Learning by Doing in Public Construction Projects

Maximiliano González

MAPSS - The University of Chicago

mjgonzalez@uchicago.edu

May 21, 2021

Context and Research Questions : Public Procurement

- More than 700 government units in Chile procure different types of buildings and construction-related services.
- Most are awarded through competitive bidding.
- Some of these require considerable domain-specific expertise, like hospital buildings.

Research Questions

1. Does past experience causally improve the probability to win contracts in the future?
2. What types of project show more returns to experience?
3. How do experience's advantages operate?

Literature Review

The research questions relate mainly to three strands of literature:

1. Learning by doing on firms: (Fudenberg and Tirole, 1983), (Dasgupta and Stiglitz, 1988).
2. New entrants' behaviour in auctions: (Fu, Drew, and Lo, 2002), (Li and Philips, 2012), (Estache and Iimi, 2010).
3. Empirical analysis of incumbent power: (Lee, Moretti, and Butler, 2004).

Gaps in empirical work: biases, restriction to bids amounts, little data.

Empirical Strategy I

Define p_1 as an arbitrary period and p_2 as the following period. We can investigate the association between experience gained in period 1 and success in period 2:

1. Winning Share

$$WIN_{i2} = \mathbb{1}_{EXP_{i1}>0} + CONTROLS_i + \varepsilon$$
$$WIN_{i2} = EXP_{i1} + CONTROLS_i + \varepsilon$$
$$WIN_{i2} = poly(EXP_{i1}, 2) + CONTROLS_i + \varepsilon$$

Where:

WIN_{i2} = Winning share of firm i on period 2.
EXP_{i1} = Experience gained by firm i on period 1.

Computing Experience: two alternatives

1. Rolling experience: for each period of outcomes, we look back two years to compute experience.
2. Annualized full experience: for each period of outcomes, we look back into the full past and annualize the experience.

Empirical Strategy II: Instrument Variables

- It is probable that success is endogenous in the previous models due to omitted cost variables.
- To adress this problem I propose to employ and IV approach with narrowly close submitted bids as the source of random variation.
- The design would be as follows:

Split data in two periods 1 and 2	$\Rightarrow$	Count narrowly won C_{i1} contracts in period 1 by firm i	$\Rightarrow$	Instrument EXP_{i1} with C_{i1}	$\Rightarrow$	Estimate Causal Effect of EXP_{i1}.

Problem

Not all contracts are assigned to the lowest bid

Construction projects dataset (32,200 contracts)

- Data consists in ten years of public auctions for construction projects, publicily avalaible.

Name	N	Complete Cases	mean	std	max	min
Bid (all)	119000	1	1.73e+10	5.8e+12	2e+15	1e+07
Winning Bid	32200	1	2.27e+08	2.54e+09	2.47e+11	1e+07
Difference between 1st bid and 2nd (%)	32200	0.686	0.0638	0.0859	0.984	0
Number of Bidders per Contract	32200	1	3.18	2.25	23	1
Year	32200	1	2015	2.89	2020	2010
Offers made by Firm	13800	1	8.64	18.5	846	1
Win prob. by Firm	13800	1	0.213	0.294	1	0
Offers won by Firm	13800	1	2.33	5.66	111	0

Note: The table shows sample summary statistics for the public construction dataset after size filtering has been applied. The difference between lowest and 2nd lowest bid is only available in approx. 70% of the contracts, with two or more bidders.

Exploratory Discontinuity Analysis

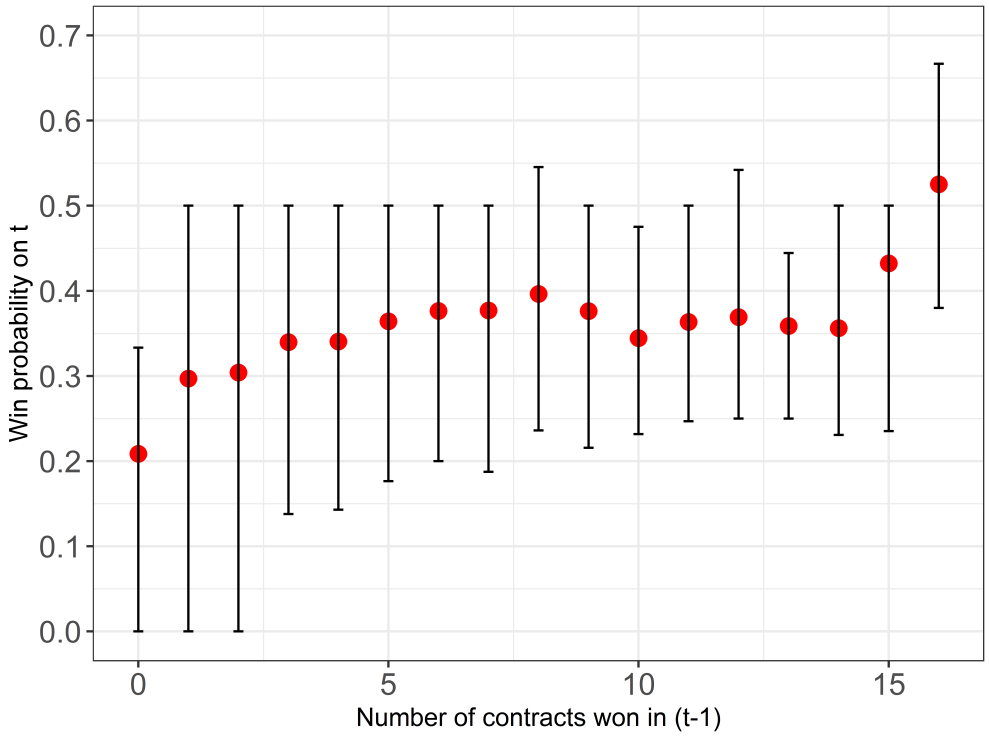

Figure: *Relationship between contracts won on $t-1$ and mean winning probability across contractors in t. Here, t and $t-1$ correspond to two periods of two years each. Error bars correspond to the interquartile range. Only x for which the number of observations is greater than ten are shown.*

Exploratory Discontinuity Analysis

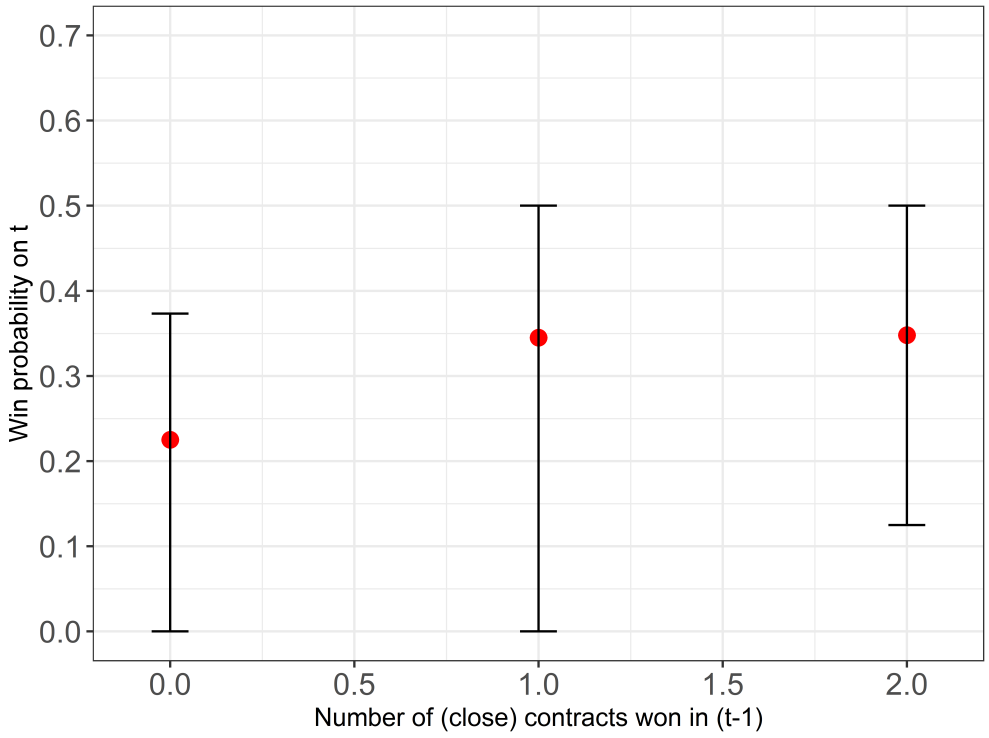

Figure: Effect of one closely won contract in period 1 on winning probability on period 2. For this graphic, the only contractors considered are those who either i) won closely one or more close contracts on period 1 or ii) did bid but not win a contract in period 1. Error bars correspond to interquartile range.

Preliminary Empirical Results: Regressions

	Dependent variable:					
	Share of Contracts won in t					
	OLS				*instrumental variable*	
	(1)	(2)	(3)	(4)	(5)	(6)
Experience in (t-1) (Binary)	0.119***			0.178***		
	(0.004)			(0.010)		
Experience in (t-1) (Linear)		0.022***	0.038***		0.022***	0.039***
		(0.002)	(0.001)		(0.001)	(0.003)
(Experience in (t-1)) (Squared)			−0.001***			−0.001***
			(0.0001)			(0.0003)
Constant	0.260***	0.276***	0.270***	0.245***	0.276***	0.269***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Fixed effects By period	Yes	Yes	Yes	Yes	Yes	Yes
Observations	37,959	37,959	37,959	37,959	37,959	37,959
R^2	0.035	0.028	0.034	0.028	0.028	0.034
Residual Std. Error	0.316 (df = 37950)	0.317 (df = 37950)	0.316 (df = 37949)	0.317 (df = 37950)	0.317 (df = 37950)	0.316 (df = 37949)
Note:						*p<0.1; **p<0.05; ***p<0.01

Figure: Results for OLS and IV specifications with robust SE in parenthesis. The dependent variable is the share of contracts won out of contract bid for by a firm (in period t). Panels 1-3 are OLS specifications with binary indicator, linear and quadratic functional forms for experience (respectively) acccrued during period (t-1) . Panels 4-6 replicate the same specifications but instrument experience with closely won contracts in period (t-1).

Analyzing type and firm size heterogeneity

1. How to adress differences in firm size?
   - Should we expect differences by firm size? Is size an endogenous variable?
   - Strategy: I employ a directory of firms maintained by the Tax Bureau which classifies firms in 13 categories according to their sales (warning: only available for 30% of firms).
   - I calculate within-categories estimates of the effect of experience.
2. How should the previous result differ by project type?
   - Hypothesis: bigger, more complex projects should show higher coefficients on experience.
   - I employ frequent unigrams and bigrams present in the name of the contract to identify types of contracts.
   - Like before, calculate within category estimates (acid test)

Differences by firm size: 10 categories of increasing size

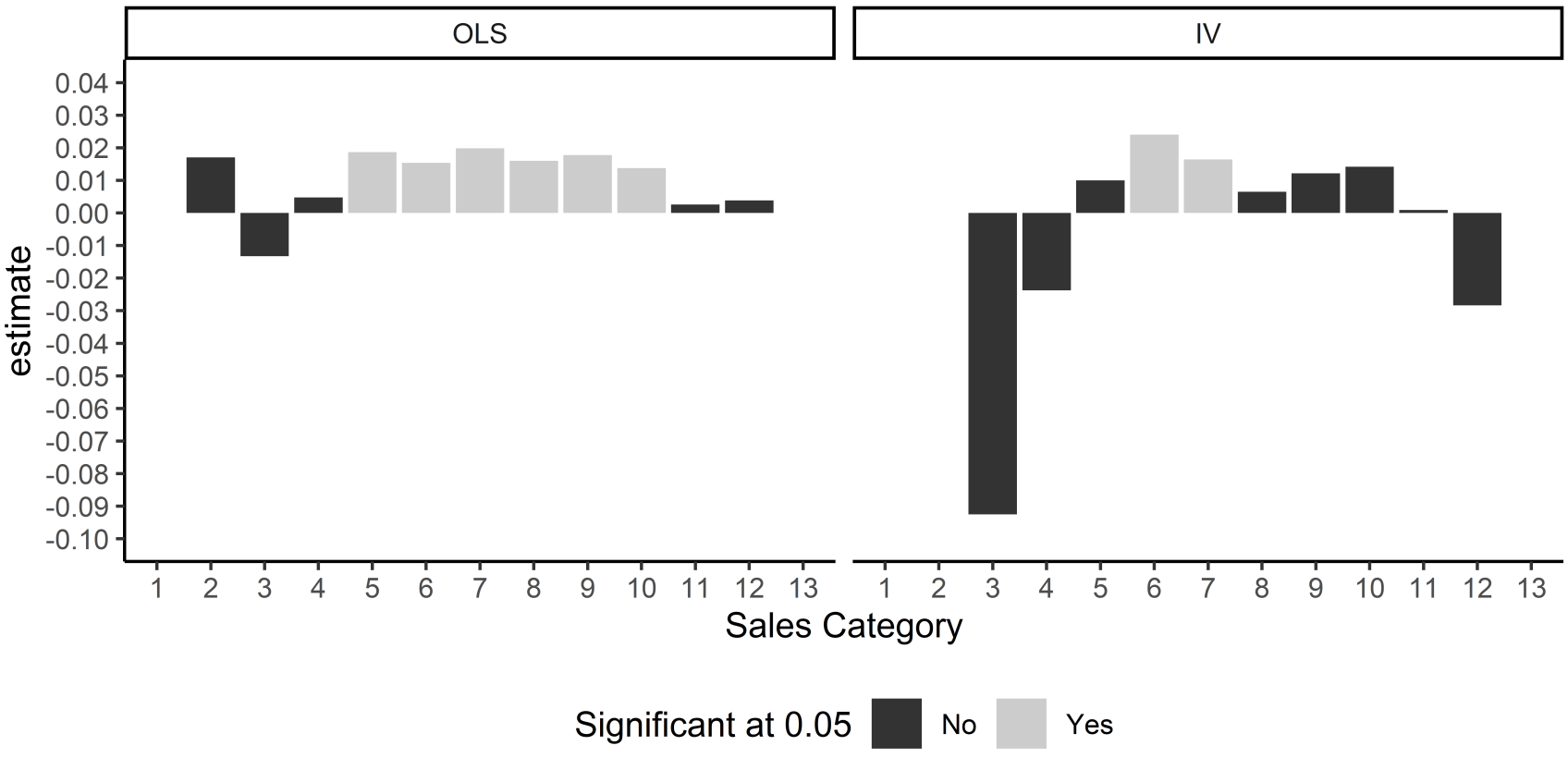

Figure: *Relationship experience and outcomes for firms of different sizes. The graph show the coefficient of linear experience on the y-axis. The x-axis show 10 categories of firms, where 3 is the smallest and 12 is the biggest. Left Panel show OLS estimates and right panel shows IV estimates as in specifications (2) and (5) of the previous regression analysis. Category 2's IV is 0.5 and insignificant and thus excluded because it would distort the visualization.*

Differences by project type

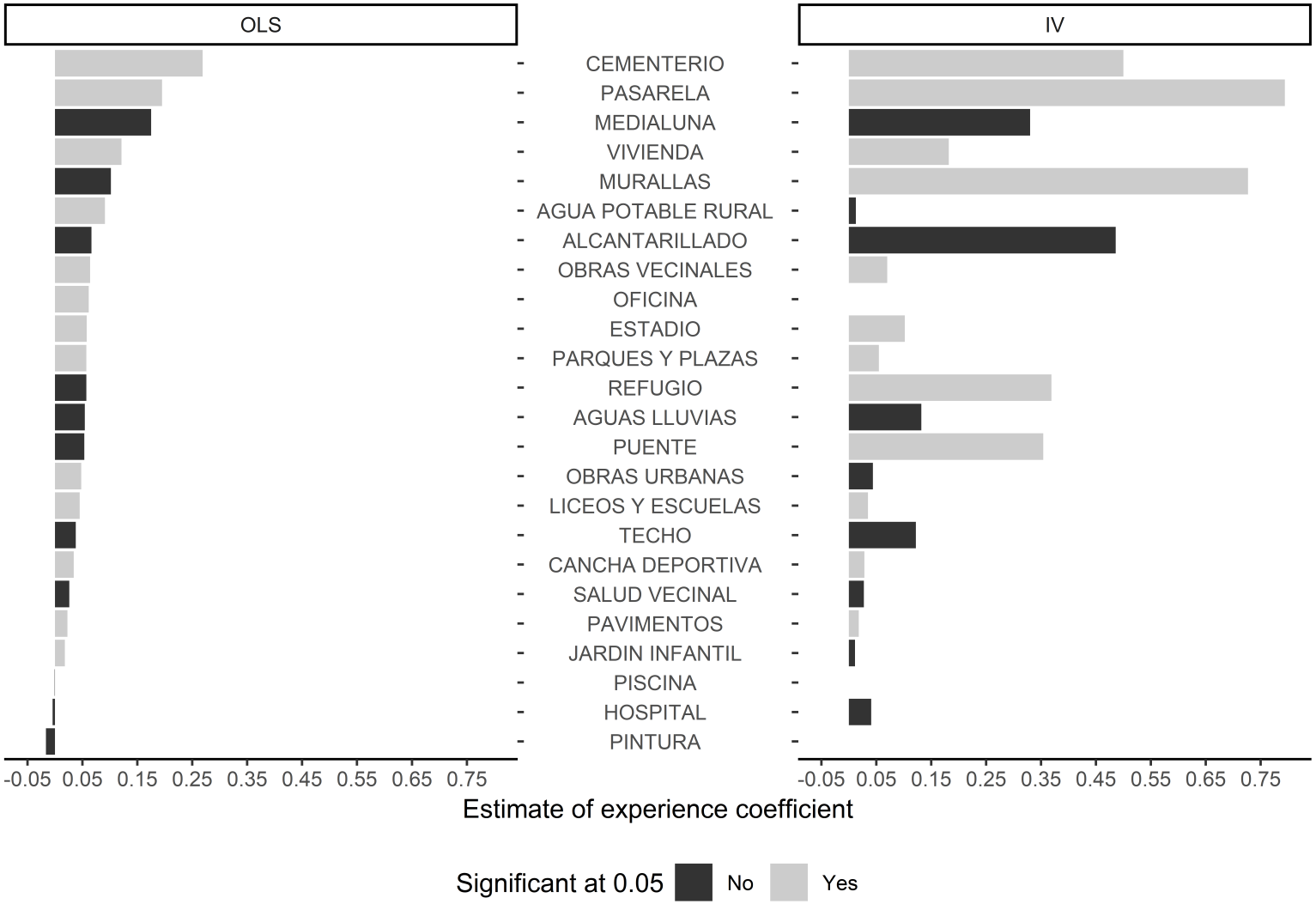

Figure: *Relationship experience and outcomes for different types of projects. Left panel show IV estimates and right panel shows OLS estimates as in specifications (2) and (5) of the previous regression analysis.*

Robustness checks developed

1. Sensibility to different thresholds of close wins.
2. Different ways to compute experience.
3. Computing outcomes in one and three year periods.
4. Computing annualized experience

Going forward

1. How do these advantages operate? Possible hypothesis are better costs, expanded feasible production space, market know-how. (Corruption?)
   - Cost advantages: analyze bids.
   - Expanded horizon: check number of bidders in low-experience vs high-experience.
   - Corruption, knowing the market: check rural projects vs. urban projects.
2. Are effects big enough to induce market changes (e.g. concentration)?
3. Is the assignment of contracts to non-lowest bidders a problem?
   - Exploring an ELO-ranked system to find close wins.

Appendix

Robustness:period of outcomes

Table: Robustness checks for duration of outcomes period of interest

	Contracts Won/Contracts Bid in Outcome Period					
	Outcome period of length (years):					
	1	2 (Original)	3	1	2 (Original)	3
Experience	0.022***	0.020***	0.022***			
	(0.001)	(0.001)	(0.001)			
Annualized Cumulative Experience				0.059***	0.057***	0.060***
				(0.002)	(0.002)	(0.002)
Constant	0.276***	0.318***	0.263***	0.289***	0.263***	0.267***
	(0.005)	(0.005)	(0.004)	(0.005)	(0.008)	(0.004)
Observations	37,959	28,893	42,533	36,844	27,737	41,471
R^2	0.028	0.032	0.026	0.027	0.028	0.023
Residual Std. Error	0.317 (df = 37950)	0.340 (df = 28883)	0.307 (df = 42525)	0.322 (df = 36835)	0.344 (df = 27727)	0.311 (df = 41463)
Note:						*p<0.1; **p<0.05; ***p<0.01

Robustness: threshold close wins

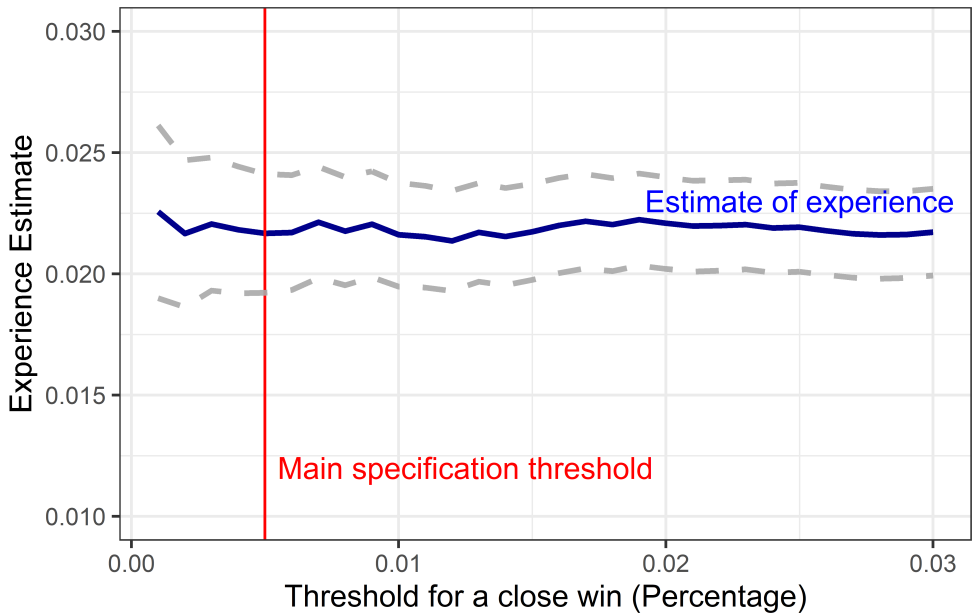

Figure: *Estimate of experience vs threshold for close wins.*

Sales Categories

Table: Sample Firm descriptive statistics with statutory sales thresholds per category

Category	Number of Firms	Sample Average Annual Sales (CLP UF)	Statutory Sales Minimum (CLP UF)	Statutotry Sales Maximum (CLP UF)
1	98	402	NA	NA
2	107	453	0	200
3	139	4317	200	600
4	496	661	600	2400
5	541	856	2400	5000
6	636	1598	5000	10000
7	852	2163	10000	25000
8	519	4567	25000	50000
9	301	9941	50000	100000
10	197	12449	100000	200000
11	153	22209	200000	600000
12	36	57219	600000	1000000
13	72	70506	1000000	NA