STATISTICAL COMPUTING

Zubeyir Eser - 0266777

May 2019

Prof. Stefano Gagliarducci
Assignment for Python and Stata
2018-2019 Academic Year

1 Introduction

In this assignment, the main purpose is analysing the US congress member's bills issued at the house and its relationships with other parameters such as cosponsorships, the state of sponsors, etc.This paper used the congress data from 110th congress up to 116th congress and also all of the data collected from **https://www.congress.gov** via Python, used the "BeautifulSoup", "pandas" and "request", in order to analyse the us congress bill analysis in terms of OLS framework.

This assignment divided into 5 parts. The first part provides the descriptive statistics. The second part will provide a graphical analysis of the variation in time of the number of cosponsors. The third part provides the analysis of the determinancy of cosponsorship at the bill level, in an OLS framework. The fourth part provides a linear probability model, to analyze the relationship between the number of cosponsors and the probability of a bill the House. The last part provides implementation of an instrumental variable(IV) strategy for the effect of cosponsorship on the probability of a bill passing the House using as an instrument for cosponsorship the number of same party representatives from the sponsor's state of election.

2 Descriptive Statistics

In this chapter we will analyses the descriptive statistics of our data. Before explain the each variable we need to indicate that information that used in the data analysis are collect from **https://www.congress.gov**. Data that we used in analysis are compononents of 116th Congress(2019-2020),115th Congress(2017-2018),114th Congress(2015-2016),113th Congress(2013-2014),112th Congress(2011-2012),111th Congress(2009-2010) and 110th Congress (2007-2008) and we have 52,558 observation over 12 year.

First of all, we need to explain all the variables that we use in our analysis. In our data our variables are ***billname***,***billdate*** ***partyname***, ***demorrep***, ***noofdemcos***, ***noofrepcos***, ***progress***, ***cosponsor***.

The ***billname***, refers to that composed of proposed bill's by sponsors and its issued number and issued congress, i.e., H.R. 0123 - 116th Congress (2019-2020) and ***billdate*** refers to the congress number. The ***demorrep*** refers to sponsor's party affiliation. In order to put this ***demorrep*** in the data analysis we need to assign a number,namely dummy variable, so that we can analyse the effect of party affiliation of each proposed bill sponsor party affiliation'. In this perspective, in the data 1 refers to Democratic Party and 2 refers to Republican Party. The ***congressman*** is a string variable and includes the name of the representative of the sponsor,its selected state and district number. The ***noofdemcos*** refers to number of cosponsorship signed by Democratic Party representative.The ***noofdemrep*** refers to number of cosponsorship signed by Republican Party representative. The ***progress*** represent the bill passing (i.e., being approved by) the House and lastly, ***cosponsor*** is the sum of the ***noofdemrep*** and ***noofdemcos***.

partyname	Freq.	Percent	Cum
Democrats	28,149	54.88	54.88
Republicans	23,139	45.12	100.0
Total	52,558	100.00	

Table 1: ***demorrep***

In the Table 1,we can see that first moment of the ***demorrep*** variable. According to the these tables last 12 year Democratic Party representative, was higher than Republican Party representatives and our sample mean is 1.448533 and standard deviation is almost 0.5.

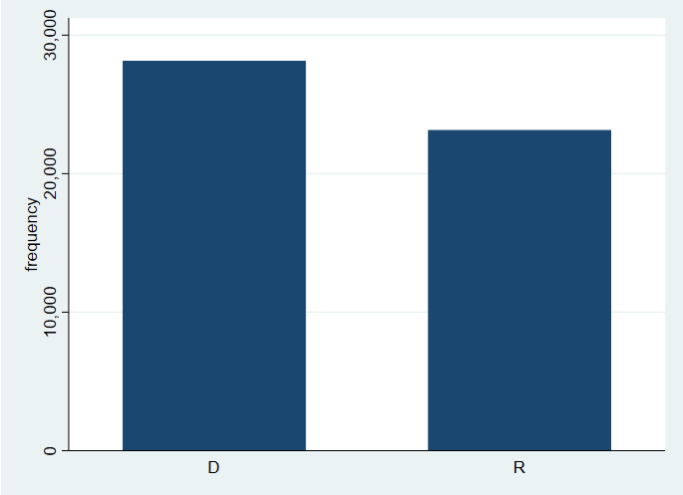

Figure 1: ***demorrep***

In the Table 3 we can see that first moment of the ***cosponsor*** variable. According to the this table average of the cosponsor on the bill is 16.20927, and its standard deviation is 34.3566. The range of cosponsor is 432.

Variable	Obs	Mean	Std. Dev.	Min	Max
cosponsor	51,288	16.20927	34.3566	0	432

Table 2: ***cosponsor***

Also it is useful to show the first moment of the partyname cosponsors ***demsuprep repsupdem***. The ***demsuprep*** indicate the number of democrats who supported(being cosponsor of republican bills) the republican and the ***repsupdem*** also indicate the number of republican who supported the democrats(being cosponsor of democrats bills).In Figure 2 we can see the mean, standard deviation minimum and maximum values respectively to the table.

All of the congress the standard deviation of the ***repsupdem*** and ***demsuprep*** are different from each other but during the 110th Congress(2009-2010) the sd of the ***repsupdem*** and ***demsuprep*** are close to one another.

billdate	demorrep	cospon~s	demsup~p	repsup~m
2007-2008)	1.364172	17.56641	2.568607	1.38577
	.481225	35.67813	10.14279	7.55224
	1	0	0	0
	2	406	184	203
2009-2010)	1.306737	17.76413	2.32795	1.333856
	.4611672	34.59576	9.061364	8.526854
	1	0	0	0
	2	425	171	250
2011-2012)	1.529772	14.04871	.8926431	2.120362
	.4991447	33.06304	7.459981	10.35024
	1	0	0	0
	2	432	236	179
2013-2014)	1.503694	16.38293	.7960199	2.99005
	.5000241	35.24977	4.427083	13.69642
	1	0	0	0
	2	380	139	194
2015-2016)	1.535601	15.93724	.9170622	2.741561
	.4987638	33.63238	5.301298	12.47393
	1	0	0	0
	2	335	156	184
2017-2018)	1.520543	15.5142	.9361986	2.68901
	.4996056	33.66412	5.61499	13.00272
	1	0	0	0
	2	385	213	196
2019-2020)	1.36695	16.04162	1.874848	1.044046
	.482046	34.61745	8.772323	7.215539
	1	0	0	0
	2	333	184	160
Total	1.451158	16.20927	1.469545	2.103942
	.4976136	34.3566	7.539367	10.88362
	1	0	0	0
	2	432	236	250

Figure 2: The first moments of the ***demorrep***, ***demsuprep*** , ***repsupdem***, ***cosponsor***

3 Graphical Analysis

This section will provide a graphical analysis of the variation in time of the number of cosponsors components such as ***demsuprep***, ***repsupdem***, ***repsuprep***, ***demsupdem*** at bill level.

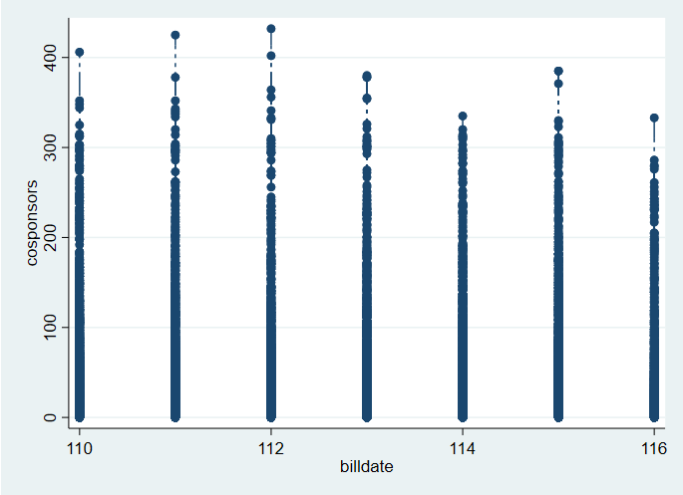

Figure 3: The range of the cosponsors over time

In the figure 3 we can see the range of the cosponsors over the time which is the number simply represent the congress,i.e., 110th congress. But this graph is not enough to evaluate the time variant of the cosponsor. We need to also take consideration of the its standard deviation and mean.

In the figure four we will the each congress year average of the cosponsors and the standard deviation added and substracted version at 95 percent of the confidence interval. Averages are shown by point symbols and intervals by capped bars.

As we can see from the Figure 4, the cosponsor average goes down if we exclude the 112th congress.The range of averages of the congress are almost between 18 and 16 if we again exclude the 112th congress. The reason 112th might be the election year of the USA president, but we cannot say that obviously because 2016 was also the election year of the US presidency.

In the figure 5 you can see that combine graph of cosponsors with the ***demsuprep***, ***repsupdem***, ***repsuprep***, ***demsupdem***. As we can see from Figure 5, the

number of cosponsors democrat are willing to support democrats. On the other hand on average republican are less willing to be a cosponsor for individual republican bills.

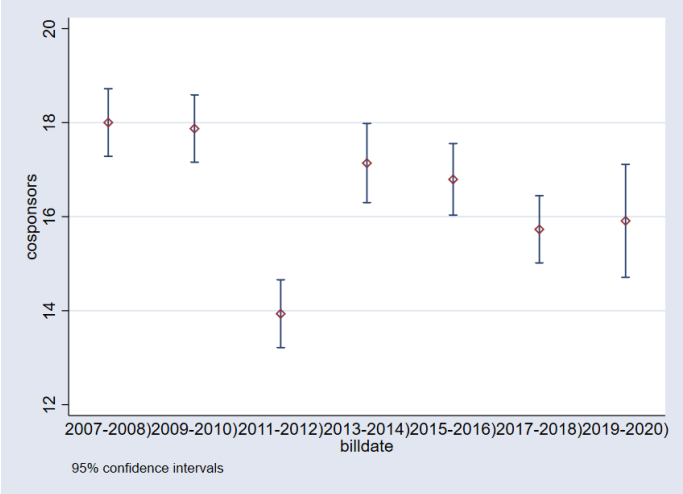

Figure 4: The first moments of the cosponsors over time

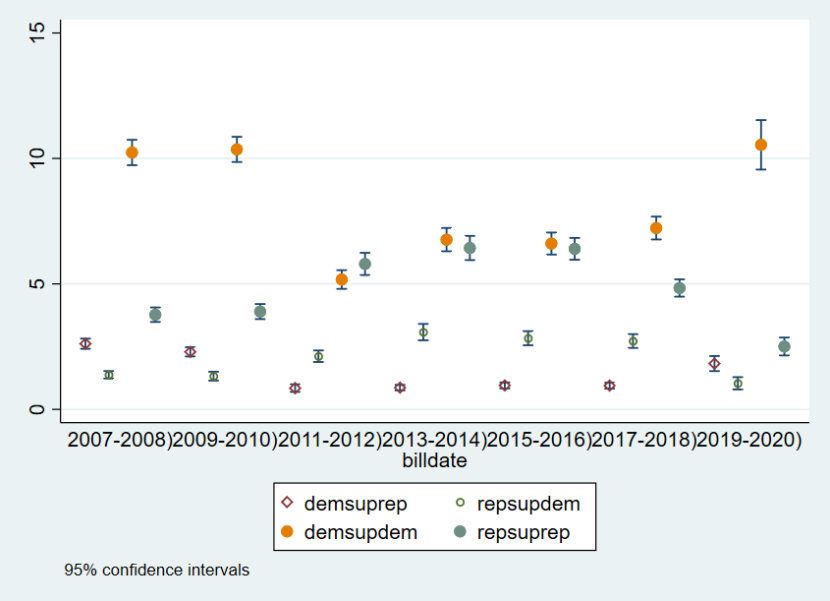

Figure 5: Cosponsorship within the parties

In figure 5 we can see that during the 110th-111th and 116th congress democrats support the democrats is more than twice on average and for this periods democrats support republican is higher than republican support democrats. On the other hand 112th-113th-114th-115th the same party representative support is more or less same on average and also during this period the average of democrats support republican is less than republican support democrats.

4 Determinancy of the Cosponsorships

In this section I used the simple OLS; cosponsor on dummy variable republican and districtno. The distrcitno is not, however, define the above. The districtno indicates the sponsor's elected district number in the state where he/she was candidate. In order to avoid dummy variable trap, I just added only one n-1 which is only one dummy variable. Before looking at the OLS result we need to consider what we are expecting from our OLS that regress cosponsor on republican.

Before analyses the OLS result we know that democratic party representatives are almost half-half (55 %) overall average of the 110th congress to 116th congress.Hence, our OLS result we cannot say that we are expecting that the coefficient of the republican will be negative or positive.

	(1)
	cosponsors
republican	-1.473***
	(-4.82)
districtno	0.0248**
	(3.14)
_cons	16.28***
	(59.47)
N	51288

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The dummy variable and districtno on cosponsor are statistically very significant

but this is not enough, because we are sure that our model is misspecified.However, if we believe that our model correctly specified under the classical assumption, we can rely on the coefficient of the republican and districtno. On the other hand district number of each sponsors state seems irrelevant to explain the cosponsors.

5 LPM

In this section we will analyse the relationship between the number of cosponsors and the probability of a bill passing (i.e., being approved by) the House via linear probability model.

	(1) cosponsors	(2) progsituation
republican	-1.473***	0.0449***
	(-4.82)	(14.08)
districtno	0.0248**	0.000106
	(3.14)	(1.29)
cosponsors		0.000896***
		(16.15)
_cons	16.28***	0.113***
	(59.47)	(39.68)
N	51288	51288

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In the above table regression (2) means that progress situation regressed on the republican, districtno and cosponsors. In this regression 2 our main interest is that the the relationship between number of cosponsors and the *progsituation* which is defined as if bill statue just introduced it is equivalent to 0, otherwise 1. 1 means that at least the bill approved by the house.

To sump up, in our linear probability regression we can say that the chances of bill by the house increase by 0.0896 percentage points for every extra cosponsor and that

on average republican are 4.3 percentage points more likely to pass the bill by house, even after allowing for cosponsors.

However, we know that LPM estimates by the OLS when the dependent variable is binary, it will allow the predicted probability to outside of the 0 and 1. For this reason it does not make any sense when we predict the outside of the range. The other problem is that our variance is depend on independent variables, hence our model will suffer from heteroskedasticity, but we can avoid the heteroscedasticity using WLS or adding robust option in the stata codes. The final problem is when we are estimating our model via the LPM we know that our model will not distrubuted as normal. In order to obtain reliable inferences we need to use non-linear model, i.e, Probit or Logit.

6 IV regression

In this section we implement the instrumental variable strategy for the effect of cosponsorship on the probability of a bill passing the House using as an instrument for cosponsorship the number of same party representatives from the sponsor's state of election. Our Instrumented variable is cosponsors and instruments are republican, districtno, and number of cosponsor from same party of sponsors and same state of sponsor.

First of all in order to implement IV strategy we need to inspect that our interested variable which is cosponsor might be endogenous because of the correlation with error term. Hence we need an instrument for cosponsor in order to purged of its correlation with error term and IV strategy will exactly purge cosponsor correlation with error term.

In the two stage least square estimation, we bacially first regressing our instrumental variable on cosponsors, after that we are analyzing the second outcomes.

	(1) cosponsors	(2) progsituation	(3) progsituation
republican	-1.473***	0.0449***	0.0465***
	(-4.82)	(14.08)	(14.51)
districtno	0.0248**	0.000106	0.0000797
	(3.14)	(1.29)	(0.96)
cosponsors		0.000896***	0.00196***
		(16.15)	(16.63)
_cons	16.28***	0.113***	0.0954***
	(59.47)	(39.68)	(29.63)
N	51288	51288	51288

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Secondly, we are inspecting that our instrument is probably highly correlated with cosponsor and uncorrelated with error term. For this reason if we assume that our model satisfy the exogenous and relevance assumption, we can implement the IV strategy.

Our instrumental analysis results are above table in (3). As we mentioned above we were expecting the district no is irrelevant about our regression and it is statistically insignificant and we can see that the t-statistic of districtno decreased. For the coefficient and t-statistic of the cosponsor is increased and also confidence interval of the cosponsor changed from (.0007869 // .0010043) to (.0017329 // .0021959). Our instrument seems highly relevant because our confindence interval is getting tighter than simple OLS.