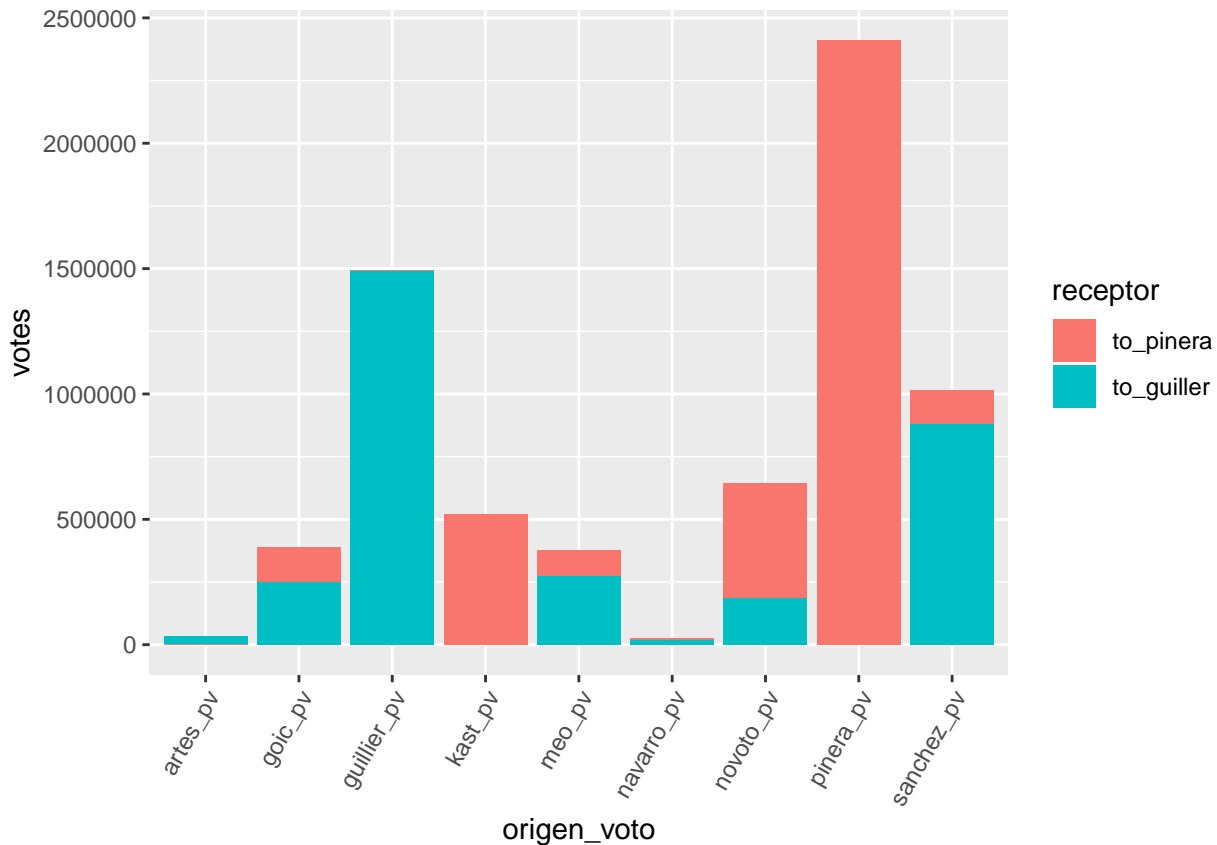
Table 7: Equation for Second Round for AG

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

Now, for the equation for AG. In the case of the nonvoters from the first round, he managed to get 2.4%. From GC voters, 64% vote for him in the second round. From Kast and SP zero. From himself and Artes basically all. From BS, 2/3 of the voters went for him in the second round. From Meo, 72% of his voters went for AG. Finally, Navarro's voters only 78% voted for AG in the second round.



In this third graph, we see the result instead of percentages in a total of votes; we can see how the vote of SP is composed in the second round through votes of all candidates expect Artes.



One problem with this methodology is that it is not easy to decompose the vote for other variables. For example, it is difficult to know how this distribute by age. We could try to rebuild the high aggregates throughout microdata and then re-run this regression. Another issue is why it's concentrating all the no voters in the second-round that voted in the first round in the candidate BS. One reason could explain only by bad motives if we see that the unconstrained regression of the previous part. The highest parameter value of the equation for nonvote in the second round is for the candidate BS, and this tells us that the algorithm for constraining the data probably it's not optimizing correctly because of concentrate all the result in one single variable. One of the reasons why we choose only to run jointly two of three models is to ease the convergence, but we have a trade-off, that is that the optimization process decides to go in the more simple local optima.

## Conclusions and discussion.

Our results tell us a history where we can see how the first round voters decide what to do in the second-round election. We have seen that there is some evidence that voters for a particular candidate do not guarantee some ideological loyalty. It is telling us that there are 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. Also, we see that in an overall election, we have a flow of new voters in the second round is 636.000 new voters and around 300.000 voters that vote only in the first round election, showing that is some space for campings in this type of election to attract or new voters.

Nevertheless, it is important to consider some limitations in the analysis. First, we are working with aggregate data, meaning that it is difficult how to interpret the data and the results of this regression, we have to take into consideration the possibility of ecological fallacy. Second, the lack of free variables for identifying the system of equations of the OLS part, ideally, we would like to estimate jointly estimate the three equations. Still, for doing that, we need three variables that help to identify the equation system correctly. But as we know, that variable is difficult to come by, and we need more elements that probably at this level of aggregation are difficult to get. Third, in the case of the constrained regression, we have seen that it 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 chooses the candidate which voters have the highest estimator in the unconstrained version, this opens a discussion that maybe the constraint force to take higher values that should be.

On the other hand, some of these problems could be solution rated at least partially. In the case of looking for free variables, it can help to identify the effect we could exploit elements as the expenditure of the campaign in the geographic area or number of visits to that Comune of the candidates or if the speeches point to specific demographics through social network. The critical point is to look at what differentiates the strategy of the candidate and build variables that reflect this process. In the case of the constrained regression problem, maybe it could be solved if it is implemented using ridge regression.

Even with this consideration, we can see from the unconstrained analysis that there are patron and behavior of the voters of the different candidates that at least even taking into consideration the limitations in the interpretation with aggregate data over individual behavior. We see the specific mobilization of voters to different candidates and this flow of voters from one to another. Nevertheless, we think that we do not have enough information to see if which hypothesis prevails. Doing this requires to go to analyze the variance of this result to commune and demographic level that this kind of information does not allow to do.