Estimating the Effect of Discretionary Spending on Corruption: Evidence from Brazilian Municipalities*

Andre Assumpcao[†] Ciro Biderman[‡] George Avelino[§]

June 10, 2018

Abstract

Placeholder

Keywords: corruption, procurement, discretion.

JEL Codes: D73, K10, H57, H72, O12, P48.

*abc

[†]Department of Public Policy, The University of North Carolina at Chapel Hill. Email: aassumpcao@unc.edu
[‡]Center of Politics and Economics of the Public Sector, Getúlio Vargas Foundation. Email: ciro.biderman@fgv.br
[§]Center of Politics and Economics of the Public Sector, Getúlio Vargas Foundation. Email: george.avelino@fgv.br

1 Introduction

2 Background and Data

2.1 Data

2.2 Empirical Strategy

3 Results

3.1 Falsification Tests

4 Conclusion

4.1 Cost-benefit Analysis

References

Assumpcao, A. (2018). Textfind: A Data-Driven Text Analysis Tool for Stata.

Grimmer, J. and Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(3):267–297.

Hopkins, D. and King, G. (2009). A Method of Automated Nonparametric Content Analysis for Social Science. *American Journal of Political Science*, 54(1):229–247.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698–714.

A Appendix: Service Order Classification

Service orders issued by CGU investigated different uses of public resources in addition to procurement, e.g. for officials compensation, for school activities, or for community monitoring of public policies. The discretion measure proposed here, however, is exclusive to procurement expenditures made under Law 8,666/93. The ideal dataset for this study would contain explicit procurement information collected by CGU auditors, but unfortunately this is not the case. The reporting of procurement processes is implicit, via descriptions of investigations or findings of violations to Law 8,666/93. Thus, we isolate service orders which investigated procurement processes from the rest by implementing an classification system based on the information retrieval and natural-language processing literatures.

The system uses each service order's description to identify if it is procurement-related. In these descriptions, CGU auditors report the purpose of their investigation, e.g. whether they are looking into painkiller purchases, whether the municipality has used the funds within designated goals, or whether primary school teachers were hired for the implementation of a school program. Using these textual descriptions as bag-of-words models, we implement a method similar to that of Hopkins and King (2009): we stem and combine unigrams to form search patterns that identify a service order as procurement-related. There are two broad types of procurement in Law 8,666/93: (i) ordinary procurement of goods and services, which we call *purchases*; and (ii) procurement of goods and services used for public works, which we call *works*. There are different search patterns for each type.

An example is useful for understanding our classification process. Unigram "aquisição" (*acquisition* in English) is stemmed to "aquisi" to form a search pattern for the *purchases*-type procurement; unigrams "adequação" and "habitacional" are stemmed and combined to form "adequa(.)*habitac"[1] search pattern for *works*-type procurement. This bigram picks up variations in main keywords as well as coding mistakes due to, for instance, multiple whitespace between the two unigrams or due to coding Portuguese special characters ("adequação" vs. "adequacao").

Table 1: Procurement Search Terms

<table>
<tr><th>Type</th><th>Search Terms</th></tr>
<tr><td>Purchases</td><td>"aquisi" "execu" "equipame" "ve[íi]culo" "despesa" "aplica[çc]" "medicamento(.)*peaf" "compra" "recurso(.)*financ" "unidade(.)*m[óo]ve(.)*sa[úu]de" "pnate" "transporte(.)*escola" "desenv(.)*ensino" "kit" "siafi" "implementa[çc]" "adquir" "pme(.)*2004" "aparelhamento"</td></tr>
<tr><td>Works</td><td>"co(ns|sn)tru" "obra" "implant" "infra(.)*estrut" "amplia" "abasteci(.)*d(.)*[áa]gua" "reforma" "(melhoria|adequa)+(.)*(f[íi]sica|escolar|habitac|sanit[áa]ria)+" "esgot" "adutora|dessaliniz|reservat[óo]" "sanit[áa]ri[ao]" "poço" "aperfei[çc]oa" "saneamento" "res[íi]duo(.)*s[óo]lido" "conclus[ãa]o"</td></tr>
</table>

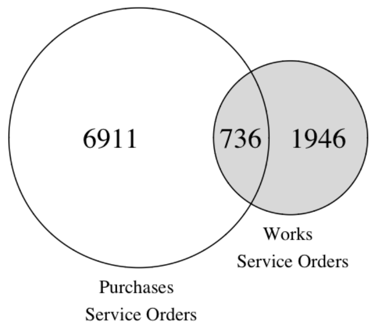
The final list contains 19 n-grams for identification of purchases and 17 n-grams for works.[2]

[1] All seach patterns are regular expressions.

[2] One of these keywords in the works search pattern is an "exclusion keyword," which removes service orders that contain the "exclusion keyword" in their description from the sample identified by the other 16 n-grams.

When any of these words is found, we include the service order into the purchases or the works group. Since all public works projects procure goods and services but not all public purchases are works-related, whenever the search patterns matches service orders to both groups, we include the service order only in the works group but not in the purchases group. Public works procurements are a subset of all public procurements in Brazilian municipalities. The search patterns here identify a total of 9,593 procurement-related service orders.

Figure 1: Sets of Procurement Service Orders



As Grimmer and Stewart (2013) rightly point out, no text analysis algorithm is perfect and only relying on keyword matches could potentially lead to misclassification of service orders. Let us suppose that one description reads "expenditures made in accordance with primary education program." Using unigram "expenditure" would yield a match for this service order to the purchases group, but in fact auditors might be looking at bonus payments for high-performing teachers. These resources could also be directed for school construction. In the first case, the service order should not be have been included in any group because it does not carry any procurement component. In the second case, it should have also been marked as public works.

We address these classification problems in three ways: (i) using means comparison tests of match quality discussed in Assumpcao (2018); (ii) comparing the performance of the same search patterns on another textual description for a subset of service orders; (iii) finally, comparing the results from the textual classification algorithm to that of procurement violations reported by CGU auditors. We discuss these three tests in turns in the following sections.

A.1 Means Tests

The first test on match quality is the means comparison test presented in Assumpcao (2018), whose reasoning is simple. Increasing the number of procurement-related terms in the search pattern is not necessarily good practice as we increase the chance of misclassifying service orders as procurement when in fact they are not; words can take on different meanings depending on their contexts, so the more search terms we use the more likely type I error is. Ideally, we would want to use as few n-grams as possible while still identifying all possible procurement matches. In order to do this, what Assumpcao (2018) suggests is testing match quality by incrementally comparing sample means identified by n vs. n - 1 keywords. This method translates into a check on whether the sample identified by one additional keyword is significantly better than the previous sample with one fewer term. The program developed by Assumpcao (2018) does this for us and we report the results in the tables below:

Table 2: Purchases Search Results

	Total	Average				Means
	Finds	Find	Length	Position	TF-IDF	test p-value
“aquisi”	3716	1.052	27.757	4.649	0.084	.
“execu”	2261	1.190	47.662	13.173	0.075	0.000
“equipame”	1117	1.005	60.645	39.853	0.168	0.000
“ve[íi]culo”	717	0.713	38.969	11.948	0.094	0.000
“despesa”	667	1.006	40.856	19.474	0.110	0.000
“aplica[çc]”	604	0.846	24.603	11.389	0.135	0.000
“medicamento(.)*peaf”	570	3.367	13.167	.	0.794	0.000
“compra”	449	1.002	5.178	2.323	2.305	0.000
“recurso(.)*financ”	425	1.599	33.416	.	0.183	0.000
“unidade(.)*m[óo]ve(.)*sa[úu]de”	364	0.897	9.365	.	0.384	0.000
“pnate”	283	1.000	22.544	21.484	2.186	0.000
“transporte(.)*escola”	201	1.360	18.493	.	0.411	0.000
“desenv(.)*ensino”	167	5.054	37.168	.	0.658	0.000
“kit”	134	1.067	7.836	3.590	1.292	0.000
“siafi”	124	1.016	18.008	7.298	1.333	0.000
“implementa[çc]”	94	0.794	32.032	4.904	0.130	0.000
“adquir”	68	1.338	29.015	17.250	0.355	0.000
“pme(.)*2004”	67	0.727	5.657	.	1.711	0.000
“aparelhamento”	4	1.000	11.250	2.500	0.716	0.045
Total	7647	.	29.263	.	.	0.000

The search terms are sorted in descending order by the number of service orders they identify (column 1). Column 6 displays p-values for means tests across samples, where each mean is the sum of observations found by *any* of the search items before, and inclusive of, any particular row over the total number of observations.[3] The means test thus compares whether the sample identified by all search terms up to any row is significantly different from the the sample identified by all rows before. For instance, the evidence presented in row four of table 2 is that the inclusion of search item “ve[íi]culo” significantly improves (at the 5% level) the identification of the purchases sample when compared to the sample which only includes the previous three search words.

The works sample is a third of the size of the purchases group and two of its search items do not significantly identify a new sample (“saneamento” and “conclus[ãa]o”). Despite having positive individual finds reported in column 1, table 3, the means test in column 6 suggests that these finds are not new service orders in addition to what had already been identified by the the previous search terms.[4]

Means tests are important to map out the relationship between search items, both within and

[3]This is also known as an alternative search where all search conditions are connected by an “or” statement.

[4]The search without these terms (available upon request) yields 2,679 service orders, just three short of the total in table 3. Nevertheless, we keep the two items in the search algorithm for additional tests discussed in section A.2.

Table 3: Works Search Results

	Total	Average				Means
	Finds	Find	Length	Position	TF-IDF	test p-value
“co(ns|sn)tru”	954	0.597	21.822	4.283	0.153	.
“obra”	877	1.003	12.754	7.023	1.658	0.000
“implant”	767	1.021	50.811	4.001	0.074	0.000
“infra(.)*estrut”	614	0.859	88.894	22.000	0.055	0.000
“amplia”	366	1.000	39.109	6.615	0.144	0.000
“abasteci(.)*d(.)*[áa]gua”	333	0.996	31.156	.	0.175	0.000
“reforma”	307	1.029	14.704	6.316	0.429	0.000
“(melhoria|adequa)+(.)* (f[íi]sica|escolar|habitac|sanit[áa]ria)+”	279	1.360	38.315	.	0.128	0.000
“esgot”	255	1.024	37.035	31.412	0.187	0.000
“adutora|dessaliniz|reservat[óo]”	170	0.303	48.871	20.253	0.031	0.045
“sanit[áa]ri[ao]”	541	0.626	29.115	9.839	0.141	0.000
“poço”	58	1.000	47.017	14.190	0.135	0.025
“aperfei[çc]oa”	35	0.769	33.257	19.029	0.141	0.000
“saneamento”	24	1.000	38.000	23.083	0.755	0.317
“res[íi]duo(.)*s[óo]lido”	21	4.455	62.619	.	0.429	0.045
“conclus[ãa]o”	4	0.750	25.750	8.000	0.276	0.157
Total	2682	.	34.882	.	.	0.000

across groups, but they do not tell us anything about the relationship between search items and their latent procurement groups. In other words, the search terms might be picking up groups that are internally consistent but that do not map onto the procurement types in Law 8,666/93. We discuss these issues in sections A.2 and A.3.

A.2 Textual Descriptions

CGU service orders can best be described as investigations on the use of public resources transferred from the federal government to Brazilian municipalities. There are six transfer types and each service order investigates only one type at a time. Since the procurement categories set out in Law 8,666/93 apply to all public procurements at all government levels, transfer types are irrelevant for constructing our discretion measure. Nonetheless, one type of these transfers helps test our classification algorithm.

Federal grants (*convênios* in Portuguese) are narrow transfer agreements signed by the federal government, its agencies, states and municipalities for the delivery of governmental programs. They are voluntary, time-limited transfers implementing policies at the local level, such as vaccinations and the construction of community health clinics. The most important feature of these grants, however, is that each of them also has an individual textual description of its purpose, e.g. a tractor purchase for a rural community in a given municipality. Thus, for a subset of service orders that are investigations of the use of these federal grants by Brazilian municipalities, we have two different textual descriptions of resource use: CGU’s, from their audit report, and the federal government's, available online at the Transparency Portal.[5]

Table 4: Classification by Grant Description

Panel A: Purchases Group				
		Grant Description		
		No	Yes	Total
Service Order Description	No	115	144	259
	Yes	83	1473	1556
	Total	198	1617	1815

Panel B: Works Group				
		Grant Description		
		No	Yes	Total
Service Order Description	No	1546	269	1815
	Yes	404	1649	2053
	Total	1950	1918	3868

There is a total of 3,868 service orders for which we have descriptions both from CGU and from the federal government. In table 4, we report the results of the search algorithm both in the service order (row-wise) and the transfer (column-wise) descriptions. We evaluate the performance of the search algorithm by checking whether it assigns the same service order to the same procurement group *regardless of the description in which it searches for the key terms.* In other words, the smaller the number of times that the algorithm assigns any service order to a different group when it switches to another textual description, the better.

This is a particularly important point for the classification method proposed here. The means test conducted in section A.1 provides internal consistency because it compares and checks whether more observations are matched when more search terms are included; the tabulation across descriptions here provides external consistency because it compares and checks if the classification algorithm is independent of search target (description). It resembles a false positive (type I error) test because we can roughly calculate the percentage of misclassification of service orders. In panel A, the service order description search assigns 1,556 to the purchases group, out of which 83 were not simultaneously assigned to the same group in the grant description search, yielding a 5.3% false positive rate. In panel B, the service order search marks 2,053 observations to the works group, where 404 are not simultaneously marked when the search is performed in the grant description (a 19.7% type I error rate).[6]

[5] `http://www.portaltransparencia.gov.br/`

[6] The inverse misclassification rates are also reassuring: false positives are 8.9% and 14.0% for purchases and works respectively when we first classify observations using grant descriptions and then move on to service order descriptions.

A.3 Procurement Violations

Though section A.2 supports external validity by showing that the service order classification is consistent across textual descriptions, we run the last robustness check here using the actual procurement violations reported by CGU.

The findings reported by auditors are coded into 35 infractions of the use of public resources, nine of which violations of procurement rules and one violation of public works rules. Thus, we know with certainty that service orders for which there are any of the nine procurement violations (ten if public works) are in fact procurement-related and should be classified either as purchases, works, or both. As opposed to section A.2, this resembles a false negative (type II error) test on yet another subset of observations for which certain infractions were reported.[7]

Table 5: Classification by Procurement Code

Panel A: Purchases Group

		Procurement Code		
		No	Yes	Total
Service Order Description	No	2487	319	2806
	Yes	6137	3456	9593
	Total	8624	3775	12399

Panel B: Works Group

		Procurement Code		
		No	Yes	Total
Service Order Description	No	2462	344	2806
	Yes	5791	3802	9593
	Total	8253	4146	12399

The total number of service orders with at least one procurement infraction is 3,775 (4,146 if we include the public works infraction), which is the sum of column 2 in table 5, panels A and B. The false negative rate is 8.5% and 9.9%, respectively, for purchases-only and works procurements. This means that 319 and 344 service orders should have been classified as procurement by our textual search algorithm but were not.

Although no text analysis mechanism is perfect, the evidence presented here supports our choice of classification algorithm. The identification of procurement orders is internally consistent (section

[7] The reason why this is a type II error test, instead of type I, resides on the way the test samples are defined. In section A.2, both sample assignments (by matching procurement keywords in the service order or grant description) can be the "correct" procurement sample against which the match on the alternative description might yield false positives. In this section, we know with certainty that the sample identified by procurement infractions is in fact the correctly identified sample, since there cannot exist a procurement violation where no procurement has occurred. It makes the unidentified observations false negatives because they should have been classified as procurement-related service orders. This sample is clearly underidentified, as there are many procurement-related service orders that simply followed Law 8,666/93 and thus carry no infraction, but still, within this subset of all CGU investigations, it provides us with a good conterfactual against which to test our classification mechanism.

A.1), there are very few incorrect assignments of service orders to procurement (section A.2), and the sample which was identified as procurement maps well onto the latent categories in the Brazilian procurement legislation (section A.3).

B CEPESP Infraction Classification

C Service Amount Maniputation

The identification strategy in this paper relies on the assumption that municipal officials do not (completely) manipulate public expenditure amounts in order to avoid stricter procurement rules. In other words, the public procurement processes carried out just below and above any discretion threshold, which is uniquely determined by procurement amount, are equal except for the rules set out in Law 8,666/93 – thus they are good counterfactuals for testing the effect of expenditure discretion on corruption. We present below the McCrary (2008) test for manipulation of running variable for all six discretion cutoffs in the Brazilian procurement legislation.

Except for purchases cutoff three, these tests show that there is no substantial manipulation of expenditure amount as there are no significant differences in the density of service orders around the procurement cutoffs defined in Law 8,666/93. As discussed in section 2.2, we drop purchases cutoff three from the analysis and move ahead with the fuzzy RD as the main identification strategy.

Figure 2: Cutoff Manipulation Tests