Optimal Audit Targeting with Machine Learning: Evidence from Pakistan

Nicholas Lacoste[1] and Zehra Farooq[2]
[1]Tulane University, [2]Federal Board of Revenue, Pakistan

Abstract

This paper bridges welfare economics and machine learning econometrics to develop empirically implementable algorithms for optimal audit targeting. We derive a sufficient statistic-based targeting algorithm that depends on three individualized causal effects: the immediate revenue recovered from an audit, the causal effect of an audit on long-run tax revenue, and the marginal administrative cost of an audit. We estimate these effects with a variety of machine learners comparing causal forests, LASSO, gradient boosted trees, and neural networks using the universe of Pakistani income tax returns, exploiting years in which audits were assigned completely at random. We implement our targeting algorithms in out-of-bag years, comparing them to the real-world policy when audits were partially or entirely targeted. We show that the real-world audit program in Pakistan lost almost 173,000 Rs (about $1,700) in net revenue per-audit, while our optimal policy generates 285,000 Rs (about $2,800) in expected net revenue per-audit. We also find that targeting audits based on immediate recoup is sub-optimal to targeting on long-run deterrence in this setting. Moving forward, our framework offers a general approach to empirical welfare maximization using machine learning in resource-constrained policy settings.

Introduction

Motivation

- Governments typically target income tax audits on predicted evasion likelihood
- This ignores (1) long-run behavioral responses to audits, (2) heterogeneous administrative costs, and (3) taxpayer burden from audits

Objectives

- Derive a targeting algorithm to maximize social welfare
- Apply that algorithm in a real-world setting, compare to traditional approach

Theory/Targeting Algorithm

$$\theta_i = \eta_i \frac{\widehat{WTP_i}}{\Delta\widehat{C_i}} + \lambda \frac{\Delta\widehat{R_i^m} + \Delta\widehat{R_i^f}}{\Delta\widehat{C_i}}$$

- θ_i is a sufficient statistic for optimal audit targeting
- Under mild assumptions, 3 individualized causal effects are needed to target audits optimally:
  - 1) ΔR_i^m = immediate revenue recoup from auditing person i
  - 2) ΔR_i^f = effect of an audit on person i's future tax revenue
  - 3) ΔC_i = administrative costs to audit person i

Table 1

Year	(1) Total Filers	(2) First Time Filers	(3) Audit Regime	(4) Number Audited	(5) Share Audited	(6) Share Evaders	(7) % Liability Evaded (Avg.)
2014	1,108,177	–	Random	52,683	4.8%	38.9%	39.8%
2015	1,350,325	367,732	Random	69,732	5.2%	34.9%	30.4%
2016	1,599,220	330,375	Parametric	30,777	1.9%	28.9%	25.5%
2017	1,930,069	398,529	Parametric	9,140	0.5%	29.6%	21.0%
2018	2,830,049	949,199	RAMS	9,211	0.3%	14.2%	11.5%
2019	3,142,855	560,352	RAMS	Unobserved	–	–	–
2020	3,286,852	306,589	RAMS	Unobserved	–	–	–
2021	2,871,756	74,579	RAMS	Unobserved	–	–	–

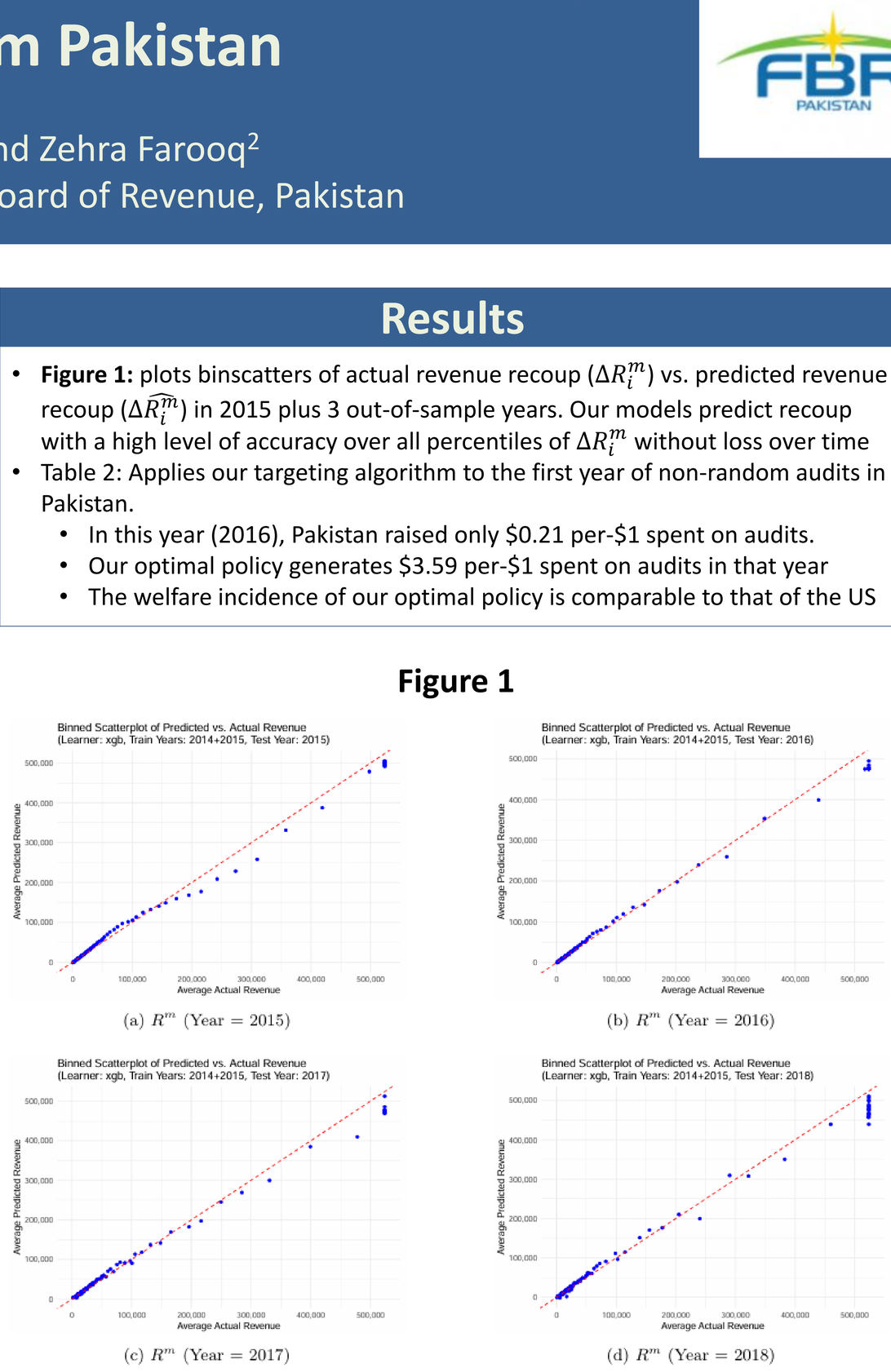
Data and ML Estimators

- **Table 1:** We observe all Pakistani income tax returns and audits from 2014 – 2021
- We also observe regional tax office expenditures and administer surveys of officer time-use to estimate $\Delta\widehat{C_i}$
- In 2014 and 2015, audits were **completely randomized** in Pakistan, allowing us to estimate $\Delta\widehat{R_i^m}$ and $\Delta\widehat{R_i^f}$ with causal machine learning
- Train cross-validated and debiased ML estimators of individualized treatment effects:
  - (1) LASSO
  - (2) XGBoost
  - (3) Neural Networks with L2-regularization

Results

- **Figure 1:** plots binscatters of actual revenue recoup (ΔR_i^m) vs. predicted revenue recoup ($\Delta\widehat{R_i^m}$) in 2015 plus 3 out-of-sample years. Our models predict recoup with a high level of accuracy over all percentiles of ΔR_i^m without loss over time
- Table 2: Applies our targeting algorithm to the first year of non-random audits in Pakistan.
  - In this year (2016), Pakistan raised only $0.21 per-$1 spent on audits.
  - Our optimal policy generates $3.59 per-$1 spent on audits in that year
  - The welfare incidence of our optimal policy is comparable to that of the US

Figure 1

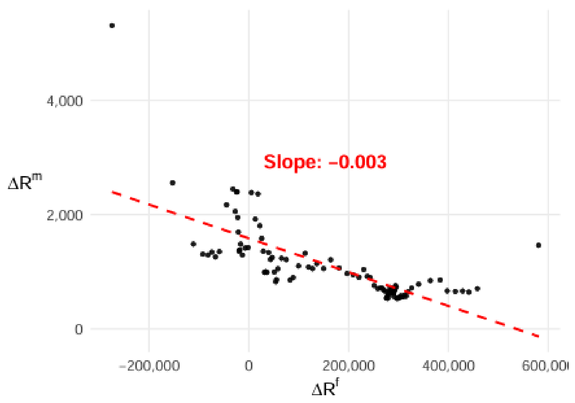


(a) R^m (Year = 2015) (b) R^m (Year = 2016) (c) R^m (Year = 2017) (d) R^m (Year = 2018)

Table 2

Metric	Observed Policy	Machine Learners: Optimal Policy (Best Learners)	Optimal Policy (Causal Forest)	Optimal Policy (LASSO)	Optimal Policy (XGBoost)	Optimal Policy (Neural Network)
2016						
Mean ΔR_i^m	23,233.92	1,232.59	158,762.60	6,410.33	2,616.49	3,805.91
Mean ΔR_i^f	23,308.52	393,949.90	64,394.88	254,429.00	283,183.25	392,868.28
Mean ΔC_i	219,341.60	110,012.21	129,071.16	112,488.92	132,375.99	110,124.21
MVPF	-0.904	1.579	3.058	2.137	2.294	1.576
NSB	0.212	3.592	1.729	2.319	2.159	3.602
Number Audited	30,747	58,822	3,632	57,527	48,885	58,763

Figure 2



- **Table 1:** There is a tradeoff between upfront recoup and long-run deterrence
- Evidence that audits push people into the informal sector if large amounts of evasion are detected upfront.

Conclusions

- We present a framework for optimal audit targeting with machine learning and corresponding empirical evidence of its efficacy
- In Pakistan (and possibly many other countries), we argue that targeting audits on immediate recoup likelihood is a poor approach to growing the long-run tax base. We provide evidence that the tax base can be grown substantially by following our ML-based targeting algorithms

Contact

Nicholas Lacoste
Tulane University
Email: nlacoste1@tulane.edu
Website: https://nickmlacoste.github.io/

Scan To Read Full Paper