

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

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

August 11, 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 type (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 categories and still getting by with wrongdoing if they so wish. This interpretation is true even if the indifference towards procurement categories comes from (i) learning how to run a public call for goods and services or (ii) from the deflated thresholds discontinuously changing procurement requirements – besides an update in June 2018, 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 5).

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 identified 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 while section IV discusses the estimation strategy and hypotheses. We present and discuss results in section V. Section VI concludes.

2 Background and Data

2.1 CGU Audit Program

In 2003, the Brazilian federal government, through the Office of the Comptroller-General (CGU), introduced a nationwide program to fight corruption and mismanagement of public resources. The program ran through 2015 with a relatively consistent design, whereby municipalities were drawn at random for expenditure auditing. After the selection via the national public lottery system, CGU teams would visit each municipality equipped with a list of *service orders* which should be the object of their investigation. Typically, this meant inspections on all federal transfers to the

municipality within the preceding three years prior to the lottery.

This program is not new to the political economy literature and has been extensively explored in various research projects. The most widely known study is Ferraz and Finan (2008), where the authors look at whether the disclosure of corruption information has had any effect on electoral outcomes in the 2004 municipal elections in Brazil. Controlling for the number of corruption allegations revealed by auditors, Ferraz and Finan (2008) report a significant decrease of seven percentage points in the likelihood of reelection for incumbent mayors with at least two corruption infractions disclosed in audit reports. In a follow-up study, Ferraz and Finan (2011) reveal that mayors running for reelection are 27 percent less likely to misappropriate federal resources, likely in response to the predicted electoral punishment for corruption if they are detected. Brollo et al. (2013) suggest a *political resource curse*, according to which an exogenous, unexpected resource windfall intensifies the political agency problem and decreases the quality of political candidates in Brazilian local elections.

Most recently, Finan et al. (2018) and Zamboni and Litschig (2018) use changes in the auditing probability in CGU's program across time to evaluate the impact of monitoring on political corruption and the delivery of public services, respectively. In Finan et al. (2018), the authors claim that repeating municipal auditing reduces political corruption by eight percent and increases the likelihood of legal action against government officials by 20 percent. Zamboni and Litschig (2018) show that an increase of 20 percentage points in audit risk reduces corruption in procurement by 10 percentage points but has no effect on health services delivery nor on fraud in the *Bolsa Família* cash transfer program enrollment process.

Similarly to these studies, we also use CGU audit reports to construct corruption and program management indicators. Unlike perception or experience-based measures of performance, corruption in particular, audit outcomes are the result of objective evaluation on the correct use of funds. For every municipal audit, CGU staff produces a comprehensive report on whether federal transfers were in fact employed to their intended goal. Auditors spent, on average, two weeks at each municipality collecting evidence on local expenditures and have no incentive to misreport findings in their reports. CGU officers are tenured civil servants, appointed to their position via approval in a national competitive exam whose performance is independent of the number of infractions they find.¹

Besides forwarding their findings to the relevant authorities (local legislatures, ministries, state and federal prosecutors), CGU also makes reports available online for wider use, which we use to build our performance measures. At the Getúlio Vargas Foundation, we coded 1,139 municipal reports between 2004-2010 covering 14,518 health or education service orders, out of which 9,593 have investigated the use of federal resources for procuring goods, services, or public works in Brazilian municipalities. The coding method is similar to that of Ferraz and Finan (2008), where we assign each report to two research assistants and ask them to read and interpret each finding

¹Ferraz and Finan (2008, 2011); Zamboni and Litschig (2018) provide more evidence confirming the integrity of the auditing process.

within each service order using a guide with codes for 35 possible infractions.² To avoid any individual bias in the interpretation of service orders, a senior research assistant resolved all coding conflicts resulting from the double-blind classification and met with coders every other week to update classification guidelines.

Table 1: Number of Municipal Audits by Ministry

	Health		
	No	Yes	Total
Education	No	0	65
	Yes	34	1040
	Total	34	1105
			1139

Source: CGU and CEPESP-FGV. The total across rows and columns is equivalent to the number of municipalities in the sample.

There are 11 corruption and 24 program management-related infractions (appendix 2).³ We use these infractions to create three measures of government performance: (i) whether the transfer has one corruption or mismanagement violation (*binary* indicator); (ii) the share of corruption or mismanagement infractions over all infractions (*share* indicator); (iii) the amount potentially lost to corruption or wasted as a result of mismanagement (*amount* indicator), which is the product of the share of infractions in (ii) times the total amount of the transfer.

An additional source of bias in the performance indicators proposed in this study would come from the fact that CGU only includes service orders in online reports if there were corruption or mismanagement findings after the auditing took place. Ferraz and Finan (2008) rightfully use only these service orders because they are looking at the effect of the publicly *disclosed* information on election outcomes. If there were no infractions, then there is no corruption information to disclose. However, since we are interested in the effect of discretion on performance that might well go unnoticed to the general public, ignoring clean service orders inflates our performance indicators by restricting the sample to transfers which had problems – instead of the entire universe of transfers investigated at any single municipality. To avoid this form of bias, and in the same way as Ferraz and Finan (2011), we use the entire sample of service orders investigated at every municipality, provided by CGU, including transfers for which no problem was reported. Thus, we have all education and health transfer investigations for all 1,139 municipalities in our sample, even those that do not show up in online reports.

²There are three codes for *clean*, non-infraction findings. The first justification for these codes are cases in which, despite the infraction being resolved with further clarification via analysis or documentation, the auditors still have to document their initial infraction finding. The other justification are cases in which the city administration has not observed policy guidelines because of mistakes made by third-parties. For instance, banks managing program funds are not authorized to charge account maintenance fees, but some banks do (by mistake, clearly) and auditors have to include these payments as nonobservance of program requirements.

³Despite the larger number of corruption infractions, they map onto the same problems included in the indicators in Ferraz and Finan (2008, 2011).

Table 2: Corruption Infraction Codes

Code #	Code Description
Procurement	
(04)	Public tender was not publicized.
(05)	Tender winner presented forged price estimates.
(06)	Shell companies have participated in tender.
(08)	Tender documentation was forged.
(09)	Tender participant received special treatment.
(30)	Wrong category was 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.

Source: CEPESP-FGV. All other codes available in appendix C.

2.2 Procurement Law 8,666/93

Brazil included government contracting provisions in its 1988 Constitution. Article 37, Paragraph XXI, specifies that, for all but a few legal exceptions, all government contracting of goods, services, and public works shall be done via public, open procurement calls. Passed in 1993, Law 8,666 became known as the *procurement law* because it laid out such rules and established them as binding to all government branches, including public companies and agencies, at all levels.

Corruption prevention inspired Law 8,666/93. The federal government wanted to reduce the manipulation of procurement processes favoring political or business figureheads by increasing contracting transparency and predictability. Corruption detection became easier as authorities responsible for tracking and prosecuting wrongdoing were given a consistent set of rules against which to check contracting calls. In addition, the law established criminal punishment for officials involved in procurement fraud, ranging from six months jail time when officials tried to prevent or stop an open call for goods/services to five years in cases where they should have carried out a procurement call but did not.⁴ Therefore, both the likelihood of detection and the severity of punishment for corruption crimes changed when Law 8,666/93 was adopted.

There are four procurement categories relevant to this study in the law:⁵ (0) direct contracting (1) invitational bidding; (2) price comparison bidding; (3) competitive bidding. These categories are decreasing in the discretion officials enjoy to spend public resources. In category zero, the government and its agencies are exempt from opening a procurement call and can purchase goods and services directly from suppliers. In category one, officials invite at least three suppliers to partici-

⁴The higher six-year punishment mentioned in Zamboni and Litschig (2018) is applicable to contractors rather than civil servants.

⁵The is another category for government awards, generally in arts and sciences, which is not applicable to our object of study.

pate in the procurement call and, if any of three decides not to participate, the administration has to resend the invite out until it finds three businesses willing to participate in the call. For category two, only companies who had pre-registered with the government as suppliers can participate in the call, which should remain open for at least 15 working days. Participation in procurement calls under category three are open to any domestic or international supplier and the calls should remain open for a minimum of 30 days before the deadline for the submission of bids.

In fact, the application of rules is a function of procurement object and contract value: there are two types of objects (any good or service, which we call *purchases*, and goods or services used in public works projects, which we call *works*) and three increasing monetary thresholds for each type. Table 3 summarizes these categories. The discretion categories are constant across columns (i.e. the same regardless of type) but decrease discontinuously at each monetary cutoff in Law 8,666/93.

Table 3: Procurement Categories

	Type	Goods/Services Purchases		Public Works	
		<i>Lower Bound</i>	<i>Higher Bound</i>	<i>Lower Bound</i>	<i>Higher Bound</i>
Category 0	Direct contracting	-	R\$ 8,000	-	R\$15,000
Category 1	Invitational	R\$ 8,000.01	R\$ 80,000	R\$15,000.01	R\$15,000
Category 2	Price Comparison	R\$ 80,000.01	R\$ 650,000	R\$150,000.01	R\$1,500,000
Category 3	Competitive	R\$ 650,000.01	-	R\$1,500,000.01	-

Source: Law 8,666/93. These values were in place between 1998 and 2018, when they have been adjusted for inflation over the same time period. Our sample covers federal transfers between 2001 and 2010, therefore it is unaffected by this change.

Besides comparing discretion by interpreting procurement requirements at each category, three other factors substantiate our claim for decreasing discretion. First, agents who are found guilty of procurement fraud have to pay a fine ranging from 2% to 5% of total contract value. This provision results in higher penalties for corruption at higher categories (albeit at a decreasing rate). Second, CGU auditors specifically report instances where a single larger procurement call was split into smaller value, fewer goods and services calls to avoid higher procurement categories (code #30 in table 2). 5.5% of the transfers in our sample have been reported as containing evidence that the wrong procurement category has been applied. Finally, Article 23, §4 of Law 8,666/93 allows for the adoption of rules from higher categories in procurement calls of lower amounts (but not the other way around). For instance, a government call for the purchase of R\$65,000 worth of textbooks should use *invitational bidding* rules (category two) but officials could apply *price comparison bidding* (category three) if they so wish; the only possible explanation is that higher categories impose fairer, more transparent, harder ways to contract with private businesses. In sections 2.3 and 4.1, we also use Art. 23, §4 to limit problems from not directly observing procurement contracts.

2.3 Data

In order to operationalize this research project, we gather data from various public sources maintained by government agencies. There are two data aggregation levels for our sample units: (i) at the service order level, which is the investigation conducted by CGU on a single type of federal transfer to municipal governments; (ii) at the municipality level, where we find the variables to control for municipal characteristics that could also be correlated with municipal corruption and poor program management.

The outcomes are measured at the service order level. We bring together data from (i) online audit reports, posted on CGU’s website and from (ii) CGU’s record of all services orders investigated at the same municipalities for which we have coded online reports. The reasoning here is simple: since the online reports only display federal transfers where there is evidence of corruption or mismanagement, we also need to include transfers for which there were no infractions, otherwise our sample would be biased towards problem-ridden transfers. CGU has kindly provided us the spreadsheet of all service orders between audit lotteries 8 and 33, which we match to a random sample of 1,139 municipal audits from lotteries 8 and 31.⁶ As described in section 2.1, there are 14,518 health and education investigations in our sample, or about 12 per municipality, and we use 9,593 investigations which can be linked to having looked into procurement processes. There are three performance variables for either outcome group (corruption or mismanagement): (i) binary, (ii) share, and (iii) amount measures. For example, if auditors found one evidence of corruption out of two infractions in a given federal transfer worth R\$10,000, the three outcome values are, respectively, 1, 0.5, and R\$5,000, where (iii) is simply the product of (ii) and the transfer amount.

The main independent variable is procurement category, which summarizes discretion in public spending. For each federal transfer, we construct a three-level categorical variable reflecting the categories in Law 8,666/93 based on the the procurement type and amount. The baseline level is direct contracting and other categories follow the order and types in table 3. Unfortunately, however, we do not directly observe the procurement contracts, i.e. we do not have access to individual municipal contracting calls. There is no organized, central database of all procurement processes across 5,500+ Brazilian municipalities. Many of these processes only exist in hard copies and had happened two or three years before the first CGU audits were conducted in 2004, at a time where monitoring and record keeping were still incipient across Brazil. Unless scholars have designed an experiment in conjunction with CGU, such as Zamboni and Litschig (2018), in which auditors specifically collected individual data from procurement contracts (and yet for a narrow timespan), this is an unrealistic sample to obtain. The measurement error in our discretion variable, however, is limited. Since Art. 23, §4 determines that no procurement call can occur following rules applicable to calls of lower amount, we know with certainty that no single federal transfer in our sample has had a procurement contract with more discretionary rules (lower-level categories); there is only upward measurement error bias and, as such, using directly observed procurement

⁶Our coding system was designed to randomly pull audit reports from CGU’s database so as to remain a good representation of the CGU investigation universe.

calls would only monotonically increase the magnitude of the effect of discretion on performance⁷ (we discuss this at length in section 4.1).

Since the procurement category is also conditional on the object being procured, we also need to split each service order into two broad groups: (i) *public purchases*, which involve the procurement of any good or service, and (ii) *public works*, which only include the purchase of goods and services used in public infrastructure projects. Not directly observing procurement contracts also requires an inference method to code transfers into either type. However, we can consistently identify each group and run robustness checks (documented in appendix A) using text analysis algorithms developed in Grimmer and Stewart (2013) and Assumpcao (2018). These algorithms search for user-defined keywords in textual descriptions of each federal transfer and return whether they are a match to the purchases, works, or both groups. Using measures of match quality similar to type I and II errors, we find potential false positive rates at 5.3% and 19.7% for purchases and works and false negatives at 8.5% and 9.9% respectively for each group.

There are two other important variables at the service order level, the monetary amount investigated and the overall municipal corruption *for all other federal transfers*, which we use to control for additional transfer and municipal characteristics. They also come from CGU’s audit reports and inspection record. The monetary amount in each service order is reported in Brazilian Reais (R\$) at nominal levels. Since these values determined the procurement category and the decision to incur in corruption then, there is no reason to update them to real levels. The overall municipal corruption is constructed using outcome (ii), the share of corruption infractions, aggregating all municipal service orders corruption share except for the one in question. It captures the overall corruption relative to the single transfer at stake. If municipal corruption is high, it also creates the incentive for corruption at that single transfer.

We use the municipal level variables as covariates to control for observable municipal characteristics. Data come from the Brazilian Office of Statistics (IBGE) and the Electoral Court (TSE). All socioeconomic data comes from the 2010 IBGE Brazil Census and the electoral data come from the election immediately before the municipality received the audit team (either 2000, 2004, or 2008). From IBGE, we pull data on the share of urban population, the female share, the illiteracy rate, income inequality as measured by the GINI index, the human development index, and poverty rate. These are the literature-standards for municipal controls. In addition, we include variables used from similar studies in which CGU audits serve as the basis for corruption outcomes, such as presence of the AM radio in municipality (binary), the establishment of health and education bottom-up monitoring councils (binary), whether the municipality is seat of a local office of the Judiciary Branch (binary), and the vote margin (as share of valid votes) and mayor reelection rates (binary). Table 4 displays the descriptive statistics for the first group of variables at the service

⁷Except, of course, the transfers for which auditors reported that inapplicable rules were used, which make up 5.5% of the sample as per section 2.2. We cannot tell with certainty, however, if these transfers have been split into amounts low enough to change procurement categories or have been split into amounts that still kept them in the same category. Since our results do not change when we exclude these transfers, it is highly unlikely that we are experiencing the former problem.

order level in Panel A and the variables at the municipal level in Panel B.

Table 4: Descriptive Statistics

<i>Panel A: Service Order Level</i>							
	<i>N</i>	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Amount (in R\$)	9,593	449,858	3,060,374	65	36,000	204,721	236,198,658
Infraction Count	9,593	2.398	2.172	0	1	3	18
Corruption Indicator I (Binary)	9,593	0.398	0.489	0	0	1	1
Corruption Indicator II (Share of Total Infractions)	9,593	0.195	0.294	0	0	0.3	1
Corruption Indicator III (Amount)	9,593	125,695	954,252	0	0	29,427	49,282,832
Mismanagement Indicator I (Binary)	9,593	0.746	0.435	0	0	1	1
Mismanagement Indicator II (Share of Total Infractions)	9,593	0.619	0.407	0	0	1	1
Mismanagement Indicator III (Amount)	9,593	268,168	2,618,568	0	0	122,000	236,198,658
<i>Panel B: Municipal Level</i>							
	<i>N</i>	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Urban Population (Share)	1,139	0.642	0.221	0.042	0.476	0.826	1
Female (Share)	1,139	0.505	0.015	0.461	0.495	0.512	0.658
Illiteracy Rate	1,139	0.168	0.099	0.016	0.083	0.254	0.428
GDP per capita	1,139	11,890	11,696	2,463	5,046	14,749	153,770
Gini Index	1,139	0.512	0.066	0.318	0.469	0.555	0.783
Human Development Index	1,139	0.654	0.072	0.469	0.592	0.714	0.862
Poverty Rate	1,139	0.250	0.184	0.003	0.078	0.404	0.755
Presence of AM Radio	1,139	0.237	0.425	0	0	0	1
Education Council Established	1,139	0.781	0.413	0	1	1	1
Health Council Established	1,139	0.969	0.173	0	1	1	1
Seat of Judiciary Branch	1,139	0.514	0.500	0	0	1	1
Vote Margin	1,139	0.168	0.188	-0.046	0.047	0.211	1
Mayor Reelection Rate	1,139	0.293	0.451	0	0	1	1

Sources: CGU, CEPESP-FGV, IBGE, and TSE. Panel A contains variables measured at the service order level coded by CEPESP-FGV straight out of CGU audit reports, including the six corruption and mismanagement outcomes. Panel B contains covariates at the municipal level measured in 2010 by the Brazilian Office of Statistics (IBGE) and electoral data from the Electoral Court (TSE) for municipal elections in 2000, 2004, and 2008. 1.8% of the two election covariates had missing values and were recoded to the overall mean as per Donald Green’s lab Statement of Purpose.

The service level statistics are not directly comparable to that of Ferraz and Finan (2008, 2011) and Zamboni and Litschig (2018) for two reasons: first, we analyze corruption at the servicer order (i.e. federal transfer) level, instead of aggregating them up to the municipal level; second, we are looking at larger universe of 1,139 municipalities versus 373 in Ferraz and Finan (2008), 476 in Ferraz and Finan (2011), and 120 in Zamboni and Litschig (2018). Though the mean transfer

value is roughly R\$450,000, the 75th percentile is close to R\$205,000 (or approximately \$107,000 in 2010 dollars, the last year in our sample). About 40% have at least one corruption infraction and 75% have one or more program management infractions. Seventy-five percent of all transfers have amounts potentially lost to corruption and mismanagement respectively at R\$29,427 and R\$122,000 (or lower).

Panel B, however, displays municipal characteristics consistent with similar studies. The average population is 64% urban, 50% female, and 16.4% are illiterate. GDP per capita is on par with the Brazilian average at R\$11,890 but the index is not equally spread as the average GINI coefficient stands at 0.512. The Human Development Index is low (0.654) relative to Brazil’s overall HDI 0.724, which is a result of CGU’s sampling strategy of excluding larger (higher HDI) municipalities from the random lottery. The average poverty rate is 25%. Twenty-three percent of municipalities have a local ratio station, and over 78.1% have created either the education or health bottom-up monitoring councils, a requirement for participation in some social policy programs. About half of the municipalities in our sample also host a local office for Judiciary branch. Vote margin and the reelection rates are 16.8% and 29.3%. The characteristics in Panel B are used in the estimating equations for robustness checks and for controlling observable municipal characteristics, therefore we do not engage in further the discussion of their relationship with government performance.

3 Theoretical Framework

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:

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

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.⁸

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 2.3. 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 4.2.

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.

After including these factors into equation 1, the corruption function then becomes:⁹

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

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).¹⁰

⁸We additionally assume that officials are risk-averse since the larger the contract the larger the public attention for any given procurement call.

⁹There 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.

¹⁰Article 37 of the Brazilian Constitution, Law 8,666/93, and the related legislation lay out the punishment for

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 categories are less effective at reducing corruption than lower-order categories. These relationships are described below.

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

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 categories 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 category 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 to their actions. In the section 4.1, 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.

$$m = f(w, v, p, d, x, l, r) \qquad (4)$$

$$\nabla f'(-, +, -, -, +, -, +)$$

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 4.1 soaks up any severity effects.

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

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.

4 Estimation Strategy

4.1 Research Design

The ideal experiment to answer the question on the causal effect between discretion and performance would randomly assign rules to an unbiased sample of government procurement calls. In addition, the effect on corruption or program management would be measured *ex post* procurement by an independent, homogeneous group of auditors any bias in outcomes. Municipalities should not anticipate monitoring and, thus, would not try to hide any wrongdoing before the audit team arrives in town. For ethical and budgetary reasons, however, this is an unrealistic experiment – the approach adopted in this paper is its best approximation yet.

We identify the causal effect of discretion on performance by employing a research discontinuity design (RD) and comparing outcomes for federal transfers whose value falls in the vicinities of Law 8,666/93’s monetary thresholds that determine procurement rules. The procurement amount is the assignment variable and the treatment condition is having to procure goods and services under less discretionary, stricter rules. In the vicinity of each cutoff, transfers are similar in observable and unobservable characteristics, including the variables in models in sections 3.1 and 3.2, except for discretion in public spending.

As discussed previously, a main challenge in this project is strictly matching each transfer to procurement categories and types as summarized in Law 8,666/93. Therefore, we proceed with caution and instead of drawing conclusions about the local average treatment effect (LATE), typical of the RD design, we focus on the (local) intention to treat effect (ITT). An unique feature of Law 8,666/93, however, makes this a special type of ITT. Since the legislation allows the adoption of stricter procurement rules even if the call is of lower monetary value, but not the other way around, we know with certainty that no procurement process has been adopted with *less strict* rules. The result is that the ITT is actually the minimum local treatment effect or, in other words, the lower bound of the LATE.¹¹ What we are doing here is estimating a *fuzzy* RD as *sharp*. As fuzzy designs

¹¹This is a property of the LATE and the instrumental variables strategy, more generally. The (unobserved) denominator is the actual procurement category (the first stage) but known to be positive – the first term is always one but the second term can be one or zero. The numerator equals the intention to treat effect and is observed, the assumed procured category from the transfer value. Thus, the LATE is bounded below by the ITT. In other words,

are subject to manipulation of treatment category, meaning public officials would try to move down a category to avoid stricter procurement rules, we run the McCrary (2008) tests for each discretion cutoff in Law 8,666/93. They are reported in appendix C, where we see no signs of substantial manipulation preventing the adoption of the RD identification strategy.

$$y_i = \alpha + \gamma_1 x + \gamma_2 x^2 + \delta_1 r + \delta_2 r^2 + \rho_1 l_1 + \rho_2 l_2 + \rho_3 l_3 + \varepsilon \quad (7)$$

$$y_{i,j} = \alpha + \gamma_1(x - c_j) + \gamma_2(x - c_j)^2 + \rho_1(x \geq c_j) + \rho_2(x - c_j)(x \geq c_j) + \varepsilon \quad (8)$$

4.2 Hypotheses

- H1:** *Comparing procurement calls, stricter procurement rules reduce corruption as measured by all three corruption indicators, all else equal.*
- H2:** *Comparing procurement calls, stricter procurement rules reduce corruption at a decreasing rate as measured by all three corruption indicators, all else equal.*
- H3:** *Comparing procurement calls, stricter procurement rules reduce mismanagement as measured by all three mismanagement indicators, all else equal.*
- H4:** *Comparing procurement calls, stricter procurement rules reduce mismanagement at a decreasing rate as measured by all three mismanagement indicators, all else equal.*

5 Results

5.1 Falsification Tests

6 Conclusion

6.1 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.

the Wald estimator reduces to:

$$\rho = \frac{E[y_i | l_i(x) = 1] - E[y_i | l_i(x) = 0]}{E[s_i(x) | l_i(x) = 1] - E[s_i(x) | l_i(x) = 0]} = \frac{E[y_i | l_i(x) = 1] - E[y_i | l_i(x) = 0]}{1 - E[s_i(x) | l_i(x) = 0]} \quad (6)$$

Where y is the outcome of interest, l is the assumed procurement category, and s is the actual procurement category for transfer i . Since we assume the second term to be zero in equation 6, knowing the actual procurement category s can only turn zeros into ones, therefore increasing the magnitude of parameter ρ .

- Becker, G. S. and Stigler, G. J. (1974). Law Enforcement, Malfeasance, and Compensation of Enforcers. *The Journal of Legal Studies*, 3(1):1–18.
- Brollo, F., Nannicini, T., Perotti, R., and Tabellini, G. (2013). The Political Resource Curse. *American Economic Review*, 103:1759–1796.
- 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 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 group.

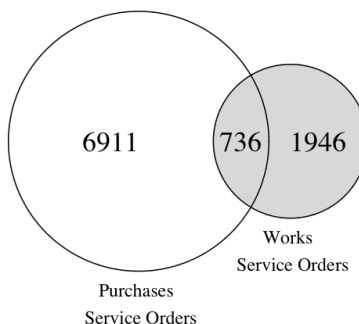
An example is useful for understanding our classification process. Unigram “aquisição” (*acquisition* in English) is stemmed to “aquisi” to form a search pattern for the *purchases*-type procurement; unigrams “adequação” and “habitacional” (roughly translated as *housing adequacy*) are stemmed and combined 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 5: Procurement Search Terms (in regular expression format)

Type	Search Terms
Purchases	“aquisi” “execu” “equipame” “ve[í]culo” “despesa” “aplica[ç]” “medicamento(.)*peaf” “compra” “recurso(.)*financ” “unidade(.)*m[ó]ve(.)*sa[ú]de” “pnate” “transporte(.)*escola” “desenv(.)*ensino” “kit” “siafi” “implementa[ç]” “adquir” “pme(.)*2004” “aparelhamento”
Works	“co(ns sn)tru” “obra” “implant” “infra(.)*estrut” “amplia” “abasteci(.)*d(.)*[á]gua” “reforma” “(melhoria adequa)+(.)*(f[í]sica escolar habitac sanit[á]ria)+” “esgot” “adutora dessaliniz reservat[ó]o” “sanit[á]ri[ao]” “poço” “aperfei[ç]oa” “saneamento” “res[í]duo(.)*s[ó]lido” “conclus[ã]o”

The final list contains 19 n -grams for identification of purchases and 17 n -grams for works.¹² 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 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

¹²One 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 6: Purchases Search Results

	Total Finds	Average				Means Test p-value
		Find	Length	Position	TF-IDF	
“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[fi]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[ó]ve(.)*sa[ú]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.¹³ 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.

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

¹³This is also known as an alternative search where all search conditions are connected by an “or” statement.

Table 7: Works Search Results

	Total Finds	Average				Means Test p-value
		Find	Length	Position	TF-IDF	
“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[i]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[fi]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

individual finds reported in column 1, table 7, 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.¹⁴

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.

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.¹⁵

There is a total of 3,868 service orders for which we have descriptions both from CGU and from the federal government. In table 8, 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 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

¹⁴The search without these terms (available upon request) yields 2,679 service orders, just three short of the total in table 7. Nevertheless, we keep the two items in the search algorithm for additional tests discussed in section A.2.

¹⁵<http://www.portaltransparencia.gov.br/>

Table 8: Classification by Grant Description

<i>Panel A: Purchases Group</i>				
Service Order Description	Grant Description			Total
	No	Yes		
	No	115	144	259
	Yes	83	1473	1556
	Total	198	1617	1815
<i>Panel B: Works Group</i>				
Service Order Description	Grant Description			Total
	No	Yes		
	No	1546	269	1815
	Yes	404	1649	2053
	Total	1950	1918	3868

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).¹⁶

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.¹⁷

¹⁶The 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.

¹⁷The 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

Table 9: Classification by Procurement Code

<i>Panel A: Purchases Group</i>				
Service Order Description	Procurement Code			Total
	No	Yes		
	No	2487	319	2806
	Yes	6137	3456	9593
	Total	8624	3775	12399
<i>Panel B: Works Group</i>				
Service Order Description	Procurement Code			Total
	No	Yes		
	No	2462	344	2806
	Yes	5791	3802	9593
	Total	8253	4146	12399

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 9, 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 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 types in the Brazilian procurement legislation (section A.3).

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.

B CEPESP Coding of Service Orders

Table 10: 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 transferred 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 category was 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, which show no significant difference between service order density just below and above discretion cutoffs imposed by Law 8,666/93.

Figure 2: Cutoff Manipulation Tests

