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

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

September 4, 2018

Abstract

In this study, we measure the causal effect of discretion in public procurement on corruption and program management in Brazilian municipalities. Our results show that discretion does not reduce corruption and only has a limited effect on mismanagement in public works (41.6 percentage point less likely to display management infractions, 29.1 percentage points less in the share of management problems over all problems, R\$4,611 less resources potentially lost to poor management practices). Our results are robust to a variety of specifications and two falsification tests. The welfare effects of discretion are small; procurement rules have only been able to prevent the misplacement of R\$2M out of a R\$32M total.

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

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

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 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 less 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) performance 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, such as how local representatives are elected, the level of political participation in elections, the policy attributions of local governments, etc.

To answer the question whether less 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 and agencies, including public companies, at all levels. 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 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, the second contribution we make is the introduction of three new corruption factors: (i) the discretion in public spending, our most important variable, 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 program management model is parallel to the corruption model and mismanagement is also a function of discretion in public spending, the global contract value, and municipal corruption levels. Management issues do constitute, however, a secondary outcome since the corruption is an issue of higher salience in the Brazilian political environment.

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 used correctly. 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); 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 would 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 would be 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 (41.6 percentage points), the share of mismanagement problems (29.1 percentage points) over all problems, and the amount potentially affected by mismanagement (R\$4,600, a third of contract value) for any federal transfer used for procuring inputs for public works projects. 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). A back-of-the-envelope welfare calculation suggests that Law 8,666/93 prevents 5.98% of overall resource misplacement and 26.7% in 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. Nevertheless, we offer two alternative interpretations for the null effect on corruption. First, procurement rules might not be 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. We show two pieces of evidence in support of these pathways of procurement categories meaninglessness: (i) there is a relative stability of corruption infractions over time, which is telling of the ease with which officials navigate procurement rules, and (ii) the average number of infractions is higher when contract values increase, revealing simpler procurement calls (easier to comply with) at lower categories. In addition, the second interpretation 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 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 of voluntary federal funds transferred to the municipality in the three years preceding the lottery date.

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 (higher) predicted electoral punishment for corruption if they are detected. Brollo et al. (2013) used CGU audit to suggest a *political resource course*, 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 enrollment process for the *Bolsa Família* cash transfer program.

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 evaluations on the correct use of funds. For every municipal audit, CGU staff produces a comprehensive report on whether federal transfers were in fact employed in 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] An example investigation helps picture such audits. Municipality Planalto, SP, randomly drawn at the 28th lottery, received a team of auditors between May 20 and July 13, 2009 and they looked into three education transfers for procuring school buses, lunches, and for distributing school books. On the school buses transfer, they reported paperwork problems of not informing the Ministry of Education back of the receipt of such resources (mismanagement) and procurement calls for the purchase of buses which did not comply with competition requirements (corruption), amongst other things. CGU auditors, however, have no prosecuting power, so they have to forward these allegations to relevant authorities (local legislatures, ministries, state and federal prosecutors) before any action is taken; otherwise, the report is the as far as they go.

These reports are available online for wider use, which we use to build our performance measures. 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 supplies 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 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

| | | Education | | |
		No	Yes	Total
Health	No	0	65	65
	Yes	34	1040	1074
	Total	34	1105	1139

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 at least 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. Multiple outcomes also help disentangle the differential effects on performance. The binary outcomes represent the extensive

---

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

[2]There 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 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 checking account fees, but some banks do (by mistake, clearly) and auditors have to include these payments as noncompliance of program requirements.

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

margin effect, as any single infraction marks a transfer as subject to performance issues, while the share outcome is an intensive margin measure as it translates the extent to which any transfer is filled with one of the performance problems.

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.

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 in which 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 would inflate 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.

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

6

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 participate 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 vendors 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 bid submission deadline.

Within the scope of Law 8,666/93, 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. For example, in municipality Nova Olinda, CE, auditors investigated a transfer for the construction of 72 restrooms worth a total of R$76,792.23, which qualifies this transfer as public works type and invitational bidding category.

Table 3: Procurement Categories

		Goods/Services Purchases		Public Works	
		Lower Bound	*Higher Bound*	*Lower Bound*	*Higher Bound*
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$150,000
Category 2	Price Comparison	R$ 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.

---

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

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

Besides suggesting that discretion is decreasing by looking at the requirements for each procurement category, three other factors substantiate our claim. 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), likely making agents less willing to engage in corruption. Second, CGU auditors specifically report instances where a single larger procurement call was split into smaller value calls to avoid higher procurement categories. In fact, we have a dedicated code to these cases (code #30 in table 2) and they occur in 5.5% of the investigations in our sample. Finally, Article 23 §4 of Law 8,666/93 allows 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 one) but officials could apply *price comparison bidding* (category two) or *competitive 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 investigation 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 transfers containing problems. 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 measures, which turn on when the investigation has found at least one evidence of performance problems; (ii) share measures, which compute the share of corruption or management issues over the total issues for any single federal transfer; and, finally, (iii) amount measures, which multiply

---

[6]Our 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.

the share of either problem by the total amount of the transfer.

Performance measures are complementary. The binary outcomes are representations of the extensive effect of discretion, as any problem flags the federal transfer as having compliance issues. The more transfers with problems, the lower municipal performance is. The intensive margin is captured in the share measures, for which we compute the proportion of problems in each transfer. Using share measures, municipalities can be sorted into mismanagement or corruption-prone groups. The amount outcomes provide a check on the important of performance problems and how much misallocation can be prevented by stricter discretion rules. Suppose auditors investigate five transfers in any given municipality and find corruption evidence in two. Suppose further that, in these two investigations, 100% of their findings is corruption-related and these are the largest transfers received by the municipality in the preceding three years. For this city government, we can then conclude that 40% of all transfers (extensive margin) are subject to intensive corruption (100% of all findings) and that discretionary rules would have had a large effect by preventing the largest misallocations of resources.

The main independent variable is procurement category, which summarizes discretion in public spending. For each federal transfer, we construct a three-level ordinal 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 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+ municipalities. Spreading over the vast Brazilian territory, many of these municipalities only keep hard copies of their procurement contracts, if any, or might have discarded the actual documents following the open calls. 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.

Not observing the treatment variable creates measurement error in our treatment variable. It is possible that each federal transfer we assume as having generated a procurement call in fact has not and, as such, could not really be matched to a procurement category from Law 8,666/93. If this is the case, then we are not identifying any causal effect and thus our design is flawed beyond repair. Fortunately, however, we have very little evidence to support this.

First, we identify transfers resulting in procurement calls by using text classification algorithms discussed in Grimmer and Stewart (2013) and Assumpcao (2018). Since each transfer is accompanied by a textual description of its intended use (for instance, *"education grant to buy textbooks"*), we can turn these descriptions into bag-of-words models in which to search for procurement keywords. If the description meets all criteria in the search algorithm, we can then assign each transfer to either group (*purchases* or *public works*). In appendix A, we extensively test the algorithm and can consistently match transfers to either procurement group with little error. 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.

Second, the very design of Law 8,666/93 supports assuming the transfer value as the procurement value in determining each procurement category. Since Art. 23, §4 determines that no call for vendors can occur following rules applicable to calls of lower amount, i.e. procuring goods and services worth R$20,000 as direct contracting (category zero) instead of invitational bidding (category one), we know with certainty that no single federal transfer in our sample has had a procurement contract with more discretionary rules (lower-level categories). Therefore, there is only upward measurement error bias, if any, as government officials could have procured items following harder (but not softer) rules than should have been applied (we discuss this at length in section 4.1).[7]

There are two other important variables at the service order level, the monetary amount investigated and the overall municipal corruption, which we use to control for additional transfer and municipal characteristics. They also come from CGU's audit reports and inspection records. 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 by aggregating outcome (ii), the share of corruption infractions, for all other federal transfers at any municipality. For a municipality with ten transfers, the level of municipal corruption for transfer one is computed over transfers two through ten, then for transfer two is one plus three through ten, transfer three is one, two, plus four through ten, and so on.

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

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

---

[7]It is possible that officials carried on procuring goods and services at lower categories in plain violation of Law 8,666/93. With that in mind, we created infraction code #30 to specifically pick up on such instances and check whether their frequency would jeopardize our research design. Fortunately, however, only 5.5% of investigations report any such problem, so we are confident in our strategy for identifying the discretion effect.

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).

Table 4: Descriptive Statistics

Panel A: Service Order Level							
	N	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)	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)	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

Panel B: Municipal Level							
	N	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	0.0003	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.

Panel B, however, displays municipal characteristics consistent with similar studies. The average population is 64% urban, 50% female, and 16.4% illiterate. GDP per capita is on par with the Brazilian average at R$11,890 but income is not evenly distributed 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 monitoring councils, a requirement for participation in some social programs. About half of the municipalities in our sample also host a local office for Judiciary branch. Winning vote margin and 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. Besides, the identification strategy we employ addresses observational differences across sample units.

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:

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

The first important change we make to the Olken and Pande (2012) 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 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 detection p and cost 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

---

[8]We additionally assume that officials are risk-averse since the larger the contract the larger the public attention for any given procurement call. That is why we suggest a proportionally smaller cut as bribe when procurement values go up.

in section 2.3. Procurement categories establish the rules each official has to follow and, as a consequence, also determine the opportunity cost of corruption. While we expect that procurement amount will positively influence the level of individual corruption, procurement categories should increase the opportunity cost of corruption and offset, at least partially, the corruption incentive from larger contract values. For instance, an official who is considering corruption at a large procurement call will take into account the difficulty with which she can hide any wrongdoing in addition to how much she can take home as kickbacks.

The last change to Olken and Pande (2012) 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 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. In all these studies, and our own, individuals model their attitude towards corruption, either in accepting or punishing it, as a function of *others' corruption practices*. Though yet little explored in the literature, this is a reasonable hypothesis for an additional factor shaping individual dishonest behavior.

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]

---

[9]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.

[10]Article 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 and we anticipate two positive and one negative effect on corruption. As discussed previously, we expect that procurement amount x and overall municipal corruption r are positively correlated with one's decision to engage in corruption. The more money is available for corruption and the more people are doing it, the more costly will be for anyone to remain honest. On the other hand, the harder it is to take out such illegal payments, the less likely any corruption will take place (discretion variable l). These effects are captured by first derivatives of the b function with respect to each variable, and are illustrated in the first terms of items (i) through (iii) in equation (3).

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

Second derivatives should have the opposite sign for all three variables, translating the different returns to scale for each effect. Procurement amount x and municipal corruption r, though positive with respect to corruption, should increase at decreasing rates to reflect official's risk-aversion to getting caught red-handed. Therefore, as we move up in the joint distributions of procurement amount, municipal corruption, and individual procurement corruption, we should see smaller effects of each determinant on the corruption outcome. For discretion, we expect a negative second derivative, which translates into a lower ability to prevent corruption even if discretion rules are monotonically harder.[11]

Though the analysis in this paper focuses on the effect of (ii) in equation (3), the effects of procurement amount x and discretion l are endogenous and prevent us from using simple OLS models to calculate the partial effects on corruption. The first evidence toward that end comes from Law 8,666/93, wherein contracting rules are a deterministic function of procurement amount. In fact, if we do not control for procurement amounts across the distribution of discretion categories, we would be, at best, computing the total effect of discretion instead of partialling out procurement amounts. In the section 4.1, however, we discuss the identification strategy that allows us to isolate the causal effect 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. Second, it could reduce mismanagement of resources by standardizing the procurement of goods and services, avoiding court challenges, and preventing 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. Hence, although not a primary outcome, better

---

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.

[11]This effect is similar in nature to that of the crime elasticity of policing, according to which there are decreasing returns to policing as measured by crime prevention.

program management qualifies as an important side benefit from discretion-reducing policies if no corruption deterrence takes place.

Fortunately, we can measure program management in a similar fashion to corruption using CGU's audit reports. Not only CGU staff reports corruption infractions in government programs, either in or as a result of government procurement, but they also verify compliance issues. Non-corruption findings are program management issues unrelated to the "use of public office for private gain," including cases in which officials have taken long to procure goods and services, that they have kept unused program funds in the city's bank account, that they have let medication go bad instead of distributing them to the population at health clinics, etc. Therefore, we can construct measures of program *mis*-management which are also determined by the same corruption factors as in section 3.1, and they can be seen in equation (4):

$$
m = f(w, \ v, \ p, \ d, \ x, \ l, \ r)
$$
$$
\nabla f'(-, +, -, -, +, -, +)
$$

(4)

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

Where m is our measure for government mismanagement, w are efficient wages earned by the representative official, v is the alternative market wage they would earn had they chosen to sell their labor for the private sector, d are dishonesty costs, including incompetence, and x, l, and r are the same determinants as in the corruption model, i.e. procurement amount, discretion, and municipal corruption. Since corruption and mismanagement are sister problems, the hypothesized effects in the vector of first derivatives in the second row of equation 4 and equation 5 are the same as in the corruption model.

3.3   Hypotheses and Policy Implications

Based on the models in sections 3.1 and 3.2, we propose two hypotheses summarizing the effect of discretion on government performance. The hypotheses are then used for policy implications, when we present citizens and legislators in a simple principal-agent framework and deduce their positions regarding discretion-reducing policies. Though the discussion is focused on the Brazilian context, the policy takeaways are also valuable for other countries with similar procurement legislation.

Corruption prevention inspired the passing of Law 8,666/93, so the primary hypothesis in this study investigates the relationship between discretion in public spending and corruption outcomes.

> **H1:** *Comparing federal transfers used for the procurement of goods and services, the imposition of harder, stricter rules for public spending reduces corruption as measured by **all** three indicators.*

H1 posits a straightforward negative effect of discretion. If stricter procurement rules are put in place, we should expect the extension (outcome I), the intensity (outcome II), and the relative importance of corruption (outcome III) to be smaller than in the absence of such rules. It is very

strong in the sense that discretion categories should have a negative effect *on all three indicators*; however, since the legislation was designed to prevent corruption, the intended outcome would be, indeed, an unconditional and positive effect on government performance.

If hypothesis one is confirmed, then both citizens and legislators support keeping Law 8,666/93. Principals (citizens) like lower corruption because more public funds are available to be invested in more and better policies. Agents (federal legislators) can altruistically favor lower corruption, in a similar way citizens do, or could favor it because there is an electoral payoff to fighting corruption for which they can claim credit. In either case, they would be supportive of the legislation because there is a net benefit of support even if they personally do not mind higher corruption levels. Therefore, the confirmation of hypothesis one supports the current equilibrium and discretion rules are effective ways to improve government performance.

A less restrictive but still relevant version of hypothesis one is the significant reduction of corruption on *any of the three indicators at any of the discretion categories*. Here, procurement categories would have a moderate effect on corruption because they would independently reduce the extension, intensity, or relative importance of corruption at each federal transfer. If only the less restrictive version of hypothesis one is confirmed, then citizens and legislators positions are slightly different. While legislators can still claim credit for reducing corruption, and they will take it wherever they can find it, citizens might be unhappy with the law if there is only limited corruption prevention. In such cases, citizens will want to see additional, side benefits of passing legislation, e.g. better resource or policy management; this takes us to hypothesis two.

H2: *Comparing federal transfers used for the procurement of goods and services, the imposition of harder, stricter rules for public spending reduces mismanagement as measured by **any** of three mismanagement indicators.*

Summarized in hypothesis two is the side effect of the Brazilian procurement law discussed in section 3.2, wherein program management improves when procurement categories force speedier procurement processes, delivery of goods and services, and compliance with federally-mandated programs. Being a side effect of Law 8,666/93, any improvement in program management is welcome. Once corruption effects have been accounted for, citizens and legislators will turn to mismanagement gains to decide on their position regarding discretion policies. If there is a substantial gain via preventing misplacement of public resources, then both citizens are legislators might be supportive of discretion-reducing policies even if there is a null effect on corruption.

4   Estimation Strategy

4.1   Research Design

The ideal experiment to answer the question on the causal effect of discretion on 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 to avoid 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 determining the applicable procurement categories. 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 (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. Using the textual descriptions of each investigation on any federal transfer, we match each transfer to a procurement type, *purchases* or *public works*, and a category, zero through three, to construct our treatment variables. Since we do not directly observe type and category, we are subject to measurement error on our treatment variables as they might not reflect the actual type and category of each call had we observed every procurement in Brazilian municipalities between 2004-2010. In fact, some procurement calls might have occurred using discretion rules different than what their monetary value would indicate, as discussed in section 2.3. For instance, as permitted by Law 8,666/93, an official might have procured R\$50,000 in goods and services using category three, *price comparison bidding*, when they should have applied *invitational bidding* rules. In addition, procurement with monetary amounts in the vicinity of discretion breakpoints might have been manipulated by public officials to avoid stricter rules. These types of non-compliance of treatment require a *fuzzy* RD design.

In the fuzzy version of RD, we are effectively using instrumental variables (IV) for treatment status at the discontinuities of discretion (Angrist and Pischke, 2008). An unique feature of Law 8,666/93, however, makes this a special type of local average treatment effect (LATE) and allows us to use the simpler *sharp* RD design. Since the legislation allows the adoption of stricter (upper categories) 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. Since the manipulation is unidirectional, we can assume sharp RD by artificially deflating the intention to treat (ITT) denominator and producing lower-bounded LATE parameters.[12] Though not the

---

[12]Generally speaking, this is a property of the LATE and the instrumental variables strategy. 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, the Wald estimator reduces to:

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

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 of the denominator to be zero in equation (6), knowing the actual

exact magnitude of the effect of discretion on performance, any measurement error improvement should only increase the magnitude of our parameters, but not change their direction.

Despite simplifying the estimation as sharp RD, knowing that this is, in fact, a fuzzy design requires additional care with manipulation around discretion breakpoints. If officials have large control over procurement amount, then expenditures just below and above any discretion category cutoff are not as good as randomly assigned and cannot be compared to produce causal inference of discretion. To address this, we deploy the McCrary (2008) tests for manipulation of the treatment assignment variable, where we investigate whether there are significant breaks in the density of sample units on both sides of a cutoff. They are reported in appendix C, where we see no signs of significant manipulation for five of the six cutoffs, thus validating the identification strategy adopted here. We drop the analysis of the discretion change at the breakpoint where there is substantial manipulation, i.e. the purchases sample when moving from price comparison to competitive bidding (categories two and three), as the observations below and above such breakpoint are not good counterfactuals to each other.

4.2 Estimating Equations

Using the performance outcomes specific to Brazilian municipalities, summarized in functions b and m in section 3, we first estimate OLS regressions to confirm the correlations between the outcomes and their performance factors. Here, we are interested in verifying whether the proposed determinants are in fact correlated with the outcomes before making any causal claim about their relationship.

$$y_{c,m} = \alpha + \gamma_1 x + \gamma_2 x^2 + \rho_1 l_1 + \rho_2 l_2 + \rho_3 l_3 + \delta_1 r + \delta_2 r^2 + \sum_{j=1} \beta_j z_j + \sum_{k=1} \lambda_k + \varepsilon \qquad (7)$$

Where y is one of the three c corruption or m mismanagement outcomes; x is the amount procured in each transfer; l_1, l_2, l_3 are the binary indicators for whether the transfer follows rules of procurement categories one, two, or three, respectively; and r is the municipal corruption level measured by the aggregation of corruption index II (the share of corruption violations over all violations) for all other transfers. These variables are all measured at the service order (transfer) level. We control for municipal factors by including each z_j vector of municipal characteristics in panel B of table 4. The log of GDP per capita, the Gini inequality index, the Human Development Index, and the poverty rate are proxies for the income variables w and v. Though not perfect, they control for income effects that could influence government performance by shaping the incentives officials have to act dishonestly or shirk. The probability of detection p is summarized by the illiteracy rate, and the binary indicators for (i) the presence of a local AM radio station, the establishment of (ii) education and (iii) health municipal monitoring councils, and (iv) whether the municipality hosts a local office of the judiciary branch. Jointly, they map the extent to which the population is informed and absorbs government performance information. Finally, dishonesty cost

procurement category s can only turn zeros into ones, therefore increasing the magnitude of parameter ρ.

d is modeled as a function of the share of urban and female populations, the vote margin for the elected mayor and mayor reelection rates, which all shape the delivery of public policies in Brazilian municipalities. Being dishonest imposes a cost for public services and these variables moderate how service provision spreads over the general population. Other additional (unobserved) factors are captured by constant α and lottery and ministry λ_k fixed-effects.

RD equation (8) hones in on the causal effect of discretion on government performance.

$$y_{c,m} = \alpha + \gamma_1(x - c) + \gamma_2(x - c)^2 + \rho_1(x \geq c) + \rho_2(x - c)(x \geq c) + \delta r + \sum_{j=1} \beta_j z_j + \varepsilon \quad (8)$$

Where, again, y is one of the three c corruption or m mismanagement outcomes; $(x - c)$ is the procurement amount centered at each cutoff (and its second-order polynomial); $(x \geq c)$ is the treatment indicator variable, i.e. whether the transfer has involved the procurement of goods/services under stricter rules; the interaction $(x - c)(x \geq c)$ summarizes differential trends on both sides of the procurement category threshold; r is the municipal corruption level, measured the same way as in equation (7); finally, we estimate RD versions with and without the municipal characteristics matrix $\sum \beta_j z_j$. We use the bandwidth selection model from Calonico et al. (2015) for the estimation of the lower bounded LATE in equation (8). Here, the causal effect on each outcome is given by ρ_1 for each change in discretion category.

4.3 Multiple and Cumulative Discontinuities

Law 8,666/93 has two additional features that require the novel design extensively discussed in Cattaneo et al. (2016). First, though the treatment conditions are the same across procurement types *purchases* and *works*, they are assigned to each call following different monetary schedules. For instance, the first change in procurement rules from direct contracting to invitational bidding occurs at R\$ 8,000 for purchases and R\$ 15,000 for works. Treatment cutoffs vary across units in a situation much like the electoral example in Cattaneo et al. (2016), for which the assignment variable is winning an election, measured by vote share, in an electoral constituency under plurality rules. The impact of one percentage point vote share in a multi-candidate election is not the same as the one percentage point in a two-candidate race, and constitutes what Cattaneo et al. (2016) calls "hidden" heterogeneity. In fact, there might be unobservable differences in municipal procurement for public works or for general goods and services which are uncovered if we just pooled all discretion changes across both samples. Though our ultimate goal is finding an overarching effect regardless of procurement type, just looking at pooled results is an incomplete analysis of the effect of discretion on performance.

In addition, the discretion categories are cumulative within each procurement type. There are different versions of treatment for different ranges of procurement amount, which are translated by the multiple discretion categories. Thus, we cannot rule out that assignment scores (i.e. the monetary amount in each procurement call) could be related to potential outcomes. What is reassuring, however, is that the solutions to multi-cutoff RD limitations proposed by Cattaneo

et al. (2016) can all be implemented here: we estimate treatment effects both for pooled and independent samples (multiple non-cumulative cutoffs) and within (multiple cumulative cutoffs) procurement type. Since the results remain unchanged in each specification, we are confident that we have minimized the effect of unobservable differences around discretion breakpoints.

5 Results

In section 3, we proposed that government performance in Brazil, measured by corruption and mismanagement indicators, is a function of procurement rules, the monetary amount of procurement calls and the municipal level of corruption. Thus, the first test in table 5 verifies whether this formulation accurately represents local government performance in Brazil.

Table 5: Performance Determinants in Brazilian Municipalities

	Dependent variable:							
	Corruption Indicator I (Binary)				Mismanagement Indicator I (Binary)			
	Purchases		Works		Purchases		Works	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Amount (in R$)	0.000***	0.000***	0.000	0.000	-0.000	-0.000	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Amount (in R$)	-0.000	-0.000*	-0.000	-0.000	0.000	0.000	-0.000***	-0.000***
(Squared)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Municipal Corruption	1.015***	0.766***	1.303***	0.974***	0.441***	0.440***	0.479***	0.428**
	(0.112)	(0.116)	(0.179)	(0.180)	(0.112)	(0.116)	(0.179)	(0.180)
Municipal Corruption	-1.544***	-1.388***	-1.834***	-1.474***	-2.053***	-2.060***	-1.694***	-1.542***
(Squared)	(0.214)	(0.212)	(0.333)	(0.312)	(0.214)	(0.212)	(0.333)	(0.312)
Proc. Category 1	0.072***	0.101***	0.185***	0.177***	-0.034	-0.038*	0.266***	0.254***
	(0.022)	(0.022)	(0.036)	(0.034)	(0.022)	(0.022)	(0.036)	(0.034)
Proc. Category 2	0.122***	0.175***	0.322***	0.303***	-0.034	-0.035	0.449***	0.436***
	(0.023)	(0.024)	(0.037)	(0.036)	(0.023)	(0.024)	(0.037)	(0.036)
Proc. Category 3	0.172***	0.202***	0.099	0.102	0.042	0.032	0.325***	0.343***
	(0.035)	(0.034)	(0.091)	(0.089)	(0.035)	(0.034)	(0.091)	(0.089)
Municipal Controls	-	Yes	-	Yes	-	Yes	-	Yes
Ministry Fixed-Effects	-	Yes	-	Yes	-	Yes	-	Yes
Lottery Fixed-Effects	-	Yes	-	Yes	-	Yes	-	Yes
Observations	6,911	6,911	2,682	2,682	6,911	6,911	2,682	2,682
Adjusted R^2	0.032	0.076	0.053	0.105	0.048	0.067	0.087	0.175
F Statistic	33.331***	13.642***	22.417***	7.968***	51.071***	12.084***	37.633***	13.669***

Notes: This table reports the relationship between the corruption and management determinants in the context of Brazilian municipalities. Corruption Indicator I (columns 1-4) and Management Indicator I (columns 5-8) are binary variables turning on when the procurement call has at least one corruption or program management infraction. Municipal corruption is computed by aggregating corruption indicator II at the municipal level; it summarizes the relationship between the corruption level at each procurement call vs. overall municipal performance. Municipal controls are summarized in table 4 and control for other variables that are associated with individual corruption decisions. Robust standard errors are clustered at the municipal level and are reported in parentheses. *p<.1; **p<.05; ***p<.01

The main outcomes in table 5 are corruption and mismanagement indicator I. They report the likelihood that any single federal transfer, investigated by individual CGU service orders, will have at least one evidence of corruption (columns 1-4) or mismanagement (columns 5-8) of public resources. In this first table, we run regressions broken down by procurement type (purchases and works) and alternate the inclusion of municipal covariates, ministry, and CGU audit lottery fixed-effects (as described in section 4) for robustness checks.

Contrary to expected, the effect of procurement amount is not consistent nor large (in cases where statistically significant). Across the first row, we notice that procurement call amount only has a positive and statistically significant (p-value $< .01$) correlation with corruption in public purchases and mismanagement of public works resources. The second-order effect, measuring the decreasing rate at which call amount correlates with, has the expected sign (negative) but is again very small and inconsistent across different regression specifications.

Although they do not describe causal relationships, the parameters in rows one and two in table 5 also suggest that procurement amount is not an important factor in official's decision whether to pocket appropriate resources or use them inefficiently. Indeed, the very fact that amounts determining procurement rules had not been adjusted between 1998 and the time for which we have transfer records (2004-2010) lends support to this finding (we also discuss this more comprehensively when presenting the results for the RD estimations).

The municipal corruption parameters confirm it as an important determinant of government performance (as expected). An increase of 1 percentage point in municipal corruption is associated with a increase of 0.766 and 0.974 percentage points,[13] respectively, in the likelihood of finding corruption in federal transfers used for procuring goods/services and public works procurement processes (parameters in columns 2 and 4, where we additionally control for municipal, ministry and audit effects). The effect is smaller for program management issues at 0.440 and .428 percentage points for each procurement type. Except for model (8), which is statistically significant at 5%, all other municipal corruption effects are significant at the 1% level. In addition to confirming the sign of first derivatives of performance with respect to overall municipal corruption, we also confirm the direction of second derivatives; across all models, municipal corruption effects increase at decreasing rates, with magnitudes ranging from 2 percentage points (column 6) to 1.4 percentage points (column 1).

The municipal corruption effect here has a similar nature to the effect documented by Chong et al. (2015), in which voters adjust their corruption punishment in response to whether the challenger candidate is herself dishonest, or Finan et al. (2018), who document lower municipal corruption when mayors learn of audits in neighboring municipalities. These studies, along with ours, are suggesting that corruption outcomes are also conditional on whether peers, or neighbors, also engage in corruption.

Mostly important, we are interested in the relationship between procurement categories and

---

[13]The municipal corruption variable is scaled from 0 to 1 in .01 increments, hence a change of one unit represents a 100 percentage point change in municipal corruption. That is why we need to divide the parameter magnitudes by 100 for the percentage point effect on performance indicators.

performance outcomes. Except for columns (5) and (6), procurement category one is positively and statistically correlated with corruption and mismanagement. Switching over to a set of less discretionary rules is associated with increasing corruption and mismanagement; the effect generally persists over in categories two and three (except for category three effect on corruption for the works subsample in the last row, columns 3 and 4). Though the effect is contrary to the proposition in section 3, what we are capturing here is the most general relationship between these categories and the performance outcomes. Since these categories are directly linked to procurement amounts, we cannot uncover partial effects of discretion without additional treatment for collinearity – which is the contribution of the RD strategy for the causal mechanism suggested here.

Besides testing the relationship between the Brazil-specific performance determinants, a final check before moving over to the RD estimates is whether officials are not manipulating procurement amounts to avoid less discretionary spending rules. If this is the case, then the assignment to treatment is not random across cutoffs and thus prevents causal inference using the local average treatment effect framework. The manipulation tests in Appendix C lead us to drop cutoff 3 for the purchases subsample from the analysis since its densities are statistically different across the amount cutoff (p-value $< .10$). Henceforth, we only analyze five of the six cutoffs in Law 8,666/93 for the RD regression equations.

Since our analysis spans over different procurement units (purchases and works), multiple cutoffs (three) and outcomes (three each for corruption and mismanagement), we calculate a total of 36 data-driven, optimal bandwidth sizes ($2 \times 3 \times 6 = 36$) for each unique combination type-category-outcome using the strategy developed by Calonico et al. (2015). The bandwidths are reported in table 6.

Table 6: CCT Bandwidths (in R$)

		Corruption Outcomes			Management Outcomes		
		Indicator I	Indicator II	Indicator III	Indicator I	Indicator II	Indicator III
	Cutoff 1	26,564	26,327	22,671	27,393	26,554	22,364
Purchases	Cutoff 2	45,854	35,566	34,547	43,431	39,460	38,455
	Cutoff 3	323,425	281,940	289,857	205,542	213,622	226,386
	Cutoff 1	29,960	33,076	28,276	43,876	44,177	38,117
Works	Cutoff 2	74,779	65,067	67,447	60,251	55,164	56,915
	Cutoff 3	1,001,774	862,717	825,342	340,947	739,532	956,064
	Cutoff 1	23,987	24,116	18,606	18,852	19,546	17,728
Pooled	Cutoff 2	51,276	39,937	39,263	33,164	32,725	60,868
	Cutoff 3	287,165	304,498	323,916	356,322	362,109	245,302

Notes: These are the Cattaneo et al. (2016, 2018) optimal, data-driven bandwidth sizes, regardless of whether the multiple cutoffs are cumulative or not. We use the average across purchases and works bandwidths for RD manipulation and covariate balance tests.

Ignoring the multiple cumulative cutoff nature of Law 8,666/93, the independent calculation of bandwidths can generate some odd results, such as that for corruption outcome I for the works cutoff 3 (row six, column one). Here, we have a cutoff value, R$1.5M, for which the bandwidth, R$1M, spreads over the preceding cutoff, R$650,000, which makes little sense as a comparison of

sample units around one single cutoff. The reason why we do not drop this cutoff, however, is that we do not rely solely on CCT bandwidths for more granular analysis of the effect of discretion. We narrow down bandwidths for the most important results in our RD regressions, and provide parameter estimates for the entire range of bandwidths.

To lend additional support to our analysis, we conduct a means comparison test for municipal characteristics across procurement cutoffs – mimicking a covariate balance test commonly employed in randomized-controlled trials. Ideally, we would like to see no significant differences across observations just below and above the procurement types and cutoffs analyzed here. We report these results in table 7.

Table 7: Means Tests p-values Across Cutoffs

	Purchases		Works		
Municipal Variables:	Cutoff 1	Cutoff 2	Cutoff 1	Cutoff 2	Cutoff 3
Urban Population (Share)	0.756	0.775	0.297	0.702	0.617
Female (Share)	0.281	0.078*	0.005***	0.857	0.157
Illiteracy Rate	0.159	0.169	0.220	0.140	0.949
GDP	0.284	0.140	0.995	0.721	0.502
Gini Index	0.153	0.163	0.071*	0.712	0.457
Human Development Indicator	0.105	0.205	0.227	0.107	0.826
Poverty Rate	0.079*	0.109	0.454	0.097*	0.679
Presence of AM Radio	0.799	0.359	0.002***	0.315	0.388
Education Council Established	0.430	0.301	0.056*	0.275	0.523
Health Council Established	0.844	0.274	0.648	0.469	0.160
Seat of Judiciary Branch	0.002***	0.004***	0.158	0.516	0.400
Vote Margin	0.815	0.918	0.728	0.242	0.900
Mayor Reelection Rate	0.785	0.332	0.726	0.745	0.250
Sample Size (Below; Above)	(363; 835)	(877; 555)	(70; 177)	(406; 238)	(33; 15)

Notes: we used the Calonico et al. (2015) bandwidths calculated in table 6. In total, there are 36 unique bandwidths from the combinations across procurement type, outcome, and cutoff. We narrow down to one single bandwidth per procurement type and cutoff by focusing only on the most important outcome, performance indicator I, such as we had done for table 5. We compute the bandwidth for both the corruption and the mismanagement version of indicator I and use the smaller bandwidth across the two for robustness purposes, as we want like to narrow down on the samples across cutoffs as much as possible.*p<.1; **p<.05; ***p<.01

The interpretation of table 7 is straightforward. Significant differences across rows of covariates would mean that our local samples were observable different. The identification strategy would not point to causal inference of the effect of discretion on government performance. However, this is not what we see in table 7 as no covariate mean is significantly different across more than two cutoffs. In fact, precisely because municipalities can differ in these observable dimensions is why we collect municipal data in the first place and use to strengthen our RD estimations (Calonico et al., 2018). The only more concerning result is the systematic difference in those transfers just below and above cutoffs 1 and 2 for the purchases group in the judiciary seat variable (row eleven, columns one and two). The procurement calls just below each cutoff are different from the ones just above in that they generally occur at municipalities that do not have local offices of the Brazilian Judiciary (.34

and .43 averages for cutoff 1; .51 and .59 for cutoff 2). However, since the assignment of state and federal judicial offices follows population and service provision criteria, we do not believe they are correlated with local government performance level (or, for that matter, officials' ability to manipulate procurement rules out of stricter categories).

The main RD estimations are reported in table 8, where we are looking at one equation per cell, i.e. a total of (1×2) purchases cutoffs $+ (1 \times 3)$ works cutoffs $\times 6$ outcomes $= 30$ regressions.[14] In each cell, we are regressing the outcome on the column against the procurement category variable on each row (and as defined in table 3). Point estimates are the main statistics reported, followed by clustered (at the municipal level) robust standard errors in the first parentheses and sample sizes at each bandwidth in brackets. For clarity, we only report regressions that do not include covariates nor fixed-effects but the results are robust to their inclusion.[15]

The first important takeaway from table 8 is no significant effect of discretion on the likelihood of finding corruption in procurement calls (corruption outcome I). No parameter in column 1 is statistically significant, regardless of sample type (purchases, public works, or pooled sample) and procurement category. This is the most important result as it is diametrically opposed to the raison d'être of Law 8,666/93, which was passed to reduce corruption. As for outcome II (corruption share of total infractions), we only see a significant effect in the pooled estimates for cutoff 1, where moving from direct contracting to invitational bidding would reduce corruption by 5.4 percentage points. This is effectively a reduction in the intensity margin of corruption (number of times agents engage in wrongdoing per transfer) but not on the extensive margin (number of times they engage in corruption when they have a chance nor the amount they appropriate as measured by outcome III). Thus, we fail to confirm hypothesis 1 that stricter procurement rules reduce corruption.

Policy implication, however, differ. The immediate consequence is weak support for procurement legislation from both legislators and citizen, if at all. Since Law 8,666/93 does little to prevent corruption, there is no reason to keep it in place as it is an excessive burden for government agencies to procure goods and services which are inputs to their policy production function. The indirect effect, however, requires the analysis of the effect on program management (hypothesis 2).

---

[14]Since we dropped cutoff 3 for the purchases group, we are effectively dropping an entire 6-outcome row from table 6 when estimating our model.

[15]We have include all covariates and fixed-effects in regressions omitted from the text and the points estimates remain unchanged.

Table 8: Multiple Non-Cumulative Cutoff Procurement Estimates

Variable:	Corruption Outcomes			Mismanagement Outcomes		
	Indicator I	Indicator II	Indicator III	Indicator I	Indicator II	Indicator III
	(Binary)	(Share)	(Amount)	(Binary)	(Share)	(Amount)
	(1)	(2)	(3)	(4)	(5)	(6)
	Purchases Estimates					
Proc. Category 1	0.117	0.037	234	0.049	-0.006	64
	(0.095)	(0.055)	(442)	(0.078)	(0.081)	(648)
	[n = 2098]	[n = 2087]	[n = 1934]	[n = 2129]	[n = 2097]	[n = 1924]
Proc. Category 2	-0.034	0.032	2,827	-0.100*	-0.055	-4,524
	(0.067)	(0.048)	(3,838)	(0.059)	(0.059)	(4,748)
	[n = 2686]	[n = 2106]	[n = 2058]	[n = 2553]	[n = 2328]	[n = 2261]
	Works Estimates					
Proc. Category 1	-0.210	-0.039	61	-0.416**	-0.291*	-4,611*
	(0.143)	(0.078)	(1,482)	(0.190)	(0.178)	(2,802)
	[n = 330]	[n = 362]	[n = 314]	[n = 485]	[n = 485]	[n = 423]
Proc. Category 2	0.043	0.009	537	0.016	-0.006	-726
	(0.102)	(0.058)	(8,614)	(0.079)	(0.076)	(11,265)
	[n = 1185]	[n = 1014]	[n = 1050]	[n = 944]	[n = 858]	[n = 892]
Proc. Category 3	0.043	0.158	221478	0.171	-0.077	-88,510
	(0.307)	(0.183)	(281,249)	(0.212)	(0.251)	(340,566)
	[n = 313]	[n = 228]	[n = 205]	[n = 51]	[n = 157]	[n = 276]
	Pooled Estimates					
Proc. Category 1	-0.025	-0.054*	-425	0.098*	0.091*	-114
	(0.058)	(0.033)	(291)	(0.054)	(0.050)	(451)
	[n = 2256]	[n = 2263]	[n = 1964]	[n = 1977]	[n = 2019]	[n = 1905]
Proc. Category 2	-0.018	0.009	1,344	-0.064	-0.040	-572
	(0.043)	(0.030)	(3,078)	(0.044)	(0.041)	(3,486)
	[n = 3767]	[n = 2896]	[n = 2846]	[n = 2460]	[n = 2436]	[n = 4584]
Proc. Category 3	0.099	0.085	76,083	-0.006	-0.050	-1,869
	(0.128)	(0.069)	(59,995)	(0.081)	(0.079)	(79,197)
	[n = 517]	[n = 568]	[n = 626]	[n = 732]	[n = 751]	[n = 423]

Note: This table reports the lower-bounded LATE of discretion on government performance in
Brazilian municipalities. Each cell represents one non-parametric, local polynomial regression
of order two between the discretion parameter (rows on first column) and performance outcome
(columns 1-6). For each cell, we first report parameter estimates, then cluster-robust standard
errors (at the municipal level), and lastly sample sizes around each cutoff for each regression.
Corruption Indicator I (column 1) and Management Indicator I (columns 4) are binary variables
turning on when the procurement call has at least one corruption or program management
infraction; Corruption and Mismanagement Indicators II (columns 2 and 5) report the share of
corruption and mismanagement infractions over the total number of infractions; Indicators III
(columns 3 and 6) report the amount potentially lost to either problem, i.e. the transfer amount
times the share of corruption or mismanagement issues over total infractions. The regressions
in table 8 do not include municipal covariates nor fixed-effects. Robust standard errors are
clustered at the municipal level and are reported in parentheses. *p<.1; **p<.05; ***p<.01

We then turn to columns 3-6 in table 8, where we report government performance as program
management mistakes. The outcomes are parallel to those of corruption and summarize the like-
lihood any transfer will have evidence of management problems (outcome I), the share of total
problems due to mismanagement (outcome II), and the amount of money potentially lost to poor

resource management (outcome III). In the purchases group, the only significant effect is on mis-management outcome I for category two, i.e. moving over from invitational to price comparison bidding, which reduces the likelihood of mismanagement by 10 percentage points.

In addition, the most consistent and strongest effect is that of moving from direct contracting to invitational bidding when municipalities are procuring goods and services for public works. A shift in discretion causes a 41.6 percentage point reduction in the likelihood that there will be program implementation issues (outcome I), a 29.1 percentage point fall in the share of management-related problems (outcome II), and R$4,611 reduction in the amount of resources that would have been lost due to poor program management (outcome III). These effects are robust to the inclusion of covariates and changing bandwidths (as discussed in section 5.1). Thus, at least at cutoff 1, we can confirm hypothesis 2 that stricter procurement rules reduce corruption when it comes to calls for goods and services used in public works projects. This indirect effect of discretion in public spending is an important determinant for citizens' support for the Brazilian procurement law. As discussed in section 3.3, citizens are more likely to support any corruption legislation if there are additional benefits that prevent inefficient allocations of public resources and increase or improve the provisions of public policies. We investigate this effect further in remainder of this section and paper.

In figure 1, we plot the local polynomial regressions for works cutoff 1 for all outcomes. These figures are the graphical representations of the regression equations for works cutoff 1 using the mismanagement outcomes as dependent variables (row three, columns 4-6).

Figure 1: Discretion Effect on Mismanagement (Works Cutoff 1)
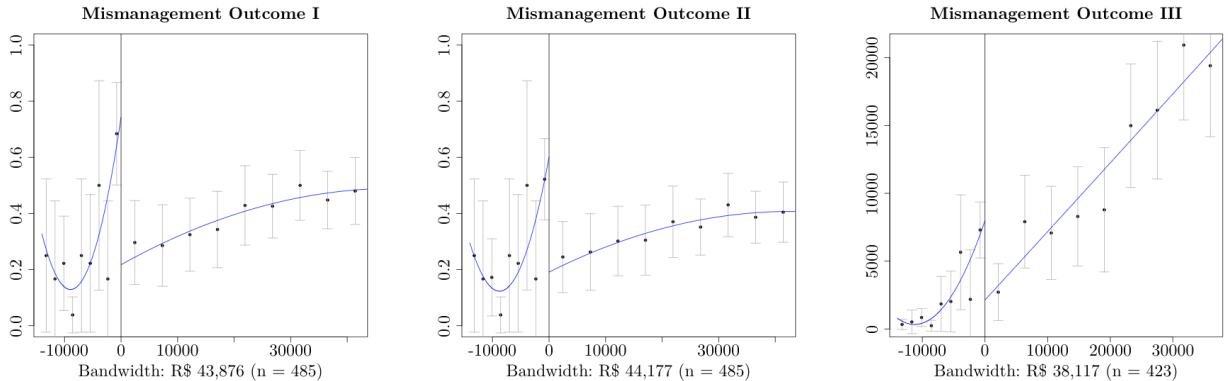
Note: The three figures in figure 1 report the minimum local causal effect of discretion (as a change in procurement rules from direct contracting to invitational bidding) on the three mismanagement outcomes. They plot the regression equations at the bandwidths specified in table 6 for mismanagement outcomes I through III when the federal transfer was used to procure goods or services used in public works. We also include 90% confidence intervals in horizontal bars for each RD bin.

Each graph uses the standard Calonico et al. (2015) bandwidths, a two-order polynomial fit for local estimating equations, and clustered-robust standard errors. We center the assignment variable (transfer amount) around zero, so all x values are differences (in R$) from the first cutoff for works at R$15,000. We see statistically significant breaks in outcome means close to the cutoff for all three variables. For outcomes I and II, mismanagement increases at decreasing rates for

observations further out to right, as the amount procured increases. The graphs for these two variables are similar, but they are complementary measures of performance. While outcome I summarizes the likelihood that any transfer sees mismanagement problems (extensive margin), outcome II looks at the share of all issues that are attributable to mismanagement *conditional on finding any mismanagement problem* (intensive margin). Finally, the last outcome summarizes the amount potentially lost to corruption by multiplying outcome II and the transfer amount. The relatively large magnitude of this parameter (R\$4,611 in a R\$15,000 cutoff) means that the procurement legislation is preventing a large amount of funds from being inefficiently applied.

Figure 2: Discretion Point Estimates (Works SOs at Cutoff 1)
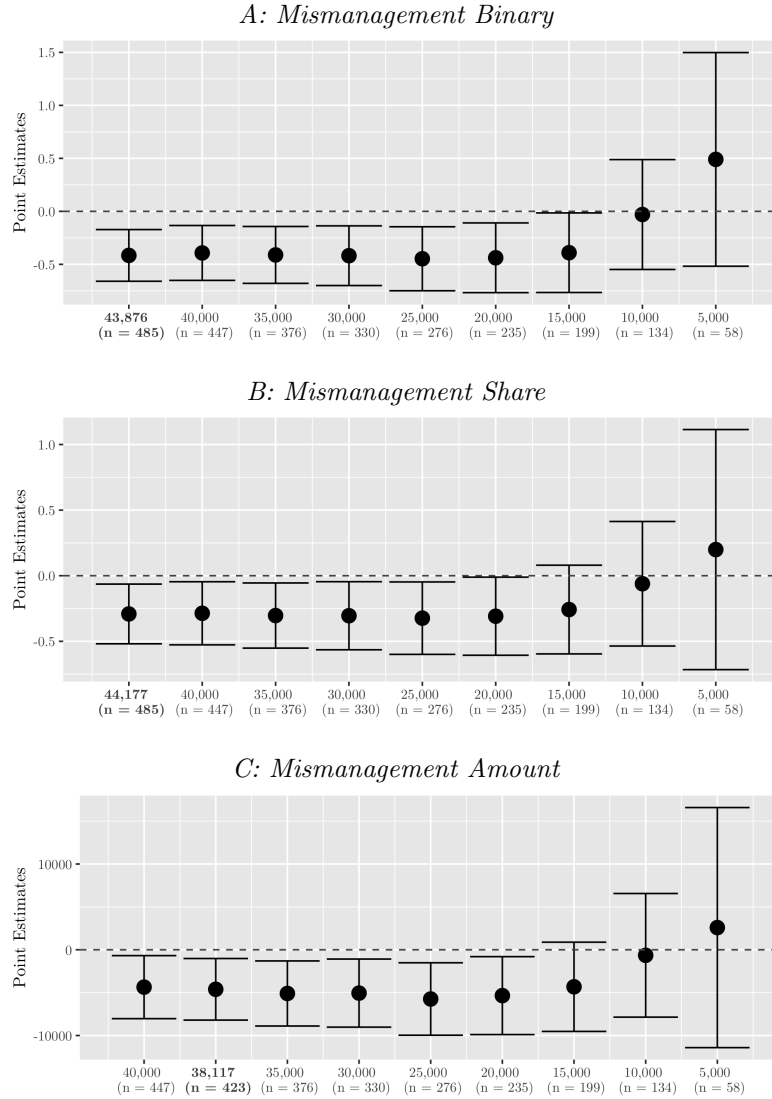
A: Mismanagement Binary

B: Mismanagement Share

C: Mismanagement Amount

Notes: These three graphs report the discretion point estimates and their 90% confidence intervals for different bandwidths across works cutoff 1. The number of observations used in each local quadratic polynomial regressions is in parentheses across the horizontal axis. Confidence intervals overlapping the dashed lines show statistically insignificant parameters. Calonico et al. (2015) bandwidths are in bold.

To further check the validity of our parameter estimates, we conduct a last test on the magnitude and significance of our discretion parameters for the works subsample at cutoff 1. In figure 2, we use smaller bandwidths at which we calculate our lower-bounded LATE starting from the Calonico et al. (2015) and moving down until we reach R$5,000 away from the cutoff on both sides. The point estimates are in the y-axis while the x-axis reports bandwidth and sample size. The point estimates are statistically different from zero for outcome I in seven out of nine bandwidths and six out of nine for outcomes II and III. The insignificant estimates at the smaller bandwidth are telling of the trade-off between significance and sample size in RD strategies. Were the samples larger, we would have seen the significant effect persist across all bandwidth sizes as standard errors would likely converge to that of significant parameters. Nevertheless, we propose two falsification tests in section 5.1 which serve as additional evidence supporting the significant reduction of management problems.
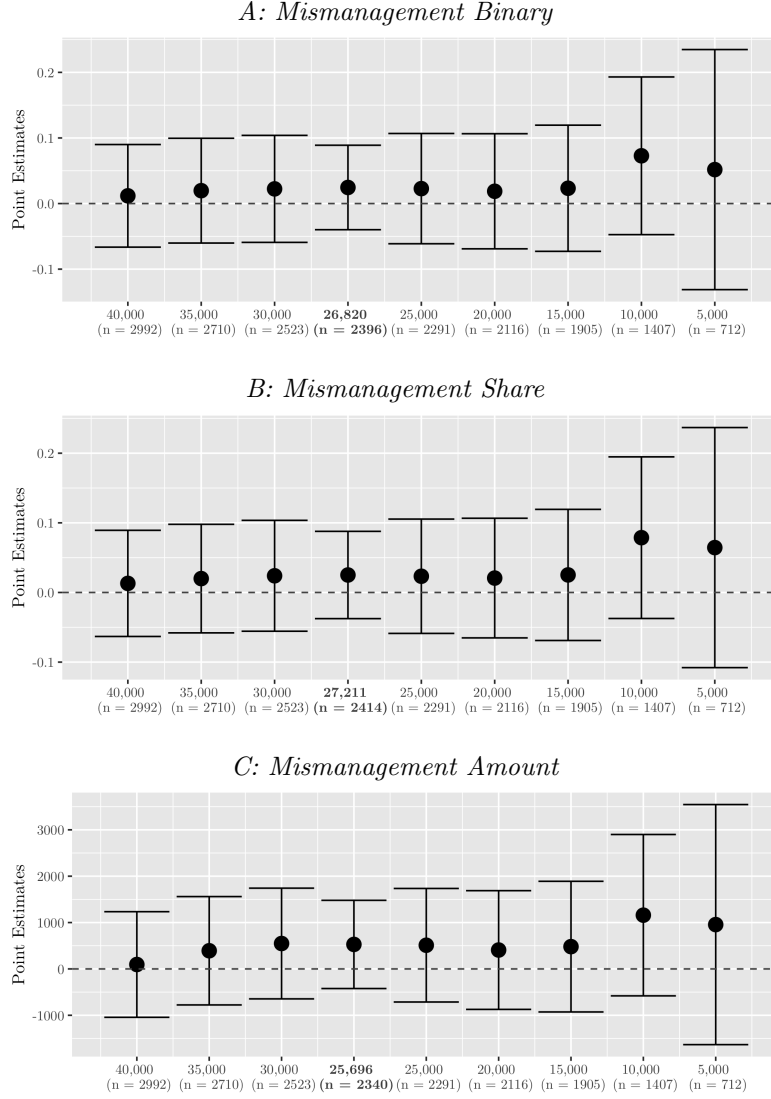
5.1 Falsification Tests

In order to confirm that the discretion effect reducing public works program management is not spurious nor due to random chance, we perform two additional tests on our sample of federal transfers.

The first test we conduct is erroneously imposing the public works cutoff 1 on purchases group observations. As a reminder, there are two types of procurement in Brazil: what we call *purchases*, which consist of any purchase of goods and services, and *public works*, which consist of purchases of goods and services to be used solely on public works projects. Since every public work requires procurement of any sort, the works group is a subsample of the overarching procurement of any good or service. If there is any significant effect of discretion at a falsified cutoff, then there is reason to believe that the effect we identified is either spurious or random, therefore meaning that procurement restrictions have no positive effect on program management.

Similarly to figure 2, we plot point estimates and 90% confidence intervals for falsified cutoffs in figure 3. In these graphs, we are only using observations from the purchases subsample and running the exact same regressions as in sections 4 and 5 with two-order polynomial fits, clustered-robust standard errors, and the same varying bandwidths as in the works-only sample.

Figure 3: Discretion Point Estimates (Purchases SOs at Works Cutoff 1)

A: Mismanagement Binary

B: Mismanagement Share

C: Mismanagement Amount

Notes: These three graphs report the discretion point estimates and their 90% confidence intervals for different bandwidths across works cutoff 1 *using purchases SO as falsification data*. The number of observations used in each local quadratic polynomial regressions is in parentheses across the horizontal axis. Confidence intervals overlapping the dashed lines show statistically insignificant parameters. Calonico et al. (2015) bandwidths are in bold.

There are no significant results across all figures at 90% confidence intervals. Even with larger samples, there is no effect of the falsified cutoff on the likelihood of finding mismanagement, nor on the share of irregularities due to management problems, nor the amount potentially lost to resource misallocation. This is an important result supporting that the effect of discretion on mismanagement of public works projects is, in fact, true. Moreover, the point estimates are either close to zero or positive, which is opposite to the expected result in hypothesis 2 and confirmed in table 8.

An additional source of concern might be on the relationship between procurement legislation and program implementation. If procurement calls are carried out once for contracting the provision of goods and services, and their rules aim at preventing corruption, there is no direct link between imposing such rules and seeing better program implementation that might occur only after procurement has concluded. We posited, however, that the work that goes into abiding by procurement rules also spreads over better resource management practices, such as speedy procurement processes, that result in better program delivery. Quality and quantity requirements, delivery times, and the duration of contract for provision of goods and services are other factors included in procurement calls that can also improve public management. Therefore, there is a non-trivial benefit of restricting discretion on resource allocation, which is the focus of the second falsification test.

In figure 4, we once again run the exact same regressions from sections 4 and 5 at varying bandwidths for the sample of *non-procurement* federal transfers erroneously imposing works cutoff 1 at R$15,000. This is a sample composed of 4,925 investigations (SOs) conducted by CGU auditors that were not identified as procurement because they met at least one of these conditions: (i) there was no textual description on the use of resources; (ii) there was no monetary amount reported by the auditor; (iii) the classification algorithm did not assign the investigation to any group. To further narrow down on non-procurement transfers, we exclude from this sample any investigation that could be matched to public works projects by grant description or by having works-related infractions as reported in audit reports (the two classification checks in appendix A). The final non-procurement sample is composed of 2,493 investigations over the use federal resources for which auditors report the total value of transfer and for which there is not a single evidence of use in public works projects.[16] This is a prime sample of transfers that have nothing to do with procurement in general, let alone procurement for public works projects, and for which we do not expect any effect of discretion.

A close look at figures 4.A through C reveals no significant relationship between false cutoff 1 on mismanagement for the non-procurement sample (at the 90% level). As we reduce the bandwidths around the false cutoff, all parameters remain indistinguishable from zero, though at smaller sample sizes we see much wider confidence intervals. Again with the case of public works true sample, more observations would help narrow down standard errors and potentially align parameter estimates along the results of larger bandwidths. Therefore, we can safely conclude that procurement requirements in fact help solve management issues that might not have been related to procurement in the first place. Standardizing procurement practices, delivering public goods and services faster, and following policy-mandated guidelines have a direct effect of improvement government performance.

---

[16]Non-procurement federal transfers are used for payment of personnel and new hires, covering budget shortfalls, for extension of procurement contracts after calls under Law 8,666/93 have been concluded, etc.

Figure 4: Discretion Point Estimates (Non-Procurement SOs at Works Cutoff 1)
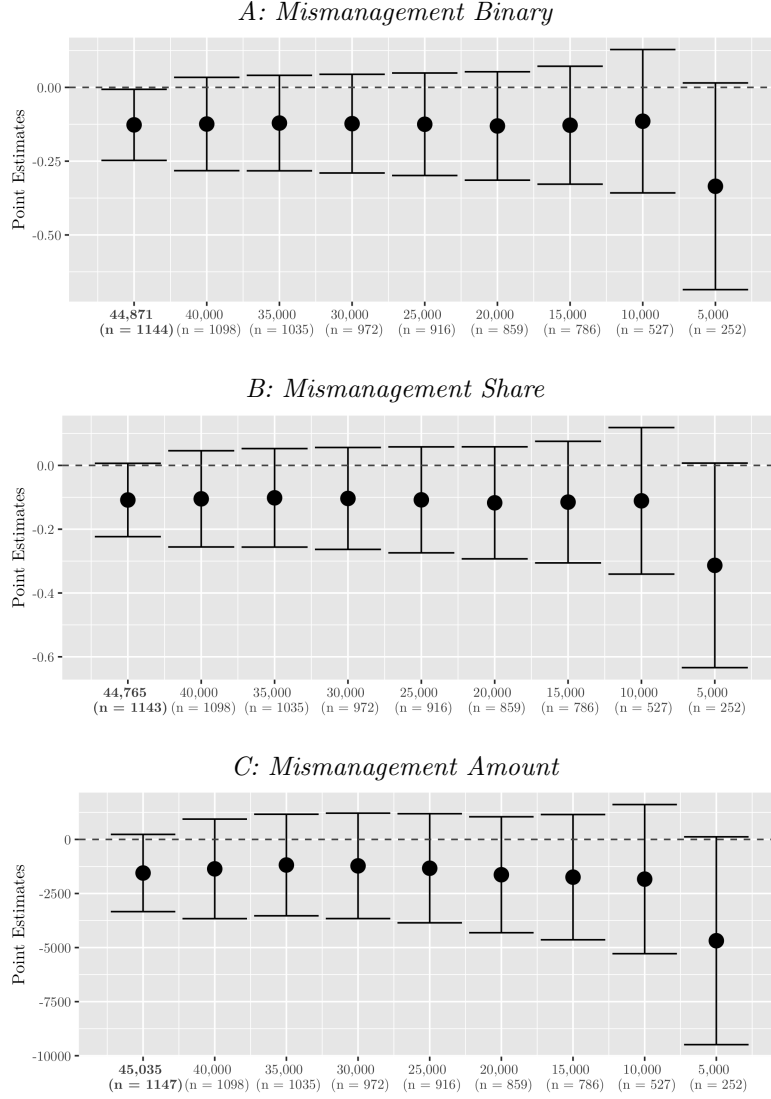
A: Mismanagement Binary

B: Mismanagement Share

C: Mismanagement Amount

Notes: These three graphs report the discretion point estimates and their 90% confidence intervals for different bandwidths across works cutoff 1 *using non-procurement SO as falsification data*. The number of observations used in each local quadratic polynomial regressions is in parentheses across the horizontal axis. Confidence intervals overlapping the dashed lines show statistically insignificant parameters. Calonico et al. (2015) bandwidths are in bold.

5.2 Interpretations of the Null Effect on Corruption

In this section we return to our primary hypothesis and suggest possible interpretations for the null effect of discretion on corruption. It is naive to think that legislators knew nothing about corruption prevention and that Law 8,666/93 is entirely flawed. A more plausible explanation, however, is that corrupt behavior has changed over time while the procurement legislation has remained static. In this case, the law was appropriate under different corruption conditions than were not present in

our period of study (2004-2010) but perhaps might have been present before, between 1993-2003.

The first potential interpretation is that public officials might have grown accustomed to procuring goods and services under the legal requirements and have thus been able to carry on engaging in corruption while avoiding its detection. If this is the case, then we should not see differential levels of municipal corruption across time. We plot anecdotal data supporting this conclusion in table 5.

Figure 5: Municipal Corruption Between 2004-2010
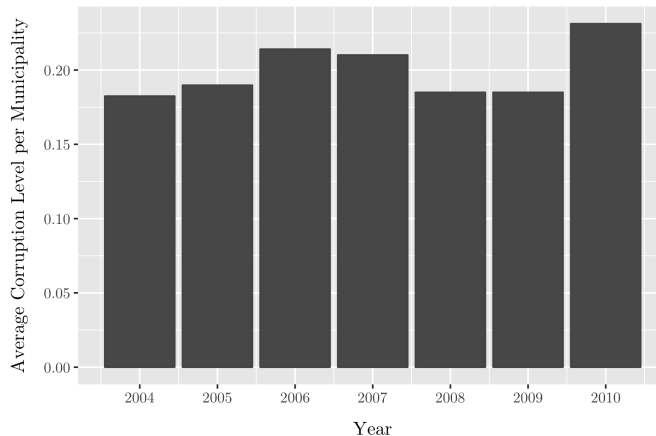
Municipal corruption levels are fairly stable over time. The average corruption share ranges between .18 to .21 of total infractions. Obviously, there are other things going on simultaneously besides learning how to work around procurement rules, but the relative stability of municipal corruption could also be a feature of the ease with which officials can sidestep rules if they want to engage in corruption. Moreover, the burden for corruption falls more heavily on mayors (Finan et al., 2018) than on their lower level staff, also indicating another incentive for why officials would keep their corruption levels stable over time.

Another explanation of the null effect is the decreasing purchasing power of public resources in the presence of inflation. Since the cutoff values across categories remain constant over the period under analysis, inflation reduces the quantity of goods and services any single procurement call can contract, thus turning small procurement amounts (regulated by lower categories of Law 8,666/93) meaningless in terms of their complexity or the corruption opportunity they create – or even the nominal value taken out as corruption. This means that compliance at lower categories of the procurement law might be easier due to the relative simplicity, lower quality, or smaller quantity of the items being procured. We propose checking this explanation by looking at the average number of infractions (both in corruption and management) across time, procurement type, and category. We present these results in table 6, where we see the average number of infractions per investigation on the y-axis, and the year of investigation on the x-axis. We further split the results by type of procurement (public purchases or works) and category (direct contracting, invitational bidding, price comparison bidding, and competitive bidding). Thus, each quadrant summarizes the average

number of infractions per year, procurement type, and category. The dashed black lines are the average number of infractions across type and category and removes the time variation.

Figure 6: Average Infraction Count Per Year, Procurement Type, and Category
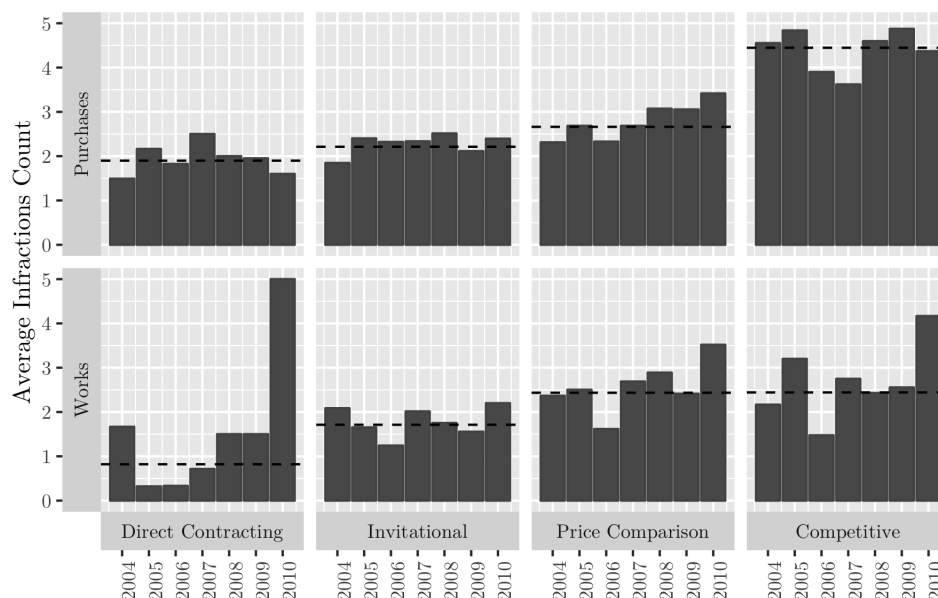
Across both types of procurement, we see the average number of infractions increase as we move up procurement categories (shown by the level of the dashed black lines read from left to right). Though not conclusive, this is suggestive of decreasing compliance when procurement increases in value and complexity. For the purchases group, the top four quadrants, we see relative stability of the average number of infractions within each procurement category, which, read along figure 5 on the stability of the corruption share over all issues, indicates little variation in the number of management problems as well. For the works group, in the bottom four graphs, we see 2006 and 2010 as outline years for lower and higher averages, respectively, but the upward trend in infractions is also consistent across procurement categories.

These are but two interpretations of the null effect on corruption and we do not aim at exhausting the alternative explanations for such unexpected effect. We suggest this puzzle as an avenue for future research on the effect of discretion on performance.

5.3   Cost-Benefit Analysis

So far, we have focused on testing our hypotheses and uncovering potential confounding effects of discretion. We now move on to discuss the welfare effects of Law 8,666/93, which should serve as a transparent metric to support or reject the procurement law.

The only significant, beneficial effect is the reduction in mismanagement for the public works sample at the change in rules from direct contracting to invitational bidding. The costs, however,

are all other corruption and mismanagement problems not prevented at the equivalent cutoffs. We conduct a back-of-the-envelope cost-benefit analysis (CBA) in table 9.

Table 9: Welfare Effects from Discretion at Cutoff 1

Cost	Type	Avg. Loss (in R$)	# Obs.	Total (in R$)
Corruption	Purchases	2,491	1,934	4,818,570
Mismanagement	Purchases	10,659	1,924	20,508,793
Corruption	Works	2,871	314	901,789
Mismanagement	Works	10,529	423	4,454,134
			Total Cost (A)	30,683,288
Benefits				
Works	Mismanagement	-4,611	423	-1,950,453
			Total Benefit (B)	-1,950,453
Welfare Effects				
$A - B$	*Total Cost (in R$) in the absence of Benefit*			32,633,741
$100 \times B/(A - B)$	*% Cost Reduction*			5.98
$A_{works} - B$	*Works Cost (in R$) in the absence of Benefit*			7,306,377
$100 \times B/(A_{works} - B)$	*% Works Cost Reduction*			26.69

In the top part of the table, we calculate the costs of Law 8,666/93 as the total misplaced amount not prevented by invitational bidding contracting rules. These costs come from both procurement types, public purchases and works, and outcomes, corruption and mismanagement. In column three, average loss in R$, we include the average amount misplaced due to either problem for all observations within the CCT bandwidth around cutoff 1. Since our parameters are lower-bounded LATE, they only constitute causal effects at the vicinity of the discretion rule discontinuity, thus the focus on these subsamples. We multiple each average cost by the number of observations and add them up to a total cost (A) of over R$30M (last row of top part).

The total benefit, on the other hand, comes from the amount that would have been misplaced had Law 8,666/93 not been in effect. On column three, in the middle part of the table, the average loss prevented is the LATE parameter at works cutoff 1 for mismanagement outcome III, which calculates the amount potentially lost by multiplying the share of mismanagement issues times the transfer amount. (B) is the total misallocation prevented, approximately R$2M.

Finally, we report welfare effects at the bottom part of the table. We calculate both misallocation prevention in absolute (R$) and relative (%) terms. Total cost of misallocation would be approximately R$32.5M and the works-only cost would be R$7.7M (compared to R$4.4M in the absence of the law). Moreover, the discretion policy reduces misallocation by 5.98% and 26.69%, respectively, for total expenditures and expenditures in public works projects. Although this is a substantial reduction in misplaced funds, R$30M still remain unaccounted for, and its misallocation could have been prevented had the procurement legislation been adapted over time. Citizens and legislators should compare this potential recovery to overall municipal budgets in order to support Law 8,666/93 or not.

6 Conclusion

In this paper, we looked at the effect of discretion in public spending on government performance. We operationalize this relationship by exploring the Brazilian procurement Law 8,666/93, which determines rules that should be followed at all government level and branches when procuring goods and services. Our observations are investigations conducted by CGU's random audit program between 2004 and 2010, from which we can create objective measures of corruption and mismanagement of public resources. We infer procurement type using a text analysis algorithm developed by one of the authors, described in the appendix, to match each investigation to procurement types and categories established by Law 8,666/93. Besides contributing to the literature by exploring discretion as determinant of government performance, we also investigate whether performance is associated with the overall procurement amount and local municipal corruption.

We cannot confirm that discretion reduces corruption. Using our preferred RD identification strategy, we find no significant effect of discretion on corruption, regardless of whether the procurement call was motivated by public works projects or any other provision of goods and services. This results has a direct, and important, policy implication regarding the support for the Brazilian procurement law. In the simple framework outlined in section 3.3, legislators and citizens would both weakly support legislation that does not deliver on its promise of reducing corruption.

Some explanations emerge for the non-significant effect on corruption. The strongest interpretation we suggest is that the inflation-unadjusted amounts determining procurement rules have become meaningless over time. Procurement compliance for items of lower amount might just be simpler than at higher contract values. In fact, figure 6 shows that the average number of infractions is increasing both across time and procurement category, suggesting that non-compliance is less frequent as the purchasing-power of procurement calls decreases. Additionally, Finan et al. (2018) show that the likelihood of punishment for corruption is low and, if at all, fall more heavily on mayors than on lower-level officials. If this is the case, then officials have relatively more incentives for non-compliance and are likely to keep the same level of corruption regardless of rules imposing less discretion (anecdotal evidence in figure 5).

The discretion effect on mismanagement is moderate at best. We find no effect for the general procurement group (*purchases*) but find a consistent effect on reducing program management problems when moving from directly contracting with private suppliers to invitational bidding in public works projects. The likelihood of finding any mismanagement problem decreases by 41.6 percentage points. Conditional on finding any issue, the share of program management infractions falls by 29.1 when implementing stricter rules and the amount potentially misplaced reduces by R\$4,611. Though our results are concentrated in one cutoff and for one type of procurement, they are important initial evidence disentangling the effect of discretion from broader measures such a political decentralization and devolution of funds. Since our estimates are lower-bounded LATE, there still remains a large field of research for identifying the precise magnitude of discretion across the entire distribution of public spending decisions.

To check these results, we carried out two falsification tests: running the same regressions that

yielded significant parameters on falsified datasets for which there should be no effect. We used the purchases and the non-procurement samples to confirm (i) that the effect is exclusive to public works procurement and (ii) that procurement is indeed relevant for government performance. There are no significant effects on mismanagement at the falsified cutoffs, regardless of the data we use.

Lastly, we provide a rough estimate of the costs and benefits of the procurement legislation in Brazil for a transparent metric with which to decide whether discretion restrictions should be kept or not. Though the legislation reduces the total amount of missing funds by 5.98%, there are still R$30M missing – equivalent to six times the yearly average of funds transferred from the federal government to municipalities in the period under analysis. Discretion restrictions, though effective at reducing mismanagement, are not unconditionally positive for government performance.

References

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton university press.

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.

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., Farrell, M. H., and Titiunik, R. (2018). Regression Discontinuity Designs Using Covariates. *The Review of Economics and Statistics*.

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 audits, 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 10: 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.[17]

---

[17]One of these keywords in the works search pattern is an "exclusion keyword," which removes service orders that

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. Out of an universe of 14,518 observations, the search patterns here identify a total of 9,593 procurement-related service orders.

Figure 7: Sets of Procurement Service Orders
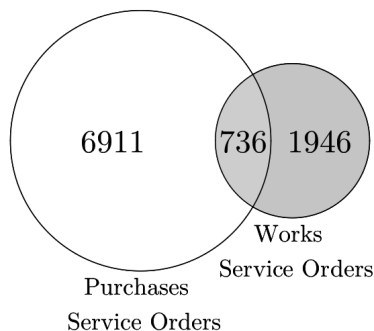
As Grimmer and Stewart (2013) rightly point out, no text analysis algorithm is perfect and only relying on keyword matches could potentially lead to misclassification of service orders. An example might help illustrate this. Let us suppose that one description reads "expenditures made in accordance with 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

---
contain the "exclusion keyword" in their description from the sample identified by the other 16 n-grams.

when in fact they are not; words can take on different meanings depending on their contexts, so 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 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 11: 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.[18] 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

---

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

not significantly identify a new sample ("saneamento" and "conclus[ãa]o"). Despite having positive individual finds reported in column 1, table 12, 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.[19]

Table 12: 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[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

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 categories set out in Law 8,666/93 apply to all public procurements at all government levels, transfer types are irrelevant for constructing our procurement discretion measure. Nonetheless, we can test the quality of our classification algorithm on one type of these transfers.

---

[19]The search without these terms (available upon request) yields 2,679 service orders, just three short of the total in table 12. Nevertheless, we keep the two items in the search algorithm for additional tests discussed in section A.2.

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 these 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.[20]

Table 13: Classification by SO and Grant Description

		Panel A: Purchases Group					*Panel B: Works Group*		
			Grant					Grant	
		No	Yes	Total			No	Yes	Total
SO	No	115	144	259	SO	No	1546	269	1815
	Yes	83	1473	1556		Yes	404	1649	2053
	Total	198	1617	1815		Total	1950	1918	3868

There is a total of 3,868 service orders[21] for which we have descriptions both from CGU and from the federal government. In table 13, 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 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 units 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).[22]

---

[20]http://www.portaltransparencia.gov.br/

[21]The reason why the total testable sample for the purchases group is 1,815 units is that we have to exclude potential works observations, in the same way discussed in the introduction of appendix A. Hence, we have to subtract the 2,053 possible works units from the total 3,868 and run the check only in the remaining 1,815 observations.

[22]The inverse misclassification rates are also reassuring: false positives are 8.9% and 14.0% for purchases and

A.3 Procurement Violations

Though section A.2 supports the external validity of our algorithm 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 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.[23]

Table 14: Classification by SO Text and Procurement Code

Panel A: Purchases Group					Panel B: Works Group			
		Procurement Code					Procurement Code		
		No	Yes	Total			No	Yes	Total
SO	No	2487	319	2806	SO	No	2462	344	2806
	Yes	6137	3456	9593		Yes	5791	3802	9593
	Total	8624	3775	12399		Total	8253	4146	12399

The total number of service orders[24] 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 14, 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).

---

works respectively when we first classify observations using grant descriptions and then move on to service order descriptions.

[23]The reason why this is a type II error test, instead of type I, resides on the way these 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 but were not. This sample is clearly underidentified, as there are many procurement-related transfers that simply followed Law 8,666/93 and thus carry no infraction, but still this subset of all CGU investigations provides us with a good counterfactual against which to test our classification mechanism.

[24]Although the universe is composed of 14,518 observations, we had to exclude the remaining 2,119 observations for which no textual description exists, hence the 12,399 universe in table 14.

B CEPESP Coding of Service Orders

Table 15: 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 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 five of the six discretion cutoffs imposed by Law 8,666/93. The only discontinuity for which variables do not seem comparable is purchases cutoff 3, which we drop from the analysis in sections 4 and 5.

Figure 8:   Manipulation Tests at SO Cutoff Amounts