

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, thus we need to isolate service orders which investigated procurement processes from the rest. Since there is no such direct reporting in CGU reports, we implement a service order classification system based on the information retrieval and natural-language processing literatures.

We use 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 & 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. The 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”).

Figure 1: Service Order Mapping

- insert table 1: keywords here -

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

Table 1: Procurement Search Terms

- insert figure 1: Venn Diagram here -

As Grimmer & 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 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 issues 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-related violations reported by CGU auditors. We discuss these three tests in turns in the following sections.

¹All search patterns are regular expressions.

²One 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 found by the other 17 *n*-grams.

A.1 Match Quality Measures

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) thus suggests is pairwise means comparison tests between the samples identified by n v. $n - 1$ 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

- insert table 2: Purchases Search Results here -

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 any row is significantly different from 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[i]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.

Table A.3: Works Search Results

- insert table 3: Works Search Results here -

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[ã]o”). Despite having positive individual finds reported in column 1, table A.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 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; however, one transfer type in particular, federal grants (*convênios* in Portuguese), provide an important counterfactual to test whether our textual classification correctly maps onto latent purchases and works groups in the procurement legislation.

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

⁴The search without these terms (available upon request) yields 2679 service orders, just three short of the total in table A.3. Nevertheless, we keep the two items in the search algorithm for additional tests discussed in section A.2.

A.3 Procurement-related Violations

A

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

Grimmer, J., & 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. <https://doi.org/10.1093/pan/mps028>

Hopkins, D., & King, G. (2009). A Method of Automated Nonparametric Content Analysis for Social Science. *American Journal of Political Science*, 54(1), 229–247. <https://doi.org/10.1111/j.1540-5907.2009.00428.x>