Reelection Incentives and Corruption: Revisiting the Evidence with LLM-Classified Audit Reports*

Ricardo Dahis
Monash University

Martin Mattsson
NUS

Nathalia Sales
PUC-Rio

May 13, 2025

Abstract

Reliable data on corruption are notoriously difficult to obtain. In this paper, we extend and reanalyze corruption data from Brazilian municipalities. We extend the data by employing a Large Language Model (LLM) to systematically classify 2,197 corruption audit reports. We first show that correlations between the LLM-generated corruption measures and manually coded assessments are comparable to correlations among the manual datasets themselves, highlighting both the relative reliability of the LLM classification and the inherent subjectivity involved in quantifying corruption from textual sources. Using this expanded corruption dataset, we revisit key findings in the literature about the impact of reelection incentives on corruption. Our results support previous findings in the literature that reelection incentives reduce corruption, although the result is only statistically significant for one out of three measures of corruption. Furthermore, we document significant heterogeneity in the effect over time and investigate several explanations for these empirical patterns, including changing composition of politicians and increasing probability of legal penalties.

*We thank participants at the MWZ Text-as-Data Workshop for comments. We thank Claudio Ferraz, Joana Naritomi, and Juan Rios for thoughtful suggestions. André Guimarães and Guilherme Soares provided excellent research assistance. Financial support was provided by the National University of Singapore. Any errors are our own.

1 Introduction

Corruption is a serious problem across the world and it is especially severe in low- and middle-income countries (Svensson, 2005). Several approaches have been proposed to combat it, including increasing the wages of potentially corrupt officials (Van Rijckeghem and Weder, 2001), reducing corruption opportunities by decreasing regulation and thus removing opportunities for corrupt behavior (Rose-Ackerman, 1998), improving both top-down and local-level monitoring (Olken, 2007; Björkman and Svensson, 2010), and implementing new technologies such as "e-governance" (Banerjee et al., 2020). To the extent that voters value honesty, reelection incentives would also be a strong force against corruption (Ferraz and Finan, 2008, 2011).

However, a major challenge in evaluating anti-corruption measures is the limited availability of reliable corruption data (Olken, 2007; Olken and Pande, 2012).[1] To address these concerns, researchers have developed more direct measures of corruption. One widely used data source is the audit reports generated by Brazil's Random Audits Anti-Corruption Program, introduced in 2003 by the *Controladoria Geral da União* (CGU). These reports have been extensively used to quantify corruption in academic research (Ferraz and Finan, 2008, 2011; Brollo et al., 2013; Avis et al., 2018; Colonnelli and Prem, 2022; Ash et al., 2025).

A key challenge in using audit reports is the infeasibility of manually reading and quantifying corruption findings across a large number of documents. In this paper, we leverage a Large Language Model (LLM) to analyze 2,197 audit reports (~185,000 pages), extending previous manual efforts. To improve readability for the LLM, we employ Retrieval-Augmented Generation (RAG) to extract relevant contextual information from each report. This extracted information is then fed into OpenAI's GPT-4. By supplying the model with the appropriate context, we enable it to answer specific questions regarding the audit findings, including whether corruption-related irregularities were identified, the proportion of audited resources associated with corruption, and the number of corruption cases.

Using this approach, we construct a new dataset on corruption covering all audit reports of Brazilian municipalities audited between 2003 and 2015. We compare this dataset with three manually coded sources: the 2003-2004 audit reports coded by Ferraz and Finan (2011), the 2003-2009 audit reports coded by Brollo et al. (2013), and a CGU-managed dataset documenting irregularities found in audits from 2005 to 2015. Comparing the LLM's corruption assessments with manually coded data, we find correlations in the share of audited resources where corruption was identified ranging between 0.44 and 0.47. These correlations are similar to those observed among the

[1] A common approach to measuring corruption relies on perception-based indicators, such as Transparency International's Corruption Perceptions Index (CPI). However, these surveys may be significantly biased by respondents' beliefs and characteristics (Olken, 2009).

manually coded datasets, which range from 0.48 to 0.71.[2]

We then apply the empirical strategy developed by Ferraz and Finan (2011) to our dataset and estimate the impact of reelection incentives on corruption. In Brazil, mayors can serve a maximum of two consecutive terms, meaning only first-term mayors face reelection incentives. We compare first-term and second-term mayors while controlling for a wide range of observable characteristics. Our findings provide evidence that reelection incentives reduce corruption, supporting previous literature. However, our estimated effect sizes are smaller and more precisely measured than those in the original study, and the effect is statistically significant for only one of the three corruption measures.[3] Additionally, alternative datasets from Brollo et al. (2013) and CGU further support our findings with estimates that are smaller and mostly statistically insignificant.

Differences in corruption estimates across datasets can largely be attributed to temporal variations in the effects of reelection incentives. When focusing on the 2003-2004 period, as studied by Ferraz and Finan (2011), all three datasets available for this period confirm that reelection incentives reduce corruption. However, for the three subsequent electoral terms for which we are the first to carry out an analysis, we find near-zero effects from 2005 to 2012, followed by large and significant effects from 2013 to 2015.[4]

This raises the question: Why did reelection incentives cease to deter corruption between 2005-2012, only for their effects to reappear in 2013? We explore four potential explanations. First, newly elected mayors in 2000 may have been particularly honest due to increased political competition, as that election was the first to allow mayoral reelection. With many incumbents running for a second term, new candidates faced stronger opponents, potentially leading to the election of higher-quality, less corrupt first-term officials. However, we do not find evidence in favor of the hypothesis that corruption levels are different when first-term mayors ran against an eligible incumbent.

Second, first-term mayors may have become more corrupt over time relative to second-term mayors. For instance, the rise of a new political party recruiting large numbers of candidates with weaker screening processes could contribute to this pattern. To test this hypothesis, we examine whether the rise of the Workers' Party (PT) in the 2004 and 2008 elections played a role. While first-term mayors are generally less corrupt, we find that PT-affiliated first-term mayors between 2005 and 2012 ex-

[2]While we use manually encoded reports to validate our methodology, we do not rely on prior encodings to "train" the LLM. Thus, our methodology does not depend on pre-existing manually encoded audit reports, eliminating concerns about in-sample vs. out-of-sample validation.

[3]For one of the three measures, we can reject the original paper's point estimate at the 95% confidence level.

[4]The Brollo et al. (2013) and CGU datasets confirm the pattern of near-zero effects from 2005 onward.

hibited higher corruption levels than other first-term mayors. This partially explains the declining gap between first- and second-term mayors in that period, though it is statistically significant only in the dataset from Brollo et al. (2013).

Third, the increasing number of legal actions against corrupt officials may have gradually overshadowed reelection incentives as a deterrent. Since both first- and second-term mayors face legal consequences, their corruption levels may have converged over time. Consistent with the findings from Avis et al. (2018), we show that the likelihood of prosecution within five years of election was higher for mayors elected in 2004 and 2008 than for those elected in 2000, before declining significantly for those elected in 2012. Thus, this is a likely contributing factor to the decline and reemergence of reelection incentives as an important deterrent to corruption.

Finally, it is possible that second-term mayors increasingly pursued state or national political careers, motivating them to maintain cleaner records despite lacking reelection incentives. However, empirical evidence does not support this hypothesis.

Our paper contributes to two strands of the corruption literature. First, we advance efforts to measure corruption (Olken, 2007), particularly through random audits (Cuneo et al., 2023). Previous research has relied on manual coding of audit reports (Ferraz and Finan, 2008, 2011; Brollo et al., 2013; Avis et al., 2018; Colonnelli and Prem, 2022).[5] We introduce an innovative approach using LLMs to encode audit reports, offering an automated, cost-effective, and scalable methodology that can be applied in other settings.

Second, we contribute to the literature on the political determinants of corruption. Ferraz and Finan (2011) argue that mayors with reelection incentives misappropriate 27 percent fewer resources than those without such incentives. Bobonis et al. (2016) find that corruption is lower in municipalities audited just before elections, though they observe no lasting effects in subsequent audits.[6] Axbard (2024) shows that convictions of corrupt officials reduce future corruption through a deterrence effect in the Philippines. Our findings underscore how the factors shaping measured corruption evolve over time, potentially obscuring the effects of electoral incentives.[7]

[5] Recent studies have used machine learning to detect corruption *out-of-sample*. For example, Ash et al. (2025) applied machine learning to municipal budgets and outcomes from Brollo et al. (2013), finding that a machine-guided audit strategy could detect nearly twice as many corrupt municipalities at the same audit rate. Colonnelli et al. (2022) used CGU audit encodings and municipal characteristics to predict corruption, highlighting financial development and human capital as key predictors. Our goal, in contrast, is to generate new *ground truth* data.

[6] A related body of research examines the direct impact of corruption audits and transparency initiatives. Ferraz and Finan (2008) show that exposing audit results significantly influenced the electoral prospects of incumbent mayors in 2004, reducing their likelihood of reelection by 17% when corruption was detected. Avis et al. (2018) find that being audited increased the probability of subsequent legal action by 20 percent and led to an 8 percent reduction in future corruption.

[7] Existing literature indicates that the effects of term limits can differ across periods. Building on the findings in Besley and Case (1995) about the fiscal impact of gubernatorial term limits in the U.S., Besley and Case (2003) document that these effects have shifted significantly over time. Initially, they

The rest of the paper proceeds as follows. Section 2 provides background information on Brazil's anti-corruption program and reviews data from previous encodings of audit reports. In Section 3 we present the methodology used to construct our corruption measures and compare our measures with previous manual classifications. Section 4 presents the identification strategy for the empirical exercise and the main findings. Section 5 tests various explanations for why the effects vary over time. Section 6 concludes.

2 Background

2.1 The Random Audits Anti-Corruption Program

In 2003 the Brazilian federal government created the *Controladoria Geral da União* (CGU), tasked with promoting transparency, preventing corruption, and enforcing integrity in public administration. As the primary oversight body, the agency is responsible for monitoring and auditing the utilization of public funds across various government agencies.

An important initiative introduced shortly after the CGU's creation was the *Programa de Fiscalização por Sorteios Públicos*. This initiative used a lottery to randomly select municipalities with populations under 500 thousand inhabitants to audit their use of federal funds. Over 13 years, the program conducted 40 lotteries and 2,197 audits across 1,910 municipalities. After 2015, the program was reformulated to include both random and non-random audits. For this reason, we restrict our analysis to the first 40 lottery rounds held between 2003-2015.

Once a municipality was selected to be audited, the CGU gathered information on all federal funds transferred to the municipal government in that political term and in some cases in the previous term. CGU auditors were then sent to the municipality to examine accounts and documents, as well as to evaluate the existence and quality of public infrastructure projects and the provision of public services.

The detailed inspections conducted by CGU auditors resulted in comprehensive reports detailing the extent of corruption and mismanagement. The reports range from 30 to 200 pages and are on average 85 pages long. These reports were submitted to the CGU headquarters after approximately one week of inspections and a few months later the summaries of the main findings were made available to the public online on CGU webpage.[8]

find that governors spent and taxed more when they could not stand for reelection. However, with data extended to the mid-1990s, this effect weakened and even reversed. Alt et al. (2011) explain this shift as a "competence effect," emerging from changes in the structure of term limits across states, as many states moved from single-term to two-term limits, allowing voters to retain more competent incumbents.

[8]The summaries can be found at https://auditoria.cgu.gov.br/.

2.2 Previous Encodings of Audit Reports

Given the limited availability of data on corruption, CGU's audit reports quickly became a popular source of data on corruption. In quantitative social science the audits were first used by Ferraz and Finan (2008). In this and subsequent papers, the reports were turned into quantitative data by manually encoding each report. Subsequently, a series of other papers used this classification as a basis and/or developed their own corruption classification (Ferraz and Finan, 2011; Brollo et al., 2013; Avis et al., 2018; Colonnelli and Prem, 2022; Ash et al., 2025).

To validate the corruption data encoded by LLM, we use the data from Ferraz and Finan (2011) (henceforth, FF) and Brollo et al. (2013) (henceforth, Brollo et al.).[9] The former manually classifies reports from 2003 to 2004 (lotteries 2 to 11), covering the 2001-2004 term, whereas the latter covers reports from 2003 to 2009 (lotteries 2 to 29) covering the period from 2001 to 2009. See Table 1 for an overview of the data used.

FF's main measure of corruption is the total amount of resources where some corruption was found, expressed as a share of the total amount of resources audited. We follow this convention and use this variable as our main benchmark when validating the encoding generated by the LLM. Additionally, we use two other variables provided by FF — a binary variable for if any corruption was found and the number of corruption cases. Similarly, Brollo et al. construct a continuous indicator, the ratio between the funds involved in the irregularities and the total amount audited, and a binary variable, whether any irregularity was found or not. The reports do not formally describe if an irregularity should be considered evidence of corruption or not. Therefore, Brollo et al. divide potential corruption cases into general (broad) irregularities, that could also be "interpreted as bad administration rather than as overt corruption", and severe (narrow) irregularities, where there is clearer evidence that an act of corruption took place.

One caveat of these reports is the possibility to audit resources transferred in the preceding political term, especially when the audit was held at the beginning of the current term. For instance, an audit held in 2005 may contain audits of resources transferred to the municipality in 2004. Therefore, we exclude the first two audits in the 2005 and 2010 mayoral terms from our analysis in Section 3.2, as a large share of the resources audited in these reports will be resources administered by the previous municipal government.[10]

In addition to the two mentioned data sources, we also gather data from the CGU,

[9] Data from FF are available in their replication package. Data from Brollo et al. can be found on Brollo's website. We appreciate their efforts in making their data accessible.

[10] Brollo et al.'s classify corruption by municipality-term instead of by municipality-audit. This is important to keep in mind when comparing Brollo et al.'s data and data encoded by the LLM, as further explained in Appendix B.

provided under the Law on Access to Public Information (LAI). This dataset contains a list of all identified irregularities for each municipality between 2006 and 2015 (lotteries 20 to 40). They are classified as administrative, medium, or serious irregularities. However, even the serious irregularity definition considers a more comprehensive classification of corruption than those from FF and Brollo et al. Beyond the corruption categories considered by them, CGU also codes cases of mismanagement as serious irregularities.[11]

3 Classifying Corruption Audit Reports with LLMs

3.1 The LLM Framework

Previous attempts to classify corruption based on audit reports, while valuable, span various time periods and employ differing manual encoding methods. As a result, there has been no unified classification framework across all 40 lotteries. This limitation is unsurprising given the large number and length of the reports. Over the 13-year period, 2,197 audits generated approximately 185,000 pages of reports. Manually analyzing all these documents would require significant time and multiple individuals, introducing variability due to subjective interpretations of corruption. In this section, we propose an alternative method to systematically interpret these reports.

To construct our corruption measures, we use a Retrieval-Augmented Generation (RAG) process to extract pertinent contextual information from the texts and apply it using OpenAI's GPT-4, a leading Large Language Model (LLM). In standard question-answering (QA) systems without RAG, models rely only on training data. Conversely, RAG enables models to incorporate additional context provided at query time, generating more accurate and contextually relevant responses.

Our framework operates as follows. First, all PDFs are converted to text files and divided into smaller text "chunks," maintaining overlap between chunks to preserve context. Next, each text chunk is transformed into embeddings—numerical representations of text—and stored in a vector database.[12] Text chunks with similar semantic content have similar embedding vectors. When a query is made, the algorithm compares the embedding of the question with stored embeddings, assigning similarity scores. The most similar chunks are selected as context for generating responses.

This entire process is managed using LangChain, an open-source framework for developing LLM applications.[13] Specifically, the process involves the following four

[11] For instance, the lack of creation of the Municipal Commission for the Eradication of Child Labor is considered a serious irregularity. In another example, the absence of mapping/diagnosis of areas of risk and social vulnerability is considered a serious irregularity.

[12] We use Chroma, an open-source database, to store embeddings (https://docs.trychroma.com).

[13] We used version 0.0.349, available as of January 2024. LangChain documentation is available at

steps. First, we input the question. Second, we transform the question into embeddings. Third, we compare the question's embeddings with all embeddings in the vector database. Fourth, we select the two most similar vectors.

Following FF's definitions of corruption, we posed five questions to GPT-4 about each audit report.[14] The first three questions address the monetary value identified in each corruption category: diversion of funds, overinvoicing, and procurement irregularities. The fourth question queries the total number of cases across these categories. Questions were originally posed in Portuguese, with English translations available in Appendix A.1. For responses to questions 1 to 3, we extract corruption values, discarding duplicates to avoid double counting.[15] These values are summed to determine the total corruption amount per report.

We define a binary corruption indicator equal to one if the total corruption amount is positive, and zero otherwise. The corruption case count is directly derived from the fourth question, which asks about the total number of cases. Finally, to compute the share of resources identified as corrupt, we asked an additional question regarding the total federal funds audited. The response serves as the denominator for our primary variable:

$$\text{Share corrupt LLM} = \frac{\text{Diversion of funds} + \text{Overinvoicing} + \text{Procurement irregularities}}{\text{Total funds audited}} \tag{1}$$

To further refine the outputs from the LLM, we conducted manual verifications based on explicit rules detailed in Appendix A.3, including illustrative examples.

3.2 Comparing LLM and Manual Classifications

In this section, we elaborate on our encoding of audit reports and compare it to three manually encoded datasets described in Section 2.2. Table 1 compares the coverage of our LLM-generated dataset to the manually classified data. Our dataset uniquely covers the entire 2003-2015 period, while others have more limited coverage, constraining direct comparisons. Our period encompasses resources from four political terms audited along thirty-five lotteries. In line with FF, we exclude from our sample the pilot lottery, which audited only five municipalities. Additionally, we excluded four lotteries conducted within the first six months of each political term, as documented in the Appendix B. It is noteworthy that the frequency of lotteries was initially higher, with seven occurring in the first year. The frequency was then reduced to just one lottery per year from 2012 until 2015.

https://python.langchain.com/docs/get_started/introduction.

[14] We set GPT-4's "temperature" to zero, ensuring direct responses and avoiding creative interpretations.

[15] A single case might simultaneously qualify as procurement irregularity and overpricing, appearing in multiple answers.

As in FF, we quantify corruption in three ways: the "share of corruption" variable described in Equation 1, the number of irregularities, and a binary variable for if there was any evidence of corruption in the report. The Brollo et al. data lack the number of irregularities found, preventing a direct comparison for this measure. Similarly, the CGU dataset lacks detailed monetary values, preventing a direct comparison for the "share of corruption" variable.

Table 2 displays the correlations between our LLM-generated measures and the corresponding variables in other datasets. The correlation of the "share of corruption" between our LLM dataset and FF's dataset is 0.47. Correlations with Brollo et al.'s narrow and broad definitions are 0.45 and 0.44, respectively, suggesting substantial consistency. These correlations fall just below the range observed among manually coded datasets (0.48 to 0.71). The imperfect correlations among manually coded data underscore the inherent subjectivity and complexity of classifying corruption.

Summary statistics for LLM-encoded variables, presented in Table 3, closely resemble manually encoded datasets. According to LLM encoding, 77% of reports show at least one corruption instance. Restricting to the initial 11 lotteries, this percentage is 70%, between the 79% reported by FF and 67% in Brollo et al.'s broad measure. Over all audits analyzed by Brollo et al., our findings align closely with their broad corruption measure (75% vs. 78%). Similarly, when restricting the time period to that covered by the CGU data, our estimate is similar to the CGU data (82% vs. 78%).

On average, our LLM-generated dataset finds that 2.3%-3.4% of audited resources involve corruption (fund diversion, overpricing, procurement fraud), depending on the period considered. FF and Brollo et al.'s broad measure estimate slightly higher shares (6.3% and 5.2%, respectively), with the narrow measure by Brollo et al. being lower (2.1%). Our LLM-based estimates consistently fall within this range.

The average number of corruption cases per report from 2003 to 2015 in our dataset is 1.21. When compared directly to equivalent audits, this is lower than FF's average (0.96 vs. 1.93) and notably lower than CGU data (1.32 vs. 7.07). The high number of cases in CGU data confirms earlier concerns about overly broad classifications, sometimes including cases of mismanagement as corruption.

4 Reelection Incentives and Corruption

4.1 Empirical Strategy

In this Section, we apply our extended corruption data to reassess a key finding in the literature, which is that mayors facing reelection are less corrupt than those not eligible for reelection. Following Ferraz and Finan (2011), we test whether reelection incentives affect the level of corruption in a municipality using the following OLS re-

gression:

$$Corruption_{ml} = \beta \textit{FirstTerm}_{ml} + \gamma Z_{ml} + State_m + \alpha_l + \varepsilon_{ml} \quad (2)$$

where $Corruption_{ml}$ is a measure of corruption in municipality m as reported in an audit from lottery l. $\textit{FirstTerm}_{ml}$ indicates whether the mayor is in their first term, while Z_{ml} represents a set of controls accounting for the mayor's observable characteristics.[16] The terms $State_m$ and α_l respectively denote state and lottery fixed effects, and ε_{ml} is the error term.

We also estimate a close election regression discontinuity (RD) to account for any unobserved municipal determinants of corruption that may differ between first and second-term mayors. We compare municipalities where incumbent mayors barely won the election, thus serving as second-term mayors in the following term, to municipalities where the incumbent barely lost the election and thus was replaced by a new mayor. This setting provides a quasi-random assignment of municipalities with a first- versus second-term mayor. The main hypothesis supporting the use of the RD is that if an election is competitive enough, who wins it is as good as random.

To estimate the effect of reelection on corruption, we subset the data, including only mayors associated with a vote margin whose absolute value is sufficiently close to zero. The optimal distance to use as bandwidth is defined according to the minimum squared error (MSE) criteria (Calonico et al., 2014). The following local linear regression is then used:

$$Corruption_{ml} = \tau \textit{FirstTerm}_{ml} + \lambda_0 MV_{ml} + \lambda_1 \textit{FirstTerm}_{ml} MV_{ml} + \gamma Z_{ml} + StateFE_m + \alpha_l + \varepsilon_{ml} \quad (3)$$

where $Corruption_{ml}$ is the corruption outcome, $\textit{FirstTerm}_{ml}$ is the indicator for first-term mayors, and Z_{ml} is a vector of mayors characteristics, as before. The terms $State_m$ and α_l, as before, respectively denotes state and lottery fixed effects. The term MV_{ml} represents the candidate's margin of victory. It is specified as the difference between the vote share of incumbent mayor minus the vote share of the challenger receiving the largest number of votes. This measure is therefore less than zero in municipalities where the incumbent was not reelected and a new mayor was elected, and greater than zero otherwise. We weigh the regression using a triangular kernel.

4.2 Results

The main finding from FF's paper is that mayors with reelection incentives misappropriate 27 percent fewer resources than those without reelection incentives. We investigate if this result holds when using our extended LLM classified data. Table 4 and Figure 1 present the estimates from the OLS regression specified in Equation 2, span-

[16] Mayor's characteristics are age, gender, education, and party afilliation.

ning the complete audits period from 2003 to 2015. All estimates include controls for mayoral characteristics, such as education level, gender, and age, as well as indicators for lotteries and state, accounting for any state-specific or lottery-specific unobservables that might have affected corruption.

Table 4 shows the first set of results using LLM data. Our preferred specification is presented in the even columns. Column 2 suggests that municipalities where mayors are eligible for reelection exhibit a 4.7 percentage point decrease in the likelihood of having a case of corruption detected when compared to municipalities with mayors in their second term. Surprisingly, this effect is smaller and statistically insignificant for the number of corruption cases and the share of corruption, the main outcome variable presented by FF.

In Figure 1 we visualize a comparison of the estimated effect size using all four encodings of the data. We conducted the regression described in Equation 2 on variables that have been divided by their own mean value, so that the coefficients can be interpreted as a percentage change of the variable mean. This normalization allows for a direct comparison of the magnitudes of the coefficients. The figure uses all available data from each dataset and is therefore comparing estimates generated from data spanning different time periods for each dataset.

The figure shows that the coefficients from our LLM-generated data on the number of corruption cases and the indicator for if any corruption was found are similar in direction and magnitude to the coefficients in FF original paper.[17] However, the coefficient of "Share corrupt", is considerably smaller and statistically significantly different from the estimate using FF's data. In addition, the coefficients from the CGU's and Brollo et al.'s data deviate significantly from those estimated using FF data, with the exception of the variable "Any narrow corruption" which is the only variable for which we find a negative and statistically significant effect in these two encodings of the data.

As a robustness test, we estimate the same regressions restricting the samples to include only reelected mayors. As pointed out by FF, if elections serve to select the most able politicians, and ability and corruption are positively correlated, we need to compare second-term mayors with the set of first-term mayors who are reelected in the subsequent election — those presumed to have greater political skills. The normalized coefficients are displayed in Figure C.2. Overall, the coefficients do not exhibit substantial differences from those presented in Figure 1.

To understand if the difference in results are due to differences in the data encoding or if they reflect differences in the time periods covered, we then narrow our analysis to audits conducted between 2003 and 2004. For these audits we have data from FF, Brollo et al. as well as our own LLM encoding. Figure 2 presents the normalized coefficients restricted to that period and shows that we successfully replicate the primary

[17] When we use FF's data we successfully replicate the findings of their paper.

finding presented in FF when using the same audit reports. While certain variables lack statistical significance, all coefficients point in the expected direction, indicating that first-term mayors were indeed less corrupt during the 2001-2004 term. Thus, the differences in the results for different encoding of the data presented in Figure 1 can to a large extent be explained by differences in the estimated effect across different time periods.

Figure 3 presents evidence on how the magnitude of the difference between first- and second-term mayors change over time. During the 2001-2004 term, two of the three LLM encoded variables indicate that first-term mayors were associated with less corruption than second-term mayors. This effect is also apparent in the 2013 term. However, we observe no difference between first and second-term mayors from 2005 to 2012. A similar pattern is observed for Brollo et al.'s corruption measures, as shown in Appendix Figure C.1. In Section 5, we discuss and test alternative explanations that could account for the change over time.

Finally, we assess the effects of reelection incentives using elections in which the incumbents won or lost by a narrow margin. The RD outlined in Equation 3 provides quasi-random assignment of first-term and second-term mayors across these competitive elections, eliminating potential confounds. The sample is conditioned on the incumbents who ran for reelection in each election. Table 5 presents the point estimates for the LLM's measures spanning all terms. All columns are estimated using the MSE optimal bandwidth (Calonico et al., 2014). In the Appendix C, we also provide robustness using half and double the optimal bandwidth. All specifications include the controls used previously: mayor's characteristics, party affiliation, state, and lottery fixed effects.

The coefficients estimated using the RD specification are all negative but small and statistically indistinguishable from zero. However, results are also statistically indistinguishable from the main results presented in Table 4. In Figure 4, we depict the results graphically. Similar results, using the data encodings by FF and Brollo et al., are shown in Appendix Table C.5 and Appendix Figures C.3 and C.4.

5 Mechanisms for Changes in the Importance of Reelection Incentives

What factors might explain the large negative effects between 2003-2004 and 2013-2015, and the absence of an effect during the 2005-2008 and 2009-2012 mayoral terms, as shown in Figure 3? We test and discuss four potential explanations that could be acting to change the results in the two terms from 2005 to 2012 and later between 2013-2015.

5.1 Differences Across Cohorts

One hypothesis follows from differences across cohorts. The Random Audits Program started in 2003, thus auditing resources from the political term that began in 2001.[18] Coincidentally, this cohort was the first generation of second-term mayors, as the 2000 election was the first to allow reelection at the municipal level. Naturally, barring any irregularities with the electoral court, all mayors who held office from 1997 to 2000 were eligible to run again. This means that many of the newly elected mayors in 2000 had faced, and won against, an incumbent politician running for reelection. If winning in this environment of increased political competition correlates with the quality of elected officials, first-term mayors elected in 2000 may have been of higher quality and potentially less corrupt than those who were in their second-term and had been initially been elected without facing an incumbent eligible for reelection. This could explain the larger difference between the first and second term mayors during the 2001-2004 period, as in later periods there were no such difference between first and second term mayors. To test this hypothesis, we estimate the following equation:

$$Corruption_{ml} = \beta FirstTerm_{ml} + \theta(FirstTerm_{ml} \times Incumbent\ Eligible_{ml}) + \gamma Z_{ml} + State_m + \alpha_l + \varepsilon_{ml} \quad (4)$$

The term of interest lies in the interaction between $FirstTerm_{ml}$ and $Incumbent\ Eligible_{ml}$. As in previous equations, $FirstTerm_{ml}$ is an indicator variable for whether the mayor in their first-term in municipality m during lottery l. The term $Incumbent\ Eligible_{ml}$ is an indicator variable for whether the incumbent was eligible to run again for the office when the current mayor was elected. In the case of the 2000 election, by definition, all incumbent mayors were eligible because they were all serving their first term. In subsequent elections, only those who were in their first term were eligible for reelection.

We report results in Table 6, with a focus on the dependent variable "Any corruption (LLM)", which exhibited significant results in Table 4. In Column 1 we present the results of Equation 4. We find that the interaction coefficient is positive, but small in magnitude, and statistically insignificant. This suggests that differences in political competition during the election of first- and second-term mayors in the 2000 election cannot explain the large difference between these mayors in terms of corruption in the 2003-2004 audits.

5.2 Candidate Screening

A second hypothesis is that the emergence and rapid growth of a new political party led to an influx of new mayoral candidates, potentially reducing the effectiveness of candidate screening processes. To explore this possibility, we examine the impact of

[18] There are few cases related to the 1997 term, according to Ferraz and Finan (2011).

the rise of the Workers' Party (PT) in the 2004 and 2008 elections. The PT's victory in the 2002 national election, where their presidential candidate won, significantly increased the party's prominence in local politics. This surge in popularity doubled the number of PT-affiliated mayors, as shown in Figure C.5. To satisfy local demand for PT candidates, the party may have hastily recruited new candidates, who could have been more prone to corruption and therefore reduced the difference in corruption between first- and second-term mayors during this period.[19] To formally investigate this hypothesis, we estimate the following equation:

$$Corruption_{ml} = \beta \textit{FirstTerm}_{ml} + \theta(\textit{FirstTerm}_{ml} \times \textit{Workers' Party}_{ml}) + \gamma Z_{ml} + State_m + \alpha_l + \varepsilon_{ml} \quad (5)$$

The main coefficient of interest is θ, which measures the difference in the effect on corruption of being a first-term mayor from the Workers' Party (PT) and of other parties. The term *Workers' Party*$_{ml}$ is an indicator variable, assigned the value one if the mayor at municipality m was a PT member during the lottery l. The explanation above would imply a positive θ.

The results are presented in Table 6. We also report the results of a joint test that considers both interactions. We start by studying audits from the full time period available between 2003 and 2015. In Columns 2 we estimate a negative and statistically insignificant coefficient of about -0.028. The same is true in Column 3 when we separately estimate a coefficient for the incumbent being eligible to run for office.

Considering that the party's growth primarily occurred during the 2004 and 2008 elections, in Columns 4 and 5 we restrict the sample to the 2005-2012 period. In this case, the coefficients for Workers' Party's first-term mayors are positive and of considerable magnitude, although still statistically insignificant. In the Appendix, we present similar regressions results using Brollo et al.'s "Any narrow corruption" measure (Table C.6). Coefficients for Workers' Party first-term mayors are substantially larger and statistically significant, suggesting that changes in party composition among first-term mayors may partially explain the lack of effect between 2005-2012.

5.3 Legal Penalties as a Deterrent

Another possible explanation is that a rising frequency of legal actions against public officials has gradually diminished the importance of reelection incentives as a deterrent to corruption. As illustrated in Figure C.9, convictions of current and former

[19] Around the same period, in 2005, the *Mensalão* scandal brought significant negative attention to PT, uncovering a widespread corruption scheme involving bribes paid to congressmen in exchange for political support. This scandal severely damaged the party's reputation and resulted in the conviction of multiple officials.

mayors increased steadily from 2000 to 2013 before declining thereafter.[20] Since both first- and second-term mayors face these legal consequences, the distinction in corrupt behavior between the two groups may have decreased over time. Supporting this interpretation, Figure 5 shows that the probability of prosecution within five years after an election was higher for mayors elected in 2004 and 2008 compared to those elected in 2000. Notably, this probability significantly declined for mayors elected in 2012, indicating a decrease in enforcement intensity. This could potentially explain why reelection incentives reemerged as an important determinant of corruption in this period.

The importance of potential legal action as a deterrent for corruption is consistent with the findings in Avis et al. (2018). They argue that the effects of audits are mostly due to increasing perceived nonelectoral costs of engaging in corruption.

5.4 Reelection Incentives in National Politics

We also examine whether any changes have occurred that altered incentives for second-term mayors or their parties. For example, an increased tendency for second-term mayors to pursue national-level politics could heighten their incentives to maintain a clean reputation and thus reduce corruption. Tables C.7 and C.8 show the percentage of second-term mayors running for or holding higher political office over time. The term "2 years later" refers to the first national or state election after the municipal election in which the mayor was reelected, meaning that those elected to higher positions would not complete their mayoral term. Conversely, "6 years later" refers to the second subsequent national or state election.[21]

The data indicate that 1.5% of second-term mayors elected in 2000 ran for higher office within two years, with only 0.7% successfully elected. Additionally, 12.7% ran for higher office after completing their mayoral terms, with 4.8% elected. However, we observe no significant increase in second-term mayors pursuing or attaining higher office either two or six years later. Indeed, the percentage of second-term mayors running in national elections remains stable from 2000 to 2004 and declines in 2008, while the percentage successfully elected to higher positions decreased in 2004, 2008, and 2008.

To further investigate this, we explore whether reelection incentives from the perspective of political parties have changed by examining party turnover in subsequent elections. Table C.9 shows that the proportion of second-term mayors whose parties retained power in the following election remained unchanged between 2000 and 2004 and decreased in 2008.

[20]The peaks in convictions observed in 2005, 2009, and 2013 correspond to years immediately following the conclusion of mayoral terms, periods when prosecution likelihood typically increases. Figure C.10, which specifically highlights convictions resulting in electoral penalties, exhibits a similar trend.

[21]Municipal and national elections in Brazil occur every four years but are staggered by two years. State elections align with national elections.

6 Conclusion

In this paper, we introduce a novel approach to measuring corruption by leveraging a Large Language Model (LLM) to analyze random audit reports from Brazilian municipalities audited between 2003 and 2015. Our methodology is automated, cost-effective, scalable, and adaptable to other contexts. When comparing our data to existing manually encoded datasets, we find similar, albeit low, correlations between the main variables. We show that manually encoded data have correlations significantly below one, indicating a degree of subjective evaluation in interpreting the audit reports and highlighting the difficulty inherent to classifying corruption from text.

We revisit the well-known evidence by FF, that mayors facing reelection incentives misappropriate fewer resources. Our results confirm this finding, but we estimate smaller and more precisely measured effects. Additionally, we provide evidence using alternative datasets that consistently support a diminished effect size across both OLS and RD estimations.

Our analysis highlights notable heterogeneity over time. Initially, our results align closely with FF's findings, showing clear evidence that reelection incentives curb corruption. However, this effect declines substantially during the 2005-2012 period, before reemerging strongly in the 2013 term. We investigate several possible explanations for these fluctuations, such as cohort differences due to increased political competition in the 2000 election, the rise of the Workers' Party, increased legal actions against corrupt officials, and changing incentives related to second-term mayors' political ambitions. Among these, we find some support for the explanations that new party dynamics, specifically the rise of the PT party, and rising threat of legal penalties for corruption contributed to changes in the importance of reelection incentives. Conversely, we find limited empirical support for other hypotheses, including shifts in political competition in 2000 and second-term mayors' ambitions for higher office.

This paper contributes to both the measurement of corruption and the understanding the political determinants of corruption. Our findings underscore the complexity of electoral incentives, highlighting how the interplay of political, legal, and institutional factors evolves over time, shaping the effectiveness of different mechanisms to reduce corruption. Future research could further refine the methodologies used for processing text data and given the rapid rate of improvement in LLM technology, we expect future iterations of our algorithm to generate even more accurate results. This could allow us to delve deeper into the evolving political and institutional determinants of corruption within local governments.

References

Alt, J., Bueno de Mesquita, E., and Rose, S. (2011). Disentangling accountability and competence in elections: evidence from us term limits. *The Journal of Politics*, 73(1):171–186.

Ash, E., Galletta, S., and Giommoni, T. (2025). A machine learning approach to analyze and support anti-corruption policy. *American Economic Journal: Economic Policy*.

Avis, E., Ferraz, C., and Finan, F. (2018). Do government audits reduce corruption? estimating the impacts of exposing corrupt politicians. *Journal of Political Economy*, 126(5):1912–1964.

Axbard, S. (2024). Convicting Corrupt Officials: Evidence from Randomly Assigned Cases. *Review of Economic Studies*.

Banerjee, A., Duflo, E., Imbert, C., Mathew, S., and Pande, R. (2020). E-governance, accountability, and leakage in public programs: Experimental evidence from a financial management reform in india. *American Economic Journal: Applied Economics*, 12(4):39–72.

Besley, T. J. and Case, A. (1995). Does Electoral Accountability Affect Economic Policy Choices? *Quarterly Journal of Economics*, 110(3):769–798.

Besley, T. J. and Case, A. (2003). Political institutions and policy choices: evidence from the united states. *Journal of Economic Literature*, 41(1):7–73.

Björkman, M. and Svensson, J. (2010). When is community-based monitoring effective? evidence from a randomized experiment in primary health in uganda. *Journal of the European Economic Association*, 8(2-3):571–581.

Bobonis, G. J., Fuertes, L. R. C., and Schwabe, R. (2016). Monitoring Corruptible Politicians. *American Economic Review*, 106(8):2371–2405.

Brollo, F., Nannicini, T., Perotti, R., and Tabellini, G. (2013). The Political Resource Curse. *American Economic Review*, 103(5):1759–1796.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

Colonnelli, E., Gallego, J., and Prem, M. (2022). What predicts corruption? *A Modern Guide to the Economics of Crime*, page 345.

Colonnelli, E. and Prem, M. (2022). Corruption and Firms. *Review of Economic Studies*, 89(2):695–732.

Cuneo, M., Leder-Luis, J., and Vannutelli, S. (2023). Government audits. Working Paper 30975, National Bureau of Economic Research.

Farenas, F. (2024). Figuring out the ideal chunk size. https://medium.com/@farenas1/fabian-7d1f90ac4cb4. Accessed: Mai 2, 2024.

Ferraz, C. and Finan, F. (2008). Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes. *Quarterly Journal of Economics*, 123(2):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.

Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., and Liang, P. (2024). Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics*, 12:157–173.

Olken, B. A. (2007). Monitoring corruption: evidence from a field experiment in indonesia. *Journal of political Economy*, 115(2):200–249.

Olken, B. A. (2009). Corruption perceptions vs. corruption reality. *Journal of Public economics*, 93(7-8):950–964.

Olken, B. A. and Pande, R. (2012). Corruption in developing countries. *Annu. Rev. Econ.*, 4(1):479–509.

Rose-Ackerman, S. (1998). Corruption and development. *Annual World Bank Conference on Development Economics 1997*, pages 35–57.

Svensson, J. (2005). Eight questions about corruption. *Journal of economic perspectives*, 19(3):19–42.

Van Rijckeghem, C. and Weder, B. (2001). Bureaucratic corruption and the rate of temptation: do wages in the civil service affect corruption, and by how much? *Journal of development economics*, 65(2):307–331.

7 Tables and Figures

Table 1: Summary Data

	FF	Brollo et al.	CGU	LLM
Lotteries 2-11 (2003-2004)	✓	✓		✓
Lotteries 12-20 (2004-2006)		✓		✓
Lotteries 20-29 (2006-2009)		✓	✓	✓
Lotteries 30-40 (2009-2015)			✓	✓
Corruption amount as a share of audited resources	✓	✓		✓
Number of irregularities related to corruption	✓		✓	✓
Dummy whether any irregularity was found	✓	✓	✓	✓

Note: This table provides an overview of the coverage of various sources of corruption data utilized in this paper. Additionally, the table displays the availability of each measure of corruption. For each range of available lotteries and variables, we mark them with a check. Variables from Brollo et al. (2013) are divided into broad and narrow irregularities as explained in Section 2.2.

Table 2: Data Correlations

	Share corrupt (FF)	Share broad (Brollo et. al.)	Share narrow (Brollo et. al.)	Share corrupt (LLM)	Any corruption (FF)	Any broad (Brollo et. al.)	Any narrow (Brollo et. al.)	Any corruption (LLM)	Any serious irregularity (CGU)
Share corrupt (FF)	1.00	0.71	0.48	0.47	.	.	.	.	.
Share broad (Brollo et. al.)	0.71	1.00	0.65	0.44	.	.	.	.	.
Share narrow (Brollo et. al.)	0.48	0.57	1.00	0.45	.	.	.	.	.
Share corrupt (LLM)	0.47	0.44	0.45	1.00	.	.	.	.	.
Any corruption (FF)	.	.	.	.	1.00	0.59	0.31	0.35	.
Any broad (Brollo et. al.)	.	.	.	.	0.59	1.00	0.57	0.25	0.31
Any narrow (Brollo et. al.)	.	.	.	.	0.31	0.57	1.00	0.20	0.29
Any corruption (LLM)	.	.	.	.	0.35	0.25	0.20	1.00	0.07
Any serious irregularity (CGU)	.	.	.	.	.	0.31	0.29	0.07	1.00

Note: This tables presents the pair-wise correlations between the LLM variables and manually coded variables from Ferraz and Finan (2011), Brollo et al. (2013), and CGU. Additionally, we present the pair-wise correlations between all variables used as reference. These variables span different periods, as illustrated in Table 1. For that reason, we can not calculate the correlation between "Any serious irregularity" and "Any corruption (FF)".

Table 3: Summary Statistics

	2003-2015	2003-2004				2003-2009			2005-2015	
	LLM	LLM	FF	Broad	Narrow	LLM	Broad	Narrow	LLM	CGU
Any corruption	0.765	0.694	0.786	0.675	0.389	0.744	0.783	0.470	0.814	0.783
	(0.424)	(0.461)	(0.411)	(0.469)	(0.488)	(0.436)	(0.412)	(0.499)	(0.389)	(0.412)
Observations	2,128	490	476	489	489	1,537	1,401	1,401	1,168	1,112
Share corrupt	0.027	0.034	0.063	0.059	0.026	0.029	0.052	0.021	0.023	
	(0.061)	(0.075)	(0.102)	(0.116)	(0.069)	(0.065)	(0.102)	(0.064)	(0.050)	
Observations	2,127	490	476	483	483	1,536	1,336	1,337	1,167	
Number of cases	1.209	0.957	1.931			1.283			1.318	7.068
	(2.258)	(1.998)	(1.707)			(2.395)			(2.410)	(9.377)
Observations	1,823	444	476			1,309			971	1,112

Note: This table presents the mean and the standard deviation of all available variables. "Broad" and "Narrow" refer to definitions from Brollo et al. The first column span all years, including all available audits (lotteries 2 to 40). The other columns are divided into different time periods to align with manually coded versions. From 2003 to 2004, we report the statistics from audits conducted within the first 11 lotteries, comparing these to the findings of both FF and Brollo et al. From 2003 to 2009, we report the statistics from audits conducted within the first 29 lotteries and compare it to Brollo et al.'s findings. Finally, from 2005 to 2015 we report the statistics from audits conducted from lottery 20 onward, comparing these to CGU variables. The reduced number of observations in the "Number of cases" from the LLM is due to missing data—cases where the algorithm does not provide an exact number. Lotteries held in the first six months of each term were excluded because most of the audited resources refers to the preceding political term. For Brollo et al. and CGU variables, we restrict the analysis to observations within the same term, excluding corruption associated to resources transferred in previous political terms. Additionally, the data is restricted to municipalities where we have information on whether the mayor is serving their first or second term.

Table 4: The Effect of Reelection Incentives on Corruption (2003-2015)

	Any corruption (LLM)		Share corrupt (LLM)		Number of cases (LLM)	
	(1)	(2)	(3)	(4)	(5)	(6)
Mayor in first term	-0.0338*	-0.0468**	-0.0022	-0.0015	-0.0277	-0.1514
	(0.0200)	(0.0202)	(0.0030)	(0.0031)	(0.1181)	(0.1152)
Mayor Characteristics	No	Yes	No	Yes	No	Yes
Lottery Dummies	No	Yes	No	Yes	No	Yes
State Dummies	No	Yes	No	Yes	No	Yes
Party Dummies	No	Yes	No	Yes	No	Yes
Mean	0.7687	0.7682	0.7691	0.7686	0.7558	0.7556
Observations	1,894	1,885	1,893	1,884	1,626	1,620

Note: This table presents the impact of reelection incentives on three corruption metrics: the probability of finding a corruption case, the proportion of audited resources associated with corruption, and the number of detected corruption cases. Each column displays the results from the OLS regression presented in Equation 2, where the respective corruption measure is regressed on an indicator variable denoting whether the mayor is in their first term. The even numbered columns include controls to Mayor's characteristics and party affiliation, as well as state and lottery intercepts. Mayor's characteristics are age, gender and education. This estimate includes municipalities audited from lotteries 2 to 40, with the exception of lotteries 15, 16, 28, and 38 —excluded due to their occurrence within the first six months of a political term (see Appendix B for further explanation). The period from 2003 to 2015 spans the four political terms with audited resources. The last term, however, does not consider the final year (2016), as the last available lottery (40) was conducted in 2015. Heteroscedasticity robust standard errors are displayed in parenthesis. P-values: * 0.10 ** 0.05 *** 0.01

Table 5: The Impact of Reelection Incentives on Corruption, RD

	Share corrupt	Any corruption	Number of cases
	(1)	(2)	(3)
Mayor in first term	-0.004	-0.074	-0.085
	(0.012)	(0.076)	(0.335)
Robust 90% CI	[-.024 ; .036]	[-.245 ; .195]	[-1.073 ; .888]
Kernel Type	Triangular	Triangular	Triangular
BW Type	CCT	CCT	CCT
BW	0.162	0.165	0.224
Observations	1025	1026	877

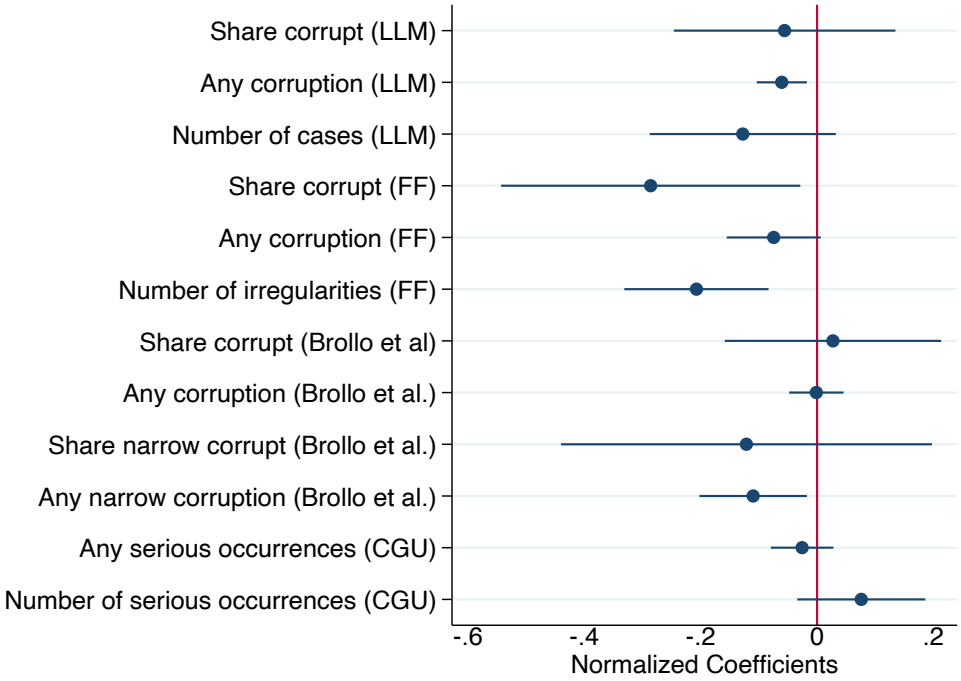
Note: This table presents the coefficients from the RD regression specified in Equation 3. We evaluate the impact of reelection incentives on three corruption metrics: the probability of finding a corruption case, the proportion of audited resources associated with corruption, and the number of detected corruption cases. All columns include controls to Mayor's characteristics and party affiliation, as well as state and lottery intercepts. Mayor's characteristics are age, gender and education. We include municipalities audited from lotteries 2 to 40 if the mayor ran for reelection. As in the previous cases, we excluded lotteries 15, 16, 28, and 38 due to their occurrence within the first six months of a political term (see Appendix B for further explanation). The BW Type indicates that the MSE optimal bandwidth was used (CCT). The BW parameter reports the respective bandwidth for each regression. Heteroscedasticity robust standard errors are displayed in parenthesis. P-values: * 0.10 ** 0.05 *** 0.01

Table 6: Alternative Explanations

	Any corruption (LLM)				
	2003-2015	2003-2015	2003-2015	2005-2012	2005-2012
	(1)	(2)	(3)	(4)	(5)
Mayor in first term	-0.0536*	-0.0471**	-0.0518*	-0.0165	-0.0242
	(0.0276)	(0.0208)	(0.0283)	(0.0273)	(0.0319)
First x Incumbent eligible	0.0090		0.0088		0.0126
	(0.0270)		(0.0271)		(0.0297)
First x Workers' Party		-0.0276	-0.0275	0.0734	0.0740
		(0.0807)	(0.0809)	(0.0902)	(0.0901)
Mayor Characteristics	Yes	Yes	Yes	Yes	Yes
Lottery Dummies	Yes	Yes	Yes	Yes	Yes
State Dummies	Yes	Yes	Yes	Yes	Yes
Party Dummies	Yes	Yes	Yes	Yes	Yes
Mean	0.7679	0.7687	0.7679	0.8086	0.8082
Observations	1,883	1,894	1,883	1,118	1,116

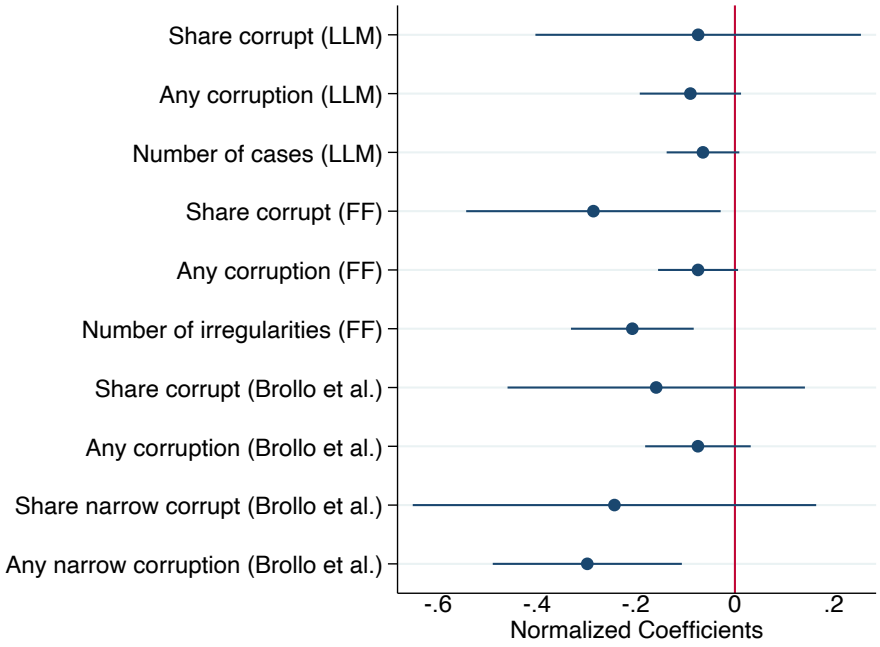
Note: This table presents the coefficients from OLS regressions specified in Equations 2, 4, and 5. All columns include controls to Mayor's characteristics and party affiliation, as well as state and lottery intercepts. Mayor's characteristics are age, gender and education. Columns 1 and 2 are derived from Equations 4 and 5, respectively, while Column 3 includes both interactions together. These regressions consider all terms, except for the final year (2016), as the last available lottery (40) was conducted in 2015. Finally, in Columns 4 and 5, we estimate Equation 5 and the combination of 4 and 5, respectively, only for the middle terms. Heteroscedasticity robust standard errors are displayed in parenthesis. P-values: * 0.10 ** 0.05 *** 0.01

Figure 1: Overall Effects of Reelection Incentives on Corruption (All Available Years)



Note: This figure depicts the coefficients from the OLS regression outlined in Equation 2. All coefficients are normalized by dividing it by the outcome variable mean. All standard errors are heteroscedasticity-robust. We estimate the impact of reelection incentives on corruption using all available measures obtained from LLM, FF and Brollo et al.'s dataset. All regressions include controls to mayor's characteristics and party affiliation, as well as state and lottery intercepts. Mayor's characteristics are age, gender and education. For each variable, we plot its source in parentheses. The regressions include different time coverage. Estimates using LLM variables includes data from 2003 to 2015 (lotteries 2 to 40). We exclude lotteries 15, 16, 28, and 38 due to their occurrence within the first six months of a political term (see Appendix B for further explanation). Brollo et al.'s variables includes data from 2003 to 2009 (lotteries 2-29), while CGU variables includes data from 2005 to 2015 (lotteries 20 to 40). For both Brollo et al.'s and CGU data, the analysis is restricted to observations within the same term. Confidence intervals are displayed at the 90% level.

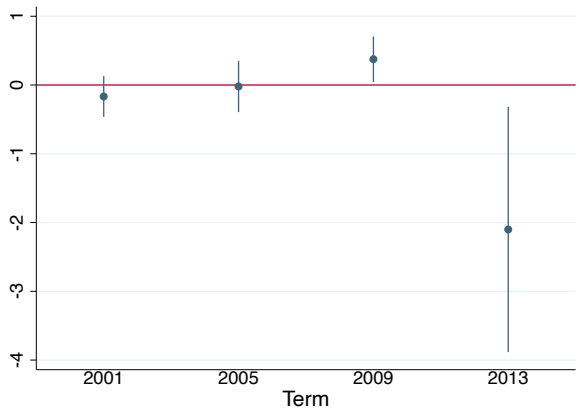
Figure 2: The Effect of Reelection Incentives on Corruption (2003-2004)



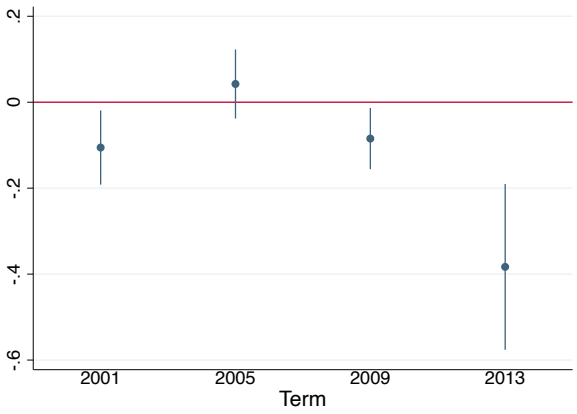
Note: This figure depicts the coefficients from the OLS regression outlined in Equation 2. All coefficients are normalized by dividing it by the outcome variable mean. All standard errors are heteroscedasticity-robust. We estimate the impact of reelection incentives on corruption using all available measures obtained from LLM, FF and Brollo et al.'s dataset. The regressions are restricted to lotteries 2 to 11, thus including only years 2003 and 2004, in the first term in which reelection was allowed at the municipal level. All regressions include controls to Mayor's characteristics and party affiliation, as well as state and lottery intercepts. Mayor's characteristics are age, gender and education. For each variable, we plot its source in parentheses. The CGU variables are not included in this figure due to the absence of data prior to lottery 20 (See Table 1 for detailed information on data coverage).Confidence intervals are displayed at the 90% level.

Figure 3: The Effects of Reelection Incentives on Corruption Over Time (LLM)

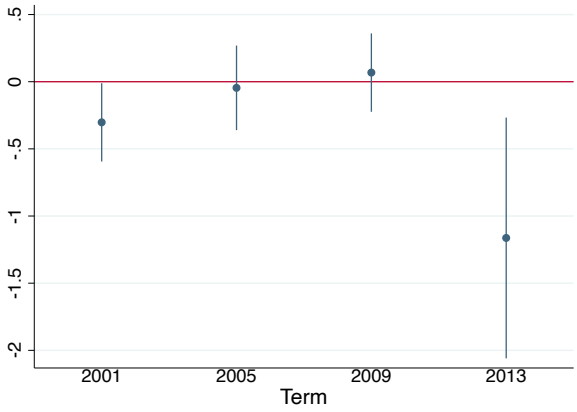
(a) Share corruption



(b) Any corruption



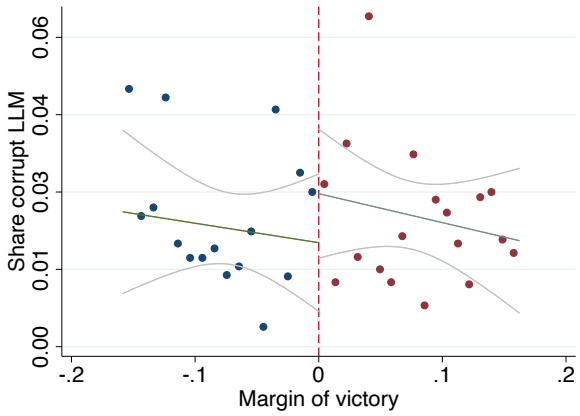
(c) Number of cases



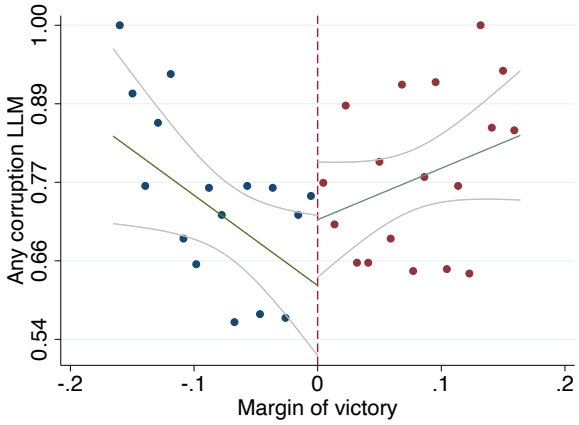
Note: This figure presents the coefficients from the OLS regression specified in Equation 2. All regressions include controls to Mayor's characteristics and party affiliation, as well as state and lottery intercepts. Mayor's characteristics are age, gender and education. The results are broken down by term. Estimates includes data from 2003 to 2015 (lotteries 2 to 40). We exclude lotteries 15, 16, 28, and 38 due to their occurrence within the first six months of a political term (see Appendix B for further explanation). Each term spans a four-year period. The last term, however, does not consider the final year (2016), as the last available lottery (40) was conducted in 2015. Confidence intervals are displayed at the 90% level.

Figure 4: The Effects of Reelection Incentives on Corruption (LLM), RD

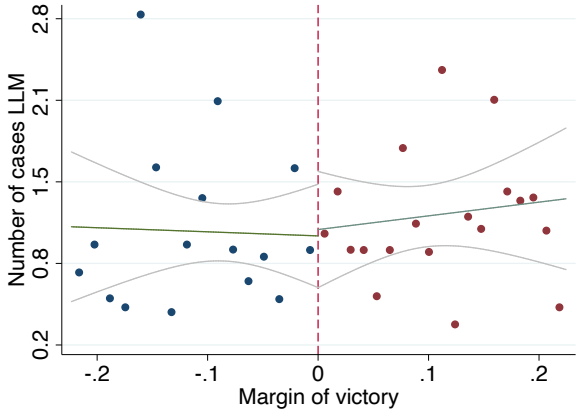
(a) Share corruption



(b) Any corruption

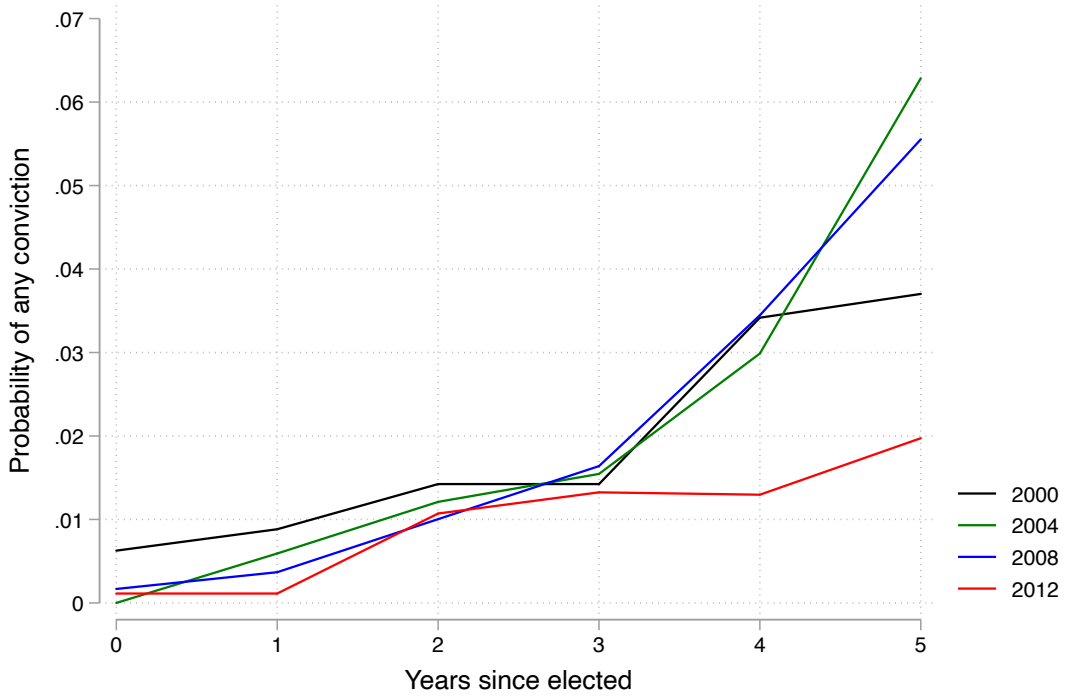


(c) Number of cases



Note: The figure shows the proportion of audited resources associated with corruption (Panel 4a), the indicator for detected corruption (Panel 4b) and the number of detected corruption cases (Panel 4c) by the margin of victory for incumbents who ran for reelection. The gray lines denote the confidence intervals for fitted lines at the 90% level. All regressions use the optimal bandwidth according to the minimum squared error (MSE) criteria (Calonico et al., 2014). We restrict the observations, such that only mayors associated with a vote margin within the interval of the optimal bandwidths are considered.

Figure 5: Probability of Any Conviction



Note: This figure presents the percentage of mayors who have at least one conviction for cases initiated between one and five years after taking office. We limit the analysis to a five-year period to allow for a more accurate comparison, as there is typically a delay in registering cases in the *Cadastro Nacional de Justiça* system (see Table C.10). We also exclude cases that took seven or more years to resolve. Each line represents a political term. The variable "Years Since Elected" denotes the difference between the filling year and the election year. For mayors elected more than once, we use the first election. We exclude cases where an individual was convicted prior to taking office as mayor, as well as cases where the conviction year precedes the filing year, as this is inconsistent with the standard process and likely indicates an error.

A Data Construction

A.1 Corruption Definition and LLM Queries

Regarding the definition of corruption, both Ferraz and Finan (2011) and Brollo et al. (2013) consider corruption to be cases in which there is diversion of funds, over-invoicing of goods and services, or illegal procurement practices. Specifically, diversion of resources may be any irregularity involving the embezzlement of public funds, such as resources that simply "disappear" from municipal bank accounts or incomplete service (unfinished construction, for example) and goods that were supposedly paid for but not delivered. In turn, over-invoicing are classified when there is evidence that goods and services were purchased at a value above market price. Finally, irregularities related to procurement involve any manipulation of the procurement process, simulation of the call for bids, use of fake receipts, and contracts being awarded to a friendly/politically connected firm or non-existing firms.

Following these definitions, we asked the LLM three different questions to identify the value of corruption in each category (questions 1-3). We also asked about the number of cases across all categories — question 4. Additionally, we asked a fifth question to determine the total amount of resources audited to calculate the share of corruption in each audit. All questions were asked in Portuguese, and their English translations are provided below:

1. Does the report mention cases of diversion of funds? If yes, what are the amounts diverted? End the response with: "The total diverted was" followed by the corresponding value.

2. Does the report mention cases of overpricing or excessive billing? If yes, what are the amounts? End the response with: "The total overpriced was" followed by the corresponding value.

3. Does the report mention cases of fraud or serious irregularities in procurement processes? If yes, what is the value of the fraud? End the response with: "The total fraud was" followed by the corresponding value.

4. How many cases of diversion of funds, overpricing, or fraud in the procurement process are mentioned in the report? End the response with "Total cases:" followed by the corresponding value.

5. What is the total value, in R$, of the audited resources? End the response with 'The total audited was' followed by the corresponding value.

As the information on the total audited value is typically presented within the initial pages and does not require much interpretation, we employ a slightly different

algorithm than the one detailed in Section 3.1. In order to extract the values, we simply transform the pdfs into text files and split only the first eight pages into smaller chunks. Then, we input these chunks directly into the GPT-4 prompt and ask the question.

A.2 Challenges in Using LLMs

Although LLMs provide a prominent framework to transform text into data, there are still some challenges with building a question-answering (QA) system, especially over large documents. The first obstacle lies in the token limits imposed on LLMs, which constrain the amount of context that can be provided. For instance, many documents exceed the capacity of 8,000 token contexts offered by GPT models. As a result, the standard practice consists of splitting the document into chunks, calculating a similarity score between those chunks and the query using embeddings, and then use the highest-scoring chunks as context for the query. This is the RAG process described in the previous section.

Another significant challenge arises from the fact that documents, such as PDFs, are naturally structured with different pages, tables, sections, and text indentation.[22] Therefore, there is great difficulty in creating robust prompts and chunking strategies that responds well to the variability among documents.

More specifically, building a RAG system involves determining the ideal chunk size for the documents processed by the retriever. The determination of the ideal chunk size involves considering various factors, including the characteristics of the data, the limitations of the retriever model, and the computational resources available (Farenas, 2024). In addition to size, splitting the document into chunks entails additional decisions. For instance, determining the degree of overlap between them and selecting the number of top scoring chunks to be considered by the LLM.

The study by Liu et al. (2024) offers insights on how well LLMs use longer context. The research evaluated a range of open and closed-source language models, including OpenAI's GPT-3.5-Turbo, across two distinct tasks: multi-document question answering and key-value retrieval. They observe that optimal performance is often achieved when relevant information is located at the beginning or end of the input context. However, performance significantly degrades when models need to access relevant information embedded within the middle of long contexts. This suggests that current language models struggle to leverage information within extended input contexts.

[22] This is very clear in the case of audits reports. Over the years and across municipalities, there is a lot of variation in how the reports are structured.

A.3 Manual Verifications Based on LLM Responses

To enhance the quality of our data, we conducted several manual verifications on Share corrupt $\text{llm}_{m,l}$. We created four rules to flag the values we should check carefully:

1. The denominator falling below the 1st percentile and above the 99th percentile of the distribution.
2. The fraction falling above the 99th percentile of the distribution.
3. The fraction is equal to zero
4. The fraction is not equal to zero but the total corruption value is less than R$500.00

Overall, instances where Question 5 did not accurately capture the total audited value mainly occurred due to formatting issues. For instance, in the case of Peritiba, SC, in lottery 5, there was a table indicated the total audited resources as R$ 1,271,260.02, whereas our algorithm only captured R$ 1.27. Another example is observed in São João das Missões, MG, in lottery 2, where the returned value was 0 because this information appeared beyond the defined 8-page interval in the algorithm, within a figure on page 12. Additionally, there were cases where the algorithm failed to return the total value because it was not explicitly stated in the report. However, we managed to obtain it by summing up the audited values for each program. In total, we identified 44 observations where the values obtained via LLM fell below the 1st percentile or above the 99th percentile. We manually inspected all these observations, and 25 cases required corrections.[23]

Regarding the remaining rules, within 2,197 observations, we have 22 cases flagged by rule number 2, 913 cases flagged by rule number 3 and 56 cases flagged by rule number 4. After we inspected the denominator using the first rule, we went to analyze the numerator. Given that we have a large number of cases flagged under these rules, we developed a method to investigate them, which is completely based on LLM responses.

Our code extracts values from responses that follow the expressions 'The total diverted was', 'The total overpriced was' and 'The total fraud was'. However, there are answers in which the value appears as undefined but they indicate suspicious cases. The example below illustrates this.[24]

LLM responses: Example 1

"The report mentions a case of irregularity in procurement, where medicines purchases were made through tender waiver in an amount higher than that set in Law

[23] We also expanded the interval to look more cases (those above 98th percentile) and all the additional values were correctly obtained by our algorithm.

[24] The response was translated to English.

No. 8,666/93. However, the report does not specify the exact value of this irregularity. Therefore, I cannot conclude with 'Total fraud was' followed by a value as the report does not provide that information." (Onça de Pitangui-MG, lottery 6)

Although the model does not provide a specific value, it does mention a case where the procurement law was not respected. In such cases, we conduct a search for associated keywords in the reports to retrieve the value. On the other side, if all the answers are generic such as: "The report does not mention any case of diversion of funds. The total is R$ 0,00.", we accept the zero value. By doing this analysis, we reduced the number of observations with zero corruption from 913 to 512.

We used the same logic to investigate observations flagged by rules 2 and 4. High percentages of corruption may arise from errors in the denominator, what we deal with the investigation of very low audited resources, or from overestimation of corruption value. In these cases, we checked whether the values mentioned in the responses were consistent with the values present in the report. From the 22 observations falling above the 99th percentile, we fixed 11.

Finally, we investigated corruption values under R$500. Some of these cases were wrongly assigned by our algorithm, particularly in cases related to overpricing. For example, it occasionally extracted a unitary price instead of the total overpricing amount.

A.4 Addressing Inconsistencies in Data

Corruption Measures

A difference between our corruption measures and those manually coded by FF and Brollo et al, as well as data provided by the CGU, lies in the level of aggregation used to identify corruption. In their data, corruption is identified by lottery and political term. The political term is given by the year in which the resources associated with the corruption were transferred, instead of the audit year.

A limitation of our algorithm is the inability to identify the year in which the resources involved in corruption were transferred. For example, if an audit from lottery 15, held in 2005, found a corruption case involving resources transferred in 2004, our model can not attribute it to either 2004 or 2005 — two different political terms. Occasionally, information about the year of resource transfer is presented as a tag named "extension of exams". However, in some cases, this information is provided as a range, such as from January/2004 to August/2005 (See Figure C.7). While precise information may occasionally be present within the text, extracting it would introduce additional complexity and noise into our model's inquiries. In addition to ask about the corruption value we would need to ask about the year the referred resource was transferred. We opt then to identify the corruption at the level of municipality-lottery.

In order to compute the correlation between LLM and Brollo et al.'s variables, we aggregate the share of broad and narrow corruption by lottery and municipality. Similarly, for the indicator variables "Any Broad" and "Any Narrow", we take the maximum by lottery and municipality. The same logic is applied to "Some serious irregularity", generated from CGU data.

Regarding the data generated from the LLM, we excluded lotteries where audits occurred within the first six months of each term, as most of the resources in such case may refer to the preceding political term. This encompasses four lotteries: 15, 16, 28, and 38. Unfortunately, we can not rule out the possibility that a corruption is wrongly associated even when the audit occurs latter in the term. Brollo et al.'s findings indicate that 63% of detected corruption cases involve irregularities that occurred during the same term as the audit, rather than in previous terms. These within-term cases also exhibit a higher average share of audited resources classified as corruption (5.4% compared to 3.1% in earlier-term cases).

Elections Data

Regarding elections data, we find a correlation of 0.97 between our variable and FF indicator for mayors in their first term. This near-perfect correlation is slightly reduced due to the poor quality of the 1996 election data[25] The differences in this variable arise

[25]The main source was the TSE harmonized data provided by *Data Basis*, but we also supplemented

from two main sources: missing candidate names or identifiers in 1996, which makes it difficult to determine their status in 2000, and cases where the mayor did not complete their term (e.g., due to death) and the vice mayor took over and was subsequently re-elected.

If no information was available for 1996 but the 2000 mayor ran for re-election in 2004, we assumed they were in their first term in 2000. If no information was available for either 1996 or 2004, we conducted Google searches to verify the information. Lastly, in cases where the original mayor did not complete their term—due to reasons such as death—and the vice mayor assumed office and was later re-elected, we categorized these cases as second-term mayors.

it with TSE original data for the 1996 election.

B CNJ Data

Data on the convictions of mayors for misconduct in public office was obtained from the Cadastro Nacional de Condenações Cíveis por ato de Improbidade Administrativa e Inelegibilidade. This dataset, managed by the *Cadastro Nacional de Justiça* (CNJ), records individuals convicted of misconduct in public office, including their names, the filing date, conviction date, and registry date. It also details the type of irregularity, the court responsible for the conviction, and the individual's position at the time of prosecution. The dataset covers all convicted officials up to 2023.

Our analysis focuses on cases where the convicted individual was a current or former mayor. Between 2000 and 2023, this amounts to 9,450 convictions involving 4,179 mayors. These data were then matched with electoral data using individuals' names. This process successfully linked 81% of the observations (convictions), allowing us to determine whether a mayor was convicted during their time in office or afterward.

C Tables and Figures

Table C.1: The Effect of Reelection Incentives on Corruption by Term

	2001-2015	2001-2004	2005-2008	2009-2012	2013-2015
	(1)	(2)	(3)	(4)	(5)
Panel A: Any corruption LLM					
Mayor in first term	-0.0450**	-0.0635*	0.0284	-0.0781**	-0.3034***
	(0.0209)	(0.0370)	(0.0400)	(0.0383)	(0.0915)
Mean	0.7682	0.6960	0.7909	0.8320	0.7890
Observations	1,885	658	636	482	109
Panel B: Share corrupt LLM					
Mayor in first term	-0.0022	-0.0055	-0.0025	0.0077*	-0.0391*
	(0.0033)	(0.0065)	(0.0063)	(0.0044)	(0.0198)
Mean	0.0269	0.0342	0.0249	0.0211	0.0197
Observations	1,884	658	636	481	109
Panel C: Number of cases LLM					
Mayor in first term	-0.1984	-0.3138*	-0.1403	0.0163	-1.0410**
	(0.1212)	(0.1794)	(0.3062)	(0.2085)	(0.4795)
Mean	1.1815	0.9779	1.5057	1.1290	0.8932
Observations	1,620	588	526	403	103
Mayor Characteristics	Yes	Yes	Yes	Yes	Yes
Lottery Dummies	Yes	Yes	Yes	Yes	Yes
State Dummies	Yes	Yes	Yes	Yes	Yes
Party Dummies	Yes	Yes	Yes	Yes	Yes

Note: This table presents the impact of reelection incentives on three corruption metrics: the probability of finding a corruption case (Panel A), the proportion of audited resources associated with corruption (Panel B), and the number of detected corruption cases (Panel C). We displays the results from the OLS regression presented in Equation 2, where the respective corruption measure is regressed on an indicator variable denoting whether the mayor is in their first term. The first column covers all four political terms with audited resources, while the subsequent columns break down each term individually. All columns include controls to Mayor's characteristics and party affiliation, as well as state and lottery intercepts. Mayor's characteristics are age, gender and education. This estimate includes municipalities audited from lotteries 2 to 40, with the exception of lotteries 15, 16, 28, and 38 —excluded due to their occurrence within the first six months of a political term (see Appendix B for further explanation). Heteroscedasticity robust standard errors are displayed in parenthesis. P-values: * 0.10 ** 0.05 *** 0.01

Table C.2: The Effect of Reelection Incentives on Corruption using Brollo et al.'s data (2003-2009)

	Any corruption		Any narrow corruption		Share corrupt		Share narrow corrupt	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mayor in first term	0.0112	-0.0010	-0.0422	-0.0523*	-0.0012	0.0015	-0.0054	-0.0026
	(0.0232)	(0.0223)	(0.0279)	(0.0267)	(0.0059)	(0.0061)	(0.0038)	(0.0042)
Mayor Characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Lottery Dummies	No	Yes	No	Yes	No	Yes	No	Yes
State Dummies	No	Yes	No	Yes	No	Yes	No	Yes
Party Dummies	No	Yes	No	Yes	No	Yes	No	Yes
Mean	0.7830	0.7823	0.4697	0.4705	0.0525	0.0522	0.0213	0.0214
Observations	1,401	1,392	1,401	1,392	1,336	1,327	1,337	1,328

Note: This table reports the effects of reelection incentives on the probability of finding a corruption case and the share of resources found to involve corruption. Broad corruption includes irregularities that could also be interpreted as bad administration rather than as overt corruption, and narrow corruption includes severe irregularities. We regress each corruption measure on an indicator variable for whether the mayor is in his first term, as specified in Equation 2. The even numbered columns include controls to Mayor's characteristics and party affiliation, as well as state and lottery intercept. Mayor's characteristics include the age, gender, education, and party affiliation. We restrict the analysis to observations within the same term, excluding corruption associated with resources transferred in previous political terms (see Appendix B for further explanation). The period from 2003 to 2009 indicates the years with audits. Heteroscedasticity robust standard errors are displayed in parenthesis. P-values: * 0.10 ** 0.05 *** 0.01

Table C.3: The Effect of Reelection Incentives on Corruption using CGU data (2005-2015)

	Number of serious occurrences		Any serious occurrences	
	(1)	(2)	(3)	(4)
Mayor in first term	0.7717	0.5642	-0.0019	-0.0203
	(0.5628)	(0.4982)	(0.0261)	(0.0258)
Mayor Characteristics	No	Yes	No	Yes
Lottery Dummies	No	Yes	No	Yes
State Dummies	No	Yes	No	Yes
Pary Dummies	No	Yes	No	Yes
Mean	7.0850	7.0850	0.7842	0.7842
Observations	1,117	1,117	1,117	1,117

Note: This table reports the effects of reelection incentives on the number of irregularities associated with corruption and on the probability of finding a serious occurrence. We regress each corruption measure on an indicator variable for whether the mayor is in his first term, as specified in Equation 2. The even numbered columns include controls to Mayor's characteristics and party affiliation, as well as state and lottery intercept. Mayor's characteristics include the age, gender, education, and party affiliation. We restrict the analysis to observations within the same term, excluding corruption associated with resources transferred in previous political terms (see Appendix B for further explanation). The period from 2005 to 2015 spans the three political terms with audited resources. The last term, however, does not consider the final year (2016), as the last available lottery (40) was conducted in 2015. Heteroscedasticity robust standard errors are displayed in parenthesis. P-values: * 0.10 ** 0.05 *** 0.01

Table C.4: The Impact of Reelection Incentives on Corruption, RD Robustness

	Share corrupt		Any corruption		Number of cases	
	(1)	(2)	(3)	(4)	(5)	(6)
Mayor in first term	0.005	-0.005	-0.034	-0.098*	-0.063	-0.083
	(0.015)	(0.009)	(0.107)	(0.058)	(0.464)	(0.253)
Robust 90% CI	[-.033 ; .048]	[-.031 ; .019]	[-.359 ; .258]	[-.25 ; .072]	[-1.081 ; 1.28]	[-.718 ; .726]
Kernel Type	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular
BW Type	.5CCT	2CCT	.5CCT	2CCT	.5CCT	2CCT
BW	0.081	0.325	0.083	0.331	0.112	0.449
Observations	1025	1025	1026	1026	877	877

Note: This table presents the coefficients from the RD regression specified in Equation 3. We evaluate the impact of reelection incentives on three corruption metrics: the probability of finding a corruption case, the proportion of audited resources associated with corruption, and the number of detected corruption cases. All columns include controls to Mayor's characteristics and party affiliation, as well as state and lottery intercepts. Mayor's characteristics are age, gender and education. We include municipalities audited from lotteries 2 to 40 if the mayor ran for reelection. As in the previous cases, we excluded lotteries 15, 16, 28, and 38 due to their occurrence within the first six months of a political term (see Appendix B for further explanation). The BW Type specifies whether half of the MSE optimal bandwidth (.5CCT) or twice the MSE optimal bandwidth (2CCT) was used. The BW parameter reports the respective bandwidth for each regression. Heteroscedasticity robust standard errors are displayed in parenthesis. P-values: * 0.10 ** 0.05 *** 0.01

Table C.5: The Impact of Reelection Incentives on Corruption, RD (FF and Brollo et al.)

	Share corrupt	Any corruption	Share corrupt	Share narrow corrupt	Any corruption	Any narrow corruption
	(1)	(2)	(3)	(4)	(5)	(6)
	FF	FF	Brollo et al.	Brollo et al.	Brollo et al.	Brollo et al.
Mayor in first term	-0.022	-0.083	-0.001	0.003	-0.003	-0.086
	(0.025)	(0.091)	(0.017)	(0.011)	(0.083)	(0.087)
Robust 90% CI	[-.064 ; .07]	[-.257 ; .192]	[-.04 ; .031]	[-.015 ; .029]	[-.225 ; .227]	[-.292 ; .193]
Kernel Type	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular
BW Type	CCT	CCT	CCT	CCT	CCT	CCT
BW	0.194	0.222	0.167	0.184	0.159	0.206
Observations	318	318	752	753	781	781

Note: This table presents the coefficients from the RD regression specified in Equation 3. We evaluate the impact of reelection incentives on the share of audited resources associated with corruption and on the probability of finding a corruption case. Columns 1 and 2 refer to FF measures and include data from 2003 to 2004 (lotteries 2-11), while Columns 3-6 refer to Brollo et al. measures, thus including data from 2003 to 2009 (lotteries 2-29). Our analysis is restricted to mayors who pursued reelection. All columns include controls to Mayor's characteristics and party affiliation, as well as state and lottery intercepts. Mayor's characteristics include age, gender and education. The BW Type indicates that the MSE optimal bandwidth was used (CCT). The BW parameter reports the respective bandwidth for each regression. Heteroscedasticity robust standard errors are displayed in parenthesis. P-values: * 0.10 ** 0.05 *** 0.01

Table C.6: The Impact of Changes in Cohort and Workers' Party Growth on Reelection Incentives (Brollo et al.)

	Any narrow corruption							
	2001-2004	2001-2009	2005-2009	2001-2009	2001-2009	2001-2009	2005-2009	2005-2009
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mayor in first term	-0.0948**	-0.0514*	-0.0061	-0.0206	-0.0705**	-0.0421	-0.0419	-0.0379
	(0.0384)	(0.0265)	(0.0384)	(0.0373)	(0.0275)	(0.0380)	(0.0404)	(0.0438)
First x Incumbent eligible				-0.0436		-0.0403		-0.0143
				(0.0361)		(0.0358)		(0.0382)
First x Workers' Party					0.3509***	0.3479***	0.4018***	0.4060***
					(0.0938)	(0.0936)	(0.1099)	(0.1104)
Mayor Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lottery Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Party Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean	0.4234	0.4697	0.5108	0.4705	0.4697	0.4705	0.5108	0.5108
Observations	659	1,401	742	1,390	1,401	1,390	742	740

Note: This table presents the coefficients from OLS regressions specified in Equations 2, 4, and 5. All columns include controls to Mayor's characteristics and party affiliation, as well as state and lottery intercepts. Mayor's characteristics include age, gender and education. Columns 1 to 3 are derived from Equation 2 and cover different time periods: the first political term, all available data, and only the middle term. Columns 4 and 5 are derived from Equations 4 and 5, respectively, while Column 6 includes both interactions together. These regressions consider all available data. Finally, in Columns 7 and 8, we estimate Equation 5 and the combination of 4 and 5, respectively, only for the middle term. Heteroscedasticity robust standard errors are displayed in parenthesis. P-values: * 0.10 ** 0.05 *** 0.01

Table C.7: Mayors Holding a Political Office After Second Term

	2000 mean/sd	2004 mean/sd	2008 mean/sd	2012 mean/sd	2016 mean/sd
Second term mayors elected 2 years later	0.007	0.006	0.005	0.003	0.004
	(0.085)	(0.079)	(0.070)	(0.056)	(0.065)
Second term mayors elected 6 years later	0.048	0.033	0.030	0.023	0.043
	(0.213)	(0.180)	(0.172)	(0.150)	(0.203)
Observations	2074	1285	2037	1254	1167

Note: This table presents the percentage of second-term mayors who were subsequently elected to higher office. The years listed in columns indicate the election year when the mayor was reelected. The term "2 years later" refers to the first national/state election following the municipal election in which the mayor was reelected. The term "6 years later" refers to the second national/state election following the municipal election in which the mayor was reelected.

Table C.8: Mayors Running for Political Office After Second Term

	2000 mean/sd	2004 mean/sd	2008 mean/sd	2012 mean/sd	2016 mean/sd
Second term mayors running 2 years later	0.015	0.017	0.010	0.009	0.013
	(0.123)	(0.130)	(0.101)	(0.093)	(0.113)
Second term mayors running 6 years later	0.127	0.125	0.095	0.097	0.143
	(0.333)	(0.330)	(0.294)	(0.296)	(0.350)
Observations	2074	1285	2037	1254	1167

Note: This table presents the percentage of second-term mayors running for higher office in subsequent elections. The years listed in columns indicate the election year when the mayor was reelected. The term "2 years later" refers to the first national/state election following the municipal election in which the mayor was reelected. The term "6 years later" refers to the second national/state election following the municipal election in which the mayor was reelected.

Table C.9: Parties Remaining in Office

	2000 mean/sd	2004 mean/sd	2008 mean/sd	2012 mean/sd	2016 mean/sd
Percentage of second term mayors					
whose parties stayed in power in the next election	0.209	0.209	0.197	0.176	0.230
	(0.407)	(0.407)	(0.398)	(0.381)	(0.421)
Percentage of second term mayors					
whose parties returned to power two elections later	0.152	0.213	0.251	0.189	.
	(0.359)	(0.409)	(0.434)	(0.392)	(.)
Observations	2098	1294	2063	1264	1177

Note: This table presents the percentage of second-term mayors whose parties remained in power in the subsequent election or returned to power 4 years later. The years listed in the columns indicate the election year when the mayor was reelected.

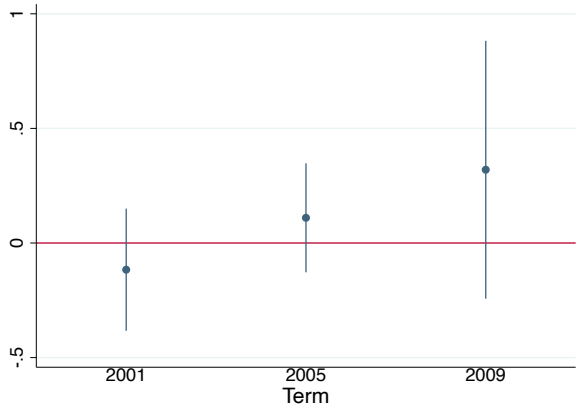
Table C.10: Summary Statistics - Convictions

	Mean	Sd	Min	Max	N
Difference between conviction and filling year	6.488	3.627	0	22	7,471
Difference between registration and filling year	7.763	3.947	0	23	7,471
Difference between filling and election year	8.256	3.641	0	26	7,471

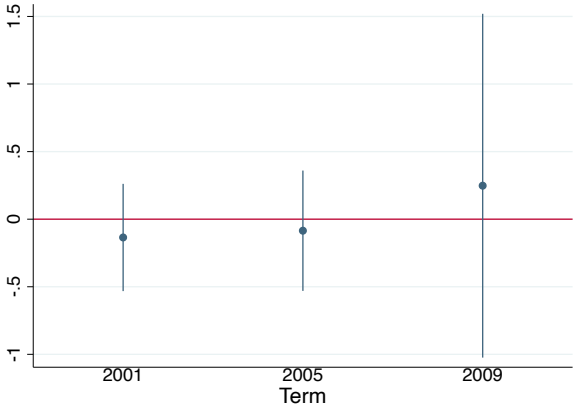
Note: The filing year refers to the year when the legal process was initiated. The conviction year is the year in which the court issued the conviction. The registry year is the year the case was entered into the *Cadastro Nacional de Justiça* system. This analysis considers cases filed between 2000 and 2023. We exclude cases where the individual was convicted before taking office as mayor. Additionally, we exclude cases where the conviction year is earlier than the filing year, as this deviates from the expected process and likely indicates an error.

Figure C.1: The Effects of Reelection Incentives on Corruption Over Time (Brollo et al.)

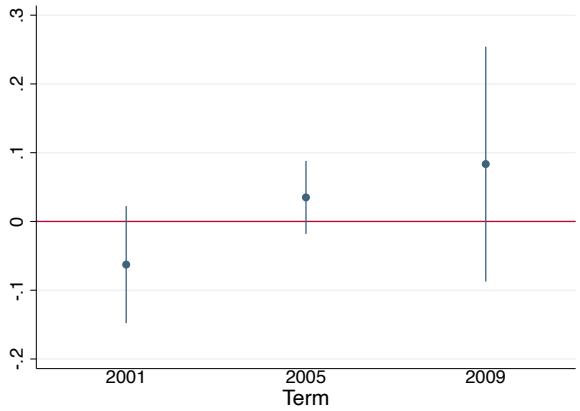
(a) Share broad corruption



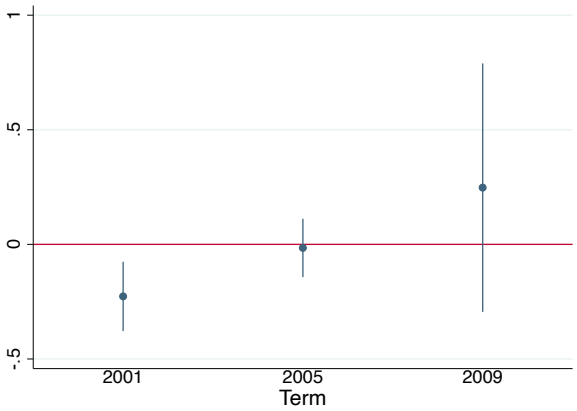
(b) Share narrow corruption



(c) Any broad corruption

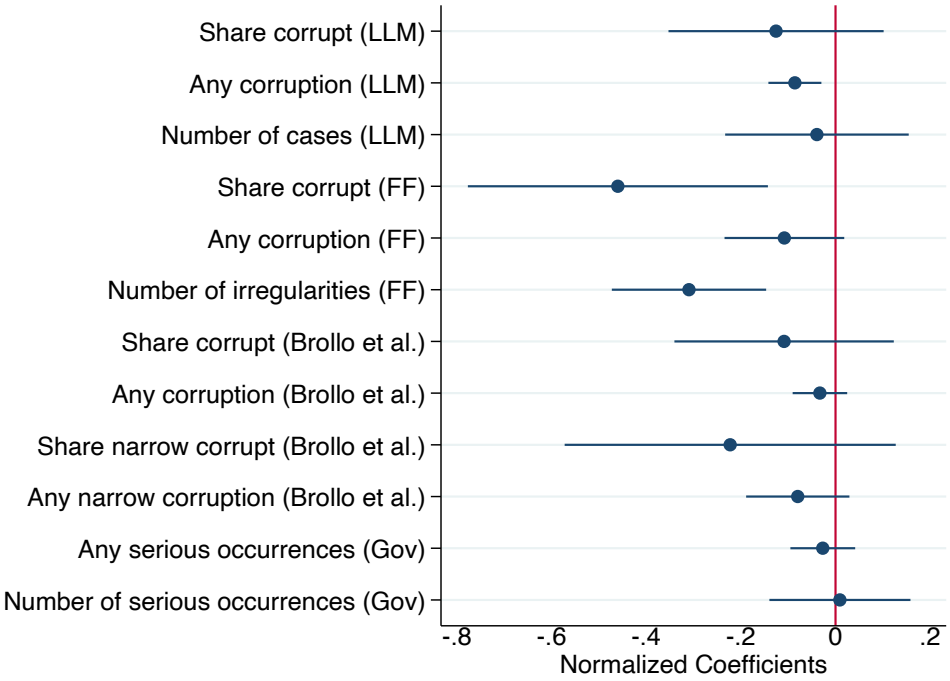


(d) Any narrow corruption



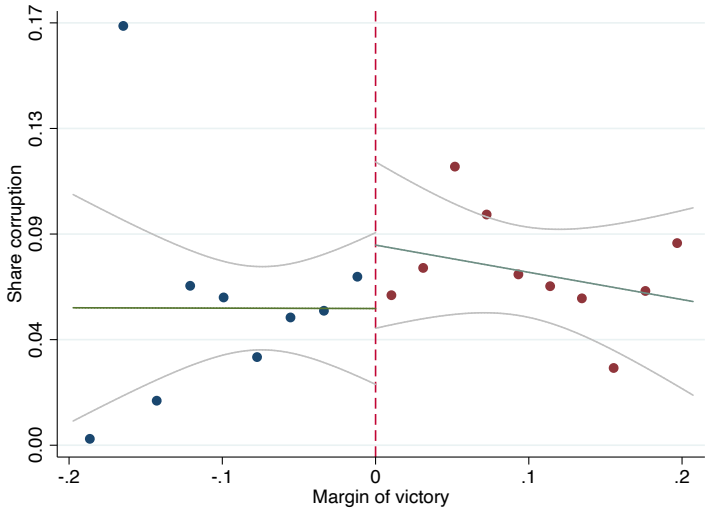
Note: This figure presents the coefficients from the OLS regression specified in Equation 2. All regressions include controls to Mayor's characteristics and party affiliation, as well as state and lottery intercepts. Mayor's characteristics are age, gender and education. The results are broken down by term. We restrict the analysis to observations within the same term, excluding corruption associated with resources transferred in previous political terms (see Appendix B for further explanation). Estimates includes data from 2003 to 2009 (lotteries 2-29). Each term spans a four-year period, with the exception of 2009, which includes only the first year due to the last available lottery (29) being conducted in that year. Confidence intervals are displayed at the 90% level.

Figure C.2: Overall Effects of Reelection Incentives on Corruption
Only Reelected Mayors (All Available Years)



Note: This figure depicts the coefficients from the OLS regression outlined in Equation 2. All coefficients are normalized by dividing it by the outcome variable mean. All standard errors are heteroscedasticity-robust. We estimate the impact of reelection incentives on corruption using all available measures obtained from LLM, FF and Brollo et al.'s dataset. We restrict the sample to consider only reelected mayors. All regressions include controls to Mayor's characteristics and party affiliation, as well as state and lottery intercepts. Mayor's characteristics are age, gender and education. For each variable, we plot its source in parentheses. The regressions include different time coverage. Estimates using LLM variables includes data from 2003 to 2015 (lotteries 2 to 40). We exclude lotteries 15, 16, 28, and 38 due to their occurrence within the first six months of a political term (see Appendix B for further explanation). FF's data cover the 2003-2004 time period. Brollo et al.'s data are from 2003 to 2009 (lotteries 2-29), while CGU data are from 2005 to 2015 (lotteries 20 to 40). For both Brollo et al.'s and CGU data, the analysis is restricted to observations within the same term. Confidence intervals are displayed at the 90% level.

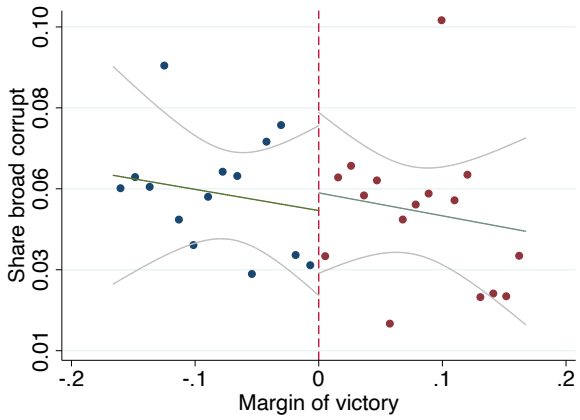
Figure C.3: The Effects of Reelection Incentives on Corruption (FF)



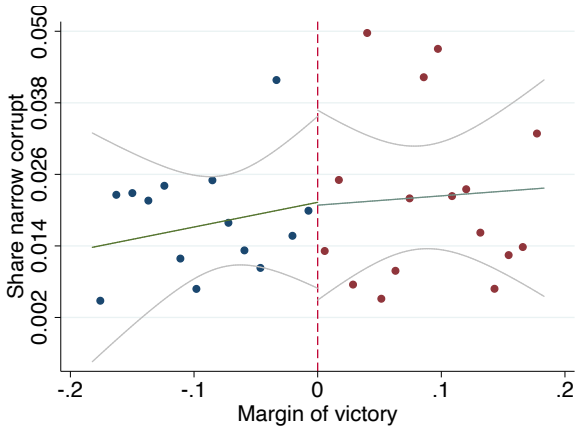
Note: The figure shows the share of audited resources involving corruption by the margin of victory for incumbents who ran for reelection in 2000. The grey lines denote the confidence intervals plotted for fitted lines at the 90% level. The regression used the optimal bandwidth according to the minimum squared error (MSE) criteria (Calonico et al., 2014). We restrict the observations, such that only mayors associated with a vote margin within the interval of the optimal bandwidths are considered.

Figure C.4: The Effects of Reelection Incentives on Corruption (Brollo et al.)

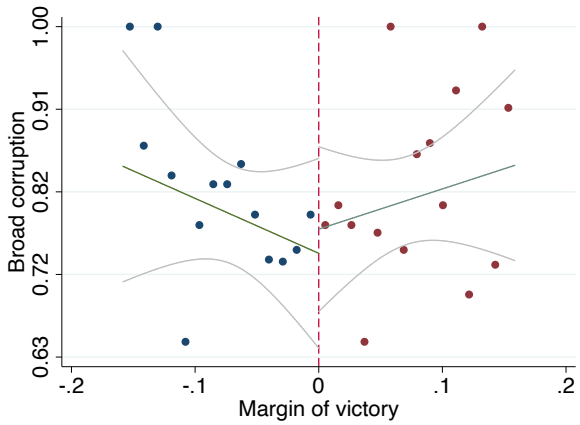
(a) Share broad corruption



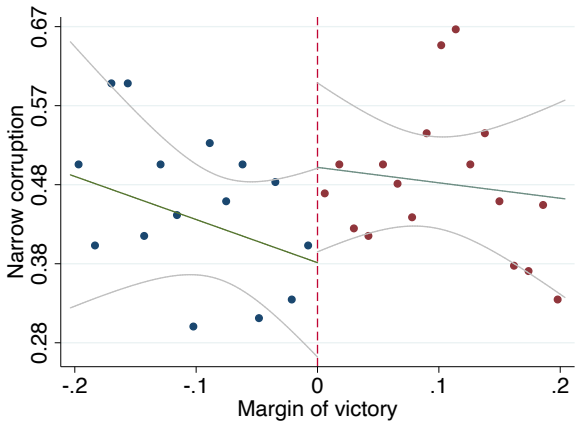
(b) Share narrow corruption



(c) Any broad corruption

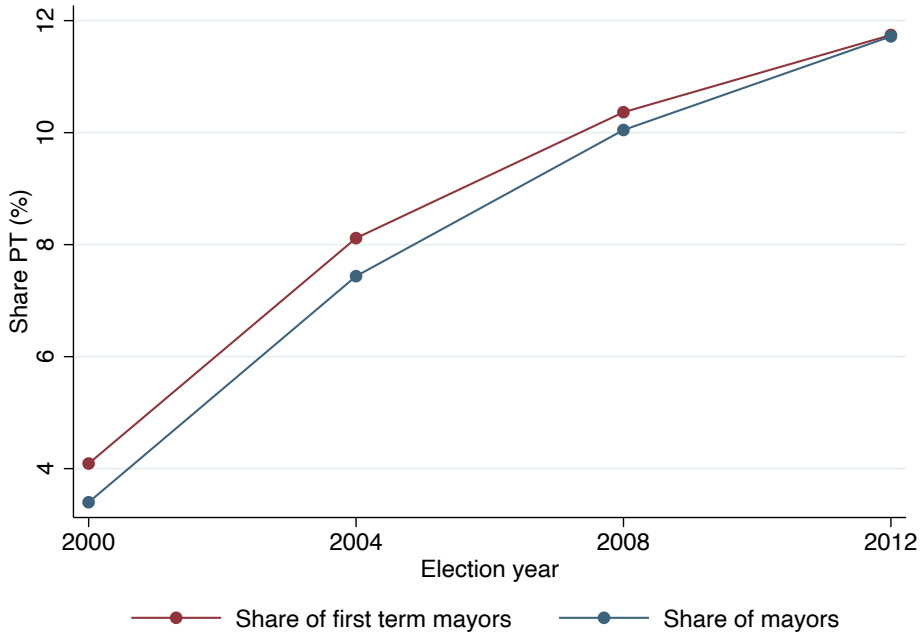


(d) Any narrow corruption



Note:The figure shows the share of audited resources involving corruption (Panels C.4a and C.4b) the indicator for detected corruption (Panels C.4c and C.4d) by the margin of victory for incumbents who ran for reelection. The grey lines denote the confidence intervals for fitted lines at the 90% level. All regressions use the optimal bandwidth according to the minimum squared error (MSE) criteria (Calonico et al., 2014). We restrict the observations, such that only mayors associated with a vote margin within the interval of the optimal bandwidths are considered.

Figure C.5: Percentage of Worker's Party (PT) Mayors Over Time



Note: This figure depicts the percentage of elected mayors and first-term mayors affiliated with the Worker's Party over time. The x-axis represents the municipal election years.

Figure C.6: Example of Total Audited Amount

PRESIDÊNCIA DA REPÚBLICA
CONTROLADORIA-GERAL DA UNIÃO
SECRETARIA FEDERAL DE CONTROLE INTERNO

RELATÓRIO DE FISCALIZAÇÃO Nº 08/2003

1. Trata o presente Relatório dos resultados das **84 ações de fiscalização** realizadas em decorrência do 3º sorteio do Projeto de Fiscalização a partir de Sorteios Públicos, no qual foi sorteado o Município de **Irauçuba-CE.**

2. As fiscalizações tiveram como objetivo analisar a aplicação dos recursos federais aplicados no Município sob a responsabilidade de órgãos federais, estaduais, municipais ou entidades legalmente habilitadas, bem como, avaliar a atuação dos Conselhos Municipais responsáveis pelo acompanhamento dos referidos Programas de Governo.

3. Os trabalhos foram realizados "in loco" no Município, no período de 30/06/2003 a 04/07/2003, sendo utilizados em sua execução: inspeções físicas, análises documentais, entrevistas, aplicação de questionários e registros fotográficos, em observância ao que foi estabelecido nas Ordens de Serviço expedidas pelas Coordenações-Gerais das Diretorias da Secretaria Federal de Controle Interno, responsáveis pelo acompanhamento da execução dos Programas de Governo fiscalizados.

4. Os Programas de Governo que foram objeto das ações de fiscalização, estão apresentados no quadro a seguir, por Ministério Supervisor, discriminando, a quantidade de fiscalizações realizadas e os recursos aproximados envolvidos, por Programa.

4.1 Recursos recebidos e quantidade de fiscalizações realizadas

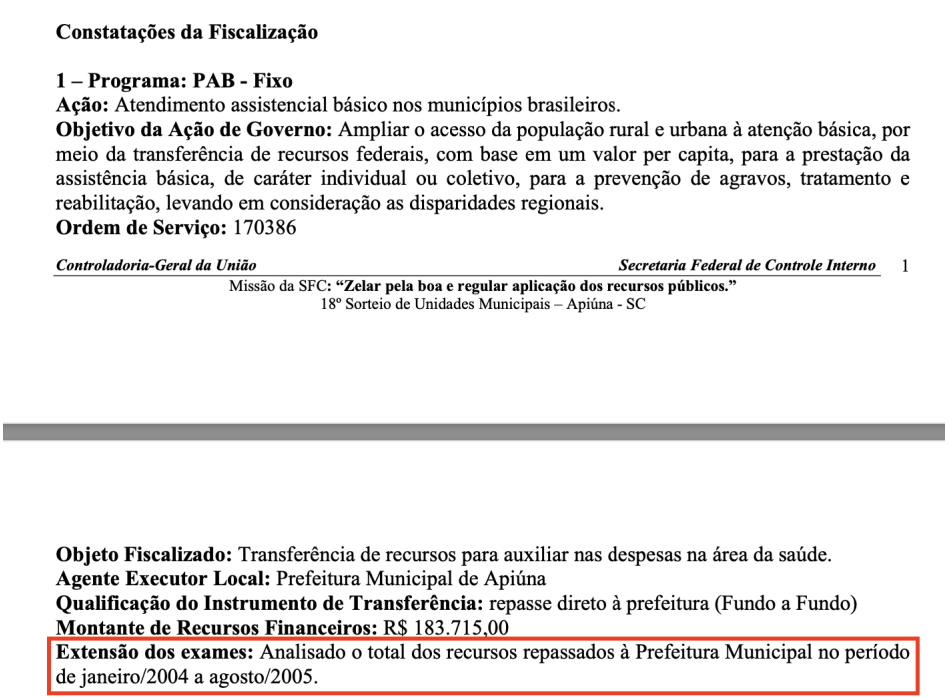
Ministério Supervisor	Objeto Fiscalizado	Quantidade de Fiscalizações	Recursos Fiscalizados
Ministério da Fazenda	Financiamento e Equalização de Juros para a Agricultura Familiar – PRONAF	01	-
Ministério da Educação	Veículos para Transporte Escolar	01	50.000,00
	Alimentação Escolar	02	179.916,60
	Garantia de Padrão Mínimo de Qualidade para o Ensino Fundamental de Jovens e Adultos – Recomeço	01	148.270,76

Missão da SFC: **"Zelar pela boa e regular aplicação dos recursos públicos."**
SAS Q. 1 Bl "A", Ed. Darcy Ribeiro, 9º andar, Brasília - DF - CEP: 70070-905 (61) 412-7115 - Fax (61) 322-1672

Ministério Supervisor	Programa / Ação	Quantidade de Fiscalizações	Recursos Fiscalizados
Ministério de Minas e Energia	Fiscalização e Controle da Produção Mineral	02	-
	Fiscalização da Distribuição e Revenda de Derivados de Petróleo e Álcool Combutível	01	-
Ministério da Integração Nacional	Financiamento aos Setores Produtivos da Região Nordeste	01	-
	Ações Emergenciais de Defesa Civil – Bolsa Renda	01	1.260,00
	Construção de Açude Público	04	515.338,87
	Ampliação de Açude Público	04	475.802,35
	Construção de Passagem Molhada	05	572.117,37
Ministério do Desenvolvimento Agrário	PRONAF	01	-
	Investimento em Infra-Estrutura Básica para Assentamentos Rurais-Nordeste - Construção de açude	01	102.693,21
Ministério do Turismo	Desenvolvimento da Infra-Estrutura Turística na Região	02	187.500,00
Ministério do Esporte	Implantação de Núcleos de Esporte	01	105.300,00
TOTAL		**84**	**4.877.991,51**

Note: This figure shows a table detailing the total audited amount for the Municipality of Irauçuba-CE, audited during the third lottery.

Figure C.7: Example of Exam Extension Information



Constatações da Fiscalização

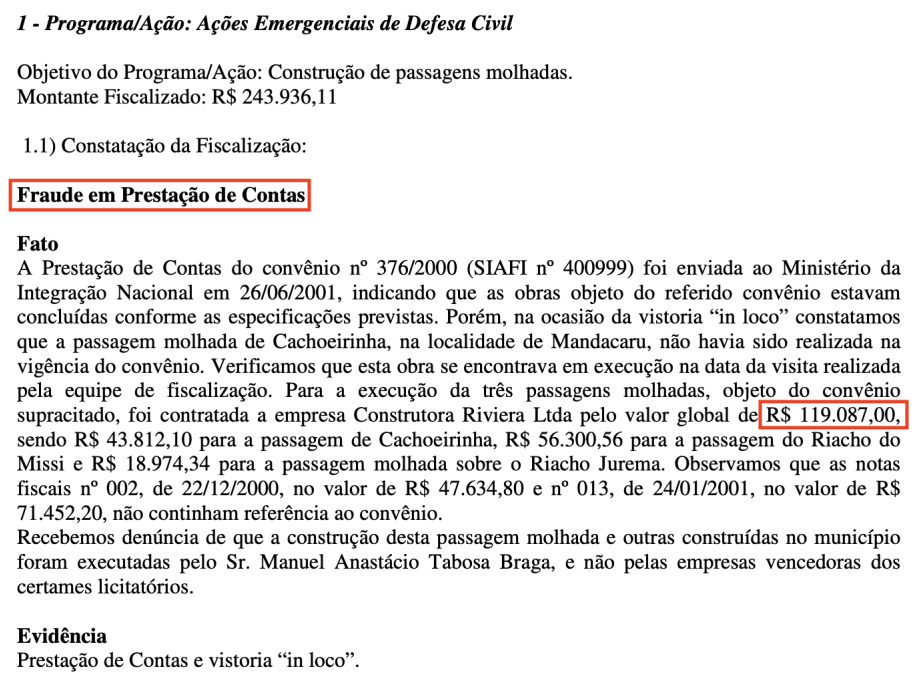
1 – Programa: PAB - Fixo
Ação: Atendimento assistencial básico nos municípios brasileiros.
Objetivo da Ação de Governo: Ampliar o acesso da população rural e urbana à atenção básica, por meio da transferência de recursos federais, com base em um valor per capita, para a prestação da assistência básica, de caráter individual ou coletivo, para a prevenção de agravos, tratamento e reabilitação, levando em consideração as disparidades regionais.
Ordem de Serviço: 170386

Controladoria-Geral da União *Secretaria Federal de Controle Interno* 1
Missão da SFC: **"Zelar pela boa e regular aplicação dos recursos públicos."**
18º Sorteio de Unidades Municipais – Apiúna - SC

Objeto Fiscalizado: Transferência de recursos para auxiliar nas despesas na área da saúde.
Agente Executor Local: Prefeitura Municipal de Apiúna
Qualificação do Instrumento de Transferência: repasse direto à prefeitura (Fundo a Fundo)
Montante de Recursos Financeiros: R$ 183.715,00
Extensão dos exames: Analisado o total dos recursos repassados à Prefeitura Municipal no período de janeiro/2004 a agosto/2005.

Note: This figure provides an example of how information on exam extensions is presented. The example is from the Municipality of Apiúna-SC, audited during the eighteenth lottery.

Figure C.8: Example of Fraud



1 - Programa/Ação: Ações Emergenciais de Defesa Civil

Objetivo do Programa/Ação: Construção de passagens molhadas.
Montante Fiscalizado: R$ 243.936,11

1.1) Constatação da Fiscalização:

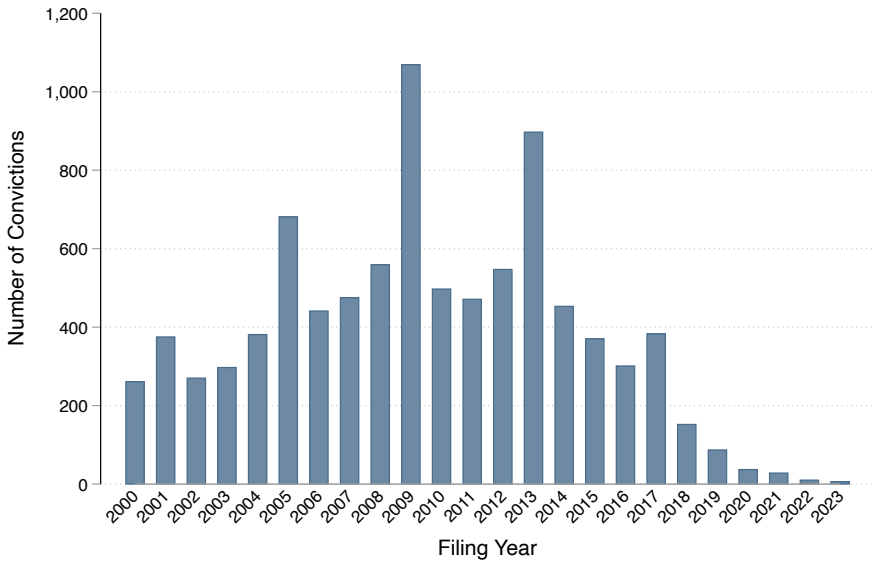
Fraude em Prestação de Contas

Fato
A Prestação de Contas do convênio nº 376/2000 (SIAFI nº 400999) foi enviada ao Ministério da Integração Nacional em 26/06/2001, indicando que as obras objeto do referido convênio estavam concluídas conforme as especificações previstas. Porém, na ocasião da vistoria "in loco" constatamos que a passagem molhada de Cachoeirinha, na localidade de Mandacaru, não havia sido realizada na vigência do convênio. Verificamos que esta obra se encontrava em execução na data da visita realizada pela equipe de fiscalização. Para a execução da três passagens molhadas, objeto do convênio supracitado, foi contratada a empresa Construtora Riviera Ltda pelo valor global de R$ 119.087,00, sendo R$ 43.812,10 para a passagem de Cachoeirinha, R$ 56.300,56 para a passagem do Riacho do Missi e R$ 18.974,34 para a passagem molhada sobre o Riacho Jurema. Observamos que as notas fiscais nº 002, de 22/12/2000, no valor de R$ 47.634,80 e nº 013, de 24/01/2001, no valor de R$ 71.452,20, não continham referência ao convênio.
Recebemos denúncia de que a construção desta passagem molhada e outras construídas no município foram executadas pelo Sr. Manuel Anastácio Tabosa Braga, e não pelas empresas vencedoras dos certames licitatórios.

Evidência
Prestação de Contas e vistoria "in loco".

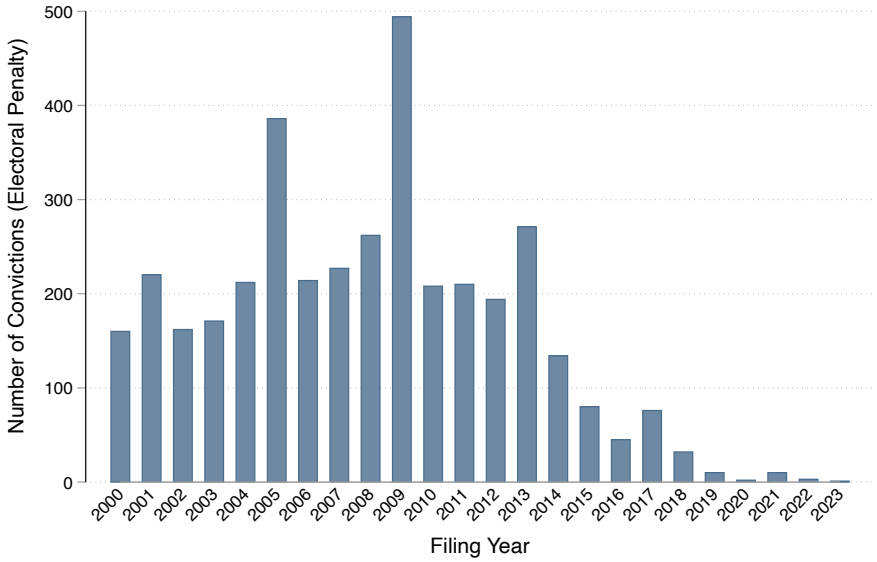
Note: This figure illustrates an example of a fraud case detected in the audit report for the Municipality of Irauçuba-CE, audited during the third lottery.

Figure C.9: Convictions by Filling Year



Note: This figure shows the distribution of convictions by filling year. The data is sourced from the *Cadastro Nacional de Justiça*, filtered to include only convictions of current and former mayors. The filing year refers to the year the legal process was initiated. We exclude cases where the conviction year is earlier than the filing year, as this deviates from the expected process and likely indicates an error. The decrease in the number of observations for 2018 is partly due to the lengthy registration process, as detailed in Table C.10.

Figure C.10: Convictions with Electoral Penalty by Filling Year



Note: This figure shows the distribution of convictions associated with electoral penalties, by filling year. The data is sourced from the *Cadastro Nacional de Justiça*, filtered to include only convictions of current and former mayors. The filing year refers to the year the legal process was initiated. We exclude cases where the conviction year is earlier than the filing year, as this deviates from the expected process and likely indicates an error. The decrease in the number of observations for 2018 is partly due to the lengthy registration process, as detailed in Table C.10.