Estimating the Effect of Discretion in Public Spending on Government Performance: Evidence from Brazilian Municipalities*

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

August 2, 2018

Abstract

Placeholder

Keywords: corruption, misallocation of resources, procurement, discretion, government performance.

JEL Codes: D73, K10, H57, H72, H83, O12, P48.

*The authors acknowledge valuable contributions from Sergio Firpo, Caio Costa, Vinicius Brunassi, Guilherme Lambais, Robert Garlick, and Marisa Domino. Excellent research assistance provided by Bruna Pizzolato and interns at the Center of Politics and Economics of the Public Sector (CEPESP) at the Getúlio Vargas Foundation (FGV). Special thanks to Ronald da Silva Balbe from the Federal Secretariat of Internal Control (SFC) at the Ministry of Transparency, former Office of the Comptroller-General (CGU), who provided us with supplementary data for this study. This work was directly supported by the University of North Carolina - Department of Public Policy, and indirectly supported by the National Council of Scientific Research (CNPq), the Coordination for Improvement of Higher Education Personnel (CAPES), and GVpesquisa. All errors are our own.

[†]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

An important discussion in developing countries is whether governments attract qualified professionals to implement policy and efficiently allocate scarce resources. When public officials are competent, there does not seem to exist much opposition to these policymakers having discretion in the allocation of public resources. This is particularly important when it comes to lower-level bureaucrats, who know more about the communities they serve and, thus, are better placed to understand and deliver on the needs of their people (Lipsky, 2010). However, it is often the case that governments attract a heterogeneous pool of workers with high and low quality professionals. Some might even be good professionals but have bad intentions. Incompetence and rent-seeking behavior help explain why officials deviate from their duties and why spending controls are called into action (Lambsdorff, 2002; Olken, 2006).

In this paper, we investigate the relationship between discretion in public spending and government performance. In particular, we look at whether lower discretion reduces corruption and poor management of public resources in Brazilian municipalities. To our knowledge, this is the first study in the economics literature specifically focusing on the effect of discretion on government performance. There are many studies looking at the relationship between economic or political decentralization and corruption (Treisman, 2000; Fisman and Gatti, 2002; Fan et al., 2009; Choudhury, 2015), but they generally face similar issues: they present cross-country evidence, unable to capture *within*-country corruption heterogeneity (Treisman, 2000; Fisman and Gatti, 2002); they employ perception-based (Treisman, 2000; Fisman and Gatti, 2002) and experienced-based (Fan et al., 2009; Choudhury, 2015) corruption measures, both susceptible to non-trivial measurement error; and, finally, they are all dealing with some sort of imperfect measurement of economic and political decentralization (devolution of fiscal revenues and number of government levels, for instance). Moreover, decentralization encompasses many context-specific factors beyond discretion that could jointly determine government performance.

To answer the question whether lower discretion reduces corruption and mismanagement, we exploit the *quasi*-experimental nature of Brazilian procurement Law 8,666, passed by the Brazilian Congress in 1993. It sets out rules for procurement of goods and services for all government branches, at every level, and public agencies, including public companies. The interesting feature of Law 8,666/93, and that which produces exogenous variation in discretion, is that the rules applicable to each procurement call are a deterministic function of the total amount being contracted. In fact, the rules change discontinuously at three monetary thresholds; the higher the global value of the contract, the less discretionary (the stricter) the procurement rules are. Thus, the identification strategy in this study is a regression discontinuity design (RD), and we deploy the novel versions of RD as discussed in Calonico et al. (2014, 2015) and Cattaneo et al. (2016, 2018) due to the nature of multiple cumulative cutoffs in Law 8,666/93.

The corruption model adopted here is similar to that of Olken and Pande (2012) and follows in the footsteps of the crime economics literature (Becker, 1968; Stigler, 1970; Becker and Stigler, 1974; Rose-Ackerman, 1975). However, departing from Olken and Pande (2012), we introduce

three new corruption factors: (i) the discretion in public spending, our most important result, which determines the ease with which public agents can divert funds without being detected; (ii) the global amount of goods and services being contracted, which makes corruption earnings vary at each procurement call; and, finally, (iii) the overall corruption level at each municipality, capturing the cost of being corrupt when everyone else is relatively more (or less) dishonest. Our identification strategy isolates the effect of (i) while controlling for endogenous factors (ii) and (iii).

The third contribution to the literature is the use of objective measures of corruption. We take the research question to the data using the random audits conducted by the Office of the Comptroller-General (CGU) in Brazil. This federal program was established in 2003 and investigates whether earmarked transfers to municipalities have been correctly used. The program has been extensively explored by other scholars in the literature to evaluate the effects of disclosing corruption information on electoral accountability (Ferraz and Finan, 2008); whether electoral incentives influence corruption levels (Ferraz and Finan, 2011); the effect of corruption on the delivery of public goods and services (Ferraz et al., 2012; Lichand et al., 2017); and the total rent extraction due to increased audit risk (Zamboni and Litschig, 2018) or in the presence of a local office of the judiciary branch (Litschig and Zamboni, 2015). We employ similar corruption measures capturing the likelihood that federal transfers have at least one corruption finding, the share of corruption infractions over all infractions for each transfer, and the amount potentially lost to corruption.

Besides testing the effect of discretion on corruption, we also look at the effect on poor management of resources. By establishing nationwide rules for government procurement, Law 8,666/93 could also reduce the mistakes in the management of public resources by making it easier for public officials to stick to policy programs and deliver policy objectives; lower-level bureaucrats face less pressure from colleagues and the general population if they just follow policy guidelines decided at higher-levels of government (which could bear some of the blame if anything went wrong). Thus, a side-effect of fighting off corruption is experiencing an improvement in the management of federal resources across local governments.

Our results show that, despite not reducing corruption, Law 8,666/93 is still a powerful deterrent of poor program management. Moving procurement rules up from direct contracting to invitational bidding significantly reduces the likelihood that auditors find mismanagement problems (41 percentage points), the share of mismanagement problems (29 percentage points) over all problems, and the amount potentially affected by mismanagement (R$4,600, a third of contract value) for any federal transfer audited by CGU. These results are robust to different functional forms, RD bandwidth size, and the inclusion of covariates in local regressions. The effect of discretion on corruption, however, is indistinguishable from zero, regardless of the procurement thresholds or procurement category (public purchases or public works).

The interpretation of these results is straightforward. Even in the presence of corruption, discretion-reducing policies could support the delivery of policy objectives by improving program management. Better program management avoids endless court challenges, delays in the delivery of contracted goods, services, and public works, and moves policy implementation closer to the

intended plan. On the other hand, there are a few explanations for the null effect on corruption. First, the positive incentive for corruption when contract values go up might offset the restrictions imposed by Law 8,666/93. Pocketing larger percentages of contract value might be worth the risk. Second, it might also be that procurement rules are not really that different along contracting amounts. In this scenario, officials would have no trouble moving up supposedly stricter procurement types and still getting by with wrongdoing if they so wish. This interpretation is true even if the indifference towards procurement types comes from (i) learning how to run a public call for goods and services or (ii) from the outdated, deflated thresholds discontinuously changing procurement requirements – the last inflation-adjustment in monetary amounts in Law 8,666/93 was 1998. Finally, the last hypothesis is that punishment for municipal corruption is relatively low for lower-level city officials while most of the burden falls more onto mayors, as it has been documented in Finan et al. (2018). We address each of these interpretations in the discussion of results (section 4).

In addition to assessing the causal relationship between discretion and government performance, controlling for additional determinants of municipal corruption in Brazil, and using objective measures of corruption and mismanagement of public resources, this paper makes a side contribution to the information retrieval literature by proposing new measures of textual match quality, used to identify federal transfers generating procurement calls across Brazilian municipalities. Since there is no organized, public database of municipal procurement calls, we use the algorithm developed in Assumpcao (2018) to infer which transfers are more likely to have generated procurement calls by running a match of procurement keywords on the textual descriptions of each federal transfer. These measures significantly identifid the sample of transfers which generated procurement calls and are robust to many different specifications. They are summarized in appendix A and provide a good alternative for scholars using text analysis to map latent concepts – such as in Lichand et al. (2017).

The remainder of this paper is organized as follows: section II presents similar studies, existing evidence on government performance, corruption in particular, and the data used in this paper. Section III introduces our theoretical model, estimation strategy, and hypotheses. We present and discuss results in section IV. Section V concludes.

2 Background and Data

2.1 Data

3 Estimation Strategy

3.1 Corruption Model

The simple model here is based on Olken and Pande (2012) and posits a straightforward relationship between discretion and corruption. A representative policymaker earns wage w from working in

government and an additional bribe b if she decides to engage in corruption. Her market wage is v, which she receives if she quits, or is fired, from her government position. If she chooses corruption, she must pay her individual dishonesty cost d and subject herself to the probability of detection p. For simplicity, we treat each corruption decision as an independent, one-time decision. Thus, officials will engage in corruption when:

$$w - v < \frac{1-p}{p} \times (b - d) \tag{1}$$

The first important change we make to the model above is the treatment of corruption gains b as a function of procurement amount x. In our context, every procurement call is a one-off opportunity for corruption where officials can earn b and leave the market upon collecting their bribe. In this case, corruption bears little resemblance to the frequent, fixed payments in exchange for continuous favors as described in Olken and Pande (2012), where corruption gains are fixed. As contract value goes up, we can expect that officials will ask for (or will be offered) larger payments to favor one bidder or another. This relationship can also be described as the opportunity cost of corruption, and posits that bribe levels are increasing in the value of goods/services being procured – a \$100,000 contract might see a \$5,000 bribe while a \$50,000 contract might see a \$3,000 bribe.[1]

Secondly, Brazilian procurement rules l impose an additional cost of corruption beyond p and d. As contract amount increases, officials are subject to stricter procurement requirements regarding the number of participants in the procurement process, the documents they have to submit, how long the call for participants should be open, and other issues as described in section 2.1. Procurement categories determine the rules each official has to follow and, as a consequence, the opportunity cost for corruption. For instance, stricter rules offset the lower opportunity cost of corruption; we discuss their relationship in section 3.4.

The last novel contribution here is modeling individual corruption decision as endogenous. In this setting, officials behave strategically: if they see their colleagues engaging in dishonest exchanges and getting by unnoticed, they are also more likely to be corrupt themselves simply because their cost of being honest is higher. Holding constant the ability to uncover and prosecute corruption cases (due to local resource constraints), the probability of detection of any wrongdoing decreases with the number of dishonest people and the number of illegal actions in each municipality. In other words, being honest in a corrupt environment is relatively costly. Ferraz and Finan (2008), Winters and Weitz-Shapiro (2013), and Chong et al. (2015) document a similar behavior according to which voters discount less the electoral punishment for dishonesty when corruption is endemic in local elections in Brazil and Mexico than when it is not. There is no reason to think that local government officials would behave differently.

[1] We additionally assume that officials are risk-averse since the larger the contract the larger the public attention for any given procurement call.

After including these factors into equation 1, the corruption function then becomes:[2]

$$\begin{aligned} b &= f(w,\ v,\ p,\ d,\ x,\ l,\ r) \\ &\nabla f'(-,+,-,-,+,-,+) \end{aligned} \tag{2}$$

Where b is a corruption measure as a function of all previous variables (w, v, p, d) plus procurement amount x, procurement rules l, and municipal level of corruption r. The representative official wants to maximize b with respect to all right-hand side variables and their hypothesized relationship is summarized by f's first derivative signs in line 2 of equation 2. The efficiency wage hypothesis predicts that higher public sector wages w would reduce corruption by increasing the returns to honesty and making illegal options less attractive. The same rationale applies to market wages v, wherein an increase in earnings would have to be met by higher returns to public office either in wages (relatively rigid) or bribes (relatively flexible) to clear public and private labor markets. Finally, the corruption literature suggests that both the probability of detection p and dishonesty costs d (which is an unobservable individual cost function including the severity of punishment) are important factors for corruption deterrence (Becker, 1968; Rose-Ackerman, 1975).[3]

The last three variables in equation 2 are the centerpiece of our model. We anticipate that procurement amount x and overall municipal corruption r are positively correlated with one's decision to engage in corruption and jointly offset the efforts to reduce corruption by imposing stricter, less discretionary rules on public spending in Brazil. Moreover, x and r are hypothesized to have negative second derivatives, representing the official's risk-aversion, while l should have a positive second derivative. In other words, higher-order procurement types are less effective at reducing corruption than lower-order types. These relationships are described below.

$$(i)\ b_x > 0; b_{xx} < 0 \qquad (ii)\ b_l < 0; b_{ll} > 0 \qquad (iii)\ b_r > 0; b_{rr} < 0 \tag{3}$$

The analysis in this paper focuses on the partial derivatives in (ii), and we are interested in potential endogenous relationships between l, x, and r. According to Law 8,666/93, contracting rules are a deterministic function of procurement amount, so we know with certainty that l is a function of x. On the other hand, procurement types in the Brazilian legislation are independent of any particular government corruption, so we know that l does not depend on r. The only effect of the existing corruption level on procurement type works through the contracting amount (in other words, $l'(r) = 0$). The direction of this effect, however, is unclear: although a more corrupt body of public officials might want larger contracting values so that their cut compensates the risk of detection, we cannot rule out a negative effect on procurement amount if corrupt, but smart, officials would know best and keep their corruption earnings low enough not to draw any attention

[2]There is no particular functional form determining the relationship among all variables in our framework. We mostly focus on the most important factors summarizing the relationship between discretion and corruption.

[3]Article 37 of the Brazilian Constitution, Law 8,666/93, and the related legislation lay out the punishment for individual corruption in the public sector. Severity, however, does not change over the period under analysis (2004-2010) and, as such, is irrelevant for the model developed here; i.e. the constant term in the regression equations in section 3.3 soaks up any severity effects.

to their actions. In the section 3.3, we discuss the identification strategy that allows us to isolate the causal effects of discretion on corruption.

3.2 Program Management Model

In addition to reducing corruption, the Brazilian procurement legislation could lead to a more efficient delivery of federal policies in two ways. First, it could reduce the volume of resources lost to corruption and, thus, make more funds available to deliver more and better federal programs. The second channel through which procurement rules could reduce poor resources management is by standardizing the procurement of goods and services and avoiding court challenges, delays in the delivery of contracted goods, services, and public works. This effectively means adhering to federal-mandated programs and to procurement requirements in Law 8,666/93. Fortunately, CGU audit reports contain both evidences of corruption and program management in the form of procurement and policy infractions; in other words, the adherence of each earmarked transfer to their federally-intended use.

$$\begin{aligned} m &= f(w,\ v,\ p,\ d,\ x,\ l,\ r) \\ \nabla f' &(-,+,-,-,+,-,+) \end{aligned} \tag{4}$$

$$(i)\ m_x > 0; m_{xx} < 0 \qquad (ii)\ m_l < 0; m_{ll} > 0 \qquad (iii)\ m_r > 0; m_{rr} < 0 \tag{5}$$

This relationship is easily translated to our model and is symmetric to the corruption framework. Since corruption and mismanagement are alternative problems in our research setting, i.e. any infraction that is not reported as corruption is reported as poor program management, we can use the same corruption determinants as factors in the mismanagement model. Equation 4 summarizes this idea, where we should expect the same results for derivatives of the m function (shown in equation 5). Therefore, we analyze two sets of outcomes, corruption and mismanagement.

3.3 Research Design

$$y_i = \alpha + \gamma_1 x + \gamma_2 x^2 + \delta_1 r + \delta_2 r^2 + \rho_1 l_1 + \rho_2 l_2 + \rho_3 l_3 + \varepsilon \tag{6}$$

$$y_{i,j} = \alpha + \gamma_1 (x - c_j) + \gamma_2 (x - c_j)^2 + \rho_1 (x \geq c_j) + \rho_2 (x - c_j)(x \geq c_j) + \varepsilon \tag{7}$$

3.4 Hypotheses

H1: *In a comparison of procurement calls (which are investigated via CGU's service orders), an increase in the tender amount is associated with an increase in all three corruption indicators.*

H2: *In a comparison of procurement calls, an increase in the tender amount is associated with decreasing returns to corruption as measured by the amount potentially lost to corruption*

(corruption indicator II) and the share of corruption findings to overall wrongdoing (corruption indicator III).

H3: *In a comparison of procurement calls (which are investigated via CGU's service orders), an increase in the municipal corruption level, measured by the share of corruption findings to the overall level of wrongdoing in any municipality (corruption indicator III aggregated by municipality), is associated with a higher probability that any individual tender will contain at least one corruption finding (corruption indicator I).*

H4: *In a comparison of procurement calls, an increase in the municipal corruption level, measured by the share of corruption findings to the overall level of wrongdoing in any municipality (corruption indicator III aggregated by municipality), is associated with decreasing returns to corruption as measure by the amount potentially lost to corruption (corruption indicator II).*

H5: *In a comparison of procurement calls (which are investigated via CGU's service orders), stricter procurement rules reduce corruption as measured by all three corruption indicators,* all else equal.

H6: *In a comparison of procurement calls, stricter procurement rules reduce corruption at a decreasing rate as measured by all three corruption indicators,* all else equal.

H7: *In a comparison of procurement calls (which are investigated via CGU's service orders), stricter procurement rules reduce mismanagement as measured by all three mismanagement indicators,* all else equal.

H8: *In a comparison of procurement calls, stricter procurement rules reduce mismanagement at a decreasing rate as measured by all three mismanagement indicators,* all else equal.

4 Results

4.1 Falsification Tests

5 Conclusion

5.1 Cost-benefit Analysis

References

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

Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy*, 76:169–217.

Becker, G. S. and Stigler, G. J. (1974). Law Enforcement, Malfeasance, and Compensation of Enforcers. *The Journal of Legal Studies*, 3(1):1–18.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 82(6):2295–2326.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2015). Optimal Data-Driven Regression Discontinuity Plots. *Journal of the American Statistical Association*, 110:1753–1769.

Cattaneo, M. D., Keele, L., Titiunik, R., and Vazquez-Bare, G. (2016). Interpreting Regression Discontinuity Designs with Multiple Cutoffs. *The Journal of Politics*, 78(4):1229–1248.

Cattaneo, M. D., Titiunik, R., and Vazquez-Bare, G. (2018). Analysis of Regression Discontinuity Designs with Multiple Cutoffs or Multiple Scores. *Working Paper*.

Chong, A., De La O, A., Karlan, D., and Wantchekon, L. (2015). Does Corruption Information Inspire the Fight or Quash the Hope? A Field Experiment in Mexico on Voter Turnout, Choice, and Party Identification. *The Journal of Politics*, 77:55–71.

Choudhury, S. (2015). Governmental Decentralization and Corruption Revisited: Accounting for Potential Endogeneity. *Economics Letters*, 136:218–222.

Fan, C. S., Lin, C., and Treisman, D. (2009). Political Decentralization and Corruption: Evidence from Around the World. *Journal of Public Economics*, 93(1):14–34.

Ferraz, C. and Finan, F. (2008). Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes. *The Quarterly Journal of Economics*, 123:703–745.

Ferraz, C. and Finan, F. (2011). Electoral Accountability and Corruption: Evidence from the Audits of Local Governments. *American Economic Review*, 101:1274–1311.

Ferraz, C., Finan, F., and Moreira, D. B. (2012). Corrupting Learning: Evidence from Missing Federal Education Funds in Brazil. *Journal of Public Economics*, 96(9):712–726.

Finan, F., Ferraz, C., and Avis, E. (2018). Do Government Audits Reduce Corruption? Estimating the Impacts of Exposing Corrupt Politicians. *Journal of Political Economy*.

Fisman, R. and Gatti, R. (2002). Decentralization and Corruption: Evidence Across Countries. *Journal of Public Economics*, 83(3):325–345.

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.

Lambsdorff, J. G. (2002). Corruption and Rent-Seeking. *Public Choice*, 113:97–125.

Lichand, G., Lopes, M. F. M., and Medeiros, M. M. (2017). Is Corruption Good for Your Health? *Working Paper*.

Lipsky, M. (2010). *Street-Level Bureaucracy, 30th Anniversary Edition: Dilemmas of the Individual in Public Service*. Russell Sage Foundation.

Litschig, S. and Zamboni, Y. (2015). Judicial Presence and Rent Extraction. *Barcelona GSE Working Paper Series Working Paper 796*.

McCrary, J. (2008). Manipulation of the Running Variable in the Regression Discontinuity Design:

A Density Test. *Journal of Econometrics*, 142(2):698–714.

Olken, B. A. (2006). Corruption and the Costs of Redistribution: Micro Evidence from Indonesia. *Journal of Public Economics*, 90:853–870.

Olken, B. A. and Pande, R. (2012). Corruption in Developing Countries. *Annual Review of Economics*, 4:479–509.

Rose-Ackerman, S. (1975). The Economics of Corruption. *Journal of Public Economics*, 4:187–203.

Stigler, G. J. (1970). The Optimum Enforcement of Laws. *Journal of Political Economy*, 78:526–536.

Treisman, D. (2000). The Causes of Corruption: A Cross-national Study. *Journal of Public Economics*, 76(3):399–457.

Winters, M. S. and Weitz-Shapiro, R. (2013). Lacking Information or Condoning Corruption: When Do Voters Support Corrupt Politicians? *Comparative Politics*, 45:418–436.

Zamboni, Y. and Litschig, S. (2018). Audit Risk and Rent Extraction: Evidence from a Randomized Evaluation in Brazil. *Journal of Development Economics*, 134:133–149.

A Appendix: Service Order Classification

Service orders issued by CGU investigated various uses of public resources in addition to procurement, e.g. for staff compensation, for school activities, or for community monitoring of public policies. The discretion measure proposed in this paper, however, is only applicable to uses of public resources that required the procurement of goods and services under Law 8,666/93. Therefore, the ideal dataset for this study would contain, for each federal transfer, the attached procurement call informing dates, goods/services procured, global contract value, who was awarded the contract, and so on. Unfortunately, there is no organized, public database of municipal contracting processes. In CGU's reports, the reporting of procurement calls is implicit, via descriptions of investigations or violations to Law 8,666/93. Thus, we isolate service orders for which there was any procurement of goods/services involved from the rest by implementing a classification system based on the information retrieval literature.

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 drug purchases, whether the municipality has used the funds within designated program 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 procurement-related keywords to form patterns to be searched for in the description of each investigation conducted by CGU. There are two broad categories 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 group.

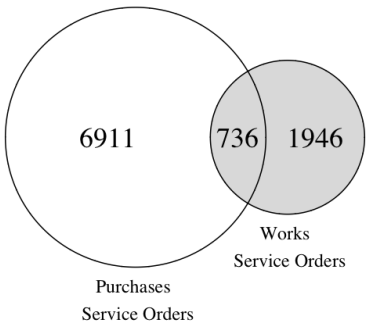
An example is useful for understanding our classification process. Unigram "aquisição" (*acquisition* in English) is stemmed to "aquisi" to form a search pattern for the *purchases*-type procurement; unigrams "adequação" and "habitacional" (roughly translated as *housing adequacy*) are stemmed and combined to form "adequa(.)*habitac" search pattern for *works*-type procurement. The use of bigrams and regular expressions, such as in the pattern above, 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 (in regular expression format)

Type	Search Terms
Purchases	"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"
Works	"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"

The final list contains 19 n-grams for identification of purchases and 17 n-grams for works.[4] 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 public 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. In other words, 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

[4] 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.

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 $n + 1$ vs. n 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) runs such check and the results are reported in the table below:

Table 2: Purchases Search Results

	Total	Average				Means Test
	Finds	Find	Length	Position	TF-IDF	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.[5] For example, the p-value in row 3 tells us that including search word "equipame" to the pattern "aquisi" or "execu" identifies a significantly different, and in this case, larger sample at the 1% level.

[5]This is also known as an alternative search where all search conditions are connected by an "or" statement.

Table 3: Works Search Results

	Total	Average				Means Test
	Finds	Find	Length	Position	TF-IDF	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

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.[6]

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. In sections A.2 and A.3, we test if these search patterns really reflect the procurement legislation in Brazil.

[6]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.

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, we can test the quality of our classification algorithm on one type of these transfers.

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 grants in 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.[7]

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

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

it switches to another textual description, the better. The diagonal (*Yes-No; No-Yes*) should be populated by only a small number of transfers compared to the overall sample.

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).[8]

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.[9]

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

[8]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.

[9]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 counterfactual against which to test our classification mechanism.

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

of classification algorithm. The identification of procurement orders is internally consistent (section 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 Coding of Service Orders

Table 6: Infraction Classification

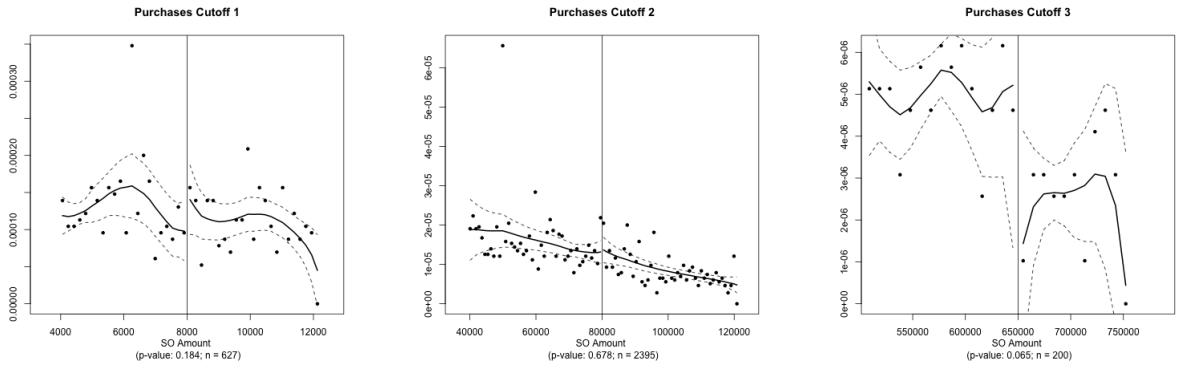
Code #	Code Description
	Bottom-up Monitoring
(01)	Citizen's committee has not been properly set up.
(02)	Committee does not monitor programs.
(03)	Committee has poor working conditions.
	Human Resources
(24)	Officials did not meet their assigned workload.
(27)	Officials received insufficient training.
(28)	Officials were not properly hired.
(32)	Officials received incorrect wage or benefit payment.
	Infrastructure
(20)	Physical infrastructure is inappropriate for program implementation.
(21)	Shortage of government goods/supplies.
(22)	Poor stock management of government goods.
(26)	Government goods/supplies were inadequately labeled.
(29)	Government goods/supplies were poorly preserved.
	Performance
(15)	Payments shifted to other government needs.
(17)	Municipality did not supplement program funding.
(18)	Program has not been entirely implemented or its goals were only partially met.
(19)	Public works have not followed construction rules.
(23)	Poor service provided to citizens.
(25)	Program documentation was wrong.
(33)	Idle funds were not transfered to savings/money market accounts.
(34)	Program participants did not receive their benefits.
(35)	Beneficiaries did not meet conditions for inclusion in program.
(36)	Poor beneficiary data management.
	Procurement
*(04)	Public tender was not publicized.
*(05)	Tender winner presented forged price estimates.
*(06)	Shell companies have participated in tender.
(07)	Tender documentation was wrong.
*(08)	Tender documentation was forged.
*(09)	Tender participant received special treatment.
(10)	Multiple (non-corruption) tender problems.
*(30)	Wrong tender rules were applied.
*(31)	Tender was incorrectly dismissed.
	Private Appropriation
*(11)	Good/service was overpriced.
*(12)	Supplier used forged receipts to claim payments.
*(13)	Payments were unaccompanied by receipts.
*(14)	Payments made to parties unrelated to policy implementation.
	Ungrouped
(00, 98, 99)	No infractions were found.

*Corruption infractions.

C Service Amount Manipulation

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 of the three discretion thresholds, which are 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 performance. We present below the McCrary (2008) test for manipulation of the running variable for all six cutoffs in the Brazilian procurement legislation.

Figure 2: Cutoff Manipulation Tests



The graphs in figure 3 show there is no significant difference between service order density just below and above discretion cutoffs imposed by Law 8,666/93.

Figure 3: Cutoff Manipulation Tests