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

E. Ash, S. Galletta, T. Giommoni

Online 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	0.1	0.5	20	0.1	5	51	10319
2	0.5	0.5	10	0.1	3	74	11666
3	0.5	0.1	20	0.1	5	52	10210
4	0.1	1	10	0.1	3	90	12390
5	0.5	2	10	0.1	1	55	16411
Mean	0.34	0.82	14	0.1	3.4	64.4	12199.2

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	1862	1196
	Corrupt	1355	886

Panel C. LASSO

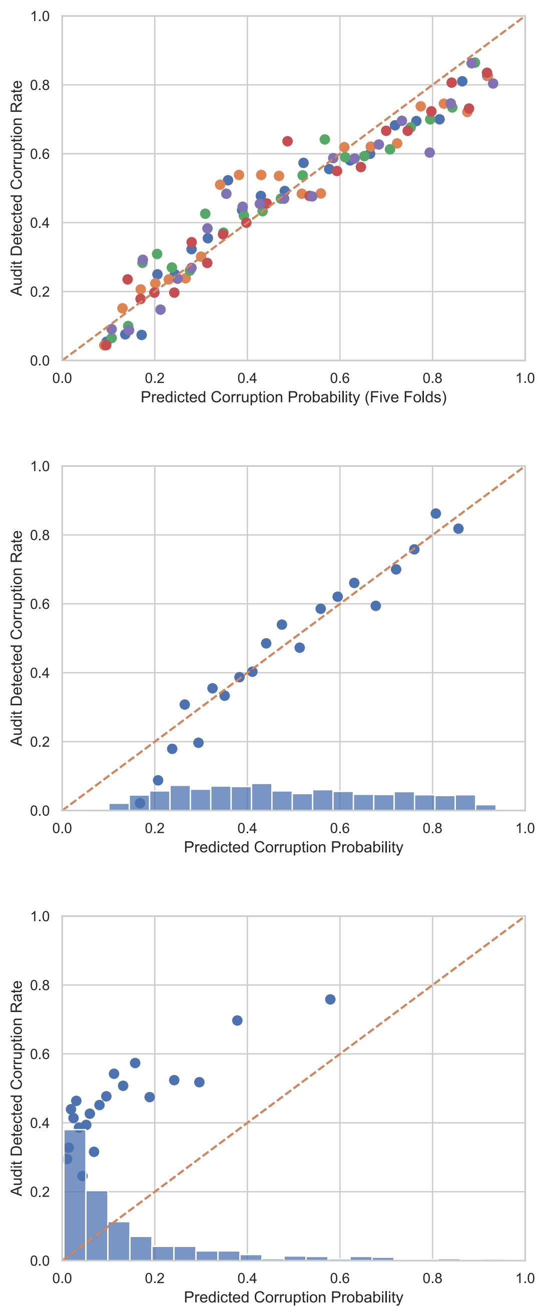
		Prediction	
		Not Corrupt	Corrupt
Truth	Not Corrupt	1392	1666
	Corrupt	1005	1236

Panel D. Logistic regression

		Prediction	
		Not Corrupt	Corrupt
Truth	Not Corrupt	2107	951
	Corrupt	1244	997

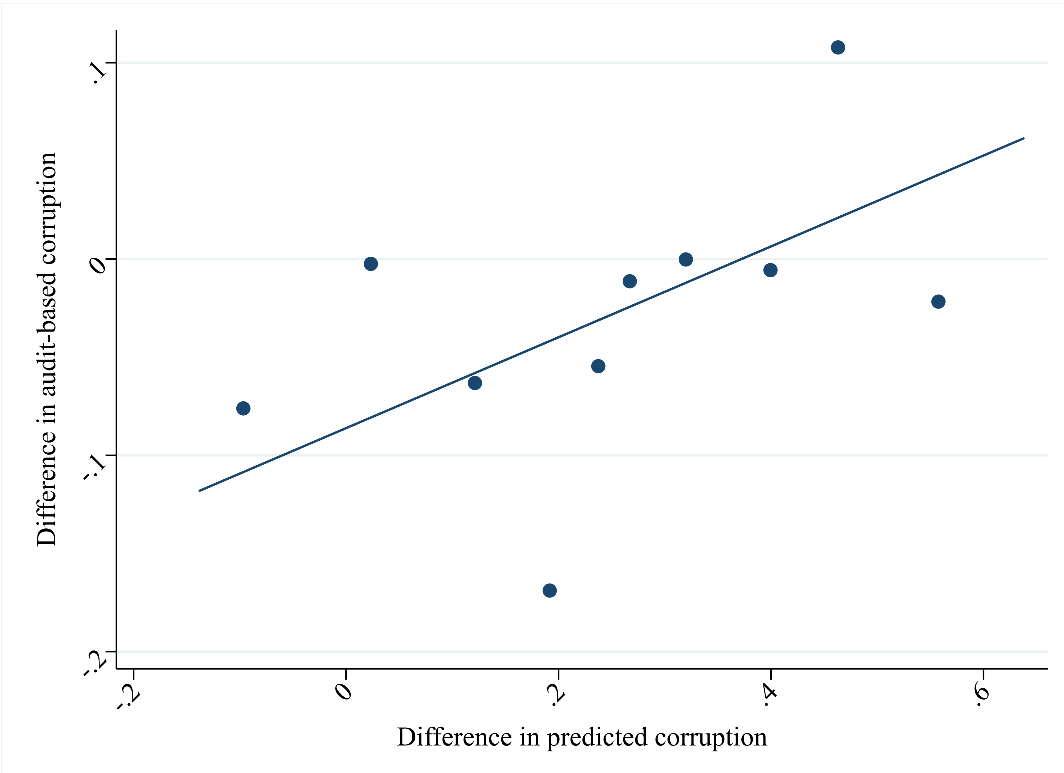
Notes: The table reports confusion matrices from the model predictions XGBoost (recall=0.56 and precision=0.72), OLS (recall=0.40 and precision=0.43), LASSO (recall= 0.55 and precision= 0.43) and Logistic regression (recall=0.44 and precision=0.51).

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: audit fixed effects, mean income, share of population employed, sector of occupation (agriculture, industry, commerce, transportation, services and public administration), share with college education, and Gini Coefficient of income. The coefficient of the corresponding regression is 0.230 (p-value 0.083). Similar results emerge in the analysis without controls (coefficient 0.239 – p-value 0.029) and without controls and fixed effects (coefficient 0.126 – p-values 0.024). 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 2, AUC-ROC = 0.638 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.

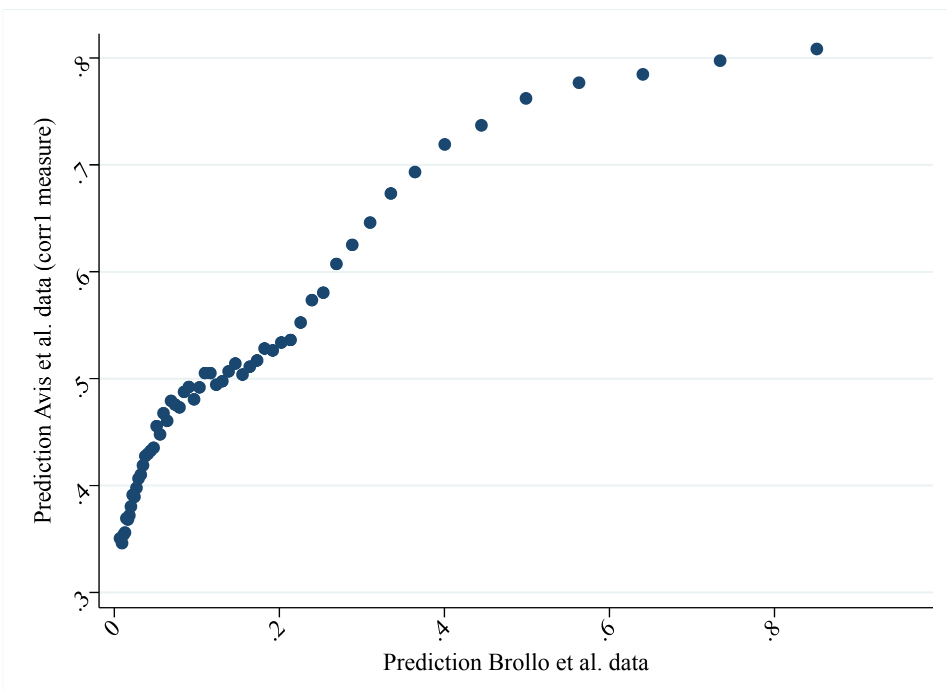
Finally, we find that most of our empirical results still hold when using the predictions from this alternative measure of corruption.

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	0.585	0.849	0.459	0.459	0.759
	[0.607-0.642]	[0.594-0.660]	[0.575-0.605]	[0.803-0.884]	[0.199-0.650]	[0.210-0.641]	[0.723-0.821]
AUC-ROC	0.631	0.636	0.594	0.900	0.517	0.474	0.648
	[0.605-0.677]	[0.605-0.698]	[0.573-0.615]	[0.851-0.904]	[0.370-0.568]	[0.392-0.609]	[0.541-0.663]
F1	0.495	0.497	0.473	0.628	0.373	0.315	0.468
	[0.462-0.522]	[0.474-0.542]	[0.443-0.507]	[0.344-0.618]	[0.175-0.354]	[0.192-0.416]	[0.212-0.457]

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 0.512.

Table C5: Most important budget features for Corruption Prediction

N.	Category	Macro Category	Weight	Perturbation Response Mean	Min	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	0.22
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.

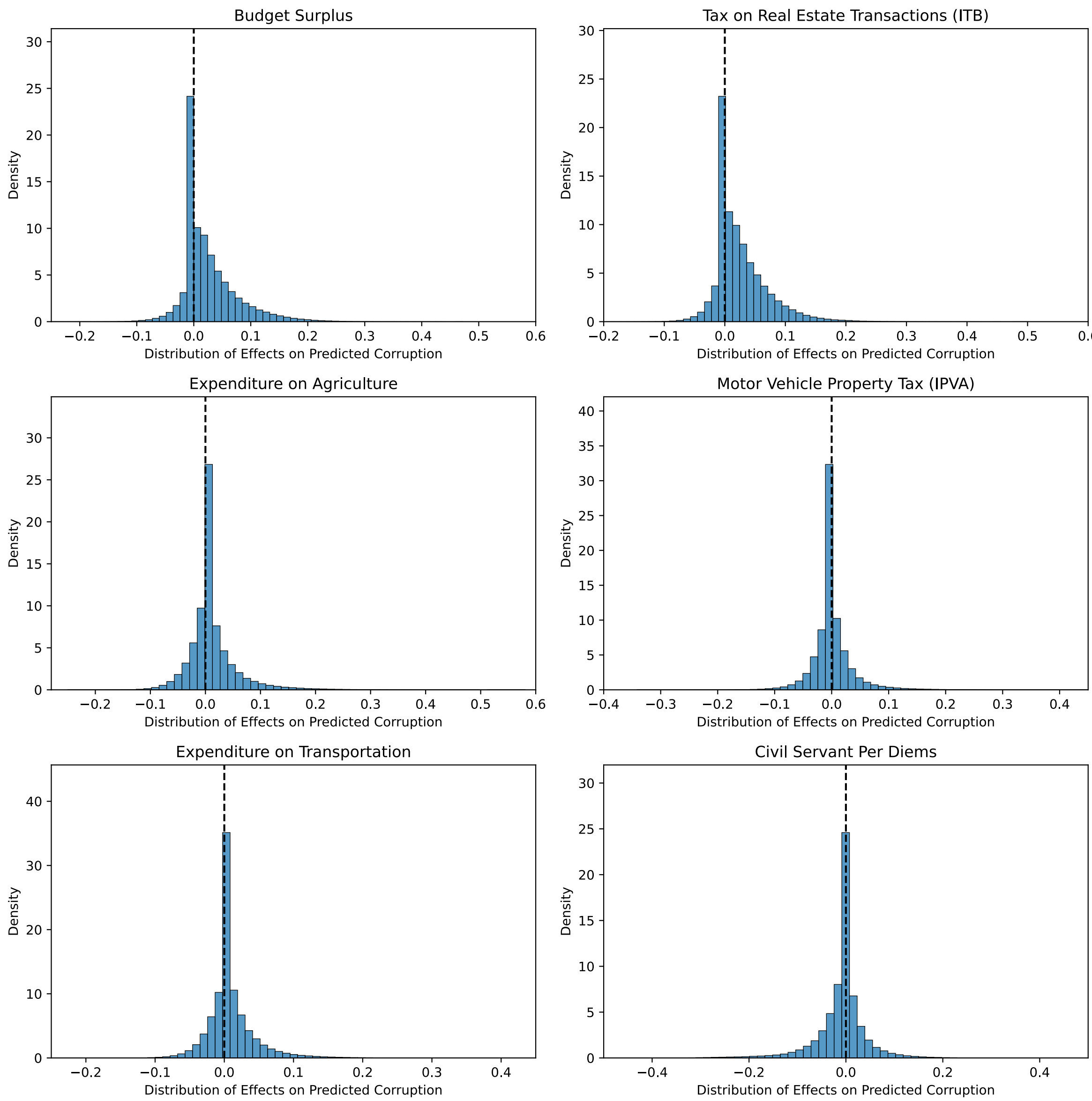
Overall, the perturbation results are consistent with a highly non-linear and contingent predictive relationship. Of the 454 predictive variables, only 31 variables have an always-weakly-positive relation, and only 24 variables have an always-weakly-negative relation. Columns 6 and 7 of Table C5 show 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

Population (1)	FPM transfers: Actual transfers (2)	FPM transfers: Theoretical transfers (3)	Predicted Corruption (4)	N (5)
6,793 − 10,188	19.655	21.200	.442	1,429
10,189 − 13,584	25.642	28.771	.501	1,076
13,585 − 16,980	31.888	36.316	.524	805
16,981 − 23,772	38.445	44.019	.540	1,083
23,773 − 30,564	44.223	51.082	.528	629
30,565 − 37,356	50.869	58.113	.521	380
37,357 − 44,148	57.376	66.468	.506	253
44,149 − 50,940	62.389	72.368	.494	154
Total	33.440	37.930	.501	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	0.6772*** (0.0579)	0.6158*** (0.0713)	0.6521*** (0.0686)	0.6405*** (0.0744)
Panel B. Reduced Form				
Theoretical transfers	0.0038*** (0.0007)	0.0032*** (0.0006)	0.0035*** (0.0007)	0.0037*** (0.0007)
Panel C. 2SLS				
Actual transfers	0.0056*** (0.0011)	0.0053*** (0.0011)	0.0054*** (0.0012)	0.0058*** (0.0012)
N. Observations	1101	1101	1101	1101

Notes: Effects of FPM transfers on (predicted) corruption measures. The four columns display the analysis focusing on four different random samples with 1,101 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	Prediction budget (without FPM)
	(1)	(2)
Panel A. Reduced Form		
Theoretical transfers	0.0000	0.0044***
	(0.0006)	(0.0003)
Panel B. 2SLS		
Actual transfers	0.0000	0.0063***
	(0.0009)	(0.0005)
N. Observations	5809	5809

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)	(0.0005)
	[0.0018 ; 0.0089]	[0.0034 ; 0.0054]	[0.003 ; 0.005]
2SLS			
Actual transfers	0.0078	0.006	0.0059
	(0.0027)	(0.0008)	(0.0007)
	[0.0025 ; 0.013]	[0.0049 ; 0.0079]	[0.0045 ; 0.0073]
N. Observations	1101	5809	4708
Panel B. Held-out sample 0			
Reduced Form			
Prescribed transfers	0.0042***	0.0037***	0.0035***
	(0.0011)	(0.0004)	(0.0004)
2SLS			
Actual transfers	0.0062***	0.0054***	0.005***
	(0.0015)	(0.0005)	(0.0006)
N. Observations	1069	5777	4708
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	1078	5786	4708

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**	-0.0070
	(0.0238)	(0.0385)	(0.1085)
Year pre3	-0.0169	-0.0372	0.0093
	(0.0163)	(0.0241)	(0.0744)
Year pre2	-0.0137	-0.0084	-0.0304
	(0.0119)	(0.0214)	(0.0383)
Audit year	-0.0292**	-0.0110	-0.0620
	(0.0136)	(0.0208)	(0.0391)
Year post1	-0.0326	-0.0611**	-0.0408
	(0.0192)	(0.0288)	(0.0487)
Year post2	-0.0287	-0.1105***	-0.0133
	(0.0262)	(0.0351)	(0.0598)
Year post3	-0.0197	-0.1287***	0.0531
	(0.0300)	(0.0451)	(0.0907)
Year post4	-0.0197	-0.1597**	0.0539
	(0.0372)	(0.0576)	(0.1208)
Year post5	-0.0214	-0.1840**	0.1030
	(0.0435)	(0.0720)	(0.1393)
Years post6 and more	-0.0229	-0.2003**	0.1135
	(0.0489)	(0.0831)	(0.1534)
N. Observations	17252	8895	3187
Adjusted R^2	0.535	0.507	0.525

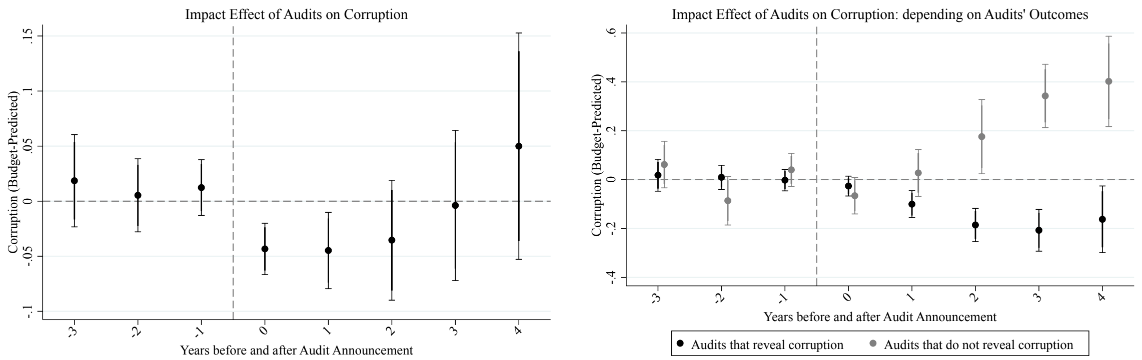
Notes: The dependent variable is (predicted) corruption measure - binary. The sample includes all the cities that receive an audit for the period 2001-2012. Column (1) includes the complete sample, Column (2) includes the sample of cities in which the audit discovered corruption (according to the definition 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	-0.024	-0.172	0.135	0.0577**	0.0033	0.1293	-0.0383	-0.0831**	0.0851
	(0.028)	(0.039)	(0.082)	(0.0264)	(0.0401)	(0.0888)	(0.0288)	(0.0379)	(0.1018)
	[-0.078 ; 0.031]	[-0.248 ; -0.096]	[-0.027 ; 0.296]						
Year pre3	0.021	-0.029	0.046	0.0516**	0.0120	0.0794	-0.0102	-0.0062	0.0562
	(0.019)	(0.026)	(0.054)	(0.0235)	(0.0347)	(0.0608)	(0.0186)	(0.0279)	(0.0658)
	[-0.016 ; 0.058]	[-0.08 ; 0.022]	[-0.06 ; 0.152]						
Year pre2	0.013	0.009	0.016	0.0121	-0.0059	0.0482	-0.0022	-0.0075	-0.0330
	(0.011)	(0.019)	(0.030)	(0.0211)	(0.0253)	(0.0428)	(0.0132)	(0.0254)	(0.0439)
	[-0.009 ; 0.035]	[-0.028 ; 0.046]	[-0.043 ; 0.075]						
Audit year	-0.075	-0.294	0.091	-0.0347*	-0.0428*	0.0054	-0.0108	-0.0384**	0.0216
	(0.015)	(0.030)	(0.035)	(0.0185)	(0.0249)	(0.0456)	(0.0155)	(0.0174)	(0.0306)
	[-0.104 ; -0.046]	[-0.353 ; -0.235]	[0.022 ; 0.16]						
Year post1	-0.081	-0.392	0.128	-0.0386*	-0.0792**	-0.0133	-0.0141	-0.1015***	0.0220
	(0.020)	(0.041)	(0.057)	(0.0201)	(0.0368)	(0.0701)	(0.0226)	(0.0178)	(0.0515)
	[-0.120 ; -0.042]	[-0.472 ; -0.312]	[0.016 ; 0.24]						
Year post2	-0.092	-0.490	0.190	-0.0304	-0.1135**	0.0440	-0.0113	-0.1239***	0.0510
	(0.026)	(0.051)	(0.078)	(0.0256)	(0.0431)	(0.0810)	(0.0252)	(0.0260)	(0.0776)
	[-0.143 ; -0.041]	[-0.590 ; -0.390]	[0.037 ; 0.343]						
Year post3	-0.101	-0.573	0.232	-0.0149	-0.1292**	0.0837	-0.0044	-0.1341***	0.0826
	(0.032)	(0.059)	(0.095)	(0.0323)	(0.0478)	(0.1259)	(0.0314)	(0.0296)	(0.0762)
	[-0.164 ; -0.038]	[-0.689 ; -0.457]	[0.046 ; 0.418]						
Year post4	-0.135	-0.672	0.233	-0.0170	-0.1313**	0.1181	-0.0110	-0.1660***	0.0848
	(0.039)	(0.067)	(0.116)	(0.0353)	(0.0532)	(0.1257)	(0.0399)	(0.0400)	(0.1158)
	[-0.211 ; -0.059]	[-0.803 ; -0.541]	[0.006 ; 0.46]						
Year post5	-0.153	-0.760	0.268	-0.0326	-0.1551**	0.1190	-0.0000	-0.1832***	0.1438
	(0.045)	(0.078)	(0.140)	(0.0457)	(0.0668)	(0.1416)	(0.0419)	(0.0502)	(0.1160)
	[-0.241 ; -0.065]	[-0.913 ; -0.607]	[-0.006 ; 0.542]						
Year post6 and more	-0.194	-0.921	0.280	-0.0332	-0.2141**	0.1465	-0.0083	-0.2256***	0.1492
	(0.056)	(0.089)	(0.158)	(0.0534)	(0.0809)	(0.1722)	(0.0594)	(0.0643)	(0.1514)
	[-0.304 ; -0.084]	[-1.095 ; -0.747]	[-0.03 ; 0.59]						
Observations	17252	8895	3187	14603	7424	2833	14602	7386	2797
Adjusted R^2	0.535	0.507	0.526						

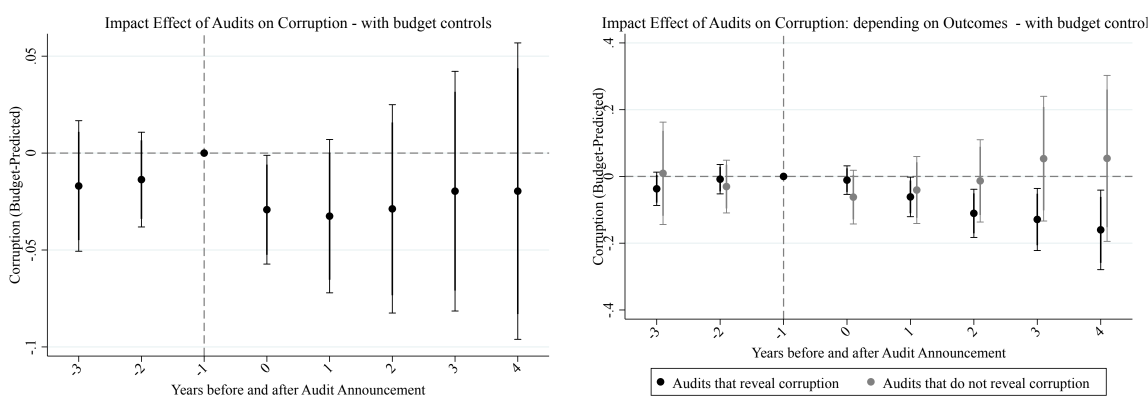
Notes: The dependent variable is (predicted) corruption measure - binary. The sample includes all the cities that receive an audit for the period 2001-2012. Columns (1), (4) and (7) include the complete sample, Columns (2), (5) and (8) include the sample of cities in which the audit discovered corruption (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)



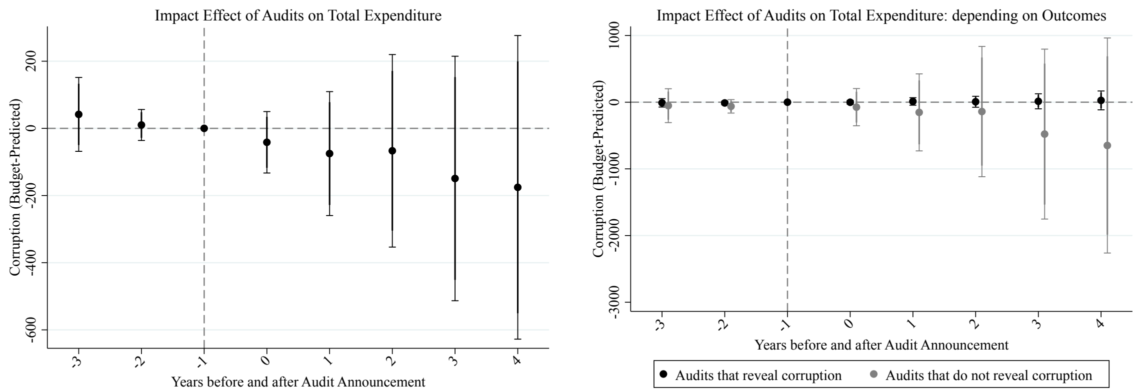
Notes: Event study estimates for dynamic effect of audits on budget-predicted corruption according to the methodology of Callaway and Sant'Anna (2021). Error spikes give 95% confidence intervals, with standard error clustered by state. Left panel: all audits; right panel: audits that found corruption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

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



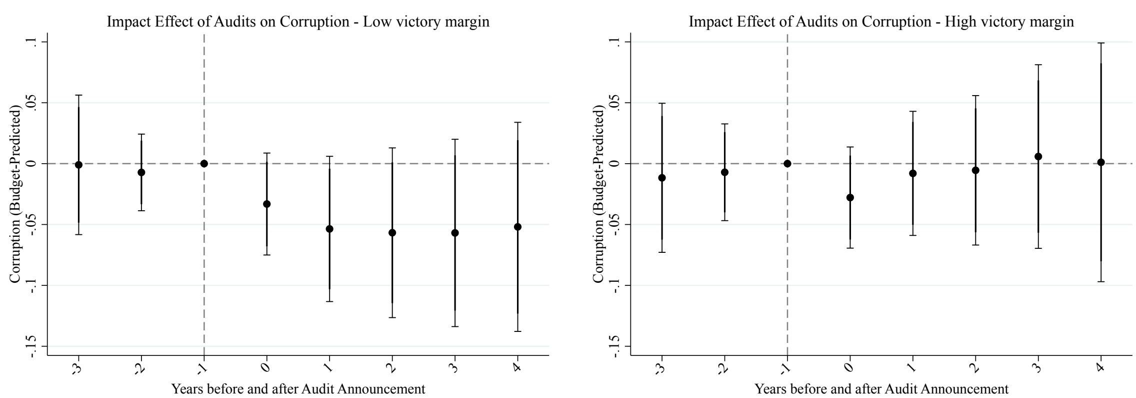
Notes: Event study estimates for dynamic effect of audits on budget-predicted corruption. Error spikes give 95% confidence intervals, with standard error clustered by state. Left panel: all audits; right panel: audits that found corruption (in black); audits that did not find corruption (in grey). This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

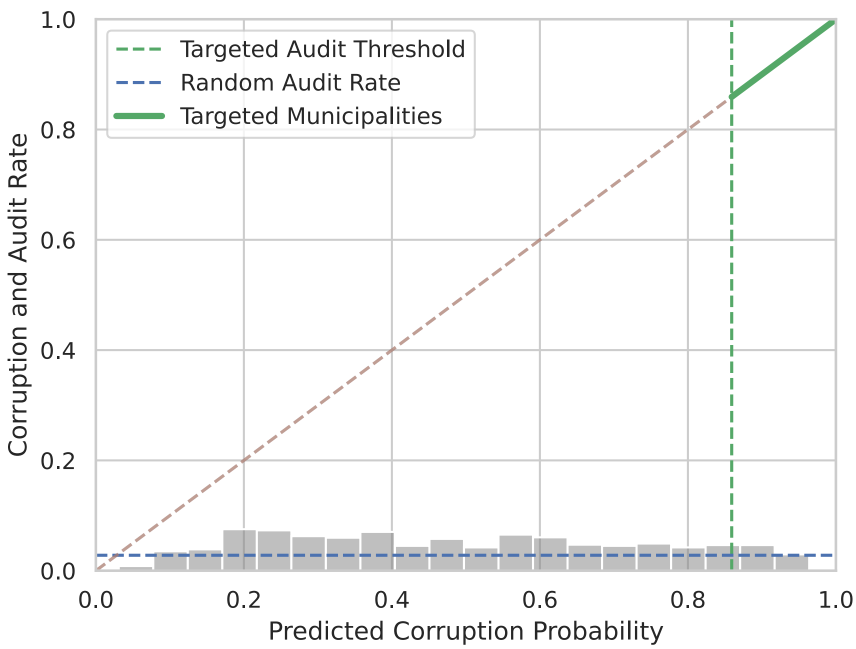
Figure E8: Dynamic effect of the audits - Margin of victory



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)

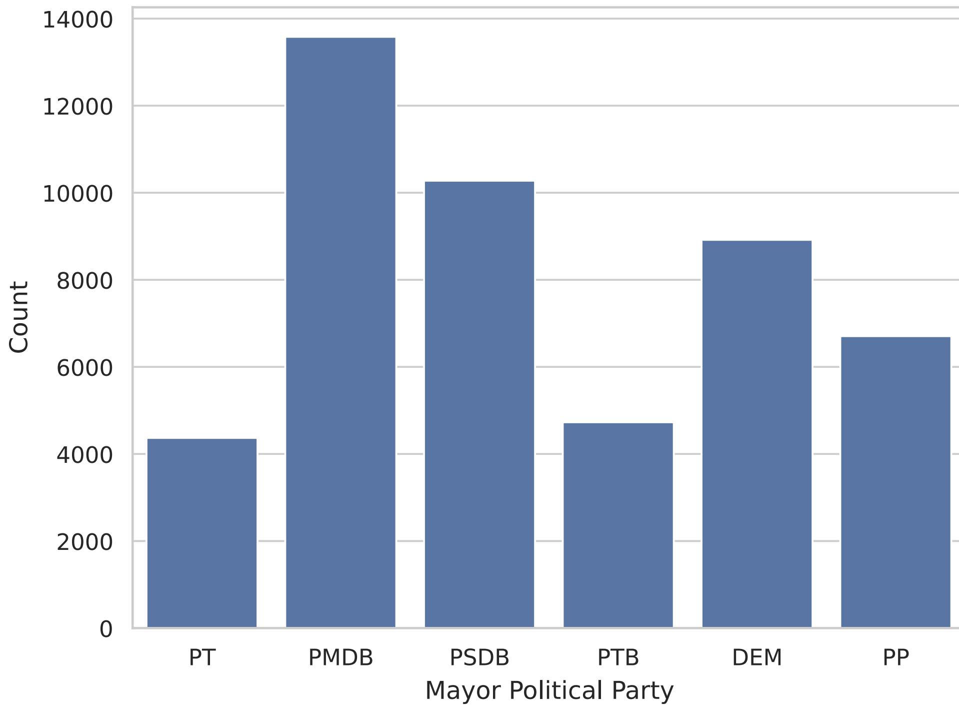
F. Additional Material on Audit Policy Support

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 (0.859) 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.