Online Data Appendix

Distortion by Audit:

Evidence from Public Procurement

Maria Paula Gerardino[*] Stephan Litschig[†] Dina Pomeranz[‡]

January 2022

[*]Inter-American Development Bank, mariage@iadb.org.
[†]National Graduate Institute for Policy Studies, s-litschig@grips.ac.jp.
[‡]University of Zurich and CEPR, dina.pomeranz@econ.uzh.ch.

A Data preparation

This appendix includes a detailed description of the data sources used in the study, the merge process across datasets, the construction of the estimation sample, additional information used to measure political affiliation (used as a control variable), and the data processing involved for the analysis of unit prices.

The main data used in this study were provided by two different Chilean government agencies: (i) Contraloría General de la República (CGR) and (ii) ChileCompra (CHC). We also use data from the Servicio Electoral de Chile (SERVEL) to construct some control variables. Table A1 describes these sources and the particular information provided by them.

Table A1:
Data sources

Source	Code	Information provided	Years
Contraloria General de la Republica	CGR	Classification of entities regarding compliance risk and relative importance	2011-2012
		Details about performed audits (e.g. date, entity, type of audit)	2010-2013
ChileCompra	CHC	Records on number of purchases and amounts purchased by modality and entity	2009-2014
		Records on conducted auctions (e.g. amount purchased, number of bidders) by auction	2009-2014
Study Audits	SA	Lists of checks performed, irregularities detected, and follow-up measures conducted	2015
Servicio Electoral de Chile	SERVEL	Political affiliation of entity	2008

The main dataset is constructed by merging information from CGR and CHC. The data from CGR includes the score of relative importance and the level of risk (high, medium or low) of each public entity. This is used in a scoring rule to determine which entities are audited in a particular year. This dataset also contains further information about the audited entities. The data provided by CHC contains information about what purchases were made, at what time and through which modalities. The CGR audits data are available for 2011-2012, while the CHC data are available for both prior and later years. The process of merging these datasets is explained in the following section.

A.1 Merging Datasets

The universe of public entities considered by CGR is different from the universe of entities operating on the CHC platform. The differences between the CGR and CHC universes are outlined below:

i) CGR only has one structural level of entities, whereas CHC has two: individual purchasing units and their superordinate institution. Purchasing units in the CHC dataset are identified through the combination (called *set identity*) of both the purchasing unit ID variable called *idunidaddecompra*, as well as the variable *idinstitucion* that identifies the superordinate institution the unit belongs to.

ii) Even when adjusting for the different levels of details, the entities employed by CHC and CGR do not overlap completely. For example, some CGR entities do not correspond to any CHC entity (i.e. public enterprises that do not purchase through CHC). Conversely, some CHC purchasing units do not correspond to any CGR entity and therefore do not have scores of risk and relative importance calculated.

In order to merge the two datasets, we need to establish a mapping of CGR entities to corresponding CHC purchasing units. Therefore, 1437 and 1709 entities scored by the CGR in 2011 and 2012 respectively had to be matched ex post to the CHC universe. This was done in two consecutive steps as described below.

Step 1: Assignment of CHC Set Identity

First, CGR staff attempted to amend the CGR dataset by assigning corresponding CHC institution and purchasing unit IDs to each CGR entity based on matching entity names. For each CGR entity, the following steps were performed:

i) Try to find an equivalent CHC purchasing unit (as identified by *idunidaddecompra*). This direct assignment was possible for 529 and 747 CGR entities in 2011 and 2012, respectively.

ii) If the assignment of a corresponding CHC purchasing unit was not possible for a given CGR entity, but a matching superordinate CHC institution (identified by *idinstitution*) was found, the latter was assigned. This was the case for 908 and 962 CGR entities in 2011 and 2012, respectively.

iii) If neither CHC purchasing unit nor institution could be assigned, it is assumed that the CGR entity did not conduct any purchases through CHC in the given time period. This was the case for 41 and 40 CGR entities in 2011 and 2012, respectively.

Due to the aforementioned lack of a full correspondence between the two universes, two issues arise after this first assignment step:

i) In those CGR entities where only a corresponding CHC institution could be assigned, the institution's ID was transcribed both as the CGR entity's corresponding CHC institution and purchasing unit ID. However, this is never the case in the original CHC dataset and will be addressed below. Overall, 908 and 962 CGR entities have identical institution and purchasing unit IDs in 2011 and 2012, respectively.

ii) Some CHC set identities have been assigned to more than one CGR entity, which renders the CHC set identity a non-unique identifier within the amended CGR dataset. The case of a non-unique set of CHC institution and purchasing unit IDs appears for 300 and 303 entities in 2011 and 2012 respectively.

Most of the entities affected by issue i) are also affected by issue ii): 276 (out of 300) and 282 (out of 303) CGR entities with non-unique IDs in 2011 and 2012 respectively could also not be assigned to a CHC purchasing unit (instead, only to a CHC institution). For example: the entities *Administrative Management of the Air Force of Chile* and *Personal Air Force Command* represent different CGR entities. However, they were both assigned the same CHC set identity (idunidaddecompra and idinstitucion). In addition, these entities also have identical values for both CHC ID variables.

Step 2: Mapping Amended CGR Data to CHC Data

In order to merge the amended CGR dataset with the CHC dataset, the CGR entities with and without assigned CHC purchasing unit have to be merged separately. This is because those entities with assigned purchasing unit ought to be merged on *idunidaddecompra*, whereas those without have to be merged on *idinstitucion* for the lack of a more refined identifier. We proceeded in two steps:

i) First, we merge the CHC dataset with the CGR dataset on *idunidaddecompra* using a *m:1*-type merge. This is because a given CHC purchasing unit ID is unique within the CHC dataset, but may be assigned multiple times in the amended CGR dataset. Only successfully matched entities are saved. CGR entities that had been assigned a CHC purchasing unit ID may still not have a match if they did not purchase at all through CHC in a given year. For these entities, purchases are set to zero (a further match on the CHC institution ID, is not attempted for these entities). Table A2 shows the frequency of these cases: First for all entities with a medium risk score, and then for the subsample within a 10 point range around the cutoff in the relative importance score (between low/medium in 2011 and medium/high in 2012).

ii) In the second step, CGR entities that were only assigned a CHC institution ID, are merged with the so far unmatched CHC purchasing units on *idinstitucion*. Both matched and unmatched CGR entities are saved. If entities were not matched in this step, then they had been assigned a CHC institution ID, but this ID was not found in the CHC dataset. For these entities, purchases are again assumed to be zero in a given year. Table A3 shows the frequency of this case, first for all entities with a medium risk score, and then for the subsample within a 10 point range around the cutoff in the relative importance score.

   Moreover, in contrast to step i), step ii) involves a *m:n*-type merge. This is necessary since CGR entities are not assigned to CHC institutions uniquely (e.g. 15 regional entities are all assigned to a big institution). Furthermore, the CHC institution also is not a unique identifier within the CHC dataset. Therefore, a merge on the CHC institution ID may match multiple CGR entities to multiple CHC purchasing units. Since about one third of the CGR entities involved in this step have non-unique CHC set identities (their CHC institution ID is shared with other CGR entities), this step creates duplicate matches.[1] In these instances, a given value associated with a

[1] The medium risk sample contains 154 entities with non-unique CHC IDs, and in the subsample with

particular CHC purchasing unit may be assigned to multiple CGR entities. Thus, the values associated with this CHC purchasing unit (e.g. total amount of purchases) would be listed multiple times in the merged dataset. To conserve the integrity of the total values of these variables across CGR entities, the values of some CHC variables are divided by the number of CGR entities mapping to said CHC purchasing unit. This assumes that each of the CGR entities sharing the reference to one particular CHC purchasing unit is accountable for an *equal share* of the CHC purchasing unit's variable value. Other CHC variables, such as shares of purchases by modality, are not affected by being matched multiple times.

Table A2:
Step (i): Merging on *Idunidaddecompra*. Entities Without Purchases

	CGR 2011					CGR 2012				
CHC	2009	2010	2011	2012	2013	2010	2011	2012	2013	2014
	Panel A: Medium risk entities (1673 entities)									
Number of entities	7	3	0	2	4	6	5	0	5	6
	Panel B: Medium risk entities in +/-10 range around cutoff (1002 entities)									
Number of entities	4	3	0	1	3	1	1	0	2	2

After merging, the monthly values of entities' CHC variables are summed up in order to collapse the dataset by CGR entity and year or quarter. 1491 entities in 2011 and 1762 entities in 2012 remain.

A.2 The Estimation Sample

The final estimation sample results from dropping the following entities:

i) CGR entities that were assigned neither a CHC purchasing unit ID nor a CHC institution ID (42 entities in 2011 and 41 entities in 2012). For these entities, no merge of CHC data was attempted in the first place.

cutoff between +/- 10 (estimation sample) there are 58 entities with non-unique IDs.

Table A3:
Step (ii): Merging on *Idinstitucion*. Entities Without Purchases

	CGR 2011					CGR 2012				
CHC	2009	2010	2011	2012	2013	2010	2011	2012	2013	2014
	Panel A: Medium risk entities (1673 entities)									
Number of entities	9	2	0	0	3	3	1	0	1	2
	Panel B: Medium risk entities in +/-10 range around cutoff (1002 entities)									
Number of entities	3	0	0	0	0	1	0	0	0	0

ii) CGR entities that matched (*m:n*-type on the assigned CHC institution) to more than *13* CHC purchasing units. Merging on the CHC institution ID can yield a large number of corresponding CHC purchasing unit matches for any given CGR entity since the institution ID is not unique within the CHC dataset. We decided to drop the most extreme instances. In particular, entities were dropped when the number of their CHC purchasing unit matches registered above the 95th percentile (13 purchasing unit matches). Thus, 52 entities were discarded in 2011 and 47 entities in 2012.

iii) Entities with a risk score of low or high as designated by CGR. This is because for these entities audit risk was generally low or high, irrespective of the relative importance score (as described in section 2.2 of the paper). Consequently, 601 entities were dropped in 2011 and 645 in 2012.

iv) Successfully matched entities that did not conduct any purchases in 2011 or 2012. This resulted in elimination of 76 entities in 2011 and 76 entities in 2012.

The final sample employed for the main analysis of this paper contains 720 entities in 2011 and 953 entities in 2012.

A.3 Additional Data

A.3.1 Audits Data

CGR provided further information on audits carried out between 2010 and 2013. This data contains variables indicating whether a given entity was audited in a given reference year (t, 2011 or 2012), one year prior (*t-1*), or one year after (*t+1*).

A.3.2 Study Audits

In order to analyze the details of the auditing process, CGR agreed to conduct a number of additional audits in the context of this study. These audits were performed identically to typical public procurement audits, with the difference that auditors recorded more information than usual. For example: which contracts were audited; which checks were conducted; which infractions were detected. This information was recorded at the purchase level.

These expanded study audits were implemented in 2015. They took place in two waves (one in July and the other one in September). 18 entities were audited and in each audit three purchases of goods and three purchases of services were inspected (for a total of 108 audited purchases). The audit protocol involved 95 different checks, most of which referred to aspects of contract awarding or contract execution.

A.3.3 Data on Political Affiliations

We account for the political affiliation of a given CGR entity as a control variable, utilizing data data from SERVEL. We assign political affiliation as follows:

i) For municipal entities, we assign the political party of the mayor of said municipality according to the SERVEL data.

ii) For entities of the regional or national administration, we assume a "right-wing" affiliation since a right-wing governing coalition was in power in 2011 and 2012 at the national level.

A.3.4 CHC Data at the Auction Level

For the competitiveness analysis, we examine auctions with at most three bidders separately from those with more than three bidders. The information on the number of bidders comes from an additional CHC dataset at the auction level. We perform the same merging process between this dataset and the CGR dataset as described in section A.1 above.

A.4 Price Analysis Data

The sample used in the price analysis has particular characteristics, which we describe in this section. For each line item of the purchase orders dataset we have information about the quantity of units that the line item contains and the monetary value of the line item. We divide the monetary amount by the quantity to obtain unitary prices. We also have a field where procurement officers have to indicate what the unit of measurement is. A challenge of this analysis is that many of the reported measurement units are non-standard. For instance, we have measures such as bottles, kits, cans or pallets. When the unit of measurement is not well defined, the quantity field is ambiguous, so unitary prices have no clear interpretation. Given this, we restrict the universe of line items to those where the unit of measurement is well defined.

As detailed below, this severely restricts the sample on which we can study effects on prices. Furthermore, when the data is restricted to purchases with clear unit of measurement, we no longer find a shift in the procurement modalities as a consequence of being audited. Given this, we proceed in the following two steps:

1. Restrict the data to purchases of products with clear unit of measurement, and then estimate the impact on procurement modalities separately for each product category.

2. Estimate the impact on unitary prices, restricting the sample of line items to products with clear unit of measurement and classified in product categories where we find a more pronounced modality shift. Specifically, we focus on product categories where the modality shift is larger than the median shift.

In what follows we describe the above procedure in more detail.

A.4.1 Sample restrictions

The universe of purchase orders carried out by entities in the middle risk category in the years 2011 and 2012 combined contains $5,509,379$ line items. We drop from this universe all the line items that do not have a clear unit of measurement, which leaves us with $328,522$ line items. The following shows which measurement units where classified as clear and not clear:

i) **Clear Unit of Measurement**: foot, inch, milimeter, centimeter, meter, mililiter, cubic centimeter, cubic meter, gallon, liter, microgram, miligram, pound, ounce, kilogram, ton, square meter and International Unit.

ii) **Not Clear Unit of Measurement**: blister, bucket, balloon, tray, bar, drum, block, bag, bottle, box, packet, drawer, load, reel, cartridge, cylinder, tablet, vaginal tablet, cone, capsule, disk, dose, flask, global, dragée, sheet, kit, can, foil, ball, pack, pallets, bread, paper, piece, plate, rack, ream, roll, sachet, sack, envelope, suppository, checkbook, drum, jar, tin, strip, tube, unit, undefined unit, hour, day, week, month, year, pair, dozen, fifteen, hundreds, person-hour, person-day, person-month, thousands.

We also drop remaining line items classified in product categories that correspond to services ($16,353$ line items), or where the product classification is missing ($10,927$ line items), as unitary prices are ambiguous in these cases.

Next, we keep only those purchases that were bought through the modalities of auctions or direct contracting, given that we do not find shifts in the modalities of framework agreement and small purchases. After all of the above restrictions, the number of line items from 2011 and 2012 combined is $269,586$, or 7.6% of the full universe of line items bought through auction or direct contracts by middle-risk entities.

A.4.2 Step 1: Estimation of Modality Shifts By Product Category

We use the above data to compute entity-level modality shares specifically for each product category in the data. That is, for purchases of, say, computers, we compute how much of the total value of computer purchases was bought through auctions and how much through direct contracting. With this information, we estimate impacts on modality shares separately for each product category, following the same regression specification as in the main analysis of the study. As a robustness check, we do this analysis restricting the sample of entities with two alternative bandwidths: ± 4 and ± 10.

Some products are bought by a very small number of entities, and insufficient degrees of freedom make it infeasible to estimate the impact on these products. This happens in a large number of product categories, but these products represent a relatively small proportion of line items. Specifically, when using entities in the ± 4 bandwidth to estimate product-specific modality shifts, we are only able to estimate the modality shift for 110 products out of $2,431$ products. However, the products for which we obtain estimates contain $206,474$ line items, out of $269,586$ line items in total. Similarly, when using a ± 10 bandwidth to estimate product-specific modality shifts, the number of products for which we obtain estimates is 195 out of $3,286$, and the number of line items contained in these products is $224,818$ out of $269,586$ line items in total.

A.4.3 Step 2: Estimation of Impact on Unitary Prices

The price analysis is carried out using a dataset at the level on lines items, which is the level at which we can observe unitary prices. The preparation of this dataset involves the following steps:

i) Restrict the universe of 2011 and 2012 line items in the same way as in Step 1, by dropping line items that (i) correspond to products with no clear unit of measurement, (ii) have no product category information, (iii) are classified as services, or (iv) are purchased through modalities other than auction and direct contract.

ii) Given that no modality shift is found in the above subset of purchases, we restrict the data to line items that belong to product categories where the modality shift is greater than the median shift. When the product-specific modality shifts are estimated using the ± 4 bandwidth, the final dataset contains 56 distinct products and $65,821$ line items. When a bandwidth of ± 10 is used, the final dataset contains 99 products and $79,420$ line items.

B Justifications for Direct Purchases

There are 21 justifications for direct purchases according to Chilean law number 19.886:

1. Acquisition originated in an unsuccessful private auction, where no bidders showed up (point 1)
2. Remnant of unfinished contract due to breach of contract from original supplier or other causes, when remnant does not exceed 75,000 USD (point 2)
3. Emergency or unforeseen urgency (point 3)
4. There is only one supplier of goods or services (point 4)
5. Contracting with foreign legal persons outside the national territory (point 5)
6. Confidential services (point 6)
7. Extension of utility contracts or ancillary services (point 7.a)
8. Procurement under representation expenses (point 7.b)
9. Hiring a specific supplier for the safety and integrity of authorities (point 7.c)
10. Hiring of consultancy services, considering the special qualities of the provider (point 7.d)
11. Contracting with holders of intellectual or industrial property (point 7.e)
12. Trust and security of providers, derived from their experience (point 7.f)
13. Replacing or complementing compatible accessories for models already acquired (point 7.g)
14. Public awareness of the tender could jeopardize the goal of the contract (point 7.h)
15. Procurement of goods to foreign suppliers for use outside the country (point 7.i)
16. Purchase is below 7,500 USD and cost of evaluating bids is disproportionate (point 7.j)
17. Goods or services for teaching or research projects (point 7.k)
18. Public auction without prior public tender offers, or inadmissible deals (point 7.l)
19. Specialized services below 75,000 USD (point 7.m)
20. Acquisition of less than or equal to 750 USD (point 8)
21. Exceptions for being required to operate in the online procurement system as specified in article 62