A Machine Learning Approach to Analyze and Support Anti-Corruption Policy*

Elliott Ash[1], Sergio Galletta[1], Tommaso Giommoni[1]

[1] *ETH Zürich*

Abstract

Can machine learning support better governance? This study uses a tree-based gradient-boosted classifier to predict corruption in Brazilian municipalities using budget data as predictors. The trained model offers a predictive measure of corruption, which we validate through replication and extension of previous corruption studies. Our policy simulations show that machine learning can significantly enhance corruption detection: compared to random audits, a machine-guided targeted policy could detect almost twice as many corrupt municipalities for the same audit rate.

JEL Classification: C53, D73, H83, K42.

Keywords: algorithmic decision-making, corruption policy, local public finance.

*We benefited from comments by participants at seminars at the Collegio Carlo Alberto Turin, Catholic University Milan, University of St. Gallen, ETH Zürich, University of Zürich, the "Advances in Economics Winter Symposium" (Bergamo), Transatlantic Workshop on the Economics of Crime (Bocconi), Bonn Law and Economics Workshop, CESifo Area Conference on Public Economics and the Online Workshop in Computational Analysis of Law. We are grateful to Francisco Cavalcanti for sharing essential data for conducting the analysis. David Cai, Matteo Pinna, and Angelica Serrano provided excellent research assistance. Thanks to Daniel Bjorkegren, Aniket Kesari, Himabindu Lakkaraju, Michael Livermore, Julian Nyarko, Alexis Perakis, and Martin Stancsics for comments on previous drafts. *This version: May 2023 (First version: April 2020).*

Email addresses: `ashe@ethz.ch` (Elliott Ash), `sergio.galletta@gess.ethz.ch` (Sergio Galletta), `giommoni@ethz.ch` (Tommaso Giommoni)

1. Introduction

Given the deep and negative impacts of corruption on citizens and society as a whole,[1] policymakers and governance researchers across the globe have prioritized anti-corruption efforts.[2] Corrupt actors have strong incentives to conceal their actions, however, so anti-corruption policies are often frustrated by the costs of detecting corruption in the first place. Hence, although several countries have introduced monitoring programs to detect wrongdoing (e.g., Olken, 2007; Ferraz and Finan, 2008; Bobonis et al., 2016), the costs of those programs are high relative to the benefits in terms of detection and deterrence.

A promising approach to making anti-corruption efforts more cost-effective is to support them with machine learning. In line with successful efforts to support policy in other domains (e.g. Bansak et al., 2018; Kleinberg et al., 2018), the idea is to use machine learning algorithms to detect patterns indicating a high risk of corruption. We explore this idea in the context of local public finances in Brazilian municipalities, where we have access to both detailed records on municipal budgets and the outcomes of relatively frequent corruption audits undertaken by Brazilian federal authorities (Ferraz and Finan, 2008; Brollo et al., 2013). Motivated by the fact that public corruption involves the (mis)management of public resources (Mauro, 1998), we use machine learning to predict corruption from the features of the budget accounts.

We find that an effective machine learning model for predicting corruption from budget features is a gradient boosted classifier, an ensemble of decision trees typically used to identify patterns in high-dimensional datasets. Using only municipal budget characteristics (797 fiscal variables) as inputs, the classifier can detect the existence and predict the intensity of corruption with high accuracy in held-out (unseen) data. In the best model, we get an accuracy of 72%, far better than guessing the modal

[1] According to recent United Nations statistics, for example, international corruption costs the global economy over 3.6 trillion USD annually. On a more micro level, social scientists have demonstrated that foul play by government actors does real harm to the average citizen. These harms lead to responses in politics and political participation (Ferraz and Finan, 2008; Chong et al., 2015), undermine trust toward institutions (Morris and Klesner, 2010), and have additional side effects on the economy (Ferraz et al., 2012; Lagaras et al., 2017).

[2] Broadly speaking, the previous research has identified two important factors. First, electoral incentives play a crucial role in discouraging misbehavior by officials (Ferraz and Finan, 2008; Winters and Weitz-Shapiro, 2013; Poblete-Cazenave, 2021). Second, an effective judicial system to prosecute offenders and enforce the law may be necessary to deter corrupt actions (Becker, 1968; Djankov et al., 2003; Vannutelli, 2021).

category or prediction using linear models.[3] We show that the model accurately ranks municipalities by the probability of corruption, reproducing the distribution of corruption in held-out data. In a dataset of municipalities audited twice, the model can predict within-municipality changes in corruption over time.

The trained prediction model is a complex forest of dozens of decision trees, containing over 10,000 decision nodes that capture decisive non-linearities and interactions between features. Because pairwise correlations between budget features and corruption risk do not give useful insights, we use a feature importance score that identifies which budget factors the model most attends to when making predictions (Hastie et al., 2009). This ranking of features provides some insight into how the model works, especially because we can compare it to the relative frequency that the features are mentioned in the published audit reports. Applying text analysis to a 55-million-word corpus of digitized reports, we show that the corruption-related features identified by the gradient boosted machine are also mentioned significantly more often in those reports. This validation provides evidence that the model captures corruption-related activities rather than just correlated proxies.

With some confidence that the model predictions measure corruption, we extrapolate to the unlabeled budgets and form a synthetic measure of corruption for all municipalities and years. To demonstrate the empirical applicability of the method, we use the predicted corruption measure to replicate previous causal results on local corruption in Brazil. First, we evaluate the result from Brollo et al. (2013) that a revenue windfall, based on population thresholds, increases corruption. In particular, we can replicate this result in a sample of municipalities never audited by the Brazilian authorities. Normalized coefficient magnitudes using the model predictions are comparable with the estimates obtained by Brollo et al. (2013) using the auditor-produced corruption label as the outcome.

As a second empirical application, we extend the analysis from Avis et al. (2018) and analyze the causal effect of auditing on corruption. Because we have a measure of corruption by year, we can implement an event study analysis. We show that audits reduce corruption in fiscal accounts over the subsequent years, with an average drop of

[3]The AUC-ROC (defined in Section 3.4 below) is 0.78. As a reference, our classifier's performance is similar to that of other papers in economics using machine learning to analyze and support decision-making (Kleinberg et al., 2018; Mullainathan and Obermeyer, 2019).

around 2.7% in the probability of malfeasance. Moreover, the effect is especially large for audits that did find corruption, with an average decline of around 18%, about half of the pre-audit mean of 39.5%. In comparison, there is no effect on our measure for audits that did not find corruption.

Beyond their direct interest, these empirical results provide some additional validation about what our machine learning model is capturing. Because predicted corruption responds the same way to treatments (revenue shocks and audits) as audit-measured corruption, that suggests that model predictions are driven by budgeting choices that local officials have some control over. If the model were picking up (mostly) non-budget factors, then we would not see a comparable response to the causal treatments. Thus, we obtain additional confidence that the model is capturing corruption-related activities, rather than just correlated proxies. While not strictly necessary for the limited task of helping to detect corruption in the existing data, this validation is reassuring that the model could generalize to new settings and time periods.

In the last part of the paper, we investigate the potential of an audit policy guided by our model's predictions on corruption risk. We show that, compared to the status quo policy of random audits, a targeted approach based on predicted corruption would be significantly more efficient in the policy goals of detecting and deterring corruption. According to our policy simulations, a targeted approach would detect about 80 percent more corrupt municipalities relative to the random lottery (for the same number of implemented audits). Similarly, by targeting the municipalities at highest risk for corruption, the audit agency could obtain the same number of corruption detections as the random-lottery system but with 47 percent fewer audits, with a corresponding reduction in administrative costs. From a deterrence perspective, it is notable that the annual audit probability, conditional on being corrupt, more than doubles under targeted audits, compared to random audits.

Finally, we consider the implementation issue that algorithmic targeting could change the audit rates across political parties, potentially leading to perceptions of bias in a politically sensitive environment. Using the party affiliations of municipal mayors, we show that this concern is relevant, as there is substantial variation across the five main parties in targeting incidence relative to random audits. To address this potential barrier to implementation, we draw on recent developments in algorithmic fairness (Barocas et al., 2019; Rambachan et al., 2020) and adjust our audit targeting policy to equalize audit rates across parties. We show that such a politically neutral targeting policy can achieve similar gains in policy effectiveness (detecting more corruption).

The findings add to several literatures in economics. Methodologically, our study adds to the emerging literature in economics applying machine learning techniques to overcome the limitations of standard datasets (Athey, 2018). Supervised learning is a type of machine learning where the algorithm is trained on labeled data, meaning that the input data is already classified or labeled. The goal is to learn a mapping function from the input to the output. In contrast, unsupervised learning is a type of machine learning where the algorithm is trained on unlabeled data, meaning that the input data is not classified or labeled. The goal is to discover hidden patterns or structures in the data. The most established technique in empirical work is to use unsupervised learning to analyze high-dimensional data. For example, Hansen et al. (2018) use Latent Dirichlet Allocation (an unsupervised machine learning algorithm) to measure topics and diversity of discussion in Central Bank committee meeting transcripts. Bandiera et al. (2020) use a similar method to detect CEO behavioral types from their work activity records. Like these papers, we use machine learning to extract relevant dimensions from high-dimensional data. However, we use supervised learning (rather than unsupervised learning) to construct these measurements. This approach is related to several papers in political economy that have used supervised learning to extract measures of partisanship from text, to show (for example) changes in polarization over time or to analyze media influence (Gentzkow and Shapiro, 2010; Ash et al., 2017; Gentzkow et al., 2019; Widmer et al., 2020).

At the intersection of machine learning and development economics, a few papers have applied machine learning methods to detect corruption. For instance, López-Iturriaga and Sanz (2018) train a model to predict the presence of a corruption case each year in 52 Spanish provinces. More at the micro level, Gallego et al. (2018) predict corruption investigations associated with a sample of 2 million public contracts in Colombia. Compared to our study both analyses have the disadvantage of relying on samples in which the two classes are not balanced and the selection of the investigated cases does not occur at random. The closest paper is Colonnelli et al. (2019), which also predicts the results of corruption audits in Brazilian municipalities but focuses on time-invariant non-budget variables (private sector activity, financial development, and human capital measures). Besides our focus on fiscal factors, the main difference in our paper is to use the predicted time-varying measure of corruption for an empirical analysis and policy simulation analysis.

Our use of machine learning to guide auditing is most relevant to the literature on AI-powered policy design (Kleinberg et al., 2015; Athey, 2018; Knaus et al., 2018; Athey and Wager, 2021). For example, Kleinberg et al. (2018) show how an algorithm can support the decisions of judges on pre-trial bail release, by identifying which offenders should be denied bail. Correspondingly, we show that machine learning can support government efforts to identify municipalities with suspicious public budgets, where further investigation is warranted. The growing corpus of work in this vein has used machine learning to detect higher-quality teachers (Rockoff et al., 2011), support physician decision-making (Kleinberg et al., 2015; Mullainathan and Obermeyer, 2019), identify restaurants for targeted health inspections (Kang et al., 2013; Glaeser et al., 2016), allocate tax rebates and tax audits (Andini et al., 2018; Battiston et al., 2020), identify crime hotspots (Mohler et al., 2015), assign refugees to their economically optimal locations (Bansak et al., 2018), demarcate areas of the Amazon for protection against deforestation (Assunção et al., 2019), or identify individuals who are most responsive to marketing nudges (Hitsch and Misra, 2018; Knittel and Stolper, 2019). Besides the new setting (corruption policy), we expand on this work in several methodological directions. First, we use model explanation to validate how the model makes its predictions. Second, we validate the empirical relevance of our machine predictions by showing that they respond appropriately as outcomes in causal regressions. Third, we adopt methods from algorithmic fairness (e.g., Rambachan et al., 2020; Kasy and Abebe, 2020) to address potential political biases in the targeted audits.

Substantively, our paper contributes to the literature studying the relationship between corruption and public finance. Many studies emphasize the connection between governmental transfers and public corruption: Brollo et al. (2013) focus on the Brazilian setting, while De Angelis et al. (2020) study the impact of European funds on rent-seeking activity. Another set of papers analyze the extent to which corruption originates from public spending (Hessami, 2014; Cheol and Mikesell, 2018), and there is evidence that policies that constrain public expenditure may reduce corruption (Daniele and Giommoni, 2020). Further, other works attend to the link between public procurement and rent-seeking (Conley and Decarolis, 2016; Coviello and Gagliarducci, 2017; Decarolis and Giorgiantonio, 2020).[4] Our results build on this work with a broader view of the entire

[4]Notably, Mexico recently introduced a corruption risk index in order to tackle corruption in public procurement (link).

budget, rather than focusing on single elements. Our method provides a proof of concept that these issues can be analyzed using machine predictions on corruption, in addition to measurements produced by human audits.

Finally, we add to the existing evidence on the efficacy of auditing programs on corruption in developing countries. This literature includes Olken (2007), which sets up an RCT with villages in Indonesia and find that the introduction of the auditing scheme decreased corruption. Bobonis et al. (2016), studying municipalities from Puerto Rico, show that audits effectively reduce corruption and rent-seeking activities by enhancing electoral accountability in the short run, but these effects do not last. In the Brazilian context, Zamboni and Litschig (2018) show that increasing the probability of being audited could reduce corruption, while Avis et al. (2018) find that the implementation of an audit in a specific city reduces future corruption levels in that city. Our event study analysis confirms the latter results, and we are the first to show the dynamics of this effect. Moreover, we find that the effect is particularly strong in cities where corruption is actually detected.[5]

The rest of the paper is organized as follows. In Section 2 we present the institutional setting and the data. Section 3 describes the prediction procedures and model performance results. Section 4 shows how the model predictions can be used in empirical analyses of corruption. Section 5 reports a set of policy simulations for guided audits supported by machine learning. Section 6 concludes.

2. Institutional Background and Data Sources

2.1. Local Government and Budgets

Brazil has a decentralized governance structure composed of 26 states and 5563 municipalities. Local municipal governments have substantial autonomy in expenditure decisions, with primary responsibility for providing healthcare, education, public transport, and infrastructure. The funding for these services comes mostly from state and federal sources via intergovernmental transfers.

Using information from the Finance Ministry's online database (tesourotransparente.gov.br), we built a dataset on the annual budgets of all Brazilian municipalities for

[5]Our study also contributes to the body of work on corruption and politics in Brazil. For instance, Ferraz and Finan (2008) show that the disclosure of scandals reduces vote shares for the incumbent. Cavalcanti et al. (2018) emphasize that exposing corrupted incumbents affects the quality of candidates selected by their party to run in the following election.

2001 through 2012. For each municipality and year, the online database reports detailed information about the categories of expenditure, revenue, active positions (assets), and passive positions (liabilities). We downloaded the raw data for each year and cleaned the variables to make them comparable across years.

After the cleaning process, our budget dataset includes 797 accounting variables. The full list is shown in Appendix Table G14. Appendix Table A1 reports counts on the number of variables by year and by category. As reporting practices change over time, an increasing number of variables are reported.

2.2. Anti-corruption policy in Brazil

The Brazilian government introduced new anti-corruption policies in 2003, motivated in part by the misuse of federally transferred funds by local authorities. A cornerstone of the program was a system of random audits, in which municipalities are randomly selected to have their fiscal accounts audited for corruption. Around 200 municipalities are chosen per year.[6]

The audits are implemented by officials from *Controladoria Geral da União* (CGU), an independent federal public agency. Every selected municipality is visited by a team of 10 to 15 auditors, who spend a couple of weeks in municipal offices collecting information to identify potential mismanagement in the use of public funds in the previous 3-4 years. The auditors then write a report on irregularities, which is made available to the public within a few months of the inspection. These audit reports provide detailed information, used before in a number of papers to produce measures of corruption (e.g. Ferraz and Finan, 2008; Brollo et al., 2013; Zamboni and Litschig, 2018).

We use the corruption measures constructed by Brollo et al. (2013) for 1,481 municipalities audited from 2003 to 2009. Our analysis focuses on a binary indicator that is based on their measure of *narrow corruption*, limited to severe violations such as illegal procurement, fraud, favoritism, and over-invoicing. A full 42% of municipalities at their first audit are tagged as having narrow corruption. Besides narrow corruption, they define a measure of *broad corruption*, including inconsistencies that could be linked to government mismanagement but not intentional misuse. Broad corruption is less useful

[6]There are about four lottery rounds per year. Separate lotteries are run for each state (meaning some states are getting slightly more lotteries per municipality than others), and cities with more than 500,000 inhabitants are excluded. Starting in 2016 (after our period of analysis), the policymakers began selecting some municipalities for audit based on a risk score, but the details of that targeting are not publicly available.

for our purposes because it is widespread: 76% of audited municipalities have broad corruption.[7] For robustness, we have access to an alternative corruption variable from Avis et al. (2018).[8] Appendix B provides supplementary analysis using this alternative measure.

2.3. Linked Dataset

We join the corruption outcome with the local budget accounts based on the years of the budget that were examined by the auditors. To this, we add demographics from the 2000 Brazilian Census, including *mean income*, *share of population employed*, *sector of occupation* (agriculture, industry, commerce, transportation, services, and public administration), *share with college education*, *poverty rate*, and *Gini Coefficient of income.* Federal-to-municipal revenue transfer data come from the Brazilian National Treasury (*Tesouro Nacional*), while population data come from the Brazilian Institute of Geography and Statistics (IBGE). Finally, we collected information about the mayor party affiliations in 2000, 2004, and 2008 elections. Summary statistics on these variables are reported in Table 1.

3. Predicting Corruption from Budget Data

Our goal is to use the information in the municipal budget to detect corruption. This section outlines how we build our dataset and machine learning model to form those predictions. We evaluate and interpret the predictive model, and then apply it to all municipalities in Brazil for use in the subsequent analysis.

[7] Brollo et al. (2013) compute the share of financial items with corruption (i.e., the ratio between the total amount of funds involved in the detected violation and the total amount audited). In unreported analysis, we also performed predictions using the continuous measure of corruption. Gradient boosting machines can be used as regressors with continuous outcomes without any issues. However, we decided not to pursue this approach after finding that the model's performance was poor relative to the classification of the discrete variable. One potential reason for the poor performance is that the distribution of the continuous measure is limited and highly skewed towards 0, predicting the continuous measure is not much more informative than what was already provided with the binary classification.

[8] As shown in Appendix Figure B3, the two measures are positively correlated, yet clearly capture different dimensions. The measure from Avis et al. (2018) is computed differently from Brollo et al's. We have for each audited municipality the share of inspection orders that presented irregularities. The Avis et al measure is for a different time period, as we have records from July 2006 through March 2013 (lotteries 22–38). With the Avis et al measure, we do not know the exact year (or term) in which the violation took place, so we assign the value to the three years before the actual audit took place.

Table 1: Summary Statistics

Variable	Mean	Std. Dev.	Min	Max	N
True corruption (term)					
Main Labels from Brollo et al. (2013)	0.424	0.494	0.000	1.000	2087
Alternative Labels from Avis et al. (2018)	0.238	0.426	0.000	1.000	1604
Budget categories (year)					
Total assets	1184.5	2272.3	-19.5	300364.4	64933
Financial assets	216.1	473.2	-3153.9	40128.9	64933
Cash	3.5	35.0	-1607.5	5017.8	64933
Financial liabilities	136.1	272.7	-3023.7	25282.2	64933
Taxes	8.9	15.9	0.0	781.5	64933
Revenues from municipal properties	21.4	127.5	-29.2	27471.4	64933
Total expenditures	1315.4	1402.5	7.1	179411.6	64933
Capital expenditure	181.2	250.2	0.0	21258.5	64933
Current expenditures	1134.1	1206.1	0.0	159532.0	64933
Budget surplus/deficit	41.0	3339.2	-3743.5	650900.8	64933
Municipal characteristics					
Mean income	593.0	319.8	29.8	3062.5	64933
Agriculture (% employed)	16.9	10.1	0.0	72.3	64933
Industry (% employed)	4.2	4.2	0.0	37.5	64933
Commerce (% employed)	7.5	3.6	0.3	27.8	64933
Transport (% employed)	1.2	0.7	0.0	5.9	64933
Service (% employed)	6.8	2.7	0.3	19.3	64933
Public administration (% employed)	2.1	1.2	0.1	16.1	64933
Employed population	38.4	8.5	9.7	79.8	64933
Graduated people	1.2	1.3	0.0	16.5	64933
Poor population	10.0	8.1	0.3	54.4	64933
Gini coefficient	0.6	0.1	0.3	0.9	64933
Audit reports mentions					
Number of mentions	1284.8	11014.3	0.0	190610.0	709

Notes: Main Labels from Brollo et al. (2013) captures the binary variable measuring the presence of corruption according to Brollo et al. (2013) (*narrow corruption* variable). *Alternative Labels from Avis et al. (2018)* captures the binary variable measuring the presence of corruption according to Avis et al. (2018). All budget variables are expressed in per-capita terms. The municipal characteristics are drawn from the 2000 Brazilian census. *Mean income* captures the average income of the working population, the variables *Agriculture*, *Industry*, *Commerce*, *Transport*, *Service* and *Public administration* capture the population employed in a specific sector. *Employed population* measures the fraction of employed population, *Graduated people* is expressed in percentage points and *Poor population* is the fraction of poor population. *Number of mentions* is the number of times each budget item is mentioned in the text of the audits reports.

3.1. Corruption Prediction Dataset

Our data consists of audit outcomes and budget predictors. The corruption label $Y \in \{0, 1\}$ equals one for years of budgets where auditors found narrow corruption and equals zero for years in which the audits did not find narrow corruption.[9] For the model training, any municipality-years that were not audited are excluded because we do not have any labels. For the set of municipalities that were (randomly) audited twice, the second audit is excluded during model training because the samples are not comparable.

For the budget features, the only pre-processing performed is to impute missing values with the mean value for the associated variable.[10] The resulting matrix X of budget factors has 797 columns, corresponding to the budget fields, and rows corresponding to each municipality and year.

In principle, the set of predictors could be further enriched to include (potentially many) non-budgetary factors. Colonnelli et al. (2019), for example, form corruption predictions using local variables on economic activity, financial development, and human capital. We focus on budget variables because they proxy directly for the opportunities and behavioral choices of local government officials who may be engaged in fiscal corruption (e.g., Hessami, 2014; Cheol and Mikesell, 2018). For comparison, we train models using local demographic and economic variables instead of the budget, and also using the union of the predictor datasets.[11]

3.2. Machine Learning Approach

We want to learn a conditional expectation function $Y(X) = \mathbb{E}\{Y|X\}$ that provides a predicted probability, based on the publicly observed budget features X, that a municipality-year would be tagged as corrupt by an audit. Given the high dimensionality of the feature set (about 800 variables) relative to the number of rows in the data set (about 1500 audits), classical statistical models like ordinary least squares or logit will tend to over-fit the training sample, obtaining arbitrarily high training-sample

[9] See Section 2.2 above for definition of narrow corruption. We train the model using budget data that come from the years in a given term before the municipality is selected by the random lottery.

[10] We got similar results when experimenting with additional pre-processing steps, including adding missing indicator variables, standardizing variables, transforming variables as per capita, or imputing missing variables in different ways.

[11] Given their high dimensionality, the budget variables likely proxy for non-budget factors such as local economic activity. Consistent with this idea, the prediction model that includes both budget and non-budget variables gets a similar performance to the model with just budget variables (see Appendix Table B4).

R^2 but failing to extrapolate to held-out test samples from the same data-generating process (e.g., Hastie et al., 2009). In addition, if there are important non-linearities or interactions in the true function $Y(X)$, simple linear models will fail to capture them and produce errors even within the training sample. Given our goal of detecting potential corruption in municipalities that have not been audited – that is, a test set of municipalities – we must solve these problems and move past the standard statistical models used in applied economics (Mullainathan and Spiess, 2017; Athey, 2018).

These problems are in the domain of machine learning (ML). ML algorithms have two key benefits that are relevant to our problem. First, they add *regularization* – algorithmic adjustments to the model-training procedure that reduce over-fitting. To illustrate this point, assume that $Y(X)$ can be parametrized as a logit, $\hat{Y}(\theta) = (1 + e^{-X'\theta})^{-1}$. In the standard (non-regularized) logit, θ is learned by minimizing the binary cross entropy,

$$-L(\theta) = \ Y \log(\hat{Y}(\theta)) + (1 - Y) \log(1 - \hat{Y}(\theta))$$

summed across all observations. In the regularized logistic regression, instead, the cost objective is

$$\min_{\theta} \sum(-L(\theta) + R(\theta))$$

where $R(\theta)$ is a regularization term. This term is also called a *penalty* because it "penalizes" larger coefficients in the training objective, which reduces over-fitting. In the standard regularized logistic regression model, $R(\theta) = \lambda|\theta|_2$, where $|\theta|_2$ is the norm of the vector of learnable coefficients. Regularization with the 2-norm is called L2 regularization, also known as the Ridge penalty. Correspondingly, regularization with the 1-norm is called L1 or Lasso regularization. The penalty term λ, is a *hyperparameter* calibrating the regularization strength. Higher values of λ will reduce over-fitting by further shrinking the learned coefficients, but only up to a point. To find the best option for hyperparameters like λ, the researcher can use *grid search* – that is, testing out different values to find the one with the best out-of-sample fit. Regularized logistic regression has just one hyperparameter to select, but more sophisticated ML models have many hyperparameters.

The second benefit of ML is allowing efficient learning of complex functional relationships. With linear/monotonic models like OLS or logit, the researcher has to manually explore potential non-linearities or interactions, by adding polynomial and interaction terms. Given the infinite set of possible polynomial and interaction terms, estimating these more subtle functional forms requires substantial researcher effort and computation time. That problem has led to the development of ensemble models, such as random forests and gradient boosted machines, as well as feed-forward neural nets, which can efficiently learn arbitrarily complex functional forms (e.g. Hastie et al., 2009; Goodfellow et al., 2016).

In particular, a state-of-the-art binary classification model for tabular datasets is gradient boosted trees (Friedman, 2001; Hastie et al., 2009).[12] Gradient boosting models consist of an ensemble of decision trees that "vote" on the predicted outcome. Each decision tree iteratively selects informative variables (in our case, e.g., agricultural expenditures), splits on a value of that variable (e.g., $x > 100$) to predict the outcome better, branches off for additional splitting, and so on, until reaching a terminal node and an associated prediction ($\hat{Y} = 0$ or $\hat{Y} = 1$). With gradient boosting, additional layers of trees are gradually added during the training process to fit residuals and fix errors in the initial layers. A number of hyperparameters, such as the number of trees, and their depth, can be selected to calibrate the regularization level.

A popular implementation of gradient boosting is XGBoost ("eXtreme Gradient Boosting"; Chen and Guestrin 2016). Besides being optimized for fast training, XGBoost has a number of computational adjustments to improve out-of-sample fit. Feurer et al. (2018) systematically compared XGBoost to many other classifiers, including a sophisticated automated ML system, and found that XGBoost consistently performed best on classification tasks with tabular data (see also Grinsztajn et al., 2022). Hence, we take XGBoost as our preferred model for predicting the corruption label based on budgetary inputs.

3.3. Model Training

ML models are intended to produce good predictions in unseen data – in our case, municipalities not selected for audit. Given the large scope of the paper, we employ three different training techniques to balance the trade-off between model performance and the time/cost of training the model.

The first approach is meant to maximize the model performance and generate predictions that would be used for the main analysis presented in the paper. Formally, we

[12] For non-tabular data, such as images, text, or audio, neural nets are often preferred (Goodfellow et al., 2016).

use a technique called *nested cross-validation*. This method involves performing multiple cross-validation procedures within each fold of an outer cross-validation loop. In this approach, the data is first divided into five equal parts or folds, and the model is trained and evaluated using each fold as a test set and the remaining four as a training set. For each of the five test sets, we perform another round of cross-validation using the remaining four folds as the training set. This inner cross-validation is used to select the best hyperparameters for the model. We perform a *grid search* over a range of hyperparameters, and the hyperparameters that produce the best performance on the validation set are chosen. In our case, we search over 288 hyperparameter grid cells.[13] Once the best hyperparameters are identified, we train a new model on the entire training set and evaluate it on the test set. This process is repeated for each of the five folds, resulting in five separate estimates of the model's performance. Nested cross-validation can be computationally expensive, but it provides a more robust and reliable estimate of the model's performance than other methods, such as simple train-test splits.

For the second approach, we use a standard 80-20 train-test split, meaning that 80% of the data is used for training and 20% for testing, without performing any pre-imposed hyperparameters tuning.[14] Importantly, we repeat this split 1,000 times, randomly shuffling the data and dividing it into new training and test sets. By repeating the process many times, we can obtain a more reliable estimate of the model's generalization performance.

Finally, in the third approach, we follow the second approach but hold out a fixed 20% such that the 80-20 train-test split is done on 80% of the full sample. This approach allows for a robust and unbiased assessment of the model's ability to make accurate predictions on unseen data.

[13] These include the following. L1 and L2 regularization penalties on the learned parameter values (each selected from $\{0.1, 0.5, 1, 2\}$), the max depth of the constituent decision trees (selected from $\{5,10,20\}$), the learning rate (selected from $\{0.1,0.5\}$), and an additional regularization constraint specifying a minimum threshold for the size of the decision tree terminal nodes (minimum child weight, selected from $\{1,3,5\}$). Hence, $4 \times 4 \times 3 \times 2 \times 3 = 288$ hyperparameter grid cells to search. See descriptions of these options and selected values in Appendix Table A2. Finally, we also use early stopping with patience of 10 training epochs.

[14] XGBoost provides default values for many hyperparameters, which are designed to provide reasonable performance across a wide range of tasks.

3.4. Model Performance

We evaluate our set of models by their scores on a set of standard classification metrics in the held-out test data. These metrics, reported by row in Table 2 Panel A, describe how well a model trained on the budget accounts can replicate the auditing agency's judgments about fiscal corruption. First, the most straightforward metric is *accuracy*, which gives the proportion of test-set observations for which the machine-predicted label matches the true label. A naive guessing model that chooses the modal category (not corrupt), shown as a weak baseline in Column 1, would obtain accuracy = 0.58. Second, we report *AUC-ROC* (area under the receiver operator characteristic curve), another standard metric in binary classification. AUC-ROC, which takes values between 0.5 (random guessing) and 1.0 (perfect accuracy), can be interpreted as the probability that a randomly sampled corrupt municipality is ranked more highly by predicted probability of corruption than a randomly sampled non-corrupt municipality.[15] Because AUC-ROC requires a ranking of predicted probabilities, it is undefined for naive guessing (Column 1). Third, we report *F1* for the corrupt class, defined as the harmonic mean of precision (proportion audited-as-corrupt within the set predicted-corrupt) and recall (proportion predicted-corrupt within the set audited-as-corrupt). F1, ranging from 0.0 (zero recall, zero precision, or both) and 1.0 (perfect accuracy for the corrupt class), penalizes both false positives and false negatives.

Columns 2 through 4 of Table 2 show the predictive performance for a set of additional baselines. For Column 2, we train ordinary least squares (OLS), or non-penalized linear regression, dropping multi-collinear predictors. Lasso (Column 3), perhaps the most familiar machine learning model to economists (e.g., Belloni et al., 2014), is a linear regression model but adds an L1 penalty that penalizes larger coefficients and outputs a sparse model.[16] In Column 4, we use penalized logistic regression. As described in more detail in Section 3.2 above, penalized logistic is a binary logit model with regularization.[17] Each table cell reports the average test-set performance metric across the five cross-

[15]Differently from the other metrics, AUC-ROC evaluates the full distribution of predicted probabilities, rather than relying on predicted labels that depend on a decision threshold. For AUC-ROC, outputs from the linear probability models (OLS and Lasso) are capped between zero and one.

[16]For both OLS and Lasso, the predicted probabilities for Y might be below zero or above one, but a decision threshold of 0.5 is used for assigning a predicted label.

[17]The logistic model from Column 4 includes both a Lasso (L1) and Ridge (L2) penalty – also known as Elastic Net regularization. For both Lasso and Logistic, the penalty is selected by cross-validation grid search in the training set. All three of the alternative baselines are implemented using the python package scikit-learn (Pedregosa et al., 2011).

Table 2: Out-of-Sample Metrics for Predicting Corruption

		Optimal (5-fold NCV)				Bootstrap	
	Guessing	OLS	Lasso	Logistic	XGBoost	XGBoost	XGBoost
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Accuracy	0.580	0.502	0.500	0.575	0.720	0.700	0.683
		[0.413-0.566]	[0.481-0.527]	[0.465-0.619]	[0.699-0.739]	(0.013)	(0.011)
AUC-ROC		0.519	0.486	0.559	0.781	0.755	0.736
		[0.473-0.569]	[0.429-0.535]	[0.475-0.600]	[0.759-0.811]	(0.014)	(0.011)
F1	0.000	0.507	0.452	0.477	0.627	0.612	0.592
		[0.300-0.582]	[0.252-0.527]	[0.306-0.550]	[0.578-0.657]	(0.018)	(0.015)

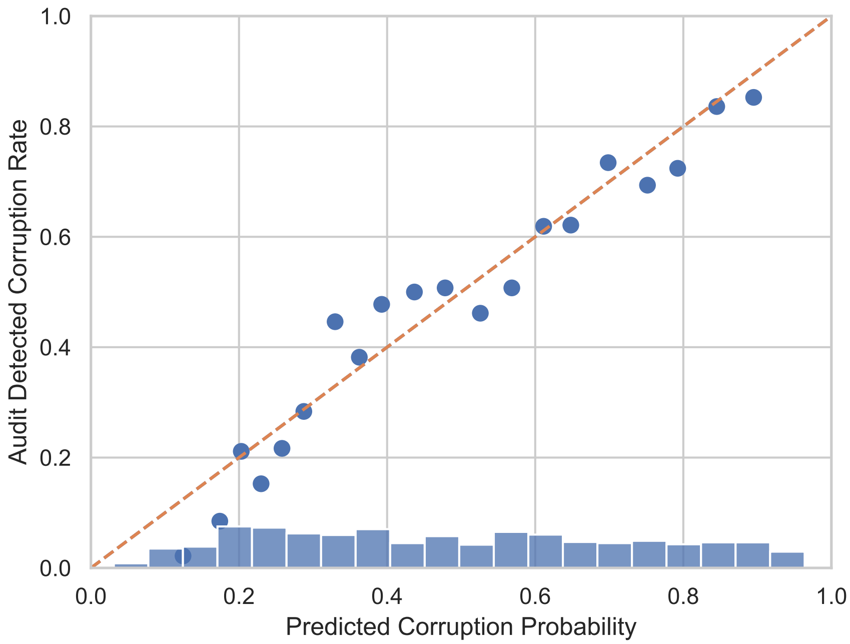
Notes: Column (1) reports the performance metrics (by row) from naively guessing the modal category ("not corrupt"). Columns (2) to (5) report the mean for the indicated test-set performance metrics (by row) across the five model runs produced using separate training-set folds with hyperparameters tuning. Columns (6) and (7) report the mean for the indicated test-set performance metrics (by row) of 1,000 bootstrapped trained models, where we keep the default value for the hyperparameters. In column (6), the performance is evaluated each time on the 20% test set of the specific iteration, while in column (7), the performance is evaluated on a complete hold-out 20% validation set. In brackets, there are the minimum and maximum of the metrics, while in parentheses, the s.d. produced from 1,000 bootstrapped trained models. Columns indicate the machine learning model used.

validated model runs, with the standard error of the mean in parentheses. OLS and Lasso (Columns 2 and 3) improve on the zero F1 from guessing (Column 1) at the expense of lower test-set accuracy. Logistic regression (Column 4) is better than the linear models on all metrics, yet still similar in overall accuracy to guessing.

With XGBoost (Column 5), meanwhile, we report a substantial lift in all metrics over all baselines. The average test set accuracy for the predictions across five nested folds is 0.720, with the minimum accuracy being 0.699 and the maximum 0.739 across folds. For AUC-ROC, the average is 0.781 (with min = 0.759 and max = 0.811), while for F1, the average is 0.627 (min = 0.587, max = 0.657).[18] Finally, in column (6), we report the mean and the standard deviation of accuracy metrics from 1,000 different predictions, where we randomly select the sample splits, train the models on 80% of the sample and test the accuracy on the remaining 20%. In column (7), instead, we report the mean and the standard deviation of accuracy metrics from 1,000 different predictions where the accuracy is tested in held out 20% of the sample. The training is done on 80% of 80% of the total sample. In both cases, as explained in the previous section, we run XGBoost without hyperparameter tuning, which would be too costly (time and computing cost).

[18]The model performance is similar to that in other recent papers using machine learning for economic analysis of human decisions. For example, the gradient boosted ensemble in Kleinberg et al. (2018) trained to predict criminal recidivism obtains an AUC-ROC of 0.707. The model in Mullainathan and Obermeyer (2019) trained to predict a cardiac medical intervention obtains AUC-ROC = 0.731.

Figure 1: Predicted Probabilities in Test Set are Well-Calibrated to Audited Corruption Rates



Notes: Calibration plot showing audit-measured corruption rates (Audit detected corruption rates, indicated by blue marks on the vertical axis), binned by predicted corruption probability (horizontal axis), in held-out test set. Blue histogram shows the density of the predicted corruption probability. Dashed 45-degree line (in orange) demarcates perfect calibration.

Interestingly we find that, as expected, there is a reduction in the accuracy level, but this is not excessively large. This is coherent with the internal step XGBoost does by default for hyperparameter tuning. Overall we confirm the increase in performance while using XGBoost, suggesting a nonlinear, interactional relationship among the predictors that the tree ensemble is better able to learn.

Beyond the binary classification metrics, another relevant metric for model performance is the *calibration* of the predicted class probabilities. That is, do the predicted corruption probabilities faithfully rank the municipalities by risk and reproduce the correct distribution of corruption rates? Figure 1 shows that the model is well-calibrated: In each of the 20 bins in the held-out test set, the predicted corruption probability in that bin corresponds closely to the audited-based corruption rate of test-set observations in that bin. Appendix Figure A1 provides additional plots, showing good calibration with alternative train/test sampling and for the alternative corruption measure.

Appendices A and B report additional evaluations of the prediction task. First, to help visualize the distribution of predictions, Appendix Table A3 shows the confusion matrices for the test-set predictions. For XGBoost, we can see good precision and recall across categories. The confusion matrices for OLS, Lasso, and logistic regression show that the linear models tend to produce many false positives (not-corrupt municipalities are often labeled as corrupt), yet fewer false negatives. Note that in our performance evaluation so far, we have treated false positives and false negatives symmetrically in terms of their policy costs. But from the perspective of auditing costs, it is not clear that they are substitutable. In practice, if false positives are valued differently from false negatives, the XGBoost model can be calibrated to account for that.

Second, we focus on the municipalities that have been audited twice and see if our prediction model can reproduce within-municipality changes in corruption over time. To that end, we regress the change in audited corruption against the change in predicted corruption, adjusting for audit year fixed effects and demographic characteristics. Appendix Figure A2 shows that there is a significant positive relationship in this regression ($p = .07$ with FE and controls). This within-municipality validation speaks to the usefulness of our measure in empirical tasks, where one would like to be able to examine changes in corruption over time.

Third, in Appendix Table B4 we report performance metrics with an alternative sampling approach and with an alternative corruption outcome. In Columns 1-3, we apply random splits between training and test set by municipality, instead of by municipality-year, which allows us to compare the model performance using budget factors, fixed demographic factors, or both. The alternative sampling approach obtains comparable performance to our baseline model and shows that a model trained using just demographic information is less accurate than a model using budget information. Finally, in Columns 4-7, we show the model performance in predicting the alternative corruption variable from Avis et al. (2018), obtaining even higher accuracy than the baseline corruption variable (AUC-ROC = 0.900 for XGBoost).

3.5. Interpreting the Predictions

Gradient boosted machines, like all ensembles and other sophisticated machine learning algorithms, are black boxes. At the end of model training, we have a dense forest of decision trees. With 797 variables being input into those trees, and thousands of splitting nodes within the forest (Appendix Table A2), it is difficult to tell how the model is making its predictions. In this subsection, we use model explanation methods to understand better how the model works.

The applied machine learning literature has discovered an advantage of gradient boosted machines that compensate for their basic lack of interpretability (e.g., Hastie et al., 2009; Molnar, 2020). One can rank the input variables by their *feature importance*, computed as the number of times the model "uses" that variable in the sense that one of the constituent decision trees splits on it. Moreover, the important features can be seen as *pivotal* in the sense that they are the most useful variables for predicting the outcome, even among clusters of highly collinear predictors.

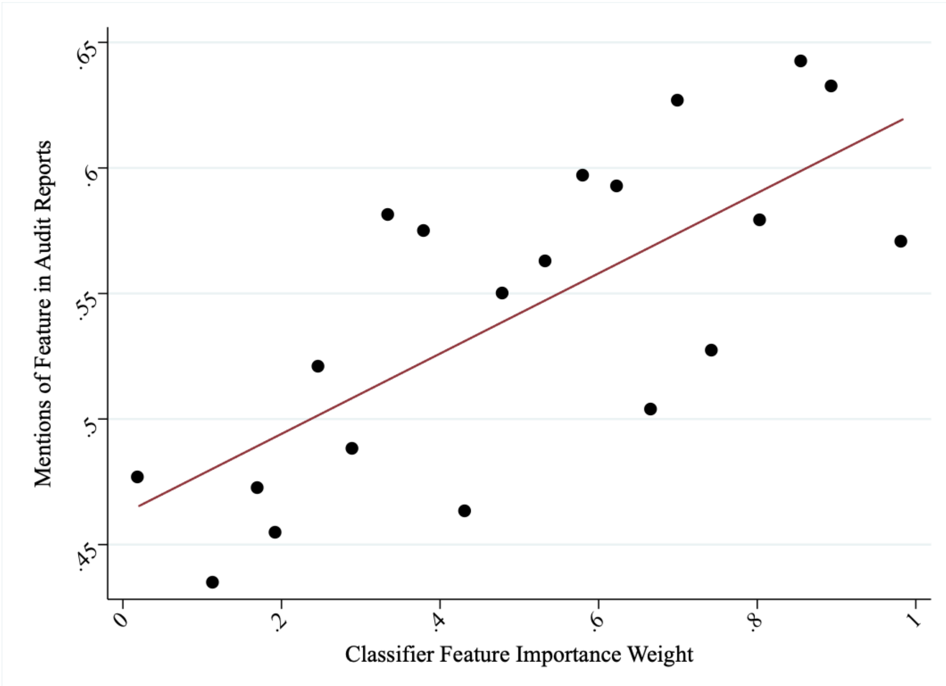
Here we use the feature importance ranking to get some insight into how our corruption detection model makes its predictions. After model training, we have feature importance scores for each of the five cross-validated models. We average the scores across folds and then rank the most important features. From the feature importance scores, we learn immediately that our dataset contains many noise predictors. Out of the 797 variables input to the ensemble, 342 are ignored by the ensemble across all five folds.

Still, that leaves 454 variables the tree ensemble finds useful for predicting corruption, implying a high-dimensional learned function. Appendix Table C5 lists the most important predictors, ranked by the feature importance score. Some of the top-ranked factors have been mentioned in the previous literature as being related to corruption, including: (4) expenditures on transportation (Hessami, 2014), (31) municipal debt (Liu et al., 2017), (7) real estate and construction (Kyriacou et al., 2015), and (15) public health services (Machoski and de Araujo, 2020).

For these and the other important variables, one could think of (perhaps many) ways they could contribute to corruption. Many of these stories would be unsatisfying, however, and they would likely be incorrect because the feature importance ranking does not identify direct links between a budget factor and corrupt behavior. The important features could be either positively or negatively correlated with the predicted corruption. Moreover, they could be important through a non-linear relation or through interactions with other variables.

Appendix C.1 provides some supporting analysis to illustrate the non-monotonicities learned by the tree ensemble. In the spirit of a partial dependence plot (e.g. Molnar, 2020), we perturb each feature individually and record the predicted change in corruption risk. Looking at the distribution of the changes, we can infer a highly non-linear, interacted functional form. Of the 454 informative variables, only 39 variables have a

Figure 2: Model Feature Importance and Mentions in Audit Report Texts



Notes: Binscatter diagram for percentile that a budget feature appears in the municipal audit reports (vertical axis) against binned percentiles of feature importance weights for each feature (horizontal axis). Pearson's correlation is 0.24. The regression coefficient is 0.159 with $p = .001$ (robust standard errors).

monotonic relationship with corruption risk. In the set of important features (according to the feature importance score), none of them have a monotonic relationship with corruption risk. Instead, perturbing budget features could have either a positive or negative relation to predicted corruption depending on the status quo values of the variables.

Given these pivotal non-linearities, a further qualitative discussion of the highest-ranked features would obtain limited insights. As a more comprehensive alternative, we ask whether the corruption-related features identified by the gradient boosted machine are also identified as corruption-related by the Brazilian auditing agency. To do that, we look for mentions of these items in the best available place – the text of the published audit reports.

As described in Appendix C.2, we obtained the text of over two thousand reports for the time period of our analysis (containing over 55 million words) and counted the mentions of each budget item in the corpus of reports. Appendix Table C6 shows the budget components with the highest number of mentions in the reports. We produced a dataset at the feature variable level, containing the percentile rank in the model feature importance score and the percentile rank in the audit-report mention count.

To compare these values, Figure 2 plots the audit mention percentiles against the feature importance percentiles. We see a strong positive relationship that is statistically significant in a univariate regression ($p = .001$). Our classifier, trained on the budget accounts with just corruption labels, identifies as important the same budget features that tend to be mentioned in the audit report documents.[19] These validation results support the view that our measure captures activities that are indeed related to corruption.[20]

3.6. Measuring Corruption in Non-Audited Municipalities

Now that we have some confidence on the accuracy of our model in held-out data, we can extrapolate it to the set of municipalities that have never been audited. Indeed, our approach allows us to measure corruption for all Brazilian municipalities and all years from 2001 to 2012. We apply the trained model directly to the budget variables and obtain a predicted probability $\hat{Y}_{it}$ that municipality i's budget at year t would be audited as corrupt.[21]

Figure 3 provides a visualization of the sample of only audited municipalities (Panel a) and how it compared to the sample of municipalities that we can analyze when using our predicted measure of corruption (Panel b). The map illustrates quite clearly the additional information produced by the machine learning method. With the machine predictions, we can then analyze corruption in municipalities (and years) regardless of whether they have been audited.

4. Empirical Applications

This section replicates and extends existing evidence from the literature on corruption in Brazil. This exercise has two purposes. First, it provides checks on the internal validity

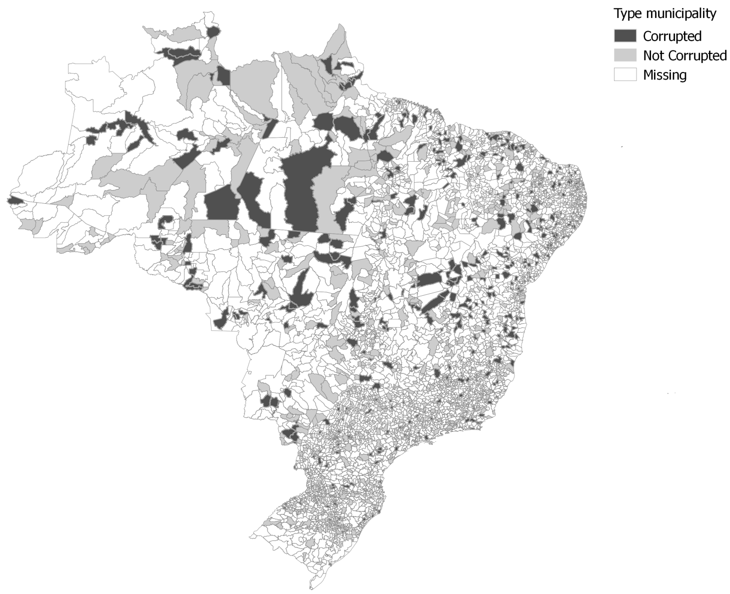
[19]Notably, we obtain similar results if we consider only those model features that tend to have a positive effect in the perturbation analysis discussed in Appendix C.1.

[20]Note that the outliers in Figure 2 – i.e., the cases in which the audited level of corruption is low, but the associated report mentions budget variables with a high feature importance ranking – could themselves be informative for additional inductive analysis. In particular, they could potentially be used to detect possible oversights by the human audit.

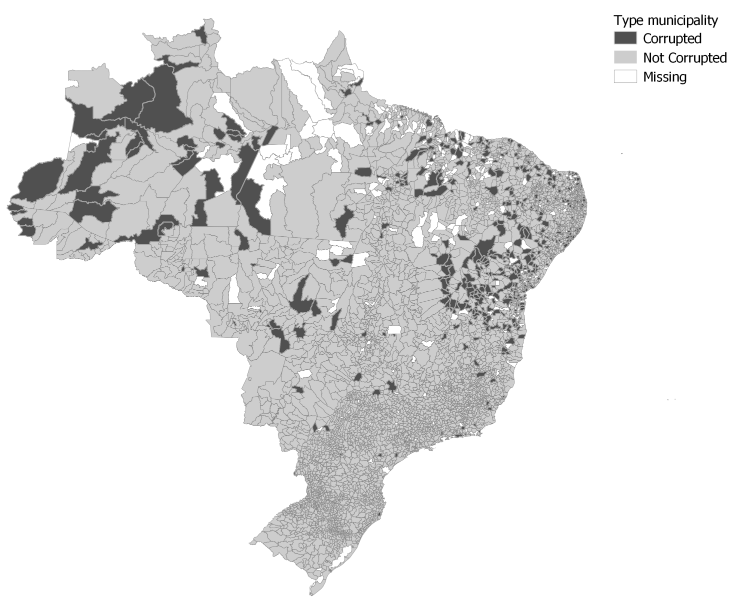
[21]Note that we have five corruption prediction models, each trained on the five outer 80% training folds as described in Section 3.3 above. Because audits are randomly assigned, these non-audited municipalities can be seen as "held-out test set" municipalities and each of the five models provides a clean prediction. Hence, we form the predicted probability from all five trained models and take the average. We will also use this to assess the sensitivity of the predictions to training-sample variation.

Figure 3: The Geography of (Predicted) Corruption

(a) Actual Corruption



(b) Predicted Corruption



Notes: Actual (Panel a) and predicted (Panel b) corruption by municipality, using budgets from 2004. A municipality is predicted to be corrupted if mean prediction is >0.5.

of our synthetic measure of corruption – that is, we can check whether it responds to causal treatments the same way as auditor-measured corruption. Second, we extend previous results using the larger sample of municipalities and the time variation of our corruption measure.

4.1. Revenue Shocks and Corruption

Our first empirical application is to use the new synthetic measure of corruption to analyze the effect of revenue shocks on corruption, as done in Brollo et al. (2013). That paper shows that a windfall of public revenues increases rent-seeking by the public administration (as measured by a subsequent surge in corruption). Here we replicate and extend those findings using our ML-predicted measure of corruption.

Federal transfers are the largest single source of municipal revenues in Brazil (around 40% of the total budget). The amount transferred through this *FPM* program (*Fundo de Participação dos Municipios*) depends on exogenous population thresholds, where municipalities in the same state and in a given population bracket are statutorily prescribed the same transfer amounts.[22] Given imperfect compliance, Brollo et al. (2013) use a fuzzy regression discontinuity design, instrumenting actual transfers with statutorily prescribed transfers.[23]

Appendix D provides the details on the estimating equations. We adapt the design from Brollo et al. (2013), using the ML-predicted corruption rather than audit-detected corruption. Besides replicating the former paper's results, we can use the ML-predicted corruption for non-audited municipalities and analyze a larger and potentially more representative sample. Therefore, our exercise is also providing a test for the external validity of their results.

Table 3 reports the regression results analyzing the effect of municipal revenue shocks on predicted corruption. For each specification, we report estimates from three samples, corresponding to columns: (1) cities that received an audit (same as Brollo et al. (2013)), (2) all cities, and (3) cities that were never audited. In Panel A, we replicate a strong first-stage effect; prescribed transfers positively affect actual transfers in all samples. In Panels B and C, we find positive and significant coefficients when estimating the

[22] Appendix Table D7 shows the corresponding population brackets and transfer multipliers. We follow Brollo et al. (2013) in constructing the sample.

[23] Due to imperfect compliance, the statutorily prescribed transfers do not perfectly determine the amounts actually transferred. This imperfect compliance is due to many factors (*e.g.,* municipalities splitting, manipulation in population figures).

Table 3: Effect of Revenue Shocks on (Predicted) Corruption

	Audited cities (1)	All cities (2)	Non-audited cities (3)
Panel A. First Stage			
Prescribed transfers	0.6805***	0.6909***	0.6996***
	(0.0205)	(0.0233)	(0.0230)
Panel B. Reduced Form			
Prescribed transfers	0.0037***	0.0039***	0.0038***
	(0.0009)	(0.0003)	(0.0003)
Panel C. 2SLS			
Actual transfers	0.0055***	0.0057***	0.0054***
	(0.0014)	(0.0005)	(0.0005)
N. Observations	1115	5808	4693

Notes: Effects of FPM transfers on (predicted) corruption measures. Panel A reports the estimates of the first-stage analysis – the dependent variable is *actual transfers*. Panel B reports the estimates of reduced-form analysis – the dependent variable is *predicted corruption*. Panel C reports the estimates of the 2SLS estimates, the dependent variable is *predicted corruption* and *actual transfers* is instrumented with *prescribed transfers*. Column headings indicate the sample of municipalities included. All regressions controls for a third-order polynomial in normalized population size, term dummies, and macro-region dummies. Robust standard errors clustered at the municipal level are in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

reduced-form or two-stage-least-squares models. The coefficients and standard errors are stable across samples. Notably, the magnitude of the standardized reduced-form coefficient is about four-fifths the size of that estimated by Brollo et al. (2013), and our 95% confidence interval contains the original coefficient. Thus even the magnitudes of empirical estimates using ML-measured corruption are comparable to those using auditor-measured corruption.

We conducted a series of checks to probe the robustness of these results. First, to show that our results are not driven just by a larger sample, we replicate the main analysis on four random samples of 1,115 municipalities, the sample size of the original analysis by Brollo et al. (2013) (Appendix Table D9). The coefficients show some variation, but they are always positive and statistically significant. Second, we show that the instrument is not correlated with the error of the prediction model, defined as the difference between the audit-based corruption level and the predicted one (p-value=0.190). This null is helpful because it suggests that the machine learning model's errors are not responding to the instrument – that is, the correlated factors besides corruption contributing to our prediction are not affected directly by revenue transfer shocks. Third, we show in Appendix Table D10 Column 1 that there is no revenue-shock effect on a corruption prediction formed with a model trained on time-invariant municipal demographic characteristics (similar to Colonnelli et al.'s (2019)). This placebo test is reassuring because the model trained on demographics does not contain budget information, so revenue shocks should not have an effect. Since our budget-trained model does have an effect, the placebo test provides additional confidence that our main model is not forming corruption predictions based on spurious correlations with demographics. Fourth, we formed predictions from our baseline model while permuting randomly the FPM transfer variable, which could be mechanically shifted by the revenue shocks instrument. The effect of revenue shocks is the same (Appendix Table D10 Column 2). Finally, in Table D11 Panel A, we report the results of a bootstrap approach, where we execute the analysis 1,000 times, using each time a different prediction, and report the mean coefficient, the standard deviation, and the confidence interval.[24] We report additional evidence in Panel B and C of Table D11 by running a hold-out exercise where the predictor is fitted only on half of the municipalities and the estimation performed on the other independent half. In this

[24]These are the predictions we produced with the second method we described in Section 3.3 whose predicting performance are reported in column 6 of Table 2.

case, the interpretation of the standard error estimates would then be conditional on the first half. We apply these approaches to account for 1) potential limitations in the reliability of the standard errors because we are using predicted variables in a regression framework and 2) to improve the generalizability of the results, as one might argue that using only one prediction for each observation, from the 5-fold cross-validation method, could be affected by the training sample used. Overall, we find the results to be generally confirmed.

Together with the validations from above (in terms of capturing within-municipality changes in corruption and with the important features being mentioned in the audit reports), these causal results give confidence about the usefulness of our measure for empirical tasks. These results go against an alternative interpretation that there are exogenous conditions that make it easier to engage in corruption – such as the presence of many infrastructure projects – and that our model picks up budgetary responses to those conditions. If the model were picking up (mostly) non-budget factors, then we would see no response to the causal treatments.

4.2. Effect of Audits on Corruption

The next empirical application uses our predicted measure of corruption to analyze the effect of auditing on subsequent corruption. This analysis complements Avis et al. (2018), who explore the same research question using the set of Brazilian municipalities that were (by random draw) audited twice. Because of the longitudinal nature of our dataset, we can capture dynamic effects and condition estimates on pre-audit levels of corruption. Further, our effects are identified by the much larger sample of municipalities that got audited at least once (rather than twice).[25]

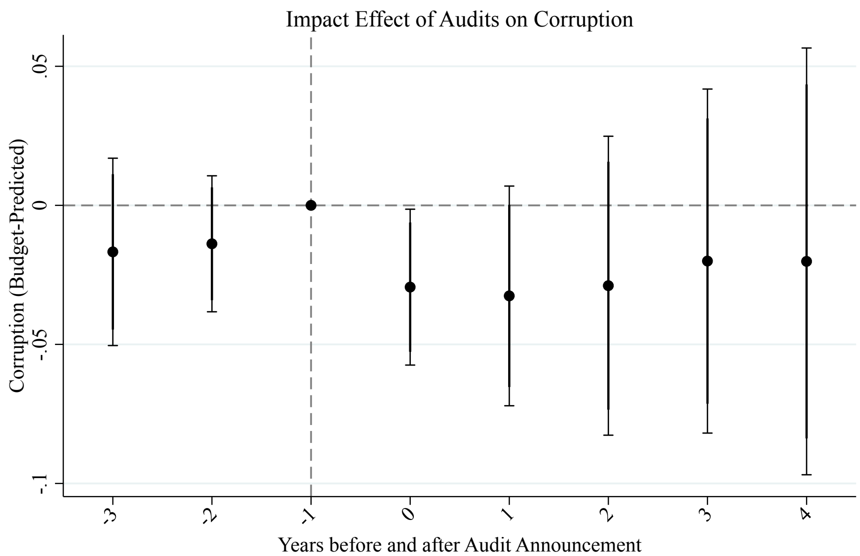
We use a standard panel event-study framework, with details and notation enumerated in Appendix E. Identification comes from randomness in the timing of audit selection. The sample includes 1,479 municipalities that have received an audit in the time period under analysis.

The event-study estimates are visualized in Figure 4 Panel (a). We can see that already in the year of the audit, there is a sharp and statistically significant drop in predicted corruption. This effect persists for another year before becoming weaker and

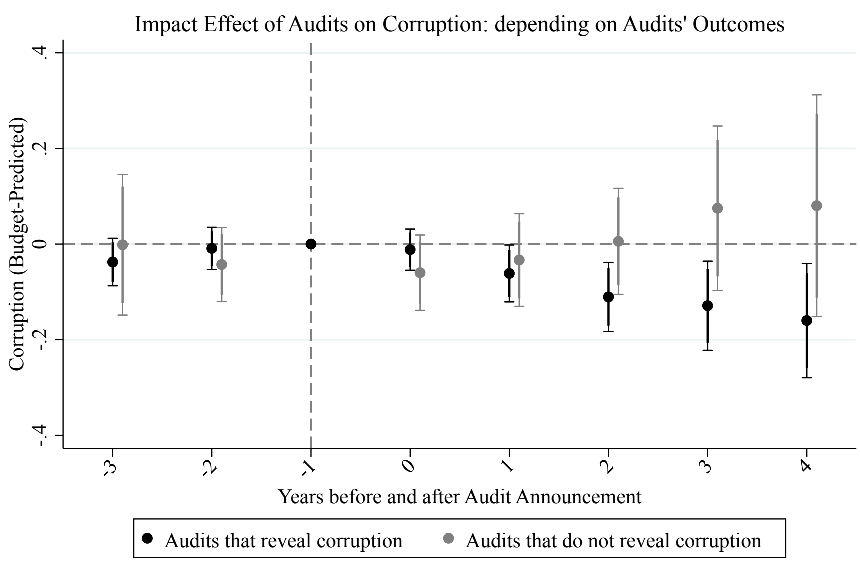
[25] Bobonis et al. (2016) study a similar research question in Puerto Rican municipalities. The authors focus on (non random) audits of municipal accounts, finding that audits do not persistently reduce corruption in that case.

Figure 4: Dynamic Effect of Audits on Fiscal Corruption

(a) All municipalities



(b) Corrupted vs. Non-corrupted



Notes: Event study estimates for dynamic effect of audits on budget-predicted corruption. Error spikes give 95% confidence intervals, with standard error clustered by state. Top panel: all audits; bottom panel: audits that found corruption (in black); audits that did not find corruption (in grey). For the analysis on all audits leads are jointly insignificant (p-value=0.63) and lags are jointly significant (p-value=0.005). For the analysis on audits that found corruption leads are jointly insignificant (p-value=0.115) and lags are jointly significant (p-value=0.013).

not statistically significant. Panel (b) reports event-study effects for the subsets of audits that find clear corruption (black points) and those that do not find corruption at all (grey points).[26] These trends look quite different. When corruption is discovered (black points), there is a much larger negative effect and the effect is persistent across subsequent years. In contrast, when the audit does not find any corruption or irregularities (grey points), there is no effect.[27]

We report supporting analyses in Appendix E. First, we check whether the results are driven mechanically by a budget reduction due to financial penalties imposed on these municipalities; that does not appear to be the case, as the occurrence of an audit does not affect expenditures per capita), and the main results do not change if we control for expenditures (per-capita) in the main regression specification. Second, we test the political accountability channel by checking whether the effect of the audit is stronger when local political competition is high. We find that the answer is yes: In cities where the mayor has been elected with a small margin of victory, the impact of the audit is larger. Finally, similar to what we provided for the previous replication, In Table E13, columns 1-3, we used bootstrapping with 1,000 iterations and different predictions each time to address potential standard errors and improve generalizability. Additional tests using a hold-out approach were carried out in columns 4-9. The results generally confirm our findings.

These results show how ML-predicted corruption measures can be used for further analysis and to provide new and relevant findings. In the event-study analysis, we go beyond Avis et al. (2018) and show more granular effects of audits on corruption, by looking over time and by the outcome of the audit. These additional insights were not available with the existing data based only on audits, given the small sample size.

5. Using Machine Learning to Guide Audit Policy

This section outlines a policy simulation for how corruption policy could be supported using machine predictions from our model. We start with a baseline targeting

[26] The former group includes those cities in which the audits discovered a positive amount of corruption (measured with the variable narrow corruption), while the latter group includes those municipalities in which the audit did not find any type of corruption.

[27] The numerical estimates reported in Appendix Table E12. These findings are robust to correction for the two-way fixed effects designs with multiple treatment cohorts (Callaway and Sant'Anna, 2021), as shown in Appendix Figure E5.

policy based on predicted corruption risk, showing that targeted audits can detect more corruption than random audits. Second, we consider the issue of political bias in the risk scoring algorithm towards different mayor party affiliations, and analyze the performance of a politically neutral targeting policy. Third, we discuss additional caveats and complications with implementing a targeted audit system.

5.1. Targeted Audit Policy

To set the stage for targeted audits, let's first consider the performance of the status quo random audit system. Limiting to places with mayors from the six main political parties (PMDB, PSDB, DEM, PP, PTB, and PT), there are 5341 municipalities in the dataset. In that subset of the data, there are about 150 audits per year, with 69 corrupt municipalities detected for the average year of audits. The resulting audit probability (and therefore detection rate, regardless of whether a municipality is corrupt) is roughly 0.028.

To improve these numbers, let's consider a policy designed to maximize the number of detected corrupt municipalities in a single round of audits. Note that taking a dynamic perspective and trying to detect the most corruption over multiple rounds of audits would involve a different policy. Similarly, our baseline approach would not necessarily optimize other objectives, such as maximizing the deterrent effect of the policy, or more generally minimizing total corruption. We return to some of these issues below.

We start by ranking the municipalities by corruption risk. That is, we apply the baseline gradient boosting model to the budget data for each municipality i from year t to produce $\hat{Y}_{it}$. Then for each year t, we have an ordinal ranking of the municipalities (1 through 5341) by predicted probability of corruption. The proposed policy is to replace random audits with audits targeted by predicted corruption risk. Rather than sampling 150 municipalities uniformly from the distribution, the agency could audit the top 150 with the highest $\hat{Y}_{it}$. These are municipalities that have a level of corruption probability higher than 0.845 in the average year. The policy is illustrated visually in Appendix Figure F9.

We assess the targeting policy by simulating it on our dataset and measuring its effectiveness in terms of detecting and deterring corruption. Effectiveness is measured by comparing to the status quo lottery. For completeness, we simulate policies using the whole population of municipalities (treating our corruption predictions as if they were observed levels) as well as limiting to the municipalities in our audited sample (using

Table 4: Performance Metrics for Targeted Auditing Policies

	Status Quo (Lottery)		Targeted Audits		Fair Targeting
	(1a)	(1b)	(2a)	(2b)	(3)
Evaluation Sample	All (Sim)	Audited	All (Sim)	Audited	All (Sim)
Corruption Rate, if Audited	0.486	0.458	0.871	0.883	0.868
↪ Ratio over Random Audits			[1.788]	[1.927]	[1.783]
Audit Rate, if Corrupt	0.036	0.036	0.076	0.119	0.074
↪ Ratio over Random Audits			[2.714]	[4.246]	[2.644]

Notes: Metrics for comparing the effectiveness of audit policies: random audits (columns 1a and 1b), targeting audits to the municipalities with the highest corruption risk (columns 2a and 2b), or targeting audits with highest corruption with the constraint that all political parties are audited at the same rate (column 3). "Corruption Rate, if Audited" is the share of audited municipalities where narrow corruption is detected, for the respective policy. "Ratio over Random Audits" is the "Corruption Rate, if Audited" value for the indicated policy, divided by that value under random audits. "Audit Rate, if Corrupt" is the expected probability of being audited, if corrupt, under the various policies. In Columns 1a and 2a, "All (Sim)" indicates that statistics are produced using the full sample of municipality-year observations, treating the machine-predicted corruption probabilities as reflecting true realized corruption rates. In columns 1b, 2b, and 3, "Audited" reflects that statistics are produced in the sample of audited municipality-terms, using the true audit results. Thus, Column 1b reports the observed rates in the data. In the other columns, statistics give the mean predicted corruption risk.

the results of the real audit for evaluation).

Table 4 reports statistics for evaluating the targeted audit policy. The predicted corruption rate in the whole sample (Column 1a, with "All (Sim)" for "simulated") is 0.486. This predicted rate is quite close to the observed corruption rate of 0.458 from the audited sample (Column 1b: "Audited").[28]

Second, Columns 2a and 2b report statistics on the expected outcomes of the targeting policy, where Column 2a (like Column 1a) simulates outcomes using predicted corruption rates from the whole sample, while Column 2b (like Column 1b) evaluates outcomes using corruption rates from the audited sample.[29] In both simulation samples, we see substantial policy gains. In the full sample (Column 2a), the detected corruption rate of 0.871 is almost double (1.788×) that of the status quo policy (0.486). In the audited sample (Column 2b), the gain is even larger (1.927×). In turn, a higher corruption

[28]Recall that the base rate in Section 3 was 0.422, rather than 0.458. The difference here is that we simplify the dataset to have a single observation per audit, and also limiting to the cities with mayors from the six main political parties. In Section 3, the dataset included all fiscal years checked by the audit, which slightly changes the mean corruption rate.

[29]The statistics from Column 3 come from a politically neutral audit targeting policy. We revisit these numbers in the next subsection.

detection rate also means a higher audit rate conditional on being corrupt. As seen in the third row, the conditional audit rate under targeting is about 0.076 (Column 2a) or 0.119 (Column 2b). That indicates an even more dramatic change in policy effectiveness, significantly more than double (2.714× in the full simulated sample; 4.246× in the true-audited sample).

Overall, targeting makes a big difference in policy effectiveness. For the same number of implemented audits (and presumably the same allocation of government resources), the targeted approach detects about 80-90% more corrupt municipalities. Because successful audits reduce corruption (see Section 4.2 above), the targeted policy would also reduce the frequency of corrupt activities in Brazil. To achieve the same number of corruption detections as the status quo policy (69 municipalities), only 79 targeted audits are needed, down from 150 random audits. This decrease of 47%, or 71 audits per year, could imply a significant reduction in administrative expenditures.[30]

5.2. Issues with Dynamic Implementation

The policy simulation considered so far has a single round of targeted audits. At least in the short run, multiple targeted audit rounds would be possible and effective if they used the public finances data from before the first audit. But a more reasonable measure of success of targeted audits is not to identify a high number of corrupt municipalities but rather to deter corruption in the future. Subsequent to the first round of audits, the broader effect of the policy depends on how agents will learn and adapt to it. The existing model, when applied to post-targeting accounting data, may produce errors that would favor the more savvy mayors.[31] On the other hand, the machine targeting might actually spur additional corruption for municipalities who realize they have low predicted corruption. Therefore, the simulated gains from a static environment likely represent an upper bound on the longer-term deterrent effect, absent additional enhancements to the system. Further work could focus on forecasting model performance over time, including through adaptive systems that trade off exploring versus exploiting information

[30] The qualitative magnitudes of these improvements is not sensitive to the details of the specification. We can simulate similar policy improvements using different train/test split approaches, using the measure of corruption from Avis et al. (2018), or adding additional non-budget features to the model.

[31] Still, it could reduce the net marginal benefit of corrupt activities by increasing the expected cost of corrupt fiscal actions that are not easily substitutable. Our setting is not amenable to the "manipulation-proof machine learning" method from Björkegren et al. (2020), which requires information on the cost function over corruption activities.

on probable corruption (e.g. Hadad et al., 2021).

A longer-term system of targeted audits could be strengthened by maintaining some random audits. In such a mixed system, targeted audits would be used to detect and deter corruption for the highest-risk municipalities. Beyond mixing random and targeting audits, a broader discussion could consider more refined implementation choices using predicted corruption risk. As one example, policymakers could adapt the scale of the audit to corruption risk, where high-risk municipalities could get a larger team of auditors. Alternatively, auditors might weight the corruption risk by other relevant factors, such as municipal population or budget size. As a final example, policymakers could exploit spillovers in audit effects on corruption (Avis et al., 2018) by targeting geographical clusters of high-risk municipalities, or by targeting more politically prominent mayors. Understanding the relevance of these factors would require additional empirical evidence, preferably through randomized interventions. The behavioral responses could be further elucidated by structural modeling to better understand their relative importance (Avis et al., 2018; Finan and Mazzocco, 2021)

5.3. Adjusting for Political Bias in Targeted Corruption Audits

A practical strength of randomized audits is their intuitive fairness in the sense that all municipalities are targeted equally in expectation. In particular, audit probability does not depend on political party control of each municipality. Given the political sensitivity of corruption, it could be that any perceptions of bias in the machine allocation of audits would pose an institutional barrier to implementation of anti-corruption policies.

This concern is not idle, as targeting would indeed change the allocation of audits across parties. Focusing on the five largest political parties, we compute the observed corruption rate (from the random audits) and the predicted corruption rate (from the algorithm). These statistics are visualized in Figure 5, Top Panel, ranked roughly from most left-wing to most right-wing (Power and Rodrigues-Silveira, 2019): PT, PMDB, PSDB, PTB, and DEM (formerly PFL).[32] We can see that observed corruption (green bars) varies somewhat across parties. For example, PT has a relatively low corruption rate, while DEM has a relatively high corruption rate. The differences across parties are mostly reproduced in the model's predictions (orange bars), reflecting that the model is designed to treat parties the same. A notable exception is PMDB, for whom the

[32]The distribution of municipality-terms by party is shown in Appendix Figure F10.

algorithm somewhat understates the risk of corruption relative to the audit-measured rate.

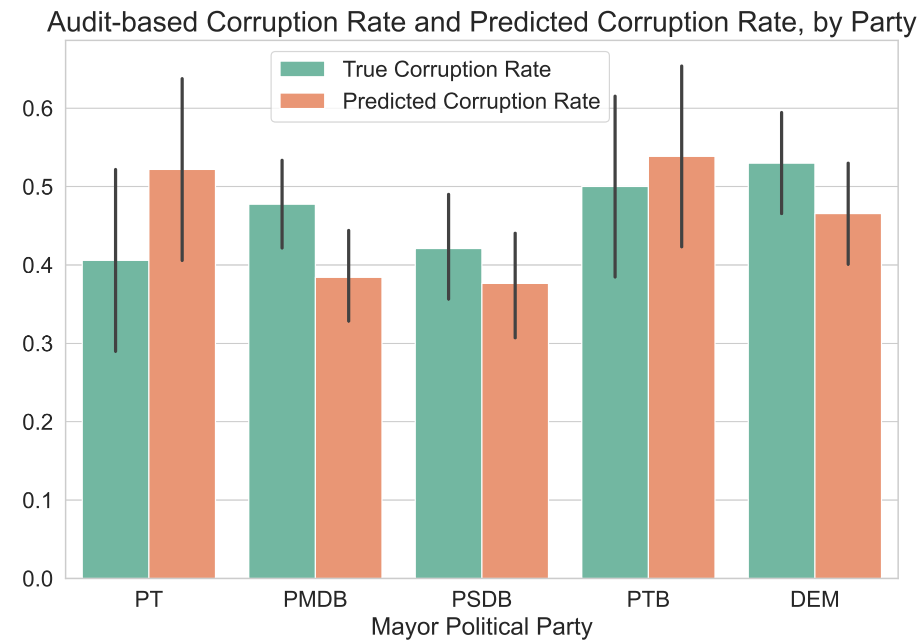
These differences in predicted corruption risk would be reflected in different audit rates under a targeting system, perhaps leading to perceptions of bias and accompanying skepticism about such a program. One way to address such concerns is to adapt tools from algorithmic fairness (Barocas et al., 2019), a burgeoning literature in computer science that provides statistical definitions of fairness for evaluating automated decision processes. The classic case study is criminal risk scoring and how the scores vary across racial or ethnic groups (Chouldechova, 2017; Berk et al., 2018; Kasy and Abebe, 2020). Statistical fairness metrics assess divergences across groups in a model's predictions (e.g., error rates) or its decisions (e.g., jail/release rates). While the literature has proposed a number of statistical fairness metrics, a standard criterion is *statistical parity.* Statistical parity requires that the probability of the negative outcome – in our case, audit rates – is equalized across groups. Given the political economy issues underlying corruption audits, statistical parity is a sensible criterion because it approximates the status quo in terms of the incidence of audits across political parties.

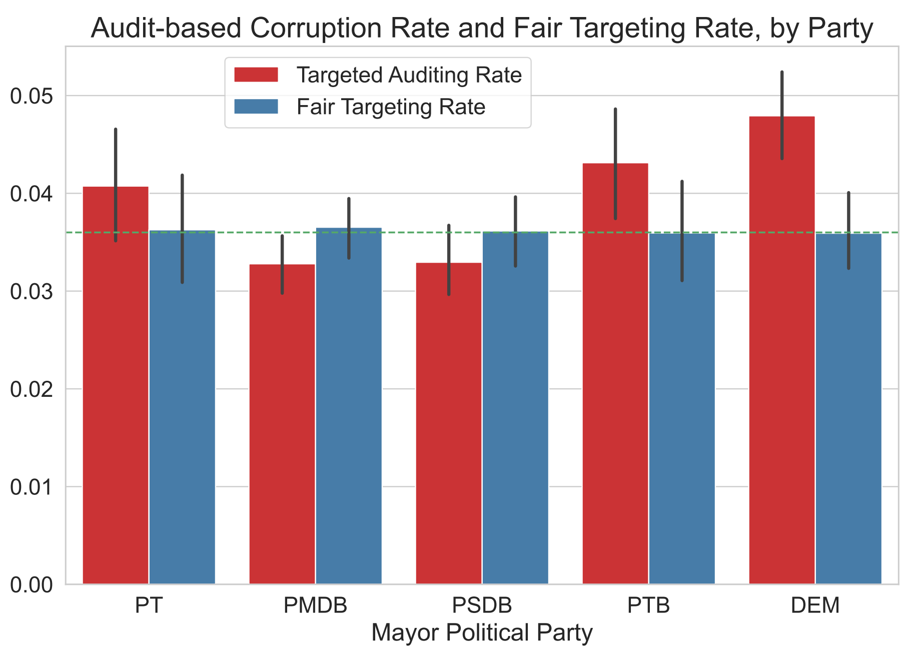
An intuitive approach to guarantee statistical parity is called *post-processing*, which starts by separating the problem into a prediction step and a decision step. Rambachan et al. (2020) show that equity concerns can be addressed solely at the decision step, with the prediction step being untouched.[33] The starting point is the same, with $\hat{Y}_{it}$ for each municipality-year and the resulting corruption-risk ranking for all municipalities in a given year. Instead of taking the highest-ranked municipalities from the whole set, however, we produce separate rankings for each political party. Within each party, we audit the same share of municipalities. Then by construction, the incidence of audits is equal across parties.

Figure 5 Bottom Panel shows the impact of fair targeting (blue bars) relative to unconstrained targeting (red bars). As intended, the fair audits have identical frequencies for each party (up to a rounding error). Yet this fairness adjustment has significant

[33] This *post-processing* approach is distinct from the more technically complex *pre-processing* or *constrained optimization* approaches that are explored in the computer science literature (see Barocas et al., 2019, ch. 3). The advantage of the latter methods is that the model does not need access to the sensitive covariate – normally, race/ethnicity – in order to produce a fair decision. In our setting, the sensitive covariate (city mayor party affiliation) is not that sensitive after all, and it will always be available in practice. Thus we take the post-processing approach.

Figure 5: Corruption Risk and Targeted Auditing, by Party





Notes: Top Panel reports the corruption rate in the audit data (in green bars) next to the predicted corruption rate from our XGBoost classifier (in orange bars), separately by the five political parties (meaning control of the mayor's office). Bottom Panel compares the auditing rates by party, under unconstrained targeting (red bars) and constrained targeting that equalizes audit rates across parties (blue bars). Horizontal dashed line gives the average audit rate in the sample. In both plots, 95% confidence interval spikes constructed by bootstrapping.

impacts on the distribution of audits. On the one hand, PTB and DEM benefit from the introduction of fair targeting and are audited less often compared to the standard policy. On the other hand, fair targeting increases the audit risk for PMDB-controlled municipalities.

The next question is how fair targeting changes the overall effectiveness of audits, relative to unconstrained targeting. Revisiting Table 4, we see in Column 3 that the discovered corruption rate for audited municipalities is 0.868, still far higher than the random baseline (0.486). Discovered corruption is just slightly lower than the main targeting policy. In terms of deterrence – the audit rate, conditional on corruption – the nonpartisan policy still maintains similar policy effectiveness gains: 0.074, which is still 2.644× higher than the audit rate under random assignment. Overall, adjusting targeted anti-corruption policies to equalize audit rates across political parties does not substantially undermine the effectiveness of those policies.

6. Conclusion

This paper has shown that corruption in local governments can be reliably detected, predicted, and measured using public budget accounts data. We have shown that the resulting synthetic measurements can then be used in downstream empirical analysis, as we can produce the same empirical results using corruption predictions in municipalities that were never audited. In the future, we hope that the expanded datasets built with machine prediction could be broadly useful for social scientists interested in corruption, as well as other variables that the method could produce.

Beyond expanding on empirical work, the corruption predictions can be used to guide policy responses to corruption. Our counterfactual policy estimates indicate substantial gains from such a policy, even when constraining the algorithm to treat each political party equitably. We hope that this proof of concept leads to further exploration and experimentation by researchers, development organizations, and government agencies.

This research adds to the emerging literature using machine learning and other tools from data science to explore new datasets and questions (Kleinberg et al., 2015; Athey, 2018). Our method of detecting corruption has the potential to substantially expand the stock of datasets available for economists studying development, political economy, and public finance. Within Brazil, researchers will no longer be constrained to the relatively small set of municipalities that were audited. Outside of Brazil, the method could in principle be applied in any context with auditor-provided labels for corruption. Something that can and should be explored is whether the corruption predictions produced in Brazil could be valid for other countries and settings.

References

Andini, M., E. Ciani, G. de Blasio, A. D'Ignazio, and V. Salvestrini (2018). Targeting with machine learning: An application to a tax rebate program in italy. *Journal of Economic Behavior & Organization 156*, 86–102.

Ash, E., M. Morelli, and R. Van Weelden (2017). Elections and divisiveness: Theory and evidence. *The Journal of Politics 79*(4), 1268–1285.

Assunção, J., R. McMillan, J. Murphy, and E. Souza-Rodrigues (2019). Optimal environmental targeting in the amazon rainforest. Technical report, National Bureau of Economic Research.

Athey, S. (2018). The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda*. University of Chicago Press.

Athey, S. and S. Wager (2021). Policy learning with observational data. *Econometrica 89*(1), 133–161.

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

Bandiera, O., A. Prat, S. Hansen, and R. Sadun (2020). Ceo behavior and firm performance. *Journal of Political Economy 0*(0), 000–000.

Bansak, K., J. Ferwerda, J. Hainmueller, A. Dillon, D. Hangartner, D. Lawrence, and J. Weinstein (2018). Improving refugee integration through data-driven algorithmic assignment. *Science 359*(6373), 325–329.

Barocas, S., M. Hardt, and A. Narayanan (2019). *Fairness and machine learning: Limitations and Opportunities.*

Battiston, P., S. Gamba, and A. Santoro (2020). Optimizing tax administration policies with machine learning. *University of Milan Bicocca Department of Economics, Management and Statistics Working Paper* (436).

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy 76*(2), 169–217.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives 28*(2), 29–50.

Berk, R., H. Heidari, S. Jabbari, M. Kearns, and A. Roth (2018). Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, 0049124118782533.

Björkegren, D., J. E. Blumenstock, and S. Knight (2020). Manipulation-proof machine learning. *arXiv preprint arXiv:2004.03865*.

Bobonis, G. J., L. R. Cámara Fuertes, and R. Schwabe (2016). Monitoring corruptible politicians. *American Economic Review 106*(8), 2371–2405.

Brollo, F., T. Nannicini, R. Perotti, and G. Tabellini (2013). The political resource curse. *American Economic Review 103*(5), 1759–96.

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics 225*(2), 200–230.

Cavalcanti, F., G. Daniele, and S. Galletta (2018). Popularity shocks and political selection. *Journal of Public Economics 165*, 201–216.

Chen, T. and C. Guestrin (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.

Cheol, L. and J. Mikesell (2018). The impact of public officials' corruption on the size and allocation of u.s. state spending. *Public Administration Review*, 346–359.

Chong, A., A. L. De La O, D. Karlan, and L. Wantchekon (2015). Does corruption information inspire the fight or quash the hope? a field experiment in mexico on voter turnout, choice, and party identification. *The Journal of Politics 77*(1), 55–71.

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data 5*(2), 153–163.

Colonnelli, E., J. A. Gallego, and M. Prem (2019). What predicts corruption? *Available at SSRN 3330651*.

Conley, T. G. and F. Decarolis (2016). Detecting bidders groups in collusive auctions. *American Economic Journal: Microeconomics 8*(2), 1–38.

Coviello, D. and S. Gagliarducci (2017). Tenure in office and public procurement. *American Economic Journal: Economic Policy 9*(3), 59–105.

Daniele, G. and T. Giommoni (2020). Corruption under austerity. *BAFFI CAREFIN Centre Research Paper No. 2020-131*.

De Angelis, I., G. de Blasio, and L. Rizzica (2020). Lost in corruption. evidence from eu funding to southern italy. *Italian Economic Journal*, 1–23.

Decarolis, F. and C. Giorgiantonio (2020). Corruption red flags in public procurement:

new evidence from italian calls for tenders. *Questioni di Economia e Finanza, Occasional Papers* (544).

Djankov, S., R. La Porta, F. Lopez-de Silanes, and A. Shleifer (2003). Courts. *The Quarterly Journal of Economics 118*(2), 453–517.

Ferraz, C. and F. Finan (2008). Exposing corrupt politicians: The effects of Brazil's publicly released audits on electoral outcomes. *The Quarterly Journal of Economics 123*(2), 703–745.

Ferraz, C., F. Finan, and D. B. Moreira (2012). Corrupting learning: Evidence from missing federal education funds in brazil. *Journal of Public Economics 96*(9-10), 712–726.

Feurer, M., K. Eggensperger, S. Falkner, M. Lindauer, and F. Hutter (2018). Practical automated machine learning for the automl challenge 2018. In *International Workshop on Automatic Machine Learning at ICML*, pp. 1189–1232.

Finan, F. and M. Mazzocco (2021). Electoral incentives and the allocation of public funds. *Journal of the European Economic Association 19*(5), 2467–2512.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.

Gallego, J., G. Rivero, J. D. Martínez, et al. (2018). Preventing rather than punishing: An early warning model of malfeasance in public procurement. Technical report.

Gentzkow, M. and J. M. Shapiro (2010). What drives media slant? evidence from us daily newspapers. *Econometrica 78*(1), 35–71.

Gentzkow, M., J. M. Shapiro, and M. Taddy (2019). Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica 87*(4), 1307–1340.

Glaeser, E. L., A. Hillis, S. D. Kominers, and M. Luca (2016). Crowdsourcing city government: Using tournaments to improve inspection accuracy. *American Economic Review 106*(5), 114–18.

Goodfellow, I., Y. Bengio, and A. Courville (2016). *Deep learning*. MIT press.

Grinsztajn, L., E. Oyallon, and G. Varoquaux (2022). Why do tree-based models still outperform deep learning on tabular data? *arXiv preprint arXiv:2207.08815*.

Hadad, V., D. A. Hirshberg, R. Zhan, S. Wager, and S. Athey (2021). Confidence intervals for policy evaluation in adaptive experiments. *Proceedings of the National Academy of Sciences 118*(15).

Hansen, S., M. McMahon, and A. Prat (2018). Transparency and deliberation within the fomc: a computational linguistics approach. *The Quarterly Journal of Economics 133*(2), 801–870.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction.* Springer Science & Business Media.

Hessami, Z. (2014). Political corruption, public procurement, and budget composition: Theory and evidence from oecd countries. *European Journal of Political Economy 34*(C), 372–389.

Hitsch, G. J. and S. Misra (2018). Heterogeneous treatment effects and optimal targeting policy evaluation. *Available at SSRN 3111957*.

Kang, J. S., P. Kuznetsova, M. Luca, and Y. Choi (2013). Where not to eat? improving public policy by predicting hygiene inspections using online reviews. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pp. 1443–1448.

Kasy, M. and R. Abebe (2020). Fairness, equality, and power in algorithmic decision making. In *ICML Workshop on Participatory Approaches to Machine Learning.*

Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan (2018). Human decisions and machine predictions. *The quarterly journal of economics 133*(1), 237–293.

Kleinberg, J., J. Ludwig, S. Mullainathan, and Z. Obermeyer (2015). Prediction policy problems. *American Economic Review 105*(5), 491–95.

Knaus, M. C., M. Lechner, and A. Strittmatter (2018). Machine learning estimation of heterogeneous causal effects: Empirical monte carlo evidence. *The Econometrics Journal*.

Knittel, C. R. and S. Stolper (2019). Using machine learning to target treatment: The case of household energy use. Technical report, National Bureau of Economic Research.

Kyriacou, A. P., L. Muinelo-Gallo, and O. Roca-Sagalés (2015). Construction corrupts: Empirical evidence from a panel of 42 countries. *Public Choice 165*(1), 123–145.

Lagaras, S., J. Ponticelli, and M. Tsoutsoura (2017). Caught with the hand in the cookie jar: Firm growth and labor reallocation after exposure of corrupt practices.

Liu, C., T. T. Moldogaziev, and J. L. Mikesell (2017). Corruption and state and local government debt expansion. *Public Administration Review 77*(5), 681–690.

López-Iturriaga, F. J. and I. P. Sanz (2018). Predicting public corruption with neural networks: An analysis of spanish provinces. *Social Indicators Research 140*(3), 975–998.

Machoski, E. and J. M. de Araujo (2020). Corruption in public health and its effects on the economic growth of brazilian municipalities. *The European Journal of Health Economics*, 1–19.

Mauro, P. (1998). Corruption and the composition of government expenditure. *Journal of Public economics 69*(2), 263–279.

Mohler, G. O., M. B. Short, S. Malinowski, M. Johnson, G. E. Tita, A. L. Bertozzi, and P. J. Brantingham (2015). Randomized controlled field trials of predictive policing. *Journal of the American statistical association 110*(512), 1399–1411.

Molnar, C. (2020). *Interpretable machine learning.* Lulu. com.

Morris, S. D. and J. L. Klesner (2010). Corruption and trust: Theoretical considerations and evidence from mexico. *Comparative Political Studies 43*(10), 1258–1285.

Mullainathan, S. and Z. Obermeyer (2019). A machine learning approach to low-value health care: wasted tests, missed heart attacks and mis-predictions. Technical report, National Bureau of Economic Research.

Mullainathan, S. and J. Spiess (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives 31*(2), 87–106.

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

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research 12*, 2825–2830.

Poblete-Cazenave, R. (2021). Reputation shocks and strategic responses in electoral campaigns.

Power, T. J. and R. Rodrigues-Silveira (2019). Mapping ideological preferences in brazilian elections, 1994-2018: a municipal-level study. *Brazilian Political Science Review 13*(1).

Rambachan, A., J. Kleinberg, S. Mullainathan, and J. Ludwig (2020). An economic approach to regulating algorithms. Technical report, National Bureau of Economic Research.

Rockoff, J. E., B. A. Jacob, T. J. Kane, and D. O. Staiger (2011). Can you recognize an effective teacher when you recruit one? *Education finance and Policy 6*(1), 43–74.

Vannutelli, S. (2021). From lapdogs to watchdogs: Random auditor assignment and municipal fiscal performance in italy. *Job Market Paper: Yale University*.

Widmer, P., S. Galletta, and E. Ash (2020). Media slant is contagious. *Center for Law & Economics Working Paper Series 14*.

Winters, M. S. and R. Weitz-Shapiro (2013). Lacking information or condoning corruption: When do voters support corrupt politicians? *Comparative Politics 45*(4), 418–436.

Zamboni, Y. and S. Litschig (2018). Audit risk and rent extraction: Evidence from a randomized evaluation in brazil. *Journal of Development Economics 134*, 133 – 149.

A Machine Learning Approach to Analyze and Support Anti-Corruption Policy

APPENDIX

A. Additional Material on Model Training and Evaluation

Table A1: Balance sheets components

Year	N. of Categories								
	Assets		Liabilities		Expenditures		Revenues		
	All	Selected	All	Selected	All	Selected	All	Selected	N. of Audits
2001	56	43	46	37	43	43	52	51	0
2002	56	43	46	37	101	78	90	76	0
2003	57	44	48	39	100	77	90	76	276
2004	59	46	49	38	295	155	146	106	340
2005	63	43	52	38	298	158	151	105	300
2006	63	43	52	38	301	161	155	106	180
2007	64	43	52	38	309	161	170	108	180
2008	64	43	52	38	310	162	170	108	120
2009	80	41	57	37	331	161	198	108	180
2010	88	41	69	37	334	161	219	109	180
2011	89	41	69	37	335	161	219	109	120
2012	89	41	69	37	334	161	219	109	120

Notes: This table report the summary tabulations by year and by macro category on the total number of components of the municipal budget, the the number of components selected by XGBoost to form predictions, and the number of audits by year.

Table A2: Selected Hyperparameters and Learned Model Size

Fold	L1 Penalty	L2 Penalty	Max Tree Depth	Learning Rate	Min. Child Weight	Tree Count	Node Count
1	1	0.1	10	0.1	5	72	9461
2	1	0.1	10	0.1	3	71	11874
3	0.5	0.5	10	0.1	1	46	13135
4	2	2	10	0.1	5	97	13604
5	1	0.5	10	0.1	3	70	12467
mean	1.1	.64	10	0.1	3.4	71.4	12108.2

Notes: This table reports the hyperparameters selected for each of the 5 folds model training. Rows give the folds. L1 and L2 Penalty are regularization terms on the splitting decision that encourage smaller trees. Max Tree Depth is the max number of splits before a terminal node. The learning rate is how quickly parameters are updated during training. Minimum Child Weight is another regularization term, corresponding to the minimum number of observations required at each node. The grid is as follows: L1 and L2 regularization penalties on the learned parameter values (each selected from {0.1, 0.5, 1, 2}), max depth of the constituent decision trees (selected from {5,10,20}), learning rate (selected from {0.1,0.5}), minimum child weight (selected from {1,3,5}). Hence, $4 \times 4 \times 3 \times 2 \times 3 = 288$ hyperparameter grid cells to search. Tree Count is the number of trees grown in the resulting forest. Node Count is the total number of variable splitting nodes in the forest.

Table A3: Confusion Matrices

Panel A. XGBoost

		Prediction	
		Not Corrupt	Corrupt
Truth	Not Corrupt	2572	486
	Corrupt	993	1248

Panel B. OLS

		Prediction	
		Not Corrupt	Corrupt
Truth	Not Corrupt	1303	1755
	Corrupt	884	1357

Panel C. LASSO

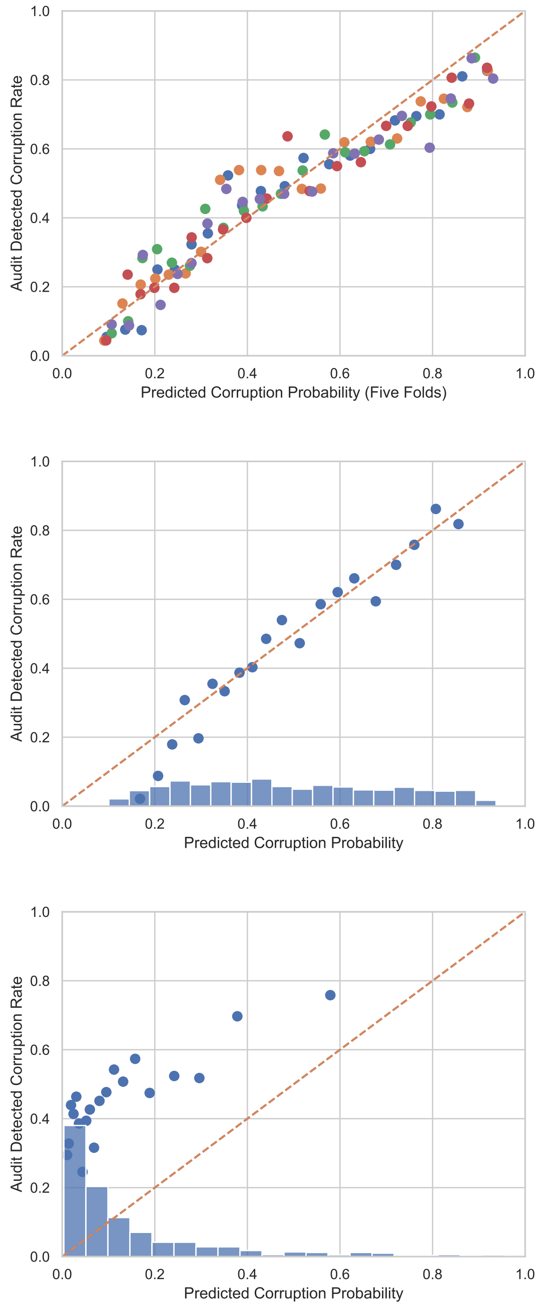
		Prediction	
		Not Corrupt	Corrupt
Truth	Not Corrupt	1555	1503
	Corrupt	1147	1094

Panel D. Logistic regression

		Prediction	
		Not Corrupt	Corrupt
Truth	Not Corrupt	2018	1040
	Corrupt	1211	1030

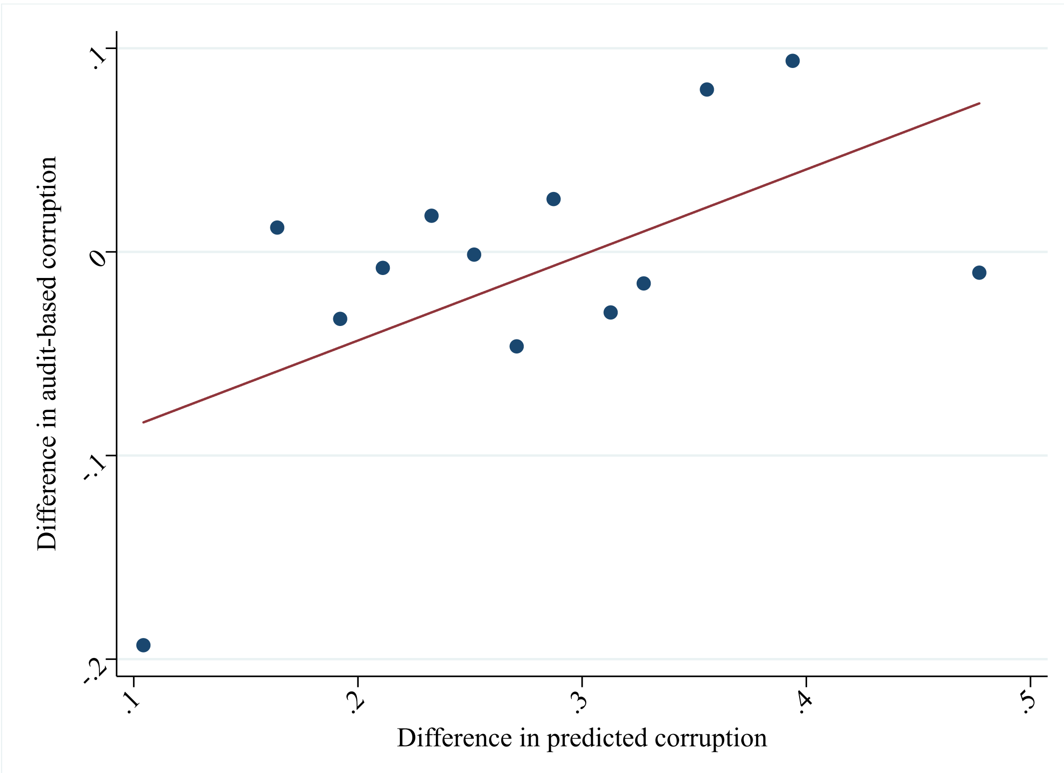
Notes: The table reports confusion matrices from the model predictions XGBoost (recall=0.56 and precision=0.72), OLS (recall=0.61 and precision=0.44), LASSO (recall=0.49 and precision=0.42) and Logistic regression (recall=0.46 and precision=0.50).

Figure A1: Additional Calibration Plots: Audited Corruption Rate vs. Predicted Corruption Risk



Notes: Calibration plots showing audit corruption rates (marks on the vertical axis), binned by predicted corruption probability (horizontal axis). The dashed 45-degree line (in orange) demarcates perfect calibration. In the top panel, we use the baseline model but show the calibration plot for each of the five models trained separately on different training folds. In the other panels, the blue histogram shows the density of the predicted corruption probability. The middle panel uses the Brollo et al. (2013) corruption measure with group sampling by municipality. The bottom panel uses the Avis et al. (2018) corruption measure using random sampling.

Figure A2: Difference in observed and predicted corruption for cities audited twice



Notes: The figure focuses on cities that have been audited twice and it shows a binscatter between the difference over time in the audited levels of corruption using the data from Brollo et al. (2013) and the predicted levels of corruption. The analysis includes the following list of fixed effects and controls: first audit year and second audit year fixed effects, mean income, share of population employed, sector of occupation (agriculture, industry, commerce, transportation, services and public administration), share with college education, poverty rate, and Gini Coefficient of income. The coefficient of the corresponding regression is 0.603 (p-value 0.070). Similar results emerge in the analysis without controls (coefficient 0.348 – p-value 0.107) and without controls and fixed effects (coefficient 0.168 – p-values 0.031). While there is a clear upward trend, it is not a one-to-one relationship. This could be due to missing information in the audit-measured corruption, which is binary. It could also be due to the fact that second audits are not included in our machine learning dataset, which would add error to the predictions for second audits.

B. Alternative Specifications for Corruption Prediction

This appendix reports the performance metrics from some alternative corruption prediction specifications. First, to compare XGBoost model performance using not only budget factors but also fixed demographic factors, we apply random splits between training and test set by municipality, instead of by municipality-year. Appendix Table B4 shows the relative performance when we use budget data (column 1), when we add demographic characteristics (column 2), or when we use only demographic characteristics (column 3). The more conservative sampling specification in Column 1 reduces accuracy compared to the main-text specification, but it is still capturing significant predictive signal (in Column 3, AUC-ROC = 0.636 with budget and demographics). Comparing Column 1 to Column 2, we see that budget information is more predictive of corruption than demographic information.

Second, we replicate our predictive results by using the corruption measure from Avis et al. (2018). As already discussed, there are structural differences between these two original measures of corruption. First, this alternative measure is continuous, rather than binary. We have for each audited municipality the share of inspection orders that presented irregularities. Second, we are missing the first audits, as we have information only from July 2006 through March 2013 (lotteries 22–38). Third, this alternative measure does not provide the exact year (or term) in which the irregularity took place. To overcome this limitation, we treat as audited the three years before the actual audit took place. Finally, to create a binary label from the continuous variable we identified as corrupted those municipalities with a share of irregularities in the top quartile of the distribution.

Despite these differences, Figure B3 shows that the predictions using the alternative corruption label are similar to those from the main analysis. They show similar rankings on average. The performance metrics are reported columns (4-7) of Appendix Table B4. Again, we find that XGBoost outperforms all the other methods. Indeed, we find accuracy metrics that are higher than those from the main analysis.

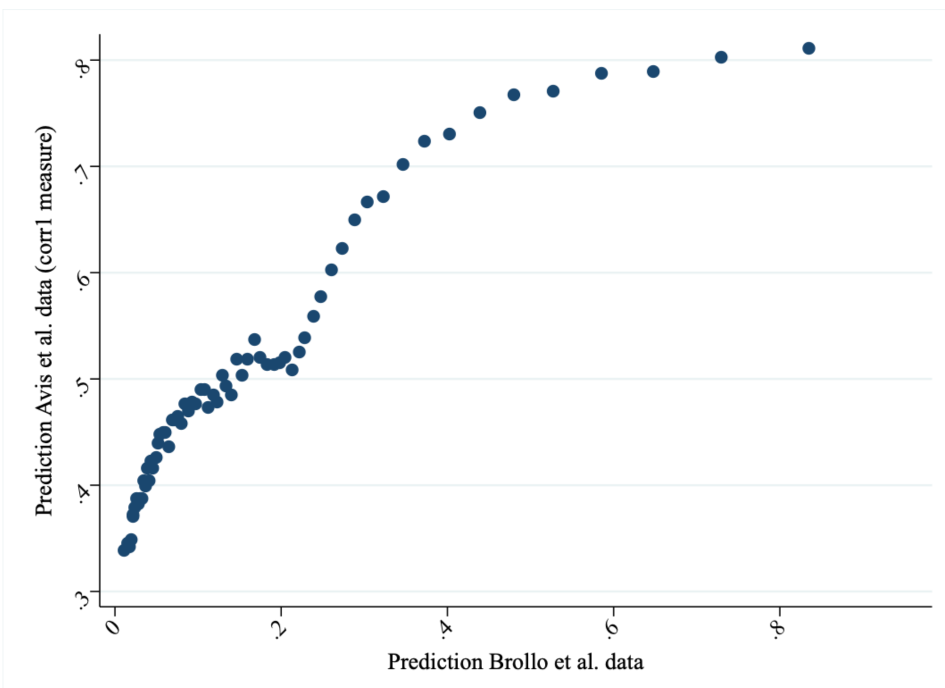
Finally, we find that most of our empirical results still hold when using the predictions from this alternative measure of corruption.

Table B4: Additional models performance

	XGBoost (municipal sampling)			Avis et al. (2018) data			
	Budget (1)	Budget + Demo (2)	Demo (3)	XGBoost (4)	OLS (5)	LASSO (6)	Logistic (7)
Accuracy	0.625	0.619	0.594	0.849	0.482	0.591	0.743
	[0.607-0.642]	[0.594-0.660]	[0.575-0.628]	[0.826-0.870]	[0.379-0.546]	[0.429-0.747]	[0.647-0.786]
AUC-ROC	0.636	0.638	0.601	0.900	0.476	0.520	0.653
	[0.605-0.677]	[0.605-0.698]	[0.573-0.657]	[0.889-0.917]	[0.336 -0.636]	[0.427-0.639]	[0.601-0.693]
F1	0.495	0.497	0.484	0.627	0.288	0.217	0.473
	[0.462-0.522]	[0.474-0.542]	[0.443-0.511]	[0.532-0.660]	[0.185-0.410]	[0.000-0.410]	[0.358-0.534]

Notes: The table provides the mean and standard error (in parentheses) across five values for the prediction performance, produced using different training-set folds. In columns (1-3) we use XGBoost models with municipal sampling, and different sets of predictors: only budget components in column (1), budget components and demographic characteristics in column (2) and only demographic characteristics in column (3). In columns (4-8) we report the predictions performance as in Table 2, but using the corruption data from Avis et al. (2018).

Figure B3: Predictions from Avis et al. (2018) vs. Predictions from Brollo et al. (2013)



Notes: The figure shows a binscatter between the predictions formed using the data from Avis et al. (2018) and the ones formed using the data from Brollo et al. (2013) for all municipality-year. The correlation between the two variables is 0.531.

Table C5: Most important budget features for Corruption Prediction

N.	Category	Macro Category	Weight	Perturbation Response Mean	Perturbation Response Min	Perturbation Response Max
1	Spending in agriculture	Expenditure	114	0.010	-0.25	0.58
2	Tax on agricultural territorial property (ITR) (compartecipation)	Revenue	96	0.022	-0.24	0.45
3	Tax on export of industrialized products (IPI) (compartecipation)	Revenue	93	0.023	-0.42	0.53
4	Spending in transportation	Expenditure	92	0.008	-0.22	0.43
5	Taxes	Revenue	82	0.011	-0.41	0.65
6	Motor vehicle property tax (IPVA) (compartecipation)	Revenue	80	0.001	-0.34	0.45
7	Tax on real estate transactions (ITB)	Revenue	76	0.026	-0.20	0.50
8	Cash	Assets	75	0.010	-0.23	0.43
9	Income Tax (IRRF)	Revenue	73	0.003	-0.21	0.26
10	Tax on real estate (IPTU)	Revenue	73	0.024	-0.26	0.41
11	Budget Surplus/Deficit		72	0.027	-0.25	0.47
12	Revenue from assets	Revenue	72	0.008	-0.23	0.31
13	Deposits	Assets	71	-0.007	-0.32	0.21
14	Transfers from tax on circ. of goods/services (Law 87-96)	Revenue	70	0.009	-0.20	0.31
15	Transfers for the health system	Revenue	69	-0.016	-0.44	0.30
16	Spending for legislative procedure	Expenditure	67	-0.009	-0.42	0.30
17	Civil servant per diems	Expenditure	67	-0.009	-0.50	0.45
18	Financial and non-financial liabilities	Liabilities	66	-0.003	-0.22	0.35
19	Transfers from tax on circulation of goods/services (compartecipation)	Revenue	61	-0.007	-0.31	0.24
20	Supplies (current year)	Liabilities	61	0.006	-0.20	0.24
21	Liquid assets	Assets	61	0.005	-0.19	0.26
22	Capital expenditure	Expenditure	60	-0.003	-0.33	0.15
23	Processed Outstanding Liabilities	Liabilities	59	-0.007	-0.28	0.19
24	Banks	Assets	58	0.006	-0.20	0.37
25	Tax to fund police authority	Revenue	57	-0.013	-0.30	0.20
26	Direct spending (previous years)	Expenditure	56	-0.040	-0.67	0.26
27	Financial assets	Assets	55	0.006	-0.15	0.37
28	Federal transfers	Revenue	53	-0.014	-0.61	0.39
29	Direct spending for consulting	Expenditure	52	-0.021	-0.56	0.224
30	Liquid assets	Assets	51	0.014	-0.19	0.43
31	Outstanding debt	Liabilities	50	0.001	-0.16	0.33
32	FPM (comparticipation)	Assets	50	-0.008	-0.34	0.21
33	Supplies (previous years)	Liabilities	48	-0.001	-0.29	0.19
34	Other revenues	Revenue	48	-0.023	-0.38	0.20
35	Financial liabilities	Liabilities	47	-0.009	-0.27	0.20

Notes: List of the most important features. Weight ranks the features (budget components) by how often they are included in a decision tree contained in the ensemble classifier, averaged across the five training folds. The last three columns show the mean, minimum, and maximum values computed from the perturbation analysis described in Appendix C.1, evaluating how each individual feature affects the predicted probability of being corrupted.

C. Additional material for model interpretation

This appendix contains additional material for interpreting the predictions of the tree ensemble. Table C5 shows the list of variables (Column 2), ranked by their feature importance weight (Column 4). This weight is the average across five models (using different training-set folds) of the number of times that a constituent decision tree splits on a variable. For example, the model "uses" spending on agriculture (second row) about 103 times.

C.1. Perturbation-Based Partial Dependence Analysis

Here we illustrate the important non-linearities and interactions encoded by the tree ensemble. We take a perturbation-based partial dependence approach (see, e.g., Friedman, 2001), which works as follows. Iterating over each observation i in the dataset, we form the predicted change in $\hat{Y}_i$ from perturbing a budget feature j by one standard deviation, either up or down. This perturbation is done for each model m (that is, models trained on a different training-set fold), so we obtain a dataset of values $\Delta\hat{Y}_{ijm}$ (with values inverted for negative perturbations).

With five folds and a positive/negative perturbation per fold, we observe ten deltas for each observation in the dataset. With 5563 municipalities and 12 years of data per municipality, we produce about 650,000 deltas in total for each of 797 feature variables. What can these distributions of predicted changes in corruption rates tell us? If corruption were linearly related to features, we would expect the distribution of $\Delta\hat{Y}_{ijm}$ to have the same sign for a given feature j. If the algorithm captures important non-linearities, non-monotonicities, and/or interactions, then the responses would have both positive and negative density.

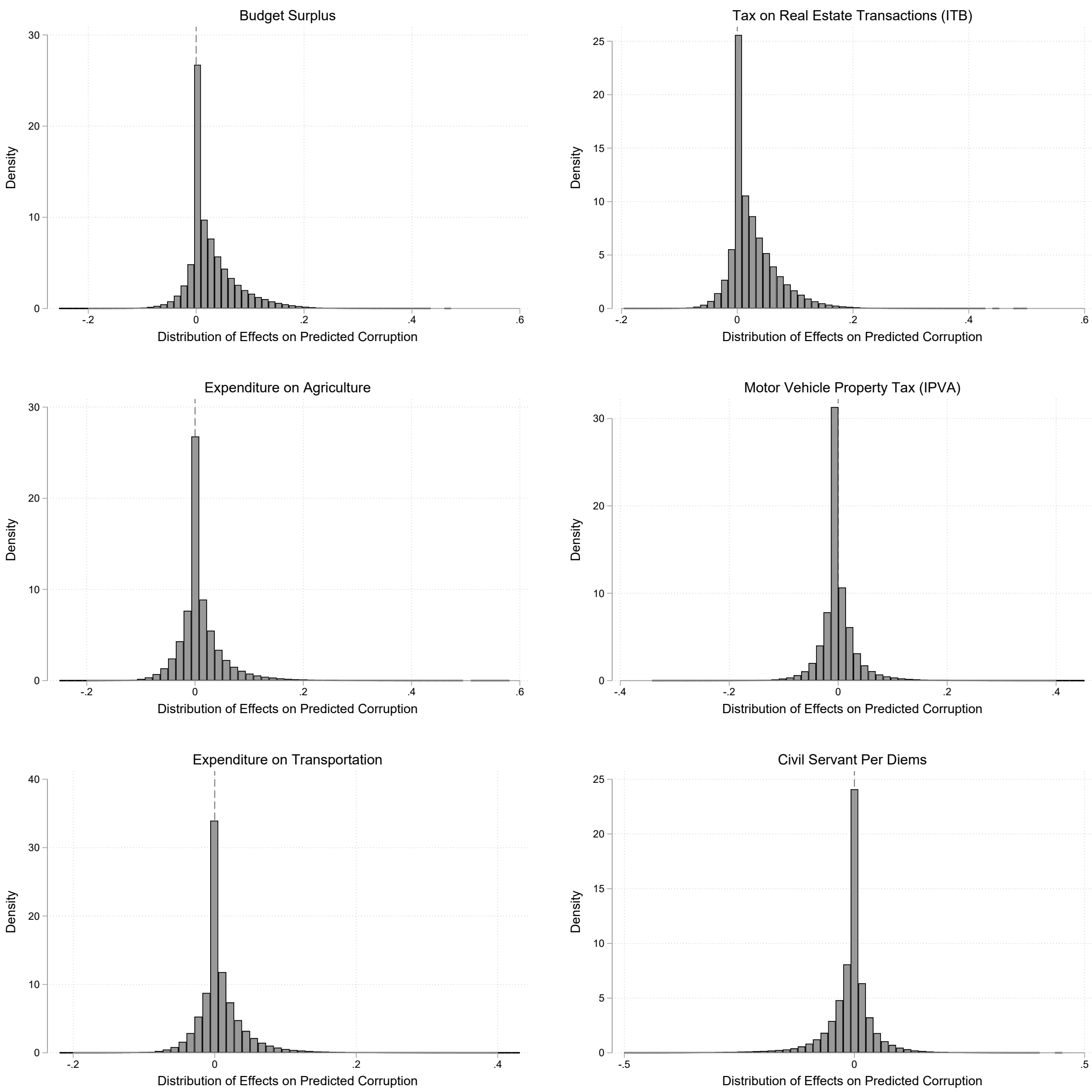
Overall, the perturbation results are consistent with a highly non-linear and contingent predictive relationship. Of the 454 predictive variables, only 31 variables have an always-weakly-positive relation, and only 24 variables have an always-weakly-negative relation. Columns 6 and 7 of Table C5 show the minimum and maximum values for the perturbation responses $\Delta\hat{Y}_{ijm}$ by variable for the most important features. We can see that for all of these variables, the minimum is negative, and the maximum is positive. So for the most pivotal variables, shifts could have either a positive or negative relation to predicted corruption depending on the status quo values.

To help illustrate this contingent partial dependence, Figure C4 shows distributions of the perturbation response $\Delta\hat{Y}_{ijm}$ for a selection of variables j. Each graph contains a relatively wide distribution of possible responses, indicating a non-linear, interaction-heavy relationship. For example, it is intuitive that judiciary spending (bottom left panel) is mostly negatively related to corruption risk. However, the effect is non-monotonic and there are some positive values in the response distribution.

In Table C5 Column 5, we report the mean value of the feature perturbation response. Considering all of the caveats mentioned so far, this column shows the average direction and magnitude of the model's perceived association between the indicated variable and corruption risk. Positive values indicate that higher values of this variable tend to reflect greater corruption risk, while negative values indicate that higher values of this variable tend to reflect lower corruption risk.

As mentioned in the main text, it is consistent with some previous literature that

Figure C4: Distributions of Predicted Responses to Perturbing Model Features



Notes: The figure shows histograms of $\Delta\hat{Y}_{ijm}$, produced using the perturbation approach described in the text, for a selection of variables j.

(4) expenditures on transportation (Hessami, 2014) and (7) real estate and construction (Kyriacou et al., 2015) are positively correlated with corruption. In addition, it is intuitive that higher (25) spending on policing is negatively related to corruption. But other positive variables are inconsistent with the previous literature. For example, our model suggests that having a budget surplus rather than deficit (11) is positively associated with corruption, which goes against the findings in Liu et al. (2017). Overall, these additional results do not modify the central point that the model's functional form is complex and one cannot identify one-to-one relationships between a budget factor and predicted corruption.

C.2. Counting Budget Feature Mentions in Audit Reports

The municipal audit reports are published on the web site of the CGU, `auditoria.cgu.gov.br`, in a search engine interface. We programmatically downloaded the full library of audit reports for our time period as PDF files. The corpus contains 2,062 reports. The PDFs were in machine-readable Portuguese and therefore straightforward to extract as plain text using the python package pdfminer.

We performed some mild pre-processing on the report texts. Punctuation and capitalization were removed. The resulting pre-processed corpus consists of 2,062 documents, each containing on average 26,743 words and 173,648 characters. In total, the corpus contains over 55 million words.

The next step is to identify mentions of relevant budget factors. Our dataset of budget features has a codebook with a variable label and short description for each budget item. For example, the budget item "Outras TrConvMun" is described as "Outras Transferências de Convênios dos Municípios" ("Other Transfers from Municipalities"). Both the label and the description are included in our pattern matching lexicon, after being pre-processed in the same way as the corpus (punctuation and capitalization removed). The lexicon contains 1,141 items as sometimes the label and description are the same. On average, the pre-processed items contain 28 characters and are 4 words long.

Table C6: Budget features most often mentioned in the audit reports

N.	Category	Macro Category	Mention
1	Health expenditure	Expenditure	190,610
2	Assets	Assets	69,328
3	Spending in labour	Expenditure	59,835
4	Spending in education	Expenditure	56,563
5	Spending in adminstration	Expenditure	49,858
6	Cash	Assets	35,553
7	Spending in transportation	Expenditure	32,176
8	Spending in social services	Expenditure	29,633
9	Spending in basic health	Expenditure	28,499
10	National fund for education development	Revenue	19,148
11	Spending in culture	Expenditure	15,776
12	Spending in primary education	Expenditure	15,049
13	Supply spending	Expenditure	10,427
14	Spending in agriculture	Expenditure	9,615
15	Permanent assets	Assets	8,132
16	Spending in communication	Expenditure	8,062
17	Spending in social security	Expenditure	8,015
18	Spending in sanitation	Expenditure	6,739
19	Spending in the employment fund	Expenditure	5,501
20	Current spending in other contributions	Expenditure	4,510
21	Spending in transfers	Expenditure	4,388
22	Spending in telecommunication	Expenditure	4,273
23	Spending in energy	Expenditure	3,961
24	Spending in kindergarten	Expenditure	3,799
25	Stocks	Assets	3,758
26	Spending in tourism	Expenditure	3,527
27	Spending in high school	Expenditure	2,791
28	Transportation services	Revenue	2,100
29	Spending in health surveillance	Expenditure	2,096
30	Spending in adult education	Expenditure	1,944
31	Taxes	Revenues	1,942
32	Spending in electric energy	Expenditure	1,881
33	Spending in industry	Expenditure	1,848
34	Spending in leisure	Expenditure	1,819
35	Spending in urban infrastructures	Expenditure	1,796

Notes: List of the features most often mentioned in the audit reports, as described in Appendix C.2.

Finally, we counted the total mentions of each budget feature in the corpus of reports, limiting to exact matches of the pre-processed strings. Summary statistics on these matches are reported in Table C6. For example, health expenditures are mention almost 200,000 times. 28% of the budget variables are mentioned in the reports. Conditional on being mentioned at all, a budget factor is mentioned 4,465 times on average, or about twice per report.

D. Additional Material: Effect of Revenue Shocks on Corruption

Following Brollo et al. (2013), we focus on the initial seven brackets and restrict the sample to cities with a population below 50,940. Furthermore, we restrict the sample, for the sake of symmetry, to municipalities from 3,396 below the first threshold to 6,792 above the seventh threshold. This sample represents about 90 percent of Brazilian municipalities. The amount of revenues received by municipality i in state k follows the allocation mechanism: $FPM_i^k = \frac{FPM_k \lambda_i}{\sum_{i \in k} \lambda_i}$, where FPM_k is the total amount allocated in state k and λ_i is the municipality-specific coefficient, as shown in Table D7.

Formally, we have the first stage

$$\tau_i = g(P_i) + \alpha_\tau \hat{\tau}_i + \delta_t + \gamma_p + u_i \tag{1}$$

and reduced form

$$y_i = g(P_i) + \alpha_y \hat{\tau}_i + \delta_t + \gamma_p + \eta_i \tag{2}$$

where y_i is corruption, $g(\cdot)$ is a high order polynomial in P_i (the population of city i), δ_t contains time fixed effects, γ_p contains state fixed effects, and u_i and η_i are the error terms. The coefficients α_τ and α_y capture the effects of prescribed transfers on actual transfers and (predicted) corruption, respectively. For the two-stage-least squares analysis, we estimate the second stage

$$y_i = g(P_i) + \beta_y \tau_i + \delta_t + \gamma_p + \epsilon_i \tag{3}$$

where prescribed transfers $\hat{\tau}_i$ are used as an instrument for actual transfers τ_i and all other terms are defined as above. The coefficient β_y captures the causal effect of actual transfers on (predicted) corruption. For inference, standard errors are clustered by municipality. See Brollo et al. (2013) for a detailed discussion and testing of the econometric assumptions in this setting.

Our data cover the two mayoral terms, January 2001–December 2004 and January 2005–December 2008. For the sake of brevity we only replicate the analysis on the overall effect, omitting the threshold-specific analysis. Appendix Table D8 shows the descriptive statistics by population bracket. Brazilian municipalities in our sample receive, on average, \$3.3M BRL (about \$610K USD), while prescribed transfers are somewhat higher at \$3.7M BRL (about \$680K USD). The average level of (predicted) corruption is around 0.5 and its level does not change significantly as we move to larger cities.

Table D7: Population thresholds for Inter-Government Transfers

Population interval	FPM coefficient
Below 10,189	0.6
10,189−13,584	0.8
13,585−16,980	1
16,981−23,772	1.2
23,773−30,564	1.4
30,565−37,356	1.6
37,357−44,148	1.8
44,149−50,940	2
Above 50,940	from 2.2 to 4

Notes: These coefficients have been introduced by *Decreto-lei* n. 1,881, 27 august 1981.

Table D8: Descriptive statistics for the Revenue Shocks Analysis

	FPM transfers			
Population	Actual transfers	Theoretical transfers	Predicted Corruption	N
(1)	(2)	(3)	(4)	(5)
6,793 − 10,188	19.655	21.200	.442	1,429
10,189 − 13,584	25.642	28.771	.500	1,076
13,585 − 16,980	31.888	36.316	.527	805
16,981 − 23,772	38.445	44.019	.543	1,083
23,773 − 30,564	44.223	51.082	.529	629
30,565 − 37,356	50.869	58.113	.521	380
37,357 − 44,148	57.376	66.468	.510	253
44,149 − 50,940	62.389	72.368	.498	154
Total	33.440	37.930	.502	5,809

Notes: The sample includes all Brazilian municipalities with a population in the interval 6,793-50,940. Population is the number of inhabitants. Actual and theoretical FPM transfers expressed in R$100,000 at 2000 prices.

Table D9: Replication Brollo et al. (2013) with random samples

Random sample:	First (1)	Second (2)	Third (3)	Fourth (4)
Panel A. First Stage				
Theoretical transfers	0.7100***	0.6810***	0.7344***	0.6794***
	(0.0247)	(0.0485)	(0.0177)	(0.0439)
Panel B. Reduced Form				
Theoretical transfers	0.0044***	0.0046***	0.0034***	0.0043***
	(0.0007)	(0.0007)	(0.0008)	(0.0007)
Panel C. 2SLS				
Actual transfers	0.0061***	0.0068***	0.0046***	0.0063***
	(0.0010)	(0.0012)	(0.0011)	(0.0011)
N. Observations	1115	1115	1115	1115

Notes: Effects of FPM transfers on (predicted) corruption measures. The four columns display the analysis focusing on four different random samples with 1,115 observations. Panel A reports the estimates of the first-stage analysis, the dependent variable is *actual transfers*. Panel B reports the estimates of reduced form analysis, the dependent variable is *predicted corruption*. Panel C reports the estimates of the 2sls estimates, the dependent variable is *predicted corruption* and *actual transfers* is instrumented with *theoretical transfers*. Column headings indicate the sample of municipalities included. All regressions controls for a third-order polynomial in normalized population size, term dummies, and macro-region dummies. Robust standard errors clustered at the municipal level are in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D10: Effect of Revenue Shocks on Corruption - Alternative Predictions

Dep. var.: Predicted corruption	All cities	
	Prediction demographics (1)	Prediction budget (without FPM) (2)
Panel A. Reduced Form		
Theoretical transfers	-0.0001	0.0044***
	(0.0007)	(0.0003)
Panel B. 2SLS		
Actual transfers	-0.0002	0.0064***
	(0.0009)	(0.0005)
N. Observations	5808	5808

Notes: Effects of FPM transfers on (predicted) corruption measures: column (1) contains the analysis with the predictions built using as predictors a set of municipal demographic characteristics, and column (2) contains the analysis with the predictions built with budget predictors where FPM transfers are permuted randomly. Panel A reports the estimates of reduced form analysis, the dependent variable is *predicted corruption.* Panel B reports the estimates of the 2sls estimates, the dependent variable is *predicted corruption* and *actual transfers* is instrumented with *theoretical transfers.* The sample includes all Brazilian municipalities. All regressions controls for a third-order polynomial in normalized population size, term dummies, and macro-region dummies. Robust standard errors clustered at the municipal level are in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D11: Robustness checks on predicted corruption measure

	Audited cities (1)	All cities (2)	Non-audited cities (3)
Panel A. Bootstrap predictions			
Reduced Form			
Prescribed transfers	0.0053	0.0044	0.004
	(0.0018)	(0.0005)	(0.0004)
	[0.0017 ; 0.0088]	[0.0034 ; 0.0053]	[0.0032 ; 0.0047]
2SLS			
Actual transfers	0.0078	0.006	0.0059
	(0.0026)	(0.0007)	(0.0006)
	[0.0027 ; 0.0128]	[0.0046 ; 0.0073]	[0.0047 ; 0.007]
N. Observations	1115	5808	4693
Panel B. Held-out sample 0			
Reduced Form			
Prescribed transfers	0.0043***	0.0037***	0.0035***
	(0.0011)	(0.0004)	(0.0004)
2SLS			
Actual transfers	0.0062***	0.0054***	0.0051***
	(0.0015)	(0.0005)	(0.0006)
N. Observations	1083	5776	4693
Panel C. Held-out sample 1			
Reduced Form			
Prescribed transfers	0.0039***	0.0044***	0.0044***
	(0.0011)	(0.0004)	(0.0004)
2SLS			
Actual transfers	0.0058***	0.0064***	0.0062***
	(0.0016)	(0.0005)	(0.0005)
N. Observations	1092	5785	4693

Notes: The table reports the analysis of Table 3 using different versions of the measure of predicted corruption. Panel A reports the results using the predictions through a bootstrap method. Panel B reports the results using the predictions with the held-out method in the first sample. Panel C reports the results using the predictions with the held-out method in the second sample. Column headings indicate the sample of municipalities included. All regressions controls for a third-order polynomial in normalized population size, term dummies, and macro-region dummies. The analyses do not report the first stage which is as the one reported in Table 3. Panel A reports 95% confidence intervals in squared brackets. Robust standard errors clustered at the municipal level are in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

E. Additional Material: Effect of Audits on Corruption

Using the annual corruption prediction y_{ist} in municipality i of state s at year t, we take a standard event study approach and estimate the within-municipality effects of a (randomly assigned) audit. Let D_{ist}^k be a dummy variable for k years before and after an audit. We estimate

$$y_{ist} = \sum_{k=-3,k\neq-1}^{5} \beta_k D_{ist}^k + \delta_i + \lambda_t + W_{ist}'\phi + \epsilon_{ist} \tag{4}$$

where we have municipality fixed effects δ_i, year fixed effects λ_t and other controls W_{ist}, which in particular includes dummy variables indicating periods distant from when the audit took place. Because $k \neq -1$ (the year before the audit), the β_k's estimate the dynamic effects relative to the year before the audit. The identifying assumption hinges on randomness in the timing of selection into the audit program. We cluster standard errors by state, although results are qualitatively similar when clustering by municipality and when using bootstrapped standard errors. The sample includes 1,479 municipalities that have received an audit in the time period under analysis.

Now, we discuss a series of additional results for the event study analysis. First, we replicate the main results with the correction for the two-way fixed effects designs with many groups and periods. To do this, we estimate the main model according to the methodology of Callaway and Sant'Anna (2021). These findings are displayed in Figure E5 and are very similar to the main results of Figure 4. A difference is a positive effect on corruption for municipalities where the audit did not found irregularities.

Second, we check whether post-audit budget adjustments may explain the decline in predicted corruption levels after the audit. We provide two tests. First, we estimate the main model controlling for total expenditure, expressed in per-capita terms. This test is reported in Figure E6 and the results are similar to the ones of the main model, reported in Figure 4. Secondly, we estimate the main model using as dependent variable the amount of total expenditure (per-capita): Figure E7 shows this test and it suggests that the audit does not have any significant effect on future levels of total expenditure. This result holds for the full sample and for the sample of corrupted and non-corrupted cities.

Third, we test the channel of political accountability. In particular, we aim to study whether the effect of the audit on future corruption is stronger where local political accountability is high and we focus on the variable margin of victory. This test is shown in Figure E8, which reports the analyses conducted with the full sample. The figures show that the effect is stronger in cities where the mayor won with a small margin of victory – below the median level – compared to cities where she won with a high margin – above the median level. This result suggests that the audit has a larger impact where the electoral competition is more pronounced. Overall, these results provide some evidence that political accountability affects the impact of an audit on future corruption.

Table E12: Coefficient Estimates for Event Study Analysis

	All cities (1)	Cities with corruption (2)	Cities without corruption (3)
Year pre4 and behind	-0.0293	-0.0816**	-0.0241
	(0.0238)	(0.0385)	(0.1042)
Year pre3	-0.0169	-0.0372	-0.0016
	(0.0163)	(0.0241)	(0.0714)
Year pre2	-0.0137	-0.0084	-0.0428
	(0.0119)	(0.0214)	(0.0375)
Audit year	-0.0292**	-0.0110	-0.0598
	(0.0136)	(0.0208)	(0.0383)
Year post1	-0.0326	-0.0611**	-0.0333
	(0.0192)	(0.0288)	(0.0471)
Year post2	-0.0287	-0.1105***	0.0057
	(0.0262)	(0.0351)	(0.0539)
Year post3	-0.0197	-0.1287***	0.0749
	(0.0300)	(0.0451)	(0.0835)
Year post4	-0.0197	-0.1597**	0.0802
	(0.0372)	(0.0576)	(0.1126)
Year post5	-0.0214	-0.1840**	0.1385
	(0.0435)	(0.0720)	(0.1327)
Years post6 and more	-0.0229	-0.2003**	0.1448
	(0.0489)	(0.0831)	(0.1459)
N. Observations	17252	8895	3086
Adjusted R^2	0.535	0.510	0.538

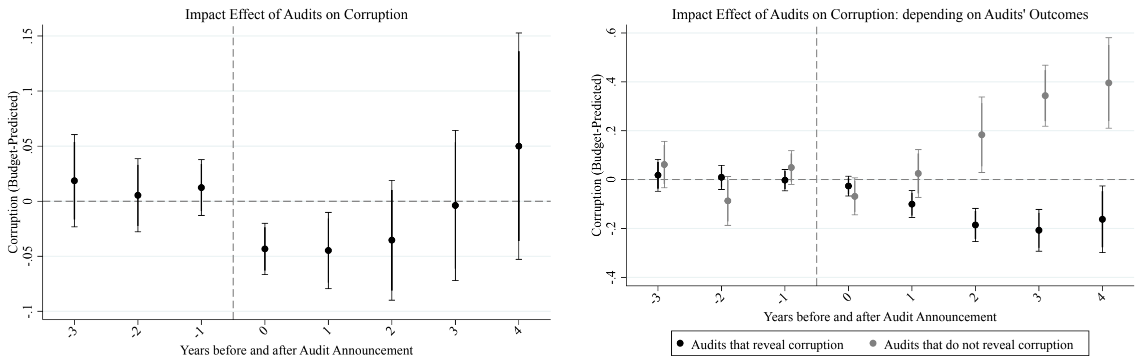
Notes: The dependent variable is (predicted) corruption measure - binary. The sample includes all the cities that receive an audit for the period 2001-2012. Column (1) includes the complete sample, Column (2) includes the sample of cities in which the audit discovered corruption (according to the definition of narrow corruption) and Column (3) includes the sample of cities in which the audit did not discover any type of corruption. The specification includes city and year fixed effects. Robust standard errors clustered at the state level are in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table E13: Robustness checks on predicted corruption measure - Event Study Analysis

	Bootstrap			Held-out sample 0			Held-out sample 1		
	All cities	Cities with corruption	Cities without corruption	All cities	Cities with corruption	Cities without corruption	All cities	Cities with corruption	Cities without corruption
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Year pre4 and behind	-0.023	-0.172	0.123	0.0577**	0.0033	0.1035	-0.0383	-0.0831**	0.0736
	(0.028)	(0.039)	(0.082)	(0.0264)	(0.0401)	(0.0881)	(0.0288)	(0.0379)	(0.0977)
	[-0.077 ; 0.031]	[-0.248 ; -0.095]	[-0.037 ; 0.283]						
Year pre3	0.020	-0.028	0.039	0.0516**	0.0120	0.0629	-0.0102	-0.0062	0.0498
	(0.018)	(0.025)	(0.053)	(0.0235)	(0.0347)	(0.0606)	(0.0186)	(0.0279)	(0.0634)
	[-0.015 ; 0.055]	[-0.077 ; 0.021]	[-0.064 ; 0.142]						
Year pre2	0.013	0.009	0.011	0.0121	-0.0059	0.0405	-0.0022	-0.0075	-0.0352
	(0.011)	(0.018)	(0.029)	(0.0211)	(0.0253)	(0.0455)	(0.0132)	(0.0254)	(0.0444)
	[-0.008 ; 0.034]	[-0.026 ; 0.044]	[-0.045 ; 0.067]						
Audit year	-0.074	-0.293	0.097	-0.0347*	-0.0428*	0.0116	-0.0108	-0.0384**	0.0248
	(0.014)	(0.030)	(0.034)	(0.0185)	(0.0249)	(0.0431)	(0.0155)	(0.0174)	(0.0321)
	[-0.101 ; -0.046]	[-0.351 ; -0.234]	[0.030 ; 0.163]						
Year post1	-0.080	-0.391	0.139	-0.0386*	-0.0792**	-0.0045	-0.0141	-0.1015***	0.0293
	(0.020)	(0.041)	(0.056)	(0.0201)	(0.0368)	(0.0666)	(0.0226)	(0.0178)	(0.0504)
	[-0.119 ; -0.040]	[-0.471 ; -0.310]	[0.029 ; 0.248]						
Year post2	-0.091	-0.489	0.207	-0.0304	-0.1135**	0.0612	-0.0113	-0.1239***	0.0688
	(0.026)	(0.050)	(0.077)	(0.0256)	(0.0431)	(0.0776)	(0.0252)	(0.0260)	(0.0726)
	[-0.141 ; -0.040]	[-0.587 ; -0.391]	[0.056 ; 0.357]						
Year post3	-0.101	-0.572	0.252	-0.0149	-0.1292**	0.1090	-0.0044	-0.1341***	0.1004
	(0.032)	(0.058)	(0.095)	(0.0323)	(0.0478)	(0.1155)	(0.0314)	(0.0296)	(0.0713)
	[-0.163 ; -0.038]	[-0.685 ; -0.458]	[0.065 ; 0.438]						
Year post4	-0.135	-0.671	0.260	-0.0170	-0.1313**	0.1492	-0.0110	-0.1660***	0.0994
	(0.038)	(0.066)	(0.117)	(0.0353)	(0.0532)	(0.1170)	(0.0399)	(0.0400)	(0.1098)
	[-0.209 ; -0.060]	[-0.800 ; -0.541]	[0.030 ; 0.489]						
Year post5	-0.153	-0.759	0.296	-0.0326	-0.1551**	0.1527	-0.0000	-0.1832***	0.1624
	(0.044)	(0.077)	(0.137)	(0.0457)	(0.0668)	(0.1291)	(0.0419)	(0.0502)	(0.1097)
	[-0.239 ; -0.066]	[-0.909 ; -0.608]	[0.027 ; 0.564]						
Year post6 and more	-0.194	-0.921	0.311	-0.0332	-0.2141**	0.1823	-0.0083	-0.2256***	0.1640
	(0.055)	(0.089)	(0.160)	(0.0534)	(0.0809)	(0.1650)	(0.0594)	(0.0643)	(0.1437)
	[-0.301 ; -0.086]	[-1.095 ; -0.746]	[-0.002 ; 0.624]						
Observations	17252	8895	3086	14603	7424	2745	14602	7386	2703
Adjusted R^2	0.535	0.507	0.526						

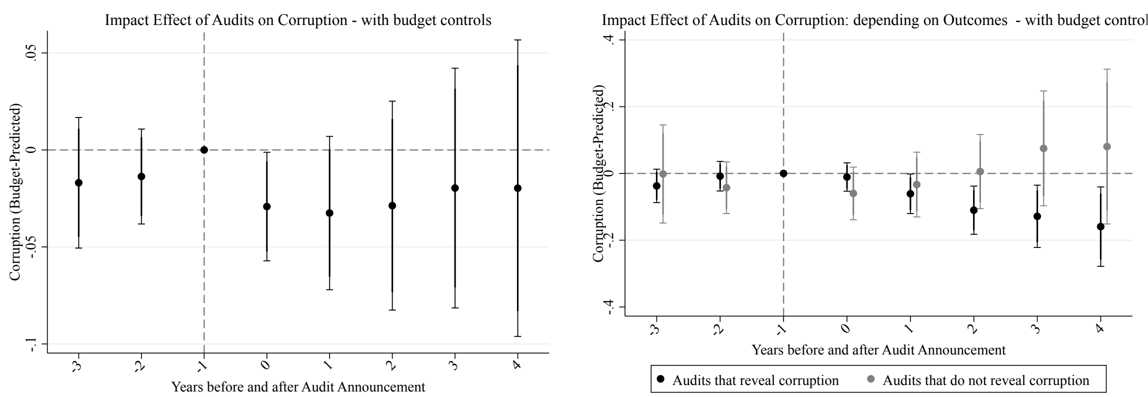
Notes: The dependent variable is (predicted) corruption measure - binary. The sample includes all the cities that receive an audit for the period 2001-2012. Columns (1), (4) and (7) include the complete sample, Columns (2), (5) and (8) include the sample of cities in which the audit discovered corruption (according to the definition of narrow corruption) and Columns (3), (6) and (9) include the sample of cities in which the audit did not discover any type of corruption. Columns (1), (4) and (7) report the results using the predictions through a bootstrap method. Columns (2), (5) and (8) report the results using the predictions with the held-out method in the first sample. Columns (3), (6) and (9) report the results using the predictions with the held-out method in the second sample. The specification includes city and year fixed effects. Columns (1-3) report 95% confidence intervals in squared brackets. Robust standard errors clustered at the state level are in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure E5: Dynamic effect of the audits - Callaway, Sant'Anna (2021)



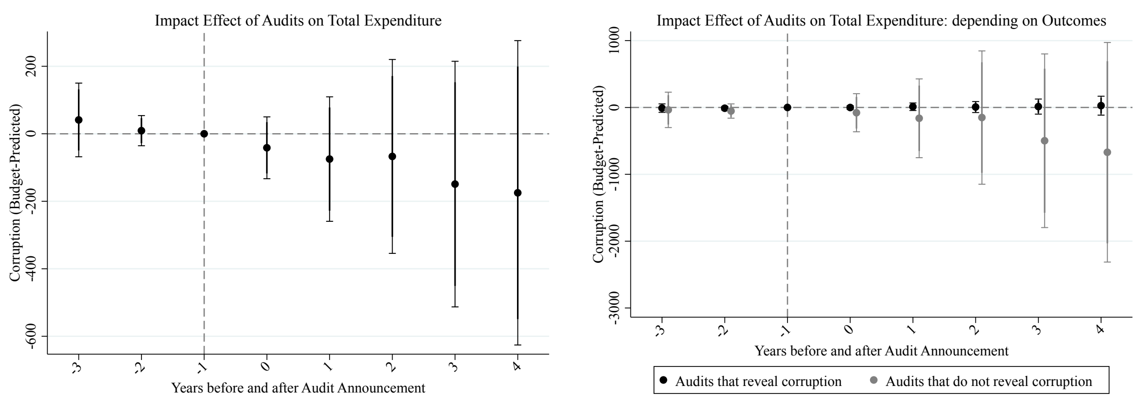
Notes: Event study estimates for dynamic effect of audits on budget-predicted corruption according to the methodology of Callaway and Sant'Anna (2021). Error spikes give 95% confidence intervals, with standard error clustered by state. Left panel: all audits; right panel: audits that found corruption (in black); audits that did not find corruption (in grey).

Figure E6: Dynamic effect of the audits - Controlling for total expenditure



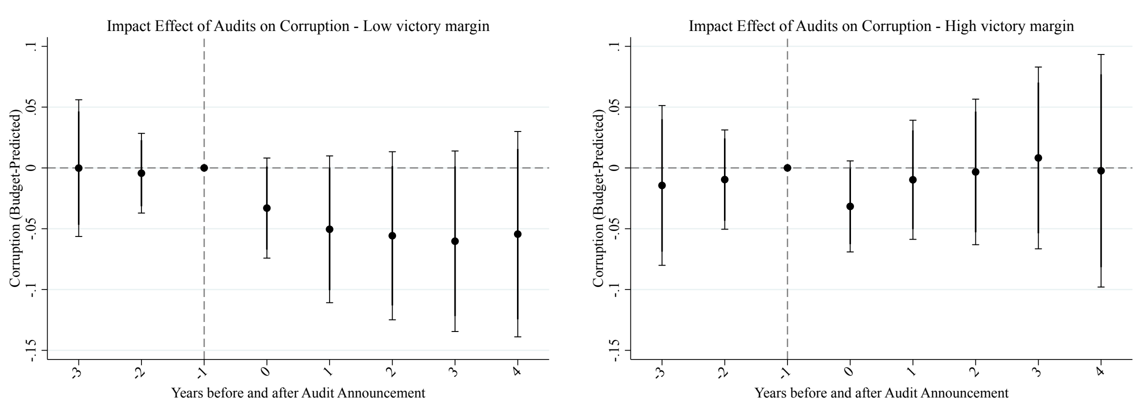
Notes: Event study estimates for dynamic effect of audits on budget-predicted corruption. Error spikes give 95% confidence intervals, with standard error clustered by state. Left panel: all audits; right panel: audits that found corruption (in black); audits that did not find corruption (in grey). This regressions include as additional control municipal total expenditure.

Figure E7: Dynamic effect of the audits on total expenditure



Notes: Event study estimates for dynamic effect of audits on municipal total expenditure. Error spikes give 95% confidence intervals, with standard error clustered by state. Left panel: all audits; right panel: audits that found corruption (in black); audits that did not find corruption (in grey).

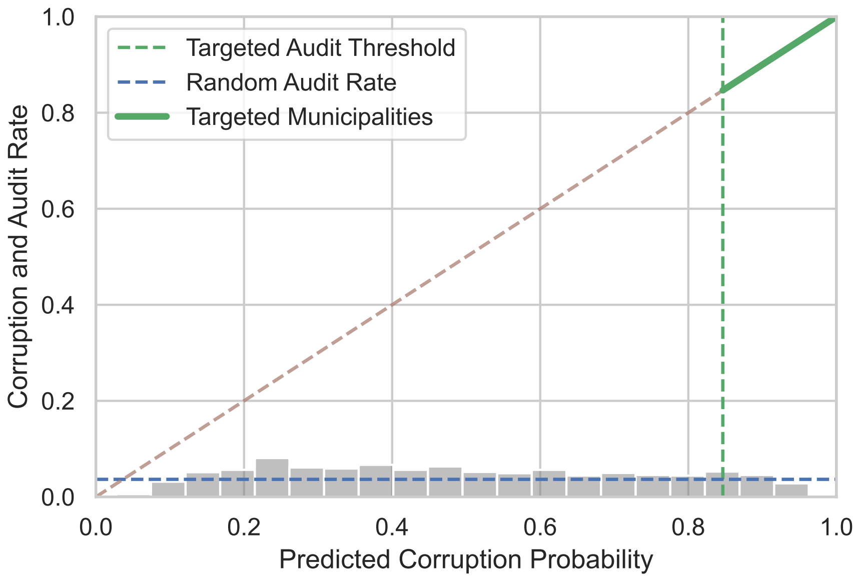
Figure E8: Dynamic effect of the audits - Margin of victory



Notes: Event study estimates for dynamic effect of audits on budget-predicted corruption. Error spikes give 95% confidence intervals, with standard error clustered by state. In the left panel are considered only municipalities where the mayor won with a low margin of victory (below the median); In the right panel are considered only municipalities where the mayor won with a high margin of victory (above the median)

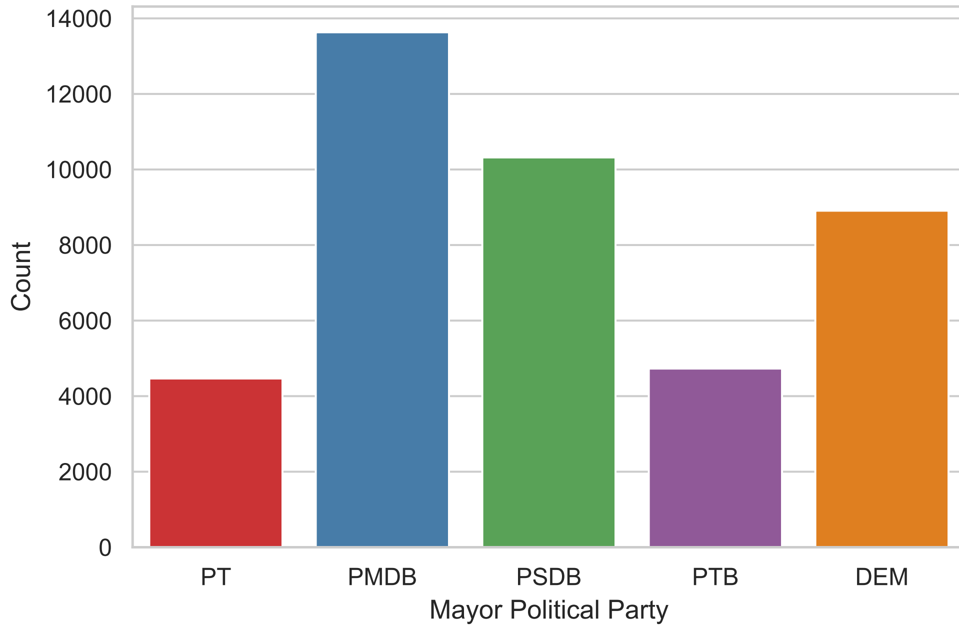
F. Additional Material on Audit Policy Support

Figure F9: Targeted Auditing Based on Corruption Risk



Notes: Illustration of targeted auditing policy. The underlying gray histogram indicates the distribution of the corruption risk predictions, with the top two bins containing the approximately 203 municipalities to be targeted. The diagonal line is at 45 degrees and indicates the predicted corruption rate at any spot in the risk distribution. The horizontal blue dashed line at 0.036 gives the audit probability under random audits, while the vertical green dashed line indicates the across-year average threshold corruption risk (0.846) above which municipalities are targeted for audit.

Figure F10: Distribution of Party Control of Municipalities



Notes: Number of municipality-year observations for each party, in terms of the affiliation of the mayor in that municipality.

G. Additional Material on Budget variables

Table G14: Full list of Budget Features

Category	Category	Category	Category
Ativo	Ativo Financeiro	Disponivel	Caixa
Bancos	Aplic Financ	Creditos Circul AF	Cred a Receber
Dep Realiz CP	Out Valor Realiz	Ativo N Fin	Realiz CP
Creditos Circul ANF	Fornec Receber	Creditos Parcelados	Diversos Respons
Emprest e Financ	Adiantam Concedidos	Recursos vinculados	Out Cred Circul
Provisao Dev Duvid	Bens e Val Circul	Estoques	Out Bens Val Circul
Valores Pendentes CP	Desp Antecipadas	Valores Diferidos	Out Val Pend CP
Realizavel LP	Dep Realiz LP	Dep Compuls	Recursos Vincul
Cred Realiz LP	Divida Ativa	Deved Entid Agente	Emprest Financ
Creditos a Receber	Prov Perdas Provav LP	Permanente	Investimentos
Partic Societaria	Out Investimentos	Prov Perdas Provav Perm	Imobilizado
Bens Moveis Imoveis	Titulos Valores	Deprec Amort Ex	Diferido
Desp Diferidas	Amortiz Acumulada	Ativo Real	Ativo Compensado
Resp Titulo Val	Garantias	Convenios Contratos	Out Compensac
Prov Perdas Provav CP	AF Val Pendente CP	AF a Curto Prazo	AF a Longo Prazo
Prov Dev Duv Fornecimentos	Emprest e Financ CP	Recursos vinculados CP	Titulos e Valores
Prov Perdas Provav OBV	Recursos Vincul LP	Prov Perdas Divida Ativa	Emprest Financ LP
Creditos a Receber LP	Partic Soc Emp Depend	Titulos Valores Intangiveis	Disponivel Moeda Estrangeira
Invest Regime Prop Previd	Invest Segmento Renda Fixa	Invest Segmento Renda Variável	Títulos e Valores Mobiliários
Invest Taxa de Adm do RPPS	Emprést Recur Prev a Receber	Provisão Perdas em Invest	Invest do Regime Próp Prev
Invest em Segmento de Imóveis	Bens Móveis	Máquinas e Equipamentos	Outros Bens Móveis
Bens Imóveis	Edifícios e Instalações	Terras e Terrenos	Outros Bens Imóveis
Títulos e Valores	Bens Intangíveis	Disponível Moeda Nacional	Títulos
Fundos de Aplicação Financeira	Poupanças	Outras aplicações Financeiras	Demais Disponibilidades
Prov Devedores Duvidosos LP	Intangível	Deprec Amort Ex Ac Imobil	Deprec Amort Ex Ac Intang
Despesas Orçamentárias	Desp Correntes	Desp de Custeio	Desp de Pessoal
Pessoal Ativos	Obrigações Patronais	Demais Desp de Pessoal	Terceirização de Mão de Obra
Outras Desp de Pessoal	Serviços de Terceiros e Encargos	Outros Custeio	Desp com Transf Correntes
Transf a Pessoas	Pessoal Inativos	Pessoal Pensionistas	Salário Família
Outra Trans a Pessoas	Contr Form PASEP	Juros e Encargos da Dívida	Demais Desp Transf Correntes
Outras Desp Correntes	Despesas de Capital	Inversões Financeiras	Desp Transf de Capital
Amortizações	Outras Desp Transf Capital	SUPERAVIT ou DEFICIT	Legislativa
Judiciária	Planejamento	Agricultura	Educação e Cultura
Habitação e Urbanismo	Indústria e Comércio	Saúde e Saneamento	Assistência e Previdência
Transporte	Segurança Pública	Desenvolvimento Regional	Energia e Recursos Minerais
Comunicações	Outras	Pessoal e Encarg SocPES	PES Transf a Estados DF
PES Transf ao Exterior	PES Aplicações Diretas	PESAD Aposent e Reformas	PESAD Pensões
PESAD Contrat Tempo Determ	PESAD Contrib Entid Fec Previd	PESAD SalárioFamília	PESAD Vencimentos Pes Civil
PESAD Vencimentos Pes Mil	PESAD Obrig Patronais	PESAD Out Desp Variáveis PC	PESAD Out Desp Variáveis PM
PESAD Out Desp Pes Terceiriz	PESAD Dep Compulsórios	PESAD Sentenças Judiciais	PESAD Desp Exerc Anteriores
PESAD Indeniz Res Trabalhistas	PESAD Ressarc Desp Pes Req	PES Out Desp Pessoal e Enc	Juros e Encargos Dívida
Out Desp CorrentesODC	ODC Transf à União	ODC Transf a Estados DF	ODC Transf a Municípios
ODC Transf Inst Priv s Fins Lucr	ODC Transf Inst Priv c Fins Lucr	ODC Transf Inst Multigov Nac	ODC Transf ao Exterior
ODC Aplicações Diretas	ODCAD Aposent e Reformas	ODCAD Pensões	ODCAD Contrat Tempo Determ
ODCAD Out Benef Previdenc	ODCAD Benef Deficiente e Idoso	ODCAD Out Benef Assistenciais	ODCAD SalárioFamília
ODCAD Out Benef Nat Social	ODCAD Diárias Civil	ODCAD Diárias Militar	ODCAD Aux Fin Estudantes
ODCAD Auxílio-Fardamento	ODCAD Aux Fin Pesquisadores	ODCAD Obrig Política Monetária	ODCAD Encargos pela Honra
ODCAD Remun Cotas Fund Autárq	ODCAD Mat Consumo	ODCAD Mat Distribuição Gratuita	ODCAD Pass Desp Locomoção
ODCAD Serv Consultoria	ODCAD Out Serviç Terceiros PF	ODCAD Locação Mão-de-Obra	ODCAD Arrendamento Mercantil
ODCAD Out Serviç Terceiros PJ	ODCAD Equaliz Preços Taxas	ODCAD Auxílio-Alimentação	ODCAD Obrig Tribut e Contrib
ODCAD Out Aux Financeiros PF	ODCAD Auxílio-Transporte	ODCAD Dep Compulsórios	ODCAD Sentenças Judiciais
ODCAD Desp Exerc Anteriores	ODCAD Indeniz e Restituições	ODCAD Inden Trabalhos Campo	ODC Out Despesas Correntes
Amortização da Dívida	Essencial à Justiça	Administração	Defesa Nacional
Relações Exteriores	Assistência Social	Previdência Social	Saúde
Trabalho	Educação	Cultura	Direitos da Cidadania
Urbanismo	Habitação	Saneamento	Gestão Ambiental
Ciência e Tecnologia	Organização Agrária	Indústria	Comércio E Serviços
Energia	Desporto e Lazer	Encargos Especiais	ODCAD Premiaçoes Diversas
Juros e Encargos DívidaJED	JED Aplicações Diretas	JED Juros Div pContrato	JED OutEncDivContratada
JED JurDesagios Mobiliaria	JED OutEncDivMobiliaria	JED Encargos ARO	JED Sentencas Judiciais
JED DespExercAnteriores	JED Inden e Restituições	I TransfUniao	I TransfEstadoDF
I TransfMunicipios	I TransfInsPrivadaSFL	I TransfInsPrivadaCFL	I TransfMultigovNacionais
I Transf Exterior	I Aplicações Diretas	IAD Contrat Tempo Determ	IAD Diarias Civil
IAD Out Desp Variáveis PM	IAD Aux Fin Pesquisadores	IAD Material Consumo	IAD Pass Desp Locomoção
IAD Serv Consultoria	IAD Out Serv Teceiros PF	IAD Locação Mão-de-Obra	IAD Out Serv Teceiros PJ
IAD Obras e Instalações	IAD Equipam Mat Perm	IAD Aquisição de Imóveis	IAD Sentenças Judiciais
IAD Desp Exerc Anteriores	IAD Inden e Restituições	IF Transf EstDF	IF Transf Municipios
IF Transf Privada SFL	IF Transf Exterior	IF Aplicações Diretas	IFAD Aquisição Imóveis

Table G14: Full list of Budget Features (*cont.*)

Category	Category	Category	Category
IFAD Aquis Prod Revenda	IFAD Aquis Titulos Credito	IFAD Aquis Tit Cap Integral	IFAD Const Aum Capital
IFAD Concessao Emprestimo	IFAD Dep Compulsorios	IFAD Sentenças Judiciais	IFAD Desp Exerc Anteriores
IFAD Inden e Restituições	AD Aplicações diretas	ADAD PrincipalDivContratual	ADAD PrincipalDivMobiliaria
ADAD CorreçãoDivContratua	ADAD CorreçãoDivMobiliaria	ADAD CorreçãoDivARO	ADAD PrincCorrigidoDivCont
ADAD PrincCorrigidoDivMob	ADAD Sentenças Judiciais	ADAD Desp Exerc Anteriores	ADAD Inden e Restituições
ODCAD Contribuicoes	ODCAD Subvencoes	PES Transf à União	PES Transf a Municípios
IF Transf à União	PES Transf a Consórcios Públicos	PES AD Operação entre Órgãos	ODC Transf a Consórcios Públicos
ODC AD Entre Órgãos	I Transf a Consórcios Públicos	I AD Operações entre Órgãos	IF Transf Consórcios Públicos
IF AD Operação entre Órgãos	IF Transf Privada CFL	PESAD Outros Benef Previdenc	PESAD Outros Benef Assist
PESAD FGTS	PESAD Contrib Previd - INSS	PESAD Plano de Seg Soc do Serv - Pes Ativo	PESAD Outras Obrig Patronais
PESAD Demais Obrigações Patronais	PESAD Demais Aplicações Diretas	PES AD Obrigaç Patronais Intraorç	PES AD Contrib Patron RPPS Intraorç
PES AD Outras Obrigaç Patron Intraorç	PES AD Demais Obrigaç Patron Intraorç	PES AD Demais Desp Pes Intraorç	JED Demais Aplicaç Diretas
ODCAD Demais Aplicações Diretas	IAD Obras em Andamento	IAD Demais Obras e Instalações	IAD Demais Aplicações Diretas
IFAD Demais Aplicações Diretas	ADAD Demais Aplicações Diretas	Reserva do RPPS	Reserva de Contingência
PES Transf a Inst Fin Sem Fins Lucrativos	Passivo	Passivo Financeiro	Depositos
Consignacoes	Depositos Diversos	Obrigac em Circulacao	Restos a pagar Processados
Fornecedores Ex	Fornecedores Ant	Convenios a pagar	Pessoal a pagar Ex
Pesoal a pagar Ant	Encargos Sociais RC	Provisoes Diversas	Obrigacoes Tributarias
Debitos Divers aPG	Restos a Pagar NP	Restos a Liquidar	Credores diversos
Adiantamentos recebidos	Outras Obrig a PG	Passivo Nao Financeiro	Obrigacoes em circ
Provisoes	Opc Internas	Opc Externas	Adiantam Div Receb
Outros Debitos a Pagar	Val Pend Curto Prazo	Valores Pendentes	Exigivel Longo Prazo
Dep Exig Longo P	Obrig ex longo P	LP OPC Internas	LP OPC Externas
Obrig Legais e Trib	LP Obrig a Pagar	Outras Exigibilidades	Result Futuros
Passivo Real	Patrimonio Liquido	Patrim Capital	Reservas
Resultado Acumulado	Passivo Compensado	Precatórios	Precatórios Obrig Circul
PF Valores Pendentes	PF a Curto Prazo	Obrigac em Circulacao PF	Pessoal a pagar Ant
Precatórios PF	Valores Pendentes CP PF	Opc Internas em Circul	Opc Externas em Circul
Obrigac aPagar em Circul	Precatórios Passivo NF	Precatórios Pre2000	Precatórios Pos2000
Val Pend Curto Prazo PNF	Provisões Matem Previdenc	Prov Benefícios Concedidos	Prov Benefícios a Conceder
Provisões Amortizadas	Prov Atuariais para Ajustes do Plano	OPCI em Títulos	OPCI em Contratos
OPCI Financiamentos	OPCE em Títulos	OPCE em Contratos	OPCE Financiamentos
LP OPCI em Títulos	LP OPCI em Contratos	LP OPCI Financiamentos	LP OPCE em Títulos
LP OPCE em Contratos	LP OPCE Financiamentos	Rec Orçamentária	Rec Correntes
Rec Tributária	Impostos	IPTU	ISS
ITBI	Taxas	Tx Poder de Policia	TX Prestação de Servicos
Contr de Melhoria	Rec de Contribuição	Contrib Custeio Previdência	Comp Fin 201 CF
Outras Rec de Contribuição	Rec Patrimonial	Rec Financeiras	Outras rec Patrimoniais
Rec Industrial	Rec Agropecuária	Rec de Serviços	Rec Transf Correntes
Transf Intergov da União	Cota FPM	IRRF	Cota ITR
Cota IOF Ouro	LC 8796 ICMS	Cota Salário Educação União	Fundef União
SUS União	Outras Transf da União	Transf Intergov do Estado	Cota ICMS
Cota IPVA	Cota IPI Exportação	Cota Salário Educação Estado	Fundef Estado
SUS Estado	Outras Transf Estado	Outras Transf Correntes	Demais Rec Correntes
Rec Divida Ativa	Outras Rec Correntes	Rec de Capital	Operações de Crédito
Alienação	Rec Transf de Capital	Rec Transf de Capital União	Rec Transf de Capital Estado
Outras Rec Transf Capital	Outras rec de Capital	ISSQN	Tx Prestação de Servicos
Contribuições Sociais	Contribuições Econômicas	Rec Imobiliárias	Rec Valores Mobiliários
Rec Concessões e Permissões	Transf Cor Intergovern	Comp Extrac Mineral	Cota Petróleo
FNAS	FNDE	Demais Transfer União	Transf Intergov Estado
Cota Salário Educação	Transf dos Municípios	SUS Municípios	Out Transf Municípios
Transf Multigovernamentais	Transf Multigov FUNDEF	Transf Multigov FUNDEF Comp	Transf Instit Privadas
Transf Exterior	Transf Pessoas	Transf Convênios	Transf Convênios União
Transf Convênios Estados DF	Transf Convênios Municípios	Transf Convênios Inst Privadas	Out Rec Correntes
Multas e Juros de Mora	Indeniz e Restituições	Receitas Diversas	OPC Internas
OPC Externas	Alienação de Bens	Alien Bens Móveis	Alien Bens Imóveis
Amortização de Empréstimos	Transf Cap Intergovern	Transf Cap Inter União	Transf Cap Inter Estados
Transf Cap Inter Municípios	Transf Cap de Inst Privadas	Transf Cap Exterior	Transf Cap Pessoas
Transf Cap Out Inst Públicas	Transf Cap Convênios	Transf Cap Conv União	Transf Cap Conv Estados
Transf Cap Conv Municípios	Transf Cap Conv Inst Privadas	Outras Rec Capital	Deduções Rec Corrente
Dedução FUNDEF FPM	Dedução FUNDEF LC8796	Dedução FUNDEF ICMS	Dedução FUNDEF IPI Exp
Imp s Patrimonio e Renda	Imp s Renda e Proventos	Imp s Produção e Circulação	Participação Rec União
Transf Uni CompFinanc	Cota Royalties Excedente	Cota Royalties Part Especial	Outras Transf U ComFin
Participação Rec Estados	Cota CIDE	Outras Part Rec Estado	Transf Est CompFinanc
Cota ComFin Rec Hidricos	Cota ComFin Rec Minerais	Cota Royalties Produção	Outras Transf E CompFin

Table G14: Full list of Budget Features (*cont.*)

Category	Category	Category	Category
Transf Est Saude Fundo	Outras Transf Estados	Outras Transf Multigov	TrConvUn SUS
TrConvUn Educação	TrConvUn Assist Social	TrConvUn Combate Fome	Outras TrConvUn
TrConvEst SUS	TrConvEst Educação	Outras TrConvEst	TrConvMun SUS
TrConvMun Educação	Outras TrConvMun	Transf pCombate Fome	TrCF Exterior
TrCF PJ	TrCF PF	TrCF Não Identificado	TrCapU SUS
TrCapU Educação	Outras TrCapU	TrCapEst SUS	TrCapEst Educação
Outras TrCapEst	TrCapMun SUS	TrCapMun Educação	Outras TrCapMun
TrCapConvU SUS	TrCapConvU Educação	Outras TrCapConvU	TrCapConvEst SUS
TrCapConvEst Educação	Outras TrCapConvEst	TrCapConvMun SUS	TrCapConvMun Educação
OutrasTrCapConvMun	Transf Cap Combate à Fome	TrCapCF Exterior	TrCapCF PJ
TrCapCF PF	TrCapCF Não Identificado	TrConvUn Saneamento Basico	TrCapConvU Saneamento
TrCapConvU Meio Ambiente	TrCapConvU Transporte	TrCapConvEst Saneamento	TrCapConvEst Meio Ambiente
TrCapConvEst Transporte	Compensações Financeiras	Transf Convênios Exterior	Transf Cap Inst Publicas
Transf Cap Conv Exterior	IRRF Trabalho	IRRF Outros Rendimentos	Cota-parte Comp Fin RecHídricos
Cota-parte CFEM	Cota-parte Royalties Petróleo	Transf União Consórcios Públicos	Outras Transfer União
Transf Est Consórcios Públicos	Transf Mun Consórcios Públicos	Transf Multigov FUNDEB	Transf Multigov FUNDEB Comp
TrCapU Consórcios Públicos	TrCapEst Consórcios Públicos	TrCapMun Consórcios Públicos	Dedução Rec Tr União
Dedução FUNDEB FPM	Dedução FUNDEB ITR	Dedução FUNDEB LC8796	Dedução Rec Tr Estado
Dedução FUNDEB ICMS	Dedução FUNDEB IPVA	Dedução FUNDEB IPI Exp	Rec Cor Intra-Orçamentárias
Rec Capital Intra-Orçamentárias	Contr para o Reg Prop de Prev do Serv Público	Contr Patr de Serv Ativo Civil para o Reg Prop	Contr Patr de Serv Ativo Militar
Contr Patr - Inativo Civil	Contr Patr - Inativo Militar	Contr Patr - Pens Civil	Contr Patr - Pens MIlitar
Contr do Serv Ativo Civil para o Reg Prop	Contr de Serv Ativo Militar	Contr do Serv Inativo Civil para o Reg Prop	Contr de Serv Inativo Militar
Contr de Pens Civil Reg Prop	Contr de Pens Militar	Contr Prev Amortiz do Déficit Atuarial	Contr Prev em Reg de Parcel de Débitos
Outras Contr Sociais	Demais Contr Sociais	Juros de Títulos de Renda	Dividendos
Participações	Remuneração de Dep Bancários	Remuneração de Dep Especiais	Remuneração de Saldos de Recur Não Desembol
Remun Invest Reg Próprio Previd Servidor	Outras Receitas de Valores Mobiliários	Rec Divida Ativa Tributária	Rec Divida Ativa Não Trib
Demais Deduções da Receita	Aluguéis	Arrendamentos	Foros
Laudêmios	Taxa de Ocupação de Imóveis	Outras Receitas Imobiliárias	Serviços Financeiros
Serviços de Transporte	Serviços de Saúde	Serviços de Processamento de Dados	Serviços Administrativos
Serviços Educacionais	Serviços de Fornecimento de Água	Demais Receitas de Serviços	MeJM dos Tributos
MeJM das Contribuições	MeJM da Divida Ativa dos Tributos	MeJM da Dívida Ativa das Contribuições	MeJM da Dívida Ativa de Outras Receitas
MeJM de Outras Receitas	Multas de Outras Origens	Ação Legislativa	Controle Externo
Outras Desp na Função Legislativa	Ação Judiciária	Defesa do Interesse Público	Outras Desp na Função Judiciária
Defesa da Ordem Jurídica	Representação Juridica	Outras Desp na Função Justiça	Planejamento e Orçamento
Administração Geral	Administração Financeira	Controle Interno	Normatização e Fiscalização
Tecnologia da Informação	Ordenamento Territorial	Formação de Recursos Humanos	Administração de Receitas
Administração de Concessões	Comunicação Social	Outras Desp na Função Administração	Defesa Aérea
Defesa Naval	Defesa Terrestre	OUtras Desp na Função Desfesa	Policiamento
Defesa Civil	Informação e Inteligência	Outras Desp na Função Segurança Pública	Relações Diplomáticas
Cooperação Internacional	Outras Desp na Função Relações Exteriores	Assistência ao Idoso	Assistência ao Deficiência
Assistência à Criança	Assistência Comunitária	OUtras Desp na Função Assistência Social	Previdência Básica
Previdência do Regime Estatutário	Previdência Complementar	Previdência Especial	Outras Desp na Função Previdência Social
Atenção Básica	Assistência Hospitalar	Suporte Profilático	Vigilância Sanitária
Vigilância Epidemiológica	Alimentação e Nutrição	Outras Desp na Função Saúde	Proteção ao Trabalhador
Relações de Trabalho	Empregabilidade	Fomento ao Trabalho	Outras Desp na Função Trabalho
Ensino Fundamental	Ensino Médio	Ensino Profissional	Ensino Superior
Educação Infantil	Educação de Jovens e Adultos	Educação Especial	Outras Desp na Função Educação
Patrimônio Cultural	Difusão Cultural	Outras Desp na Função Cultura	Custódia e Reintegração Social
Direitos Humanos	Assistência Povos Indígenas	Outras Desp na Função Cidadania	Infra-Estrutura Urbana
Serviços Urbanos	Transportes Coletivos Urbanos	Outras Desp na Função Urbanismo	Habitação Rural
Habitação Urbana	Outras Desp na Função Habitação	Saneamento Básico Rural	Saneamento Básico Urbano
Outras Desp na Função Saneamento	Preservação Ambiental	Controle Ambiental	Recuperação Areas Degradadas
Recursos Hídricos	Meteorologia	Outras Desp na Função Gestão Ambiental	Desenvolvimento Científico
Desenvolvimento Tecnológico	Difusão do Conhecimento Científico	Outras Desp na Função Ciência e Tecnologia	Promoção da Produção Vegetal
Promoção da Produção Animal	Defesa Sanitária Vegetal	Defesa Sanitária Animal	Abastecimento
Extensão Rural	Irrigação	Outras Desp na Função Agricultura	Reforma Agrária
Colonização	Outras Desp na Função Organização Agrária	Promoção Industrial	Produção Industrial
Mineração	Propriedade Industrial	Normalização e Qualidade	Outras Desp na Função Indústria
Comércio e Serviços	Promoção Comercial	Comercialização	Comércio Exterior
Turismo	Outras Desp na Função Comércio e Serviços	Comunicações Postais	Telecomunicações
Outras Desp na Função Comunicações	Conservação de Energia	Energia Elétrica	Petróleo
Álcool	Outras Desp na Função Energia	Transporte Aéreo	Transporte Rodoviário
Transporte Ferroviário	Transporte Hidroviário	Transportes Especiais	Outras Desp na Função Transporte
Desporto de Rendimento	Desporto Comunitário	Lazer	Outras Desp na Função Desportos e Lazer
Refinanciamento da Dívida Interna	Refinanciamento da Dívida Externa	Serviço da Dívida Interna	Serviço da Dívida Externa
Transferências	Outros Encargos Especiais	Outras Desp na Função Encargos Especiais	Outras Desp na Função Desfesa

Notes: This table provide the list of all budget features used for the prediction task. More details are provided in the ministry website: https://www.tesourotransparente.gov.br/publicacoes/finbra-dados-contabeis-dos-municipios-1989-a-2012/2012/26.