Catching Value Added Tax evaders in Delhi using Machine Learning

Aprajit Mahajan (UC Berkeley)
Shekhar Mittal (UCLA)
Ofir Reich (Data Scientist, CEGA)

The problem - Value Added Tax evasion via bogus firms

National Capital Territory of **Delhi**, India

Bogus firms exist only on paper, and make money by falsely reporting transactions with genuine firms (more on this soon)

Media reports estimate annual revenue loss around $300 Million (₹2000 crore)

Hard to locate offenders, limited labor to inspect. When found their license revoked.

We show a way of better targeting inspections and finding bogus firms using tax data, show very high accuracy and estimate $30 Million in potential additional revenue.

In discussions to replicate this in Tamil Nadu, Mexico, Dominican Republic.

The project in a nutshell

Value Added Tax (VAT) returns of all registered private firms

Who sold to whom, for how much, what tax rate? Quarterly. *Anonymized.*

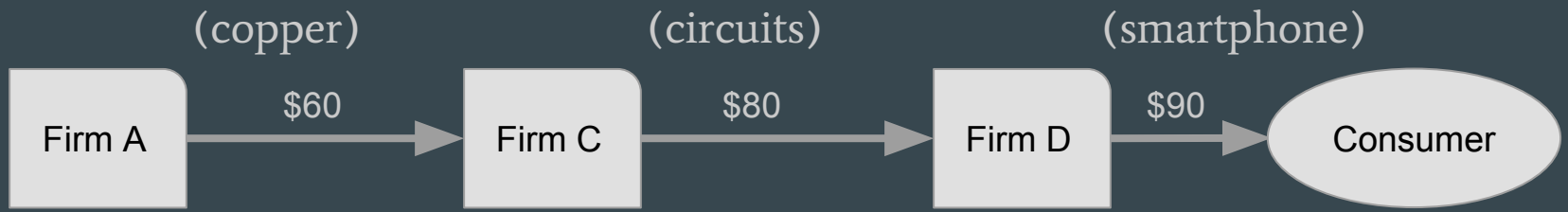
Automatically process the data and identify firms suspected of being bogus

Target them for physical inspections

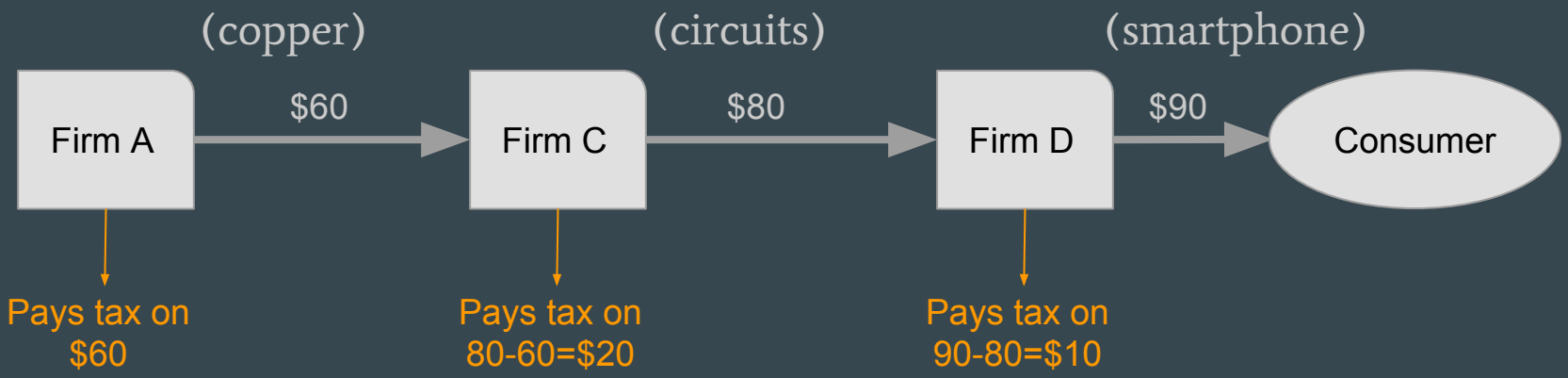
Machine Learning Approach - we use firms that were found to be bogus in the past to identify suspicious behavior in the data and target firms that display similar behavior in the present data.

Past bogus firms -> what is suspicious behavior -> similar behavior in present -> target

How VAT works

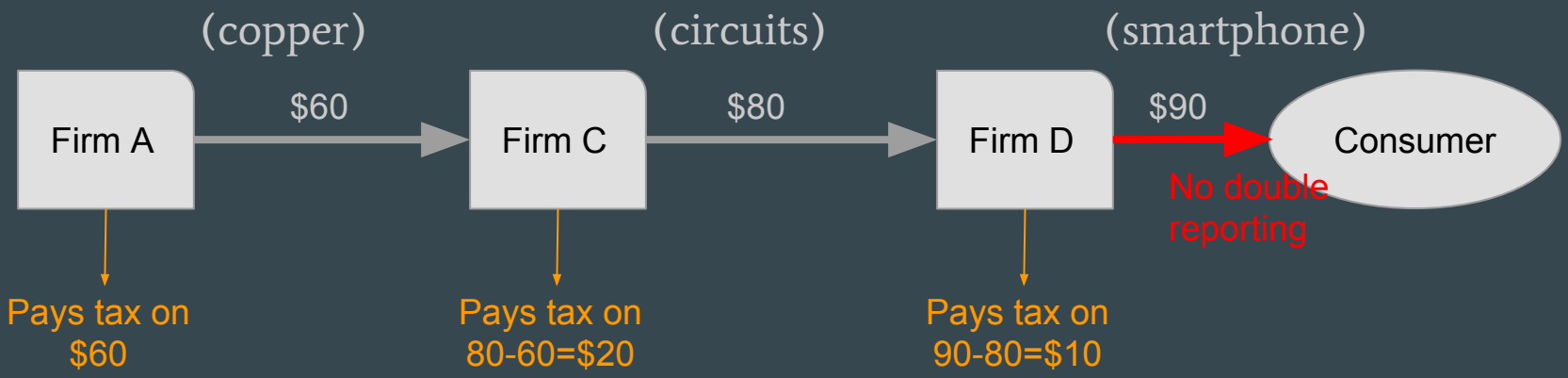


How VAT works



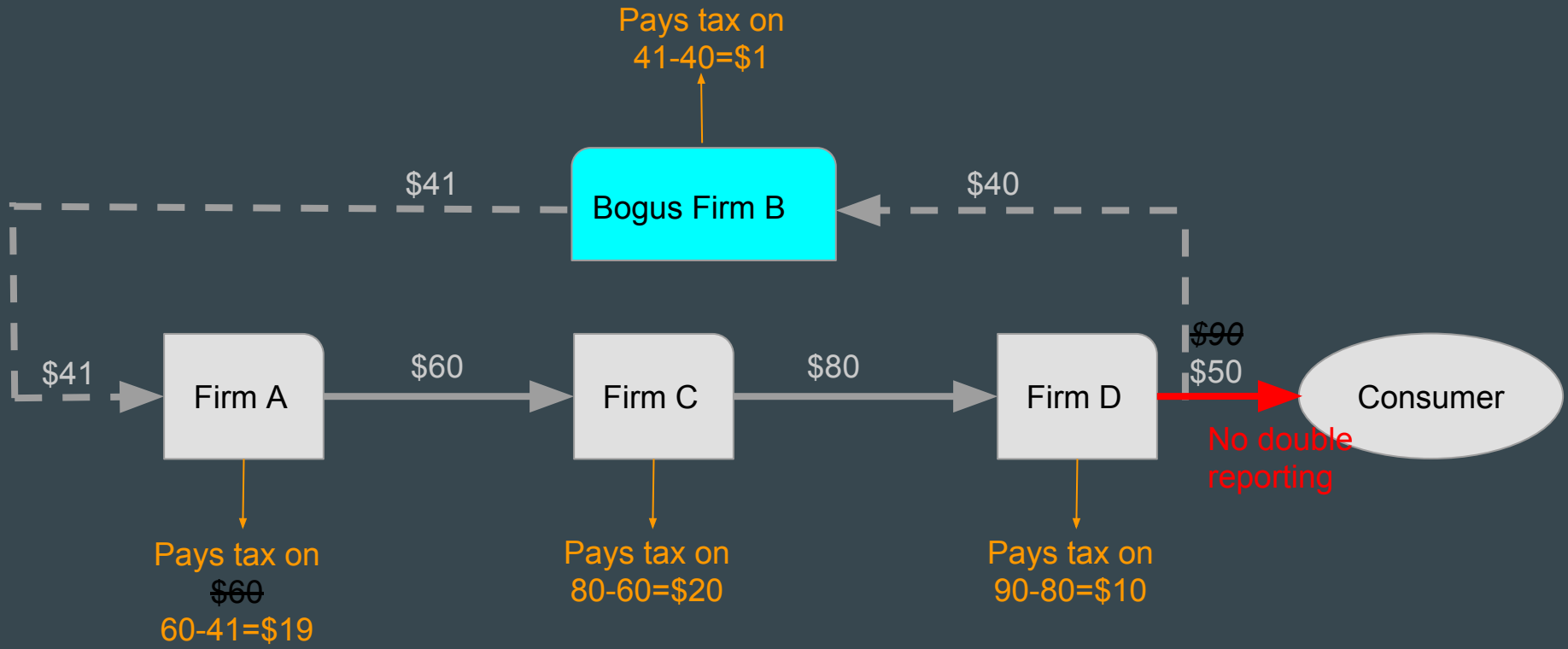
Government receives tax on $90 value added

How VAT works



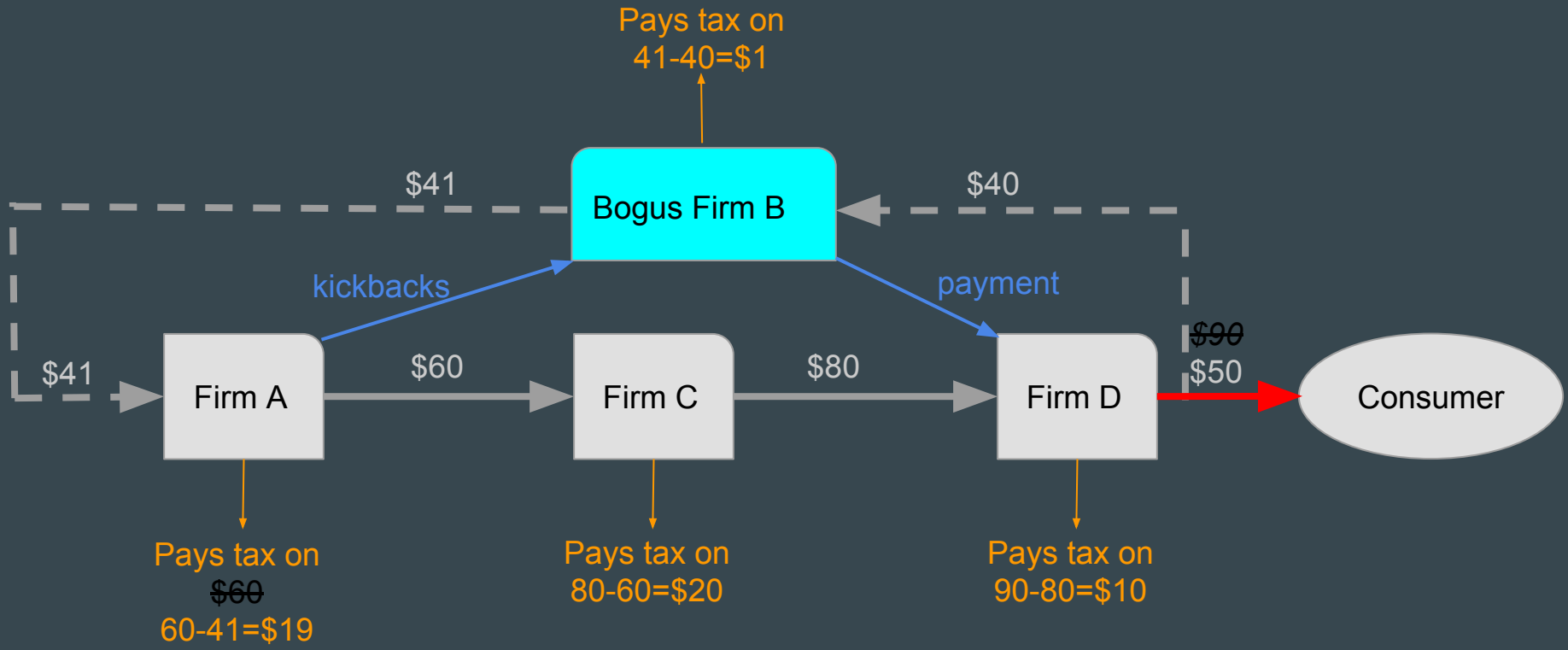
Government receives tax on $90 value added

How VAT evasion works



Government receives tax on $40 less value added

How VAT evasion works

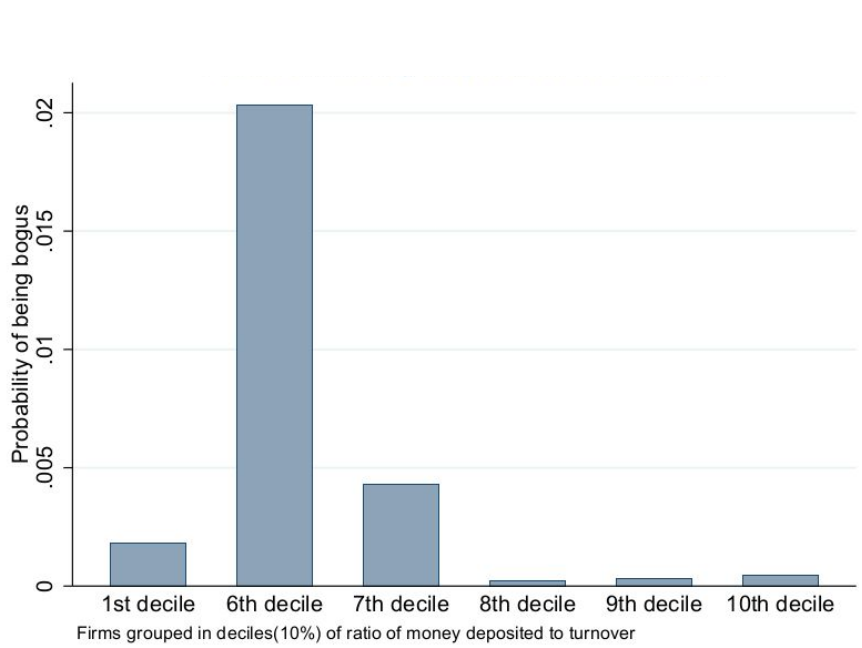


Government receives tax on $40 less value added. Surplus is divided between offenders.

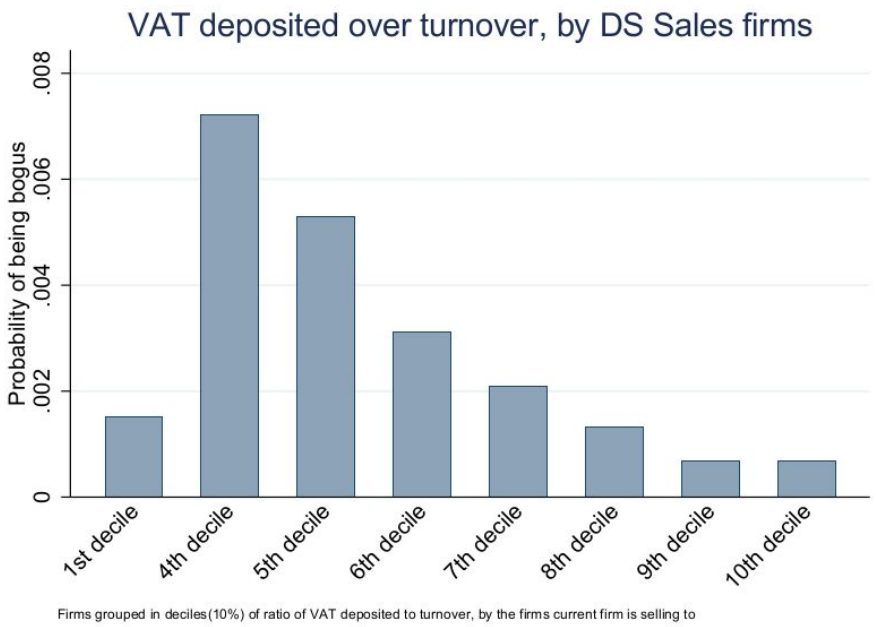
Our approach: rely on data

Past bogus firms -> what is suspicious behavior -> similar behavior in present -> target

Bogus firms tend to have low profit margins

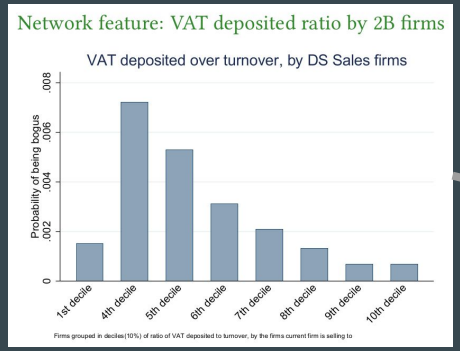


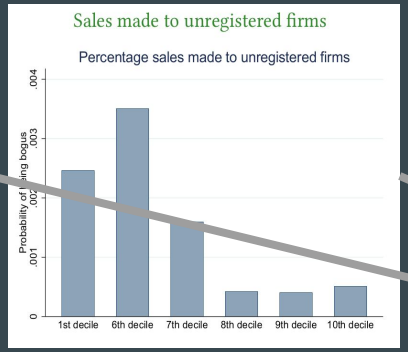
… and trading partners with low profit margins

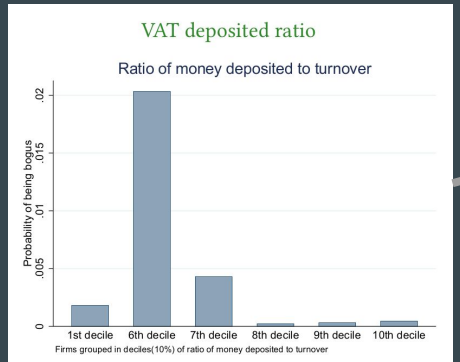


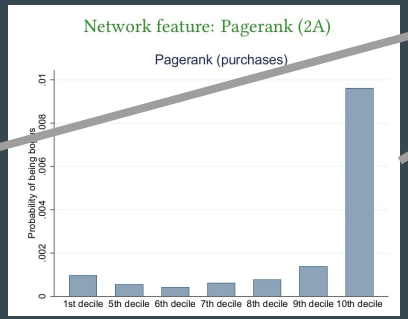
Our approach

Past bogus firms -> **what is suspicious behavior -> similar behavior in present** -> target



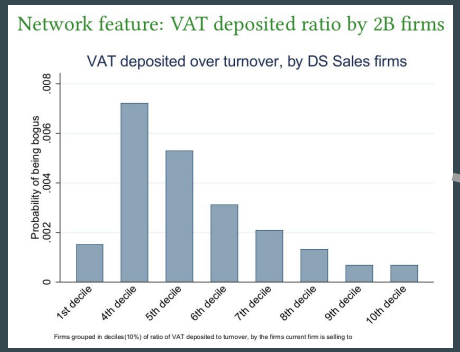


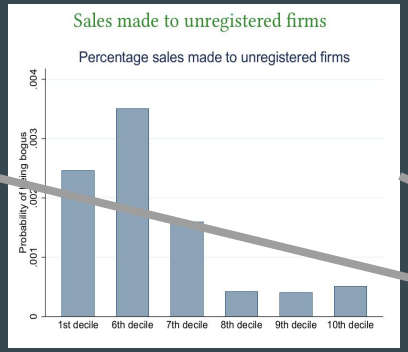


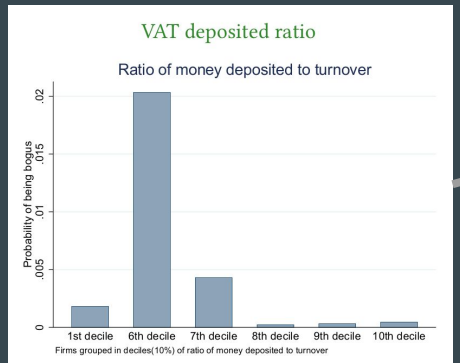


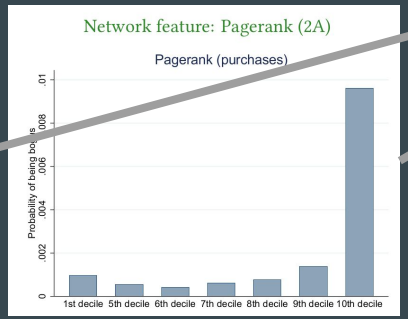
Our approach

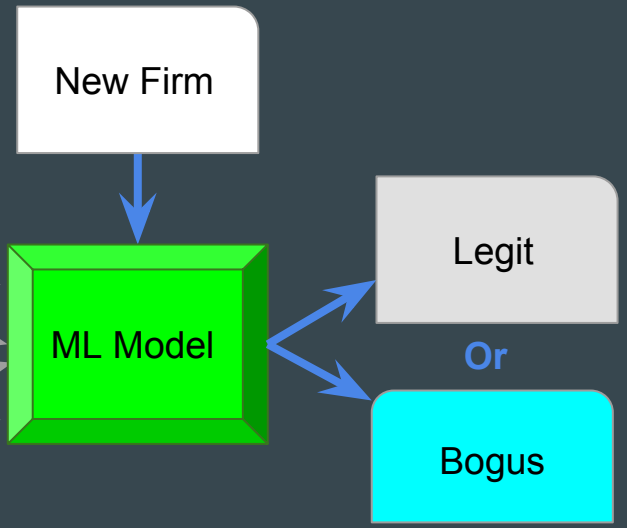
Past bogus firms -> what is suspicious behavior -> **similar behavior in present -> target**











Our approach

Past bogus firms -> what is suspicious behavior -> **similar behavior in present -> target**

results of inspections

Target suspicious firms (by the model prediction) for inspection by the tax authority

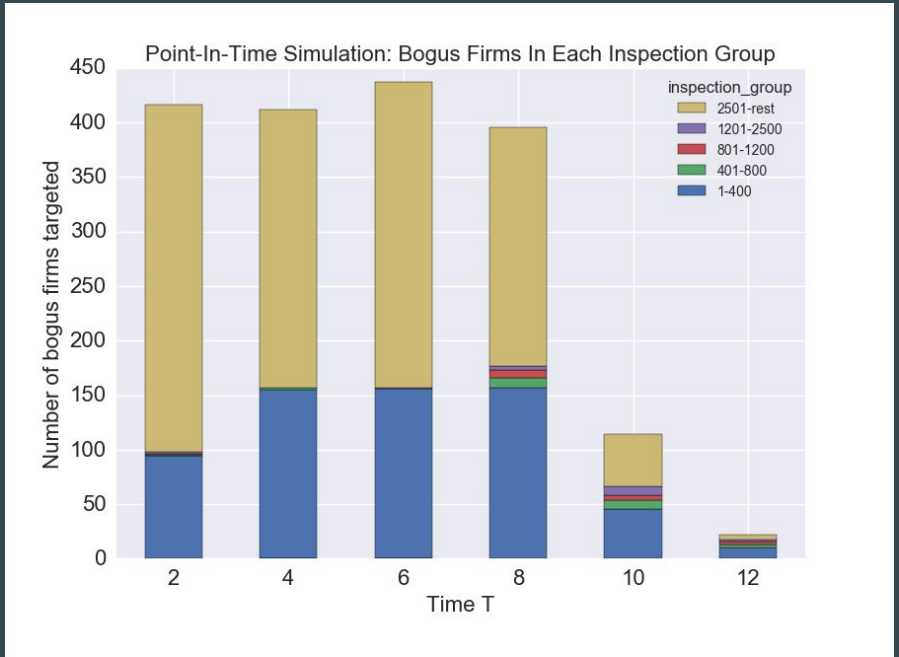
Inspection results -> feed back into the system, improve future prediction

Evaluate impact

Added benefit: objective, fair targeting of inspections

Results

Of the top 400 suspicious firms our model finds, we expect at least 30% to be bogus.



Conclusion

Machine Learning on VAT data identifies rare tax evaders.

- High accuracy. In Delhi, Potential cost savings of 30 Million USD.
- We work with tax authorities: Delhi, Tamil Nadu, Mexico, Dominican Republic.

Approach can be applied to many other tax evasions and tax data

- increase revenue, optimize use of scarce resources (audits, inspections)
- Income tax evasion, VAT evasion, property tax misreporting, ...

Requirements

- Data on many tax transactions (anonymized, censored). Preferably digitized.
- A clear problem to solve, evasion or other issue

We have many other ideas on working with revenue authorities on their data - talk to us!

E-auditing

E-auditing:

Digital “paper trail” + ML => monitoring of service provision

Teacher attendance - mobile phone call records
Health workers give vaccines - electronic immunization cards/app
Welfare payments delivered - Aadhaar records
Collusion in public procurement - public records of auctions

. . .

Thanks!

ofir@precisionag.org