

Chapter 1

Data

1.1 Source and characteristics

1.1.1 Source

Our main source of data is a set of proposals submitted by firms in public auctions procured by government units in Chile between 2010 and 2021. Each observation corresponds to a proposal submitted by a firm to an auction. Each observation includes project characterization variables, auction characterization variables, and firm characterization variables.¹

Raw data for public purchases developed via Mercado Publico is publicly available in the Open Data Portal of the Directorate of Public Purchases². As mentioned in the institutional context section, most government units are mandated by the law to develop their procurement process via the Mercado Público platform. Most of the Data comes from auctions developed through this portal. Additionally, units who do not use the portal are mandated by law to publish a basic set of information to the database. Given the law requirements for firms to submit purchasing data to the platform, we expect this dataset to include all purchases made by government units in the construction type save for some exceptions.

¹Note that first two sets of variables are the same within bids for the same contract

²<https://datosabiertos.chilecompra.cl/>, last visited, july 2021

We expect information missing in contracts related to national security, for example, the construction of naval bases. However, we still see quite an amount of contracts procured by the military, but without national security connotations. Second, we do not have complete data for the Ministry of Public Works. This Ministry is exempt from the specific rules of law related to public purchases since it has its own set of regulations governing procurement of projects in road, airport, and other types of projects. Although the law mandates that even in this case the Ministry should publish basic information to the digital platform mentioned in the previous section, in practice we observe that information is only partial, especially before 2014. Third, because the Directorate of Concessiones depends of the Ministry of Public Works, we expect to see little or no contracts developed as Public-Private Partnerships (PPPs). This is not detrimental to our objective since these contracts are usually developed by extremely experienced firms who also form alliances.

The Open Data Platform has available data on public purchases in .csv files covering each one year-month of purchases. The .csv files were downloaded and merged together to form the initial raw dataset of around 10,000,000 observations, which span a much wider array of different types of goods purchased by the government.

The dataset has the following set of relevant variables:

- Project characterization variables: auctions' date, geographic region, the product category, legal size classification, procuring government unit, government estimate of cost.
- Firm characterization variables: unique tax identifier, amount offered, amount awarded, bid amount, acceptance status of the bid, awarding status of the bid.

We employ as the auction's date the StartDate field, which shows when the auction started. This can be later than when it was published.

The firm unique identifier can be of two types depending on the firm. For firms constituted as juridic entities separate from final taxpayers(i.e. individuals), the unique identifier is the unique tax number given by the internal tax bureau. For

firms identified with a final taxpayer, the unique identifier is the personal unique ID (RUN) that uniquely identifies every person in Chile. Therefore, the variable that allows us to follow entities across the years and contracts has very little noise in it and is subject to almost no errors. Government Unit's IDs are also tax identifiers, which save for extraordinary circumstances also should stay the same over the years.

The acceptance/awarded bid variables indicate whether the bid did pass the first screening for formal correctness and whether it was awarded the project.

We not detail further filtering steps employed. We employed one of the product category standardized classification ("RUBRO2") field to filter only projects in the Category "Construction Projects and Services" or "WORKS". The vast majority (almost 90%) of our data comes from observations in the first category. The second category begun being employed in 2017 to identify auctions from the Ministry of Public Works so we included it as well. This filtering steps renders observations 272,104 observations.

Furthermore, to obtain our analysis sample we perform additional filtering steps. We filter projects where more than one item is awarded to a single contractor or contractors offer more than one item o consider only single item contracts. We also drop contracts with a government estimate of less than 20,000,000CLP and where the maximum bid is less than 15,000,000CLP (if there is no government estimate we do not take that condition into account). This step aims to exclude excessively simple projects, like small repairs, which do not entail either relevant subject-matter or public-specific domain expertise. Finally, we observe firms in the dataset with more than one offer for the same contract, since contractors are allowed to modify their proposals until the end of the timeframe. We filter only the last offer by the same contractor in the same project. We end up with 163,626 observations, submitted for 49,481 unique projects.

1.1.2 Analysis Sample

The current section describes our main dataset. Although different analysis in the chapter perform small adjustments that further filter the sample, here we characterize

the main sample employed in the next chapter.

First we characterize the buyers, Table 1.1 shows relevant statistics. We have 928 unique government organisms developing on average 53 auctions across the 12 year period. Note that the average number of years in the sample is six, which speaks about good time coverage for each government body. Next, we characterize the types of government bodies, by matching strings to the unit's name. We find the distribution of units in table 1.2 . It can be seen that municipalities make the most of the projects in the sample, followed by ministries. We observe some universities owned by state as buyers as well.

Table 1.1: Government Bodies Descriptive Statistics

name	N	Complete Cases	mean	std	max	min
Number of Auctions Performed	928	1	53.3	118	2780	1
Total Firms Submitting Proposals	928	1	61.5	80.5	1170	1
Years in the dataset	928	1	6.01	4.24	12	1
Average Firms per Auction	928	1	4.36	15.3	466	1

Table 1.2: Types of Government Bodies Developing contracts

Type of Government Body	Number of Contracts Performed	Percentage	Cumulative Percentage
Municipality	36359	74%	74%
Ministry	8382	16%	90%
Other	1550	4%	94%
Child School Board	1335	2%	96%
University	892	2%	98%
Police, Investigations	728	2%	100%
Regional Government	135	0%	100%
Army,Navy	100	0%	100%

Next, we first describe bids and sellers. Table ?? shows descriptive statistics for bids and firms found in the analysis sample. The average firm bids in ten projects and wins approximately two, which gives a mean winning share of around .21. This shows that winning projects is not immediate for winning firms.

The time dimension is essential in the current investigation. Table displays the number of observations, unique firms and unique contracts for each year of the sample, along with key variables. As expected, contracts have increased over the years.

Table 1.3: Sample Descriptive Statistics

name	N	Complete Cases	mean	std	max	min
Bid (all)	145000	1	1.44e+10	5.26e+12	2e+15	0.5
Winning Bid	36200	1	2.62e+08	2.47e+09	2.47e+11	1
Difference between 1st bid and 2nd (%)	36200	0.697	0.084	0.151	1	0
Number of Bidders per Contract	36200	1	3.26	2.32	23	1
Year	36200	1	2016	3.1	2021	2010
Offers made by Firm	15000	1	9.66	27.4	1930	1
Win prob. by Firm	15000	1	0.209	0.293	1	0
Offers won by Firm	15000	1	2.42	5.93	140	0

Table 1.4: Number of firms and contract per sample year

Variable	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Number of firms	2501	2765	3038	3025	3155	3978	3677	3583	3436	3530	3700	1828
Number of Auctions	2372	3661	4710	3759	4110	5352	4640	4360	4280	4922	5313	2002
Firms per Auctions	7.27	5.67	4.87	4.26	4.43	17.5	4.61	5.34	4.94	4.94	5.77	4.23

We attempt to characterize what are the types of projects in our dataset. The name variable is usually descriptive of the nature of the project and can be used to generate an initial characterization of the types of projects. We match the individual name variable with two pre-defined lists of words, the first containing words containing the type of work (construction, maintenance, etc,) and the second the type of project (hospital, park, etc.). If a project has more than one word it is considered in both categories. The results are shown in the heatmap of Figure. We find that.

1.1.3 Awarding Criteria Data

Does not contain data on criteria used.

We employ the API to search for information on the url of the acta of adjudication. we then employ a bot to download and parse the full page and extract the criteria data.

We find good success on the process, as we match % of the sample to a url and then % of the sample to the corresponding criteria.

we create the following variables: and indicator and the weight of criteria related to experience, quality and price. we do this by matching common strings to criteria, which is non-standardized text. Statistics for these criteria in the sample are shown

in Table. -indic, complete cases, mean , sd, etc. Plt shows the distribution of the three weights. In our analysis, this data is key to identify which contracts already include experience as awarding factor and thus the ones we drop from our outcome computation.

1.2 Experience and Firms