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 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 both variations from the 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).

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