Using Divide-and-Conquer to Improve Tax Collection

Samuel Kapon ⓡ Lucia Del Carpio ⓡ Sylvain Chassang[*,†]

UC Berkeley, Haas — INSEAD — Princeton University

April 28, 2024

Abstract

Tax collection with limited enforcement capacity may be consistent with both high and low delinquency regimes: high delinquency reduces the effectiveness of threats, thereby reinforcing high delinquency. We explore the practical challenges of unraveling the high delinquency equilibrium using a mechanism design insight known as "divide-and-conquer." Our preferred mechanism takes the form of Prioritized Iterative Enforcement (PIE). Taxpayers are ranked using the ratio of expected collection to capacity use. Collection threats are issued in small batches to ensure high credibility and induce high compliance. Following repayments, liberated capacity is used to issue the next round of threats. In collaboration with a district of Lima, we experimentally assess PIE in a sample of 13,432 property taxpayers. The data both validate and refine our theoretical framework. A semi-structural model suggests that keeping collection actions fixed, PIE would increase tax revenue by roughly 10%.

Keywords: prioritized iterative enforcement, divide-and-conquer, tax collection, limited government capacity.

---

[*]Kapon: skapon@berkeley.edu; Del Carpio: lucia.delcarpio@insead.edu; Chassang: chassang@princeton.edu.
Research funding from INSEAD's Dean's Innovation Fund and Princeton's Griswold Center for Economic Policy Studies is gratefully acknowledged.

[†]We are particularly grateful to the municipality of Jesús María, and especially Pablo Paredes and Patricia Corrales for being patient and supportive partners under exceptionally difficult circumstances. We thank seminar participants at Baruch College, Chicago Booth, Goethe University, Indiana University at Bloomington, Lausanne, LSE, MIT, Northwestern, NYU, Princeton, PSE, Seminars in Economic Theory, Simon Fraser University, SITE (2021), ThReD (2021), the University of Alabama, the University of British Columbia, the University of Calgary, the University of Pennsylvania, and Yale University for very helpful comments.

1 Introduction

Tax collection with limited enforcement capacity may be consistent with both high and low delinquency regimes: if delinquency is low, limited coercive collection capacity is enough to discipline the majority of potential individual deviators; if delinquency is high, coercive collection capacity is spread more thinly reducing the effectiveness of threats. Bassetto and Phelan (2008) refer to the latter possibility as a tax riot. In principle, divide-and-conquer, a theoretically important but untested principle from mechanism design, can be used to unravel the undesirable high-delinquency equilibrium. The main idea is to assign taxpayers a known priority in advance, and collect from non-compliant taxpayers in strict order of priority. In theory, this should lead the highest priority taxpayer to comply, inducing the second highest to comply, and so on. This paper seeks to figure out the engineering challenges of making such a mechanism work in practice.

Our benchmark model adapts known insights about divide-and-conquer to tax collection, and expands on them to deal with issues of incomplete information and bounded rationality. Our preferred implementation, Prioritized Iterative Enforcement (PIE), embeds divide-and-conquer in an extensive form game in which small groups of taxpayers are iteratively threatened with fast-track collection if they fail to settle, i.e. if they fail to make the payment required to be compliant within deadline. Upon repayment, enforcement capacity tied up by past threats is released and used to issue collection threats to a new group of taxpayers. An attractive property of PIE is that the threats it issues are systematically enforced.

We partnered with a district of Lima (Peru) to experimentally evaluate the impact of PIE on the collection of property taxes from 13432 delinquent taxpayers. Using a semi-structural model to simulate relevant counterfactuals, our preferred estimate is that implemented at scale, PIE would increase tax revenue between 9% and 12%. The data highlights a novel trade-off ignored by the theory. Successful implementation of PIE requires carefully tuning the flow of threats: fewer but more potent threats expedite settlements but also narrow the pool of taxpayers likely to settle at any moment. We hope that our work facilitates the application of divide-and-conquer in other contexts, including debt collection, law enforcement, and organizational governance.

The model. We consider a government entitled to collect an amount of taxes due D_i from taxpayers labeled $i \in \{1, \cdots, N\}$. With prior probability q_i, taxpayer i may not be able to settle (i.e. make a required payment in order to be treated as compliant), and their ability to settle is private information.[1] The government is able to coercively collect the amount D_i but doing so is costly in terms of time and resources. The difficulty is that the government is able to perform at most αN coercive collections (i.e. collect directly from taxpayers that do not make voluntary payments), with $\alpha \in (0, 1)$. Instead of coercively collecting taxes, the government can offer agents to settle their taxes by paying a given price $P_i < D_i$. Agents who settle are not collected on.

In a random enforcement mechanism, the government coercively collects from αN taxpayers drawn uniformly from non-settlers. This can result in multiple equilibria, exhibiting both high and low collection levels: if most taxpayers comply, minimal enforcement capacity is enough to deter unilateral deviators; if many taxpayers don't comply, enforcement rates are low, and collection threats are less dissuasive. PIE guarantees that the government achieves the highest possible revenue provided taxpayers are minimally rational. Specifically, when the number of taxpayers is large, under any undominated strategy profile, PIE achieves the highest possible Bayes-Nash equilibrium revenue under any mechanism.

Under PIE, taxpayers are ranked according to a scoring rule trading-off expected collection $(1 - q_i)D_i$ and expected capacity use q_i. Taxpayers are then iteratively threatened with collection unless they settle at a price that increases with delay. At any point in time, the number of taxpayers being threatened is equal to the remaining collection capacity, so that

[1] Taxpayers may experience a liquidity shock, or a personal crisis. We assume that the amount of tax due D_i is known to the government, which is true of property taxes. We show in Online Appendix OD that the analysis extends when the taxpayer has private information about the amount D_i of taxes they owe.

settling is a dominant action whenever feasible. This causes the low-settlement equilibrium to unravel.[2] Taxpayers that do not settle face coercive collection.

Field implementation. From April 2021 to September 2021, we partnered with the municipality of Jesús María, a relatively affluent district of Lima (Peru), to evaluate the impact of PIE on the collection of property-related taxes from 13432 taxpayers delinquent in their first-quarter (Q1) payment. The experiment ended with the municipality's decision to adopt a simplified version of PIE.[3]

Jesús María typically enjoys high ultimate collection rates, but expends significant resources on tax-collection. Roughly 30% of taxpayers are delinquent at some point of the annual collection process. Lenient enforcement in 2020 (due to Covid 19) also raised concerns of reduced tax compliance in 2021. Consultation with the city's collection unit revealed a bottleneck in tax collection: a costly garnishment process requiring legal and bank involvement. The city estimated its capacity to issue garnishment orders to roughly 400 per month (this corresponds to the available coercive collection capacity in our model). Taxpayers need to be notified with a formal writ before garnishment can take place, but the city is capable of issuing several thousand such writs per month.

At the end of Q1 2021, we randomly assigned delinquent taxpayers to two treatment arms. A control arm implemented the default collection policy used by the city. As Figure 1 illustrates, it proceeds in steps: (1) delinquent taxpayers are informed that they are delinquent all at once; (2) writs are sent to majority of taxpayers; (3) if repayment does not occur over the next quarter, garnishment orders are issued, overweighting taxpayers that owe the most taxes. While the city tends to prioritize taxpayers with higher taxes due at the garnishment stage, there is no explicit priority shared with taxpayers. In fact, depend-

[2]This relates to the point made by Lazear (2006) and Eeckhout et al. (2010) that when government capacity is limited, random public crackdowns may be more effective than the thinly spread incentives provided by uniform enforcement. Here, focused incentives also speed up compliance. Once repayment occurs, enforcement capacity tied-up by threats can be redeployed to induce other agents to comply.

[3]We describe and evaluate this simplified policy in Section 6. One difference is that only two rounds of threats are issued, whereas PIE issues new collection threats as soon as capacity is freed up.

ing on subjective evaluations of expected returns from administrators, some taxpayers with moderate taxes dues may be garnished before taxpayers with higher taxes due.

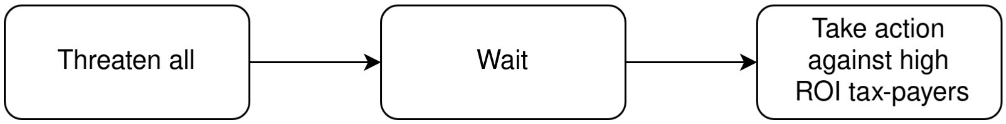


Figure 1: Collection in the control group

Note: This figure describes collection in the control group. Delinquent taxpayers are informed they are delinquent all at once and writs are sent to the majority (*Threaten all*). If repayment does not occur over the next quarter (*Wait*), then garnishment orders are issued, prioritizing those who owe the most (*Take action against high ROI tax-payers*).

Our treatment arm implements PIE: (1) at any one time, we issue a small number of targeted threats, with clear short-term deadlines designed to induce prompt repayment; (2) notifications and writs are promptly sent to targeted taxpayers; (3) whenever repayment occurs, freed-up threat capacity is applied to the next batch of taxpayers. Repayment propensities $1 - q_i$ used to rank taxpayers were predicted using historical data. Each week, the 400 highest-ranked taxpayers who had not paid more than 50% of their taxes were in priority group G1, the next 400 were assigned to priority group G2, and the rest were assigned to priority group G3. Group membership was updated weekly based on repayments. Members of priority group G1 were given a clear deadline for repayment, set on average at 6 weeks, failing which garnishment would occur. Members of group G2 were promised that their income would be garnished within 12 weeks if taxes remained unpaid. In addition, they were informed that they could be moved to group G1 at any time. Members of group G3 did not receive collection promises. They were informed of the taxes they owed, of the penalty for late payment, and that they could be moved to group G2 at any time. Taxpayers assigned to the control group (C) received a similar notification of the taxes they owed.

As Figure 2 illustrates, the core idea of PIE is to iteratively issue high-powered threats to a thin-but-moving slice of the tax-base. The effectiveness of PIE hinges on whether the

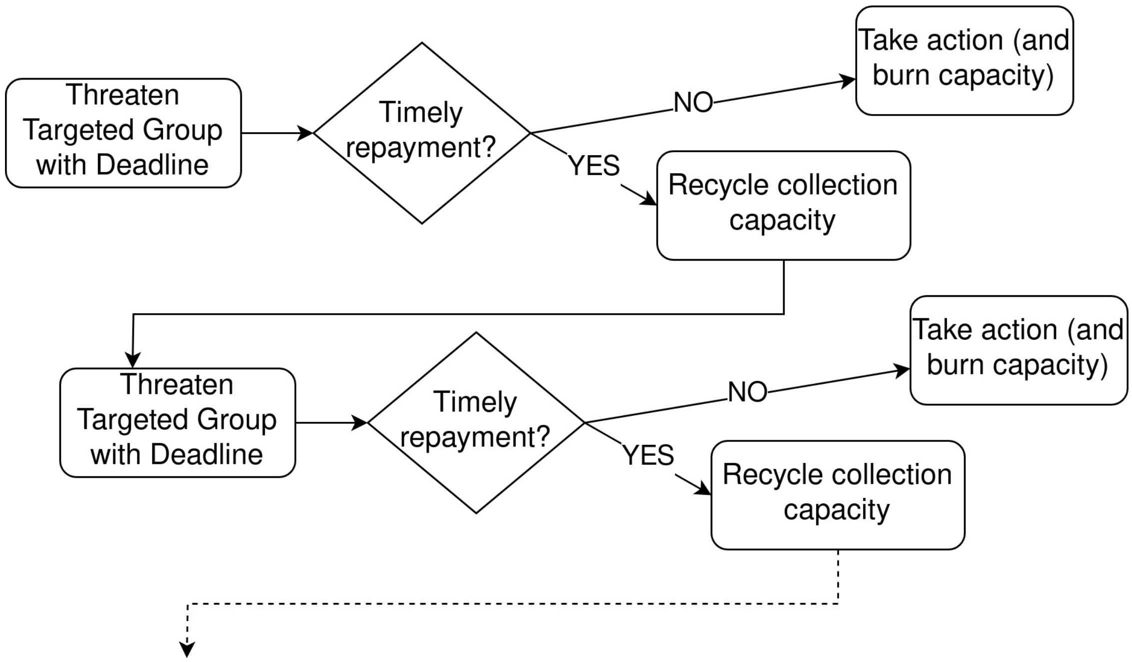


Figure 2: Collection under Prioritized Iterative Enforcement

Note: This figure describes collection under prioritized iterative enforcement. A small number of targeted taxpayers receives an explicit collection promise with a short-term deadline (*Threaten Targeted Group with Deadline*), and is promptly issued notifications and writs. Whenever repayment occurs, freed-up threat capacity is applied to the next batch of taxpayers (*Recycle collection capacity*). If a taxpayer in the threatened group does not pay, a garnishment action must be taken against them (*Take action (and burn capacity)*).

increase in settlement speed due to high-powered incentives allows us to recycle collection capacity enough to compensates for the reduced number of taxpayers targeted at any time. There is no trade-off under our benchmark theory because taxpayers settle immediately when it is dominant to do so. In practice, delays in best-response make the trade-off less obvious: it may be optimal to threaten groups larger than the available collection capacity.

Importantly, both arms were allocated the same budget of costly garnishment actions (up to 1000 over five months), and realized garnishments are very close: 533 garnishment actions were taken in the treatment arm, while 531 were taken in the control arm. Differences in outcomes between treatment and control are driven by differences in the way enforcement capacity is used, rather than differences in allocated capacity.

Empirical findings. We report raw findings, as well as estimates from a semi-structural model that identifies the impact of different collection steps on behavior, but does not seek to estimate a microfounded model of taxpayer choice. This model permits the evaluation of some counterfactual policies, but not all.

Raw findings show that the key ingredients needed for prioritized enforcement to be effective were present: clear short-term promises significantly increased the repayment propensity of taxpayers, and repayment propensities were meaningfully predicted by our scoring rule. In addition, our specification of PIE (parameterized by the size of priority groups and the deadlines they are given) was effective in increasing the efficiency of collection. Over a five months period, taxes collected in the treatment group were 9.2% higher than in the control group (a more robust estimate using our semi-structural model to correct for the impact of large payments suggests a treatment effect of 7.5%). In addition, the number of collection actions other than garnishment (notifications, and legal writs) taken for the treatment group was three times smaller than in the treatment group, saving significant labor costs.[4]

We build a semi-structural model of taxpayers' behavior as a function of collection actions taken. The model does not microfound choices made by taxpayers. As a result, we can only evaluate counterfactual mechanisms that keep the nature of threats the same, and only change the process of issuing threats.[5] Estimates confirm that G1 priorities considerably increase taxpayers' repayment propensities. In contrast, G2 priorities, and G3 priorities respectively had small positive and negative effects on repayment propensities. Regarding collection actions, receiving a legal writ had a large positive effect, similar to that of receiving priority G1, while receiving an initial notification had a negligible effect.

Simulated treatment effects for the experiment as it was implemented suggest an expected

[4] We have reliable data on collection actions taken, but limited records of time use by city employees. Time cost estimates for different actions provided by the municipality suggest that control collection actions took up 60% of employee time, versus 40% for treatment. See Online Appendix OF for details.

[5] A fully structural model would allow us to evaluate mechanisms that fail to deliver on promises at some rate. Our data does not let us commit confidently to such a model. Strong responses to weakly enforced writs suggests that optimization under rational expectations may not be the correct model of taxpayer behavior.

increase in revenue of 7.5%, at the cost of 558 expected garnishments. Because issuing writs has a large impact on repayment, we evaluate a counterfactual in which the number of writs sent across treatment and control are equated. This variant of PIE would increase revenue over control by 10.8%, at the cost of 547 expected garnishments. Ranking taxpayers according to taxes due alone turns out to increase collected revenue to 12.3% but also increases the expected number of garnishments issued to 595. A simplified policy, corresponding to the one adopted by the municipality, maintains a broad use of writs, but issues only two rounds of G1 priorities to 500 taxpayers, with a seven weeks deadline. It yields a predicted increase in revenue of 11.3%, but induces 642 garnishments on average.

Related literature. We believe this paper is the first experimental evaluation of divide-and-conquer mechanisms in the field.[6] There is a rich and growing theoretical literature on the use of divide-and-conquer to implement desirable social outcomes under all rationalizable strategy profiles (Abreu and Matsushima, 1992, Segal and Whinston, 2000, Spiegler, 2000, Segal, 2003, Winter, 2004, Dal Bó, 2007, Eliaz and Spiegler, 2015, Halac et al., 2020, 2021). We bridge the gap between this theoretical literature and practical implementation. The evidence is encouraging to the theory, but highlights the importance of frictions such as bounded rationality, and delay in best-response. We hope that this improved understanding stimulates other efforts to take divide-and-conquer to the field.

The paper contributes to the economic literature on tax-compliance reviewed in Hallsworth (2014), Mascagni (2018), Alm (2019) and Slemrod (2019). It relates to letter-based experiments in which researchers have partnered with tax authorities to evaluate how deterrence in the form of audits and penalties affects tax compliance. Slemrod et al. (2001), Kleven et al. (2011) and more recently De Neve et al. (2021) evaluate the impact of auditing threats on taxpayers' compliance, finding a meaningful impact of threats, especially on taxpayers

[6]The insight of divide-and-conquer naturally shows up in policy. A well documented case is the homicide reduction program Operation Ceasefire (Braga et al., 2001, Kennedy, 2011, 2012). It publicly prioritizes the allocation of police resources to homicides in order of occurrence, dissuading gangs to initiate violence.

for whom third party information is not available. An important aspect of PIE is that it focuses on the optimal deployment of a fixed collection capacity. As a result, enforcement promises issued under PIE are always credible, even if PIE were to be implemented at scale. This would not necessarily be the case for experiments that significantly increase the use of costly deterrent steps. We show that even under strict credibility constraints, it is possible to increase revenue compared to a control collection process allowed to use weakly enforced threats.

Our treatment is closely connected to communication interventions in which researchers remind taxpayers of the possible consequences of non-compliance. For instance, Bergolo et al. (2023) provide small- and medium-sized firms in Uruguay information about the costs of non-compliance. They find that reminders of existing policies have a large effect compared to variation in the likelihood of audits.[7] Still, Alm (2019) cautions that the impact of information likely depends on taxpayers' baseline expectations. If taxpayers overestimate the likelihood of enforcement, information may reduce compliance. This more qualified view is consistent with the differential impact of priority groups G1, G2 and G3 on settlement in our experiment.

Our paper belongs to a body of work applying mechanism design insights to tax collection by capacity constrained governments. One popular class of mechanisms uses cross validation to induce truthful reporting. Among others, Pomeranz (2015) and Pomeranz and Vila-Belda (2019) highlight the role that value-added taxes can play in obtaining third party reports of firms' income. Naritomi (2019) evaluates a cross-validation mechanism in which taxpayers are encouraged to submit sales receipts to the tax authority in exchange for various fiscal benefits. Balan et al. (2022) studies the value of using village chiefs as intermediaries for tax collection.

Finally, an active literature has examined determinants of tax compliance other than

[7]Other recent work also finds that increasing the salience of penalties increases compliance, at least in the short term (Fellner et al., 2013, Castro and Scartascini, 2015, Chirico et al., 2016, Meiselman, 2018).

deterrence. Following Alm et al. (1992)'s seminal observation that observed tax-compliance levels are often inconsistent with narrow rational self interest for credible degrees of risk aversion, an active literature has studied tax-morale and the importance of intrinsic versus extrinsic incentives in achieving compliance (Luttmer and Singhal, 2014, Del Carpio, 2014, Dwenger et al., 2016, see for instance). De Neve et al. (2021) studies the value of simplifying communications between tax-collection agencies and taxpayers. Bergeron et al. (2021) shows that tax rates may be above revenue maximizing rates because high taxes dissuade compliance.

Structure. Section 2 sets up our benchmark model. Sections 3 and 4 describe our experimental context and experimental design. Section 5 reports raw outcomes of interest and confirms that the key ingredients needed for PIE to be effective are present. Section 6 estimates a model of taxpayer behavior and uses it to evaluate counterfactual policies of interest. An appendix and an online appendix assemble proofs, experimental materials, and further empirical exploration, as well as findings from a pilot of PIE in a laboratory setting.

2 Framework

We clarify the point of divide and conquer in a stylized model. We then turn to a more realistic framework allowing for heterogeneity, incomplete information, and bounded rationality.

2.1 A stylized model

N taxpayers indexed by $i \in \{1, \cdots, N\}$ each owe the government a fixed amount D. The taxpayers and the government are all risk-neutral. If a taxpayer fails to repay on time, the government can coercively collect amount D – in our experimental setting, by garnishing bank accounts. The difficulty is that the government has limited capacity: it can coercively collect from only $\alpha N \geq 1$ taxpayers with $\alpha \in (0, 1)$. To induce taxpayers to pay voluntarily, the government can make settlement offers and commit to an enforcement rule according to the following extensive-form game:

1. The government gives each taxpayer the possibility to settle by paying a fixed price $P \in (\alpha D, D)$. Taxpayers who settle are spared from coercive collection.

2. Taxpayers simultaneously decide whether or not to settle and pay price P.

3. The government coercively collects D from taxpayers who do not settle according to a known enforcement rule.

We contrast two possible enforcement rules (the next section studies arbitrary mechanisms):

- Random enforcement: In period 3, up to αN taxpayers are drawn with uniform probability from the set of non-compliant taxpayers, and assigned to coercive collection.

- Prioritized static enforcement: Taxpayers are given a known priority rank in period 1. In period 3, up to αN non-compliant taxpayers are subjected to coercive collection in order of their preassigned rank. By convention, we rank taxpayers in descending order of their index $i \in \{1, \cdots, N\}$, so that taxpayer 1 has the highest priority.

The value of prioritized enforcement. The following result clarifies the value of prioritized enforcement: it selects a high collection equilibrium as the unique strategy profile surviving the iterated elimination of dominated strategies. In contrast, random enforcement induces multiple equilibria involving both high and low collection levels.

Proposition 1. *Fix a choice of P by the government in period 1.*

(i) Under random enforcement, there exists a Nash equilibrium such that all taxpayers settle, and a Nash equilibrium such that all taxpayers refuse to settle.

(ii) Under prioritized static enforcement, a unique strategy profile survives iterated elimination of dominated strategies: all taxpayers settle.

Under random enforcement, a small collection capacity can deter individual deviations, but only when overall compliance is high. If many taxpayers do not pay their taxes, then available capacity is thinly spread and fails to dissuade tax-evasion. Prioritized enforcement unravels this last equilibrium by focusing capacity on a marginal set of taxpayers. It is dominant for the αN highest ranked taxpayers to settle their taxes. Anticipating this, it is a best response for taxpayers with rank up to $2\alpha N$ to settle, and so on.

This model assumes away important frictions. First, taxpayers' ability to repay is complete information. Second, taxpayers exhibit a common belief in rationality: many rounds of deletion of dominated strategies are needed to unravel the low settlement equilibrium. Realistically, laboratory experiments suggest that people rarely engage in more than 2 to 3 rounds of strategic introspection.

2.2 Modeling realistic frictions

We first introduce incomplete information about ability to pay, heterogeneity across taxpayers, and provide bounds on collection under any mechanism in any Bayes Nash equilibrium. We then propose an extensive form mechanism, PIE, that attains this performance bound in large populations, even when players are boundedly rational.

We allow taxes due D_i to vary with taxpayer identity i. Collection costs are also heterogeneous: coercive collection against agent i consumes $\lambda_i \in [\underline{\lambda}, \overline{\lambda}] \subset (0, \infty)$ units from the principal's enforcement capacity αN. With probability $q_i \in [\underline{q}, \overline{q}] \subset (0, 1]$, taxpayer i is exogenously unable or unwilling to repay their taxes, say because they are experiencing a liquidity shock. The ability to repay, denoted by $\xi_i \in \{0, 1\}$, is private information: a taxpayer knows whether they are able to repay, but the government does not.

Bounds on any mechanism. We establish bounds on any incentive compatible collection by considering partial implementation in direct, truthful, and obedient mechanisms. Taxpayers send a message $m_i \in \{0, 1\}$ revealing whether they are capable of making payments; the government then sends price offers $P_i \in [0, D_i]$ and settlement recommendations $\widehat{s}_i \in \{0, 1\}$; finally, the government implements an enforcement action $a_i \in \{0, 1\}$, with $a_i = 1$ denoting coercive collection. Note that settlement offers P_i, recommendations $\widehat{s}_i$, and enforcement actions a_i are correlated random variables across taxpayers. Realized coercive collection must satisfy capacity constraint: $\sum_{i=1}^{N} a_i \lambda_i \leq \alpha N$. The government seeks to maximizes tax-revenue Π collected through either voluntary settlement or coercive collection,

$$\Pi \equiv \frac{1}{N} \sum_{i \in I} s_i P_i + \rho \times \frac{1}{N} \sum_{i \in I} a_i \xi_i (D_i - s_i P_i)$$

where s_i denotes i's settlement decision, and $\rho \in [0, 1]$ is the collection recovery rate.[8]

Proposition 2 (upper-bound on equilibrium revenue)**.** *Under any mechanism, in Bayes Nash equilibrium, expected per capita tax revenue is bounded above by*

$$\max \left\{ \frac{1}{N} \sum_{i=1}^{N} \delta_i (1 - q_i) D_i \;\middle|\; (\delta_i)_{i \in \{1, \cdots, N\}} \in [0, 1]^N \textit{ such that } \sum_{i=1}^{N} \delta_i q_i \lambda_i \leq \alpha N \right\}. \tag{1}$$

Problem (1) is a linear optimization problem with a single constraint. Term δ_i is agent i's probability of settling when they can. The marginal benefit of increasing δ_i is $(1 - q_i) D_i$ while the marginal shadow cost is $\mu q_i \lambda_i$ where μ is the Lagrangian multiplier associated with the capacity constraint. Hence it is optimal to set $\delta_i = 1$ for all agents such that $(1 - q_i) D_i / q_i \lambda_i > \mu$ and $\delta_i = 0$ for all agents such that $(1 - q_i) D_i / q_i \lambda_i < \mu$.

We show in Online Appendix OD that Proposition 2 extends nearly as is when taxes due D_i are privately observed by agents but are uncertain to the principal. The government can achieve no better collection than bound (1), with non-payment rates q_i depending on take-it-or-leave-it settlement price offers P_i chosen by the principal.

[8] When $\rho < 1$ coercive collection is wasteful. Since taxpayers observe their own payment ability, optimal mechanisms never implement coercive collection on taxpayers who reveal that they can make payments.

Prioritized iterative enforcement. We now describe PIE, an extensive-form mechanism that attains the bound of Proposition 2 when the population N is large, even when taxpayers are boundedly rational. The underlying payoffs and types are unchanged. Settlement takes place over time $t \in [0,1]$, and taxpayers can choose to settle their taxes at any time.[9] The government commits to the following settlement and collection process:

(i) Taxpayers $i \in \{1, \cdots, N\}$ are ranked in decreasing order of score

$$z_i \equiv \frac{(1-q_i)D_i}{q_i \lambda_i}. \tag{2}$$

(ii) In each period t where they haven't settled, taxpayer i receives a settlement offer $P_{i,t} = D_{i,t} - \nu(1-t)$ with $\nu > 0$.[10]

(iii) In each period t, taxpayers are informed of their effective rank, taking into account the settlement behavior of others. Specifically, they receive signal $x_{i,t} = i - \sum_{j<i} s_{j,t}$.

Taxpayers who have not settled face collection in decreasing order of rank at time $t = 1$.

We model bounded rationality using Li (2017)'s notion of non-obviously dominated play. Let h_i denote a private history of taxpayer i, and $\sigma_i : h_i \mapsto s_i \in \{0,1\}$ a feasible strategy. Denote by σ_{-i} strategy profiles by players other than i, and by ω possible moves of nature. Let $u_i(\sigma_i, \sigma_{-i}, \omega | h_i)$ denote the realized payoff of agent i given history h_i, their own behavior σ_i, the behavior of others σ_{-i}, and realized moves of nature ω (here agents' ability to pay).

Definition 1. *A strategy σ_i obviously dominates a strategy σ_i' if and only if, for every history h_i potentially on the equilibrium path, at which strategies σ_i and σ_i' first differ,*

$$\sup_{\sigma_{-i},\omega} u_i(\sigma_i', \sigma_{-i}, \omega | h_i) \leq \inf_{\sigma_{-i},\omega} u_i(\sigma_i, \sigma_{-i}, \omega | h_i).$$

[9] The analysis is unchanged if taxpayers become capable of repaying their taxes at random times.

[10] This price schedule is chosen for simplicity. Any strictly increasing price schedule would be as effective.

Strategy σ_i is non-obviously dominated if no strategy obviously dominates it. Laboratory evidence reported in Online Appendix OE, as well as evidence from auctions (Kagel et al., 1987, Kagel and Levin, 2001) suggests that mechanisms achieving implementation under non-obviously dominated strategies behave closer to equilibrium predictions in practice.

Proposition 3 (Revenue under PIE)**.** *Assume that taxes are collected using PIE. Fix $\eta > 0$. With probability approaching 1 as N gets large, for any profile of non-obviously dominated strategies,*

(i) taxpayers with rank j such that $\frac{1}{N}\sum_{i\leq j} q_i\lambda_i \geq \alpha + \eta$ do not settle;

(ii) taxpayers with rank j such that $\frac{1}{N}\sum_{i\leq j} q_i\lambda_i \leq \alpha - \eta$ settle;

(iii) per capita tax revenue approaches

$$\max\left\{\frac{1}{N}\sum_{i=1}^{N}\delta_i(1-q_i)(D_i-\nu)\;\middle|\;(\delta_i)_{i\in\{1,\cdots,N\}}\in[0,1]^N\ s.t.\ \sum_{i=1}^{N}\delta_i q_i\lambda_i\leq\alpha N\right\}. \tag{3}$$

Since the slope of settlement offers $\nu > 0$ can be made arbitrarily small, this implies that PIE approaches the bound of Proposition 2 under weak assumptions about rationality.

2.3 Known limits and design implications

Section 2.2 suggests a benchmark design that performed well in the lab (see Online Appendix OE). Nonetheless it exhibits known limits that our field design tries to address.

2.3.1 Commitment

Sections 2.1 and 2.2 assume that the government has commitment power: it keeps feasible collection promises, and taxpayers believe that it will. In practice, taxpayers need not take enforcement threats seriously since local governments do not always follow-through.

Because Jesús María had taken specific steps to ban tax amnesties in the two years before our experiment, it is plausible that its enforcement promises are at least partly credible.

We enhance the government's commitment power by specifying collection priorities, and setting clear deadlines. The goal is to make unfulfilled threats noticeable.[11] We choose deadlines sufficiently distant in time to give taxpayers the opportunity to repay, yet not so distant that it undermines the importance of compliance.

2.3.2 Delay in decision making

Under non-obviously dominated play, taxpayers settle if they can as soon as they understand it is a dominant action. In practice, even when an action becomes dominant, it may take taxpayers time to actually change their behavior. One reason for delay may be liquidity shocks. Alternatively, delay may occur if the payoff gain from prompt repayment is small. We estimate a model of delayed repayment in Section 6 but the following approximate accounting equality helps clarify the implications of delay for design:

$$\begin{aligned}
\text{Revenue} =&\ \text{Num. Taxpayers Threatened} \times \text{Settlement Probability} \\
\text{Num. Taxpayers Threatened} =&\ \min \left\{ \text{Population}, \right. \\
&\left. \text{Batch Size of Threat Group} \times \frac{\text{Total Collection Period}}{\text{Repayment Delay after Threat}} \right\}.
\end{aligned}$$

Term (Total Collection Period /Repayment Delay after Threat) corresponds to the number of times the same collection capacity can be used to issue threats. If the collection period is one quarter, and delay is a month, we can recycle capacity three times.

Larger threat groups. The first implication of delay is that it may be useful to use threat groups that are larger than available collection capacity. In a model with vanishing delay,

[11] We formalize this argument in Online Appendix A.

capacity is recycled sufficiently many times that batch size is not a limiting factor: the entire addressable population receives a threat. With delay, the number of times capacity can be recycled is moderate, so that increases in batch size could be beneficial. The trade-off is that greater batch size may reduce settlement probability.

As an example, imagine there are 500 taxpayers, 50 collection days, that repayment occurs with a 10 day delay, and that the city can coercively collect from 20 people every 10 days. If the city threatens 20 people every 10 days, threats can be fulfilled under any scenario. Assume that in this case, 100% of taxpayers comply. Then, $20 \times 1 \times 5 = 100$ people voluntarily settle their taxes. If the city threatens 50 people every 10 days, threats are less credible. Assume that only 60% of people comply in this case. Still, $50 \times .6 \times 5 = 150$ people settle their taxes, and threats are fulfilled in equilibrium. With a 10 day delay, issuing more threats than available collection capacity increases revenue.

In our experimental design we issue roughly twice as many direct threats as available garnishment capacity. However, this remains a conservative choice in equilibrium: the realized number of threats we need to enforce is half of available capacity.

Longer deadlines. When it takes people time to pay, short deadlines can be counterproductive if people simply can't raise necessary funds in time: it triggers costly garnishments on people who may have otherwise repaid. For this reason, longer deadlines may increase settlement probability enough to compensate for a potential increase in repayment delay.

Consider again a city with 500 taxpayers, 50 collection days, and able to perform 100 coercive collections in aggregate. For simplicity, assume that the city is conservative: at any given time the number of threats issued is equal to the remaining available collection capacity. Repayment is now stochastic: 60% of people can repay within 10 days, and 40% can repay within 25. People repay by the end of the deadline they are given. If the city sets a deadline of 25 days, the repayment rate is 100% so that the city can issue 100 threats twice, and 200 taxpayers settle voluntarily. If the city sets a deadline of 10 days, 60% of people settle, and 40% of the collection capacity is consumed, allowing the city to issue $100 \times .6$ threats on the second round, followed by $100 \times .6^2$, $100 \times .6^3$ and $100 \times .6^4$ threats on the remaining rounds. In expectation, roughly 140 people will settle their taxes voluntarily. Setting a longer deadline is beneficial.

In our field experiment, collection deadlines have an interquartile range of 4 weeks, running between 4 and 8 weeks, with a mean of 6 weeks. The variation in deadline reflects specific circumstances of the taxpayer, including the nature of the debt, and whether they are local residents.

This discussion defines a family of PIE mechanisms indexed by length of deadline and size of threat-group (see Figure 3). The standard collection process corresponds to a large threat group and long deadlines.[12] Under both rationalizability and non-obviously dominated play, small threat groups and short delays are weakly optimal. However, when repayment takes time, a poor choice of parameters may cause PIE to be ineffective. Our experiment identifies parameters for which PIE improves over standard collection.

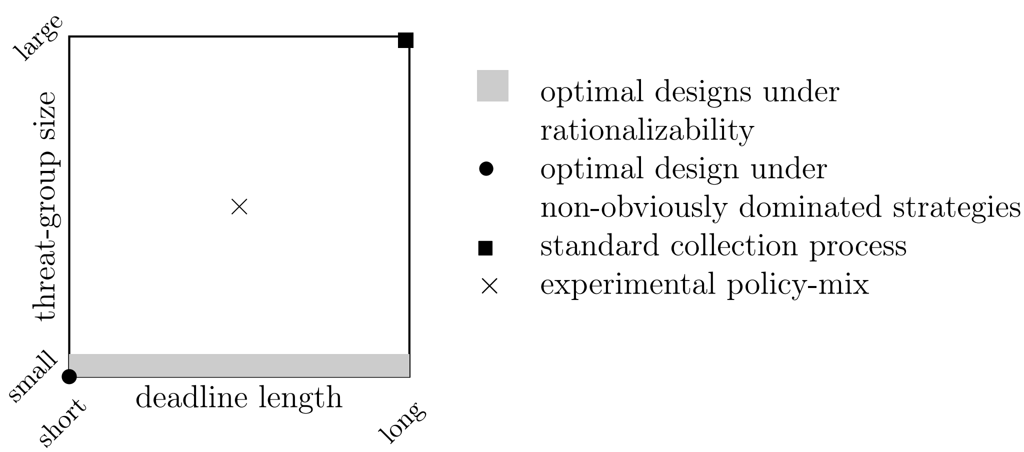


Figure 3: The range of PIE mechanisms

Note: Deadline length refers to the deadline given to members of group G1 after which garnishment begins. Threat-group size refers to the number of taxpayers included in group G1 at any time.

[12] Figure 6 in Section 4 plots the number of collection actions taken over time. It clarifies that standard collection processes debtors in parallel, while PIE processes debtors serially in small batches.

3 Experimental Context

From April to September of 2021, we partnered with Jesús María, a district of Lima (Peru), to collect property-related taxes from 13,432 delinquent taxpayers. This section details the context for our experiment, and why this context seemed well suited to evaluate PIE.

3.1 General context

Property taxes and user charges. Our study targets the two most important municipal taxes in Peru: (i) property taxes, based on assessed property values, with tax rates between 0.2% and 1% of assessed value, and (ii) user charges, paying for public goods such as trash collection and public safety. In 2020, they represented almost 50% of municipal revenues.

Jesús María. Jesús María is one of 43 municipal districts of Lima, and belongs to the top quartile in terms of income. As of 2020, there were above 60,000 properties in the district, 90% of which were residential units. The average assessed value of properties amounted to Peruvian soles S/. 110,000 (around US$30,000).

Properties are linked to over 35,000 registered taxpayers, 90% of which live in the district. In 2020, total municipal taxes due stood at US$15.8 million, with a per capita average equal to US$435. The distribution of taxes dues is skewed to the right. The ten largest taxpayers own commercial properties (e.g. shopping malls) and represent 16% of total taxes due, while the top 500 taxpayers accounted for 42% of total taxes due.[13]

By the end of 2020 Jesús María had a delinquency rate of 12% for property taxes and 24% for user charges. Jesús María's annual collection costs are roughly US$1 million.

Suitability for experimentation and external validity. Concerns that Covid 19 would reduce tax compliance created an opportunity for experimentation. Because tax collection

[13]The top 10 taxpayers tend to pay taxes on time. Only one entered our sample of delinquent taxpayers (it was assigned to the treatment group) and for a relatively small amount of taxes due (2000 soles).

costs were already high the collection budget was held fixed.[14] Hence, any improvement would require deploying available collection capacity more effectively.[15] Experimentation was facilitated by the fact that within legal constraints, Peruvian municipalities have significant degrees of freedom over their collection process. In addition, Jesús María had banned the use of tax amnesties in the previous two years, making collection threats more credible.

Jesús María exhibits fairly high enforcement capacity compared to other settings in which PIE may be applied. This was a conscious choice, motivated by the logistical challenges of implementing divide-and-conquer in the field. This affects the external validity of our findings in two ways. Organizations with less administrative resources may find it difficult to implement PIE. At the same time external validity is improved by the fact that we relied solely on existing municipal employees, rather than hiring workers of our own (see Online Appendix OF for a description of task allocation). The upside of PIE may also be higher in settings with lower capacity. Finally, we show in Section 6 that implementations of PIE can be simplified at little cost, potentially facilitating adoption.

3.2 The standard collection process

Collection steps. Property taxes and user charges are enforced jointly, on a quarterly basis. Figure 4 summarizes the timeline of collection steps.

Collection consists of two stages: *ordinary* collection and *coercive* collection. Ordinary collection starts right after the payment deadline and involves: (i) sending bulk reminders (mostly through emails and text messages) to all taxpayers who missed the deadline, (ii) calling roughly the top 50% of delinquent debtors with the highest tax due to remind them of their liabilities, and (iii) a formal letter notifying the amount owed ("valor"). This letter triggers a countdown at the end of which coercive collection can begin.

[14]Jesús María's collection costs are roughly 8% of revenue compared to 5-6% for neighboring municipalities with similarly high rates of tax compliance.

[15]PIE can improve collection taking capacity budgets as given. Increasing capacity may also be worthwhile depending on the context. Initial implementations of PIE will consume organizational bandwidth, which is not being accounted for in our analysis.

Coercive collection for property taxes (resp. user charges) can only begin 1 (resp. 20) working day(s) after the taxpayer is formally notified. If the city chooses to initiate coercive collection, a legal writ (sometimes referred to as the 'REC1') issued by employees with formal legal training must be sent to the taxpayer. Collection actions can only begin 7 working days after the taxpayer is notified of the writ.

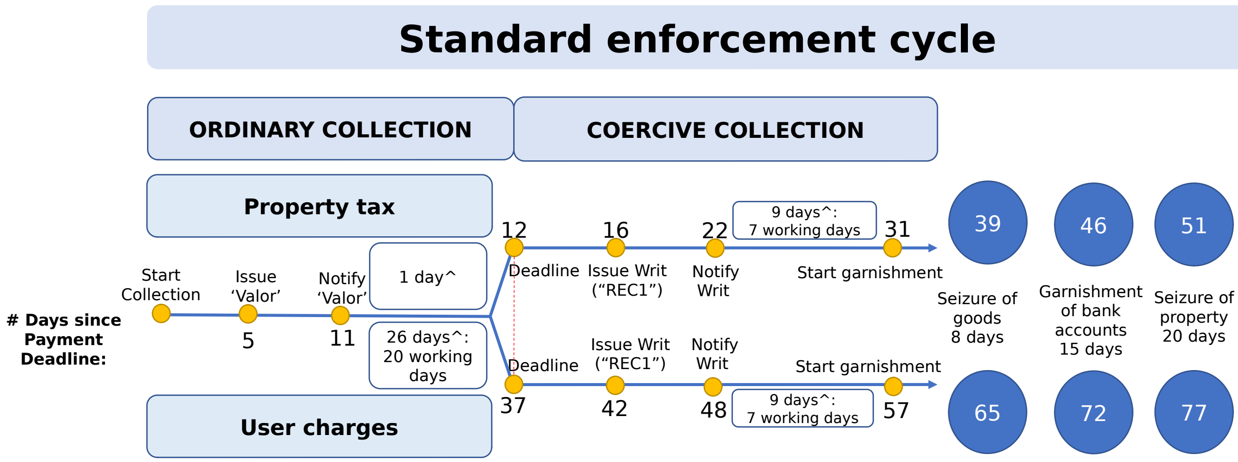


Figure 4: Standard collection timeline

Note: Standard collection timeline for Jesús María in 2021. *Valor* is the notification of delinquency. The *Writ* initiates coercive collection. *Garnishment* is the seizure of bank accounts. The number of days legally required between different actions (e.g., Notify "Valor" to Coercive Collection) depends on whether delinquency is over property taxes or user charges.

In principle the city government has three options for collection: garnishing bank accounts; seizing goods at the property; and placing a lien on the property. Although some taxpayers do not use banks, garnishing accounts is by far the most effective measure. Due to the Covid 19 pandemic, garnishment was the only collection step taken in 2020 and 2021.

Some collection steps depend on the amount of taxes due. The largest 500 taxpayers are assigned a dedicated collection agent that manages their account.[16] For debtors owing less than 100 soles, debts across different quarters are pooled and enforced with low intensity

[16] We balanced the assignment of these 500 taxpayers to treatment and control, with the same collection agent performing collection duties for both arms.

once or twice per year. They amount to a small share of taxes due, and we exclude them from our analysis.

Penalties. Daily interest corresponding to an annual rate of 10.8% is applied to all delinquent debt. When coercive collection begins, a 10% penalty is added and the taxpayer is charged for some of the collection costs.[17] In addition, taxpayers under coercive collection are registered with a credit-risk agency and show up as delinquent in national databases.

Capacity constraints. The collection unit consists of 15 city employees, led by a head and deputy head. Five staff handle regular collection, with one dedicated to the top 500 taxpayers. Three staff manage coercive collection, two handle notifications, one oversees IT, and two offer general support.

	Monthly capacity (units)	Unit cost (soles)
Phone calls	5237	1.60
SMS	16000	0.16
E-mails	16000	0.18
'Valor' issue	10687	0.90
'Valor' notification	10687	1.83
Writ ("REC1") issue	5990	2.68
Writ ("REC1") notification	5990	1.92
Garnishment issue	400	60.80
Garnishment notification	400	6.37

Table 1: Operational capacity and unit costs

Note: Monthly capacity and cost (in Peruvian S/.) of actions are shown for collection unit staff in Jesús María in 2021. "Valor" is the notification of delinquency. The *Writ* initiates coercive collection. *Garnishment* is the seizure of bank accounts.

Collection efforts are constrained by workforce, budget, and service provider capacity (e.g. banks). City officials' estimates reported in Table 1 show that issuing notifications,

[17]Collection charges average to US$35 per delinquent taxpayer.

and issuing formal writs are inexpensive and have a high monthly capacity (5,000-16,000). In contrast, garnishment is limited to 400 actions per month, and represent a bottleneck. Accordingly, our experiment focused on the effective use of garnishment capacity. Treatment and control arms were each allocated a budget of 1000 garnishments over 5 months. Realized garnishments are also very close: 533 garnishment actions were taken under treatment, and 531 were taken under the control arm.

4 Experimental Design

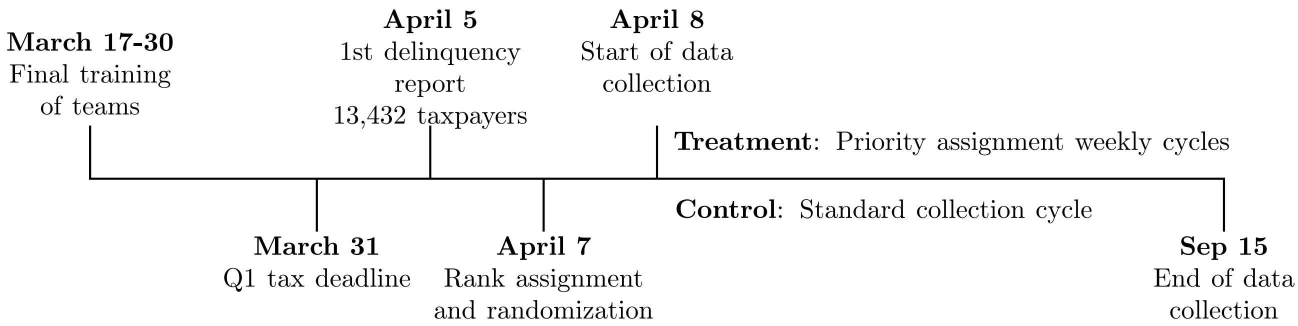
4.1 Scope and treatment arms

The experiment was pre-registered with the American Economic Association's Randomized Controlled Trial registry under number 7305. We did not commit to a pre-analysis plan because uncertainty over the logistics of implementation made it difficult to anticipate the correct analytic framework. The sample population consisted of taxpayers delinquent on their first quarter (Q1) property tax or user charges by April 5th, 2021.[18] Figure 5 summarizes the experiment's timeline.

Following the payment deadline, 13,432 taxpayers who had not paid their Q1 2021 taxes as of April 5th and had taxes due above Peruvian soles S/.100 (around US$25) entered our experimental sample. Smaller debts were excluded. Debtors were all assigned a priority rank based on scores z_i defined using a statistical model of repayment described below.

Half of taxpayers were randomly assigned to a version of PIE described below; remaining taxpayers were assigned to the city's standard collection procedure. Following Banerjee et al. (2020), we drew our treatment assignment uniformly from the 10% most balanced assignments under the Mahalanobis distance with respect to age, taxes due, top 500 taxpayer

[18] The regular tax payment deadline of February 28th was extended to March 31st due to the Covid 19 pandemic. No enforcement measure was taken before that date.



Note: *Rank assignment and randomization* refers to the computation of rank z_i in PIE (see equation 2) and the assignment of taxpayers to treatment and control, respectively.

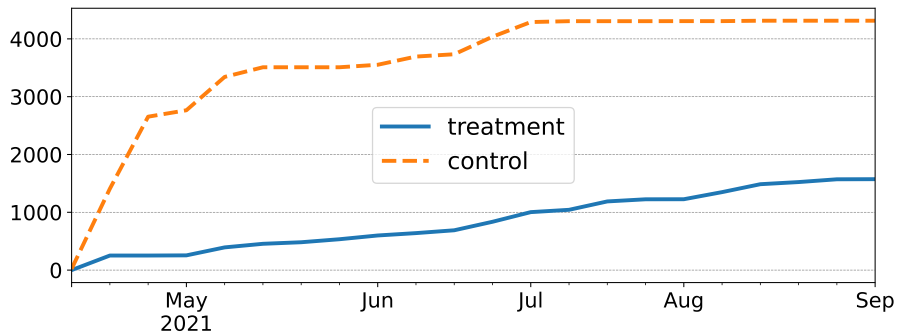
Figure 5: Experiment timeline

status, and predicted repayment. Table E.1 of Appendix E provides summary statistics.

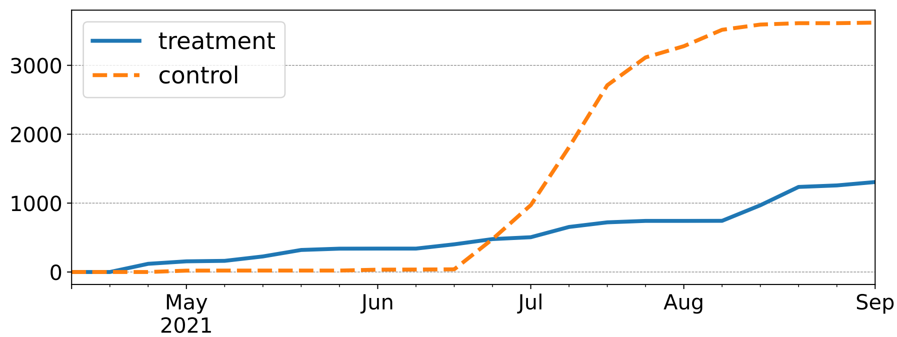
Control and Treatment. The control arm follows the collection process detailed in Figure 4 and described at length in Section 3.2: the bulk of taxpayers are concurrently issued identical notifications; after some delay, the majority of taxpayers are issued legal writs; after further delay a number of taxpayers are placed in garnishment. Taxpayers go through each stage of collection as a group, and are not issued prioritized threats with tight deadlines.

Implementation details for the treatment arm are provided in Sections 4.2 and 4.3. We implement a version of PIE in which small groups of taxpayers are iteratively issued assertive collection threats with a deadline set on average at 6 weeks. Notifications and writs are issued only to threatened taxpayers, under the shortest schedule consistent with the law. Garnishment takes place if taxpayers fail to meet minimum payments by the deadline they were given. New threats are only issued when capacity is freed up either by taxpayers settling their taxes, or by garnishments being executed.

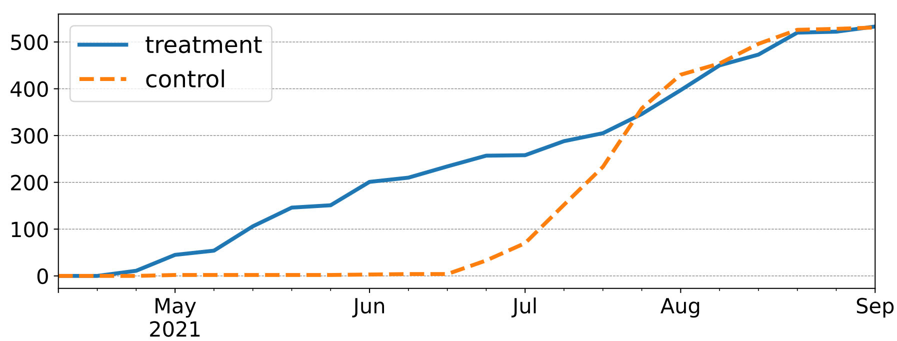
Figure 6 plots the number of collection actions taken over time and clarifies the difference between treatment and control. Both are associated with roughly the same final number of garnishments, but the processes leading up to these final numbers differ. Under control,



(a) Number of notifications issued.



(b) Number of writs issued.



(c) Number of garnishment orders issued.

Figure 6: Number of collection actions taken

Note: Panels (a), (b), and (c) plot the number of notifications, writs, and garnishment orders issued in treatment and control arms of the experiment over the course of data collection, April to September 2021.

notifications and writs are issued in bulk, and taxpayers go through each collection stage as a group. As a result, garnishment takes place late in the collection period. Under treatment, notification and writs are only issued following targeted threats. Collection steps, including garnishment, take place continuously. This enables faster recycling of collection capacity upon repayment. In short, standard collection processes debtors in parallel, while PIE processes debtors serially in small batches.

Differences between treatment and control are further clarified by Table 2. For different quantiles of taxes due, we report the *writ-enforcement* rate for taxpayers that were issued a writ before July 19^{th}.[19] That is, the share of taxpayers under garnishment among taxpayers that received a writ before July 19^{th}, but have made no payment by the end of the experiment, or received a garnishment by the end of the experiment. On average, more than 97% of writs are enforced under treatment. In contrast, the writ-enforcement rate in the control group is 31%. I addition, while the likelihood of enforcement is higher for taxpayers with higher taxes due, enforcement does not follow a strict priority rule as a function of taxes due. This is because in the control group, tax-collectors can prioritize based on subjective assessments. As result, while enforcement in the control group is not uniformly random, control taxpayers only face weakly prioritized enforcement.

Tax due quantiles	Writ-enforcement rate		# Unpaid Writ and No Garnishment		# Unpaid Writ or Garnishment	
	Control	Treatment	Control	Treatment	Control	Treatment
0-20%	0.208	1.000	156	0	197	21
20-40%	0.219	1.000	196	0	251	18
40-60%	0.193	0.970	356	1	441	33
60-80%	0.321	0.980	298	2	439	99
80-100%	0.509	0.995	202	2	411	367

Table 2: The writ enforcement rate is the share of taxpayers under garnishment among taxpayers that received a writ before July 19^{th}, but have made no payment by the end of the experiment or received a garnishment by the end of the experiment.

[19] By July 19th, a substantial share of writs that are eventually issued had been issued in both treatment and control.

Failure of SUTVA. Although we randomly assign a large number of taxpayers to treatment and control, we are also dealing with a single implementation organization: Jesús María's tax-collection unit. Due to its small size, we could not guarantee balance with respect to employee characteristics, and chose to rotate staff across treatment and control. Prioritized enforcement also proved less labor intensive than standard collection.[20] As a result, employees nominally assigned treatment regularly assisted employees assigned to the control group. This enabled the city to issue a large number of writs under the control arm.

This failure of SUTVA implies that comparing treatment and control arms likely underestimates the effect of PIE implemented at scale.[21] To draw more meaningful policy comparisons we interpret raw findings reported in Section 5 using a semi-structural model in Section 6.

4.2 Rank assignment

This section and the next describe the logistics of running PIE in the field. Score s_i, defined by (2), depends on taxes due D_i, repayment odds $1 - q_i$, and collection cost λ_i. In our case, D_i is observed and collection cost λ_i can be treated as roughly constant, so that normalizing it to 1 does not change the ranking of taxpayers. The challenge is to estimate $1 - q_i$.

We do so by averaging predictions from OLS, LASSO, and Random Forest models, using 2019 and 2020 administrative data (see Online Appendix OB). The main takeaways are that:

[20] Informal estimates from city officials suggest that standard collection was 50% more time consuming than prioritized enforcement. Online Appendix OF reports additional accounting details. For standardized tasks assigned to specialized collection agents, such as issuing writs, or garnishments, our design is plausibly near blind. We cannot fully rule out the possibility that some employees exerted differential effort across treatment and control arms in more informal interactions with taxpayers.

[21] Another possible concern for SUTVA is information sharing between treatment and control taxpayers. The impact of information sharing is potentially ambiguous. If control taxpayers update that the municipality is taking enforcement more seriously, then control taxpayers are more likely to comply. If instead, they infer that they are not at risk of being collected on, then they may be less likely to comply. The evidence of Drago et al. (2020) suggests that the former effect may dominate. Online Appendix OA reports additional evidence specific to our context. Pilot surveys for Del Carpio (2014) suggest that roughly 10% of taxpayers discuss their taxes or the tax-collection process with their neighbors. An attempt to replicate Drago et al. (2020) using block-level geolocation yields noisy but positive spillover estimates.

- Higher taxes due predicts greater repayment odds, so that score $(1-q_i)D_i/q_i$ prioritizes large taxpayers. This enhances the progressivity of tax collection.

- We estimate two models: an "endogenous model" using past repayments as a covariate and an "exogenous model" excluding past behavior. The endogenous model predicts repayment better, but potentially disicentivizes repayment. We predict each taxpayers' repayment odds using either our endogenous or exogenous model, with equal probabilities.

4.3 Prioritized iterative enforcement in the field

A total of 6704 taxpayers were assigned to the treatment group. Our implementation of PIE reflects legal constraints on the collection process, as well as concerns over commitment power, and delay in taxpayers' reactions.

Priority groups. At every point in time, taxpayers were assigned to one of three priority groups: G1, G2, and G3. The top 400 highest ranked taxpayers who had not paid more than 50% of their taxes were assigned to group G1, the next top 400 were assigned to group G2, and the remainder were assigned to group G3. Group membership was updated weekly. New members of a given group were sent a physical card and an email specifying the corresponding collection promise. A translated information letter for group G1 is reproduced in Table 3.[22]

G1 taxpayers faced a repayment deadline of 6 weeks on average, with some variation depending on individual circumstances.[23] Failure to pay would result in prompt garnishment. G2 members had a 12-week garnishment deadline and were informed that they could be moved to group G1 at any time. The rationale for group G2 was to induce taxpayers to prepare for payment early, but the data shows it had a limited impact. G3 members

[22] Translated and original information letters are provided in Appendix D and Online Appendix OF.
[23] See Online Appendix OB for a description of how deadlines correlate with repayment behavior.

NOTICE OF IMMINENT COLLECTION	
We remind you that you have the following debt outstanding with the municipality:	**Amount**
The coercive collection process will start at the latest on:	**Today + 6 weeks**
and it can start at any time and without prior warning.	
If the coercive collection process is started your debt will include penalties and administrative expenses regulated by law amounting to:	**Amount * .1 +US$35**
In addition to accruing a weekly interest of:	**Interest**
We remind you that it is on your own interest to pay immediately to avoid higher expenses. You can use any of the payment options listed below.	

Table 3: Information letter for priority group G1

Note: Text is translated from original spanish to english. Information letter is sent to new members of G1.

received no fixed garnishment promise but were warned about tax amounts, late penalties, and potential promotion to group G2. This implicitly informed taxpayers of their low priority, and data suggests this may have reduced repayment propensities.

Taxpayers assigned to control group C received a notification of the amount of tax they owed, via a letter of similar complexity to that sent to group G3 (see Online Appendix OF).

Letters sent to groups G1 and G2 specify a clear deadline for collection and are more assertive than letters to group G3 and group C. This is part of treatment. Prompt repayment allows the recycling of collection capacity at the core of PIE. To ensure that threats are effective, proper calibration is key: sufficiently many threats must be issued, but not so many that they cannot be acted on. Section 6 shows that only G1 letters increased taxpayer compliance. G2 and G3 letters had limited effects compared to control group C. In short, the impact of letters on repayment arises only from prioritization at the heart of PIE.

We deviated from assigning the 400 top-ranked delinquent taxpayers to group G1 in two ways. For the first assignment (April 5th, 2021), 200 G1 spots were assigned to the 200 top-ranked taxpayers, and 200 G1 spots were uniformly assigned to taxpayers with rank below the top 200. This provided an early estimate of the impact of G1 membership and validated two premises of PIE: specific short-term threats increase repayment intensity; and estimated repayment predicts actual repayment. The second deviation was to expand group G1 to 600 taxpayers in June 2021, reflecting the fact that the number of garnishments issued fell well short of available capacity.

Collection actions. In the treatment arm, only taxpayers assigned to group G1 were subjected to collection actions like notifications and writs. This strategy allowed us to expedite the collection process and set short repayment deadlines (see Figure D.1). For groups G2 and G3, we limited actions to sending an initial letter and making reminder calls, as in the control group. Although we initially thought garnishment was the only impactful collection step, Section 6 reveals that sending legal writs significantly increases taxpayer compliance, even if most writs have no consequences. This benefited the control group, which issued three times as many writs compared to the treatment group (see Figure 6).

Excess promise making and delay. In line with Section 2.3, we issued more threats than could be processed if all taxpayers defaulted. With a 4-week garnishment capacity of 200, extending to 300 over 6 weeks, breaking G1 promises becomes inevitable if over 75% of taxpayers default. However, high realized settlement rates keep actual garnishments well below this threshold. Across both treatment and control, approximately 1100 garnishments were issued over 5 months, against a theoretical capacity of 2000.

G1 promises were largely upheld. Of the 805 taxpayers in G1 due to Q1 debt, 371 missed payment deadlines and were slated for garnishment. Of these, 353 were actually garnished, yielding a 95% fulfillment rate. However, there was an average delay of 19 days in the actual implementation of garnishments.

5 Raw Findings

We report raw outcomes of interest, including tax revenue, the frequency of payment events, and the number of collection actions taken by the municipality. We also document that the basic requirements for PIE to be effective are satisfied: G1 priorities increase the settlement rate of taxpayers; and it is possible to predict repayment propensity.

5.1 Main outcomes

Tax collection. Figure 7 displays cumulative 2021 tax collection for the treatment and control groups over the five months following the 2021 Q1 tax deadline. We include all 2021 property taxes paid during that period, even if they correspond to Q2, Q3 or Q4 taxes. A similar figure restricted to Q1 taxes only is provided in Appendix OB.

By September 6, 2021, total collection in the treatment group was 9.2% higher than in the control group. The speed of collection is also higher under treatment than control. Because the distribution of taxes due has a long right-tail total revenue is potentially sensitive to the behavior of large taxpayers. In Section 6, we use a more robust model identified from binary payment decisions rather than payment amounts, and estimate that expected collection is 7.5% higher under treatment than control.

Collection actions. Table 4 (and Figure 6) report the number of collection actions taken in the treatment and collection arms. Although the number of garnishments issued is roughly the same across treatment and control, the municipality issued 3 times as many notifications and writs in the control arm as in the treatment arm.

Reduced-form impact on payment intensities. Table 5 reports coefficients from the OLS regression (4) of payment events in the next week on indicator functions for the most

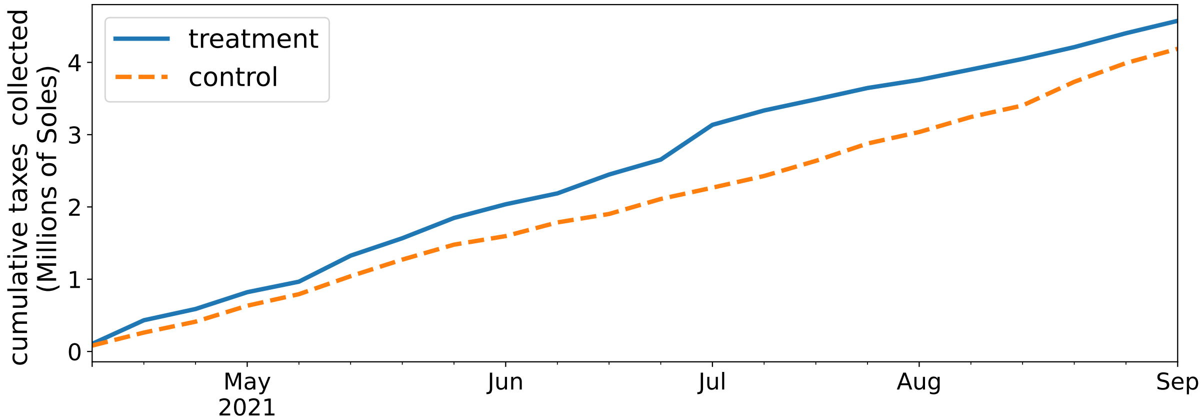


Figure 7: Cumulative Tax Collected April to September 2021

Note: Cumulative taxes collected (in millions of Peruvian S/.) include both voluntary payments by taxpayers and collection through garnishment.

	Number of taxpayers who have received		
	Notification	Writ	Garnishment
Treatment	1,573	1,306	533
Control	4,314	3,620	531

Note: *Notification* is the notice of delinquency. The *Writ* initiates coercive collection. *Garnishment* is the seizure of bank accounts.

Table 4: Number of collection actions taken

recent priority group, the most recent actions taken, and covariates X.[24]

$$\mathbf{1}\{\text{Payment Next Week}\} = C_0 + \sum_{\substack{G\in\{G1,\\ G2,G3\}}} \beta_G \mathbf{1}_G + \sum_{\substack{A\in\{Writ,\\ Garnishment\}}} \beta_A \mathbf{1}_A + \beta_{G1,A}\mathbf{1}_{G1}\mathbf{1}_A + \beta^\top X + \varepsilon \tag{4}$$

Priority G1 and writs have significant positive impacts on payment propensities, while the impact of priorities G2 and G3 is insignificant or negative. There is no crowding out between priority G1 and writs.

Regression (4) assumes that the impact of priorities and collection actions does not

[24]I.e. $\mathbf{1}_G = 1$ if G is the most recent priority group, and $\mathbf{1}_A = 1$ if A is the most recent collection action taken. Our preferred regression includes interaction effects between G_1 and collection actions, results are robust to including more covariates (see Online Appendix OA).

	Payment Next Week	
	Estimate (s.e.)	Estimate (s.e.)
G1	0.021 (0.002)	0.023 (0.002)
G2	−0.003 (0.002)	−0.000 (0.002)
G3	−0.005 (0.001)	−0.005 (0.001)
Writ	0.015 (0.001)	0.016 (0.001)
Garnishment	0.005 (0.003)	0.008 (0.003)
G1 & Writ	−0.003 (0.004)	−0.004 (0.004)
G1 & Garnishment	−0.013 (0.005)	−0.014 (0.005)
Est. Repayment Prob	0.072 (0.002)	0.033 (0.003)
Some Repayment	0.045 (0.002)	0.045 (0.002)
Share Repaid	−0.024 (0.000)	−0.024 (0.000)
Prev. Year Share Repaid at 3M	–	0.025 (0.002)
Constant	0.009 (0.001)	0.012 (0.001)
N	295504	295504

Table 5: OLS Regression of Payment Events on Priorities, Actions and Covariates

Notes. Robust standard errors in parentheses. Covariates include or not the share of taxes repaid 3 months after the Q1 deadline in the previous year, which fails balance tests (see Appendix E)

depend on the assignment history. This is incorrect and biases estimates. Taxpayers who remain delinquent late in the collection process are adversely selected. To account for this we estimate a model allowing for hidden types in Section 6.

Progressivity of tax-collection. Because priorities are increasing in order of taxes due, PIE enhances the progressivity of tax-collection. Figure 8 plots the share of tax revenue raised from taxpayers in the lower quantile q of taxes due. Treatment shifts the curve to the right. Those who owe more pay a larger share of total taxes under treatment than control.

5.2 Evidence on mechanisms

Raw findings confirm that the ingredients needed for PIE to improve collection are present:

- G1 priorities significantly increase the settlement rate of taxpayers;

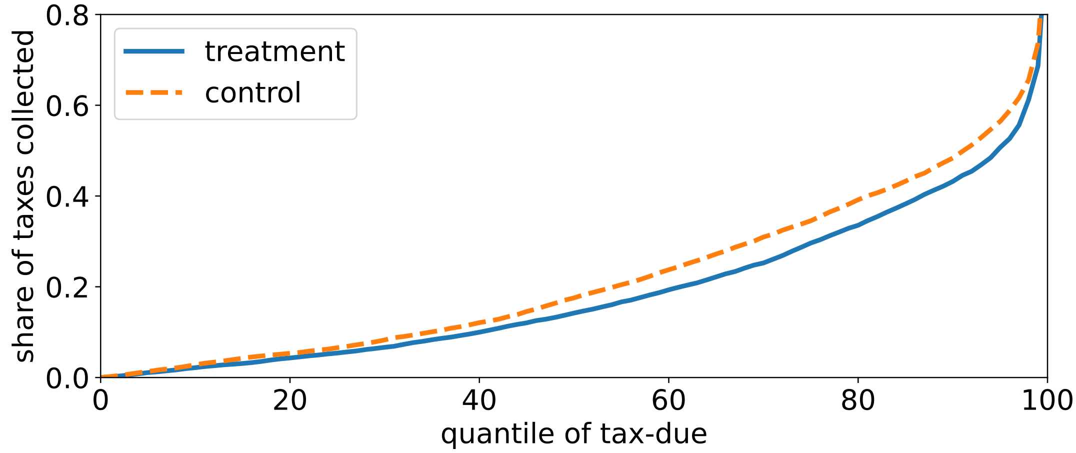


Figure 8: Share of total tax revenue collected as a function of quantile of taxes due.

Note: To compute the share of taxes collected from taxpayers at quantile q of taxes due, we divide taxes paid by taxpayers below the q^{th} quantile of taxes due, with taxes paid by all taxpayers.

- our ranking of taxpayers predicts repayment behavior.

In addition, we provide suggestive evidence that the impact of prioritized enforcement may get stronger over time, as the government's reputation for delivering on promises grows.

Impact of G1 priorities on settlement. Figure 9 plots the share of taxpayers who have repaid at least 50% of tax due for taxpayers in the first batch of group G1 assignments and a suitable comparison group from control.[25]

Taxpayers exhibit a significantly higher settlement rate under treatment than control. This is true even in the first few weeks during which no collection actions take place. The early impact of G1 membership is entirely driven by threats, rather than collection actions.

Predictability of repayment behavior. Figure 10 compares the settlement behavior of the top and bottom thirds of taxpayers across three predictive models: one incorporating historical repayment data, one excluding it, and one ranking taxpayers solely by total taxes

[25]The first batch of G1 assignments includes the 200 taxpayers with the highest rank, and 200 uniformly selected lower-ranked taxpayers. At this stage, taxpayers have not been differentially selected across treatment and control and we can use control taxpayers with similarly distributed scores as a comparison.

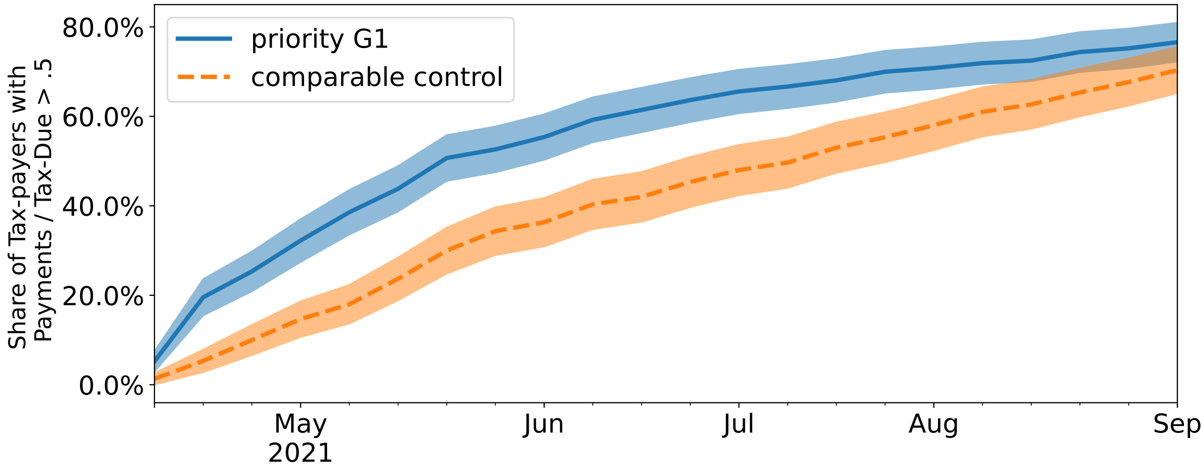
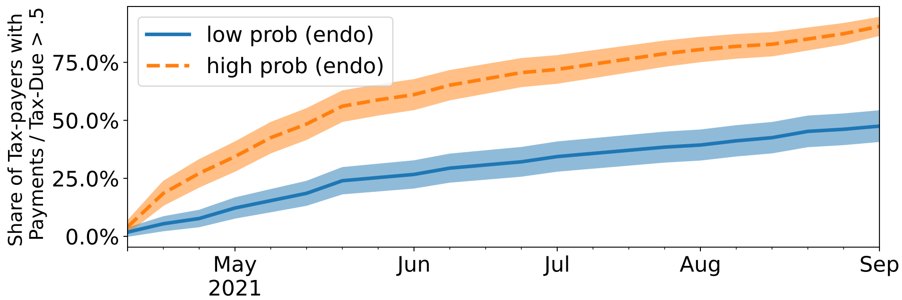


Figure 9: Repayment in G1 vs Control, First G1 Batch

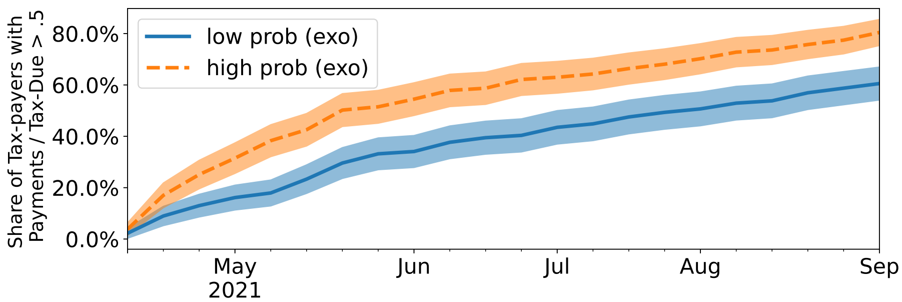
Note: The solid and dashed lines represent the share of taxpayers who have paid at least 50% of their tax due for the treatment and control groups. Shaded regions are 95% robust confidence intervals for coefficients $\beta_{T,t}$ and $\beta_{C,t}$, in the regression $\mathbf{1}\{\Pi_{i,t} \geq .5\} \sim \beta_{T,t}\mathbf{1}\{i \text{ in Treatment}\} + \beta_{C,t}\mathbf{1}\{i \text{ in Control}\}$ where $\Pi_{i,t}$ is taxpayer i's cumulated payment share of taxes due by week t.

due. The figure indicates that all models are predictive, but the inclusion of past repayment data enhances precision. Note that better classification does not necessarily translate into better performance. If the discrepancy between estimated precision and realized precision is worse for more complex models, then scoring based on complex models overweighs repayment probability and underweighs taxes due as scoring inputs.

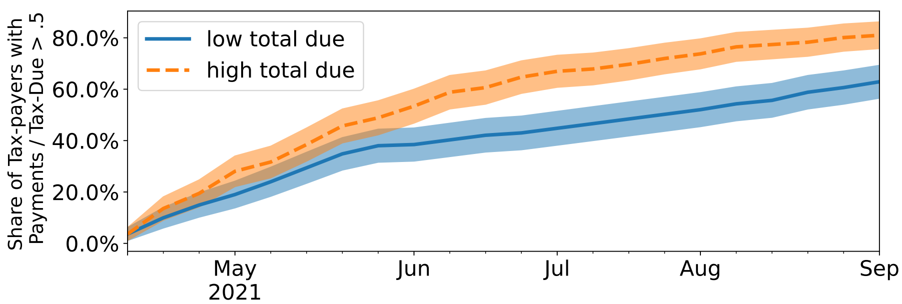
Reputation formation. We assess the impact of treatment on the propensity of taxpayers to be delinquent in subsequent quarters. For all taxpayers delinquent in the first quarter, we observe the amount by which they are delinquent in Q1 (their Q1 Debt), and whether they had any debt related to second quarter taxes (Q2 Debt). If the taxpayer is not delinquent with respect to Q2 taxes, then their Q2 Debt is set to 0. In order to control for the fact that taxpayers included in G1 are selected, we control for whether the taxpayer's score is high enough to be included in G1 in Q1. For the sample of taxpayers delinquent in Q1, we



(a) Predicted repayment uses past repayment behavior



(b) Predicted repayment excludes past repayment behavior



(c) Predicted repayment uses only taxes due

Figure 10: Share of population having repaid more than 50% of taxes due, by top third, and bottom third of predicted repayment probability, with 95% confidence bounds.

estimate the following linear model via OLS:

$$\frac{\text{Q2 Debt}}{\text{Q1 Debt}} = \text{Cst} + \beta_0 \times \mathbf{1}\{\text{Treatment}\} + \beta_1 \times \mathbf{1}\{\text{Potential G1}\} + \beta_2 \times \mathbf{1}\{\text{Assignment to G1}\} \times \mathbf{1}\{\text{Potential G1}\} + \varepsilon. \quad (5)$$

Findings reported in Table 6 suggest that among taxpayers with comparably high scores, being assigned to group G1 reduced delinquency rates by 23.8% in Q2.

	Q2 Debt/Q1 debt
Constant	1.401 (0.012)
Treatment	0.016 (0.015)
Potential Assignment to G1	-0.170 (0.027)
Potential Assignment to G1 × Treatment	-0.238 (0.041)
N	13432

Notes: Robust standard errors in parentheses.

Table 6: Impact of treatment on subsequent delinquency.

6 Counterfactuals

6.1 A semi-structural model

We estimate a model of settlement behavior as a function of threats and collection actions, that allows us evaluate key counterfactual policies. Those include an implementation of PIE at scale, an optimized policy reflecting insights from our experiment, and the simplified version of PIE ultimately adopted by the municipality.

Our modeling choices seek to address three empirical challenges. The first is adverse selection. Because treatment and control groups face a different process of threats and actions, delinquent taxpayers in each group become differentially selected over time. We include unobserved types to account for this differential evolution.

The second challenge is the thick right tail in the distribution of both taxes due and payment amounts. This makes inference about the impact of treatment on average payments difficult. As a result, we focus on the impact of priorities and collection actions on decisions to make a payment or not, and assume that conditional on a payment being made, the amount being repaid is independent of treatment status and collection actions taken.[26]

[26]This is a conservative choice. Repayment amounts tend to be larger under treatment, and treatment

The third challenge is to realistically capture individual choices. As Table 2 shows, the likelihood of enforcement associated with a writ is low under control. This makes us reluctant to assume that taxpayers' responses to writs reflects informed rational optimization.[27] Given this concern, we simply seek to fit behavioral responses to priorities and collection actions, without seeking to estimate a microfounded model of behavior. This is sufficient to let us evaluate counterfactual policies in which threats are similar to threats used in our experiment (in terms of deadlines and credibility), and only the process of threat issuance varies. Having a reliable micro-founded model of taxpayers' preferences would let us evaluate threats with significantly different deadlines and credibility.

The remainder of this section focuses on describing the model, and associated findings. Online Appendix OB investigates in-sample fit and robustness to specification changes.

Repayment behavior. We assume that each taxpayer i is associated with a persistent observed characteristic $\xi_i \in \mathbb{R}$ and a persistent unobserved type $\theta_i \in \mathbb{R}$, drawn i.i.d. across taxpayers from a Gaussian distribution $\mathcal{N}(0, \sigma^2)$. In our main implementation, we use as observed characteristic the taxpayer's predicted repayment probability from our most predictive model (which includes past repayment behavior).[28] Unobserved type θ_i serves to explain correlation in repayment behavior across periods, and captures the impact of selection over time: taxpayers who have not made repayments after 2 weeks are systematically different from taxpayers who have not made repayments after 3 months.

At the beginning of each period t (before payment actions are taken), the city government assigns the taxpayer a priority $g_t \in \{G1, G2, G3, \emptyset\}$ and takes a collection action $a_t \in \{\text{garnishment}, \text{writ}, \text{notification}, \emptyset\}$. Priorities and actions are ordered – $\emptyset \prec G3 \prec G2 \prec G1$ and $\emptyset \prec$ notification $\prec$ writ $\prec$ garnishment – and increase over time.

effects are larger in a specification that allows payment amounts to depend on treatment status (see Online Appendix OB).

[27]This is consistent with Alm et al. (1992)'s argument that for plausible degrees of risk aversion, narrow self interest with rational expectations fails to explain a significant share of observed compliance behavior.

[28]This does not affect individual incentives, since this data is not used to specify the taxpayer's individual rank, but rather to control for heterogeneity in our analysis of overall taxpayer behavior.

Each period t, a taxpayer i makes a payment with Poisson intensity $\kappa_{i,t}$. Let $s_{i,t} \in \{0,1\}$ denote whether or not the taxpayer makes a payment. We assume that conditional on making a payment, the share of taxes due repaid by this payment (referred to as normalized payment), $\pi_{i,t} \in [0,1]$, is drawn from a fixed distribution $f_{\pi,i}$ that depends only on the taxes owed by taxpayer i.[29] Let $T_i(t)$ denote the set of taxpayer i's payment times occurring strictly before t. We denote by $\Pi_{i,t} = \sum_{s \in T_i(t)} \pi_{i,s}$ the sum of normalized payments made up to period t. A taxpayer's total payments should rarely exceed 4 times quarterly taxes due.[30]

Let $X_{i,t}$ denote the vector of covariates

$$X_{i,t} = \begin{bmatrix} \mathbf{1}_{\Pi_{i,t}>0} \\ \Pi_{i,t} \\ \mathbf{1}_{g_{i,t}=g} \text{ for } g \in \{G1, G2, G3\} \\ \mathbf{1}_{a_{i,t}=a} \text{ for } a \in \{\text{garnishment}, \text{writ}, \text{notification}\} \\ \mathbf{1}_{g_{i,t}=G1 \text{ and } a_{i,t}=a} \text{ for } a \in \{\text{garnishment}, \text{writ}\} \\ \xi_i \end{bmatrix}.$$

We assume that in each period, taxpayer i makes a payment with Poisson intensity $\kappa_{i,t}$ taking the form

$$\kappa_{i,t}(\theta_i, \beta) = \max\{10^{-3}, \phi(\langle X_{i,t}, \beta\rangle + \theta_i) \times \mathbf{1}_{\Pi_{i,t}<4}\} \tag{6}$$

where $\langle \cdot, \cdot \rangle$ is the usual dot product, and ϕ is a non-decreasing S-shaped function, parameterized by $\varphi \in \mathbb{R}^2$, specified below. Conditional on type θ_i, the intensity of payment at time t depends only on payments made, the current priority group $g_{i,t}$, and the latest collection action taken $a_{i,t}$. We allow for interaction effects between priority G1 and collection actions. For instance, it may be that receiving a priority G1, and receiving a writ act as substitutes.

[29] Taxpayers are binned based on amount due, and given a bin, $\pi_{i,t}$ is drawn from the empirical distribution of normalized payments associated with that bin of taxpayers. Counterfactuals for different models of payment amounts conditional on payment events are provided in Online Appendix OB.

[30] For most taxpayers, the amount of taxes owed each quarter in a year is approximately the same.

Collection actions and priorities. Let us denote by $h_{i,t} = (\xi_i, a_{i,s}, g_{i,s}, \pi_{i,s})_{s \leq t}$ the public history of actions, priorities, and payments made, associated with taxpayer i at time t.

Assumption 1. *We assume that the distribution of priority assignments $g_{i,t}$ and collection actions $a_{i,t}$ are functions of public data $h_{i,t-1}$ alone, and independent of type θ_i.*

We denote by $G(\cdot|h_{i,t}) \in \Delta(\{G1, G2, G3, \emptyset\} \times \{\text{garnishment, writ, notification}, \emptyset\})$ the joint distribution of $g_{i,t}$ and $a_{i,t}$ conditional on public history $h_{i,t}$.

The assumption that priorities $g_{i,t}$ and actions $a_{i,t}$ are functions of public data alone is true by construction in the treatment arm: we assigned priorities and collection actions on the basis of data shared by the city government. In principle, collection actions taken by the government in the control arm could depend on signals of θ_i unavailable to us. We have no evidence that such signals play a role. Assumption 1 formally rules them out.

In the language of Engle et al. (1983), Assumption 1 guarantees that priorities $g_{i,t}$ and actions $a_{i,t}$ are weakly exogenous to parameters (φ, β, σ), so that we don't need to explicitly specify the data-generating process for priorities and actions in order to estimate (φ, β, σ).

Let $K_{i,t} \equiv 1 - \exp(-\kappa_{i,t})$ denote taxpayer i's period t payment probability. For any final history $h_{i,T}$, its likelihood can be factorized as follows.

$$\begin{aligned}\mathsf{prob}(h_{i,T}|\varphi, \beta, \sigma) = &\prod_{t=1}^{T} G(g_{i,t}, a_{i,t}|h_{i,t-1}) \times \prod_{t=1}^{T} f_{\pi,i}(\pi_{i,t})^{s_{i,t}} \\ &\times \underbrace{\int_{\theta \sim \mathcal{N}(0,\sigma)} \prod_{t=1}^{T} K_{i,t}(\theta, \beta)^{s_{i,t}} \times (1 - K_{i,t}(\theta, \beta))^{1-s_{i,t}}}_{\equiv \Psi(h_{i,T}|\varphi,\beta,\sigma)}.\end{aligned}$$

The first two factors do not depend on parameters of interest φ, β, σ. This implies that parameters φ, β, σ can be efficiently estimated using the conditional log-likelihood.

$$\mathcal{L}(h_T|\varphi, \beta, \sigma) \equiv \sum_{i \in I} \log(\Psi(h_{i,T}|\varphi, \beta, \sigma)). \tag{7}$$

In turn, $f_{\pi,i}$ can be estimated using conditional payment data in the event a payment is made, for taxpayers with tax due amounts similar to taxpayer i.

By construction, the estimation of parameters of interest (φ, β, σ) is not driven by taxpayers with large amounts of tax due. Instead, parameters of interest are estimated using only taxpayers' binary decisions to make a payment or not in any period. This allows us to form more robust estimates of treatment effects than those obtained from raw averages.

Implementation. Altogether, we seek to recover 14 parameters:

$$\begin{aligned}
&\beta_{\Pi_{i,t}>0}, \beta_{\Pi_{i,t}}, \\
&\beta_{G1}, \beta_{G2}, \beta_{G3}, \\
&\beta_{\text{garnishment}}, \beta_{\text{writ}}, \beta_{\text{notification}}, \\
&\beta_{\text{G1-garnishment}}, \beta_{\text{G1-writ}}, \\
&\beta_{\xi}, \sigma, \underline{\varphi}, \overline{\varphi}.
\end{aligned}$$

Given parameters $\varphi = (\underline{\varphi}, \overline{\varphi}) \in \mathbb{R}^2$, we specify function ϕ mapping covariates and persistent types to payment intensity $\kappa_{i,t}$ in (6) as

$$\phi(x) = \min\left\{\overline{\varphi} - \underline{\varphi}, \max\{x - \underline{\varphi}, 10^{-3}\}\right\}.$$

We aggregate payments at the weekly level. The log-likelihood of observed payment behavior (see (7)) induces a posterior distribution over parameters φ, β, σ computed using Markov Chain Monte Carlos (Chernozhukov and Hong, 2003).[31]

Our framework is essentially a fixed-effect Poisson model (Hausman et al., 1984). In-

[31] Our preferred specification imposes that the coefficient $\beta_{\text{notification}}$ associated with notifications be non-negative. This is an intuitive restriction: every collection process needs to start with a notification, so receiving a notification should increase perceived incentives to repay. Our data partially challenges this prior restriction: during the first 2 months of the experiment, taxpayers in the control group that receive a formal notification tend to make payments at a lower rate than taxpayers who have not received a notification. The pattern is not present in the treatment group, or in the control group during the second half of the experiment. We discuss the data, possible explanations (other than noise), and their implication for design in Online Appendix OB. Removing this restriction does not qualitatively change inferences.

tuitively, coefficients β are identified from variation in non-collinear covariates $X_{i,t}$ in the cross-section of taxpayers. The variance σ of types θ_i is identified from persistent variation in payment propensity across taxpayers with similar covariates. Coefficients $\underline{\varphi}$ and $\overline{\varphi}$ place upper and lower bounds on payment intensities. Although our model is surely misspecified, we note that likelihood-maximizing parameters achieve the best fit between the true and simulated processes under the Kullback-Leibler divergence (White, 1982).

6.2 Findings

Table 7 reports posterior means and standard deviations for parameters of interest. Robustness checks are provided in Online Appendix OB. To a first order, these coefficients can be directly interpreted as percentage point changes in the likelihood of a payment event in the next week.[32]

Consistent with the reduced-form evidence reported in Section 5, inclusion in group G1 increases payment intensity. The coefficient associated with priority G2 is much smaller. Although we highlighted to members of group G2 that they could be moved to group G1 at any time, this did not seem to effectively engage taxpayers' anticipation and planning, possibly because it informed taxpayers that they were a relatively lower collection priority. Inclusion in group G3 reduced payment intensities compared to control, possibly because informing G3 members of potential promotion to group G2 reduced their expectations of prompt collection.

Estimated settlement probability ξ_i is predictive of actual settlement behavior, confirming reduced-form findings reported in Section 5. Our estimate of coefficient σ suggests meaningfully large amounts of unobserved heterogeneity in types. Coefficients on payment variables suggest that taxpayers who have made some payments are subsequently more likely to make further payments, but less so if the cumulated share of taxes due paid is larger.

[32]This is because weekly payment intensities are small, so that $K_{i,t} \equiv 1 - \exp(-\kappa_{i,t}) \simeq \kappa_{i,t}$, and because mapping ϕ has a slope of 1, whenever it is not constant.

	Mean	(Std. Dev.)
β_{G1}	$3.62 \cdot 10^{-2}$	$(3.33 \cdot 10^{-3})$
β_{G2}	$0.43 \cdot 10^{-2}$	$(2.93 \cdot 10^{-3})$
β_{G3}	$-0.65 \cdot 10^{-2}$	$(1.33 \cdot 10^{-3})$
$\beta_{\text{garnishment}}$	$2.20 \cdot 10^{-2}$	$(4.67 \cdot 10^{-3})$
β_{writ}	$3.27 \cdot 10^{-2}$	$(2.34 \cdot 10^{-3})$
$\beta_{\text{notification}}$	$0.01 \cdot 10^{-2}$	$(0.14 \cdot 10^{-3})$
$\beta_{\text{G1-garnishment}}$	$-2.61 \cdot 10^{-2}$	$(6.86 \cdot 10^{-3})$
$\beta_{\text{G1-writ}}$	$-1.16 \cdot 10^{-2}$	$(5.91 \cdot 10^{-3})$
$\beta_{\Pi_{i,t}>0}$	$2.86 \cdot 10^{-2}$	$(2.82 \cdot 10^{-3})$
$\beta_{\Pi_{i,t}}$	$-3.50 \cdot 10^{-2}$	$(1.30 \cdot 10^{-3})$
β_{ξ}	$1.29 \cdot 10^{-1}$	$(4.21 \cdot 10^{-3})$
σ	$0.50 \cdot 10^{-1}$	$(0.23 \cdot 10^{-2})$
$\underline{\varphi}$	$1.29 \cdot 10^{-2}$	$(2.03 \cdot 10^{-3})$
$\overline{\varphi}$	$3.44 \cdot 10^{-1}$	$(0.90 \cdot 10^{-1})$

Table 7: Estimated parameters for our model of taxpayer payment choice.

Note: The first (second) column reports the mean (standard deviation) of parameter estimates from the MCMC procedure described in Section 6.1.

Issuing formal writs has a large impact on settlement, comparable to that of being included in priority group G1.[33] The effect of writs is somewhat smaller when associated with a G1 priority, but G1 priorities and writs are only weak substitutes. The effectiveness of writs within the control group was a surprise to us given that writ enforcement rates in the control group are low (Table 2).

6.3 Counterfactuals

We evaluate counterfactual policies under the following assumption.

[33] The coefficients associated with each collection action should not be added to get the cumulative impact of collection actions. The coefficient associated with each action summarizes the aggregate effect of the current action and preceding required collection steps.

Assumption 2 (valid extrapolation). *Provided that collection promises are kept, changing the process for priorities $g_{i,t}$ and actions $a_{i,t}$ does not affect the settlement behavior of taxpayers given priorities and collection actions.*

Under Assumption 2, estimated parameters β, σ, φ let us evaluate counterfactual mechanisms assigning actions and priorities as a function of public histories, provided that collection promises continue to be kept under the counterfactual. In the language of Engle et al. (1983), priority assignment and collection action processes are super-exogenous to settlement behavior over the restricted class of mechanisms that keep collection promises.

Relevant counterfactuals. We are interested in the following counterfactuals whose outcomes are summarized in Table 8.[34]

CF1. *Experiment as implemented.* We first replicate collection policies from our experiment. Taxpayers are ranked using the same score, and 3450 writs are issued under our simulated control, vs. 1101 under simulated treatment. We evaluate the experiment *as it was implemented*, which means that garnishments are systematically implemented on non-compliant members of group G1, but with an average implementation time of 9 weeks rather than 6 weeks.[35] In addition we replicate the increase of group G1 from 400 to 600 halfway into the experiment. We find that treatment increases expected collection by approximately 7.5%, at the cost of 558 expected garnishments.

Online Appendix OB further documents the fit between simulated and realized outcomes.

[34]In each case we sample 100 parameters from the posterior distribution of parameters obtained from MCMC. For each parameter we perform 400 simulations of counterfactual treatment outcomes on the population of treated taxpayers. Note that for the control simulation, we use 3450 writs, rather than 3620 as in Table 4. Any taxpayer who receives a writ but is already under garnishment for a previous quarter's debt is recorded as immediately in garnishment upon receiving the new writ. Since we aggregate to the weekly level for estimation, such taxpayers appear to move directly from a notification to garnishment, and we match this in our simulation.

[35]To get 9 weeks, we add the average delay from promise to garnishment in Q1, where we can observe almost all promised garnishments being fulfilled eventually, to the average promised deadline over the entire sample.

CF2. *Experiment as intended.* Second, we simulate a version of treatment as it was intended: the payment deadline of 6 weeks is strictly enforced. Because this stricter deadline naturally induces more garnishments, we scale down the size of group G1 until the number of garnishments issued roughly matches the number of garnishments issued in the actual experiment.

We find that in this scenario, treatment increases expected collection by 1.8%. This reduction in treatment effect, compared to treatment as implemented illustrates the importance of setting deadlines correctly as highlighted in Section 2.3.

CF3. *8 weeks deadlines.* The increased delay in the experiment as implemented raises the question of whether group G1 would be less effective if taxpayers had anticipated a 3 week delay in the implementation of garnishments. For this reason we evaluate a version of treatment in which the garnishment deadline is set at 8 weeks (rather than 9), and the size of group G1 is fixed at 400 throughout the experiment. We maintain the assumption that control issues approximately three times as many writs as treatment.

We estimate that under this scenario, treatment would increase expected revenue by 5.9%, at the cost of issuing an average of 552 garnishments.

CF4. *Matching writs used in control.* In this counterfactual, we replicate the design of counterfactual CF3, but issue writs in a manner similar to control. In this case, we find that PIE improves collection by 10.8% over treatment. On average, this collection scheme issues 1523 G1 priorities, 3450 writs, and 547 garnishments.[36]

Online Appendix OC shows that similar effects can be obtained if writs are issued early, and only to taxpayers that will be issued G1 priorities in the future. In that case roughly 1500 writs are issued, all of which are credible.

[36]Notifications are only sent to taxpayers that are assigned a G1 priority or receive a writ. It turns out that matching writs issued under control also induces our counterfactual to match notifications issued under control: on average, 4549 notifications are issued under counterfactual CF4, versus 4314 under the actual control. We note that changing the number of notifications has little impact on simulated treatment effects since the estimated impact of notifications on repayment propensities is very small (see Table 7).

CF5. *Rank by Tax Due.* We replicate counterfactual CF4, but rank taxpayers according to their taxes due. Under this simpler scoring rule, treatment increases revenue by 12.3% over control, and issues 595 garnishments on average. The increase in garnishments is not surprising since taxes dues is a less discriminating predictor of repayment behavior than the scores used in the experiment.

CF6. *Adopted policy.* Finally we simulate the effect of the collection policy adopted by the municipality: it uses tax due alone as a rank; it issues only 2 rounds of G1 priorities, one at the beginning of the collection period, and one seven weeks later, each with a deadline of seven weeks. Unlike other PIE designs G1 priorities are not reissued each week. Instead, G1 priorities are recycled twice. This policy combination increases revenue by 11.3% compared to control and issues 642 garnishments on average.

Counterfactual Policy	% Change in Revenue Mean Effect	(95% CI)	#G1	#Writs	#Garnished
CF1. Experiment as implemented	7.5	(3.9, 10.5)	1862	1101	558
CF2. Experiment as intended	1.8	(-1.1, 4.6)	1101	699	536
CF3. 8 weeks deadline	5.9	(2.5, 9.0)	1499	911	552
CF4. CF3 + Matching writs in control	10.8	(7.3, 13.9)	1523	3450	547
CF5. CF4 + Rank by taxes due	12.3	(8.6, 15.5)	1451	3450	595
CF6. Adopted policy	11.3	(7.8, 14.5)	1000	3450	642

Table 8: Evaluation of counterfactual policies. In our actual experiment, treatment issued 1838 G1 priorities, 1306 writs, and 533 garnishments.

Online Appendix OC reports additional counterfactuals. We note that our implementation of PIE is a policy bundle: members of group G1 are also issued a writ as promptly as feasible. We disentangle the impact of PIE that can be attributed to an earlier use of writs, and the impact that can be attributed to priority G1 alone. We find that roughly 2/3 of the effect can be attributed to priority G1. We also show that the distribution of taxes due plays a key role in mediating the effectiveness of PIE, refining our understanding of the effectiveness of PIE in different environments.

7 Discussion

Prioritized Iterative Enforcement seeks to improve the way a fixed enforcement capacity is deployed. The approach is well suited to environments in which agents are boundedly rational or have private information. Importantly, similar divide-and-conquer strategies can be applied to settings other than tax-collection in which policy enforcement is hindered by limited capacity (e.g. see Braga et al., 2001, for an application to homicide prevention).

Our field experience identifies delay in best-response as a key friction affecting the effectiveness of PIE. Under a simple baseline model setting short deadlines and keeping threat groups small is a dominant design. This is no longer the case when repayment takes time. Longer deadlines and threat groups that are larger than available collection capacity may be optimal. Our data identifies a policy combination that improves over the benchmark collection process and also lets us simulate out counterfactual designs. Plausible designs yield increases in revenue between 7% and 12%. Importantly, we are able to evaluate trade-offs between simplicity of design and revenue. We show that the simplified version of PIE ultimately adopted by the municipality yields the majority of feasible collection benefits.

Building on those insights, Appendix B provides a brief checklist of key steps required to adopt PIE.

Limits. Our experiment has important limitations. We fail SUTVA: treatment is less labor intensive, yielding a reallocation of capacity to control. In addition, although our sample of taxpayers is large, we implement PIE in a single location in partnership with an unusually collaborative municipality.

Some aspects of our design also failed to have the effect we hoped for. Priority group G2 did not increase settlement, suggesting that our letters did not effectively engage the higher order strategic thinking of taxpayers. Worse, priority G3 may have reduced taxpayers' propensity to repay. Design variants that communicate incentive information more clearly or differently may be able to address these issues.

Our semi-structural model also has limits: it only permits counterfactuals in which the nature of enforcement promises is not changed in significant ways, and does not elucidate taxpayers' decision-making. The fact that weakly enforced writs are effective even on a control population that is familiar with them potentially indicates a persistent failure of learning. At the same time, the fact that taxpayers assigned to priority group G3 settle at a lower rate than similar taxpayers in the control group suggests that taxpayers respond to information about the likelihood of enforcement. It is plausible to us that effective policy design should seek to strategically slowdown learning. For instance, one may consider disclosing less information to low-priority taxpayers as well as keeping garnishment rates low but bounded away from zero.

External validity. The value of PIE as a collection strategy may depend non-monotonically on government capacity. Because it requires commitment power, as well as some organizational competence (especially during initial implementations) PIE may not be suitable in contexts with very low capacity. At the other end of the spectrum, efficiency gains from using PIE may vanish in high capacity environments. To that extent, Jesús María, a relatively high capacity district in a middle income country, may be representative of locations where PIE has a significant impact.

On the positive side, external validity is improved by the fact that we operate only with the human resources available to the municipality. In addition, our semi-structural model allows the evaluation of PIE under counterfactual treatments and counterfactual populations. For instance, Online Appendix OC clarifies the sensitivity of PIE to the distribution of taxes due. We show that PIE yields the highest benefits in environments with moderate inequality in taxes due.[37]

Appendix

A Commitment power

In our experimental setting we maximize the government's commitment power by making collection threats with clearly specified implementation dates, set not too far in time. This allows the government to better leverage its limited reputational capital by making failures to deliver on threats more detectable. In contrast, collection promises with distant deadlines are likely to be forgotten, or made irrelevant by policy and government changes.

This argument can be formalized as follows. Let us denote by V_{failure} and $V_{\text{no failure}}$ the value of the government's reputation vis à vis a taxpayer, depending on whether or not the government fails to deliver on a promise to collect. This value may reflect government officials' value for their public image, their reputation for being effective, as well as the ongoing benefits of inducing trust in public messaging. Let p denote the probability that a failure would be detected, and c the taxpayer's perception of the government's opportunity cost of delivering on a promise.

The government's expected value if it chooses not to deliver on a promise is $pV_{\text{failure}} + (1 - p)V_{\text{no failure}}$, depending on whether or not failure to deliver is observed. If the government delivers on a promise, its value is $V_{\text{no failure}}$. Hence, a promise is credible if and only if

$$p\left(V_{\text{no failure}} - V_{\text{failure}}\right) \geq c. \tag{A.1}$$

B Adoption steps

Our experimental design choices were made to permit measurement rather than maximize operational effectiveness. Such choices include randomizing over two scoring rules, and including a representative sample of lower ranked taxpayers in the first G1 priority group. As a result, tax collection authorities interested in implementing PIE should be able to adopt designs that are both simpler and more effective than the one used in our experiment. This

[37] Under arbitrarily high inequality, capacity is effectively not binding, since a few taxpayers owe the bulk of taxes. Under moderate inequality, additional taxpayers encouraged to repay by PIE owe more taxes than lower ranked taxpayers not encouraged by PIE.

appendix clarifies the steps required to adopt PIE, and provides some guidance on how to approach design choices. A caveat is that common sense adjustments, motivated by a precise understanding of context, likely trump the general guidelines we provide.

Pick a suitable order. The first step is to pick a suitable order for collection. In our experience, targeting taxpayers in order of taxes due is simple and effective. It also has the merit of being a legitimate sorting variable, and can be simply explained to stakeholders. It also increases the chances that PIE improves the progressivity of tax collection.

If different groups have very different repayment behavior, require markedly different procedures, or if equity concerns demand that different groups be collected on at similar rates, it may be reasonable to apply PIE within each group rather than over the aggregate population.

Pick a collection deadline. The repayment deadline given to prioritized taxpayers should be long enough to given solvent taxpayers enough time to repay, but not so long that payments are highly delayed. This choice must be informed by experience (getting inputs from local tax collectors is essential), or data (surveys of taxpayers may be helpful, past repayment delays may also provide a rough estimate). An adaptive approach may be practical: if many tax-payers fail to repay by the deadline, extend the deadline given to the next batch of high priority taxpayers; if instead the deadline does not appear to be binding, shorten the deadline given to the next batch of taxpayers.

Calibrate capacity and threat flow. A core component of PIE is to limit the number of threats issued at a given point in time so that available collection capacity is sufficient to implement promised collection actions. Practically, the first step is to estimate collection capacity. The second is to set a target for threat flows equal to a multiple of estimated flow collection capacity.

In our application we issue roughly twice as many threats as available capacity, which ends up being a conservative choice, leading us to increase threats to three times available capacity in the second half of the experiment. This suggests a prudent but adaptive approach: start with a multiplier that is slightly on the conservative side, and increase it if a significant share of available collection capacity is unused.

Draft prioritized collection letters. Our findings provide only limited guidance on how to formulate collection letters associated with different priority groups. While priority G1 was effective, priority G2 was not, and priority G3 seemed to reduce settlement propensities. We suspect that letters associated with G2 and G3 priorities made excessively clear to taxpayers that they were assigned lower collection priorities.

A practical way to proceed would be to maintain priority G1, eliminate priority G2, and send a generic letter to taxpayers that have not been assigned priority G1. Such a letter should mention that taxes are due, and what the penalties for non compliance are, but not emphasize to taxpayers that they are a lower priority group. In addition, we suggest enforcing collection for some share of lower priority taxpayers in order to maintain the norm that taxes should be repaid.

Define a threat issuance process. In our experiment, we issued new threats on a weekly basis, as soon as collection capacity tied up with past threats was freed up, either by payments or by the execution of garnishments. This allows for a just-in-time use of threat capacity, but is somewhat administratively demanding. An alternative is to schedule a few threat issuance dates at regular intervals throughout the collection period. The policy ultimately adopted by Jesús María issued two rounds of threats, which simplified the administration of PIE. The cost is that less threats can be issued in aggregate.

Other considerations. Other aspects of PIE would likely require adaptation depending on specific circumstances.

A surprising finding from our experiment is that collection steps that are cheap and weakly associated with enforcement can still have a large impact on settlement propensities. In our case, such actions took the form of formal writs, but different collection actions may be relevant in other contexts. Although our experiment does not elucidate the mechanism through which cheap but salient collection actions operate, we suggest to continue their use in field implementations of PIE.

We implement PIE in the context of property taxes where the amount of taxes due is known. We provide a theoretical extension to income taxes in Online Appendix OD. Although the analysis is similar, it involves a key additional challenge which is to set the price offer at which taxpayers would be allowed to settle their taxes. More experimental evidence is needed to make informed suggestions about how to make this choice. One speculative possibility is to compute a conservative estimate of taxes due, say by running a quantile regression of past audit outcomes on observables, for a relatively low quantile (say between 25% and 40%), and use that as a settlement price offer.

C Proofs

Proof of Proposition 1. Consider first the case of random uniform enforcement. Assume that all agents settle. Then a deviator who refuses to settle faces enforcement with probability 1. Since $P < D$ and $P > \alpha D$, it is indeed individually optimal for a taxpayer to settle. Assume now that all agents refuse to settle. Then in equilibrium, an agent faces enforcement with probability α, yielding expected payoff $-\alpha D$. Settling yields payoff $-D$. Since $\alpha < 1$, it is individually optimal for an agent not to settle.

Consider now prioritized static enforcement. We show that it is iteratively dominant for all agents to settle, so that the principal raises tax revenue NP. The proof is by induction on the priority of agents. The induction hypothesis is that in all strategy profiles that survive k-iterations of elimination of dominated strategies, all agents with priority higher than k choose to settle. The induction hypothesis holds for $k = 1$ since the highest priority agent faces collection with probability 1 in the event they do not settle. In turn, if the hypothesis holds for $k \geq 1$, then an agent of rank $k + 1$ that does not comply is audited with probability 1. Hence, it is iteratively dominant for an agent of rank $k + 1$ to comply, which establishes the induction step. ■

Proof of Proposition 2. We begin by observing that it is never optimal to initiate coercive collection $a_i = 1$ after a taxpayer sends message $m_i = 1$ (they are able to pay), and follows the mechanism designer's settlement recommendation ($s_i = \widehat{s}_i$). Indeed, if a mechanism implements collection after such an event, then the alternative mechanism that demands settlement of the full amount due D_i instead of implementing coercive collection collects a weakly greater tax-revenue (strictly greater if $\rho < 1$) and satisfies the same IC constraints. For this reason coercive collection collects zero revenue under an optimal mechanism, and it is without loss of generality to ignore revenue from coercive collection, and focus on the case where $\rho = 0$.

Under a truthful and obedient equilibrium, conditional on submitting a message $m_i = 1$, the expected utility of taxpayer i is bounded above by $-\mathbb{E}[P_i \widehat{s}_i | m_i = 1]$. Since a taxpayer can always choose to submit messages $m_i = 0$ and take settlement decision $s_i = 0$, it follows from incentive compatibility that for any taxpayer i,

$$-\mathbb{E}[P_i \widehat{s}_i | m_i = 1] \geq -\mathbb{E}[a_i D_i | m_i = 0]. \tag{C.1}$$

Because of capacity constraints, it must be that $\sum_{i=1}^{N} a_i \lambda_i \leq \alpha N$. This implies that

$$\sum_{i=1}^{N} q_i \lambda_i \mathbb{E}[a_i | m_i = 0] \leq \alpha N. \tag{C.2}$$

Together (C.1) and (C.2) imply that

$$\sum_{i=1}^{N} q_i \lambda_i \mathbb{E}\left[\frac{P_i}{D_i}\widehat{s}_i \,\Big|\, m_i = 1\right] \leq \alpha N. \tag{C.3}$$

In turn total expected revenue is equal to $\sum_{i=1}^{N}(1-q_i)\mathbb{E}[P_i \widehat{s}_i | m_i = 1]$. Let $\delta_i \equiv \mathbb{E}\left[\frac{P_i}{D_i}\widehat{s}_i | m_i = 1\right] \in [0,1]$. In equilibrium, expected collection is equal to $\sum_{i=1}^{N} \delta_i (1-q_i) D_i$. Condition (C.3) implies that weights $(\delta_i)_{i \in \{1,\cdots,N\}}$ satisfy

$$\sum_{i=1}^{N} \delta_i q_i \lambda_i \leq \alpha N.$$

This concludes the proof. ■

Proof of Proposition 3. We begin with point (i). For any $i \in \{1, \cdots, N\}$, define

$$A(i) \equiv \frac{1}{N}\sum_{j=1}^{i-1} q_j \lambda_j, \quad \text{and} \quad \widehat{A}(i) \equiv \frac{1}{N}\sum_{j=1}^{i-1} \gamma_j \lambda_j$$

where $(\gamma_n)_{n \in \{1,\cdots,N\}}$ is a sequence of independent Bernoulli random variables with parameters $(q_n)_{n \in \{1,\cdots,N\}}$.

Take $\epsilon > 0$ as given. Concentration inequalities for martingales (the Azuma-Hoeffding theorem) imply that,

$$\mathsf{prob}\left(\max_{n \in \{1,\cdots,N\}} |A(n) - \widehat{A}(n)| < \epsilon\right) \to_N 1$$

uniformly over sequences $(\lambda_n, q_n)_{n \in \{1,\cdots,N\}} \in \left([\underline{\lambda}, \overline{\lambda}] \times [\underline{q}, \overline{q}]\right)^N$.

Consider any taxpayer with rank i such that $A(i) \geq \alpha + \epsilon$. Since taxpayer i settles with probability less than $1 - q_i$ in any equilibrium, the capacity used to investigate taxpayers with rank $j < i$ stochastically dominates $\widehat{A}(i)$. Hence, the probability that taxpayer i gets investigated approaches 0 for N large. This implies that whenever $P_{i,0} > 0$, it is dominant for taxpayer i not to repay taxes.

We turn to point (ii). The proof proceeds by iterating over groups of taxpayers. We begin by defining a sequence of thresholds for taxpayer ranks. We define

$$B(i) \equiv \frac{1}{N} \sum_{j=1}^{i-1} \lambda_j$$

and for any strictly increasing function $f : \{1, \cdots, N\} \to [0, 1]$,

$$\forall x \in [0,1], \quad f^{-1}(x) \equiv \max\{n \in \{1, \cdots, N\},\ s.t.\ f(n) \leq x\}.$$

The following properties hold: f^{-1} is increasing and for all $n \in \{1, \cdots, N\}$, and $x \in [0, 1]$,

$$f^{-1}(f(n)) = n, \quad f(f^{-1}(x)) \leq x \quad \text{and} \quad f(f^{-1}(x) + 1) > x.$$

We define the sequence $(n_k)_{k \in \mathbb{N}}$ by

$$\begin{aligned} n_0 &\equiv B^{-1}(\alpha - \epsilon) \\ n_k &\equiv B^{-1}(B(n_{k-1}) + [\alpha - A(n_{k-1}) - \epsilon]^+). \end{aligned}$$

where for all $\Delta \in \mathbb{R}$, $[\Delta]^+ = \max(0, \Delta)$. By construction, $(n_k)_{k \in \mathbb{N}}$ is weakly increasing, and bounded above by N. In addition if $n_k = n_{k-1}$, then $n_{k+1} = n_k$. We show that for K and N large enough, uniformly over $(\lambda_n, q_n)_{n \in \{1, \cdots, N\}} \in \left([\underline{\lambda}, \overline{\lambda}] \times [\underline{q}, \overline{q}]\right)^N$,

$$A(n_K) \leq \alpha - \epsilon \quad \text{and} \quad A(n_K) \geq \alpha - 2\epsilon.$$

We first prove by induction that for all k, $A(n_k) \leq \alpha - \epsilon$. This is true for n_0 since $A(n_0) \leq B(n_0) \leq \alpha - \epsilon$. Assume this is true for n_k. By construction

$$B(n_{k+1}) \leq B(n_k) + \alpha - A(n_k) - \epsilon.$$

Since $A(n_{k+1}) - A(n_k) \leq B(n_{k+1}) - B(n_k)$, it follows that

$$A(n_{k+1}) - A(n_k) \leq \alpha - A(n_k) - \epsilon \Rightarrow A(n_{k+1}) \leq \alpha - \epsilon.$$

We now prove that $A(n_K) \geq \alpha - 2\epsilon$. Since $\sum_{k \leq K} B(n_{k+1}) - B(n_k) \leq 1$, and $B(n_k)$ is increasing in k, there exists $k \leq K$ such that $B(n_{k+1}) - B(n_k) \leq \frac{1}{K}$. In addition, by construction

$$
\begin{aligned}
B(n_{k+1}+1) &\geq B(n_k) + \alpha - A(n_k) - \epsilon \\
B(n_{k+1}+1) &\leq B(n_{k+1}) + \frac{\overline{\lambda}}{N}.
\end{aligned}
$$

This implies that

$$
\begin{aligned}
&\alpha - A(n_k) - \epsilon \leq \frac{\overline{\lambda}}{N} + B(n_{k+1}) - B(n_k) \leq \frac{\overline{\lambda}}{N} + \frac{1}{K} \\
\Rightarrow &A(n_k) \geq \alpha - \epsilon - \frac{\overline{\lambda}}{N} - \frac{1}{K}.
\end{aligned}
$$

This implies that for N and K large enough, $A(n_K) \geq \alpha - 2\epsilon$. We now take K fixed, and let N grow arbitrarily large. Since K is fixed, and values $(n_k)_{k\in\mathbb{N}}$ are deterministic, it follows that for N large enough, with probability approaching 1,

$$
\forall k \leq K, \quad |\widehat{A}(n_k) - A(n_k)| \leq \epsilon. \tag{C.4}
$$

We condition on the event that (C.4) holds, and iteratively consider batches of taxpayers $\{0, \cdots, n_0\}, \cdots, \{n_{k-1}+1, \cdots, n_k\}, \cdots, \{n_{K-1}+1, \cdots, n_K\}$ at times $k/(K+1)$. We show that by time $(k+1)/K$, it is obviously dominant for taxpayers $\{n_k+1, \cdots, n_{k+1}\}$ to settle their taxes if they are able to, and have not done so already. Indeed, since $B(n_0) \leq \alpha - \epsilon$ and since settlement offers $P_{i,t}$ are strictly increasing in time t, for N large enough it is obviously dominant for taxpayers in group $0, \cdots, n_0$ to settle before time $1/(K+1)$.

Assume that by time $k/(K+1)$ all groups $\{n_{k'-1}+1, \cdots, n_{k'}\}$ with $k' \leq k$ have settled if they can. This means that taxpayers in group $\{n_k+1, \cdots, n_{k+1}\}$ are informed that the collection capacity expended on taxpayers with rank less than n_k is $\widehat{A}(n_k) \leq A(n_k) + \epsilon$. In the worst-case scenario where none of the taxpayers in group $\{n_k+1, \cdots, n_{k+1}\}$ settle, the capacity needed to collect on all of them is $B(n_{k+1}) - B(n_k)$. In turn, the capacity available for collection on taxpayers with rank higher than n_k is $\alpha - A(n_k)$. By construction

$$
B(n_{k+1}) - B(n_k) \leq \alpha - A(n_k) - \epsilon \leq \alpha - \widehat{A}(n_k).
$$

Hence, by time $(k+1)/(K+1)$ taxpayers in group $\{n_k+1, \cdots, n_{k+1}\}$ all know that they will be investigated with certainty if they don't settle. Since settlement offers $P_{i,t}$ increase strictly over time, it is obviously dominant for them to settle. This implies that with probability

1 as N gets large, in any non-obviously dominated strategy profile, taxpayers with rank n such that $A(n) \leq \alpha - 2\epsilon$ all settle their taxes. Point (ii) follows by taking ϵ small enough.

The proof of point (iii) follows from points (i) and (ii), as well as the fact that the solution to (3) takes the form $\delta_i = 1$ for all $i < i^*$ and $\delta_i = 0$ for all $i > i^*$, with i^* such that $\sum_{i<i^*} q_i\lambda_i \leq \alpha N$ and $\sum_{i\leq i^*} q_i\lambda_i \geq \alpha N$. ■

D Experimental Materials

Figure D.1 illustrates the scheduling of collection actions satisfying legal constraints, allowing to achieve tight processing deadlines.

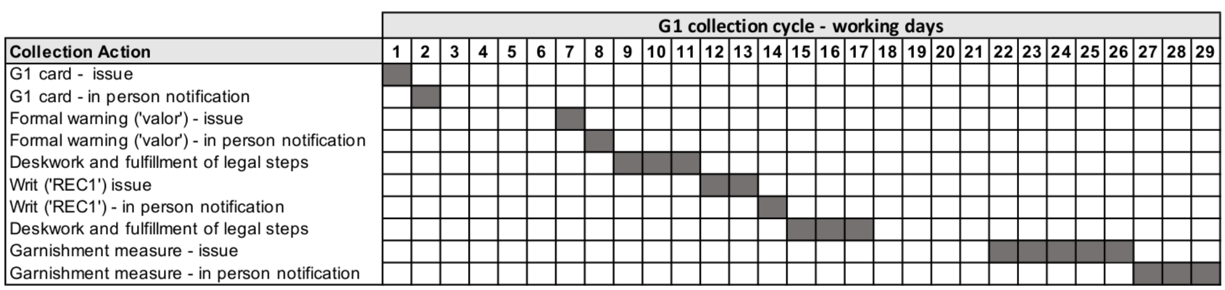


Figure D.1: Schedule of collection actions for G1 taxpayers

Note: A gray box indicates the number of days since assignment to G1 before the collection action can be taken.

Tables D.1 and D.2 illustrate the information letters sent to taxpayers in priority groups G2 and G3. Spanish original of all letters sent are reported in Online Appendix OF.

NOTICE OF IMMINENT COLLECTION

We remind you that you have the following debt outstanding with the municipality:	**Amount**
The coercive collection process will start at the latest on:	**Today + 12 weeks**
and you can be promoted at any time and without prior warning to the top priority group (which will imply the start of the coercive collection in maximum 6 weeks).	
If the coercive collection process is started your debt will include the penalties and administrative expenses regulated by law and will amount to:	**Amount * .1 + US$35**
In addition to accruing a weekly interest of:	**Interest**
We remind you that it is on your own interest to pay immediately to avoid higher expenses. You can use any of our payment options listed below.	

Table D.1: Information letter, priority group G2

Note: Text is translated from original spanish to english. Information letter is sent to new members of G2.

NOTICE OF DEBT OUTSTANDING

We remind you that you have the following debt outstanding with the municipality:	**Amount**
and that you can be promoted at any time and without prior warning to the high priority group (which will imply the start of the coercive collection process in maximum 12 weeks).	
If the coercive collection process is started your debt will include the penalties and administrative expenses regulated by law and will amount to:	**Amount * .1 +US$35**
In addition to accruing a weekly interest of:	**Interest**
We remind you that it is on your own interest to pay immediately to avoid higher expenses. You can use any of our payment options listed below.	

Table D.2: Information letter, priority group G3

Note: Text is translated from original spanish to english. Information letter is sent to new members of G3.

E Balance Check

Table E.1 reports summary statistics and balance checks for our initial assignment.

	Control	Treatment	Difference		
			Mean	S.E.	p-value
Q1 total tax due	374.5	377.5	-3.00	18.33	0.87
Q1 property tax due	138.1	129.6	8.49	9.09	0.35
Q1 user charges due	236.4	247.9	-11.49	13.23	0.39
Total tax due (quarterly)	402.1	402.0	0.09	21.3	0.99
Property tax due (quarterly)	150.8	141.9	8.98	11.86	0.45
User charges due (quarterly)	251.2	260.1	-8.88	14.07	0.63
Exo. score	459.5	460.0	-0.65	28.51	0.98
Endo. score	545.0	555.2	-10.18	39.94	0.80
Last year repayment share	0.498	0.515	-0.02	0.007	0.02
Is Pricos	0.020	0.020	.0002	0.002	0.93
Has Employer	0.448	0.444	0.003	0.009	0.69
Has Education	0.199	0.205	-0.007	0.007	0.32
Has Email	0.652	0.653	-.0002	0.008	0.98
Has Cellular	0.792	0.788	0.003	0.007	0.63
Salary	2,862.50	2,900.13	-37.63	62.26	0.54
Age	58.0	57.8	0.23	0.22	0.28
Male	0.49	0.49	0.005	0.009	0.6
Lives in district	0.9	0.9	-0.007	0.005	0.13
Num Observations	6728	6704			

Notes: *Q1 total tax, Q1 property tax* and *Q1 user charges* are debts for Q1-2021 as of April 5, 2021 (in Peruvian S/). *Total tax due (quarterly), Property tax due (quarterly)* and *user charges due (quarterly)* are debts for Q1 plus amounts due for Q2-Q4, divided by 4, as of April 5, 2021 (in Peruvian S/). *Last year repayment share* is the share of tax debt in the previous year that was paid within three months of the deadline. *Exo. and Endo. scores* are the scoring rules estimated from data (*Exo.* includes only exogenous covariates; *Endo* includes also *Last year repayment share*). *Is Pricos* is an indicator used by the tax administration for the 500 top tax amounts owed. *Salary* in Peruvian S/.

Table E.1: Summary statistics by treatment status

We note that last year's repayment share and age are imperfectly balanced. Online

Appendix OB estimates specifications that control explicitly for unbalanced covariates. Parameters of interest are unchanged.

References

Abreu, D. and H. Matsushima (1992): "Virtual implementation in iteratively undominated strategies: complete information," *Econometrica: Journal of the Econometric Society*, 993–1008.

Alm, J. (2019): "What motivates tax compliance?" *Journal of economic surveys*, 33, 353–388.

Alm, J., G. H. McClelland, and W. D. Schulze (1992): "Why do people pay taxes?" *Journal of public Economics*, 48, 21–38.

Balan, P., A. Bergeron, G. Tourek, and J. L. Weigel (2022): "Local Elites as State Capacity: How City Chiefs Use Local Information to Increase Tax Compliance in the Democratic Republic of the Congo," *American Economic Review*, 112, 762–97.

Banerjee, A. V., S. Chassang, S. Montero, and E. Snowberg (2020): "A theory of experimenters: Robustness, randomization, and balance," *American Economic Review*, 110, 1206–30.

Bassetto, M. and C. Phelan (2008): "Tax riots," *The Review of Economic Studies*, 75, 649–669.

Bergeron, A., G. Tourek, and J. Weigel (2021): "The state capacity ceiling on tax rates: Evidence from randomized tax abatements in the DRC," .

Bergolo, M., R. Ceni, G. Cruces, M. Giaccobasso, and R. Perez-Truglia (2023): "Tax audits as scarecrows: Evidence from a large-scale field experiment," *American Economic Journal: Economic Policy*, 15, 110–153.

Braga, A. A., D. M. Kennedy, E. J. Waring, and A. M. Piehl (2001): "Problem-oriented policing, deterrence, and youth violence: An evaluation of Boston's Operation Ceasefire," *Journal of research in crime and delinquency*, 38, 195–225.

Castro, L. and C. Scartascini (2015): "Tax compliance and enforcement in the pampas evidence from a field experiment," *Journal of Economic Behavior & Organization*, 116, 65–82.

Chernozhukov, V. and H. Hong (2003): "An MCMC approach to classical estimation," *Journal of Econometrics*, 115, 293–346.

Chirico, M., R. P. Inman, C. Loeffler, J. MacDonald, and H. Sieg (2016): "An experimental evaluation of notification strategies to increase property tax compliance: free-riding in the city of brotherly love," *Tax Policy and the Economy*, 30, 129–161.

Dal Bó, E. (2007): "Bribing voters," *American Journal of Political Science*, 51, 789–803.

De Neve, J.-E., C. Imbert, J. Spinnewijn, T. Tsankova, and M. Luts (2021): "How to improve tax compliance? Evidence from population-wide experiments in Belgium," *Journal of Political Economy*, 129, 1425–1463.

Del Carpio, L. (2014): "Are the neighbors cheating? Evidence from a social norm experiment on property taxes in Peru," *Unpublished Manuscript, Princeton University.*

Drago, F., F. Mengel, and C. Traxler (2020): "Compliance behavior in networks: Evidence from a field experiment," *American Economic Journal: Applied Economics*, 12, 96–133.

Dwenger, N., H. Kleven, I. Rasul, and J. Rincke (2016): "Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany," *American Economic Journal: Economic Policy*, 8, 203–32.

Eeckhout, J., N. Persico, and P. E. Todd (2010): "A theory of optimal random crackdowns," *American Economic Review*, 100, 1104–35.

Eliaz, K. and R. Spiegler (2015): "X-games," *Games and Economic Behavior*, 89, 93–100.

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983): "Exogeneity," *Econometrica: Journal of the Econometric Society*, 277–304.

Fellner, G., R. Sausgruber, and C. Traxler (2013): "Testing enforcement strategies in the field: Threat, moral appeal and social information," *Journal of the European Economic Association*, 11, 634–660.

Halac, M., I. Kremer, and E. Winter (2020): "Raising capital from heterogeneous investors," *American Economic Review*.

Halac, M., E. Lipnowski, and D. Rappoport (2021): "Rank Uncertainty in Organizations," *Available at SSRN 3553935*.

Hallsworth, M. (2014): "The use of field experiments to increase tax compliance," *Oxford Review of Economic Policy*, 30, 658–679.

Hausman, J., B. H. Hall, Z. Griliches, et al. (1984): "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica*, 52, 909–938.

Kagel, J. H., R. M. Harstad, and D. Levin (1987): "Information impact and allocation rules in auctions with affiliated private values: A laboratory study," *Econometrica: Journal of the Econometric Society*, 1275–1304.

Kagel, J. H. and D. Levin (2001): "Behavior in multi-unit demand auctions: experiments with uniform price and dynamic Vickrey auctions," *Econometrica*, 69, 413–454.

Kennedy, D. M. (2011): *Don't shoot: one man, a street fellowship, and the end of violence in inner-city America*, Bloomsbury Publishing USA.

——— (2012): *Deterrence and crime prevention: Reconsidering the prospect of sanction*, Routledge.

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011): "Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark," *Econometrica*, 79, 651–692.

Lazear, E. P. (2006): "Speeding, terrorism, and teaching to the test," *The Quarterly Journal of Economics*, 121, 1029–1061.

Li, S. (2017): "Obviously strategy-proof mechanisms," *American Economic Review*, 107, 3257–87.

Luttmer, E. F. and M. Singhal (2014): "Tax morale," *Journal of economic perspectives*, 28, 149–168.

Mascagni, G. (2018): "From the lab to the field: A review of tax experiments," *Journal of Economic Surveys*, 32, 273–301.

Meiselman, B. S. (2018): "Ghostbusting in Detroit: Evidence on nonfilers from a controlled field experiment," *Journal of Public Economics*, 158, 180–193.

Naritomi, J. (2019): "Consumers as tax auditors," *American Economic Review*, 109, 3031–3072.

Pomeranz, D. (2015): "No taxation without information: Deterrence and self-enforcement in the value added tax," *American Economic Review*, 105, 2539–69.

Pomeranz, D. and J. Vila-Belda (2019): "Taking state-capacity research to the field: Insights from collaborations with tax authorities," *Annual Review of Economics*, 11, 755–781.

Segal, I. (2003): "Coordination and discrimination in contracting with externalities: Divide and conquer?" *Journal of Economic Theory*, 113, 147–181.

Segal, I. R. and M. D. Whinston (2000): "Naked exclusion: comment," *American Economic Review*, 90, 296–309.

Slemrod, J. (2019): "Tax compliance and enforcement," *Journal of Economic Literature*, 57, 904–54.

Slemrod, J., M. Blumenthal, and C. Christian (2001): "Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota," *Journal of public economics*, 79, 455–483.

Spiegler, R. (2000): "Extracting interaction-created surplus," *Games and Economic Behavior*, 30, 142–162.

White, H. (1982): "Maximum likelihood estimation of misspecified models," *Econometrica: Journal of the econometric society*, 1–25.

Winter, E. (2004): "Incentives and discrimination," *American Economic Review*, 94, 764–773.