Using Divide-and-Conquer to Improve Tax Collection

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

Princeton University INSEAD Princeton University

November 22, 2022

Abstract

Tax collection by capacity constrained governments may exhibit multiple equilibria: if delinquency is low, limited enforcement capacity is enough to discipline deviators; if delinquency is high, limited enforcement capacity is overstretched and no longer dissuasive. In principle, divide-and-conquer, a theoretically important but untested principle from mechanism design, can be used to unravel the undesirable high-delinquency equilibrium. We investigate the challenge of doing so in practice.

Our preferred mechanism takes the form of Prioritized Iterative Enforcement (PIE). Tax-payers are assigned a rank trading-off expected collection and expected capacity use. Tax-payers are then iteratively threatened in small groups for which collection capacity is sufficient to induce compliance. After repayment occurs, unused collection capacity is released to issue the next round of threats.

In partnership with a district of Lima (Peru) we experimentally evaluate the impact of PIE on the collection of property taxes from 13432 tax-payers. Reduced-form evidence both validates and refines the theoretical benchmark. A structural model of tax-payer behavior suggests that, keeping the number of collection actions fixed, PIE would increase tax revenue by 11.3%.

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

---

*Kapon: kapon@princeton.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, and the University of Pennsylvania for very helpful comments.

1 Introduction

Tax collection by capacity constrained governments may exhibit multiple equilibria: if delinquency is low, limited enforcement capacity is enough to discipline individual deviators; if delinquency is high, limited capacity is overstretched and no longer dissuasive. In principle, divide-and-conquer, a theoretically important but untested principle from mechanism design, can be used to unravel the undesirable high-delinquency equilibrium. This paper seeks to figure out the engineering challenges of doing so 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 tax-payers are iteratively threatened with fast-track collection if they fail to make prompt payments. Upon repayment, enforcement capacity tied up by past threats is released and used to issue collection threats to a new group of tax-payers.

We partnered with a district of Lima (Peru) to experimentally evaluate the impact of PIE on the collection of property taxes from 13432 delinquent tax-payers. Using a structural model to simulate relevant counterfactuals, our preferred estimate is that implemented at scale, PIE would increase tax revenue by 11.3%. Our field experience also provides a more granular understanding of the mechanics of divide-and-conquer in the field. The success of PIE relies on carefully tuning the flow of threats being issued to manage the following trade-off. Sending a smaller number of sharp credible threats increases settlement speed, allowing threat-capacity to be redeployed faster. However, it also reduces the number of tax payers potentially settling at any given time. We hope that the lessons from our experience facilitate the applications of divide-and-conquer in other contexts, including debt collection, policing, addressing corruption in organizations, etc.

The model. We consider a government entitled to collect an amount of taxes-due D_i from tax-payers labeled $i \in \{1, \cdots, N\}$. With prior probability q_i tax-payer i may not be able to settle, and their ability to settle is private information.[1] The government is able to forcefully 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 forceful collections, with $\alpha \in (0, 1)$. Instead of forcefully 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.

Random enforcement mechanisms, in which the government collects from αN uniformly drawn tax-payers from the group of non-settlers, can result in multiple equilibria, exhibiting both high and low collection levels. PIE guarantees that the government achieves the highest possible revenue provided tax-payers are minimally rational. Specifically, when the number of tax-payers is large, under any non-obviously dominated strategy profile (Li, 2017), PIE achieves the highest possible Bayes-Nash equilibrium revenue under any mechanism.

Under PIE, tax-payers are ranked according to a specific scoring rule, and iteratively threatened with collection unless they settle at a price that increases with delay. At any point in time, the number of tax-payers being threatened is equal to the remaining collection capacity, so that settling is an obviously dominant action whenever tax-payers are able to do so. This causes the low-settlement equilibrium to unravel.[2] Tax-payers that do not settle face forceful collection. The optimal priority rule ranks tax-payers according to individualized score

$$z_i \equiv \frac{(1 - q_i)D_i}{q_i},$$

reflecting the trade-off between expected collection $(1 - q_i)D_i$, and expected capacity use q_i.

[1]Tax-payers may suffer a liquidity shock, or face 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 OB that the analysis extends essentially as is to the context of income taxes where the tax-payer may have private information about the amount D_i of taxes they owe.

[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, focusing incentives has further benefits: once agents comply, the enforcement capacity needed to ensure compliance can be redeployed to induce other agents to comply.

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 tax-payers delinquent in their first-quarter (Q1) payment. The experiment ended with the municipality's decision to adopt PIE for its regular collection process.

Jesús María typically enjoys high ultimate collection rates, but it expends significant resources on tax-collection, and roughly 30% of tax-payers are delinquent at some point during the annual collection process. In addition, collection administrators were concerned that lenient enforcement in 2020, due to Covid 19, may induce tax-payers to skip payments, expecting other tax-payers to do the same. Consultation with the city's collection unit identified a specific bottleneck in their capacity to directly collect taxes from delinquent tax-payers: a costly garnishment procedure that requires the involvement of legal professionals and bank cooperation. It was established that the city's capacity to issue garnishment orders was roughly 400 per month. Tax-payers need to be notified and issued a formal writ before garnishment can take place, but these actions are relatively cheap and the city is capable of issuing several thousand per month.

At the end of Q1 2021, we randomly assigned delinquent tax-payers to two treatment arms. A control arm implemented a common collection policy, previously used by the city: delinquent tax-payers are informed that they are delinquent all at once, preliminary collection actions are taken against most of them, and if repayment does not occur over the next quarter, collection steps are taken against tax-payers that owe the most taxes. We note that the city has an implicit ranking over tax-payers, but does not make it explicit, or inform tax-payers of their rank. As Figure 1 illustrates, the tax authority issues all threats at once and gives all of them similar deadlines. At the deadline, there is one round of enforcement on non-compliers with the greatest collection potential.

Our treatment arm implements PIE: (1) at any one time, we issue only a small number

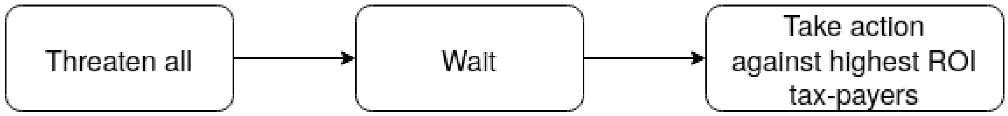


Figure 1: Collection in the control group

of targeted specific threats, with clear short-term deadlines designed to induce prompt repayment; (2) if prompt repayment occurs, we recycle freed-up threat capacity and apply it to the next small group of targeted tax-payers. Concretely, repayment propensities $1 - q_i$ used in scores z_i were predicted based on covariates using previous years' collection data. At any given point in time, the top 400 highest-ranked tax-payers who had not paid more than 50% of their taxes were assigned to priority group G1, the next top 400 were assigned to priority group G2, and the remainder of our treatment sample was assigned to priority group G3. Group membership was updated on a weekly basis, depending on repayment behavior. Members of priority group G1 were given a clear promise that income would be garnished within 6 weeks if taxes remained unpaid. 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 a definite promise. They were informed of the amount of tax they owed, of the penalty for late payment, and that they could be moved to group G2 at any time. Tax-payers assigned to the control group (group N, for “no promises") received a similar notification of the amount of tax 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. Whether PIE is effective depends on whether the increase in settlement speed afforded by high-powered G1 incentives (which could not be credibly provided to the entire tax-payer population) allows us to recycle collection capacity and reissue threats sufficiently many times that it compensates for the smaller number of targeted tax-payers at any given time. There is no trade-off under our benchmark theory because tax-payers settle immediately whenever it is dominant to do so. In practice, delays

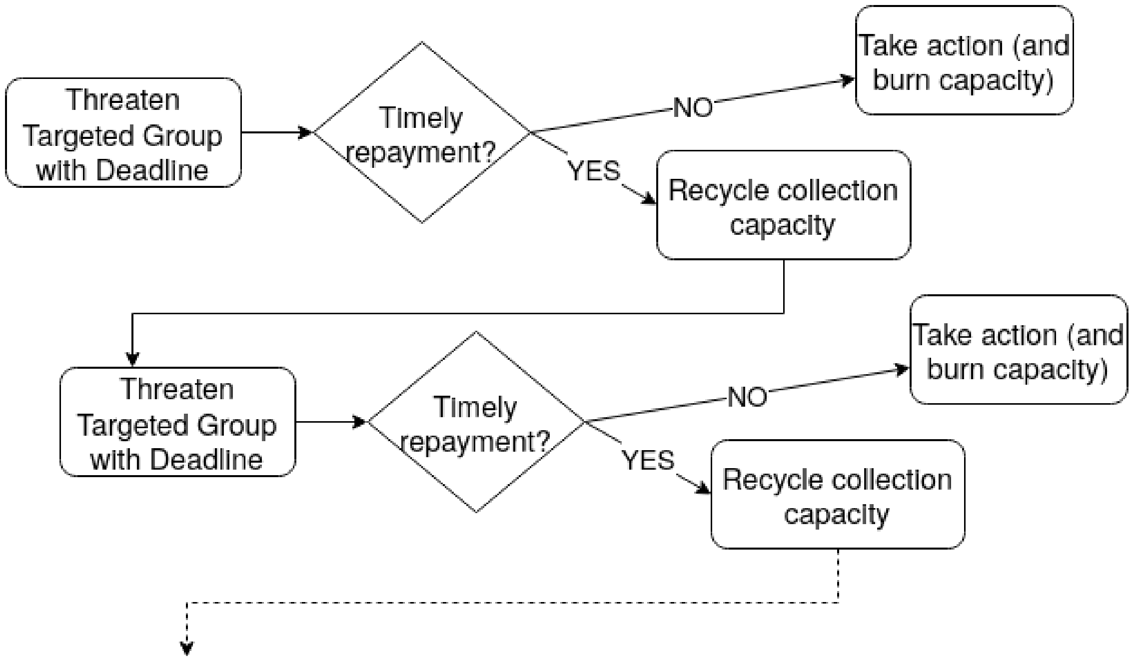


Figure 2: Collection under Prioritized Iterative Enforcement

in best-response make the trade-off less obvious. In particular, it may be optimal to threaten groups larger than the available collection capacity.

Empirical findings. We report raw findings, as well as estimates from a semi-structural model that identifies the impact of different collection steps, and permits counterfactuals.

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 tax-payers, 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 9.4% 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 2.8%). 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.[3]

We build a semi-structural model of tax-payers' behavior that permits the counterfactual evaluation of other PIE mechanisms, provided that they do not compromise the city's ability to deliver on collection threats.[4] Our model estimates confirm that G1 priorities considerably increase tax-payers' 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.

Because our control arm involved issuing many more notifications and legal writs than treatment, our reduced-form findings provide a lower bound for the impact of PIE on tax revenue. Counterfactual evaluation suggests that keeping the number writs equal across treatment and control, PIE would increase revenue over control by 11.3%. This is our preferred estimate of the impact of PIE implemented at scale.

Related literature. As far as we are aware, this paper constitutes the first experimental evaluation of divide-and-conquer mechanisms in the field.[5] There is a rich and growing theoretical literature on the use of divide-and-conquer mechanisms 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., 2019, 2020). Our contribution is to help bridge the gap between this theoretical literature and practical implementation. The evidence is encouraging to the

[3] While we have reliable data on formal collection actions taken, we do not have records of internal time use by city employees: the tax collection office consists of only 15 employees with diverse duties and task assignment is fluid. Administrators estimate that the equivalent of 2 days a week were spent implementing treatment-arm steps, and 3 days a week were spent implementing control-arm steps.

[4] A fully structural model would allow us to evaluate mechanisms that fail to deliver on promises at some rate. Our data does not inform such a model.

[5] The insight behind divide-and-conquer naturally shows up in policy. One recent notable example is Operation Ceasefire (Braga et al., 2001, Kennedy, 2011, 2012), a multi-city homicide reduction program that explicitly prioritizes the assignment of law enforcement capabilities to homicides in the order in which they are committed, thereby dissuading gangs to initiate gang wars.

theory, but highlights the importance of taking seriously realistic 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 literature on the economics of tax-compliance reviewed in Slemrod (2019). It relates to letter-based randomized control trials in which researchers have partnered with tax-collection authorities to evaluate how different tax-collection policies affect compliance. Our main concern is not just the nature of the letters that are sent (threatening or not), but also the dynamic process used to issue threats consistent with limited capacity. Slemrod et al. (2001), Kleven et al. (2011) and more recently De Neve et al. (2021) evaluate the impact of auditing threats on tax-payers' compliance, finding a meaningful impact of threats, especially on tax-payers for whom third party information is not available. Del Carpio (2014), Dwenger et al. (2016), De Neve et al. (2021) study tax-morale and evaluate the importance of intrinsic versus extrinsic incentives in achieving compliance. De Neve et al. (2021) studies the value of reducing compliance costs by simplifying communication between tax-collection agencies and tax-payers. Our results are of the same order of magnitude compared to that literature (other effects ranging from 0.7% to 23%).

The paper also contributes to a growing literature seeking to improve tax collection in developing countries. In such settings, low enforcement capacity makes optimizing design all the more important. Pomeranz (2015) uses Chilean data to establish the informational power of value-added taxes: by giving businesses incentives to report one another's revenue (to reduce their own tax burden), they generate information that tax-authorities can use to curb tax evasion. Balán et al. (2022) illustrates the informational value of using village chiefs as intermediaries for tax collection. Bergeron et al. (2021) shows that tax rates may often be above the revenue maximizing rates, not because taxes dissuade economic activity, but because tax-payers' do not have sufficiently high powered incentives to repay taxes.[6]

The paper is structured as follows. Section 2 sets up our benchmark model. Sections

[6] More broadly, see the large range of projects undertaken under the umbrella of the ICTD.

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 semi-structural model of tax-payer behavior and uses it to evaluate counterfactual policies of interest. Appendices A, B, C, and D respectively discuss commitment by the principal, collect proofs, document experimental materials, and collects figures and tables excluded from the main text. Extensive online appendices OA, OB, OC, and OD respectively report further empirical findings, extend our analytic framework to the case where tax-due amounts are not known by the tax authority, provide laboratory evidence regarding the effectiveness of divide-and-conquer mechanisms other than PIE, and provide copies of the letters actually sent to tax-payers.

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 tax-payers indexed by $i \in \{1, \cdots, N\}$ each owe the government a fixed amount D. The tax-payers and the government are all risk-neutral. If a tax-payer fails to repay on time, the government can potentially collect amount D through direct intervention – in our experimental setting, garnishing bank accounts. The difficulty is that the government has limited enforcement capacity: the government can directly collect from only $\alpha N \geq 1$ tax-payers with $\alpha \in (0, 1)$.[7] To induce tax-payers to settle their taxes voluntarily, the government can make settlement offers and commit to an enforcement rule according to the following extensive-form game:

[7]This could be because forceful collection requires resources (e.g. physically seizing assets is difficult), or because due process steps must be taken.

1. The government gives each tax-payer the possibility to settle by paying a fixed price $P \in (\alpha D, D)$. Tax-payers who settle are spared from forceful collection.

2. Tax-payers simultaneously decide whether or not to settle and pay price P.

3. The government forcefully collects D from tax-payers 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 tax-payers are drawn with uniform probability from the set of non-compliant tax-payers, and designated for collection.

- Prioritized static enforcement: tax-payers are given a known priority rank in period 1. In period 3, up to αN non-compliant tax-payers are targeted for collection in order of their preassigned rank. For simplicity, we assume that tax-payers are ranked in descending order of their index $i \in \{1, \cdots, N\}$ (i.e. tax-payer 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) Consider the case of random enforcement. There exists a Nash equilibrium such that all tax-payers settle, and a Nash equilibrium such that all tax-payers refuse to settle.

(ii) Consider the case of prioritized static enforcement. A unique strategy profile survives iterated elimination of dominated strategies: all tax-payers settle.

Under random enforcement, if most tax-payers pay their taxes, then even a small collection capacity is enough to deter unilateral deviations. However, if most tax-payers do not pay their taxes, then available capacity is thinly spread and fails to dissuade tax-evasion. Prioritized enforcement causes this last equilibrium to unravel by ensuring that available capacity is focused on a marginal set of tax-payers. It is dominant for the αN highest ranked tax-payers to settle their taxes. Anticipating this, it is a best response for tax-payers with rank up to $2\alpha N$ to settle, and so on.

This stylized model of tax collection ignores many frictions likely to matter in practice. First there is complete information about the tax-payers' ability to pay taxes. Second, we assume that tax-payers' behavior reflects common belief in rationality: many rounds of deletion of dominated strategies are needed to unravel the low settlement equilibrium. In experimental settings, players rarely seem to apply more than three rounds of iterative strategic thinking.

2.2 Modeling realistic frictions

We first introduce incomplete information about ability to pay, heterogeneity across tax-payers, 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 now assume that taxes due D_i are indexed by tax-payer identity $i \in \{1, \cdots, N\}$. Collection costs may vary across agents: forceful collection against agent i consumes $\lambda_i \in [\underline{\lambda}, \overline{\lambda}] \subset (0, \infty)$ units from the principal's total enforcement capacity αN. With probability $q_i \in [\underline{q}, \overline{q}] \subset (0, 1]$, tax-payer i is exogenously unable or unwilling to repay their taxes, say because they are experiencing a liquidity shock. The ability to repay is private information: a tax-payer 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. Tax-payers 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 $\widehat{s}_i \in \{0,1\}$; the government implements an enforcement action $a_i \in \{0,1\}$, with $a_i = 1$ denoting forceful collection. Note that settlement offers P_i, recommendations $\widehat{s}_i$, and enforcement actions a_i are correlated random variables across tax-payers. In particular, realized enforcement actions must satisfy the capacity constraint $\sum_{i=1}^N a_i \lambda_i \leq \alpha N$. The government seeks to maximize tax-revenue Π collected through settlement,

$$\Pi \equiv \frac{1}{N} \sum_{i \in I} s_i P_i$$

where s_i denotes i's settlement decision.

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

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

Problem (1) is a linear optimization problem with a single constraint. The marginal benefit of increasing agent i's probability of settlement 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 OB 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 friction rates q_i depending on take-it-or-leave-it settlement price offers P_i chosen by the principal.

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 tax-payers are boundedly rational. The underlying payoffs and types are unchanged. Settlement takes place over time $t \in [0,1]$, and tax-payers can choose to settle their tax at any time.[8] The government commits to the following settlement and collection process:

(i) Tax-payers $i \in \{1, \cdots, N\}$ 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, tax-payer i receives a settlement offer $P_{i,t} = D_{i,t} - \nu(1-t)$ with $\nu > 0$.[9]

(iii) In each period t, tax-payers are informed of their effective rank, taking into account the settlement behavior of others. Specifically, they receive signal

$$x_{i,t} = i - \sum_{j<i} s_{j,t}.$$

Tax-payers who have not settled are collected on in decreasing order of rank at time $t = 1$.

We model bounded rationality using Li (2017)'s notion of non-obviously dominated play. Let us denote by h_i private histories of tax-payer i, and by $\sigma_i : h_i \mapsto s_i \in \{0,1\}$ a feasible strategy. Denote by σ_{-i} strategy profiles by players other than i, and by ω the underlying 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 corresponding to 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,*

[8] The analysis is unchanged if tax-payers become capable of repaying their taxes at random times.

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

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

Strategy σ_i is non-obviously dominated if no strategy obviously dominates it. The laboratory evidence reported in Online Appendix OC, as well as evidence from auctions (Kagel et al., 1987, Kagel and Levin, 2001) suggests that mechanisms are more likely to be successful in practice if they achieve implementation under non-obviously dominated strategies.

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

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

(iii) aggregate revenue approaches

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

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

The approach of Section 2.2 tackles some realistic frictions and provides a useful guide to design that performed well in the lab (see Online Appendix OC). Nonetheless it exhibits significant limits. Some anticipated limits are reflected in our field design.

Commitment. The analysis of Sections 2.1 and 2.2 assumes that the government has commitment power: it keeps feasible collection promises, and tax-payers believe that it will.

In practice, tax-payers do not always take enforcement threats seriously: local governments do not always have great reputation for follow-through.

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, promises over actions far into the future are likely to be forgotten, or made irrelevant by policy and government changes. This is confirmed by the data: short-term threats are much more effective than long-dated ones. We formalize this argument in Online Appendix A.

Delay in decision making. Non-obviously dominated play assumes that whenever tax-payers learn that it is dominant for them to repay, they do so as soon as they are able. In practice, it may take time for tax-payers to make payments even if they have funds available: if the slope ν of the price schedule described in Section 2.2 is small, then incentives to act fast are small.

When there is delay in best-reply, total revenue collected can be approximately expressed as follows:

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

The term (Total Collection Period / Repayment Delay after Threat) roughly corresponds to the number of times the same collection capacity can be reused to issue a threat. For instance, if the collection period is one quarter, and delay is a month, we may hope to recycle capacity three times.

In this context, setting the number of tax-payers threatened in each iteration of PIE involves a non-trivial trade-off. Reducing the number of tax-payers increases the settlement probability conditional on receiving a threat: smaller groups receive higher powered incentives. However, it may also affect the total number of tax-payers eventually threatened: when delay is bounded away from zero, the number of times collection capacity can be recycled is not infinite.[10]

Because of this trade-off, our experimental design reflects a compromise. The number of short-term collection promises that we issue is greater than available flow enforcement capacity (i.e. acting on threats is infeasible in a worst case settlement scenario), but much lower than the maximum feasible flow of promises given equilibrium settlement rates.

3 Experimental Context

From April to September of 2021, we partnered with the municipality of Jesús María, a district of Lima (Peru), to collect property-related taxes from 13,432 tax-payers delinquent in their first quarterly payment. This section details the context for our experiment, and why this context seemed well suited to evaluate prioritized enforcement.

3.1 General context

Property taxes and user charges. Our study targets the two most important municipal taxes in Peru, which are both property-related: (i) property taxes, based on land values as well as assessed building construction costs, with progressive tax rates ranging from 0.2% to 1% of total assessed value, and (ii) user charges, covering the provision of public goods such as trash collection, maintenance of green areas, and public safety, charged to each property

[10]Under the assumption that agents use non-obviously dominated strategies there is no trade-off. Once settling is dominant, tax-payers settle immediately, so that repayment delay after a threat is equal to 0, provided the tax-payer can pay their taxes. As a result, small threat groups do not reduce the total number of tax-payers threatened, since it is possible to iterate very fast.

and also varying depending on the quantity and quality of public goods provided. In 2020, property-tax and user charges represented almost 50% of total municipal revenues.

Jesús María. Jesús María is one of 43 municipal districts of Lima. It belongs to the top quartile of districts both in terms of income and educational attainment. 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).[11]

Properties are linked to over 35,000 registered taxpayers, of which 90% live in the district. In 2020, total annual taxes due, including property taxes and user charges, stood at US$15.8 million, while the average annual tax due amounted to US$435. The distribution of taxes dues is skewed to the right. The ten largest taxpayers (mostly tax-payers with commercial properties including shopping malls and real estate agencies), represented 16% of total tax due, while the top 500 tax-payers accounted for 42% of total tax due.[12]

In 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. The impetus for experimentation was partly driven by the municipality's concerns over collection in 2021. It appeared plausible that the economic shock associated with the Covid 19 pandemic may push taxpayers to a low settlement equilibrium. In addition, because tax collection costs were already high, the city council was unable to increase tax collection budgets. This motivated the tax collection authority to seek ways to deploy limited collection capacity more effectively.

Experimentation was facilitated by several other facts. First, within constraints set by national law, Peruvian municipalities have significant degrees of freedom in how they

[11]Properties are assessed using construction costs rather than commercial values. Official construction rates per sqm are provided by the national government.

[12]We note that the top 10 tax payers tend to pay taxes on time. For this reason, only one entered our sample of delinquent tax-payers (it was assigned to the treatment group) and for a relatively small amount of taxes due (2000 soles).

administer their tax collection process. Second, Jesús María had taken specific steps over the previous 2 years to enhance its reputation vis à vis tax-payers, including banning the use of tax amnesties. This means that collection promises would a priori be taken seriously.

We note that 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 running what we believe is the first experimental implementation of divide-and-conquer in the field. This potentially affects the external validity of our findings in two ways. On the one hand, the upside of PIE may be higher in settings with lower capacity. On the other hand, more constrained organizations may find it difficult to implement PIE. On the whole, we tend towards optimism. External validity is improved by the fact that we relied solely on existing municipal employees, rather than hiring workers of our own. In addition, our experience suggests that operations can be streamlined over time. For instance, counterfactuals provided in Section 6 suggest that the scoring rule may be simplified at little cost to efficiency.

3.2 The standard tax collection framework

Collection steps. Property taxes and user charges are enforced jointly, on a quarterly basis. They require specific preliminary collection steps before garnishment can take place.

Figure 3 summarizes the key collection steps as well as the usual collection timeline following quarterly deadlines. Collection consists of two main stages: the preliminary or *ordinary* collection and the formal process or *coercive* collection. Ordinary collection starts right after the payment deadline and involves: (i) sending bulk reminders (mostly through emails and sms) to all taxpayers who have 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 notification (sent through a letter) with the amount owed ("valor"), which also triggers a countdown at the end of which, legal collection procedures can proceed ("the coercive process").

The coercive process for property taxes (resp. user charges) can only begin 1 (resp. 21) working day(s) after the tax-payer is formally notified. The government does not automatically initiate the coercive process when allowed to do so. Once the government initiates 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 tax-payer receives notification of the writ.

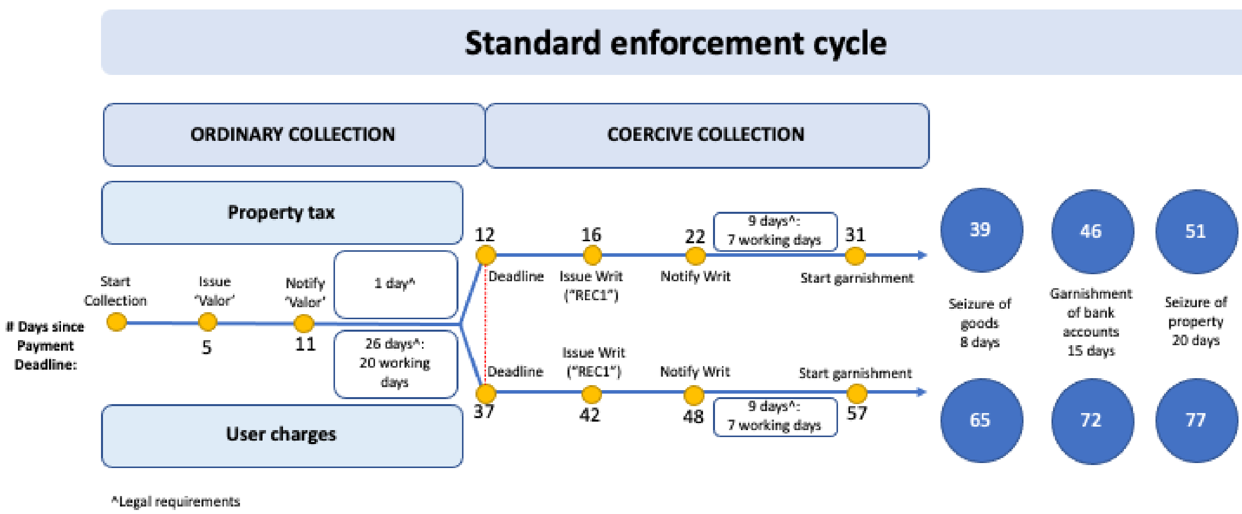


Figure 3: Standard collection timeline

In general, the city government has three main options for collection: (i) garnishing bank accounts, (ii) seizing goods at the property, and (iii) placing a lien on the property itself. Garnishing bank accounts is by far the most effective measure, but not every tax-payer has a bank account. Seizing durable goods from the property is used for smaller debts. In this case, a formal notification is delivered first, and then a municipal truck is sent to the property to seize the goods. Placing a lien on the property is used rarely, and only for very high debts. In this case the government asserts a right of first-repayment if and when the property is sold, but usually doesn't provoke the sale itself. Due to the Covid 19 pandemic, garnishment was the only collection step taken in 2020 and 2021.

The collection steps described above are the same for all tax-payers, though some additional steps depend on the size of the debt. In particular, the largest 500 tax-payers by annual amount owed are also assigned a dedicated collection agent that manages their account.[13] In addition, for smaller debtors, debts across different quarters are pooled and enforced with low intensity once or twice per year. They amount to a small share of taxes due, and we exclude them from our analysis.

Penalties. Daily interest rates, corresponding to an annual rate of 10.8%, is applied to all delinquent debt. A penalty of 10% of debt due is added when the coercive process begins. In addition, the tax-payer is charged for some of the collection expenses incurred by the municipality, averaging to US$35 per delinquent tax-payer. When coercive collection begins, the municipality registers all tax-payer debt with a credit-risk agency, which lists tax-payers as delinquent in national databases.

Capacity constraints. Collection is conducted by 15 city employees coordinated by the head and the deputy head of the collection unit. Five employees are responsible for ordinary collection (one is dedicated to the top 500 tax-payers) and three employees are responsible for coercive collection. Two employees are in charge of delivering notifications, one employee is in charge of IT, and two employees provide overall support.

Collection steps are limited by the available workforce, limited budget, and the capacity of service providers (e.g. banks). Table 1 depicts total monthly enforcement capacity by collection action, as estimated by city officials. The city has very large capacity for cheap messaging and collection steps, including issuing formal writs (between 5000 and 16000 a month), and much lower capacity for actual garnishment (400 a month). This represents a bottleneck in the city's collection capacity. The effective use of limited garnishment capacity

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

	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

was therefore the focus of our experimental treatment.

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.[14] The sample population for our experiment consisted of tax-payers delinquent on their first quarter (Q1) property tax or user charges by April 5th, 2021.[15] Figure 4 summarizes the experiment’s timeline.

Following the payment deadline, 13,432 tax-payers who had not paid their Q1 2021 taxes as of April 5th and had a tax 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 tax-payers were randomly assigned to a prioritized enforcement mechanism described below, while the remaining tax-payers were assigned to the standard collection proce-

[14]We did not commit to a pre-analysis plan.

[15]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.

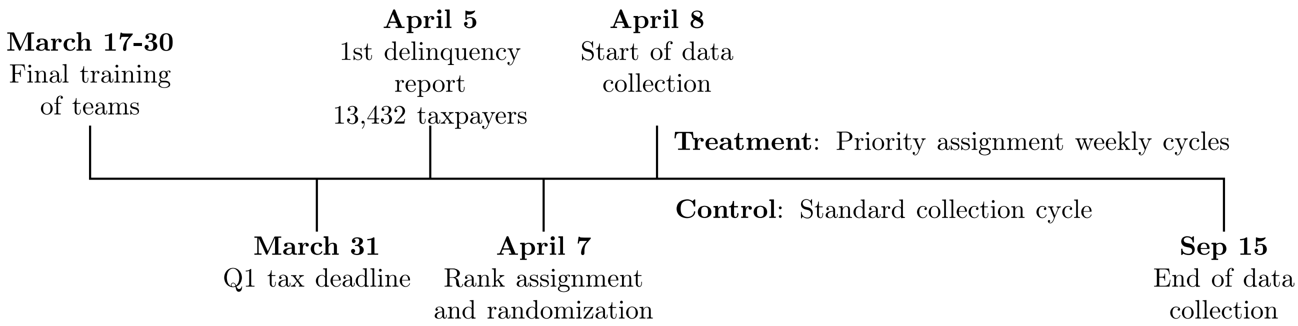


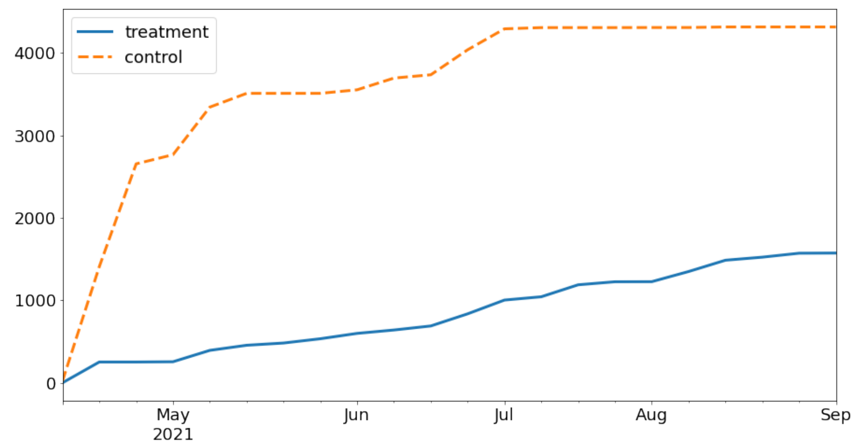
Figure 4: Experiment timeline

dure used by Jesús María. Following Banerjee et al. (2020), we drew our sample assignment uniformly from the set of 10% most balanced samples under the Mahalanobis distance, targeting balance on tax-payer age, tax due, status as a top 500 tax-payer, and expected repayment probability. Table D.1 provides summary statistics.

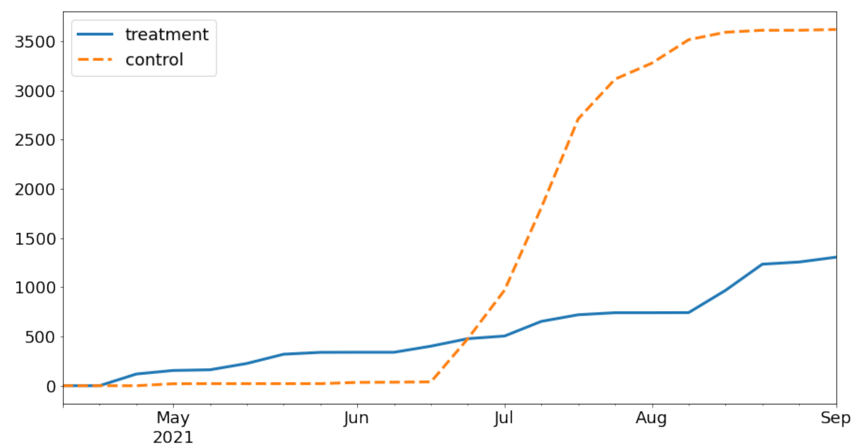
Control and Treatment. The control arm follows the collection process stylized Figure 3 and described at length in Section 3.2: the bulk of tax-payers are concurrently issued identical notifications; after some delay, the bulk of tax-payers are issued legal writs; after further delay a number of tax-payers are placed in garnishment. Tax-payers go through each stage of collection as a group, and are not issued prioritized threats with tight deadlines.

The treatment arm follows the structure of PIE, illustrated in Figure 2. Small groups of tax-payers are iteratively issued short-term (6 weeks) assertive collection threats. Notifications and writs are issued only to threatened tax-payers, under the shortest schedule consistent with the law. Garnishment takes place if tax-payers fail to meet minimum payments by the deadline they were given. New threats are only issued in proportion to capacity freed up either by tax-payers settling their taxes, or by garnishments being executed.

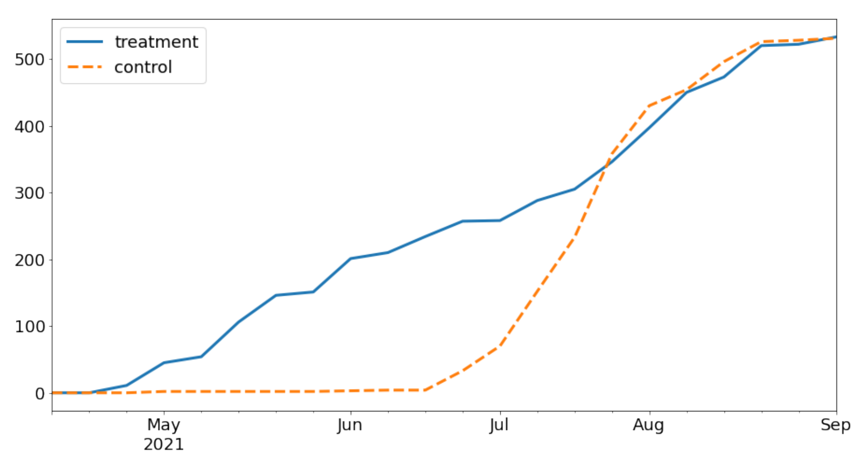
Figure 5 anticipates our discussion of raw outcomes and shows the number of collection actions taken over time. Control and treatment are associated with roughly the same final



(a) Notifications issued.



(b) Writs issued.



(c) Garnishment orders issued.

Figure 5: Number of collection actions taken

number of garnishments. However, the processes leading up to these final figures differ. As described, under the control, notifications and writs are issued in bulk, and tax-payers go through each collection stage as a group. In contrast, under treatment, notification and writs are only issued following targeted threats. All collection steps, including garnishment, take place continuously, in a manner consistent with short-term threats, allowing the recycling of collection capacity once repayment occurs.

Failure of SUTVA. Although we are able to randomly assign large number of tax-payers to treatment and control, we are also dealing with a single implementation organization: the tax-collection department of Jesús María. As the collection unit is a small organization, we could not guarantee balance with respect to employee characteristics, and so chose to rotate employees across treatment and control. In addition, prioritized enforcement ended up being less labor intensive than the city's usual collection process (this is reflected by collection actions illustrated in Figure 5; informal estimates from our implementation partners suggest that the control collection process was 50% more time consuming than prioritized enforcement). As a result, employees nominally assigned to the treatment group spent part of their work-week helping employees assigned to the control treatment. This contributed to the large number of writs issued by the city under the control arm.

As a result, simply comparing treatment and control arms likely underestimates the effect of PIE implemented at scale. To draw more meaningful policy comparisons we interpret the raw findings reported in Section 5 using the semi-structural model of Section 6. We estimate a model of individual settlement behavior that allows us to simulate informative counterfactuals under various scenarios.

Sections 4.2 and 4.3 provide further details on the logistics of running PIE in the field.

4.2 Rank assignment

The score s_i defined by (2) requires knowledge of taxes due D_i, the likelihood of making a repayment $1 - q_i$, and the cost of collection λ_i. In our application, taxes due D_i are observed. The cost of collection λ_i is assumed to be roughly constant and can be normalized to 1 without changing the ranking of tax-payers. The main challenge is to estimate $1 - q_i$.

We do so by building a simple model of repayment using administrative data for 2019 and 2020, averaging out highly correlated predictions from OLS, LASSO and Random Forest estimators. Implementation details are provided in Online Appendix OA. The key points are the following:

- Predicted repayment probability is increasing in tax-due. This implies that score $(1 - q_i)D_i/q_i$ is increasing in tax-due D_i, so that the optimal score puts a higher priority on collecting from large tax-payers. This means that PIE should improve the progressivity of tax-collection.

- We build two predictive models of repayment. A more precise model that uses endogenous past repayment behavior as a covariate (referred to as the endogenous model), and a somewhat less precise model that does not (referred to as the exogenous model). In principle, the endogenous model is problematic: if past behavior is used to prioritize tax payers, this may give tax-payers incentives not to repay taxes to lower their priority. Everything else equal, we would prefer to use models based on exogenous covariates, but we are interested in evaluating the short-term loss of revenue from doing so. We show in Section 6 that scores estimated using exogenous covariates can be used at no loss of revenue.

4.3 Prioritized iterative enforcement in the field

Our field implementation of PIE reflects legal constraints on the timing of notification, writ, and collection steps, as well as concerns over commitment power, and delay in tax-payers' reactions. A total of 6704 tax-payers were assigned to this collection process.

Priority groups. At any given point in time during the experiment, we grouped tax-payers in three priority groups, G1, G2, and G3, corresponding to distinct collection promises allowed by total capacity: the top 400 highest ranked tax-payers 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 of our treatment sample was assigned to group G3. Group membership was updated on a weekly basis. New members of a given group were sent a physical card, as well as an email clarifying the collection promise applying to them. A translated information letter for group G1 is reproduced in Table 2. Translated information letters for other groups are reproduced in Appendix OA, along with Spanish originals.

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 the penalties and administrative expenses regulated by law and will amount to:	**Amount*1.1 +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 2: Information letter for priority group G1

Priority group G1 was given a promise that income would be garnished within 6 weeks if taxes remained unpaid. Information letters sent to priority group G2 were similar. Members of group G2 were promised that their income would be garnished within 12 weeks in the absence of tax payments. In addition, they were informed that they could be moved to group

G1 at any time. The rationale for group G2 was to engage tax-payers' forward thinking, and get this second group to start getting ready to make payments earlier. In contrast to group G2, members of group G3 did not receive a definite promise. They were informed of the amount of tax they owed, of the penalty for late payment, and that they could be moved to group G2 at any time. Tax-payers assigned to the control group (referred to as group N, for "no promises") received a notification of the amount of tax they owed, of similar complexity (see Online Appendix OD, Figures OD.1, OD.2, OD.3, and OD.4).

We note that letters sent to groups G1 and G2 are more assertive than letters sent to group G3 and group N: they specify a clear deadline for collection. This is part of treatment. As Figure 2 highlighted, ensuring prompt repayment is at the core of the logic of PIE: this allows us to recycle collection capacity. In addition, these specific threats are not free: it is not credible to issue such threats to the entire population, and doing so would erode confidence in the government. Assertive threats are only feasible if they are issued at a manageable pace. Anticipating on the semi-structural estimation of Section 6, we note that only group G1 letters increased tax-payers' settlement intensity, while letters sent to group G2, and G3, had little impact over letters sent to control group N. Letters did not affect settlement through mechanisms other than the tighter deadlines at the heart of PIE.

We deviated from the general rule of assigning the 400 highest tax-payers to group G1 in two ways. The first time assignment took place (April 5th, 2021), 200 G1 spots were assigned to the highest ranked 200 tax-payers, and 200 G1 spots were randomly assigned to tax-payers with rank below the top 200. This allowed us to get an early estimate of the impact of getting a G1 collection promise versus a G3 or N collection promise, validating one of the key assumptions needed for prioritized enforcement to work: specific short-term promises significantly increase the settlement rate of tax-payers; and estimated repayment propensity predicts actual repayment propensity. A second deviation is that we increased the size of group 1 to 600 in June 2021, reflecting the fact that the number of garnishment orders issued remained significantly below the available capacity.

Excess promise making. We note that along the lines discussed in Section 2.3, we ended up issuing a higher flow of collection threats than the city government could really process in the worst case scenario where no tax-payer repaid their taxes. Indeed, in principle only a half of the garnishment capacity is available to the treatment arm, this corresponds to a capacity of 200 tax-payers over 4 weeks, so roughly 300 tax-payers over 6 weeks. Therefore, if more than 3/4th of tax payers do not make required payments within 6 weeks, we would break promises made to members of group G1. This could potentially lead to multiple equilibria. As it turns out, even with this excess promise making, we do not consume all available garnishment capacity (across treatment and control, we end up issuing roughly 1100 garnishment orders over 5 months, instead of a theoretical capacity of 2000).

Collection actions. To minimize the time-horizon of promises made to G1 members, in the treatment arm, collection actions were only taken if a G1 collection promise was issued. This led us to establish a fast processing schedule achieving the minimum delay in promise-delivery compatible with regulation. It is illustrated by Figure C.1.

We did not implement collection actions for members of groups G2 and G3 apart from sending an initial information letter, and making reminder phone calls to the same proportion of delinquent tax-payers as in the control group. This choice was motivated by the fact that garnishment is the only collection step that has direct real consequences to tax-payers. We show in Section 6 this assumption turned out to be wrong: sending legal writs has a large impact on tax-payer behavior, even if most cannot actually be acted on by the city. This benefited collection in the control arm, where greater resources were spent on issuing writs.

5 Raw Findings

We use as our main outcome both the total tax revenue as well as the number of collection actions taken by the city government. In addition we document that the basic premises required for PIE to be effective are present: receiving clear short-term collection promises increases the settlement rate of tax-payers; and it is possible to predict repayment propensity.

5.1 Main outcomes

Tax collection. Figure 6 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 OA.

As of September 15, 2021, total tax collection in the treatment group was 9.4% higher when compared to the control group. The speed of collection is also higher under treatment than control throughout the experiment. These raw findings require qualification because the distribution of taxes-due has a long right-tail. In Section 6, we estimate a collection increase of 2.8% using a more robust model estimated using only binary payment decisions, rather than payment amounts. However, to evaluate implementation at scale, we must take into account differences in the number of collection actions taken.

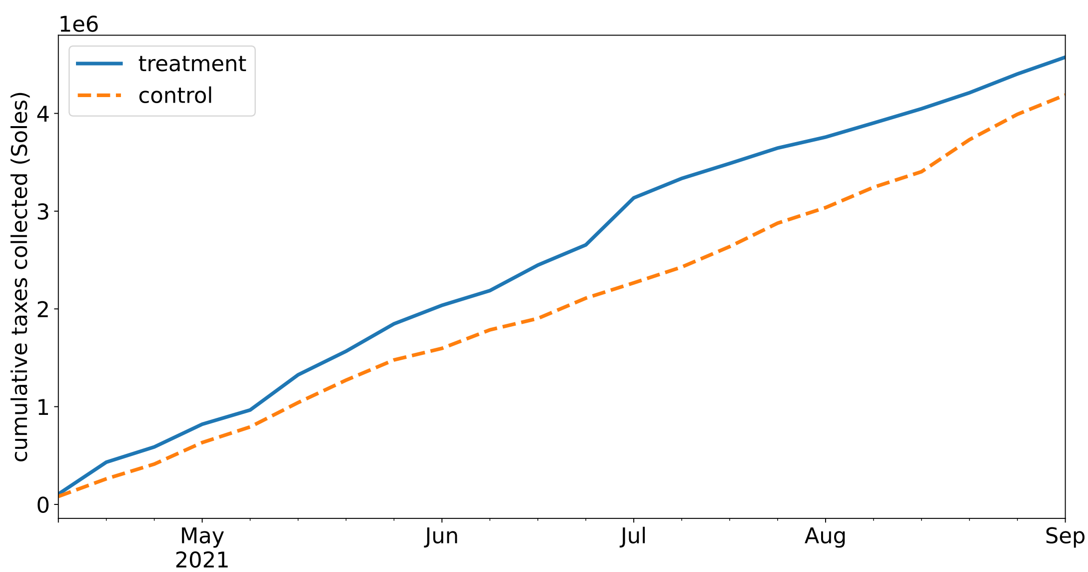


Figure 6: Cumulative Tax Collected April - September 2021

Collection actions. Table 3 (and Figure 5) report the number of collection actions taken in both the treatment and collection arms during the entire experimental period. Although the number of garnishment orders issued is roughly the same across treatment and control, the city government issued 3 times as many notifications and writs in the control arm as in the treatment arm.

Number of tax-payers who have received			
	Notification	Writ	Garnishment
Treatment	1,534	1,283	537
Control	4,301	3,581	528

Note: *Notification* is the initial notice informing the taxpayer of their delinquency. *Writ* is the legal document indicating the beginning of coercive collection. *Garnishment* refers to the process of collecting payment from the bank account of the taxpayer.

Table 3: Number of collection actions taken

This reduced use of notifications and writs in the treatment group was driven by our desire to keep garnishment deadlines for group G1 short: we only issued notifications and writs to tax-payers in group G1. We study the impact of this greater use of relatively cheap collection actions in Section 6.

Progressivity of tax-collection. Because the predicted likelihood of repayment is an increasing function of taxes due, prioritized enforcement enhances the progressivity of tax-collection. This is illustrated by Figure 7 which plots the share of the total tax collected raised from tax-payers who fall within the bottom $q\%$ of the distribution of amount of taxes due, for increasing quantiles q. Treatment shifts the curve to the right, indicating that tax-payers who owe large amounts of taxes pay a larger share of total taxes under treatment than control.

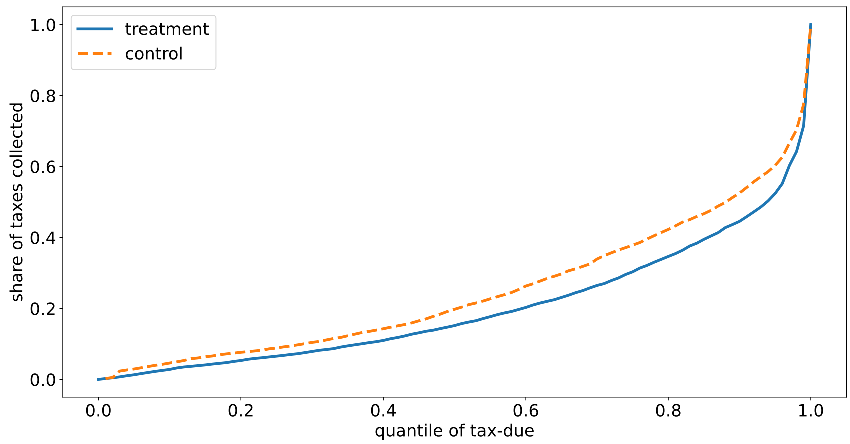


Figure 7: Share of total tax revenue collected as a function of quantile of taxes due.

5.2 Evidence on mechanisms

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

- clear short-term promises significantly increase the settlement rate of tax-payers;
- our ranking of tax-payers usefully predicts repayment behavior.

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

Impact of short-term promises on settlement. Figure 8 focuses on tax-payers with rank less than 200 included in group G1 of the treatment arm as part of the first batch of group G1 assignments. It plots the share of tax-payers who have repaid at least 50% of tax-due. We use control tax-payers with similarly distributed scores as a comparison.

Tax-payers exhibit a significantly higher settlement rate under treatment than control. We emphasize that this is true even in the first few weeks of priority group assignment. Since no collection actions take place during this period, the early impact of G1 membership is entirely driven by collection threats, rather than collection actions.

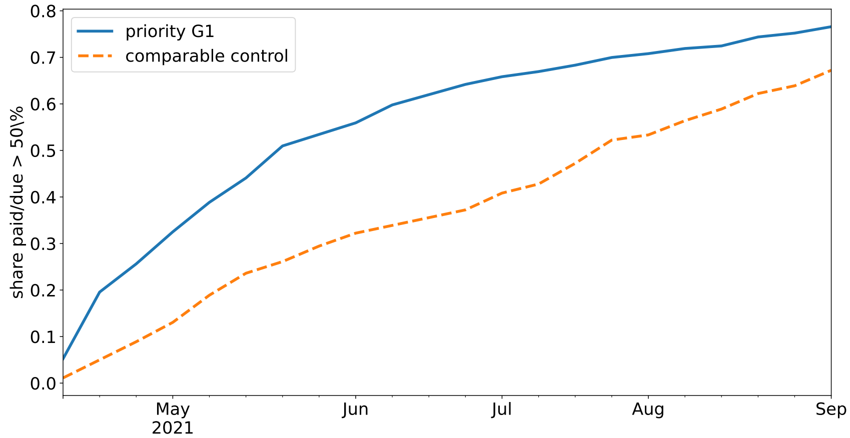
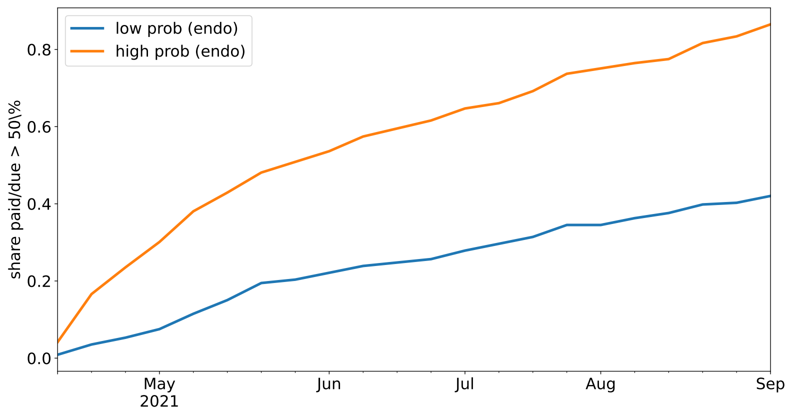


Figure 8: Repayment G1 vs Control, Rank<200

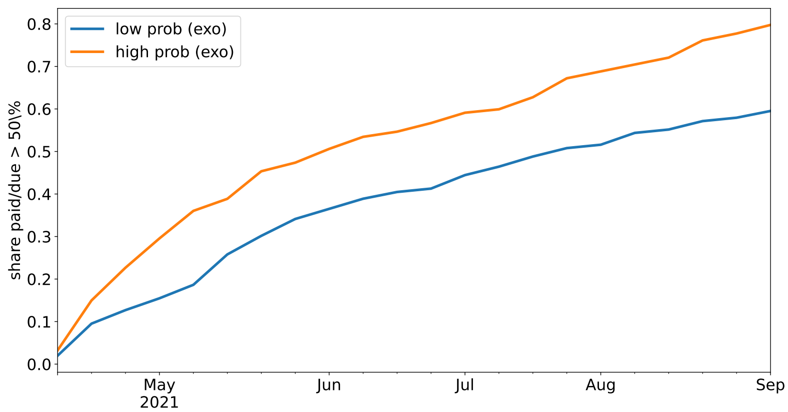
Predictability of repayment behavior. Figure 9 illustrates differential settlement behavior for the top 33% and the bottom 33% of tax-payers with respect to different predictors of repayment: predicted repayments from a model using endogenous past repayment, predicted repayments from a model excluding past repayments, and predicted repayments ranking tax-payers based on total taxes-due alone. Figure 9 suggests that all three rankings predict repayment behavior, but using endogenous data improves precision.

Note that better classification does not necessarily translate into better performance. Indeed, the scoring rule only affects collection to the extent that it induces the optimal set of tax-payers to be settlers vs. non-settlers. Changing the ranking of tax-payers within the group of settlers does not improve collection. We return to this point in Section 6 where we evaluate the impact of simpler scoring rules.

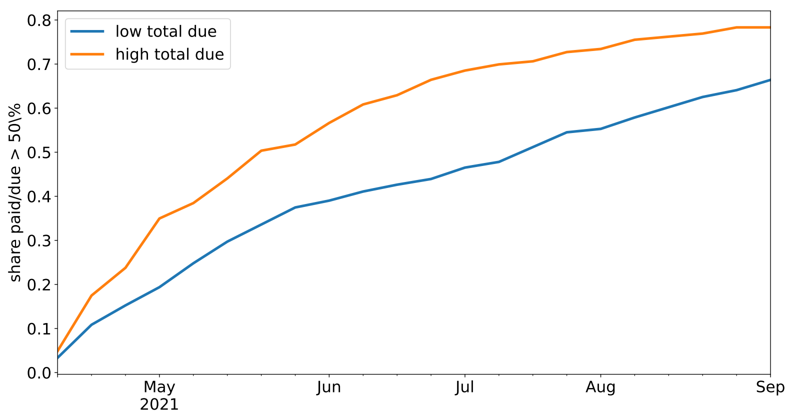
Reputation formation. We assess the impact of treatment on the propensity of tax-payers to be delinquent in subsequent quarters. For all tax-payers delinquent in the first quarter, we observe the amount by which they are delinquent in Q1 (their Q1 Debt), and whether they had any debt related to second quarter taxes (Q2 Debt). If the tax-payer is not delinquent with respect to Q2 taxes, then their Q2 Debt is set to 0. For the sample of



(a) predicted repayment, using past repayment behavior.



(b) predicted repayment excluding past repayment behavior.



(c) predicted repayment by taxes due.

Figure 9: Share of population having repaid more than 50% of taxes due, by top third, and bottom third of predicted repayment probability

tax-payers delinquent in Q1, we estimate the following linear model via OLS:

$$\frac{\text{Q2 Debt}}{\text{Q1 Debt}} \sim 1 \oplus \text{Treatment} \oplus \text{Assignment to G1}. \tag{4}$$

Estimated coefficients are reported in Table 4. Being assigned to treatment does not reduce Q2 Debt, but being assigned to group G1 does. This suggests that the effect of treatment may grow over time, as the number of tax-payers assigned to group G1 at some point expands.

	Q2 Debt
Constant	1.3771***
	(0.0106)
Treatment	0.0075
	(0.0141)
Assignment to G1	-0.2181***
	(0.0354)
Observations	13432

Notes: Delinquency in terms of Q2 Debt following assignment to treatment and G1 in the first quarter. Robust standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Impact of treatment on subsequent delinquency.

6 Counterfactuals

6.1 A Semi-structural model

We now estimate a model of settlement behavior that takes into account threats made, and collection actions taken. This is the core of our empirical analysis. The model is semi-structural in the sense that it allows us to evaluate many relevant, but not all, counterfactual policies. In particular, we cannot estimate how failures to deliver on promises on time affects the settlement rate of tax-payers. For this reason we only consider counterfactual policies that do not affect the city government's ability to deliver on promises. This includes an estimate of the impact of PIE, keeping the number of writs issued the same across treatment and control.

Repayment behavior. We assume that each tax-payer i is associated with a persistent observed characteristic $\xi_i \in \mathbb{R}$ and an unobserved persistent type $\theta_i \in \mathbb{R}$, drawn i.i.d. across tax-payers from a Gaussian distribution $\mathcal{N}(0, \sigma^2)$. In our implementation, we use as observed characteristic the tax-payer's predicted repayment probability from our most predictive model (which includes past repayment behavior).[16] Unobserved type θ_i serves to explain correlation in repayment behavior across periods, and captures the impact of selection over time: tax-payers who have not made repayments after 3 months are systematically different from tax-payers who have not made repayments after 2 weeks.

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

In each period t, a tax-payer i makes a payment with Poisson intensity $\kappa_{i,t}$. We denote by $s_{i,t} = 1$ the event that the tax-payer makes a payment, and by $s_{i,t} = 0$ the event that they don't. We assume that conditional on making a payment, the share of taxes-due repaid with this payment (or normalized payment), $\pi_t \in [0, 1]$, is drawn from a fixed distribution $f_{\pi,i}$ that depends only on the taxes owed by tax-payer i.[17] Let $T_i(t)$ denote the set of tax-payer i's payment times occurring strictly before t. We denote by $\Pi_{i,t} = \sum_{s \in T_i(t)} \pi_{i,s}$ the running sum of normalized payments made up to period t. Except for small penalties, a tax-payer's

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

[17] Tax-payers are placed into one of 13 bins based on amount due, and π_t for a taxpayer in a bin is drawn from the empirical distribution of payments associated with that group of tax-payers.

total payments should rarely exceed 4 times quarterly taxes-due.[18]

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

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

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

$$\kappa_{i,t}(\theta_i, \beta) = \max\{10^{-3}, \phi(\langle X_{i,t}, \beta \rangle + \theta_i) \times \mathbf{1}_{\Pi_{i,t}<4}\} \tag{5}$$

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. Note that conditional on type θ_i, the intensity of payment behavior at time t depends only on the current priority group $g_{i,t}$, and the latest collection action taken $a_{i,t}$. The past only affects expected settlement intensity through the posterior distribution over types θ_i.

The per-period payment probability associated with intensity $\kappa_{i,t}$ is denoted by $K_{i,t} \equiv 1 - \exp(-\kappa_{i,t}) \simeq \kappa_{i,t}$ (when $\kappa_{i,t}$ is small).

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, priority assignments, and payments made, associated with tax-payer 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 $h_{i,t}$ alone.*

We denote by $G(\cdot | h_{i,t}) \in \Delta(\{G1, G2, G3, N\} \times \{\text{garnishment, writ, notification, N}\})$ the

[18] For most tax-payers, the amount of taxes owed each quarter in a year is approximately the same.

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). Specifically, the likelihood of final histories $h_{i,T}$ can be factorized as follows.

For any final history $h_{i,T}$,

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

Importantly, 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.

$$\mathcal{L}(h_T|\varphi,\beta,\sigma) \equiv \sum_{i\in I} \log(\Psi(h_{i,T}|\varphi,\beta,\sigma)). \tag{6}$$

In turn, $f_{\pi,i}$ can be estimated parametrically or non-parametrically, using conditional payment data in the event a payment is made, for tax-payers with tax-due amounts similar to tax-payer i.

We note that by construction, the estimation of parameters of interest (φ, β, σ) is not driven by tax-payers with large amounts of tax-due. Instead, parameters of interest are estimated using only tax-payers' 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 12 parameters:

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

We do so by computing a posterior distribution over parameters φ, β, σ using Markov Chain Monte Carlos (Chernozhukov and Hong, 2003).[19]

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 (5) as

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

We aggregate payments at the weekly level, so that $1 - \exp(-\kappa_{i,t}) \simeq \kappa_{i,t}$ is the probability tax-payer i makes a payment in week t.

[19] 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. However, our data partially challenges this prior restriction: during the first 2 months of the experiment, tax-payers in the control group that receive a formal notification tend to make payments at a lower rate than tax-payers 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 OA. Removing this prior-restriction does not qualitatively change the inferences we draw from data.

6.2 Findings

Table 5 reports posterior means and standard deviations for parameters of interest. Corner plots, as well as estimates computed considering only the payment of Q1-dated taxes only, are provided in Online Appendix OA.

Consistent with the reduced-form evidence reported in Section 5, inclusion in group G1 has a large impact on payment intensity, and predicted settlement probability ξ_i is indeed predictive of settlement behavior. In contrast, the coefficient associated with priority G2 is much smaller: the short deadlines associated with group G1 are needed to induce a noticeable behavioral response. Although we highlighted that members of group G2 could be moved to group G1 at any time, this did not seem to effectively engage tax-payers' anticipation and planning, possibly because that claim was too vague to be credible.

Issuing formal writs has a meaningfully large impact on settlement.[20] This confirms that the systematic use of writs contributes positively and significantly to settlement rates in the control sample. Our estimate of coefficient σ suggests meaningfully large amounts of unobserved heterogeneity in types. Coefficients on payment variables suggest that tax-payers who have made some payments are subsequently more likely to make further payments, but less so if the cumulated normalized amount paid is larger.

6.3 Counterfactuals

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

This implies that estimated parameters φ, β, σ allow us to evaluate counterfactual mechanisms assigning actions and priorities as a function of public histories, provided that promises

[20] Note that the coefficients associated with each collection action should not be added to get the cumulated impact of collection actions. Instead, the coefficient associated with each collection action summarizes the aggregate effect of the current action and preceding required collection steps.

	Mean	(std. dev.)
$\beta_{\Pi_{i,t}>0}$	$3.22 \cdot 10^{-2}$	$(2.62 \cdot 10^{-3})$
$\beta_{\Pi_{i,t}}$	$-3.78 \cdot 10^{-2}$	$(1.10 \cdot 10^{-3})$
β_{G1}	$3.06 \cdot 10^{-2}$	$(2.55 \cdot 10^{-3})$
β_{G2}	$0.50 \cdot 10^{-2}$	$(2.87 \cdot 10^{-3})$
β_{G3}	$-0.65 \cdot 10^{-2}$	$(1.37 \cdot 10^{-3})$
$\beta_{\text{garnishment}}$	$1.08 \cdot 10^{-2}$	$(3.38 \cdot 10^{-3})$
β_{writ}	$2.96 \cdot 10^{-2}$	$(2.15 \cdot 10^{-3})$
$\beta_{\text{notification}}$	$0.16 \cdot 10^{-3}$	$(0.15 \cdot 10^{-3})$
β_{ξ}	$1.20 \cdot 10^{-1}$	$(4.05 \cdot 10^{-3})$
$\underline{\varphi}$	$0.86 \cdot 10^{-2}$	$(1.92 \cdot 10^{-3})$
$\overline{\varphi}$	$3.07 \cdot 10^{-1}$	$(0.74 \cdot 10^{-1})$
σ	$0.49 \cdot 10^{-1}$	$(2.30 \cdot 10^{-3})$

Table 5: Estimating the settlement behavior of tax-payers.

to collect continue to be kept under the counterfactual, keeping fixed the settlement behavior of tax-payers.

In the language of Engle et al. (1983), this ensures that the priority assignment and collection action process is super-exogenous to settlement behavior over the restricted class of mechanisms that maintain promises. Throughout the rest of this section, counterfactuals are chosen so that they do not break the capacity constraint of 200 garnishments per month in simulations.[21] We provide simulations of capacity use in Online Appendix OA.

Relevant counterfactuals. We are interested in the following counterfactuals whose results are summarized in Table 6. In each case we use our model to simulate behavior under both the control policy and a version of PIE.

1. *Replicating experimental findings in a manner robust to large repayments.*

[21] To ensure our counterfactuals do not break capacity constraints, we promote tax-payers into group G1 gradually: we initialize 200 tax-payers in G1 and all other tax payers in G3, and restrict promotion into G1 to a maximum of 70 a week. This ensures that less than 200 garnishments take place every month.

The weight of large repayments means that raw comparisons of revenue across arms are very noisy. Because parameters reported in Table 5 are estimated using only binary payment events, rather than payment amounts, they are not sensitive to large repayments. Using these estimates, we simulate out collection rates for treatment and control under collection actions similar to the ones used in our actual experiment. Running many such simulations allows us to average out the noise associated with large repayments. This exercise suggests a 2.8% increase in expected repayments, taking as given the very asymmetric use of collection actions across treatment and control.

2. *PIE versus control, setting the number of writs to match the control group.*[22]

   This lets us evaluate the uncontaminated counterfactual treatment effect of introducing PIE without changing the number of relatively cheap collection actions taken across the treatment and control groups.

   As Table 6 reports, we estimate that keeping the number of writs constant across arms, introducing PIE would increase collection by 11.3%. *This is our preferred estimate of the impact of PIE implemented at scale.*

3. *Ranking tax-payers based on endogenous scores, exogenous scores, total tax due, and a uniform random order.*

   This addresses several policy relevant questions. Is the loss from using exogenous data to rank tax-payers large? Second, can we simplify the scoring rule, and use only tax-due as a basis for ranking? Third, is ranking players at all important? Could we use a random order which may be perceived as a fairer procedure?

   As Table 6 reports, whether we use a rank based on endogenous scores, exogenous scores, or total tax-due has a very limited impact on tax-revenue. To a first order, what matters is how the ranking splits tax-payers between those who are assigned a

[22]We target issuing 3000 notification within the first 4 weeks, and 3000 writs from week 4 to week 8 under both treatment and control.

counterfactual policy	change in tax revenue vs. control
PIE with reduced # of writs	+2.8%
PIE with matching # of writs	+11.3%
PIE with endogenous rank (reduced # of writs)	+2.6%
PIE with exogenous rank (reduced # of writs)	+3.7%
PIE with tax-due rank (reduced # of writs)	+3.6%
PIE with random rank (reduced # of writs)	−12.5%

Table 6: Counterfactual treatment effects.

priority G1 at some point, and those who aren't. Changes in the relative ranking within the group of tax-payers ultimately assigned priority G1, and within the group of tax-payers never assigned priority G1 matter less. In contrast, implementing PIE using a uniformly random rank would revenue losses of -12.5% compared to the control group.

The takeaway is that ranking matters, but can be considerably simplified at little loss of efficiency.

7 Conclusion

We argue that PIE offers governments with limited capabilities an effective way to enforce tax-collection. We show in a benchmark model that PIE is well suited to deal with issues of incomplete information (see Online Appendix OB for an extension to the case where the amount of taxes due D_i is private), and some forms of bounded rationality.

Field evidence helps refine our understanding of threat-management, and of the optimal sizing of threats. Credible high-powered threats, targeted to a small share of tax-payers, can be effective, but only if they lead to a sufficient increase in the speed of settlement. Delay in best-response implies that in contrast to our theoretical benchmark, issuing threats to a number of tax-payers no greater than available capacity (making settlement dominant) may cause losses of revenue.

Nonetheless, our data suggests that the elements needed for PIE to be effective are present: high-powered targeted threats increase tax-payers' settlement intensity, and repayment behavior can be predicted. In addition, we show that seemingly ineffective collection actions, i.e. legal writs not followed by actual garnishment, have a meaningful impact on settlement. In our context, we estimate the impact of PIE implemented at scale on collection to be on the order of 11%.

A limit of our work is that our data comes from a single implementation partner. We believe that our findings justify experimenting with PIE in a broader range of settings. Further design improvements may be available. For instance, it is possible that versions of PIE operating within smaller groups of tax-payers would be more effective in encouraging forward-thinking planning by tax-payers in priority group G2.

Finally, we hope that our work stimulates the evaluation of divide-and-conquer methods as a tool to leverage limited enforcement capacity in the field. Possible applications include debt collection, policing, and organizational change.

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

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

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

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

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

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

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 $\delta_i \equiv \mathbb{E}\left[\frac{P_i}{D_i}\widehat{s}_i\middle|m_i=1\right] \in [0,1]$. In equilibrium, expected collection is equal to $\sum_{i=1}^N \delta_i(1-q_i)D_i$. Condition (B.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,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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,

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

We condition on the event that (B.4) holds, and iteratively consider batches of tax-payers $\{0, \cdots, n_0\}, \cdots, \{n_{k-1}+1, \cdots, n_k\}, \cdots, \{n_{K-1}+1, \cdots, n_K\}$ at times $k/(K+1)$. We show that by time $(k+1)/K$, it is obviously dominant for tax-payers $\{n_k+1, \cdots, n_{k+1}\}$ 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 tax-payers 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 tax-payers in group $\{n_k+1, \cdots, n_{k+1}\}$ are informed that the collection capacity expended on tax-payers 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 tax-payers 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 tax-payers 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 $(k+1)/(K+1)$ tax-payers 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, tax-payers 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 $\sum_{i \leq i^*} q_i \lambda_i \geq \alpha N$. ■

C Experimental Materials

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

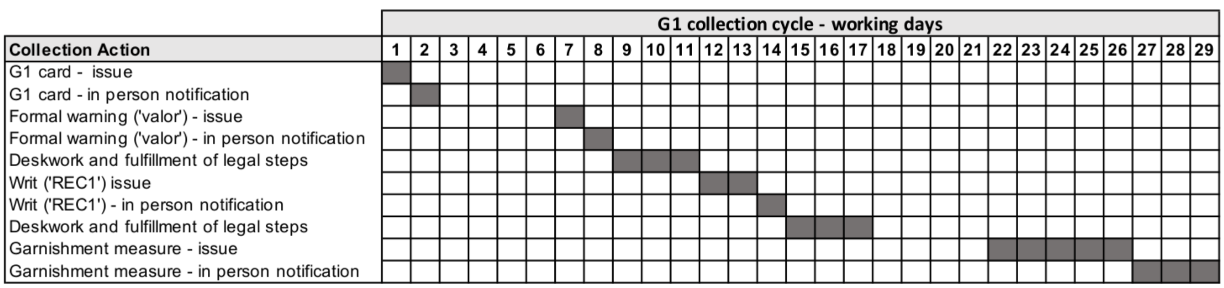


Figure C.1: Schedule of collection actions for G1 tax-payers

Tables C.1 and C.2 illustrate the information letters sent to tax-payers in priority groups G2 and G3. Spanish original of all letters sent are reported in Online Appendix OD.

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*1.1** + **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 C.1: Information letter, priority group 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*1.1** +**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 C.2: Information letter, priority group G3

D Tables and Figures

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

	Control	Treatment	Difference		
			Mean	S.E.	p-value
Total tax due	374.5	377.5	-3.00	18.33	0.87
Property tax due	138.1	129.6	8.49	9.09	0.35
User charges due	236.4	247.9	-11.49	13.23	0.39
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	58.13	57.61	0.52	0.31	0.09
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: *Total tax, property tax* and *user charges* for 2021 due as of April 5, 2021 (in Peruvian S/). *Last year repayment share* is a dummy taking value 1 for taxpayers whose share repaid in 3 months after the deadline was above 20%. *Exo. and Endo. scores* are the optimal 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 D.1: Summary statistics by treatment status

References

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

ALLINGHAM, M. G. AND A. SANDMO (1972): "Income tax evasion: A Theoretical Analysis," *Journal of Public Economics*.

BALÁN, 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.

BERGERON, A., G. TOUREK, AND J. WEIGEL (2021): "The state capacity ceiling on tax rates: Evidence from randomized tax abatements in the DRC," .

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.

CHEN, D. L., M. SCHONGER, AND C. WICKENS (2016): "oTree—An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, 9, 88–97.

CHERNOZHUKOV, V. AND H. HONG (2003): "An MCMC approach to classical estimation," *Journal of Econometrics*, 115, 293–346.

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*.

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.

Foreman-Mackey, D. (2016): "corner.py: Scatterplot matrices in Python," *Journal of Open Source Software*, 1, 24.

Gneezy, U. and A. Rustichini (2000): "A fine is a price," *The journal of legal studies*, 29, 1–17.

Halac, M., I. Kremer, and E. Winter (2019): "Raising capital from heterogeneous investors," *American Economic Review*.

Halac, M., E. Lipnowski, and D. Rappoport (2020): "Rank Uncertainty in Organizations," *Available at SSRN 3553935*.

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.

Kapon, S. (2022): "Interactive Experiments on Amazon Mechanical Turk," Tech. rep., Princeton University Working Paper.

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.

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

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.

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