

# Chapter 1

## Data

Our main source of data is a set of proposals submitted by firms in public auctions for goods developed by government units in Chile between 2010 and 2021. Raw data for public purchases developed via Mercado Public is publicly available in the Open Data Portal of the Directorate of Public Purchases. As was mentioned before, most government units are mandated by the law to develop their procurement process via the Mercado Public platform, and even those who do not must publish a basic set of information to the database.

Each observation corresponds to the monetary amount submitted by a firm for a specific project or good. Each observation includes project characterization variables, auction characterization variables, and firm characterization variables. Note that first two sets of variables are the same within bids for the same contract. Relevant characterization variables include the auctions' date, geographic region, the type of project, and the procuring government unit. Regarding the project itself, the data includes the product units being auctioned, and an official estimate of the contract amount (for most projects). For each submitted bid we also get the corresponding firm's name, unique tax identifier of the contractor (RUT), bid amount, whether it was accepted and whether it was awarded the project or not.

The Open Data Platform has available year/month data in the form of .csv files, which were downloaded and merged to form the initial raw dataset. Here, each observation is a proposal sent by a firm to a certain public auction. This dataset has

10,000,000 observations, which span a wide array of different types of goods purchased by the government. The current investigation employed a standardized classification field to filter only projects of the Category “Construction Projects and Services”. This filtering steps renders observations.

Furthermore, we perform additional filtering steps. We filter only single item projects and drop projects below the threshold of CLP. This aims to exclude excessively simple projects, like small repairs, which do not entail either relevant bureaucracy or public-specific domain expertise. We end of with observations. We end up with observations, submitted for unique projects. We also have unique contractors, developing on average contracts each.

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. First, we expect that contracts with surrounding conditions that make published information will not be present, e.g., national security reasons. This is the case for purchases related to national security, for example, the construction of naval bases. 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. The following table show descriptive statistics for projects, bids, and firms found in the final sample. We see that there is considerable variance in project sizes.

We show in figure that the sample covers fairly all years in the sample, although as it would be expected, number of projects have been increasing since 2010. Furthermore, we see that projects span the full set of geographic Chilean administrative regions.

We advance some basic characterizations regarding the types of government units procuring these projects and the types of projects in the sample. 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.