Building a Robot Judge: Data Science for Decision-Making

10. Algorithms and Decisions II

Weekly Q&A

`https://bitly.com/BRJ_Padlet10`

Recap: Solution to Selective Labeling

See https://cs.stanford.edu/~jure/pubs/contraction-kdd17.pdf

- ***What is the selective labeling problem? ("raise hand" via zoom)***

Recap: Solution to Selective Labeling

See `https://cs.stanford.edu/~jure/pubs/contraction-kdd17.pdf`

- ***What is the selective labeling problem? ("raise hand" via zoom)***
- Take judge(s) with highest bail release rate (most lenient judges).
  - train recidivism prediction model with these judges.

Recap: Solution to Selective Labeling

See `https://cs.stanford.edu/~jure/pubs/contraction-kdd17.pdf`

- ***What is the selective labeling problem? ("raise hand" via zoom)***
- Take judge(s) with highest bail release rate (most lenient judges).
  - train recidivism prediction model with these judges.
- For the rest of the judges, use model trained on lenient-judge dataset to make recidivism predictions for the jailed defendants.

Recap: Solution to Selective Labeling

See `https://cs.stanford.edu/~jure/pubs/contraction-kdd17.pdf`

- ***What is the selective labeling problem? ("raise hand" via zoom)***
- Take judge(s) with highest bail release rate (most lenient judges).
  - train recidivism prediction model with these judges.
- For the rest of the judges, use model trained on lenient-judge dataset to make recidivism predictions for the jailed defendants.
  - Note: still cannot get unbiased recidivism predictions for defendants jailed by the lenient judges.

Recap: Solution to Selective Labeling

See `https://cs.stanford.edu/~jure/pubs/contraction-kdd17.pdf`

- ***What is the selective labeling problem? ("raise hand" via zoom)***
- Take judge(s) with highest bail release rate (most lenient judges).
  - train recidivism prediction model with these judges.
- For the rest of the judges, use model trained on lenient-judge dataset to make recidivism predictions for the jailed defendants.
  - Note: still cannot get unbiased recidivism predictions for defendants jailed by the lenient judges.
- ***Why couldn't we do this in the homework assignment? ("raise hand" via zoom)***

Recap: Comparing Machine (Left Panel) to Human Judges (Right Panel)

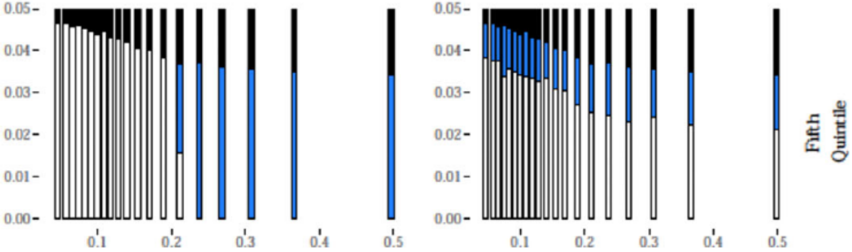


FIGURE VI

Who Do Stricter Judges Jail and Who Would the Algorithm Jail? Comparing Predicted Risk Distributions across Leniency Quintiles

- black = even most lenient judges (bottom quintile) would jail this defendant.
- blue = additional jailed by the strictest judges (top quintile). left panel = algorithm, right panel = human judges.
- white = who is released by all judges

Recap: Comparing Machine (Left Panel) to Human Judges (Right Panel)

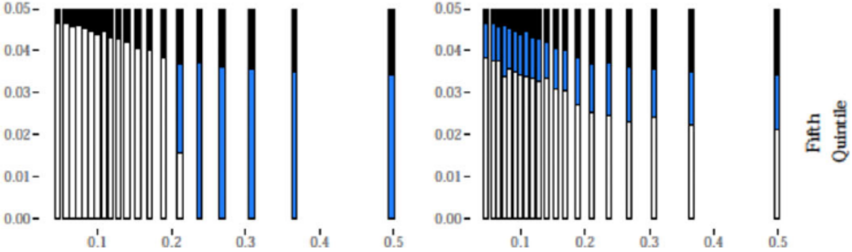


FIGURE VI

Who Do Stricter Judges Jail and Who Would the Algorithm Jail? Comparing Predicted Risk Distributions across Leniency Quintiles

- black = even most lenient judges (bottom quintile) would jail this defendant.
- blue = additional jailed by the strictest judges (top quintile). left panel = algorithm, right panel = human judges.
- white = who is released by all judges
- ***What does this graph show? ("raise hand" via zoom)***

Outline

Effects of Algorithms on Decisions

ML for Anti-Corruption Policy

- Corruption Audits as an Inspection Game
- Detecting Corruption with Machine Learning
- Empirical Applications
- Using Machine Learning to Guide Audit Policy

Behavioral responses to decisions

- Judges and criminals will change their behavior in response to adopting machine decision supports.
  - Could have unintended consequences, or create a self-reinforcing feedback loop.

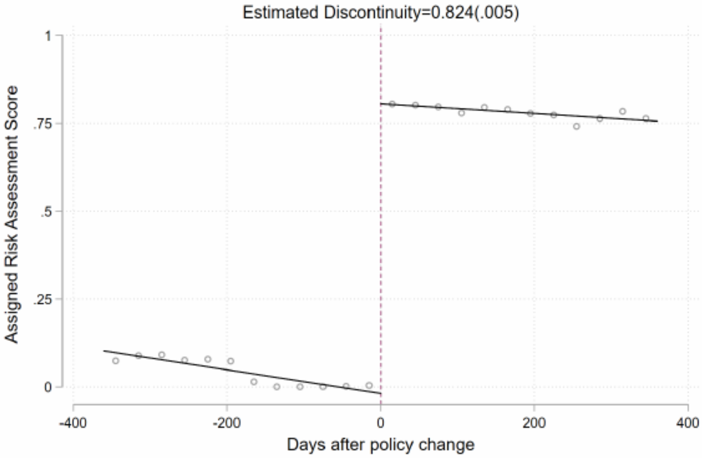
Regression Discontinuity Design

- Revisit Week 3 and Week 7 slides.

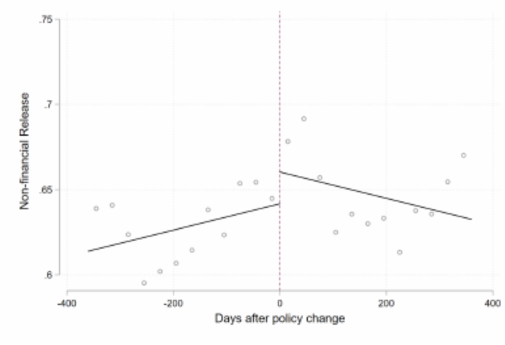
Zoom Poll: Comparing Research Designs

Sloan et al: Fuzzy RD before/after discrete introduction of risk scoring

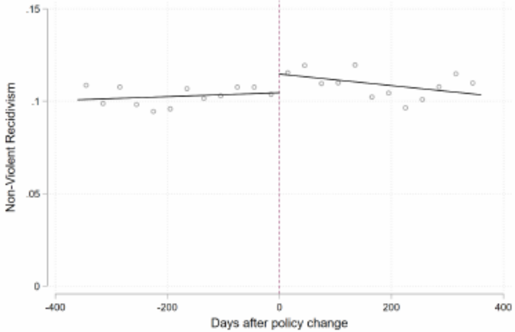
Figure 4: Regression Discontinuity Results for the Probability of Receiving a Risk Assessment Score



Sloan et al: Risk scoring increases release rates and recidivism



(a) Non-financial Bond



(a) Probability of Non-Violent Recidivism

- In response to risk scoring, judges release more poor defendants.

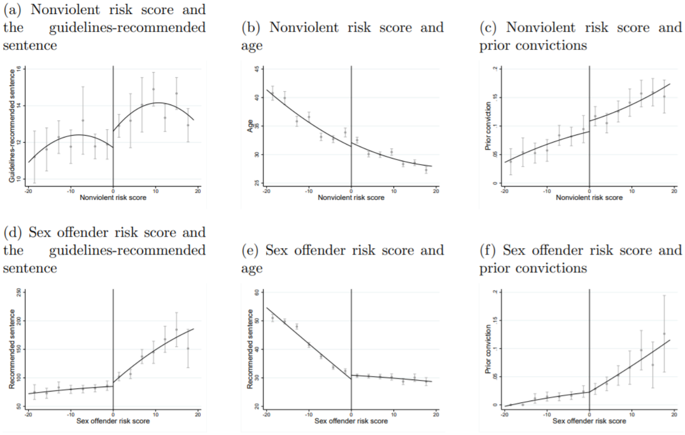
Stevenson and Doleac: Method

- RD using a continuous risk score – above a discrete cutoff, defendant is labeled "risky".

Stevenson and Doleac: Method

- RD using a continuous risk score – above a discrete cutoff, defendant is labeled "risky".
- Identification check: Other predetermined characteristics are flat around the cutoff (covariate balance):

Figure 2: Covariate balance across risk score cutoffs



(a) Nonviolent risk score and the guidelines-recommended sentence

(b) Nonviolent risk score and age

(c) Nonviolent risk score and prior convictions

(d) Sex offender risk score and the guidelines-recommended sentence

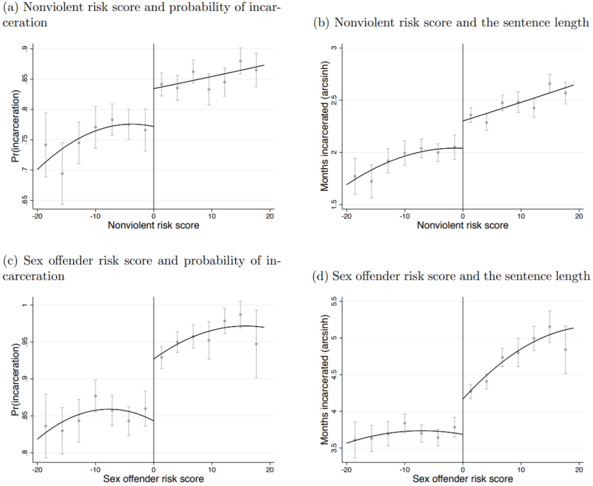
(e) Sex offender risk score and age

(f) Sex offender risk score and prior convictions

Stevenson and Doleac: Result

- Judges respond to a "is-dangerous" risk score with longer sentences:

Figure 3: Does the risk classification affect defendants' sentences at the margin?



(a) Nonviolent risk score and probability of incarceration

(b) Nonviolent risk score and the sentence length

(c) Sex offender risk score and probability of incarceration

(d) Sex offender risk score and the sentence length

- but when risk-scoring was introduced, there was no overall change in sentencing.

Activity: Break-out rooms, Detention Algorithm

`https://theintercept.com/2020/03/02/ice-algorithm-bias-detention-aclu-lawsuit/`

- In your breakout group:
  - summarize the article
  - discuss what is wrong with the system.
  - write a padlet post identifying at least two problems and how they could be solved.

`https://padlet.com/eash44/1wrxvs2srprvs0nd`

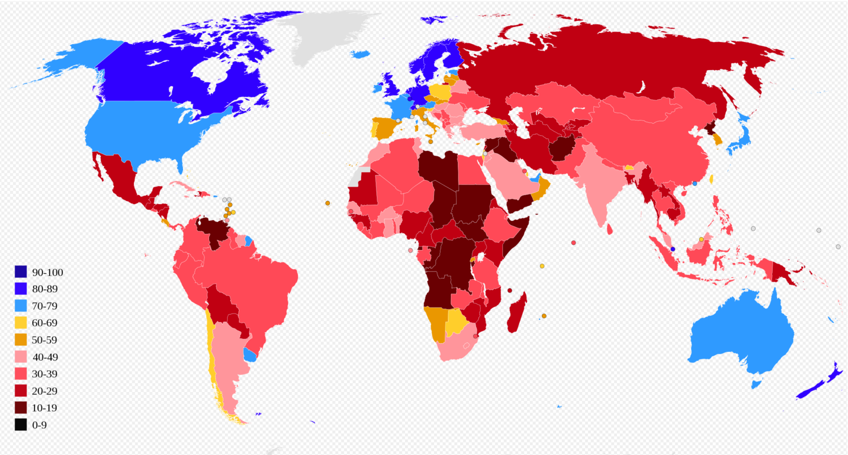
Outline

Effects of Algorithms on Decisions

ML for Anti-Corruption Policy

- Corruption Audits as an Inspection Game
- Detecting Corruption with Machine Learning
- Empirical Applications
- Using Machine Learning to Guide Audit Policy

Motivation (Ash, Galletta, Giommoni 2020)



Corruption Perceptions Index, 2018

Global costs of corruption were $2.6 trillion in 2018, according to U.N. data.
Firms and individuals spend more than $1 trillion in bribes every year.

This Paper's Goals

- **Objective 1:** Predict fiscal corruption based on public finance accounts.
  - In Brazilian municipalities, we have information on fiscal corruption from random audits.
  - We train a machine learning algorithm to detect corruption in held-out data using budget data.

This Paper's Goals

- **Objective 1:** Predict fiscal corruption based on public finance accounts.
  - In Brazilian municipalities, we have information on fiscal corruption from random audits.
  - We train a machine learning algorithm to detect corruption in held-out data using budget data.
- **Objective 2:** Construct new measure of corruption for all municipalities and years (not just those that have been audited) and use for empirical analysis.
  - Effect of public transfers on corruption (IV).
  - Effect of audits on corruption (DD).

This Paper's Goals

- **Objective 1:** Predict fiscal corruption based on public finance accounts.
  - In Brazilian municipalities, we have information on fiscal corruption from random audits.
  - We train a machine learning algorithm to detect corruption in held-out data using budget data.
- **Objective 2:** Construct new measure of corruption for all municipalities and years (not just those that have been audited) and use for empirical analysis.
  - Effect of public transfers on corruption (IV).
  - Effect of audits on corruption (DD).
- **Objective 3:** Use predictions to analyze counterfactual audit policies.
  - What can be accomplished by targeting audits to municipalities with high-risk budgets?

Brazilian municipalities

- In Brazil, local municipalities ($N = 5563$) play a central role in government services:
  - e.g., primary education, healthcare, housing, transportation.

Brazilian municipalities

- In Brazil, local municipalities ($N = 5563$) play a central role in government services:
  - e.g., primary education, healthcare, housing, transportation.
- In 2003, Brazilian government introduced innovative anti-corruption program:
  - **Audit of public spending** in **randomly selected municipalities** (through public lottery).
  - team of 10-15 auditors spend two weeks in municipal offices.
  - they write a report, send to authorities for criminal penalties and make it public.

Outline

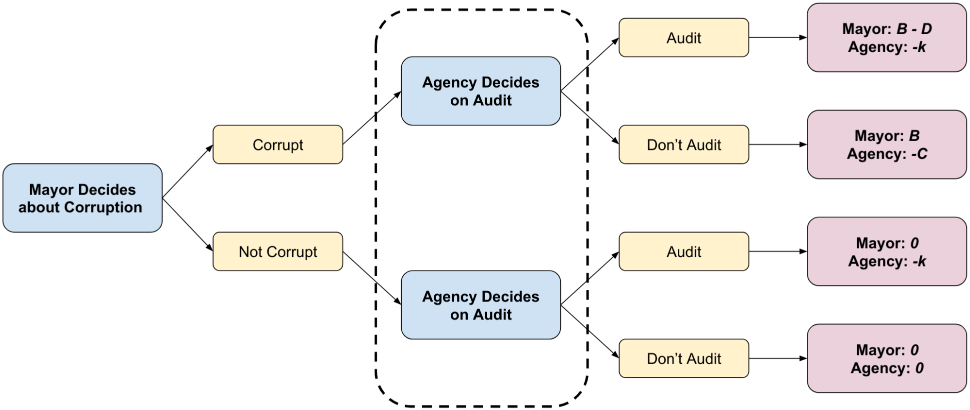
Effects of Algorithms on Decisions

ML for Anti-Corruption Policy
- Corruption Audits as an Inspection Game
- Detecting Corruption with Machine Learning
- Empirical Applications
- Using Machine Learning to Guide Audit Policy

- Stage 1: Mayor decides whether to engage in corruption.
  - if corrupt, mayor gets payoff B, society loses C (zero otherwise).

- Stage 1: Mayor decides whether to engage in corruption.
    - if corrupt, mayor gets payoff B, society loses C (zero otherwise).
- Stage 2: Agency decides whether to audit municipality i.
    - if audit, agency pays cost k, zero otherwise
    - if audit reveals corruption:
        - society does not lose C; mayor pays penalty $D > B$

- Stage 1: Mayor decides whether to engage in corruption.
  - if corrupt, mayor gets payoff B, society loses C (zero otherwise).
- Stage 2: Agency decides whether to audit municipality i.
  - if audit, agency pays cost k, zero otherwise
  - if audit reveals corruption:
    - society does not lose C; mayor pays penalty $D > B$



- In game theory, this is called an "inspection game".

Matrix Form (chalk board)

Matrix Form (chalk board)

- There is no pure-strategy Nash equilibrium (cycling).

Matrix Form (chalk board)

- There is no pure-strategy Nash equilibrium (cycling).
- Assume mixed strategies:
  - $p=$ probability of corruption, $q=$ probability of audit.
  - **Mixed strategy equilibrium**: (p^*, q^*) such that each player is indifferent between options.

Matrix Form (chalk board)

- There is no pure-strategy Nash equilibrium (cycling).
- Assume mixed strategies:
  - $p=$ probability of corruption, $q=$ probability of audit.
  - **Mixed strategy equilibrium**: (p^*, q^*) such that each player is indifferent between options.
- Payoffs for mayor:
  - no corruption: 0
  - corruption: $\underbrace{q(B-D)}_{\text{audit}} + \underbrace{(1-q)B}_{\text{no audit}} = B - qD$

Matrix Form (chalk board)

- There is no pure-strategy Nash equilibrium (cycling).
- Assume mixed strategies:
  - $p=$ probability of corruption, $q=$ probability of audit.
  - **Mixed strategy equilibrium**: (p^*, q^*) such that each player is indifferent between options.
- Payoffs for mayor:
  - no corruption: 0
  - corruption: $\underbrace{q(B-D)}_{\text{audit}} + \underbrace{(1-q)B}_{\text{no audit}} = B - qD$

$\rightarrow$ **equilibrium audit probability** $q^* = \frac{D}{B}$

Matrix Form (chalk board)

- There is no pure-strategy Nash equilibrium (cycling).
- Assume mixed strategies:
  - $p=$ probability of corruption, $q=$ probability of audit.
  - **Mixed strategy equilibrium**: (p^*, q^*) such that each player is indifferent between options.
- Payoffs for mayor:
  - no corruption: 0
  - corruption: $\underbrace{q(B-D)}_{\text{audit}} + \underbrace{(1-q)B}_{\text{no audit}} = B - qD$

  $\rightarrow$ **equilibrium audit probability** $q^* = \frac{D}{B}$
- Similarly, payoffs for agency:
  - audit: $p(-k) + (1-p)(-k) = -k$
  - no audit: $p(-C)$

Matrix Form (chalk board)

- There is no pure-strategy Nash equilibrium (cycling).
- Assume mixed strategies:
  - $p=$ probability of corruption, $q=$ probability of audit.
  - **Mixed strategy equilibrium**: (p^*, q^*) such that each player is indifferent between options.
- Payoffs for mayor:
  - no corruption: 0
  - corruption: $\underbrace{q(B-D)}_{\text{audit}} + \underbrace{(1-q)B}_{\text{no audit}} = B - qD$

  $\rightarrow$ **equilibrium audit probability** $q^* = \frac{D}{B}$
- Similarly, payoffs for agency:
  - audit: $p(-k) + (1-p)(-k) = -k$
  - no audit: $p(-C)$

  $\rightarrow$ **equilibrium corruption probability** $p^* = \frac{k}{C}$

Equilibrium Audit Policy

- Equilibrum of game:
  - **corruption probability** $p^* = \frac{k}{C}$
  - **audit probability** $q^* = \frac{D}{B}$

$\rightarrow$ Randomly assigned audits to a fraction q^* of municipalities is the equilibrium audit policy.

Equilibrium Audit Policy

- Equilibrum of game:
  - **corruption probability** $p^* = \frac{k}{C}$
  - **audit probability** $q^* = \frac{D}{B}$

→ Randomly assigned audits to a fraction q^* of municipalities is the equilibrium audit policy.

- Note that the observed corruption rate is

$$p^* = \frac{1}{N}\sum_{i=1}^{N} p_i$$

  the average of p_i, the probability of corruption for municipality i.
- Below, we will consider how this changes if agency can guess $\hat{p}(X_i)$ based on budget factors X_i.

Outline

Effects of Algorithms on Decisions

ML for Anti-Corruption Policy
- Corruption Audits as an Inspection Game
- Detecting Corruption with Machine Learning
- Empirical Applications
- Using Machine Learning to Guide Audit Policy

Corruption Audit Data

- Municipal audit reports are available from the agency web site:

☐	DOWNLOAD	TÍTULO	LINHA DE ATUAÇÃO	PUBLICADO EM	MUNICÍPIOS	TRECHOS
☐		Relatório de Fiscalização Sorteio de Municípios - Olho D'Água das Flores/AL	Fiscalização em Entes Federativos - Municípios	18/12/2018	OLHO D'ÁGUA DAS FLORES - AL	Assim como ocorre em muitos **municípios**, Olho d'Água das
☐		Relatório de fiscalização nº 201800789 - General Maynard/SE	Fiscalização em Entes Federativos - Municípios	05/12/2018	GENERAL MAYNARD - SE	junho **de** 2007, são os **municípios** os responsáveis pelo cadastramento
☐		Relatório de Fiscalização Sorteio de Municípios - Vargem Alegre/MG	Fiscalização em Entes Federativos - Municípios	17/10/2017	VARGEM ALEGRE - MG	**de** outros **municípios**.

- Brollo et al (2013) construct corruption labels from the reports for 1481 audited municipalities, 2003-2010. Their data is online.

Local Budget Data

- The annual municipality budget is available from various web sites:
  - We collected/cleaned data for 2001 through 2012 and made them comparable across years.

Local Budget Data

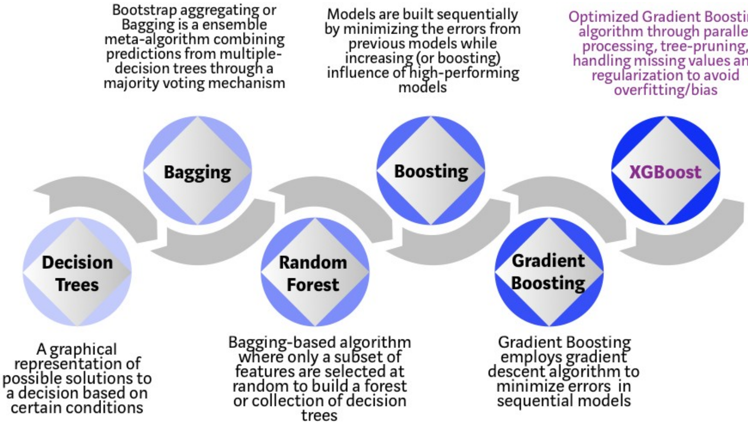
- The annual municipality budget is available from various web sites:
  - We collected/cleaned data for 2001 through 2012 and made them comparable across years.
- In total we have 797 budget variables:
  - Revenue 250, Expenditure 334, Active 100, Passive 79.

Gradient Boosted Classifier

- Gradient boosting classifier (GBC): ensemble of decision trees (Friedman, 2001; Hastie et al 2009).
  - same model used by Kleinberg et al (QJE 2018) to predict criminal recidivism.

Gradient Boosted Classifier

- Gradient boosting classifier (GBC): ensemble of decision trees (Friedman, 2001; Hastie et al 2009).
  - same model used by Kleinberg et al (QJE 2018) to predict criminal recidivism.
- We use XGBoost ("Extreme Gradient Boosting"), an optimized python implementation (Chen and Guestrin 2016).
  - Feurer et al (2018) find that XGBoost beats a sophisticated AutoML procedure with grid search over 15 classifiers and 18 data preprocessors.



Complicated in theory, easy in practice

```
from xgboost import XGBClassifier
model = XGBClassifier()

model.fit(X_train, y_train,
          early_stopping_rounds=10,
          eval_metric="logloss",
          eval_set=[(X_eval, y_eval)]
          )

y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
```

Model Training

1. Shuffle dataset into 80% training set and 20% test set
    - budget predictors standardized to mean zero and variance one in training set
2. Tuned hyperparameters in the training set using five-fold cross-validation (e.g., max depth of trees and learning rate)
    - Use early stopping to avoid over-fitting.
3. Take tuned model and get performance metrics in the test set

Model Performance in Test Set

	Guess "Not Corrupt"	*OLS*	XGBoost
Accuracy	0.58	0.594	0.750
AUC-ROC	0.5		
F1	0.0		

- Test-set accuracy of $\sim$ 75% is much better than guessing (58%) or predictions from OLS (59%)

Model Performance in Test Set

	Guess "Not Corrupt"	*OLS*	XGBoost
Accuracy	0.58	0.594	0.750
AUC-ROC	0.5	0.562	**0.814**
F1	0.0	0.413	0.665

- AUC-ROC ("Area under the receiver operating curve") is a standard metric, ranging from 0.5 (guessing) ato 1.0 (perfect accuracy).
  - Interpretation: probability that a randomly sampled corrupt municipality is ranked more highly by predicted probability of corruption than a randomly sampled non-corrupt municipality.
  - **AUC≈.81 is better than Kleinberg et al (QJE 2018) who report AUC=0.707.**

Confidence Intervals on ML Metrics

- nested cross-validation with 5 folds $\rightarrow$ produce 5 sets of perfomance metrics.

Confidence Intervals on ML Metrics

- nested cross-validation with 5 folds $\rightarrow$ produce 5 sets of perfomance metrics.

Metric	**Accuracy**	**AUC**
Mean	0.74	0.81
Median	0.74	0.82
S.D. / S.E.	0.01	0.02
95% CI's	[.73 .75]	[.79 .83]

Confidence intervals constructed as mean +/- 2$\times$S.E..

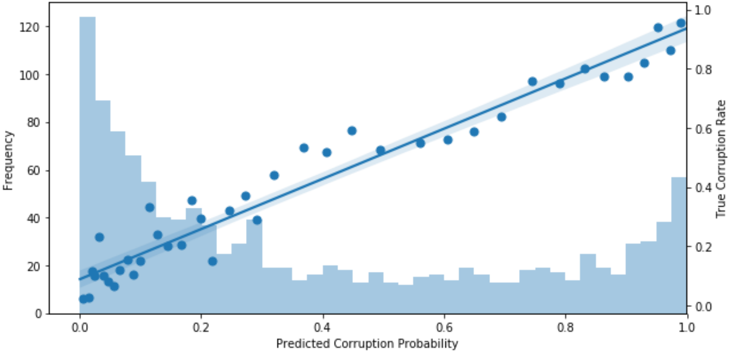
Confusion Matrix for Test-Set Predictions

	Prediction	
Truth	Not Corrupt	Corrupt
Not Corrupt	**614**	100
Corrupt	185	**313**

Confusion Matrix for Test-Set Predictions

Truth	*Prediction* Not Corrupt	Corrupt
Not Corrupt	**614**	100
Corrupt	185	**313**

True Corrupt Rate vs Predicted Prob. Corruption



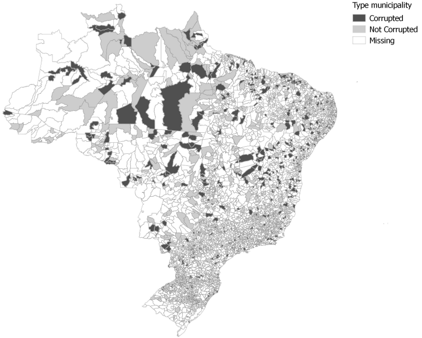
Outline

Effects of Algorithms on Decisions

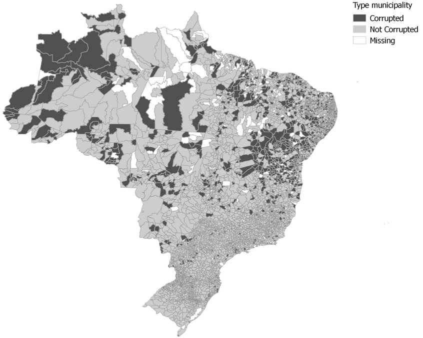
ML for Anti-Corruption Policy

- Corruption Audits as an Inspection Game
- Detecting Corruption with Machine Learning
- Empirical Applications
- Using Machine Learning to Guide Audit Policy

Applying to Full Dataset



(a) Actual Corruption



(b) Predicted Corruption

We regressed predicted corruption in pre-audit years on having an audit, and there was no difference in any specification (consistent with randomization of audits).

Analysis 1: Revenue Shocks and Corruption

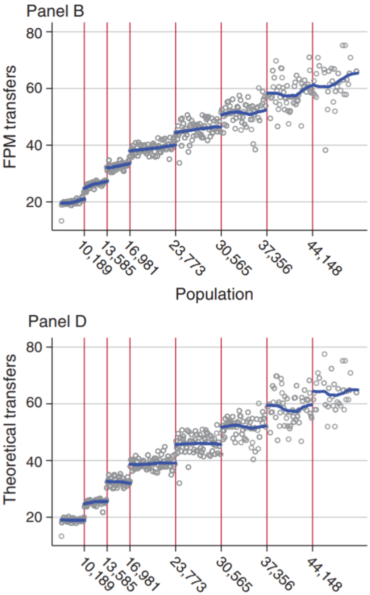
- Brollo et al (2013) find that a **windfall of public revenues** (federal transfers) leads to an increase in rent-seeking by the public administration (*i.e.* subsequent increase in corruption).

Analysis 1: Revenue Shocks and Corruption

- Brollo et al (2013) find that a **windfall of public revenues** (federal transfers) leads to an increase in rent-seeking by the public administration (*i.e.* subsequent increase in corruption).
- Empirical Strategy: Fuzzy RDD
  - Exogenous variation in transfers due to discrete population thresholds.
  - imperfect takeup, so instrument actual transfers τ_i with prescribed transfers z_i

Analysis 1: Revenue Shocks and Corruption

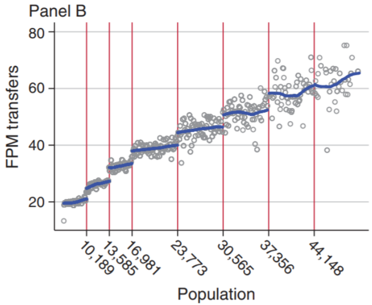
- Brollo et al (2013) find that a **windfall of public revenues** (federal transfers) leads to an increase in rent-seeking by the public administration (*i.e.* subsequent increase in corruption).
- Empirical Strategy: Fuzzy RDD
  - Exogenous variation in transfers due to discrete population thresholds.
  - imperfect takeup, so instrument actual transfers τ_i with prescribed transfers z_i

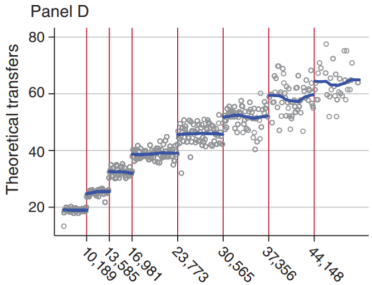


Analysis 1: Revenue Shocks and Corruption

- Brollo et al (2013) find that a **windfall of public revenues** (federal transfers) leads to an increase in rent-seeking by the public administration (*i.e.* subsequent increase in corruption).

- Empirical Strategy: Fuzzy RDD
  - Exogenous variation in transfers due to discrete population thresholds.
  - imperfect takeup, so instrument actual transfers τ_i with prescribed transfers z_i





- **<u>Our extension</u>**: Analyze **universe** of Brazilian municipalities (not only those being audited). N increases from 1115 to 5563.

Fuzzy RD (IV) Estimating Equations

- First stage: impact of prescribed transfers (z_i) on actual transfers (τ_i)

$$\tau_i = g(P_i) + \gamma z_i + u_i \tag{1}$$

- Second stage: impact of instrumented actual transfers (τ_i) on ML-predicted corruption (y_i)

$$y_i = g(P_i) + \beta \tau_i + \epsilon_i \tag{2}$$

– polynomial $g(\cdot)$ in population P_i

Activity: Exogeneity/Exclusion

https://padlet.com/eash44/cfsa9e4m4lycv33f

$$\tau_i = g(P_i) + \gamma z_i + u_i$$
$$y_i = g(P_i) + \beta \tau_i + \epsilon_i$$

- Last Name starts with A-M:
  - Articulate exogeneity assumption, and a potential violation.
- Last Name starts with N-Z:
  - Articulate exclusion restriction, and a potential violation.

Brollo et al (2013) Replication: First Stage

	Audited cities (1)	All cities (2)	Never Audited (3)
Panel A. First Stage			
Prescribed transfers	0.680***	0.687***	0.700***
	(0.021)	(0.022)	(0.023)
Observations	1115	5563	4693

Standard errors clustered at the municipal level are in parentheses: *p$<$0.10, ** p$<$0.05, *** p$<$0.01. Prescribed transfers (z_i), actual transfers (τ_i), predicted corruption (y_i). First stage: $\tau_i = g(P_i) + \alpha_\tau z_i + \delta_t + \gamma_s + u_i$; Second stage: $y_i = g(P_i) + \beta_y \tau_i + \delta_t + \gamma_s + \epsilon_i$; polynomial $g(\cdot)$ in population P_i, time fixed effects δ_t, state fixed effects γ_s (as in Brollo et al. 2013).

Brollo et al (2013) Replication: Audited Cities

	Audited cities (1)	All cities (2)	Never Audited (3)
Panel A. First Stage			
Prescribed transfers	0.680***	0.687***	0.700***
	(0.021)	(0.022)	(0.023)
Panel B. Reduced Form			
Prescribed transfers	**0.00526****		
	(0.00264)		
Panel C. 2SLS			
Actual transfers	**0.00862****		
	(0.004)		
Observations	1115	5563	4693

Standard errors clustered at the municipal level are in parentheses: *p<0.10, ** p<0.05, *** p<0.01. Prescribed transfers (z_i), actual transfers (τ_i), predicted corruption (y_i). First stage: $\tau_i = g(P_i) + \alpha_\tau z_i + \delta_t + \gamma_s + u_i$; Second stage: $y_i = g(P_i) + \beta_y \tau_i + \delta_t + \gamma_s + \epsilon_i$; polynomial $g(\cdot)$ in population P_i, time fixed effects δ_t, state fixed effects γ_s (as in Brollo et al. 2013).

Brollo et al (2013) Replication: Never-Audited Cities

	Audited cities	All cities	**Never Audited**
	(1)	(2)	(3)
Panel A. First Stage			
Prescribed transfers	0.680***	0.687***	0.700***
	(0.021)	(0.022)	(0.023)
Panel B. Reduced Form			
Prescribed transfers	0.00526**	0.00370***	**0.00294*** **
	(0.00264)	(0.001)	**(0.001)**
Panel C. 2SLS			
Actual transfers	0.00862**	0.00731***	**0.00660*** **
	(0.004)	(0.001)	**(0.001)**
Observations	1115	5563	4693

Standard errors clustered at the municipal level are in parentheses: *p<0.10, ** p<0.05, *** p<0.01. Prescribed transfers (z_i), actual transfers (τ_i), predicted corruption (y_i). First stage: $\tau_i = g(P_i) + \alpha_\tau z_i + \delta_t + \gamma_s + u_i$; Second stage: $y_i = g(P_i) + \beta_y \tau_i + \delta_t + \gamma_s + \epsilon_i$; polynomial $g(\cdot)$ in population P_i, time fixed effects δ_t, state fixed effects γ_s (as in Brollo et al. 2013).

Analysis 2: Effects of auditing on subsequent corruption

- ML-predicted corruption y_{it} in municipality i, year t:

$$y_{it} = D'_{it}\beta + \delta_i + \gamma_t + \epsilon_{it} \tag{3}$$

  - D_{it}, treatments variables for years after audit
  - δ_i, municipality FE
  - γ_t, year FE

Analysis 2: Effects of auditing on subsequent corruption

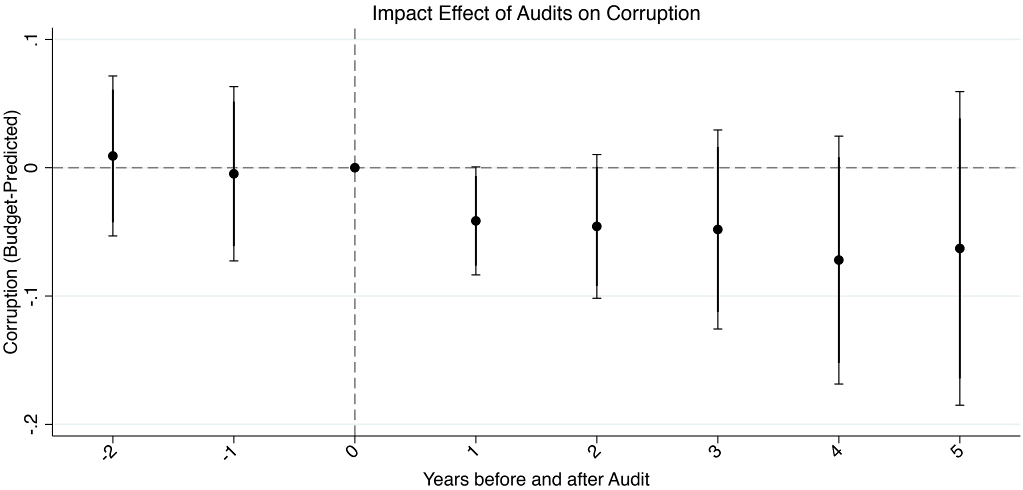
- ML-predicted corruption y_{it} in municipality i, year t:

$$y_{it} = D'_{it}\beta + \delta_i + \gamma_t + \epsilon_{it} \tag{3}$$

  - D_{it}, treatments variables for years after audit
  - δ_i, municipality FE
  - γ_t, year FE
- Empirical approach is differences-in-differences
  - What is the identification assumption for β to be consistently estimated?
  - Why is it satisfied in this case?

Event Study: Effect of Audits on Fiscal Corruption

Event Study: Effect of Audits on Fiscal Corruption

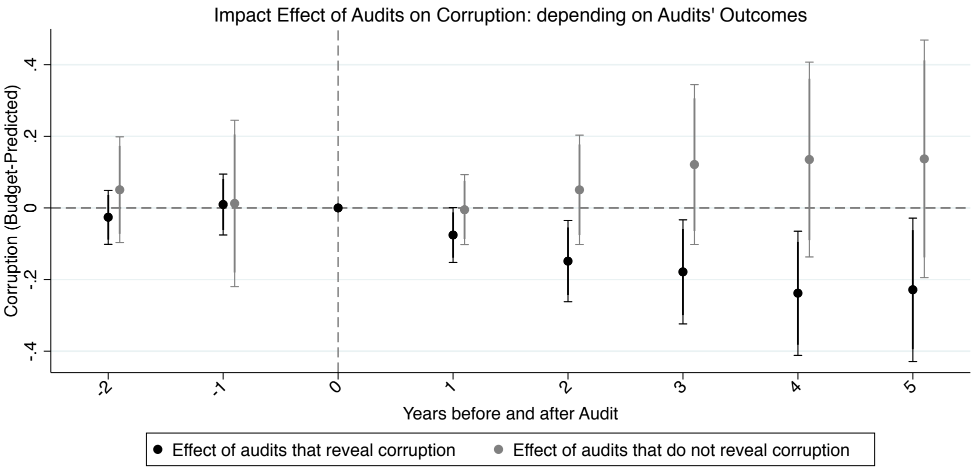


Error spikes give 95% (horizontal bars) and 90% (bold lines) confidence intervals, with standard error clustered by state.

⇒ The audit has a **disciplining effect**, inducing a reduction in corruption.

Event Study: By Audit Outcome

Event Study: By Audit Outcome

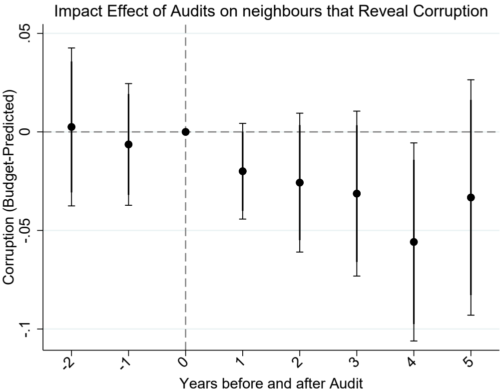


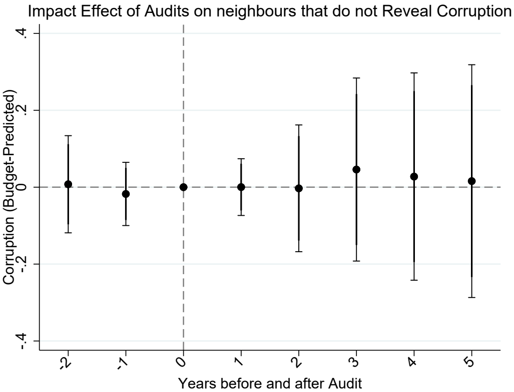
Error spikes give 95% (horizontal bars) and 90% (bold lines) confidence intervals, with standard error clustered by state.

⇒ When detected, fiscal corruption decreases by ∼24 percentage points from a mean of 47% (approx 50 percent decrease).

Spillover Effects on Neighbors: Event Study Estimates

Spillover Effects on Neighbors: Event Study Estimates





Error spikes give 95% (horizontal bars) and 90% (bold lines) confidence intervals, with standard error clustered by state.

⇒ Effect on neighbors can be interpreted as a **behavioural response**, as audit probability is unchanged.

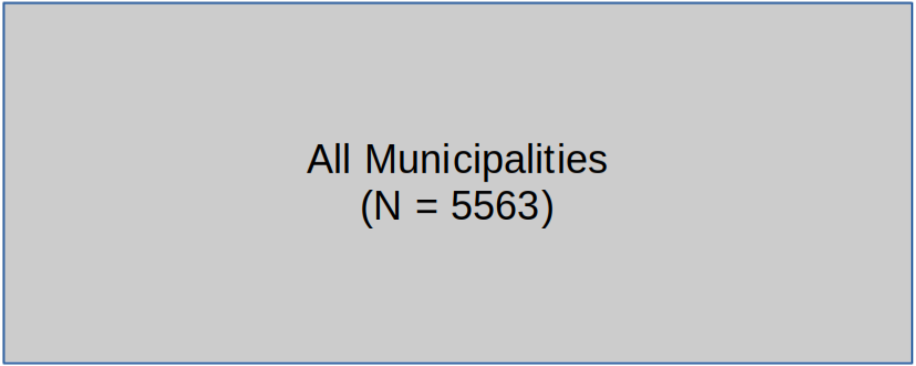
Outline

Effects of Algorithms on Decisions

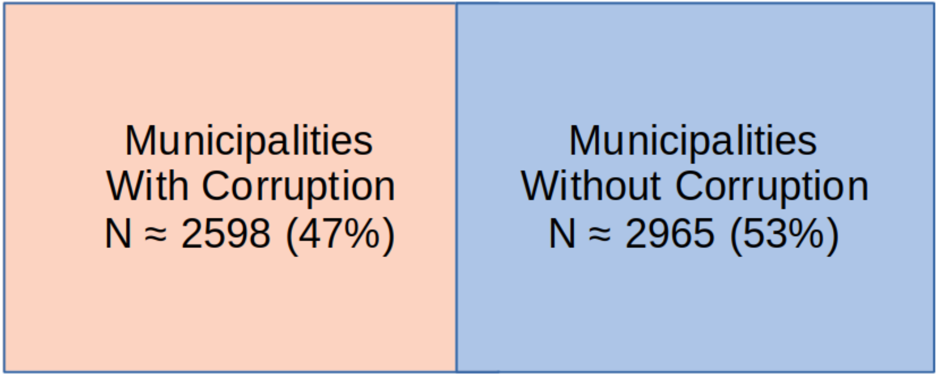
ML for Anti-Corruption Policy

- Corruption Audits as an Inspection Game
- Detecting Corruption with Machine Learning
- Empirical Applications
- Using Machine Learning to Guide Audit Policy

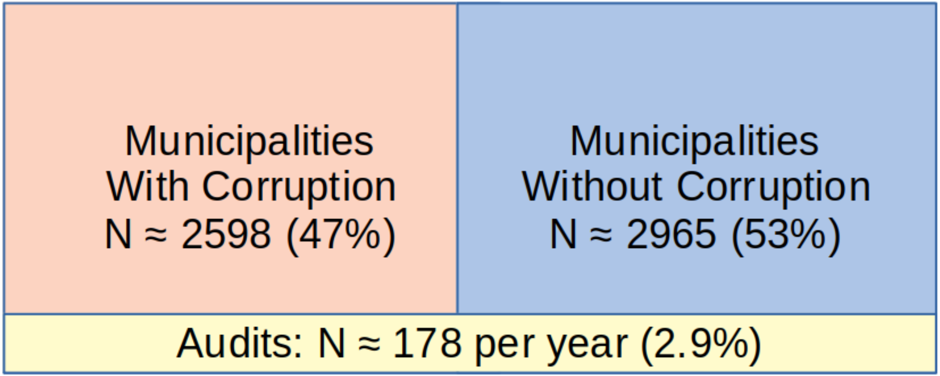
How effective are random audits?



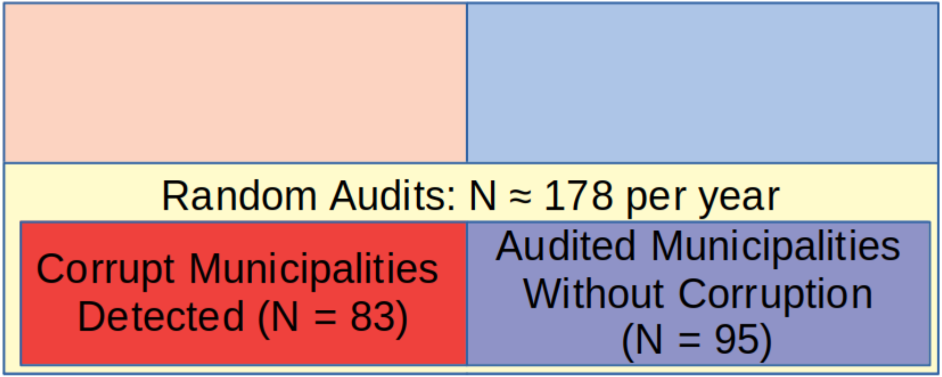
How effective are random audits?



How effective are random audits?



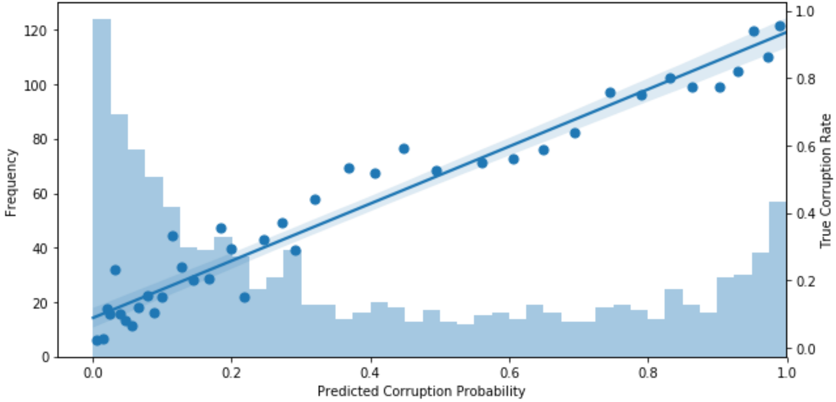
How effective are random audits?



- Under random audits, and assuming perfect detection conditional on audit, detection rate (per corrupt municipality) is equal to the audit rate (2.9%).

Targeting Audits by Corruption Risk

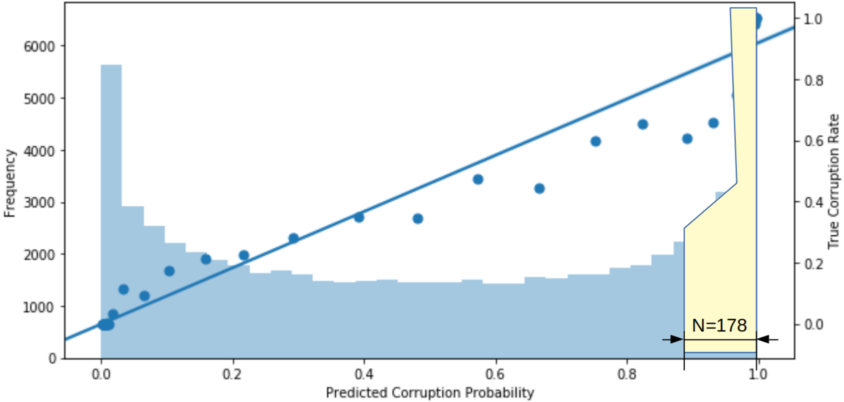
Targeting Audits by Corruption Risk



Rank municipalities by corruption risk:

- Apply model to budget data for each municipality to produce $\hat{y}_{it}$
- for each year t, get an ordinal ranking of the municipalities by predicted probability of corruption.

Targeting Audits by Corruption Risk

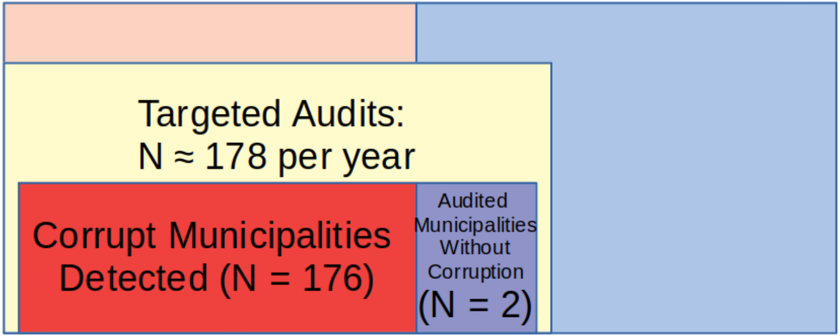


Proposed policy: Replace random audits with audits targeted by predicted corruption risk.

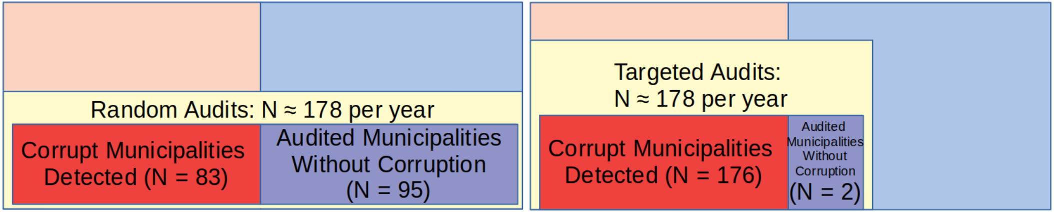
- Rather than sampling 178 municipalities uniformly from distribution, audit 178 with highest $\hat{y}_{it}$.

Targeting Audits by Corruption Risk

- ML-Targeted Auditing results in ~98% corruption detection rate.

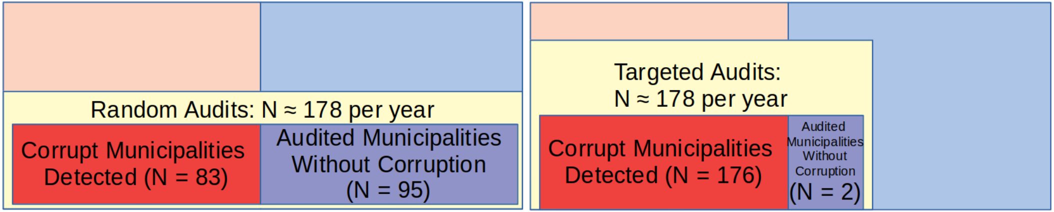


Comparing the Policies

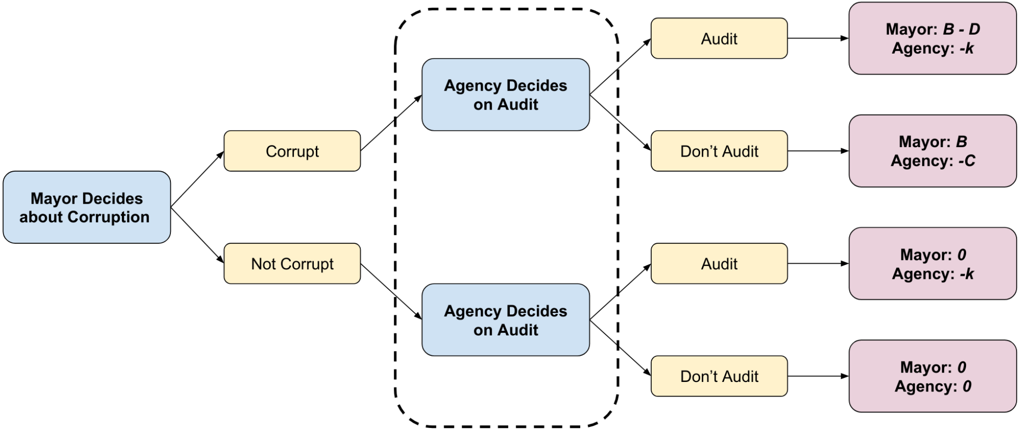


- Holding number of audits constant, targeting increases detections by 120%.
- Detection probability per corrupt municipality more than doubles – from 2.9% to 6.7%.

Comparing the Policies

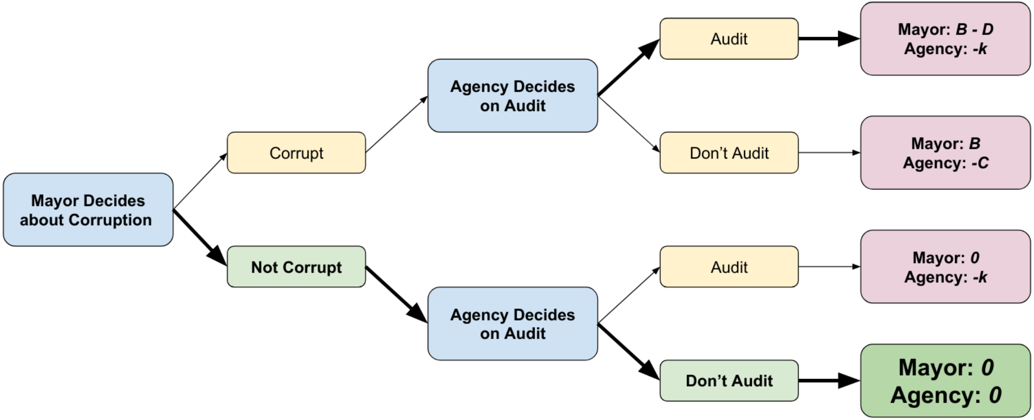


- Holding number of audits constant, targeting increases detections by 120%.
- Detection probability per corrupt municipality more than doubles – from 2.9% to 6.7%.
- To achieve same number of detections as status quo (83 municipalities), only 84 targeted audits are needed.
  - Decrease of 94 audits per year (53%), a major reduction in audit resources.
- ***Why don't we need to use the contraction method a la Kleinberg et al 2018? ("raise hand" via zoom)***



- in status quo, agency decisions are in same information set and equilibrium corruption rate is $p^* > 0$

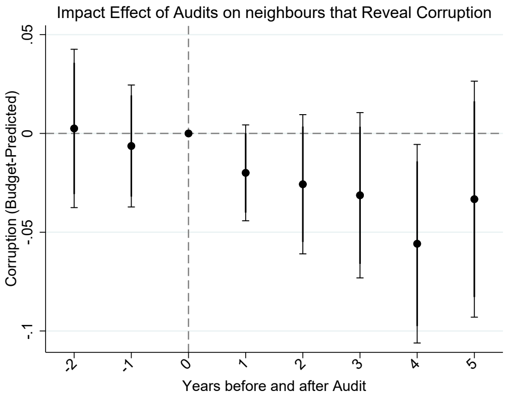
- as detection rate gets close to one, game converges to extensive form:



- by backward induction, best response is no corruption.

Behavioral AI Policy: Exploiting Spillovers

Behavioral AI Policy: Exploiting Spillovers



- According to spillover analysis, audits cut corruption by neighboring municipalities by about 10 percent (from .47 to .43) .

- Could be used to further improve policy effectiveness of targeted audits.
  - Adjust the risk ranking to target municipalities with high spillover potential.
  - For example, the policy could target the centroids of clusters of corrupt municipalities.

Breakout Groups: Open Issues / Limitations with Brazil Corruption Study

`https://bit.ly/BRJ-W10-A2`