Bad Neighbors Make Good Fences: How Politicians Mitigate the Electoral Consequences of Nearby Corruption in Brazil

Appendix

Gustavo Diaz*

May 14, 2021

A. Coding Audits Before 2006

Protocol

I use text data from the audit reports as a bridge between labeled and unlabeled cases. I use a bag-of-words approach to predict the sum of moderate and severe infractions, divided by the number of service orders. The predictors are raw word counts. I use the following protocol:

1. Match text data from the audit reports with CGU infraction labels for the 2006-2015 period. This is the period where the CGU coding is available.

2. Predictors are word counts, omitting infrequent terms (words missing in more than 99% of the documents).

3. This leaves a data set with 1226 observations and 11386 variables.

4. Randomly split data in training (75%) and test (25%) sets.

5. Fit multiple random forest on training data with a grid of tuning parameters, choose the model and tuning parameters with the lowest RMSE, create predicted variable in test set.

I chose random forests because they achieve reasonable performance with the current data. I explored including topic membership covariates from structural topic modeling to assist the algorithm, but the predictive gains

*PhD Candidate. Department of Political Science. University of Illinois at Urbana-Champaign. E-mail: diazdia2@illinois.edu

are minimal. An alternative is to use algorithms from the deep-learning family, but trial runs suggest that the sample size is too small to guarantee convergence.

One way to increase predictive power dramatically would be to turn this from a regression problem into a classification task by separating documents into findings. That is, moving from predicting numbers of infractions at the document level to predicting whether each item counts as a formal, moderate, or severe infraction. This yields a larger training set with more information, and also supervised learning algorithms tend to perform better with classification tasks than with continuous outcomes. However, because audit report formats are not stable over time, dividing documents at the finding level would require prohibitively expensive human coding.

Performance

Figure A1 reports reports performance in the test set (N = 319). In average, the predicted values are off by 1.34 infractions per service order compared to the actual values. The predictions map close to a 1:1 relationship for moderate cases of corruption, but tend to underestimate it for large outliers. This implies that models using this variable will underestimate the effect of nearby corruption on the outcomes of interest, making it harder to detect non-zero estimates.

Validation

As a validation exercise, I reproduce the findings in previous work using the machine coded categories. Rundlett (2018) shows that exposing corruption has a negative effect on incumbent vote only for the 2004 elections. Table A1 replicates the same pattern using my own data set. This is different from the main analysis in that it evaluates direct effects: Whether revealing corruption in a municipality affects votes for the incumbent in that municipality. The substantive result is the same as in previous work.

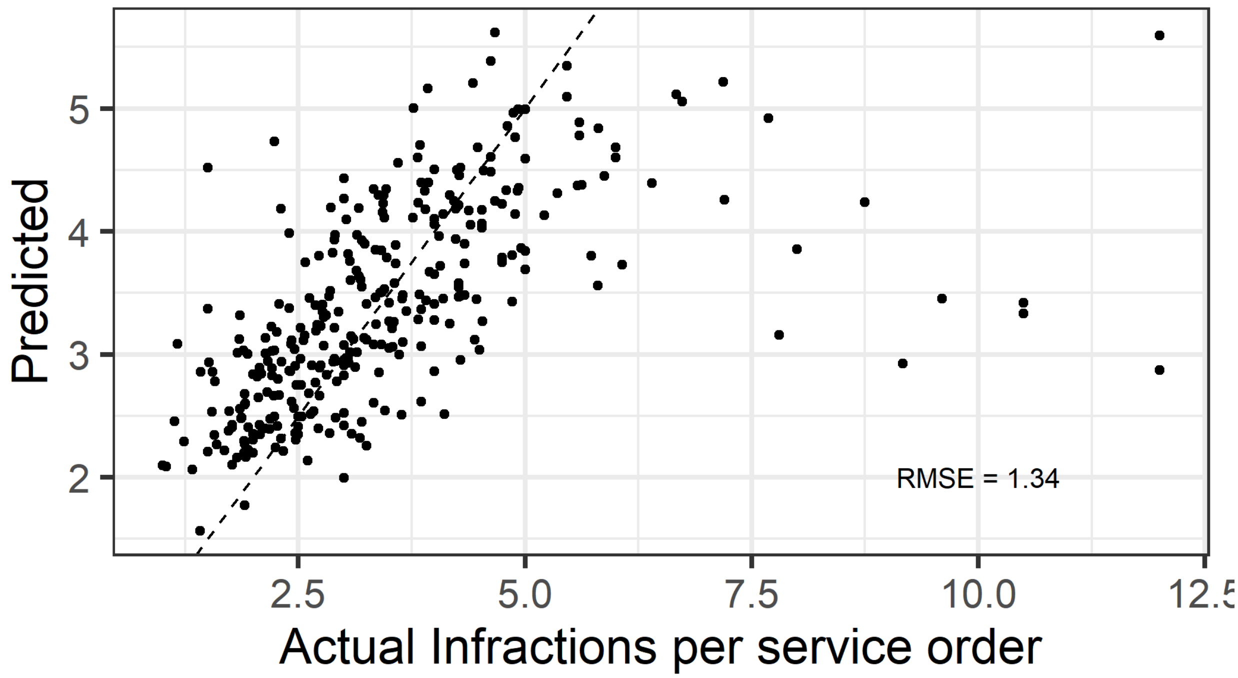

Figure A1: Actual vs. predicted corruption variable in the test set. The dashed line denotes the 1:1 relationship. The algorithm does performs well at predicting moderate levels of corruption, but tends to underestimate large outliers.

	1. Pooled	2. 2004	3. 2008	4. 2012	5. 2016
Intercept		0.54*	0.49*	0.48*	0.44*
		(0.06)	(0.03)	(0.02)	(0.06)
Infractions	0.00	−0.03	0.00	0.00	0.00
	(0.00)	(0.02)	(0.01)	(0.01)	(0.02)
R^2	0.01	0.01	0.00	0.00	0.00
Adj. R^2	0.01	0.00	−0.00	−0.00	−0.01
Num. obs.	1133	193	463	341	136
RMSE	0.16	0.14	0.17	0.14	0.21
N Clusters	4				

$^*p < 0.05$

Table A1: Replication of Rundlett (2018) with my data. The first column includes election year fixed effects and clustered standard errors by election year. The remaining columns include robust (HC1) standard errors.

B. Result tables

The main text reports results using figures. This section shows tables with numerical results that underlie those figures. The list below shows the correspondence:

- Table B1 summarizes the output of the cross-validation procedure described in section 4.3.1 in the main text
- Table B2 shows the results of Figure 2 in the main text: The effect of nearby corruption on party switching by audit status using different definitions of nearby.
- Tables B3 and B4 show the results in Figure 3: The effect of nearby corruption on seeking and winning reelection.
- Table B5 shows the results in Figure 4: The effect of nearby corruption in interaction with the proportion of same-party audited mayors on party switching. Note that Figure 3 shows simulated marginal effects at discrete margins based on this estimation, the interaction term is indistinguishable from zero.
- Table B6 shows the results of Figure 5: The effect of nearby corruption on winning reelection conditional on party switching among non-audited municipalities.

Unless otherwise specified, the columns in regression tables denote the corresponding contiguity order.

Contiguity	RMSE	SD
1	0.3449	0.0176
2	0.3392	0.0107
3	0.3373	0.0079
4	0.3374	0.0125
5	0.3348	0.0078
6	0.3354	0.0077
7	0.3357	0.0100
8	0.3359	0.0096
9	0.3359	0.0060
10	0.3378	0.0080

Table B1: Results of model selection via 10-fold cross-validation

Contiguity	Infractions	SE	p-value	Audited	SE	p-value	Interaction	SE	p-value	N	Adj. R-squared
1	0.008	0.004	0.121	0.022	0.045	0.655	-0.004	0.016	0.792	6246	0.025
2	0.007	0.002	0.024	0.027	0.009	0.061	-0.005	0.002	0.106	12025	0.021
3	0.004	0.001	0.044	0.004	0.014	0.804	-0.001	0.002	0.753	14400	0.019
4	0.005	0.001	0.031	0.010	0.013	0.486	-0.001	0.001	0.450	14873	0.020
5	0.005	0.001	0.017	0.001	0.012	0.906	-0.000	0.001	0.810	14657	0.020
6	0.005	0.001	0.016	0.016	0.010	0.196	-0.001	0.001	0.197	14269	0.021
7	0.005	0.001	0.014	0.020	0.009	0.110	-0.001	0.000	0.066	13798	0.021
8	0.005	0.001	0.012	0.034	0.007	0.017	-0.002	0.000	0.004	13246	0.021
9	0.004	0.001	0.012	0.028	0.008	0.045	-0.001	0.000	0.071	12592	0.021
10	0.004	0.001	0.012	0.037	0.017	0.119	-0.001	0.001	0.163	11890	0.021

Table B2: The effect of nearby corruption on party switching by audit status using different operationalizations of nearby

	2	3	4	5
Infractions	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Audited	−0.01	−0.03	−0.03	−0.07
	(0.02)	(0.02)	(0.02)	(0.03)
Interaction	−0.00	0.00	0.00	0.00
	(0.01)	(0.00)	(0.00)	(0.00)
R^2	0.01	0.02	0.02	0.02
Adj. R^2	0.01	0.02	0.02	0.02
Num. obs.	12025	14400	14873	14657
RMSE	0.49	0.49	0.49	0.49
N Clusters	4	4	4	4

$^*p < 0.05$

Table B3: Effect of nearby corruption on whether the incumbent mayor seeks reelection

	Model 1	Model 2	Model 3	Model 4
cornb_pre_sum_2	−0.00			
	(0.00)			
audit_pre_all	−0.01	−0.04	−0.04	−0.05
	(0.02)	(0.03)	(0.02)	(0.02)
cornb_pre_sum_2:audit_pre_all	0.00			
	(0.00)			
cornb_pre_sum_3		−0.00		
		(0.00)		
cornb_pre_sum_3:audit_pre_all		0.01		
		(0.00)		
cornb_pre_sum_4			−0.00	
			(0.00)	
cornb_pre_sum_4:audit_pre_all			0.00	
			(0.00)	
cornb_pre_sum_5				−0.00
				(0.00)
cornb_pre_sum_5:audit_pre_all				0.00
				(0.00)
R^2	0.04	0.04	0.04	0.04
Adj. R^2	0.04	0.04	0.04	0.04
Num. obs.	12025	14400	14873	14657
RMSE	0.31	0.31	0.32	0.31
N Clusters	4	4	4	4

$^*p < 0.05$

Table B4: Effect of nearby corruption on whether the incumbent mayor wins reelection

	2	3	4	5
Infractions	0.01	0.01*	0.01*	0.01*
	(0.00)	(0.00)	(0.00)	(0.00)
Prop. same party	−0.06	−0.05	−0.02	−0.08
	(0.04)	(0.03)	(0.04)	(0.05)
Interaction	−0.00	−0.01	−0.02	−0.01
	(0.01)	(0.01)	(0.01)	(0.01)
R^2	0.03	0.02	0.03	0.03
Adj. R^2	0.02	0.02	0.03	0.03
Num. obs.	10805	13032	13477	13306
RMSE	0.33	0.33	0.33	0.33
N Clusters	4	4	4	4

$^*p < 0.05$

Table B5: Effect of nearby corruption on party switching among non-audited municipalities in interaction with the proportion of audited municipalities with mayors from the same party as the incumbent

	2	3	4	5
Infractions	−0.01*	−0.01*	−0.00*	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Party switch	0.23*	0.21*	0.20*	0.20*
	(0.03)	(0.02)	(0.03)	(0.05)
Interaction	0.00	0.00	0.01	0.00
	(0.01)	(0.00)	(0.00)	(0.00)
R^2	0.11	0.10	0.11	0.11
Adj. R^2	0.11	0.10	0.11	0.11
Num. obs.	10808	13035	13480	13311
RMSE	0.30	0.31	0.31	0.31
N Clusters	4	4	4	4

$^*p < 0.05$

Table B6: Effect of nearby corruption on winning reelection among non-audited municipalities in interaction with party switching

C. Descriptive statistics and robustness checks

- Table C1 compares non-audited and audited municipalities across selected covariates. Most differences in means are indistinguishable from zero or negligible.
- Table C2 reproduces the results from Figure 2 in the main text and table B1 using logistic regression.
- Table C3 shows the effect of nearby corruption using a specification similar to Figure 2, focusing on the second contiguity order, but separating the analysis by election year. Results do not depend on the inclusion of machine-coded corruption before the 2004 election.
- Figure C1 shows the distribution of audited and total number of neighbors by contiguity order, as a complement to Figure 1 in the main text
- Figure C2 shows sensitivity analyses for the effect of nearby corruption on party switching among non-audited municipalities (cf. Figure 2 in the main text) following the partial R^2 approach of Cinelli and Hazlett (2020). The logic is to entertain how much of the residual variance, in terms of partial R^2 in a regression model, an unobserved confounder would need to explain in either the outcome or explanatory variable to bring the observed effect towards zero. The figure suggests that, given the model specification and choice contiguity upper bound, an unobserved confounder would have to explain more than 50% of the partial R^2 in either party switching or nearby corruption to eliminate the effect reported in Figure 2 of the main text.

	Non-Audited	Audited	Difference	p-value
Neighbors	20.98	20.45	0.52	-3.44
Audited neighbors	0.72	0.89	-0.17	0.44
Infractions per neighbor	1.58	1.86	-0.28	0.40
Population (thousands)	34.64	26.12	8.52	-155.56
Female population	0.49	0.49	0.00	-0.00
Rural population	0.39	0.39	-0.00	-0.00
Human Development Index	0.69	0.69	0.01	-0.02
GDP per capita	13.58	12.22	1.37	-13.44
Welfare benefits per capita	0.11	0.10	0.01	-0.02
Share illiterate	0.22	0.24	-0.01	0.02
Share with college degree	0.03	0.03	0.00	-0.00
Previous vote margin	0.15	0.15	-0.00	0.00
PT incumbent	0.08	0.07	0.01	-0.01
PSDB incumbent	0.15	0.14	0.02	-0.04
Minas Gerais	0.14	0.10	0.04	-0.16
São Paulo	0.12	0.09	0.03	-0.08
Northeast	0.35	0.39	-0.04	0.05
2004 election	0.22	0.14	0.08	-0.34
2008 election	0.29	0.50	-0.21	0.24
2012 election	0.28	0.27	0.01	-0.01
2016 election	0.21	0.09	0.13	-0.68

Table C1: Comparing non-audited and audited municipalities across selected covariates. P-values adjusted for false discovery rate

Contiguity	Infractions	SE	p-value	Audited	SE	p-value	Interaction	SE	p-value
1	0.078	0.036	0.031	0.179	0.390	0.647	-0.037	0.036	0.031
2	0.064	0.004	0.000	0.252	0.077	0.001	-0.047	0.004	0.000
3	0.040	0.010	0.000	0.039	0.141	0.785	-0.006	0.010	0.000
4	0.049	0.009	0.000	0.094	0.136	0.489	-0.011	0.009	0.000
5	0.048	0.007	0.000	0.015	0.154	0.920	-0.002	0.007	0.000
6	0.048	0.007	0.000	0.157	0.125	0.210	-0.012	0.007	0.000
7	0.045	0.006	0.000	0.196	0.123	0.109	-0.012	0.006	0.000
8	0.042	0.005	0.000	0.327	0.078	0.000	-0.017	0.005	0.000
9	0.038	0.005	0.000	0.273	0.115	0.018	-0.013	0.005	0.000
10	0.036	0.004	0.000	0.351	0.173	0.043	-0.014	0.004	0.000

Table C2: Reproducing results from Figure 2 and Table B1 using logistic regression.

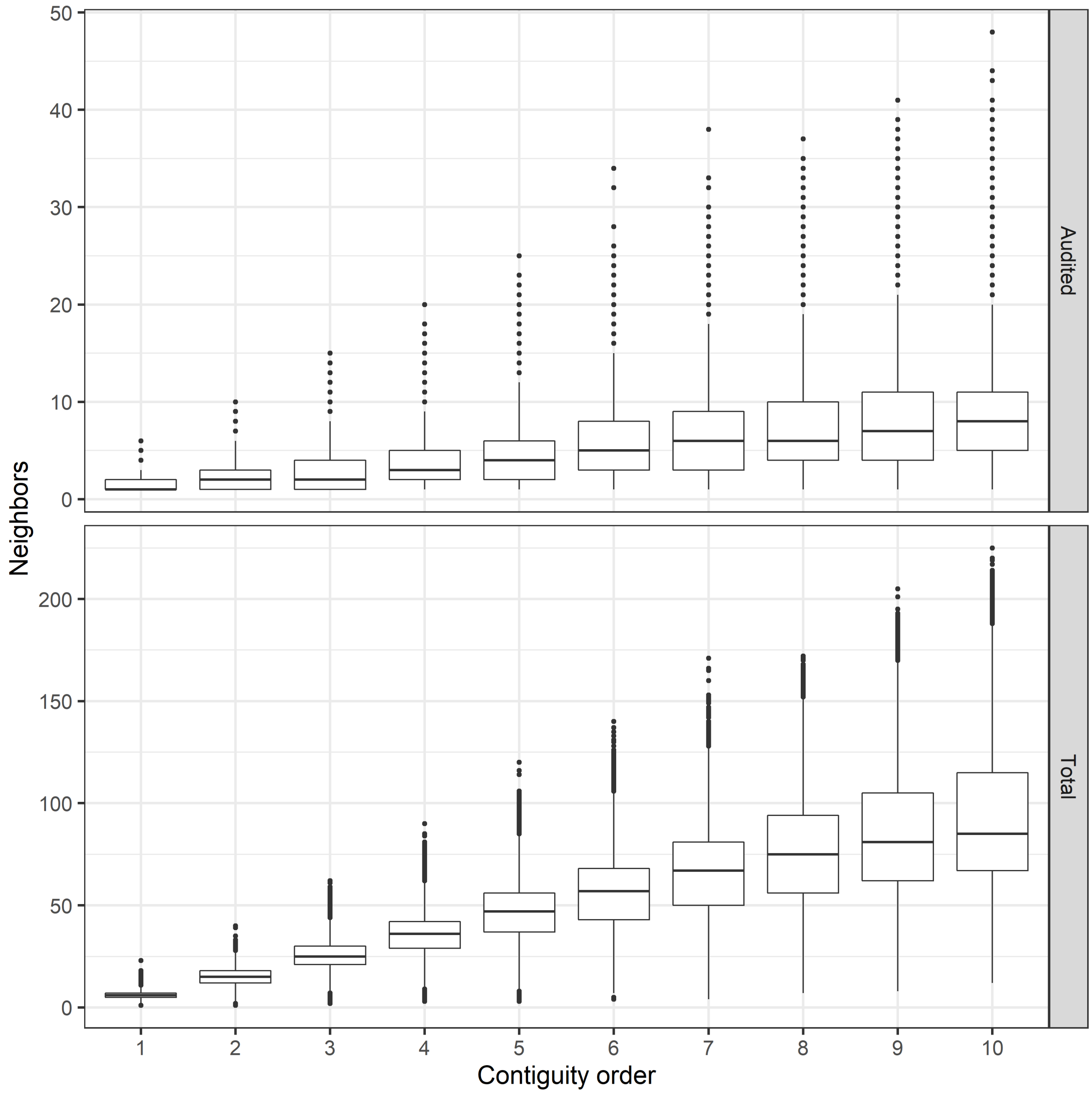

Figure C1: Distribution of audited and total number of neighbors by contiguity order.

	2004	2008	2012	2016
Intercept	0.09*	0.16*	0.05*	0.11*
	(0.02)	(0.02)	(0.01)	(0.01)
Infractions	0.01*	0.01*	0.00*	0.01*
	(0.01)	(0.01)	(0.00)	(0.00)
Audited	0.04	0.03	0.03	0.07
	(0.07)	(0.06)	(0.03)	(0.07)
Interaction	−0.01	−0.01	−0.00	−0.01
	(0.02)	(0.01)	(0.01)	(0.02)
R^2	0.00	0.00	0.00	0.00
Adj. R^2	0.00	0.00	0.00	0.00
Num. obs.	2346	2877	3674	3128
RMSE	0.33	0.41	0.26	0.35

$^*p < 0.05$

Table C3: Effect of nearby corruption at the second cumulative contiguity order on party switching by election year.

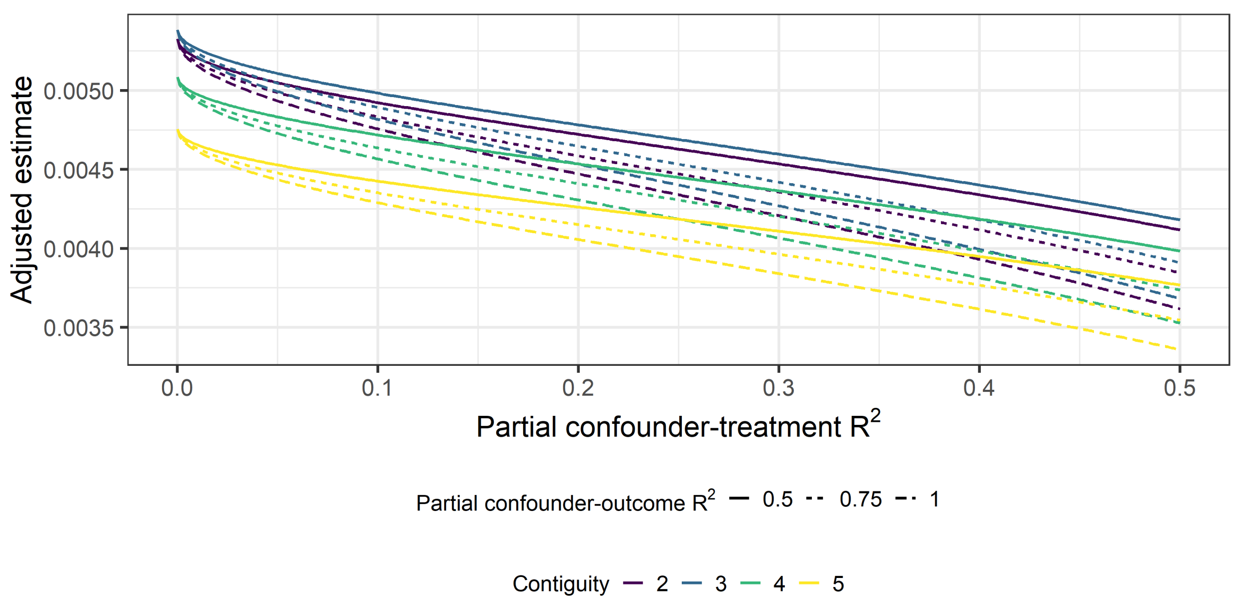

Figure C2: Sensitivity analysis for the effect of nearby corruption on party switching among non-aduited municipalities across optimal upper bounds suggested by cross-validation

References

Cinelli, Carlos, and Chad Hazlett. 2020. "Making Sense of Sensitivity: Extending Omitted Variable Bias." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82 (1): 39–67. https://doi.org/10.1111/rssb.12348.

Rundlett, Ashlea P. 2018. "The Effects of Revealed Corruption on Voter Attitudes and Participation: Evidence from Brazil." Ph.{D}. {Dissertation}, University of Illinois at Urbana-Champaign. http://hdl.handle.net/2142/101330.