Methods and Initial Results

Dan Gamarnik

May 9, 2018

Methods and construct validity

This paper studies whether decreases in government spending on minority regions increases secessionist demands. In order to do this, I use sub-national parliamentary speeches. These are good sources for a number of reasons, first, as the highest governmental authorities within the region (not counting the central government) parlimantarians are regional elites. Second, the speeches they put out are official and highly salient which will affect public discourse and policy. Third, because they meet regularly, there is a high volume of them which leads into the final benefit: the taboo nature of secession. Because calls for independence are typically not the norm, they are likely to be rare and taboo subjects for parlimentarians to give speeches about. This rarity, in an ocean of speeches, makes it more salient when secessionist demands do happen.

In order to measure the dependent variable, I use a text analysis of parliamentary speeches. Specifically, I use a negative binomial regression for counts of important stem words that indicate secessionist sentiment. In this case they are "secession" and "self-determination". This is measured in both the dominant and minority language of each region, namely English for Scotland (Gaelic is not spoken in parliament) and Spanish/Catalan for Catalonia. I use "autodeter" as a stemword for "autodeterminació" and "autodeterminación" in Spanish and Catalan as well as "self-det" for "self-determination" in Scotland. I also use "secession" with "seces as the stemword for"secession","secessió" and "secesión" in English, Catalan and Spanish respectively.

These counts are organized by month and zero indicates both that no such words were spoken or that no parliamentary secession occurred. It also uses all types of parliamentary secessions, whether or not they relate to independence. Given that this is a rare event with possible, long-tailed distributions, I will use a negative binomial regression. Slapin and Proksch (2008) have suggested using the Wordfish algorithm in order to best capture policy proposals in parliaments. However, there several reasons for why this is problematic both theoretically and empirically. First, many of these parties do not have a consistent policy for independence for the entire 1999 to 2017 period. Wordfish is designed to input policy platforms and then traces the frequency of those platform phrases in a corpus. However, this requires that parties have a consistent, stated position on a policy for the entire period which nationalists did not.

The tests are organized as such: First, I a logit of use word counts of secessionism to predict whether secessionist rhetoric increases the liklihood of independence referendums to occur. This is a conceptual check for whether the words are valid as a means of measuring secessionism. Secondly, I use a negative binomial regression of budget data to predict secessionism, my ultimate variable of interest. This suggests that as budgets decrease, secessionist demands increase in minority regions.

Data

For the parliamentary speeches, I used web scraping to download thousands of speeches from the Catalan and Scottish parliaments. These speeches are stored online by the Catalan and Scottish regional websites and are transcribed after secessions. For the project, I included all speeches regardless of the relevance of the secession: they included committees, plenary meetings and all parliamentary business in general. This was done to include the possibility that MPs talk about seceding in unrelated committees. I then used text analysis to get a word count of the words mentioned earlier and aggregated them by month from 1999 to 2017.

Table 1: Summary Statistics

Table 1: Table continues below

X	Country	Region	Year	Month
Mode:logical	Spain :228	Basque : 0	Min. :1999	Apr : 37
NA's:443	United Kingdom:215	Catalonia:228	1st Qu.:2004	Aug : 37
NA	NA	Scotland :215	Median :2008	Dec : 37

X	Country	Region	Year	Month
NA	NA	NA	Mean :2008	Feb : 37
NA	NA	NA	3rd Qu.:2013	Jul : 37
NA	NA	NA	Max. :2017	Jun : 37
NA	NA	NA	NA	(Other):221

Table 2: Table continues below

seces_word	self.det_word	tot_words	Tax_Auto
Min. : 0.0000	Min. : 0.000	Min. : 0.000	Min. :0.0000
1st Qu.: 0.0000	1st Qu.: 0.000	1st Qu.: 0.000	1st Qu.:0.0000
Median : 0.0000	Median : 0.000	Median : 0.000	Median :0.0000
Mean : 0.6298	Mean : 1.898	Mean : 2.528	Mean :0.4853
3rd Qu.: 0.0000	3rd Qu.: 1.000	3rd Qu.: 2.000	3rd Qu.:1.0000
Max. :28.0000	Max. :64.000	Max. :73.000	Max. :1.0000
NA	NA	NA	NA

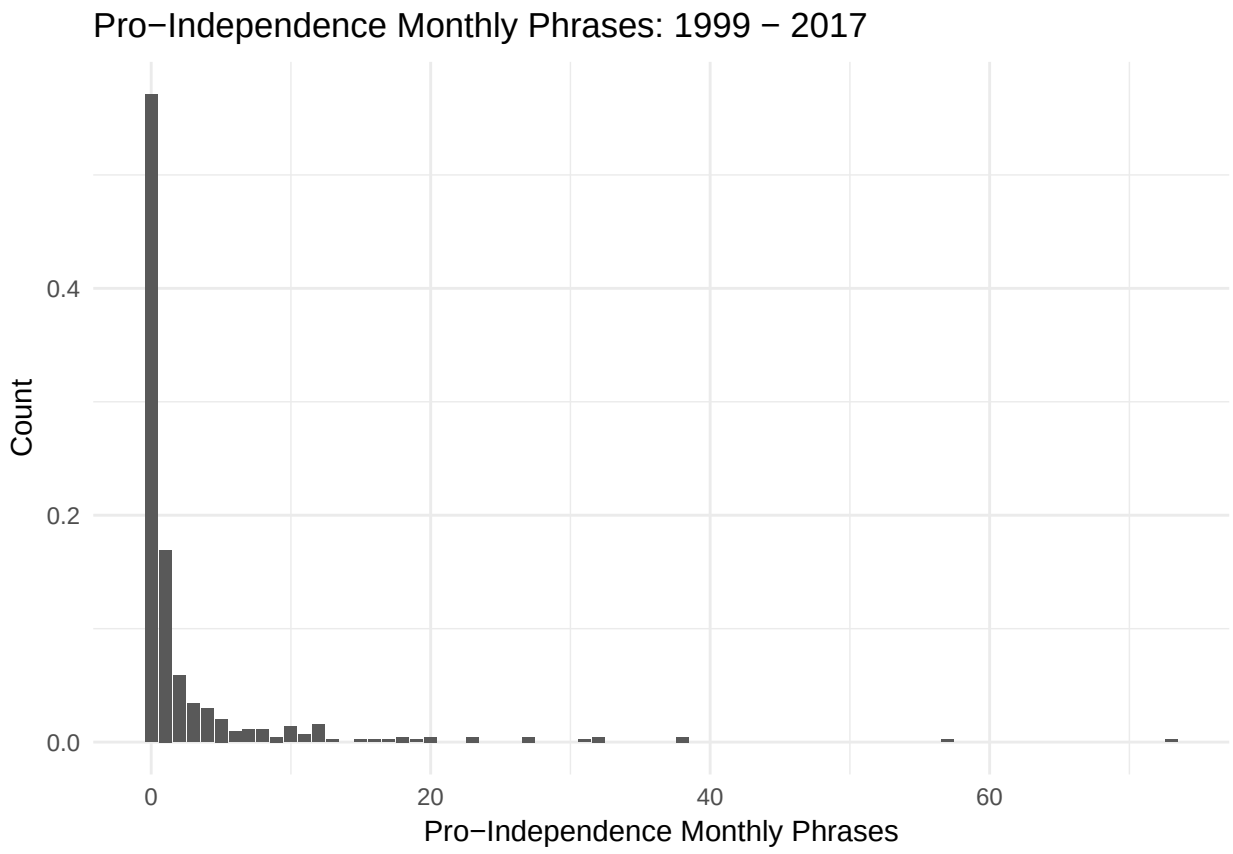
Table 3: Table continues below

Referendum	Ruling_cat	Mas_speech	Budget
Min. :0.000000	Min. :0.0000	Min. :0.0000	Min. :11.49
1st Qu.:0.000000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:26.91
Median :0.000000	Median :0.0000	Median :0.0000	Median :34.03
Mean :0.006772	Mean :0.3991	Mean :0.2807	Mean :40.20
3rd Qu.:0.000000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:59.55
Max. :1.000000	Max. :1.0000	Max. :1.0000	Max. :71.20
NA	NA's :215	NA's :215	NA's :12

Year.1	Catalonia	Scotland
Min. :1999	Min. :1	Min. :1
1st Qu.:2004	1st Qu.:1	1st Qu.:1
Median :2008	Median :1	Median :1
Mean :2008	Mean :1	Mean :1
3rd Qu.:2013	3rd Qu.:1	3rd Qu.:1
Max. :2017	Max. :1	Max. :1
NA	NA's :215	NA's :228

The data for the independent variable are the budgets. I got these from regional government websites of Scotland and Catalonia which curiate the data. These budgets indicate the level of funding that the central government gives to these minority regions and thus, I argue that it predicts the disconent of the regions and their desire to form their own countries. Other data which I currently have are major historical events that are said to have spurred Catalonia into wanting indepdence as dummy variables (that is, coding all months as ones after the events occured). I also have data on tax autonomy status for each region, also coded as a simple dummy variable.

Table 1: Pro-Independence Phrase Monthly Occurance



Initial Findings

Validity of Secessionist Phrases

Given the novelty of this design, I first wanted to test to see if these phrases were meaningful indicators of secession or independence. To do this, I use instances of referendums in Scotland and Catalonia. These were major votes to decide whether or not the regions would secede and become their own countries, occuring in both regions in 2014 and again in Catalonia in 2017. Therefore, I test whether pro-independence phrases predict the onset of these referendums.

Model specifications:

This simple model uses referendum onset as the dependent variable with the secessionist stemword variables (either "secession", "self-determination" or both) as the independent variables. Likewise, it includes robust standard errors and is clustered by region.

Table 2: Both Phrases and Onset of Referendums

Table 5: Fitting linear model: tot_words ~ Referendum

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.464	0.3229	7.631	1.453e-13
Referendum	9.536	3.923	2.431	0.01547

As the results show, these phrases did significantly increase the likelihood of a referendum happening. While the index of both phrases was significant, much of the result seems to be driven by "self-determination" rather than "secession" as the key phrase (see Appendix). Regardless, these findings suggest that using these phrases is a valid measure of secessionist sentiment among the regional elite.

Funding and Secession

Lastly, the main results tested whether funding affected parliamentarian secessionist demands. Since the data has many zeros with long tails (see figure 1) I use a negative binomial regression which generally fits this distribution.

Model specifications:

This simple model uses secessionist phrase occurance as the dependent variable (both phrases) with budgets as the independent variables. Unfortunately, I was unable to find a negative binomial regression function that also clustered by the regions and had robust standard errors.

Table 5: Funding and Secessionist Sentiment

```
##
## Call:
## glm.nb(formula = Budget ~ tot_words, data = sece, init.theta = 4.648090074,
##     link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0740  -0.7637  -0.2527   0.8423   1.3839
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.709096   0.025161 147.413   <2e-16 ***
## tot_words   -0.006407   0.003494  -1.834   0.0667 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(4.6481) family taken to be 1)
##
##     Null deviance: 449.24  on 430  degrees of freedom
## Residual deviance: 445.96  on 429  degrees of freedom
##   (12 observations deleted due to missingness)
## AIC: 3731.3
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  4.648
##           Std. Err.:  0.347
##
##  2 x log-likelihood:  -3725.325
```

The results indicate that there does seem to be a negative relationship between funding and secessionist demands. That is, as funding decreases, the calls for an independent state increase. One thing to note is thatthe relationship approaches significance with a p-value of around 0.06 but (aside from a standard of 0.1) does not quite reach it. This could be because of a lack of robust standard errors and clustering for which I did not in the negative binomial for which I did not have time to find).

However, in the tentative "robustness check", the relationship becomes highly significant but somewhat positive when controlling for if the region has tax autonomy.

Table 6: Funding and Secessionist Sentiment, Preliminary Robustness Check

```
##
## Call:
## glm.nb(formula = Budget ~ tot_words + Tax_Auto, data = sece,
##     init.theta = 27.59598207, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2210  -0.9179   0.3279   0.7758   1.6204
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 3.167554   0.020041 158.053  < 2e-16 ***
## tot_words   0.010837   0.001815   5.971 2.36e-09 ***
## Tax_Auto    0.871763   0.025352  34.386  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(27.596) family taken to be 1)
##
##     Null deviance: 1679.35  on 430  degrees of freedom
## Residual deviance:  468.76  on 428  degrees of freedom
##   (12 observations deleted due to missingness)
## AIC: 3179.8
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  27.60
##           Std. Err.:  3.53
##
##  2 x log-likelihood:  -3171.789
```

Discussion and Future Plans

Given the difficulty of just getting the data for these two regions, these results are somewhat encouraging. They suggest that (1) the phrases are valid measures of secessionist sentiment among elites because they lead to more referendums occuring and, at least based on the p-value cut-off of 0.1, there is significant negative relationship between funding and secessionist demands.

To further improve this paper, I plan on doing the same procedure on the Basque region of the Spain. Likewise, I want to include a more robust model with variables for things like opinion polls and control for things like central government revenue to test for different theories of why funding matters for secession.

Appendix

Figure 2: Funding and Secessionist Demands Over Time

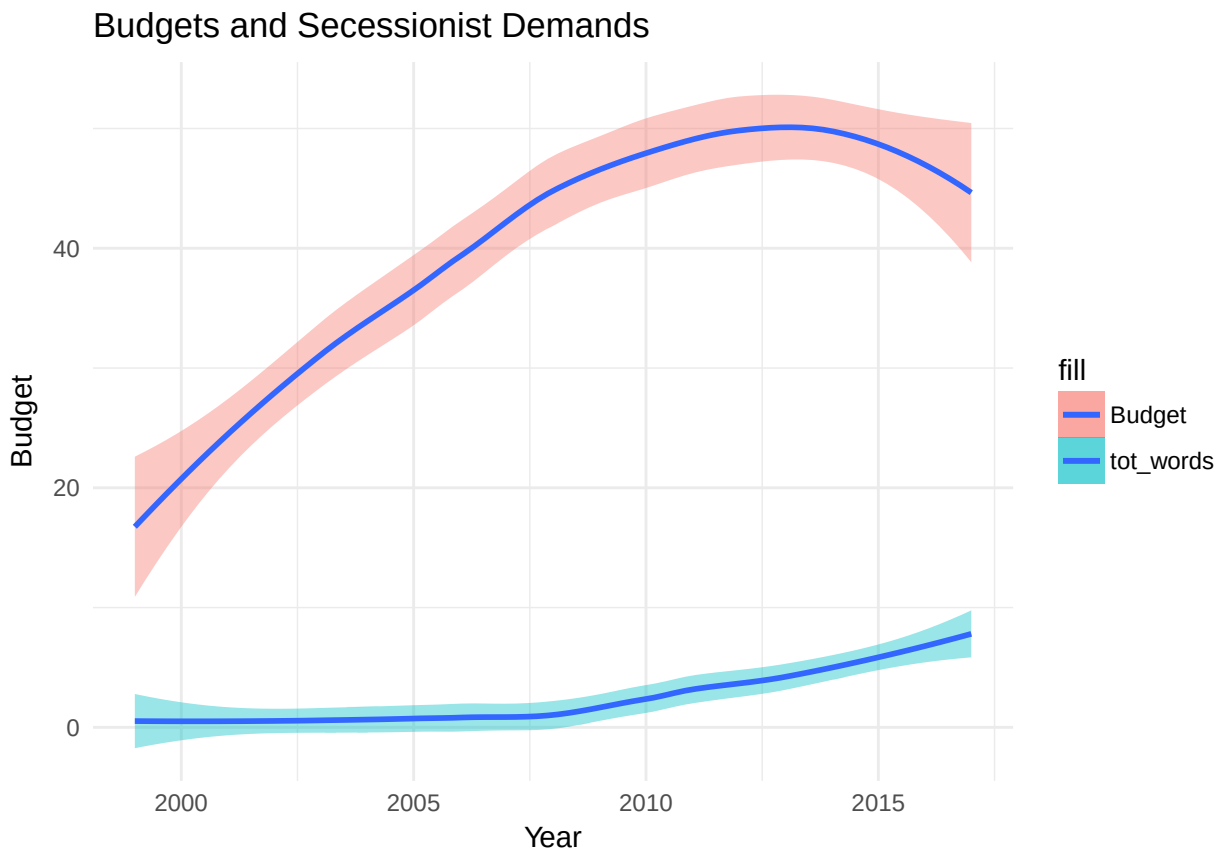


Figure 3: Secessionist Demands of Catalonia and Scotland Over Time

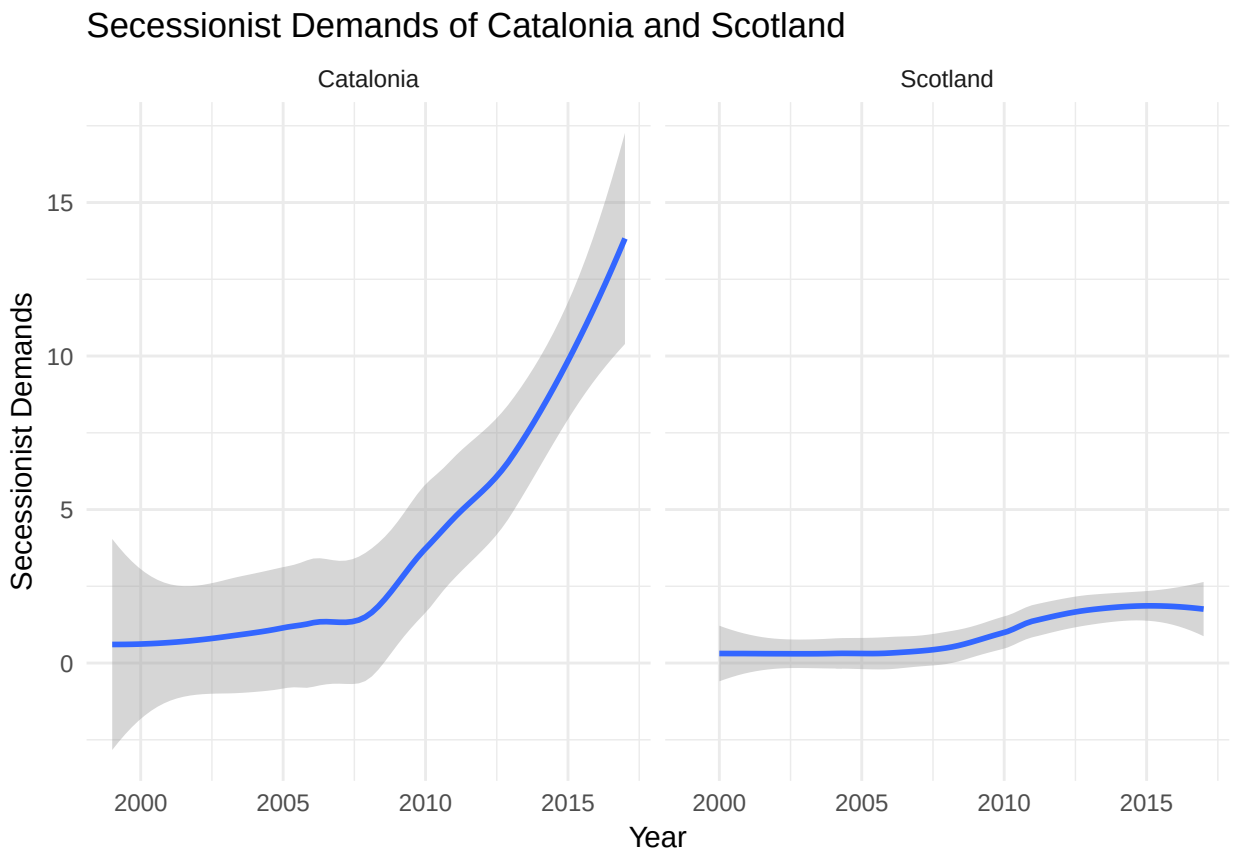


Figure 4: Regional Budgets of Catalonia and Scotland (in Billions) Over Time

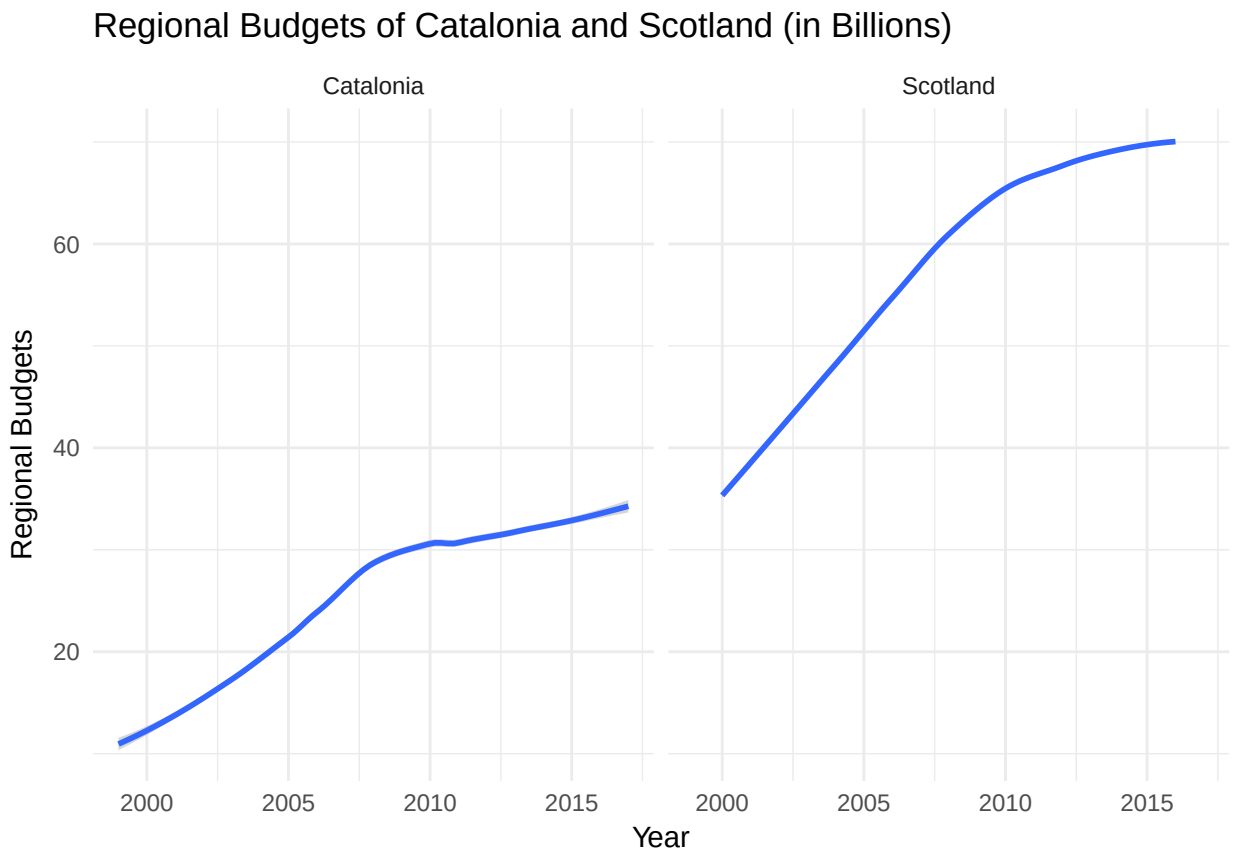


Table 3: "Secession" Phrase and Onset of Referendums

Table 6: Fitting linear model: seces_word ~ Referendum

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.6205	0.1014	6.118	2.091e-09
Referendum	1.38	1.232	1.119	0.2636

Table 4: "Self-determination" Phrase and Onset of Referendums

Table 7: Fitting linear model: self.det_word ~ Referendum

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.843	0.2727	6.76	4.391e-11
Referendum	8.157	3.313	2.462	0.01421