

## 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 advanced 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*; (ii) and 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 purchase-type procurement; unigrams “adequação” and “habitacional” are stemmed and combined to form “adequa(.)\*habitac” search pattern for works-type procurement – it picks up variations in main keywords as well as coding mistakes due to, for instance, multiple whitespace between the two unigrams or when coding Portuguese special characters (“adequação” vs “adequacao”).

- insert table 1: keywords here -

The final list contains 19 *n*-grams for identification of purchases and 17 *n*-grams for works.<sup>1</sup> 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.

- 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 but in fact auditors might be looking at bonus payments for high-performing teachers. These resources could also be directed for remodeling schools, in which case we would not know that this service order should be marked to the works group.

We address these classification issues in three ways: (i) using means comparison tests of match quality discussed in Assumpcao (2018); (ii) comparing the same search patterns on another textual description for a subset of service orders; (iii) finally, comparing the results from the textual classification against that of procurement-related violations reported by CGU auditors. The three tests are discussed below.

### A.1 Match Quality Measures

The first test on match quality is the means comparison test presented in Assumpcao (2018), whose reasoning is simple. By adding more procurement-related terms to the search pattern, we run the risk of misclassifying service orders as procurement when in fact they are not – i.e. the typical type I error. Ideally, we want to use

---

<sup>1</sup>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 found by the other 17 *n*-grams.

as few  $n$ -grams as possible while still identifying all possible procurement matches. 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 item.

- insert table 2: Purchases Search Patterns here -

- insert table 2: Works Search Patterns here -

According to tables ??

## A.2 Textual Descriptions

A

## 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>