Who is Bogus? Catching Fraudulent Firms in India

Aprajit Mahajan¹ Shekhar Mittal² Ofir Reich³

¹UC Berkeley (Dept. of ARE) & CEGA

²UC Berkeley (iSchool)

³(formerly) CEGA

August 29, 2018 EDI Research Link Up Research Presentation









Rapid Increase in Value Added Tax Adoption Since 1960



Tax levied at each stage of production or distribution (contra sales tax).

▶ 1 country in 1960 \rightarrow 50 in 1990 \rightarrow 160 in 2015.

▶ VAT requires buyer & seller to independently report each transaction.

- Opposing incentives should reduce scope for collusion and evasion.
- Whether this occurs, particularly in emerging economies, is an open question.¹
- VAT example
- ▶ Alternate evasion strategy "Bogus" firms.
 - ▶ Bogus firms are shell firms created to enable firms to lower tax bills.
 - Create (fake) paper trails of transactions with genuine firms.
 - Bogus firms example
- Precise extent and magnitudes largely unknown.
 - Media reports estimate the loss, in Delhi alone \approx \$300m.²
- Commonly reported in many VAT systems.
 - ▶ Still relevant for the newly launched Goods and Services Tax (India).
 - Early conversations in Mexico, Dominican Republic, and Zambia.

▶ VAT requires buyer & seller to independently report each transaction.

- Opposing incentives should reduce scope for collusion and evasion.
- ▶ Whether this occurs, particularly in emerging economies, is an open question.¹
- ► VAT example
- Alternate evasion strategy "Bogus" firms.
 - Bogus firms are shell firms created to enable firms to lower tax bills.
 - Create (fake) paper trails of transactions with genuine firms.
 - Bogus firms example
- Precise extent and magnitudes largely unknown.
 - Media reports estimate the loss, in Delhi alone \approx \$300m.²
- Commonly reported in many VAT systems.
 - ▶ Still relevant for the newly launched Goods and Services Tax (India).
 - Early conversations in Mexico, Dominican Republic, and Zambia.

▶ VAT requires buyer & seller to independently report each transaction.

- Opposing incentives should reduce scope for collusion and evasion.
- ▶ Whether this occurs, particularly in emerging economies, is an open question.¹
- VAT example
- Alternate evasion strategy "Bogus" firms.
 - Bogus firms are shell firms created to enable firms to lower tax bills.
 - Create (fake) paper trails of transactions with genuine firms.
 - Bogus firms example
- Precise extent and magnitudes largely unknown.
 - Media reports estimate the loss, in Delhi alone \approx \$300m.²
- Commonly reported in many VAT systems.
 - ▶ Still relevant for the newly launched Goods and Services Tax (India).
 - Early conversations in Mexico, Dominican Republic, and Zambia.

¹Limited evidence in Mittal and Mahajan (2017) ²India Today, TOI, BS.

- ▶ VAT requires buyer & seller to independently report each transaction.
 - ► Opposing incentives should reduce scope for collusion and evasion.
 - Whether this occurs, particularly in emerging economies, is an open question.¹
 - VAT example
- Alternate evasion strategy "Bogus" firms.
 - Bogus firms are shell firms created to enable firms to lower tax bills.
 - Create (fake) paper trails of transactions with genuine firms.
 - Bogus firms example
- Precise extent and magnitudes largely unknown.
 - Media reports estimate the loss, in Delhi alone \approx \$300m.²
- Commonly reported in many VAT systems.
 - ▶ Still relevant for the newly launched Goods and Services Tax (India).
 - ▶ Early conversations in Mexico, Dominican Republic, and Zambia.

¹Limited evidence in Mittal and Mahajan (2017) ²India Today, TOI, BS.

- Policy Issue: Improving tax collections central to building state capacity – key development concern.
- Question: Can we use machine learning (ML) to improve tax collections in India?
 - ▶ Use ML to ↑ tax authority's enforcement effectiveness.
 - Simulated on VAT in Delhi.
 - Expanding on GST in Punjab and Tamil Nadu.

- Policy Issue: Improving tax collections central to building state capacity – key development concern.
- Question: Can we use machine learning (ML) to improve tax collections in India?
 - Use ML to \uparrow tax authority's enforcement effectiveness.
 - Simulated on VAT in Delhi.
 - Expanding on GST in Punjab and Tamil Nadu.

Physical inspections gold standard, but resource intensive.

- Inspection resources limited (particularly in low-income countries).
- ▶ Officials in the central office create a list of "risky" firms.
 - Based on (limited set of) variables: low (VAT deposited/turnover), high turnover, high revisions, invalid address.
- ▶ Local inspectors sent out for inspections.
 - ▶ Firms deregistered ("cancelled") if inspection fails.
- ▶ Key problem: How to identify firms for inspection?
 - More of them with less effort.

Our Work

- ▶ Physical inspections gold standard, but resource intensive.
 - Inspection resources limited (particularly in low-income countries).
- Officials in the central office create a list of "risky" firms.
 - Based on (limited set of) variables: low (VAT deposited/turnover), high turnover, high revisions, invalid address.
- ▶ Local inspectors sent out for inspections.
 - ▶ Firms deregistered ("cancelled") if inspection fails.
- ▶ Key problem: How to identify firms for inspection?
 - More of them with less effort.

Our Work

- ▶ Physical inspections gold standard, but resource intensive.
 - Inspection resources limited (particularly in low-income countries).
- Officials in the central office create a list of "risky" firms.
 - Based on (limited set of) variables: low (VAT deposited/turnover), high turnover, high revisions, invalid address.
- Local inspectors sent out for inspections.
 - ► Firms deregistered ("cancelled") if inspection fails.
- ▶ Key problem: How to identify firms for inspection?
 - More of them with less effort.

Our Work

- ▶ Physical inspections gold standard, but resource intensive.
 - Inspection resources limited (particularly in low-income countries).
- Officials in the central office create a list of "risky" firms.
 - Based on (limited set of) variables: low (VAT deposited/turnover), high turnover, high revisions, invalid address.
- Local inspectors sent out for inspections.
 - ► Firms deregistered ("cancelled") if inspection fails.
- ▶ Key problem: How to identify firms for inspection?
 - More of them with less effort.

Our Work

- ▶ Physical inspections gold standard, but resource intensive.
 - Inspection resources limited (particularly in low-income countries).
- Officials in the central office create a list of "risky" firms.
 - Based on (limited set of) variables: low (VAT deposited/turnover), high turnover, high revisions, invalid address.
- Local inspectors sent out for inspections.
 - ► Firms deregistered ("cancelled") if inspection fails.
- ▶ Key problem: How to identify firms for inspection?
 - More of them with less effort.

Our Work

- ► Use VAT/GST return data & small inspection data-set to build 1st iteration.
- Improve model by carrying out inspections using the model predictions.
- Incorporate tool into the VAT/GST system to enable routine prediction.
- Quantify (revenue) gain to department from ML tool.
 - ▶ Use RCT to compare ML to "business as usual".
 - ▶ Measuring revenue changes after detection (along network) not trivial.

- ► Use VAT/GST return data & small inspection data-set to build 1st iteration.
- Improve model by carrying out inspections using the model predictions.
- Incorporate tool into the VAT/GST system to enable routine prediction.
- ▶ Quantify (revenue) gain to department from ML tool.
 - ▶ Use RCT to compare ML to "business as usual".
 - ▶ Measuring revenue changes after detection (along network) not trivial.

- ► Use VAT/GST return data & small inspection data-set to build 1st iteration.
- Improve model by carrying out inspections using the model predictions.
- Incorporate tool into the VAT/GST system to enable routine prediction.
- ▶ Quantify (revenue) gain to department from ML tool.
 - ▶ Use RCT to compare ML to "business as usual".
 - ▶ Measuring revenue changes after detection (along network) not trivial.

- ► Use VAT/GST return data & small inspection data-set to build 1st iteration.
- Improve model by carrying out inspections using the model predictions.
- Incorporate tool into the VAT/GST system to enable routine prediction.
- Quantify (revenue) gain to department from ML tool.
 - Use RCT to compare ML to "business as usual".
 - Measuring revenue changes after detection (along network) not trivial.

Random Forest Model Performance on Top 1000 Recommendations



Results similar when we control for revenue size.

Potential to gain more than 20,000\$ per inspection.

Different Classifiers

Different Feature Sets All Recommendations

Important Features

- Challenge 1: Explain the ML black box.
 - Important for inspectors to build a case around why firms are being inspected.
- Challenge 2: Actually increase tax revenue.
 - Revenue recovery from trading partners not trivial.
- Challenge 3: Firms will respond to better targeting e.g. by creating more bogus firms faster.
 - ▶ ML tool will require regular updating (more training data).
- ▶ Interest from many tax authorities, potentially useful tool in the hands of high level officials.
 - ▶ Next steps: Build and evaluate the tool in Punjab and Tamil Nadu.

- Challenge 1: Explain the ML black box.
 - Important for inspectors to build a case around why firms are being inspected.
- Challenge 2: Actually increase tax revenue.
 - Revenue recovery from trading partners not trivial.
- Challenge 3: Firms will respond to better targeting e.g. by creating more bogus firms faster.
 - ▶ ML tool will require regular updating (more training data).
- ▶ Interest from many tax authorities, potentially useful tool in the hands of high level officials.
 - ▶ Next steps: Build and evaluate the tool in Punjab and Tamil Nadu.

- Challenge 1: Explain the ML black box.
 - Important for inspectors to build a case around why firms are being inspected.
- Challenge 2: Actually increase tax revenue.
 - Revenue recovery from trading partners not trivial.
- Challenge 3: Firms will respond to better targeting e.g. by creating more bogus firms faster.
 - ML tool will require regular updating (more training data).
- ► Interest from many tax authorities, potentially useful tool in the hands of high level officials.
 - ▶ Next steps: Build and evaluate the tool in Punjab and Tamil Nadu.

- Challenge 1: Explain the ML black box.
 - Important for inspectors to build a case around why firms are being inspected.
- Challenge 2: Actually increase tax revenue.
 - Revenue recovery from trading partners not trivial.
- Challenge 3: Firms will respond to better targeting e.g. by creating more bogus firms faster.
 - ML tool will require regular updating (more training data).
- Interest from many tax authorities, potentially useful tool in the hands of high level officials.
 - ▶ Next steps: Build and evaluate the tool in Punjab and Tamil Nadu.

Thanks!

- ▶ We thank GoNCTD, IGC, CEGA, EDI, and JPAL for support.
- ▶ This project was funded with UK Aid from the UK Government.

Aprajit Mahajan aprajit@berkeley.edu UC Berkeley & CEGA

Shekhar Mittal shekhar@ischool.berkeley.edu UC Berkeley & CEGA

Ofir Reich CEGA









Appendix!

VAT: Example





Bogus Firms: Example



Government receives tax on \$50 value added. Surplus is divided between offenders.

Firms A, C and D not necessarily in the same chain.

Bogus firm can make sales to any firm which needs input credits.

Comparison of Different Classifiers



Betas curves for different feature sets



Model performance on all recommendations

Inspection Group	Firms Inspected	Total Bogus Firms Caught	Bogus Firms Caught/Inspection
1 - 400	400	305	0.76
401 - 800	400	48	0.12
801 - 1200	400	24	0.06
1201 - 2500	1300	29	0.02
2501 - rest	313229	132	0.00

Interpreting Features: Gaming Measures



Bogus firms likely to have ratio in middle indicates that they know tax authority monitors extreme values so they make sure they are not in extremes.

Network feature: Pagerank (2A)



Network feature: VAT deposited ratio by 2B firms



Network feature: Unregistered sales made by 2A firms



Firms grouped in deciles(10%) of percentage sales made to unregistered firms, By the firms current firm is purchasing from

Network feature: VAT deposited ratio by 2A firms



Firms grouped in deciles(10%) of ratio of VAT deposited to turnover, by the firms current firm is purchasing from

Point-in-time Simulations Performance



Revenue Saved

Point-in-time simulation performance for the 1-400 inspection group

Т	Total Bogus Firms Caught	Bogus Firms Caught/Inspection	Revenue Gained by Inspecting Entire Group (USD Millions)	Revenue Gained per Inspection (USD 000s)	Total Bogus Firms in the Sample	Revenue Lost from All Bogus Firms (USD Millions)
2	94	0.24	19.44	48.60	416	49.40
4	155	0.39	43.19	107.97	412	108.38
6	156	0.39	25.48	63.70	437	63.84
8	157	0.39	9.38	23.46	395	26.43
10	46	0.11	1.70	4.24	114	4.52
12	10	0.02	0	0	22	0

How genuine firms look



Total output tax reliably larger than input tax credit.

Size of problem: From explicit data



- Input credit claimed weakly greater than output tax declared
- From the limited sample, revenue loss between ₹4-6 billion, annually
 - Drop in later quarters due to missing data

From cancellation records

Size of problem: From cancellation records



From the much bigger sample, revenue loss around ₹15 billion, annually

Drop in later quarters due to missing data

Machine Learning Performance



Back to Machine Learning: Ingredients Back to Results Preview

Revenues Non-Bogus Firms



GST: Other Ideas

- Two important dimensions of tax reform
- **Equity**: who bears the incidence of consumption tax change? Several possibilities:
 - 1. Consumers through changing prices
 - 2. Firm owners through changing profits
 - 3. Workers through changing wages
- GST reform will change tax rate at the level of the state and the product. This allows us to do the following comparisons (taking prices as an example) to study the incidence of consumption taxes:
 - 1. Change in price of *a same product* across two states with different tax rate change due to different state level VAT pre-reform.
 - 2. Change in price between two products *in the same state*, where products receive different rates due to the new several tier rating system of the GST.

Network feature: Pagerank (2A)



Network feature: VAT deposited ratio by 2B firms



Network feature: Unregistered sales made by 2A firms



Firms grouped in deciles(10%) of percentage sales made to unregistered firms, By the firms current firm is purchasing from

Network feature: VAT deposited ratio by 2A firms



Firms grouped in deciles(10%) of ratio of VAT deposited to turnover, by the firms current firm is purchasing from