Placeholder*

Hyunwoo Park† John Smith‡

June 7, 2018

Abstract

Placeholder

Keywords: key1, key2, key3

JEL Codes: key1, key2, key3

*abc
†abc
‡abc

Tables

Figures

Appendix A. 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 & 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 "aquisio" (*acquisition* in English) is stemmed to "aquisi" to form a search pattern for the *purchases*-type procurement; unigrams "adequao" 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 ("adequao" vs. "adequacao").

[1] All seach patterns are regular expressions.