

Targeting in Tax Compliance Interventions

Experimental Evidence from Honduras

Giselle Del Carmen

Edgardo Enrique Espinal Hernandez

Thiago Scot



WORLD BANK GROUP

Development Economics

Development Impact Evaluation Group

March 2022

Abstract

Tax authorities often use low-cost communication with taxpayers to encourage voluntary compliance and avoid other costly interventions. This paper reports findings from an experiment with more than 30,000 taxpayers in Honduras, designed to assess how taxpayers with different risk scores respond to a communication intervention. Across several outcomes, the average effect of the intervention on compliance was null. Contrary to the expectation of experts

surveyed, only taxpayers considered to be at low risk of noncompliance increase their filing and reported income. Using rich administrative data and a causal forest algorithm, the paper finds that ex-ante predicted risk and responsiveness to the intervention are negatively correlated. These findings can inform the design of targeted interventions by tax authorities.

This paper is a product of the Development Impact Evaluation Group, Development Economics. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/prwp>. The authors may be contacted at tscot@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Targeting in Tax Compliance Interventions: Experimental Evidence from Honduras

Giselle Del Carmen

Edgardo Enrique Espinal Hernandez

Thiago Scot*

Keywords: Tax compliance, Tax evasion, Firms & Productivity, Targeting **JEL**

Codes: H26, O17

*Development Impact Evaluation (DIME), World Bank. Corresponding author at tscot@worldbank.org. Giselle del Carmen (Enodo): g.e.delcarmen@gmail.com. Edgardo Enrique Espinal Hernandez (Servicio de Administración de Rentas (SAR) de Honduras): eespinal@sar.gob.hn. We acknowledge the collaboration of the team at the Honduran Tax Authority (SAR), in particular Oziel Aron Fernandez Herrera, David Fernando Pineda Pinto, Gabriel Ricardo Perdomo Zelaya, Wilman Alonso Ponce Nuñez, Milton Alexander Maldonado Castillo, Pedro Rafael Zuniga Figueroa, Roldan Manuel Enamorado Irias, Carlos Mauricio Zelaya Guzman, Ximena Fernanda Venis Dilworth, Mario Josué Ramos Canales, Fabián Alfredo Gómez Guillen. This project also benefited from comments from Marta Ruiz-Arranz, Martin Ardanaz, Jordi Prat Cordero, Ricardo Perez-Truglia, Reed Walker, Guo Xu, Steve Tadellis, Ernesto Dal Bo, Frederico Finan, Muhammad Yasir Khan, Petr Martynov, Arianna Legovini and participants at the UC Berkeley Development Lunch. The main email experiment was submitted to UC Berkeley Internal Review Board (IRB) under ID 2019-12-12772 and was deemed "not Human Subject Research". The forecast survey was submitted to UC Berkeley IRB under ID 2020-03-13079 and approved on March 25th 2020. This study has been enrolled in the AEA Trial Registry (RCT ID: [AEARCTR-0005285](#))

1 Introduction

Tax collection in low-income countries is very different from what is observed in higher income settings since state capacity determines the ability of governments to enforce taxation. Tax revenue is equivalent to about 16% of GDP in OECD countries but less than 12% in low-income ones. The composition of collected taxes is also very different: where high-income countries collect over 50% of their total taxes through income taxes, low- and middle-income countries collect less than 30% and rely much more on taxes on goods and services.

Collecting income taxes can be particularly hard for governments given the complexity involved in correctly assessing liabilities (Gordon & Li, 2009). However, the capacity to obtain and process the type of information needed to assess income tax liabilities is rapidly increasing in several low- and middle-income countries. International organizations are committed to improving tax capacity: approximately \$200 million in Official Development Assistance (ODA) was aimed at improving tax capacity in 2014 (International Monetary Fund et al., 2016). Nonetheless, these improvements in "government intelligence" and access to technology do not translate automatically into higher compliance but require other actions by the authorities (Okunogbe & Santoro, 2021). For example, communicating to taxpayers the new information available to the tax authority might trigger higher compliance if their beliefs about the probability of punishment are updated.

In this paper we partner with the Tax Authority in Honduras (SAR, for *Servicio de Administración de Rentas*) to experimentally estimate the impact of providing information to taxpayers about SAR's knowledge on their transactions. Using a recently developed risk model, we conduct an experiment with approximately 32,000 taxpayers considered to be at-risk of noncompliance. Taxpayers in the control group receive a regular reminder about the income tax filing deadline – a usual communication provided by the tax authority. In contrast, those in the treatment group are informed that the tax authority observes specific transactions they engaged in, implying knowledge about (part) of their realized revenue. In that sense, we see our experimental treatment as providing taxpayers with data about the information set available to the tax authority. That allows us to estimate the average effect of treatment assignment on a range of compliance outcomes, measured using administrative data on tax filing.

While there is strong evidence that the average effect of these types of interventions is positive (see Mascagni (2018) and Castro & Scartascini (2015) for recent reviews), this is not inconsistent with negative or null impacts on some subset of taxpayers (De Neve et al., 2021). We argue that measuring this heterogeneity is crucial for authorities to better target this kind of intervention. While the return on investment of these interventions is often perceived to be extremely high given the very low marginal costs of letter or

email communications, a less discussed potential hidden cost is the credibility of the tax authority. When capacity is limited, it is often not possible to follow-up on this type of low cost communications by performing more expensive interventions, such as in-person audits. This poses a reputational risk: revealing knowledge about potential misreporting and not acting might lead taxpayers to believe (perhaps accurately) that the expected punishment for non-compliance is low (Slemrod et al., 2001; Gangl et al., 2015)¹.

We present heterogeneity exercises of the treatment effects on key dimensions built into the experimental design, but also explore the rich data on past behavior of taxpayers to predict which subjects are most responsive to the intervention. This is a classic targeting problem. Our goal is to predict, upon observing characteristics of taxpayers, what is their expected response to a communication intervention such as the one implemented in this study. In order to do so we borrow tools from the recent literature on causal machine learning and use a causal forest algorithm to estimate conditional average treatment effects (CATE) across the sample of taxpayers (Athey & Imbens, 2016; Wager & Athey, 2018). With those estimates at hand, we can assess how the treatment effects compare to the perceived risk level of each taxpayer, computed by the tax authority using non-experimental data from before the intervention. This is informative about future communication interventions since currently those with higher assessed risk are more likely to be targeted, but it is not clear whether these are the most responsive taxpayers².

Our main finding is that, on average, the intervention had no effect on a range of compliance outcomes. Taxpayers assigned to receive the treatment were no more likely to file income taxes, declare higher revenues, or declare higher taxable income. We estimate precise zero effects: we can rule out changes in filing probability as small as one percentage point and increases in declared revenue of more than 5%. To benchmark our results, we surveyed experts before the intervention and elicited their predictions about our experimental findings. Our estimates are significantly lower than experts' forecasts. The estimated null effects are observed both for the pooled treatment sample and for individual treatment arms that included distinct framing messages for compliance. We also show that the intervention did not increase the timeliness of tax filing or other tax compliance measures such as declared VAT sales or corrections of past filings.

The heterogeneity analysis also suggests that, across a range of dimensions, the treatment effect was mostly null. The only consistent positive treatment effect across our primary outcomes seems to be among taxpayers with low ex-ante assessed risk level. Compared to other low-risk taxpayers in the control group, those receiving the email

¹But see Bergolo et al. (2019) for evidence that these types of interventions might function as "scare-crows", generating emotional reaction that leads to compliance rather than changing the beliefs of taxpayers about enforcement capacity.

²In appendix D, we present a simple model of heterogeneous treatment effects in tax compliance and illustrate the value of additional experimental information on targeting.

with information about their revenues available to the tax authority were more likely to file taxes (+ 2.3 p.p.) and declared higher gross revenue (+ L103,000³ or 8% of the control mean). The increase in gross revenues, nonetheless, was accompanied by an increase in claimed deductions (+ L92,000) resulting in a modest increase in taxable income (L10,000 or 11% of the control mean). The offsetting increase in deductions is consistent with previous findings from [Carrillo et al. \(2017\)](#) in a similar setting in Ecuador.

The finding of significant impact only for taxpayers with lower perceived risk-level is contrary to what the experts we surveyed expected. As previously mentioned, it is also at odds with the current practice of the tax authority, which focuses communication interventions on medium- and high-risk taxpayers. Using the causal forest model, we predict treatment effects conditional on observed covariates for all taxpayers and show a similar finding: the correlation between (conditional) treatment effects and a continuous measure of risk-level assessed by the tax authority is negative and the average treatment effect across taxpayers is only positive for those in the low- and medium-low risk groups.

These results suggest that, while models to predict risk of non-compliance are an important tool for tax authorities, they are not sufficient to determine targeting. Governments face a menu of policy tools to assure compliance, ranging from "soft" communication encouraging truthful reporting to "hard" audits. Experimental evidence on the causal impact of different tools for distinct groups of taxpayers can help governments tailor interventions to maximize impact.

One important caveat, discussed in more detail throughout the paper, is that our intervention happened at the onset of the Covid-19 crisis in Honduras. Two weeks after we contacted taxpayers, the country entered a severe lock-down that depressed economic activity and the filing deadline for income taxes was twice postponed. These facts suggest that the intervention was likely much less salient in the mind of taxpayers. Furthermore, compliance in general was clearly affected by the crisis; 86% of taxpayers in our sample filed taxes in 2019 but only 62% did in 2020. This implies that the external validity of our findings should be taken with caution.

Our study contributes to the now large literature on communication interventions aimed at increasing taxpayers compliance, reviewed by [Mascagni \(2018\)](#) and [Hallsworth \(2014\)](#)⁴. Our experiment is particularly informative about exploring third-party information in communication interventions, similar to the work of [Brockmeyer et al. \(2019\)](#) in Costa Rica and [Carrillo et al. \(2017\)](#) in Ecuador. In addition to informing taxpayers about SAR's knowledge of their transactions, we also randomly vary how we frame the importance of compliance. One third of the treatment group was reminded that non-compliance carries monetary penalties; one third was reminded that SAR can deny

³Approximately USD 4,000.

⁴For recent studies on similar interventions in low-income countries, see for example [Collin et al. \(2021\)](#), [Cohen \(2020\)](#), [Okunogbe \(2021\)](#) and [Hoy et al. \(2021\)](#).

documents necessary for their operations; and the remaining group received a message appealing to tax morale, claiming that "with your taxes we build a better country". The effects of appealing to tax morale (Luttmer & Singhal, 2014) in similar interventions have so far been mixed. De Neve et al. (2021) find that appeals to the social norm of payment or the importance of public goods have null or negative effects across the compliance spectrum in Belgium, while Castro & Scartascini (2015) find null effects in Argentina for local property taxes. Kettle et al. (2016), on the other hand, find that informing late filers in Guatemala that they were in a minority of non-compliers was as effective as sending a threatening message. Although the exact treatments are not the same, it is somewhat surprising that appeals to tax morale are more effective in Guatemala than in Argentina or Belgium, given low levels of confidence in the government: according to Latinobarometro, only 22% of Guatemalans answer having some or a lot of confidence in the government between 2015-2018, lower than the 30% reported by Argentinians. The experts we surveyed systematically predicted that the tax morale treatment would likewise be less effective than the other two. In our main estimates all three treatment arms are estimated to have null effects on compliance and we cannot reject that the effect of the tax morale arm is equal to the other interventions.

Finally, our exercise using the causal forest algorithm⁵ contributes to a recent literature exploring heterogeneity analysis in experimental settings to inform program targeting, ranging from cash and asset transfers to the poor (Alatas et al., 2012) to the use of cellphones to monitor agricultural extension workers (Dal Bó et al., 2021). In a similar exercise but in a different setting, Hussam et al. (2020) worked with experimental grants to entrepreneurs in India, and showed that the causal forest algorithm can do no better in predicting returns to the program than assessment by other community members. We compare our treatment effect estimates with the baseline risk assessment of each taxpayer, showing that experimental results are a useful tool for the tax authority to improve targeting.

2 Context - Taxation in Honduras

Honduras is a lower middle-income country in Central America with GDP per capita of \$5,800 PPP in 2018. Similar to other countries with comparable income levels, the country collects less than 20% of GDP in taxes and is very reliant on taxes on goods and services, making up over 50% of total tax collection (International Monetary Fund, 2018).

The fiscal year runs from January 1st to December 31st and taxpayers must file an income tax declaration by April 30th of the following year. Corporations are taxed at

⁵See Davis & Heller (2017) and Bertrand et al. (2017) for detailed discussions on the application of causal forest algorithm in experimental settings.

a flat rate of 25% on profits (gross revenues minus cost deductions)⁶. Non-incorporated taxpayers face a progressive tax schedule: in FY2019, net income below L159,000⁷ was exempt from taxation and higher incomes are taxed with increasing marginal tax rates in three brackets of 15%, 20% and 25%⁸. All taxpayers with commercial activities (both corporations and non-incorporated) are also liable for monthly sales taxes of 15% over total goods and services sold. Individuals whose sole income sources are wages or capital income (dividends, interest) are withheld at source and do not have to file an income tax declaration. All taxpayers in our experiment, therefore, engage in some type of commercial activity that generate revenues that must be declared.

Between 2014-2019, the Honduran Tax Administration underwent a series of reforms and institutional changes aimed at strengthening the country’s fiscal system. These included improvements in operational management, recruitment of personnel, a new billing regime and the adoption of new technologies for data processing, among others. Since then tax revenues increased from 15% of GDP to over 18%, contributing to a decrease in the fiscal deficit of nearly 6 percentage points of GDP (7.9% in 2013 vs. 2.1% in 2018) ([International Monetary Fund, 2018](#)). Furthermore, SAR has been working to consolidate these institutional reforms and implement new tools to ensure fiscal compliance, including an internal risk model following international best practices.

3 Research design

3.1 Intervention

Jointly with the Tax Authority, we sent emails to approximately 32,000 taxpayers seven weeks before the original income tax filing deadline for FY2019⁹. Since our main goal is to estimate the effect of the content of emails, we contacted all taxpayers in the experimental sample, including those in the control group. This allows us to attribute any differential behavior among treated units to the content of messages and not to the simple fact of receiving a message from the government¹⁰.

We summarized our randomization design in [Figure 1](#). The control group received

⁶Since 2003, corporations must also assess a 1% tax on their net assets over L3 million and pay the largest amount between asset and income taxes. In our experimental sample only 10% of corporations (3% of all taxpayers) paid the asset tax in FY2018.

⁷Approximately USD6,400 based on an average exchange rate of 25 Lempiras per US dollar for 2018.

⁸The law also allows for a L40,000 deduction of medical costs. In practice this deduction is applied to all non-incorporated taxpayers, regardless of claiming the deduction, such that the exemption threshold is higher.

⁹Due to the Covid-19 crisis, the tax filing date was postponed from April 30th to August 31st. More details on how we dealt with these changes are provided below.

¹⁰The tax authority sends on average 200,000 emails every month to taxpayers, notifying them about various fiscal procedures. Taxpayers in our experimental sample only received the experimental email in the weeks before and after the intervention.

an email¹¹, presented in [Figure 8](#), with a reminder about the filing deadline and the importance of truthfully reporting their tax liabilities¹². It also included a link to the website of the tax administration where detailed information on how to declare taxes online was provided.

Taxpayers in the treatment group received emails containing the same informational content offered to control units, but were additionally provided with i) information available to the tax authority regarding their transactions and ii) slightly different framing messages on why they should comply with their obligations. For the majority of our experimental sample (71%) some third-party information on their transactions in FY2019 was available and that fact is included in the message. For the remaining taxpayers, no third-party information was available but their previous reporting behavior raised flags about non-compliance; that is the information included in their email. Since our treatment group includes three arms with different framings on the importance of compliance, we have in total six different types of messages, illustrated in [Figure 9](#) through [Figure 14](#)¹³.

Taxpayers for which third-party information was available ([Figure 9](#) through [Figure 11](#)) received a message informing that "In the sources of information available to the Tax Administration, your following commercial transactions for FY2019 period have been identified", followed by up to four types of transactions: sales to other taxpayers; sales through debit/credit cards; sales or services to the State; and exports¹⁴. These messages are personalized, so that each taxpayer is only informed of categories for which the TA observes their transactions (i.e. there is no deceit or bluffing involved).

Taxpayers flagged for risk of non-compliance but for whom third-party information was not available were notified ([Figure 12](#) through [Figure 14](#)) that "In the sources of information available to the Tax Administration, the following behavior has been identified in your tax returns", followed by up to three "anomalies" in their past filings: declaring three or more years with losses in the previous five fiscal periods; financial transactions incompatible with declared revenue; and declared tax liability "atypical" for tax units in similar industries and revenue size.

As seen in the messages, we have three different treatment arms in which we change a

¹¹All emails, for treatment and control taxpayers, included the same subject "Important: Notice of Tax Obligation" and were sent from an institutional email address used by the tax authority.

¹²The main part of the email reads, in English: "The Revenue Administration Service (SAR) reminds you that the obligation to file and pay the Sworn Declaration of Income Tax period 2019 expires on April 30, 2020. You are reminded that the Declaration must contain exact and truthful information, reporting all income obtained and that deductions will have to be supported by valid tax documents."

¹³To further strengthen the intervention, the content of messages were informed by previous experiments using insights from behavioral economics ([Dalton et al., 2019](#)), such as making the text simpler, personalized to each taxpayer and including actionable information (link to SAR's website). The messages were also analyzed in focus group discussions with SAR officials (including communication experts) and policymakers in Honduras.

¹⁴Unlike [Brockmeyer et al. \(2019\)](#) or [Carrillo et al. \(2017\)](#), the emails did not include monetary values on specific transactions or information on trading partners due to legal restrictions.

small part of the email, emphasizing different reasons why taxpayers should comply with their tax obligations.

Sanctions treatment: these messages emphasize the sanctions associated with non- or late-filing, by stating that "In case of not fulfilling your obligation, you will be subject to the sanctions established by the Tax Code in Articles 160 and 163."¹⁵ These are similar to other "threat" messages used in the literature, explaining or making more salient to subjects the monetary costs of non-compliance.

Procedure denial treatment: this treatment arm also includes a threatening message, but instead of mentioning possible fines it invokes the right of the TA to withhold important documents necessary for business' operations in case they are non-compliant. The additional message reads "In case of not fulfilling your obligation, you will be affected in obtaining proofs of "pagos a cuenta", solvency and fiscal documents"¹⁶.

Tax morale treatment: in this treatment arm the email contains two pieces of additional content: a motto upfront stating "For you, for your kids, for Honduras, pay your taxes!" and a paragraph stating "The Honduras we all want for our children with education, health, infrastructure and security is the fruit of the efforts of all its good citizens, thanks to their taxes we build a better country". This message appeals to the fact that taxes are used to finance public goods and it is the duty of "good citizens" to pay their taxes.

All emails were sent between March 11th-12th 2020, approximately seven weeks before the original FY2019 tax filing deadline. Due to the Covid-19 crisis and the strict lockdown implemented in the country, however, the filing date was initially postponed to June 30th and subsequently to August 31st. The email service used by the tax authority to send mass communications allows us to observe the "outcome" of every email sent, including whether they reached the taxpayers' mailbox and if they ever opened it. We consider taxpayers to be compliant, i.e. effectively receiving the treatment, if they ever open the email.

¹⁵The two articles mentioned determine fines for non-presentation or late presentation, as well as non- or late payments of tax obligations (see <https://www.sar.gob.hn/leyes/>).

¹⁶"Pagos a cuenta" refers to a special regime in which clients do not have to withhold income taxes on services offered by independent professionals - this is a benefit to these professionals since it preserves cash flow until the tax payment date. "Solvency" is a statement by the TA that the taxpayer is up to date with their obligations and is required to perform transactions with state entities. "Fiscal documents" is understood to be fiscal receipts, which need to be approved and provided by the tax authority so that firms can legally issue them.

3.2 Experimental sample and randomization

The experimental sample consists of 31,396 taxpayers considered to be at-risk of non-compliance in FY2018. The risk was assessed using SAR's internal risk model (*Modelo de Gestión de Riesgo de Honduras*, MGR-H)¹⁷ that considers both discrepancies between declared income by taxpayers and information reported by third-parties, as well as anomalies, defined as outcomes that seem inconsistent with other similar tax units. Each taxpayer is assigned a score between zero and one that determines the predicted level of non-compliance risk¹⁸. Based on the risk management model, SAR determines which treatment actions should be implemented to mitigate risks and prioritizes the allocation of enforcement resources to the high-risk taxpayers.

The tax authority uses five main sources of information on the income of taxpayers provided by third-parties. First, the Monthly Declaration of Purchases (*Declaración Mensual de Compras*, DMC) is an informative declaration filed monthly by large taxpayers. They can use declared purchases from other registered taxpayers as credits against their liabilities on sales taxes (effectively a VAT system). The Monthly Declaration of Withholding (*Declaración Mensual de Retenciones*, DMR) is also filed monthly by taxpayers designated as "withholding agents", such as firms that retain and pay income taxes of their employees. Third, the Declaration of Credit Card Administrators (*Declaración de Retenciones de las Administradoras de Tarjetas de Débito y Crédito*, ATC) is filed by credit and debit card companies about point-of-sales purchases using their system. Credit and Debit Card administrators are also withholding agents, paying a share of sales taxes due in each transaction. Finally, two sources of third-party information are provided by other government agencies: the Integrated System of Financial Administration (*Sistema de Administración Financiera Integrada*, SIAFI) provides information on all revenue made by sales to government entities, and all export sales are also informed to the tax authority.

The risk model also performs a series of risk analyses in the absence of third party information. These include, but are not limited to, flagging taxpayers reporting repeated losses and performing cluster analyses that group "similar" units and flag those with reported outcomes (such as tax liabilities) that are inconsistent with their peers. The tax authority also has access to information about financial transactions such as loans, and uses that to flag taxpayers that declare revenues inconsistent with their financial activities.

¹⁷The model follows the ISO 31000 risk management standard and international best practices regarding fiscal procedures as established by the International Monetary Fund (IMF), Organization for Economic Cooperation and Development (OECD) and the Center Inter-American Tax Administrations (CIAT).

¹⁸The model uses risks of non-compliance identified throughout the life-cycle of taxpayers (registration, presentation, payment and truthfulness) in order to maximize tax compliance. It combines probability variables (frequency with which the risk occurs) and consequence (materiality or economic damage caused by the risk).

Starting from a broader set of at-risk taxpayers, we performed two restrictions to obtain the final experimental sample. First, in order to avoid spillovers between treatment and control units, only taxpayers with a unique primary email address were included in the experimental sample¹⁹. Second, power calculations exercises suggested we could significantly increase minimum detectable effects (MDE) by dropping extremely large taxpayers. [Figure A5](#) reports MDE for our four main outcomes of interest when we trim our sample at different percentiles. Considering all primary outcomes, we decided to trim the sample at the 97th percentile of declared revenue distribution, arriving at our final sample of 31,396 taxpayers²⁰.

We implement a stratified randomization, at the taxpayer level, using 60 strata defined by whether third-party information was available or not; whether the taxpayer was a corporation; municipality of operations defined as Distrito Central (capital Tegucigalpa), San Pedro Sula (second largest city, often referred as the industrial capital) or other; and five risk levels as defined by the tax authority ($2*2*3*5 = 60$ strata). In each strata 49 percent of taxpayers were allocated to the control group and the remaining 51 percent in three equally sized treatment arms. Following [Bruhn & McKenzie \(2009\)](#) we deal with "misfits" (remaining taxpayers in each strata) by randomly assigning them to one of the four groups (control + 3 treatments) using the above weights as assignment probabilities²¹.

3.3 Benchmarking results: Expert forecast survey

In order to benchmark the magnitude of our results, we follow [DellaVigna & Pope \(2018\)](#) and [DellaVigna et al. \(2020\)](#) and surveyed academics and policymakers on their predictions about the impacts of this experiment. We focused on collecting forecasts from three groups of subjects: academic economists (faculty, PhD students and researchers at academic institutions); public sector workers in Honduras, in particular those working at the tax authority; and policymakers or researchers in international development organizations²².

The main results of the survey are presented in [Table 7](#). In terms of magnitudes of

¹⁹Approximately 4,400 taxpayers were also deemed at-risk of non-compliance but shared a primary contact email with other units, either due to joint ownership or an accounting firm as primary contact. These taxpayers were excluded from the main sample and were assessed in a separate, smaller experiment to estimate spillover effects.

²⁰Details about power calculations used in [Figure A5](#) are discussed in [Appendix C](#). Taxpayers with declared gross revenue above L19.4 million (approximately USD 780,000) in either FY2017 or FY2018 were excluded from the sample

²¹The randomization was implemented in Stata using the *randtreat* command to deal with misfits.

²²We sent the invitation to take the survey to approximately 120 academics, 100 practitioners and 50 public sector workers. We received 111 partial survey answers and dropped 35 incomplete surveys and 11 with mostly inconsistent answers. Our final sample consisted of 65 complete surveys. Details about the exclusion process, further trimming of outliers and complete survey questionnaires are provided in [Appendix E](#).

the treatment effects, experts predicted the treatment would induce an increase in filing probability of 4 p.p. among compliers, an increase of 7% in declared gross revenues and of 6% in declared taxable income. Three additional patterns in experts' forecasts are worth emphasizing. First, experts consistently predict the Sanctions and Procedures treatment arms to be more effective than the tax morale treatment. This is true across the different outcomes and, as seen in [Figure 7](#), across different categories of respondents. The average predicted effect on filing probability, for example, is 6.7 p.p. for the Sanctions arm and only 1.6 p.p. for the tax morale arm. The survey questionnaire was explicit in stating that different treatment arms included only an additional message to encourage compliance, but the main treatment was always the provision of information about taxpayers' transactions. This implies that respondents either believed the main treatment would come from the encouragement messages or that the tax morale message would have a negative effect, diluting the impact of the information provision.

Second, we also collected forecasts about heterogeneous treatment effects across risk-levels defined by the tax authority, directly eliciting perceptions of the relationship between ex-ante assessed risk-level and response to the intervention. Experts predicted how the impact on reported gross revenue would vary across the distribution of risk. Though results are noisy, on average the predictions pointed to a larger effect among high-risk taxpayers: experts predicted an increase of 9% in reported gross revenue among the top risk group vs. 6% for the lowest risk group. This expectation lines up with the current strategy of the tax authority of focusing communications to high-risk taxpayers, whether this holds or not is an empirical question. As discussed in the text and formally exemplified in the stylized model, high-risk taxpayers might be less reactive to such "soft-touch" interventions and therefore present a smaller treatment effect.

Finally, we note that academic economists are often less sanguine than both practitioners and, particularly, government employees when it comes to the magnitude of effects. With the exception of the tax morale treatment arm, which is deemed to have small effects by all experts, academic economists often predict smaller effects on average.

4 Data

4.1 Baseline descriptives and balance

We present baseline descriptive statistics of our experimental sample in [Table 1](#). As previously mentioned, all taxpayers in the study have some type of commercial activity. One-third of taxpayers are corporations, 40% are individual business (non-incorporated firms) and 6% are self-employed service providers (often professionals like lawyers or

doctors)²³. Half of all taxpayers are located in the two largest municipalities in the country, Distrito Central and San Pedro Sula.

On Panel B we present descriptive statistics of variables related to past tax filing. The majority of taxpayers were flagged to be at-risk for under declaring tax liabilities and not for non-declaration: 86% of the sample filed an income tax declaration for FY2018. Conditional on declaring, average gross revenue was L1.5 million (USD 60,000) while median gross revenue was L381,000 (USD 15,200) - even after excluding outliers the gross revenue distribution presents a long right-tail. Almost 50% of taxpayers were not liable to pay income taxes in FY2018, either because they declared taxable income below the minimum threshold to pay taxes (for non-incorporated entities) or because they declared losses. Conditional on declaring, the average taxable income was L115,000 (median = L41,000) and the average tax liability L15,000.

We also provide information on the indicators used by the tax authority to assess risk on Panel C. Third-party information on revenues is available for 71% of the experimental sample and, among those, 88% are informed on sales to other parties, 22% by credit/debit cards (point-of-sales, or POS) operators; 3% by the government and 1% by customs authority²⁴. Considering anomalies, almost half of our experimental sample is flagged for having "atypical declared revenues" when compared to peers. A much smaller share is flagged for declaring three or more years with losses in the last five fiscal periods (8%) or having financial transactions inconsistent with declared revenues (7%). These indicators, among others, are then aggregated by the TA in a global "risk factor", which is used to classify every taxpayers in five risk levels. Our sample is fairly evenly divided among the four lowest levels, with only 10% deemed to be "high-risk" by the tax authority's risk model.

In Table 2 we present balance tests between our control and treatment samples. Columns (1) and (2) present the mean and standard deviation of each variable in the control group, respectively. In Column (3) we present the difference in means between control and the pooled treated sample, and test whether we can reject the null hypothesis of equal means. Overall our sample is balanced and the few statistically significant differences are very small in magnitude. Columns (4) through (6) present differences between each of the treatment arms and the control group, and again indicate balance in observables²⁵.

²³The remaining 21% are not incorporated and also not registered as individual businesses or service providers. The nature of their transactions suggest these are mostly small businesses that never officially registered as such.

²⁴These are the categories used to fill in the message sent to taxpayers and they are not exclusive: some taxpayers are informed in these four categories.

²⁵We do not include in the balance tables variables used to stratify the sample, since they are balanced by construction.

5 Results

5.1 Primary outcomes

To obtain Intention-to-Treat (ITT) estimates on the effect of our experimental intervention on compliance, we estimate regressions of the following form:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \gamma_s + \epsilon_i \quad (1)$$

where Y_i is one of the four primary outcomes of interest in FY2019 (indicator for tax filing, amount of gross income, amount of deductions and amount of taxable income); T_i is a dummy that takes value 1 if the unit was assigned to treatment and 0 otherwise; X_i are baseline controls and γ_s are strata fixed-effects. Baseline controls include a dummy for presenting tax declaration in FY2018; amount of declared gross revenue, third-party informed gross revenue and declared taxable income in FY2018; and amount of declared sales tax revenues in FY2019. Our main coefficient of interest is β_1 , measuring the difference in mean outcomes between units in treatment and control groups. Reported outcomes for non-filing taxpayers are considered to be zero. While our continuous outcome variables are highly skewed, we control for previous FY outcomes to reduce residual variance and diminish the influence of outliers. We do not treat outliers in our main specification but perform robustness tests using winsorized outcome variables at the 99th percentile.

To account for partial compliance, we also present instrumental variable regressions (Local Average Treatment Effect or LATE estimates) where we instrument opening the email (i.e. receiving the actual treatment) with treatment assignment. We estimate regressions of the form

$$Y_i = \delta_0 + \delta_1 Open_i + \delta_2 X_i + \gamma_s + v_i \quad (2)$$

$$Open_i = \lambda_0 + \lambda_1 T_i + \lambda_2 X_i + \gamma_s + \mu_i \quad (3)$$

where $Open_i$ is an indicator of whether the taxpayer opened the email sent to them and the remaining variables are the same as above. The parameter of interest is δ_1 , which measures the LATE on taxpayers that complied with treatment assignment.

Our main results are presented in [Table 3](#), where Panel A shows the ITT estimates and Panel B LATE estimates. Across our four primary outcomes, we estimate null treatment effects. The point estimate for change in filing probability (Column 1) is -0.1 p.p. and we can rule out effects as small as 1 p.p. change in the treatment group. Similarly, the estimates for treatment effects on declared gross revenue, deductions and taxable income are all not statistically different from zero and small in economic magnitude (less than 1% of the control mean in each case). In [Figure 2](#) we present the cumulative distribution

functions of reported gross revenue for treatment and control groups, as an illustration. We observe no meaningful difference between the two distributions, consistent with the null treatment effect estimated. Since, as previously discussed, just over one third of taxpayers assigned to the treatment group actually opened the email, the LATE estimates are approximately three times as large as the ITT but similarly not statistically different from zero. We therefore cannot reject the hypothesis that the intervention had an overall null effect on the compliance behavior of taxpayers.

While the previous results pool all taxpayers assigned to the three different treatment arms together, we also estimate separately the effect of each of the treatment arms, augmenting [Equation 1](#) to estimate ITT as:

$$Y_i = \alpha + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + \beta_2 X_i + \gamma_s + \nu_i \quad (4)$$

where T_j are the dummy indicators for each of the treatment arms and we are interested in the coefficients $\beta_1, \beta_2, \beta_3$. Similarly, we estimate LATE using the different treatment arms and present both ITT and LATE in [Table 4](#). We again cannot reject the null hypothesis of zero treatment effect. ITT are very small across the treatment arms and never statistically significant²⁶. While LATE estimates are somewhat larger in magnitude they are similarly imprecisely estimated.

5.2 Secondary outcomes

While the four outcomes listed above are of primary interest, we also pre-specified evaluating the impact of the email intervention in other dimensions of compliance with tax obligations. As discussed by [Brockmeyer et al. \(2019\)](#), it is an empirical question whether interventions focused on increasing tax compliance on a specific dimension (truthful income tax declarations) will affect other behavior. On the one hand it is possible that the intervention shifts taxpayers' beliefs about the tax authority's capacity, increasing their perception of risk and inducing more compliance across all their obligations. On the other hand, if the intervention is seen as signaling an increased oversight on a narrow dimension, taxpayers might increase compliance on that specific dimension while decreasing compliance in others. Since we observed a zero treatment effect on the primary targeted outcomes, it is no surprise we also estimate null effects across the board for secondary outcomes.

²⁶In [Table A1](#) we present similar estimates trimming outcomes at the 99th percentile. We emphasize that this specification, unlike other estimates in the paper, was not pre-specified. Overall we find similar results: point estimates are small in magnitude, often negative and not statistically different from zero. The only exception is the estimate for the impact of the Moral Duty treatment on taxable income, which is positive (L3,460 or approximately 5% of the control mean) and significant at the 5% level. While estimates for the impact of that treatment arm on declared revenue is negative, deductions seem to fall by even more, increasing taxable income. We do not infer much from these estimates since they were not pre-specified and only a single coefficient is significant.

We first assess whether subjects in the treatment group were more likely to file their taxes by the established deadlines - this is an important measure of compliance since it is costly for the tax authority to follow late filers. Filing dates were twice postponed, initially from April 30th to June 30th and subsequently to August 31st, so we consider both initial deadlines. We present graphical evidence of zero treatment effect in [Figure 3](#), where we plot the cumulative share of taxpayers that filed in each week since the intervention, in control and pooled treatment (panel A) or each treatment arm (panel B). The figure indicates that approximately 25% of the experimental sample filed by April 30th, 60% by June 30th and 65% by the final August 31st deadline. There is no differential timing in filing among control and treatment taxpayers, however. In columns (1) and (2) of [Table 5](#) we present regression estimates showing no differential filing probability by each deadline.

We also assessed the impact on the amount of taxes actually paid, since taxpayers might declare tax liabilities but not pay, also generating costs for the TA to recover those due taxes. The intervention might also impact compliance with sales taxes, a higher frequency outcome since it's filed and paid monthly. We thus estimated the impact on total reported sales in the April-August period. Finally, [Brockmeyer et al. \(2019\)](#) also document that their treatment changed compliance with income tax declaration and payment in previous years: treated firms were more likely to file late income tax declarations and pay overdue taxes. We estimated whether the intervention caused a difference in rectifications of previous years' income or sales taxes. These results are presented in columns (3) through (5) in [Table 5](#). Both ITT and LATE estimates suggest no statistically significant impact of the intervention on any of the secondary outcomes examined.

5.3 Heterogeneity analysis

Our ITT estimates document, across a range of outcomes, null effects on average from our intervention. While the average effect of the email across the experimental sample is of much interest, understanding whether some taxpayers are responsive (and which taxpayers) is of crucial importance from a policy standpoint. Under limited capacity to audit and follow-up on low-cost interventions like emails, the tax authority wants to target those units that will adjust their behavior in response to the intervention and avoid losing credibility by targeting non-responsive units.

In our main set of heterogeneity results, we interacted treatment assignment with the four variables used to stratify the experimental sample²⁷: indicator for corporations; indicators for location (Distrito Central, San Pedro Sula and others); indicator of whether third-party information is available for that taxpayer; and the five risk-levels. These

²⁷These specific heterogeneity dimensions were pre-specified before the experiment.

results are presented in [Table 6](#).

We first investigate whether taxpayers for which third-party information is available respond differently from those not informed. [Pomeranz et al. \(2014\)](#) document in Chile that all the response from a similar communication intervention is driven by transactions not subject to paper trail, suggesting third-party information is sufficient to assure compliance. In an intervention similar to ours in Costa Rica, [Brockmeyer et al. \(2019\)](#), on the other hand, document that taxpayers for which third-party information is available respond to emails at least as strongly as those without information. As discussed above, third-party information seems to cover only a fraction of total revenues declared by taxpayers, which suggests there might be scope to improve compliance²⁸. In panel A we estimate that the treatment had a *negative* impact on the probability of filing by taxpayers with no third party information available (-2.4 p.p.) and a zero effect on those with information available. Impacts on reported gross revenue and deductions are estimated to be zero for both groups, while we estimate a positive but only marginally significant increase in reported taxable income for taxpayers with no third-party information. Overall, we interpret these results as evidence of zero effects for taxpayers with available third-party information and mixed evidence for those with no information available.

A second key heterogeneity dimension is whether taxpayers at different risk levels, as defined by the tax authority, respond differently to the intervention. The current practice is to send communications similar to emails in this experiment to those taxpayers considered to be at higher risk, both because that is seen as leading to maximize revenue collection and in order not to "rattle the boat" with taxpayers that are seen as compliant. In this experiment we include subjects in the entire range of positive perceived risk, from very low levels to very high, allowing us to empirically assess whether the intervention's effects are heterogeneous across the distribution of risk.

We present results in panel B. We estimate positive and statistically significant effects across all four outcomes for taxpayers with low risk-level: those in the treated group are 2.3 p.p. more likely to file a declaration and increase reported gross revenue by approximately L100,000. Consistent with previous findings in the literature, however, treated taxpayers also increase their claimed deductions by about L90,000, such that the increase in taxable income is just L10,000. For all taxpayers with higher assessed risk, estimates are not statistically different from zero and always smaller in absolute value. This finding is the opposite of what experts predicted in our survey: they expected larger impacts for high-risk taxpayers.

We interpret these results, with all the caveats related to the external validity of an experiment conducted during the Covid-19 crisis, as evidence that the risk model is not

²⁸While [Brockmeyer et al. \(2019\)](#) include a smaller treatment arm with similar messages to both groups, we cannot disentangle whether differential responses by these two groups are driven by different email content of the communication or by heterogeneity between the groups.

ideally suited for targeting email communications. One possible reason is that the risk model is not fully assessing true risk: it is possible that "low-risk" taxpayers are actually higher risk and that was not entirely captured by the model. Another possibility, in line with the discussion in the model, is that "compliance risk" is just one dimension that matters for targeting. Even if risk is correctly assessed, it is possible that a "soft" email intervention is not ideal for high-risk taxpayers but works for those with small non-compliance risk. We go back to the issue of how ex-ante assessed risk relates to ex-post reaction to the intervention when discussing the use of a causal forest model below.

Finally, we explore heterogeneity in two highly salient dimensions of taxpayers characteristics: corporate form and location. Corporations face a distinct income tax regime (flat rate of 25% on profits instead of a progressive schedule), are larger and often seen as much more sophisticated entities than non-incorporated taxpayers, so a "soft" touch intervention such as communication emails might be less effective to induce compliance. In panel C we estimate that the intervention did not affect declared revenue, deductions or taxable income for either group, but we do find significant impacts on the probability to file taxes. We estimate the intervention increased the probability of filing by 1.3 p.p. for non-incorporated taxpayers, but *decreased* the probability of filing by 3 p.p. for corporations.

Geographical differences in responses to compliance are also of relevance for the tax authorities: San Pedro Sula is the industrial center of Honduras and Distrito Central, where the capital Tegucigalpa is located, is the most populated municipality in the country. In panel D we estimate no differential impact of the treatment across regions: for all primary outcomes we estimate null treatment effects for taxpayers in all three regions.

5.4 Targeting interventions: Causal forest model

While the heterogeneity dimensions analyzed in the previous section were built into our experimental design, the tax authority has access to much richer data that can be explored to answer the question: which taxpayers are more responsive to email interventions? As discussed above, improving targeting for these types of interventions is crucial even if the monetary marginal cost of one extra email is close to zero, since there are reputation costs associated with revealing information to taxpayers and not following up with oversight.

In order to explore a larger set of possible predictors of differential treatment effects and allow for much more flexible heterogeneity²⁹, we use a causal forest algorithm (Athey & Imbens, 2016; Wager & Athey, 2018) to predict conditional average treatment effects

²⁹We use as covariate indicators for whether taxpayer was individual business, salaried worker or self-employed; indicator for corporate form; dummies for geographical region; third-party reported revenue in 2019; third-party reported and self-reported revenue in 2019; taxable income in 2018; declared sales in VAT forms in 2019; continuous measure of assessed risk in 2018; and number of filed VAT declarations in 2018.

(CATE) on subgroups in our experimental sample (De Neve et al., 2021)³⁰. Causal forests predict CATE for each taxpayer, taking the average effects estimated from several *causal trees*: similarly to decision trees, which partition data in order to find best predictions of some observed label, causal trees partition data in order to maximize heterogeneous effects across partitions (leaves). Since treatment effects are not observed but estimated from the data, Athey & Imbens (2016) recommend using "honest trees", which use separate sub samples to perform partition and estimate treatment effects.

In Figure 4 we present a histogram of predicted treatment effect on the filing probability for our entire sample³¹. The distribution of CATE is centered around zero and 90% of predicted effects fall in the range [-3.7,+3.2] percentage points. Overall, the null average treatment effect does not seem to be driven by a large positive effect for some group of taxpayers and negative effects for others, but is more consistent with null effects for the majority of subjects in the experiment.

One specific dimension of interest for targeting these interventions is whether taxpayers deemed to be high-risk are more responsive. In Figure 5, panel A, we document that is not the case. The scatter plot of predicted treatment effect using the causal forest model and the continuous risk measure used by the tax authority shows a slightly negative correlation: predicted treatment effects are positive, albeit small, for taxpayers with very low risk, and become negative for taxpayers with higher assessed risk. We summarize that relationship in panel B, where we present the average predicted treatment effect (weighted by baseline declared revenue) across the five risk-categories created from the continuous risk measure. The average is positive and small for taxpayers in the low and medium-low risk-levels, and negative for taxpayers with higher assessed risk³².

5.5 Costs of the experiment

While the marginal cost of sending emails is zero, any campaign involving thousands of taxpayers requires several activities from dozens of workers in the tax authority³³. These include meetings to define the content of messages, time to prepare databases, train frontline workers who might receive calls or visits from taxpayers, among others. Using back-of-the-envelope calculations of time invested and hourly wage of involved workers,

³⁰For a detailed explanation of the causal forest algorithm and application in economics, see Davis & Heller (2017).

³¹We show in Figure A1 that results are similar if we consider the amount of gross revenue declared as the outcome of interest.

³²Magnitudes are different from our estimated heterogeneous effects obtained in Table 6, where we present results linear regressions that interact the dummy for treatment with baseline assessed risk. In here we estimate a much more flexible model, using causal trees and a richer set of covariates, and then aggregate the predicted effects based on the baseline risk.

³³While the intervention discussed in this paper required further time from researchers outside the tax authority (e.g. developing experts' forecast survey, preparation of the pre-analysis plan and pre-registration), this cost-benefit analysis focus on what an "usual" intervention, developed entirely by the tax authority, would cost.

SAR estimates that the intervention cost approximately L155,000 or USD 6,200. Since there were 31,396 taxpayers receiving emails, the per email costs was less than USD 0.20.

While we estimate overall null effects, we do observe an increase in the probability of filing, in the amount of revenue declared and in the amount of taxable income declared for taxpayers with ex-ante low assessed risk. We can use our back-of-the-envelope calculations for costs of the intervention to estimate a return for that specific group - if we consider that the TA, going forward, decides to target those individuals. The cost of each email is approximately USD 0.20. We estimate that the intervention increased declared taxable income for the low-risk taxpayers by L10,700 or USD 430, so each USD 1 invested in the intervention led to an increase in USD 2,150 in declared taxable income by taxpayers. The actual impact on net tax payment, however, is much smaller: the point estimate for the impact among low-risk taxpayers is only L300 or USD 12, and not statistically different from zero (not reported). Taking that number at face-value, the return of USD 1 would still be a very large USD 60, but we caution about the imprecision of our estimates for those measures.

6 Conclusion and policy implications

We experimentally informed thousands of taxpayers in Honduras that the tax authority had information about their previous transactions in order to assess whether that increased tax compliance. Overall, we found null results. Across several measures of compliance, taxpayers in the treatment groups behaved no differently from those in the control group. We emphasize that the impact of the Covid-19 crisis might have significantly decreased the salience of our treatment and makes the external validity of our findings of null effects harder to assess.

In the key dimension of ex-ante predicted risk, we find that the treatment had a positive effect for taxpayers deemed to be at low-risk of non-compliance. Treated taxpayers in that group were more likely to file taxes and report higher revenue and taxable income. We complemented the heterogeneity analysis using a random forest model that allows for much more complex heterogeneity in treatment effects. Our finding of a negative correlation between predicted treatment effect and ex-ante assessed risk suggests that targeting of policies aimed at increasing compliance should not only consider the level of perceived evasion, but also the likelihood that taxpayers will respond to specific interventions. In contemporary work, [Hoy et al. \(2021\)](#) find that the only taxpayers who respond to a similar intervention in Papua New Guinea are those with low compliance costs; namely, taxpayers who had previously filed taxes and those who filed but claimed to be exempt. This is consistent with our findings and the idea that "soft" interventions such as threatening emails might only change the behavior of "marginal" non-compliers.

When deciding which taxpayers to audit, tax administrations already invest heavily in

identifying those with highest risk of non-compliance. Our findings suggest they should also think carefully about how to target their remaining tools of enforcement. Sending emails or letters to severe non-compliers – individuals or firms with large predicted tax liabilities and who do not file or severely misreport their income – might be innocuous at best or counterproductive at worst, if threats are not followed by enforcement actions. In order to improve targeting, new methods that combine experimental variation, the rich administrative data often available to tax authorities, and machine learning tools are a promising avenue for exploration.

References

- Alatas, V., Banerjee, A., Hanna, R., Olken, B. A., & Tobias, J. (2012). Targeting the Poor: Evidence from a Field Experiment in Indonesia. *American Economic Review*, 102(4), 1206–1240. 5
- Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27), 7353–7360. Publisher: National Academy of Sciences Section: Colloquium Paper. 3, 17, 18
- Bergolo, M., Ceni, R., Cruces, G., Giacobasso, M., & Perez-Truglia, R. (2019). Tax Audits as Scarecrows: Evidence from a Large-Scale Field Experiment. SSRN Scholarly Paper ID 2979096, Social Science Research Network, Rochester, NY. URL <https://papers.ssrn.com/abstract=2979096> 3, 53
- Bertrand, M., Crépon, B., Marguerie, A., & Premand, P. (2017). *Contemporaneous and Post-Program Impacts of a Public Works Program*. Other papers. World Bank. URL <https://elibrary.worldbank.org/doi/abs/10.1596/28460> 5
- Brockmeyer, A., Smith, S., Hernandez, M., & Kettle, S. (2019). Casting a Wider Tax Net: Experimental Evidence from Costa Rica. *American Economic Journal: Economic Policy*, 11(3), 55–87. 4, 7, 14, 15, 16
- Bruhn, M., & McKenzie, D. (2009). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics*, 1(4), 200–232. 10
- Carrillo, P., Pomeranz, D., & Singhal, M. (2017). Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement. *American Economic Journal: Applied Economics*, 9(2), 144–164. 4, 7
- Castro, L., & Scartascini, C. (2015). Tax compliance and enforcement in the pampas evidence from a field experiment. *Journal of Economic Behavior & Organization*, 116(C), 65–82. Publisher: Elsevier. 2, 5
- Cohen, I. (2020). Low-Cost Tax Capacity: A Randomized Evaluation with the Uganda Revenue Authority. *Working Paper*. 4
- Collin, M., Di Maro, V., Evans, D. K., & Manang, F. (2021). Property Tax Compliance in Tanzania. *Working Paper*. 4
- Dal Bó, E., Finan, F., Li, N., & Schechter, L. (2021). Information Technology and Government Decentralization: Experimental Evidence From Paraguay. *Econometrica*, 89, 677–701. 5, 53
- Dalton, A. G., Manning, L. A., Jamison, J. C., Sen, I. K., Karver, J. G., Castaneda Nunez, J. L., Guedes, L. P., & Mujica Estevez, S. B. (2019). Behavioral Insights for Tax Compliance. Tech. Rep. 144398, The World Bank. 7
- Davis, J., & Heller, S. (2017). Using Causal Forests to Predict Treatment Heterogeneity: An Application to Summer Jobs. *American Economic Review*, 107, 546–550. 5, 18

- De Neve, J.-E., Imbert, C., Spinnewijn, J., Tsankova, T., & Luts, M. (2021). How to Improve Tax Compliance? Evidence from Population-Wide Experiments in Belgium. *Journal of Political Economy*, 129(5), 1425–1463. Publisher: The University of Chicago Press. 2, 5, 18
- DellaVigna, S., Otis, N., & Vivalt, E. (2020). Forecasting the Results of Experiments: Piloting an Elicitation Strategy. *AEA Papers and Proceedings*, 110, 75–79. 10
- DellaVigna, S., & Pope, D. (2018). Predicting Experimental Results: Who Knows What? *Journal of Political Economy*, 126(6), 2410–2456. Publisher: The University of Chicago Press. 10
- Gangl, K., Kirchler, E., Lorenz, C., & Torgler, B. (2015). Wealthy Tax Non-Filers in a Developing Country: Taxpayer Knowledge, Perceived Corruption and Service Orientation in Pakistan. SSRN Scholarly Paper ID 2643456, Social Science Research Network, Rochester, NY.
URL <https://papers.ssrn.com/abstract=2643456> 3
- Gordon, R., & Li, W. (2009). Tax structures in developing countries: Many puzzles and a possible explanation. *Journal of Public Economics*, 93(7-8), 855–866. Publisher: Elsevier. 2
- Hallsworth, M. (2014). The use of field experiments to increase tax compliance. *Oxford Review of Economic Policy*, 30(4), 658–679. Publisher: Oxford Academic. 4
- Hoy, C., McKenzie, L., & Sinning, M. G. (2021). Improving Tax Compliance without Increasing Revenue: Evidence from Population-Wide Randomized Controlled Trials in Papua New Guinea. *Policy Research working paper, Washington, D.C. : World Bank Group.*(WPS 9539). 4, 19
- Hussam, R., Rigol, N., & Roth, B. N. (2020). Targeting High Ability Entrepreneurs Using Community Information: Mechanism Design in the Field. 5
- International Monetary Fund (2018). Honduras: Staff Report for the 2018 Article IV Consultation. Tech. rep., International Monetary Fund, Washington, D.C. OCLC: 1048787688. 5, 6
- International Monetary Fund, Organisation for Economic Cooperation and Development, United Nations, & World Bank Group (2016). *Enhancing the Effectiveness of External Support in Building Tax Capacity in Developing Countries*. Other papers. World Bank. 2
- Kettle, S., Hernandez Hernandez, M. A., Ruda, S., & Sanders, M. (2016). Behavioral interventions in tax compliance : evidence from Guatemala. Tech. Rep. WPS7690, The World Bank. 5
- Luttmer, E. F. P., & Singhal, M. (2014). Tax Morale. *Journal of Economic Perspectives*, 28(4), 149–168. 5
- Mascagni, G. (2018). From the Lab to the Field: A Review of Tax Experiments. *Journal of Economic Surveys*, 32(2), 273–301. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/joes.12201>. 2, 4

- Okunogbe, O. (2021). Becoming Legible to the State : The Role of Detection and Enforcement Capacity in Tax Compliance. *World Bank Policy Research working paper, no. WPS 9852*.
URL <https://documents.worldbank.org/en/publication/documents-reports/documentdetail> 4
- Okunogbe, O., & Santoro, F. (2021). The Promise and Limitations of Information Technology for Tax Mobilization. *World Bank Policy Research working paper, no. WPS 9848*.
URL <https://documents.worldbank.org/en/publication/documents-reports/documentdetail> 2
- Pomeranz, D., Marshall, C., & Castellón, P. (2014). Randomized tax enforcement messages : a policy tool for improving audit strategies. *Tax Administration Review*, (36), 1–21. Number: 36 Publisher: Inter-American Center of Tax Administrators (C I A T). 16
- Slemrod, J., Blumenthal, M., & Christian, C. (2001). Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota. *Journal of Public Economics*, 79(3), 455–483.
URL <https://www.sciencedirect.com/science/article/pii/S0047272799001073> 3
- Wager, S., & Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, 113(523), 1228–1242. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/01621459.2017.1319839>. 3, 17

Table 1: Descriptive Statistics - Final Sample

	Mean	SD	p50	N
<i>Panel A: Taxpayers' characteristics</i>				
Corporations	0.33	0.47		31,396
Individual Business	0.41	0.49		31,396
Self-employed service providers	0.06	0.23		31,396
Corporations, IB or self-employed	0.79	0.41		31,396
Distrito Central	0.27	0.45		31,396
San Pedro Sula	0.22	0.41		31,396
<i>Panel B: Past filing behavior</i>				
Reported revenue (Sales) (2019) (L1,000s)	617.79	2,876.15	39.60	31,396
Declared revenue (Sales) (2019) (L1,000s)	1,432.67	9,745.31	190.36	31,396
<i>Income Tax 2018</i>				
Declared income tax in 2018	0.86	0.34	1.00	31,396
Reported revenue (Income) (2018) (L1,000s)	544.36	2,477.44	46.78	31,396
Declared revenue (Income) (2018) (L1,000s)	1,305.15	2,670.35	274.22	31,396
Declared revenue 2018 declaring (L1,000s)	1,509.82	2,817.80	381.22	27,140
Not liable for taxes	0.47	0.50	0.00	27,140
Liable for income taxes	0.50	0.50	1.00	27,140
Liable for asset taxes	0.03	0.16	0.00	27,140
Taxable base 2018 declaring (L1,000s)	115.47	265.81	41.47	27,140
Tax liability 2018 declaring (L1,000s)	15.34	75.74	0.14	27,140
Effective tax rate	0.12	0.11	0.06	14,374
<i>Income Tax 2017</i>				
Declared income tax in 2017	0.81	0.40	1.00	31,396
Reported revenue (Income) (2017) (L1,000s)	483.26	2,538.67	5.40	31,396
Declared revenue (Income) (2017) (L1,000s)	1,233.77	2,499.54	256.28	31,396
Declared revenue 2017 declaring (L1,000s)	1,532.62	2,702.42	421.76	25,274
Taxable base 2017 declaring (L1,000s)	121.47	260.80	62.22	25,274
Tax liability 2017 declaring (L1,000s)	16.57	71.99	0.29	25,274
<i>Panel C: Third-party information</i>				
Third-party information available (2019)	0.71	0.45		31,396
Revenue reported by other taxpayers	0.88	0.33		22,423
Revenue reported by POS operators	0.22	0.42		22,423
Revenue reported by government	0.03	0.17		22,423
Revenue reported by customs	0.01	0.09		22,423
<i>Anomalies</i>				
Declared losses for five years	0.08	0.27		31,396
Atypical financial transactions	0.07	0.26		31,396
Atypical declared revenue	0.49	0.50		31,396
<i>Risk assessment</i>				
Low risk	0.21	0.40		31,396
Medium-low risk	0.28	0.45		31,396
Medium risk	0.23	0.42		31,396
Medium-high risk	0.19	0.39		31,396
High risk	0.10	0.30		31,396

Note: This table presents descriptive statistics for the entire experimental sample. The first panel presents taxpayers characteristics such as type (corporation, individual business or self-employed service providers) and location. The second panel presents descriptive statistics on past tax paying behavior, for both income and sales taxes, in FY2017 and FY2018. Finally, the last panel describes the sources of third-party information and behavioral anomalies used in the tax authority's risk model, and also the distribution of taxpayers across the five broad risk-levels used by the TA.

Table 2: Balance Table - Baseline Characteristics

	Difference in Means (t-test)					
	(1)	(2)	(3)	(4)	(5)	(6)
	Control mean	Control s.d.	Treatment v. Control diff.	Sanctions v. Control diff.	Procedures v. Control diff.	Moral duty v. Control diff.
Individual Business	0.40	(0.49)	-0.01**	-0.02**	-0.01	-0.01
Self-employed service providers	0.06	(0.24)	0.01**	0.01	0.01**	0.01
Corporations, IB or self-employed	0.79	(0.41)	-0.01	-0.01*	0.00	-0.01
Reported revenue (Sales) (2019) (L1,000s)	615.33	(3142.60)	-4.83	32.59	-46.69	-0.22
Declared revenue (Sales) (2019) (L1,000s)	1391.46	(4749.50)	-80.87	-264.16	13.36	6.86
Declared income tax in 2018	0.86	(0.35)	-0.01	-0.01*	-0.01	0.00
Reported revenue (Income) (2018) (L1,000s)	533.62	(1830.09)	-21.08	26.93	-68.67	-21.25
Declared revenue (Income) (2018) (L1,000s)	1303.26	(2672.35)	-3.70	11.18	-9.87	-12.30
Declared revenue 2018 declaring (L1,000s)	1512.46	(2823.38)	5.18	30.78	0.89	-16.18
Not liable for taxes	0.47	(0.50)	0.01	0.01	0.00	0.01
Liable for income taxes	0.50	(0.50)	-0.01	-0.01	-0.00	-0.01
Liable for asset taxes	0.03	(0.16)	0.00	-0.00	0.00	0.00
Taxable base 2018 declaring (L1,000s)	116.40	(270.68)	1.82	0.32	3.56	1.58
Tax liability 2018 declaring (L1,000s)	15.52	(73.46)	0.36	0.14	0.25	0.70
Effective tax rate	0.12	(0.11)	0.00	0.00	0.00	0.00
Declared income tax in 2017	0.80	(0.40)	-0.00	-0.01	-0.00	0.00
Reported revenue (Income) (2017) (L1,000s)	466.91	(1709.31)	-32.10	3.91	-64.74	-35.28
Declared revenue (Income) (2017) (L1,000s)	1230.89	(2513.12)	-5.64	-30.70	9.53	4.07
Declared revenue 2017 declaring (L1,000s)	1531.06	(2719.64)	-3.05	-22.44	14.46	-1.11
Taxable base 2017 declaring (L1,000s)	122.99	(272.04)	2.98	2.32	2.49	4.14
Tax liability 2017 declaring (L1,000s)	17.11	(76.76)	1.06	1.46	1.07	0.65
Revenue reported by other taxpayers	0.88	(0.33)	0.01**	0.01*	0.01*	0.00
Revenue reported by POS operators	0.22	(0.42)	-0.01	-0.00	-0.00	-0.01*
Revenue reported by government	0.03	(0.17)	-0.00	-0.00	-0.00	-0.00
Revenue reported by customs	0.01	(0.09)	0.00	0.00	0.00	0.00
Declared losses for five years	0.08	(0.27)	-0.00	0.00	-0.00	0.00
Atypical financial transactions	0.07	(0.26)	0.00	0.00	-0.00	0.00
Atypical declared revenue	0.48	(0.50)	-0.01**	-0.02**	-0.01	-0.01
Observations	15399	15399	31396	20705	20731	20758

Note: This table compares average characteristics between control and treatment arms. Column (1) presents average taxpayer characteristics and previous filing behavior in FY2018 and FY2017, while Column (2) standard errors. Column (3) presents the difference in averages between the pooled treatment group and control, while columns (4) through (5) presents the differences between taxpayers in control group and each of treatment arms (Sanctions, Procedures and Tax morale arms, respectively) and indicate whether we reject the null of equal averages. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

Table 3: Primary Outcomes - Estimating Program Effects

	(1) Filed declaration	(2) Gross Revenue	(3) Deductions	(4) Taxable Income
<i>ITT estimates</i>				
Treatment	-0.001 [-0.010,0.009]	7.284 [-48.146,62.714]	2.750 [-51.333,56.833]	0.383 [-9.868,10.635]
Observations	31,396	31,396	31,396	31,396
R-Squared	0.186	0.493	0.492	0.052
Control mean	0.65	1,267.88	1,207.31	90.61
<i>LATE estimates</i>				
Opened email	-0.002 [-0.030,0.026]	21.277 [-140.918,183.472]	7.958 [-150.301,166.217]	1.172 [-28.829,31.173]
Observations	31,390	31,390	31,390	31,390
R-Squared	0.186	0.493	0.492	0.052

Note: This table presents the estimated impact of the intervention on primary outcomes. The first panel presents Intention-to-Treat (ITT) estimates using the specification in Equation 1. The second panel presents Local Average Treatment Effect (LATE) estimates using the specification in Equation 2. The dependent variables are an indicator equal to 1 if the taxpayer filed a declaration (Column (1)), the amount of gross revenue declared (Column (2)), the amount of deductions declared (Column (3)) and the amount of taxable income declared (Column (4)). 95% confidence-intervals using robust standard errors are presented in brackets. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

Table 4: Primary Outcomes - Different Treatment Arms

	(1) Filed declaration	(2) Gross Revenue	(3) Deductions	(4) Taxable Income
ITT estimates				
Sanctions treatment	0.004 [-0.009,0.018]	24.804 [-56.611,106.218]	8.857 [-70.397,88.112]	4.078 [-10.504,18.659]
Tax procedures treatment	-0.005 [-0.019,0.008]	23.813 [-54.347,101.974]	27.387 [-50.125,104.899]	-1.905 [-15.422,11.611]
Moral duty treatment	-0.002 [-0.015,0.012]	-26.499 [-96.645,43.646]	-27.805 [-97.125,41.515]	-0.996 [-11.780,9.788]
Observations	31,396	31,396	31,396	31,396
R-Squared	0.186	0.493	0.492	0.052
Control mean	0.65	1,267.88	1,207.31	90.61
$\beta_1 = \beta_2$	0.247	0.984	0.704	0.454
$\beta_1 = \beta_3$	0.475	0.266	0.421	0.454
$\beta_2 = \beta_3$	0.652	0.264	0.221	0.890
LATE estimates				
Opened (Sanctions)	0.013 [-0.026,0.052]	71.777 [-163.608,307.162]	25.612 [-203.561,254.786]	11.803 [-30.362,53.968]
Opened (Procedures)	-0.015 [-0.054,0.023]	67.637 [-154.072,289.347]	77.789 [-142.066,297.643]	-5.413 [-43.774,32.948]
Opened (Tax morale)	-0.005 [-0.046,0.036]	-81.077 [-295.647,133.492]	-85.083 [-297.145,126.979]	-3.046 [-36.022,29.930]
Observations	31,396	31,396	31,396	31,396
R-Squared	0.185	0.493	0.492	0.052

Note: This table presents the estimated impact of each of the treatment arms on primary outcomes. The first panel presents Intention-to-Treat (ITT) estimates using the specification in Equation 4. The second panel presents Local Average Treatment Effect (LATE) estimates. The dependent variables are an indicator equal to 1 if the taxpayer filed a declaration (Column (1)), the amount of gross revenue declared (Column (2)), the amount of deductions declared (Column (3)) and the amount of taxable income declared (Column (4)). 95% confidence-intervals using robust standard errors are presented in brackets. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

Table 5: Secondary Outcomes

	(1)	(2)	(3)	(4)	(5)
	Filed 1st deadline	Filed 2nd deadline	Total taxes paid	Total sales taxes	Revised previous filing
<i>ITT estimates</i>					
Treatment	0.003 [-0.006,0.011]	-0.002 [-0.012,0.008]	-0.408 [-2477.610,2476.794]	-1024.434 [-2841.898,793.031]	0.000 [-0.002,0.002]
Observations	31,396	31,396	31,396	31,396	31,396
R-Squared	0.032	0.101	0.014	0.116	0.027
Control mean	0.18	0.47	4,189.98	13,940.39	0.01
<i>LATE estimates</i>					
Opened email	0.008 [-0.016,0.032]	-0.006 [-0.036,0.025]	-1.526 [-7251.354,7248.301]	-2994.794 [-8313.479,2323.892]	0.001 [-0.005,0.007]
Observations	31,390	31,390	31,390	31,390	31,390
R-Squared	0.033	0.101	0.014	0.115	0.027

Note: This table presents the estimated impact of each of the treatment arms on secondary outcomes. The first panel presents Intention-to-Treat (ITT) estimates using the specification in Equation 1. The second panel presents Local Average Treatment Effect (LATE) estimates using the specification in Equation 2. The dependent variables are an indicator equal to 1 if the taxpayer filed a declaration by the original deadline of April 30th (Column (1)); if they filed by the second deadline of June 30th (Column (2)); the amount of taxes paid (Column (3)), the amount of sales taxes declared between March and August (Column (4)); an indicator equal to 1 if the taxpayer rectified income tax declarations for any year in the period 2014-2018 (Column (5)). 95% confidence-intervals using robust standard errors are presented in brackets. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

Table 6: Heterogeneity Analysis

	(1) Filed declaration	(2) Gross Revenue	(3) Deductions	(4) Taxable Income
Treatment	-0.024** (0.009)	35.906 (41.373)	21.270 (44.146)	11.154* (6.515)
Treatment * Third party info available	0.032*** (0.011)	-40.069 (55.216)	-25.926 (56.281)	-15.078 (9.495)
Third party info available	0.093** (0.041)	500.667 (312.136)	620.852** (298.499)	-43.908* (26.405)
Observations	31,396	31,396	31,396	31,396
R-Squared	0.186	0.493	0.492	0.052
Third-party effect (p-value)	0.150	0.908	0.893	0.567
Treatment	0.023** (0.011)	103.221** (47.219)	92.741* (47.450)	10.740* (5.882)
Treatment * Medium-low risk	-0.026* (0.014)	-112.899* (65.735)	-120.320* (66.185)	-11.226 (9.272)
Treatment * Medium risk	-0.034** (0.015)	-149.615** (71.452)	-126.888* (72.057)	-15.872* (8.407)
Treatment * Medium-high risk	-0.041*** (0.016)	-138.890 (94.713)	-107.568 (89.265)	-19.686 (23.093)
Treatment * High risk	-0.017 (0.020)	-42.755 (135.415)	-73.463 (133.553)	0.771 (9.735)
Observations	31,396	31,396	31,396	31,396
R-Squared	0.186	0.493	0.492	0.052
Medium-low risk effect (p-value)	0.803	0.833	0.550	0.946
Medium risk effect (p-value)	0.307	0.384	0.526	0.394
Medium-high risk effect (p-value)	0.106	0.662	0.843	0.691
High risk effect (p-value)	0.673	0.633	0.877	0.138
Treatment	0.013** (0.006)	10.678 (23.005)	5.220 (22.598)	2.355 (3.129)
Treatment * Corporation	-0.043*** (0.010)	-10.381 (75.998)	-7.557 (74.023)	-6.033 (14.889)
Corporation	0.131*** (0.041)	485.694 (312.441)	611.590** (298.872)	-48.464* (26.933)
Observations	31,396	31,396	31,396	31,396
R-Squared	0.186	0.493	0.492	0.052
Corporation effect (p-value)	0.000	0.997	0.974	0.801
Treatment	0.005 (0.007)	23.814 (34.390)	24.246 (34.683)	0.921 (2.803)
Treatment * Distrito Central	-0.005 (0.012)	-42.821 (69.938)	-44.075 (66.987)	-6.895 (16.621)
Treatment * San Pedro Sula	-0.020 (0.012)	-22.412 (74.773)	-43.710 (74.066)	6.174 (8.735)
Distrito Central	-0.106*** (0.041)	-458.701 (313.115)	-585.382* (299.202)	55.018** (27.654)
San Pedro Sula	0.007 (0.031)	-48.828 (315.594)	-68.546 (307.384)	-10.617 (23.484)
Observations	31,396	31,396	31,396	31,396
R-Squared	0.186	0.493	0.492	0.052
Distrito Central effect (p-value)	0.968	0.757	0.731	0.721
San Pedro Sula effect (p-value)	0.140	0.983	0.765	0.390

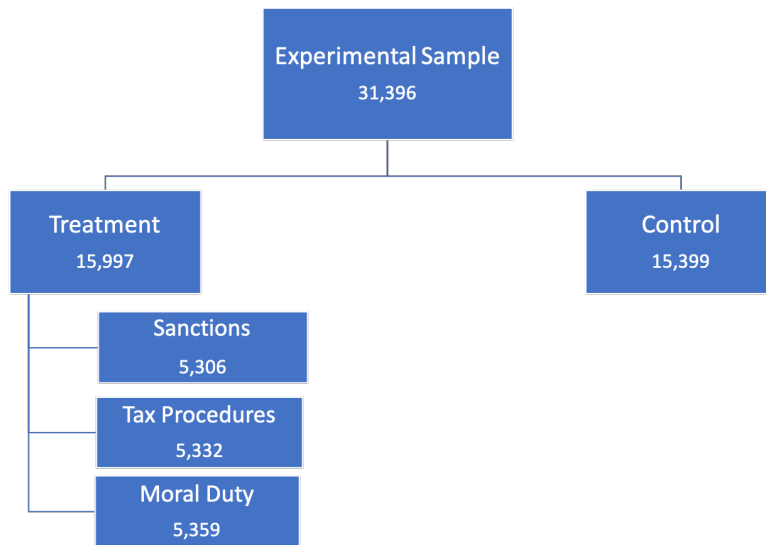
Note: This table presents the estimated impact of the intervention for different sub-populations of the experimental sample. Columns (1) through (4) use each of the primary outcomes as dependent variable and all estimates are ITT, using the same specification in Equation 1 augmented by interacting the treatment dummy with the characteristics of interest. The first panel presents estimates of differential treatment by availability of third party information; the second panel presents estimates by risk-levels; the third panel presents estimates for corporations and non-incorporated entities; and the last panel presents estimates by three different geographical regions. Robust standard errors presented in parentheses. (* p<0.1, ** p<0.05, *** p <0.01)

Table 7: Descriptive Statistics - Forecast survey

	Mean	p50	SD	Min	p25	p75	Max	N
<i>Filing probability (p.p.)</i>								
Pooled	3.9	4.0	3.4	-6.0	2.0	5.0	20.0	64
Sanctions	6.7	5.5	4.6	1.0	4.0	8.0	22.0	62
Procedures	5.3	4.0	5.5	0.0	2.0	7.0	23.0	63
Tax morale	1.6	1.0	3.8	-5.0	0.0	3.0	20.0	61
<i>Gross revenue (%)</i>								
Pooled	6.7	5.0	6.2	-3.0	2.5	10.0	25.0	64
Sanctions	11.5	9.5	10.8	-4.0	5.0	13.0	50.0	64
Procedures	9.3	5.0	12.6	-2.0	1.0	11.0	58.0	63
Tax morale	2.3	0.5	6.1	-18.0	0.0	3.0	36.0	62
Low risk	6.3	1.0	9.4	-5.0	0.0	8.0	38.0	61
Medium-low risk	6.1	4.0	8.0	-7.0	0.0	8.0	28.0	61
Medium risk	7.0	4.0	8.8	-3.0	0.0	11.0	50.0	61
Medium-high	8.8	7.0	10.6	-3.0	1.0	12.0	50.0	61
High risk	9.1	6.0	11.5	-5.0	0.0	15.0	51.0	61
<i>Taxable income (%)</i>								
Pooled	5.7	4.0	5.7	-2.0	2.0	10.0	30.0	64
Sanctions	8.6	6.0	9.2	-4.0	3.0	12.0	50.0	63
Procedures	6.5	4.0	10.0	-2.0	1.0	8.0	57.0	63
Tax morale	1.4	0.0	4.7	-12.0	0.0	2.0	21.0	63

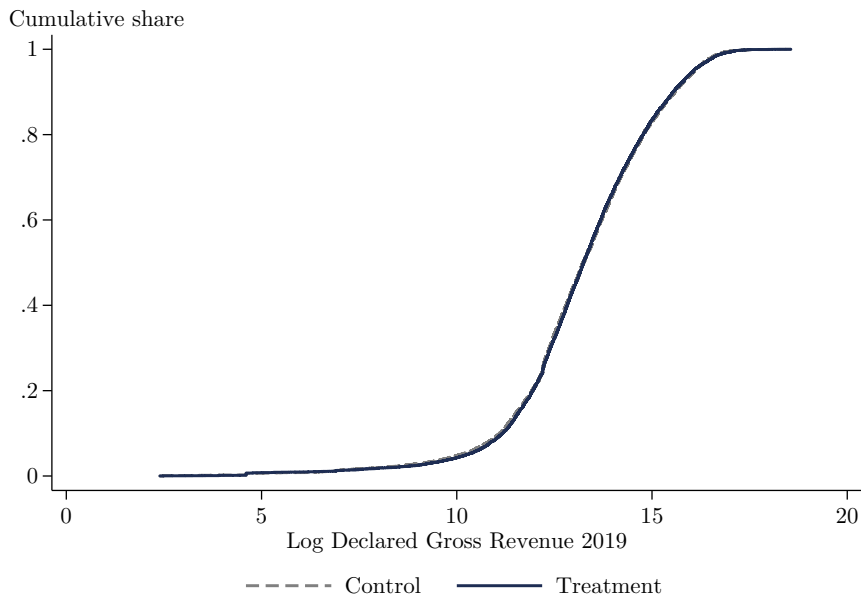
Note: This table presents descriptive statistics for the expert forecast survey. The first panel refers to forecasts of treatment effects on filing probability, the second panel on declared gross revenue and the third panel on declared taxable income. See [Appendix E](#) for details on sample restrictions.

Figure 1: Experimental Design



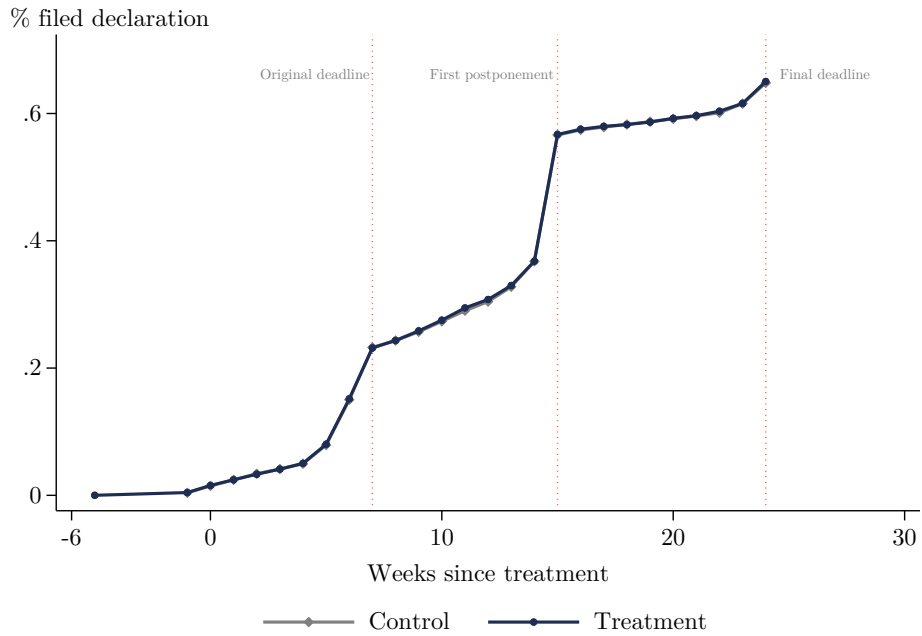
Note: This figure shows the design of the experiment and sample sizes between control and each of the treatment arms. Randomization was stratified in 60 strata defined by availability of third-party information; whether taxpayers is corporation; three geographical regions; and five risk-levels defined by the tax authority ($2 \times 2 \times 3 \times 5 = 60$ strata). Within each treatment arm, messages are personalized to each taxpayer according to the available information on their transactions.

Figure 2: Cumulative distribution function of declared gross revenue - Treatment vs. control groups

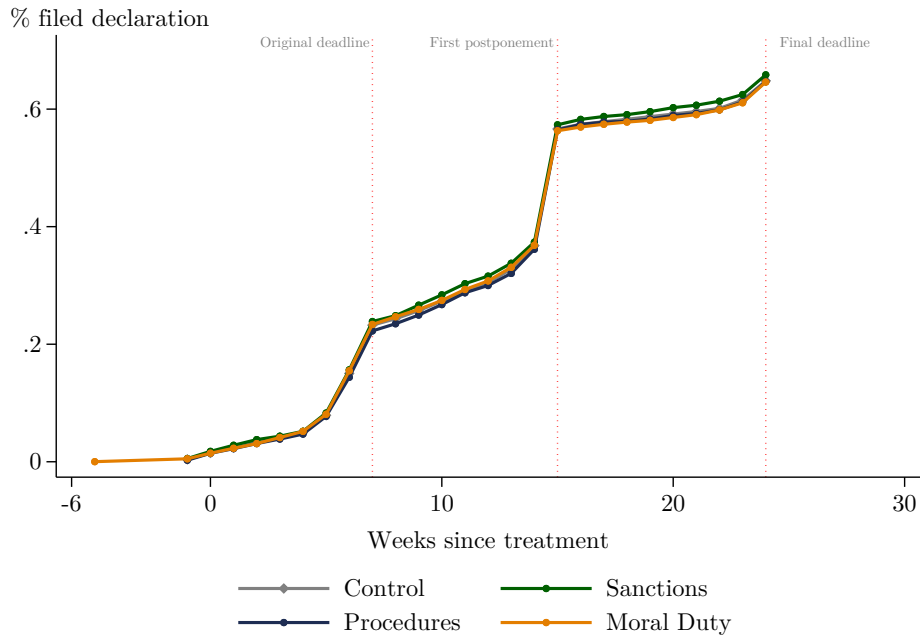


Note: This figure shows the cumulative distribution function of reported gross revenue for FY2019, separately for taxpayers assigned to treatment (blue) and control (grey). We observe no meaningful difference between the two distributions, consistent with the null effect presented in Table 3. The sample is restricted to taxpayers declaring at least L10 in gross revenue for better visualization.

Figure 3: Cumulative share of taxpayers filing per week



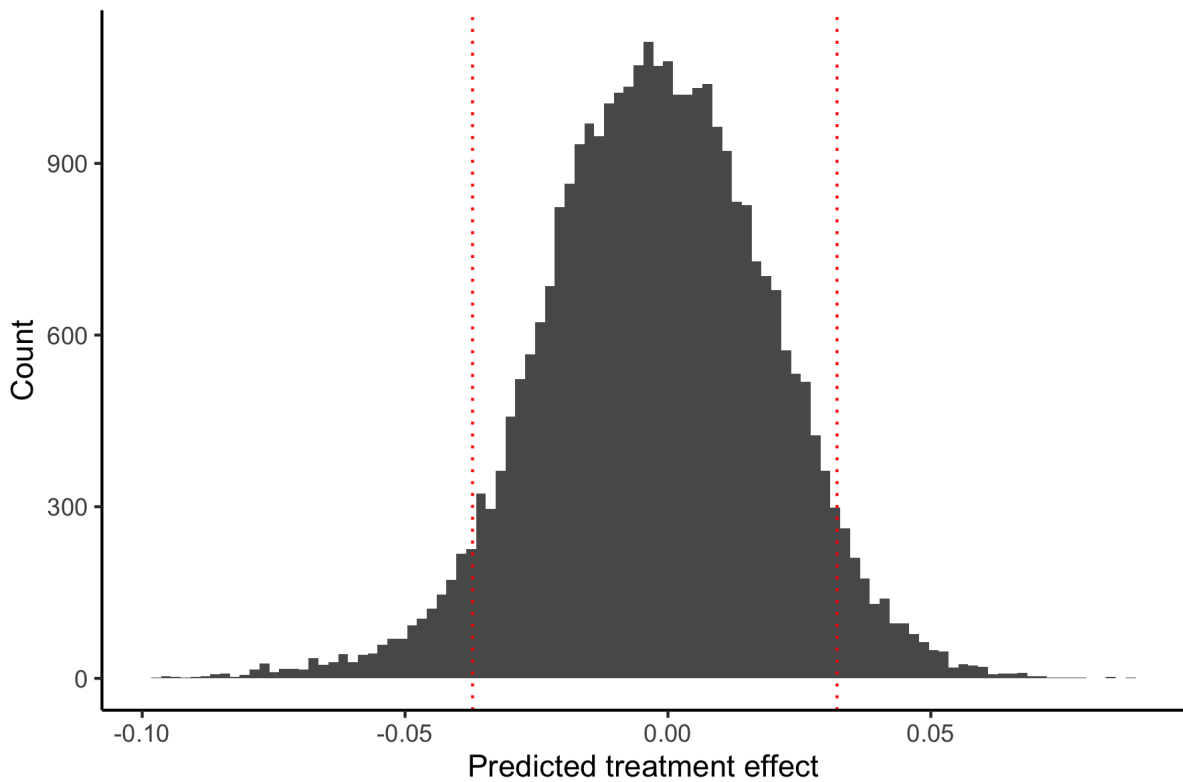
(a) Pooled treatment arms



(b) Separate treatment arms

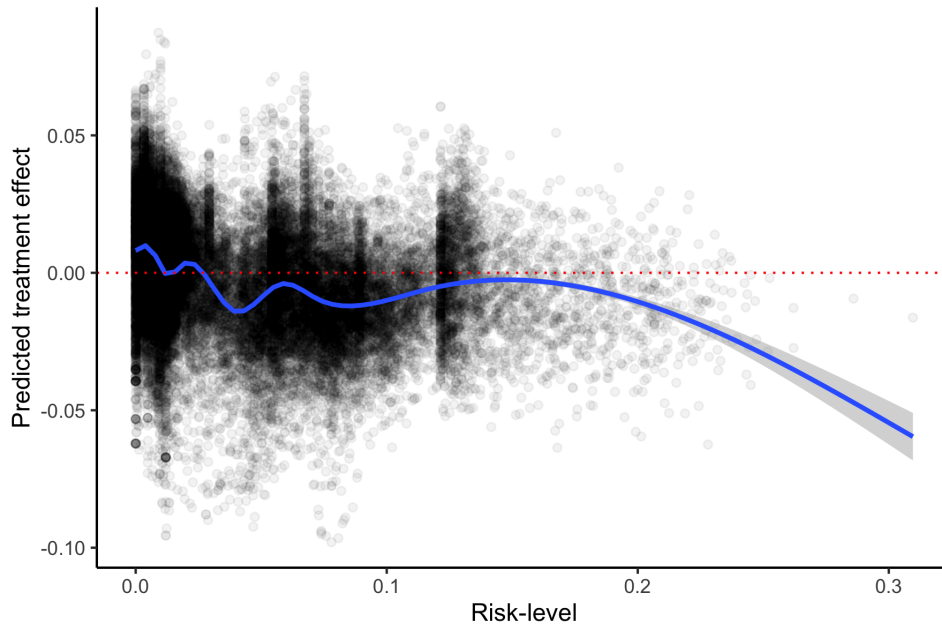
Note: This figure presents the cumulative share of taxpayers filing income tax by each week since the intervention. Panel A presents results for the pooled treatment arms vs. control group whereas Panel B presents results separately for each treatment arm. E-mails were sent on March 11-12th 2020 and the original deadline for tax filing was April 30th. That was initially postponed to June 30th ("first postponement") and subsequently to August 31th ("final deadline").

Figure 4: Estimated treatment effects on filing probability (Random Forest)

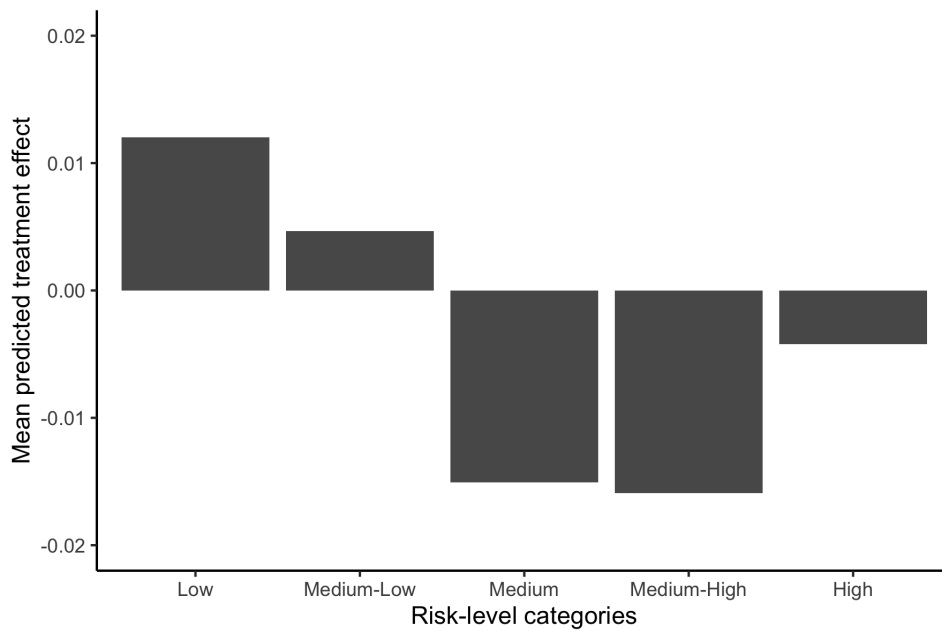


Note: This figure presents the histogram of estimated Conditional Average Treatment Effects (CATE) for the experimental sample, obtained using a random forest model. The dashed lines represent the 5th and 95th percentile of the distribution. CATE are symmetrically distributed around zero and 90% of the effects are estimated to fall in the range $[-3.7, +3.2]$ percentage points.

Figure 5: Estimated treatment effects on filing probability (Random Forest) across risk levels



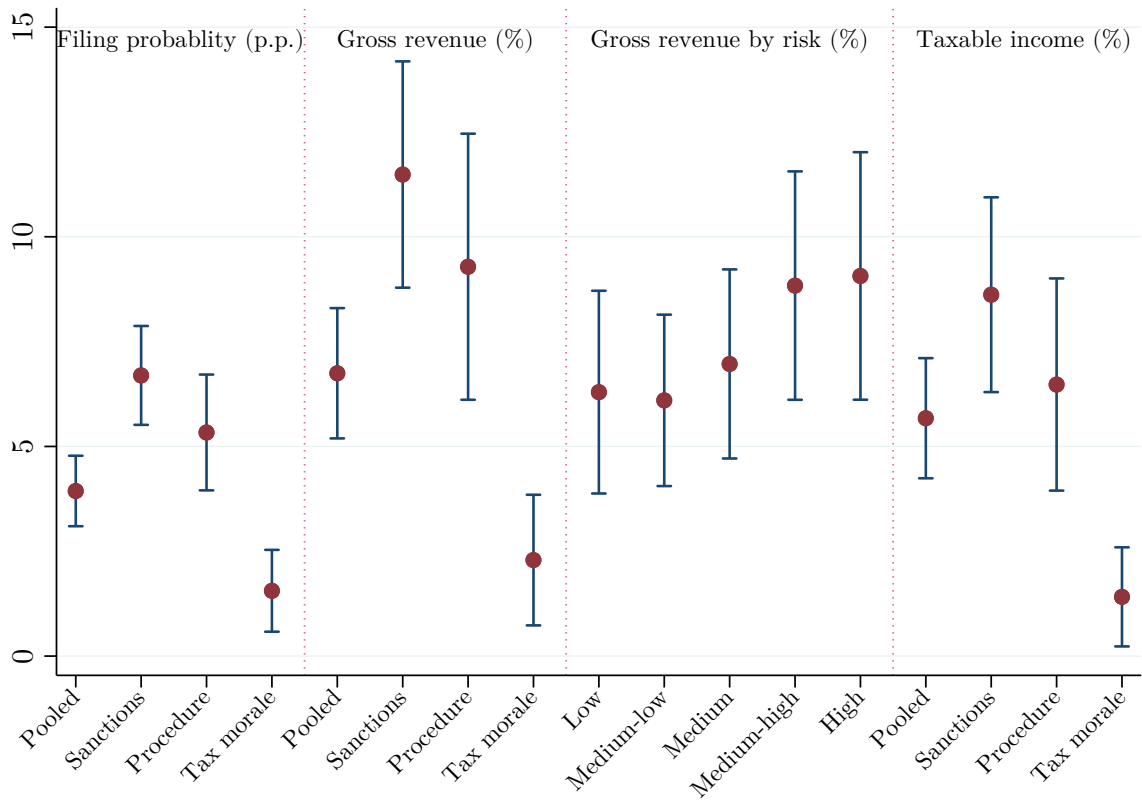
(a) Scatter plot of predicted treatment effects vs. ex-ante risk-level



(b) Mean predicted treatment effect by risk-level categories

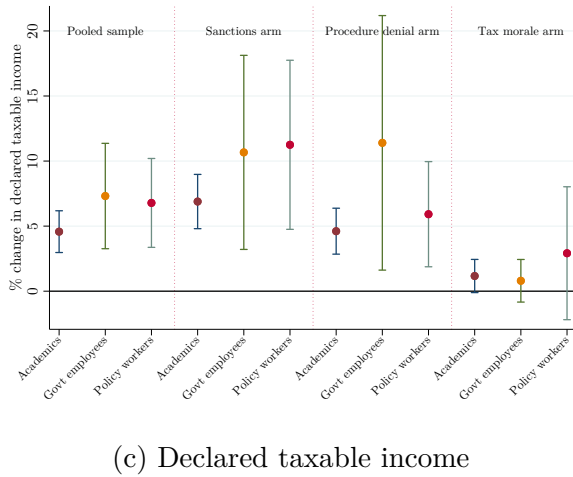
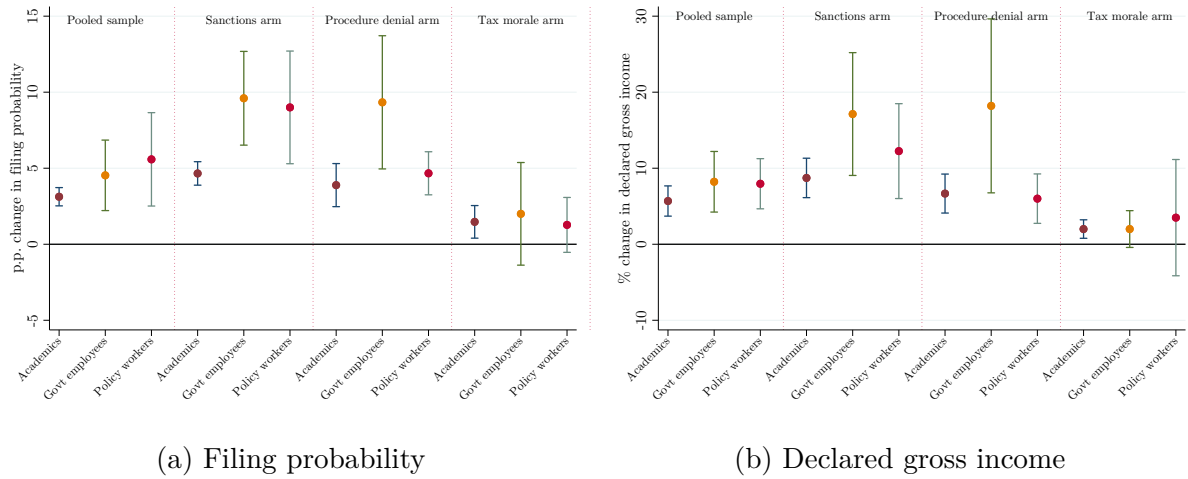
Note: This figure documents how predicted treatment effect varies according to the ex-ante assessed risk level of taxpayers. Panel A presents a scatter plot of predicted treatment effects (vertical axis) and ex-ante assessed risk level (horizontal axis), as well as a local polynomial fit. Panel B presents the mean predicted treatment effect for taxpayers in each of the five risk-level categories determined by the risk model of the tax authority. The mean is weighted by declared revenue in FY2018.

Figure 6: Survey estimates



Note: This figure presents means and 95% confidence-intervals for predictions of treatment effect-size on each of the primary outcomes. See [Appendix E](#) for details on sample restrictions.

Figure 7: Forecasts by groups of respondents



Note: This figure presents means and 95% confidence-intervals for predictions of treatment effect-size on each of the primary outcomes. See [Appendix E](#) for details on sample restrictions.

Figure 8: Letter 1 - Control group



Señor Obligado Tributario

JUAN PEREZ

RTN: XXXXXXXXXXXXXXX

El Servicio de Administración de Rentas (SAR) le recuerda que la obligación tributaria de presentar y pagar la Declaración Jurada del Impuesto Sobre la Renta período 2019, vence el 30 de abril de 2020.

Se le recuerda que la Declaración debe contener información cierta y veraz, reportando la totalidad de los ingresos obtenidos y los gastos tendrán que estar sustentados con documentos fiscales válidos.

[*Haga clic aquí: Cómo presentar su declaración*](#)

“TRIBUTAR ES PROGRESAR”

Para más información: Llamar al 2216-5800 o apersonarse a las ventanillas de Asistencia al Cumplimiento más cercana a su localidad.

www.sar.gob.hn



Figure 9: Letter 2 - Sanctions treatment arm for available third-party information



Señor Obligado Tributario

JUAN PEREZ

RTN: XXXXXXXXXXXXXXXX

En las fuentes de información disponibles en la Administración Tributaria se han identificado sus transacciones comerciales del período 2019, relacionadas a:

Ventas realizadas a otros obligados tributarios

Ventas realizadas por medios de tarjetas de crédito/débito

Ventas y/o servicios realizados al Estado de Honduras

Exportaciones identificadas en aduanas

La obligación tributaria de presentar y pagar la Declaración Jurada del Impuesto Sobre la Renta período 2019, vence el 30 de abril de 2020. Además, se le recuerda que la Declaración debe contener información cierta y veraz, reportando la totalidad de los ingresos obtenidos y los gastos tendrán que estar sustentados con documentos fiscales válidos.

En caso de no cumplir su obligación, será objeto de las sanciones establecidas por el Código Tributario en los Artículos 160 y 163.

[Haga clic aquí: Cómo presentar su declaración](#)

“TRIBUTAR ES PROGRESAR”

Para más información: Llamar al 2216-5800 o apersonarse a las ventanillas de Asistencia al Cumplimiento más cercana a su localidad.

www.sar.gob.hn



Figure 10: Letter 3 - Procedure denial treatment arm for available third-party information



Señor Obligado Tributario

JUAN PEREZ

RTN: XXXXXXXXXXXXXXXX

En las fuentes de información disponibles en la Administración Tributaria se han identificado sus transacciones comerciales del período 2019, relacionadas a:

Ventas realizadas a otros obligados tributarios

Ventas realizadas por medios de tarjetas de crédito/débito

Ventas y/o servicios realizados al Estado de Honduras

Exportaciones identificadas en aduanas

La obligación tributaria de presentar y pagar la Declaración Jurada del Impuesto Sobre la Renta período 2019, vence el 30 de abril de 2020. Además, se le recuerda que la Declaración debe contener información cierta y veraz, reportando la totalidad de los ingresos obtenidos y los gastos tendrán que estar sustentados con documentos fiscales válidos.

En caso de no cumplir su obligación, será afectado en la obtención de constancias de pagos a cuenta, solvencias y documentos fiscales.

[Haga clic aquí: Cómo presentar su declaración](#)

“TRIBUTAR ES PROGRESAR”

Para más información: Llamar al 2216-5800 o apersonarse a las ventanillas de Asistencia al Cumplimiento más cercana a su localidad.

www.sar.gob.hn



Figure 11: Letter 4 - Tax morale treatment arm for available third-party information



Señor Obligado Tributario

JUAN PEREZ

RTN: XXXXXXXXXXXXXXXX

*Por ti, por tus hijos, por Honduras,
¡Paga tus Impuestos!*

En las fuentes de información disponibles en la Administración Tributaria se han identificado sus transacciones comerciales del período 2019, relacionadas a:

Ventas realizadas a otros obligados tributarios

Ventas realizadas por medios de tarjetas de crédito/débito

Ventas y/o servicios realizados al Estado de Honduras

Exportaciones identificadas en aduanas

La obligación tributaria de presentar y pagar la Declaración Jurada del Impuesto Sobre la Renta período 2019, vence el 30 de abril de 2020. Además, se le recuerda que la Declaración debe contener información cierta y veraz, reportando la totalidad de los ingresos obtenidos y los gastos tendrán que estar sustentados con documentos fiscales válidos.

La Honduras que todos queremos para nuestros hijos con educación, salud, infraestructura y seguridad es fruto del esfuerzo de todos sus buenos ciudadanos, gracias a sus impuestos construimos un país mejor.

[Haga clic aquí: Cómo presentar su declaración](#)

“TRIBUTAR ES PROGRESAR”

Para más información: Llamar al 2216-5800 o apersonarse a las ventanillas de Asistencia al Cumplimiento más cercana a su localidad.

www.sar.gob.hn



Figure 12: Letter 5 - Sanctions treatment arm for non-available third-party information



Señor Obligado Tributario

JUAN PEREZ

RTN: XXXXXXXXXXXXXXXX

En las fuentes de información disponibles en la Administración Tributaria, se ha identificado el siguiente comportamiento en sus declaraciones fiscales:

Ha declarado pérdidas fiscales en los últimos cinco periodos de forma consecutiva o alterna

Mantiene movimientos financieros no acorde al nivel de ingresos declarados

Se identifican valores atípicos de sus montos declarados en concepto de Impuesto Sobre la Renta con relación a su industria y nivel de ingresos

La obligación tributaria de presentar y pagar la Declaración Jurada del Impuesto Sobre la Renta período 2019, vence el 30 de abril de 2020. Además, se le recuerda que la Declaración debe contener información cierta y veraz, reportando la totalidad de los ingresos obtenidos y los gastos tendrán que estar sustentados con documentos fiscales válidos.

En caso de no cumplir su obligación, será objeto de las sanciones establecidas por el Código Tributario en los Artículos 160 y 163.

[Haga clic aquí: Cómo presentar su declaración](#)

“TRIBUTAR ES PROGRESAR”

Para más información: Llamar al 2216-5800 o apersonarse a las ventanillas de Asistencia al Cumplimiento más cercana a su localidad.

www.sar.gob.hn



Figure 13: Letter 6 - Procedure denial treatment arm for non-available third-party information



Señor Obligado Tributario

JUAN PEREZ

RTN: XXXXXXXXXXXXXXXX

En las fuentes de información disponibles en la Administración Tributaria, se ha identificado el siguiente comportamiento en sus declaraciones fiscales:

Ha declarado pérdidas fiscales en los últimos cinco periodos de forma consecutiva o alterna

Mantiene movimientos financieros no acorde al nivel de ingresos declarados

Se identifican valores atípicos de sus montos declarados en concepto de Impuesto Sobre la Renta con relación a su industria y nivel de ingresos

La obligación tributaria de presentar y pagar la Declaración Jurada del Impuesto Sobre la Renta período 2019, vence el 30 de abril de 2020. Además, se le recuerda que la Declaración debe contener información cierta y veraz, reportando la totalidad de los ingresos obtenidos y los gastos tendrán que estar sustentados con documentos fiscales válidos.

En caso de no cumplir su obligación, será afectado en la obtención de constancias de pagos a cuenta, solvencias y documentos fiscales.

[Haga clic aquí: Cómo presentar su declaración](#)

“TRIBUTAR ES PROGRESAR”

Para más información: Llamar al 2216-5800 o apersonarse a las ventanillas de Asistencia al Cumplimiento más cercana a su localidad.

www.sar.gob.hn



Figure 14: Letter 7 - Tax morale treatment arm for non-available third-party information



Señor Obligado Tributario

JUAN PEREZ

RTN: XXXXXXXXXXXXXXXX

*Por ti, por tus hijos, por Honduras,
¡Paga tus Impuestos!*

En las fuentes de información disponibles en la Administración Tributaria, se ha identificado el siguiente comportamiento en sus declaraciones fiscales:

Ha declarado pérdidas fiscales en los últimos cinco periodos de forma consecutiva o alterna

Mantiene movimientos financieros no acorde al nivel de ingresos declarados

Se identifican valores atípicos de sus montos declarados en concepto de Impuesto Sobre la Renta con relación a su industria y nivel de ingresos

La obligación tributaria de presentar y pagar la Declaración Jurada del Impuesto Sobre la Renta período 2019, vence el 30 de abril de 2020. Además, se le recuerda que la Declaración debe contener información cierta y veraz, reportando la totalidad de los ingresos obtenidos y los gastos tendrán que estar sustentados con documentos fiscales válidos.

La Honduras que todos queremos para nuestros hijos con educación, salud, infraestructura y seguridad es fruto del esfuerzo de todos sus buenos ciudadanos, gracias a sus impuestos construimos un país mejor.

[Haga clic aquí: Cómo presentar su declaración](#)

“TRIBUTAR ES PROGRESAR”

Para más información: Llamar al 2216-5800 o apersonarse a las ventanillas de Asistencia al Cumplimiento más cercana a su localidad.

www.sar.gob.hn



7 Appendix

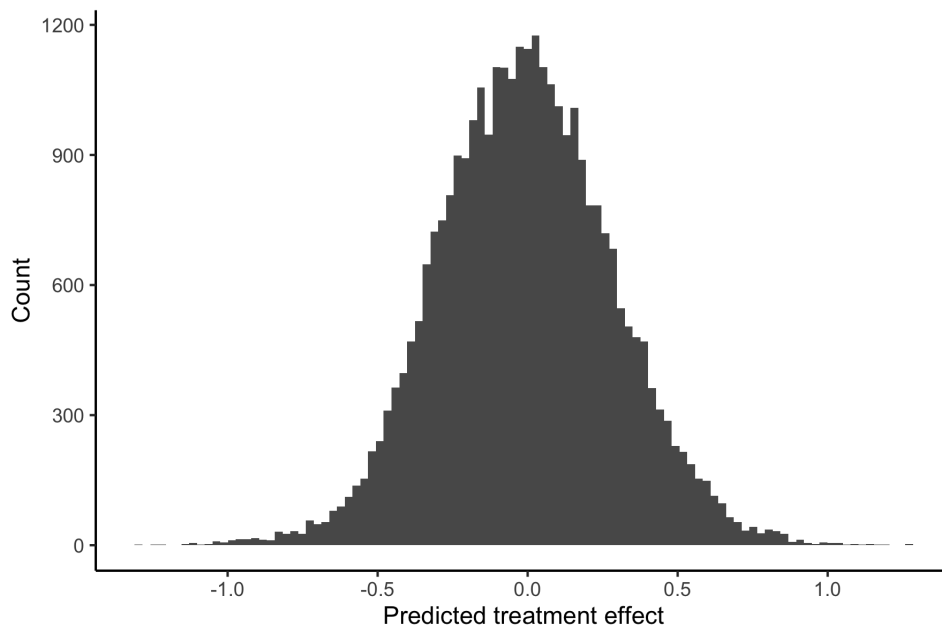
A Appendix tables and figures

Table A1: Primary outcomes trimmed at 99th percentile (non pre-specified)

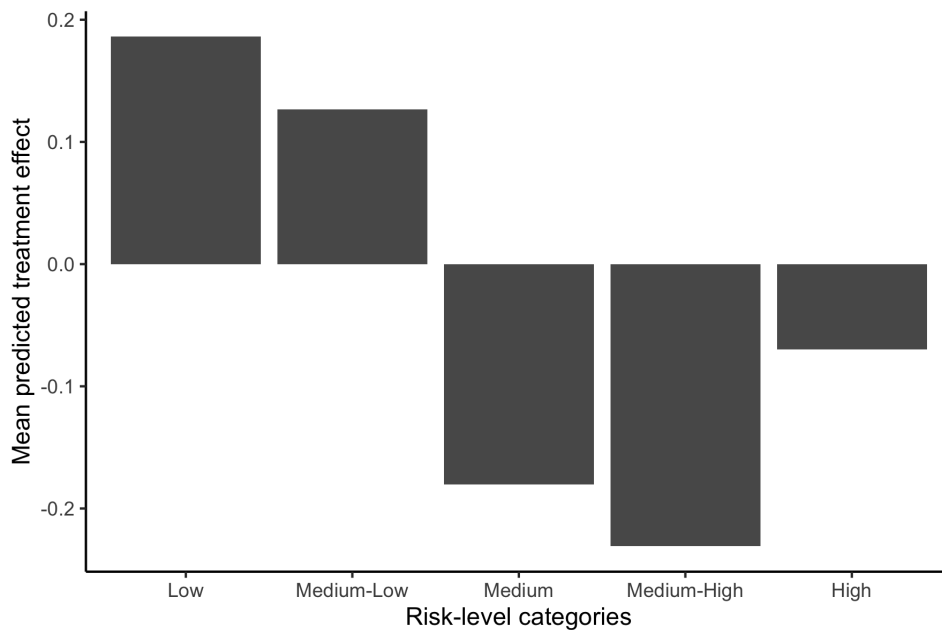
	(1) Gross Revenue	(2) Deductions	(3) Taxable Income
<i>ITT estimates</i>			
Treatment	-5.329 [-38.917,28.260]	-4.561 [-38.000,28.878]	2.024 [-0.418,4.465]
Observations	31,082	31,082	31,082
R-Squared	0.571	0.564	0.208
Control mean	1,046.67	988.63	66.61
<i>ITT estimates</i>			
Sanctions treatment	-3.100 [-48.858,42.658]	-5.132 [-50.835,40.570]	2.079 [-1.358,5.516]
Tax procedures treatment	-4.407 [-51.879,43.064]	1.612 [-45.994,49.218]	0.523 [-2.955,4.002]
Moral duty treatment	-8.447 [-56.672,39.779]	-10.126 [-57.942,37.690]	3.460** [0.041,6.879]
Observations	31,082	31,082	31,082
R-Squared	0.571	0.564	0.208

Note: This table presents the estimated impact of the intervention on primary outcomes using a sample trimmed at the 99th percentile of each outcome distribution. The first panel presents Intention-to-Treat (ITT) estimates using the specification in Equation 1. The second panel presents ITT estimates using the specification in Equation 4. The dependent variables are the amount of gross revenue declared (Column (1)), the amount of deductions declared (Column (2)) and the amount of taxable income declared (Column (3)). 95% confidence-intervals using robust standard errors are presented in brackets. (* p<0.1, ** p<0.05, *** p<0.01)

Figure A1: Estimated treatment effects on declared gross revenue (Random Forest)



(a) Histogram of predicted treatment effects



(b) Mean predicted treatment effect by risk-level categories

Note: This figure presents the distribution of predicted treatment effect on gross revenue (Panel A) and the mean predicted treatment effect for taxpayers in each of the five risk-level categories determined by the risk model of the tax authority (Panel B). The mean is weighted by declared revenue in FY2018.

B Pilot study with non-filers

In preparation for our main experiment, we implemented a pilot intervention jointly with our partners in the Honduran tax authority. Since our experiment aims at measuring the impact of email messages on compliance in FY2019 income tax filing, we could not "pilot" the same intervention before the filing deadline. Our approach was to focus on taxpayers who were believed to be at risk for non-filing in FY2018, i.e., the tax authority believes they should have filed an income tax declaration in April 2019 but they had not.

Our pilot experimental sample was comprised of 2,599 taxpayers, 1,000 of which were assigned to the treatment group and 1,599 to the control³⁴. Taxpayers in the treatment group received an email similar to the main experiment, describing which third-party information was available on their transactions, stating the tax authority believed they should have filed income taxes and giving them 10 days to do so. There were no different treatment arms, and the message included a threat of audit in case they did not file. Unlike our main experiment, the control group did not receive any notification, so we see the (differential) treatment in the pilot as much stronger.

The main impact of the experimental intervention on filing probability is illustrated in Figure A2, which presents the cumulative share of taxpayers in control and treatment status that filed a (late) declaration for FY2018 by each date. The gap between treatment and control is zero by the date of the intervention but steadily increases in the weeks after, so that six weeks post-intervention slightly more than 10% of taxpayers in the treatment group have filed a declaration vs. less than 4% among control. The impact of the intervention can also be observed in the amount of tax liability declared, presented in Figure A3: taxpayers in the treatment group declared, in aggregate, L250,000 more in tax liabilities (approximately 30% increase).

We present these results in regression form in Panel A Table A2. Odd columns present simple differences between treatment and control, while even ones include controls. For filing probability, the Intention-to-Treat (ITT) estimation is a 6 p.p. increase in the

³⁴We decided to assign most taxpayers to the control group so that they could still be part of the main experiment, while those in the treatment group were excluded.

probability of filing, from a baseline probability of only 4% among control units. For the amount of tax liability declared, the point estimate indicates an increase of approximately L500, or 100% from the control mean, but standard errors are very large and we cannot reject a null effect.

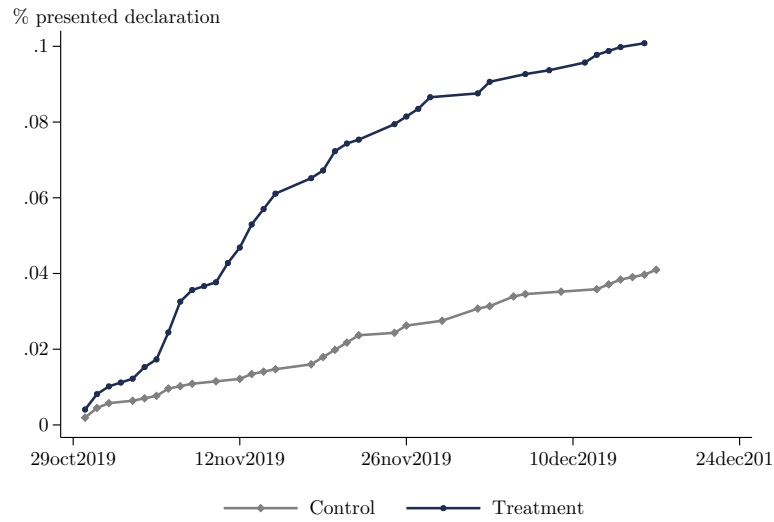
The share of taxpayers who actually clicked on the email sent to the treatment group, nonetheless, was only 33%, implying that the effect of compliers must have been even larger. We present LATE results in Panel B of Table A2, where we instrument opening the email with treatment assignment. The results suggest that clicking on the email increase filing probability by 19 p.p. The same result can be seen in non-parametric form in Figure A4, where we see that the entire increase in filing among units assigned to treatment come from those that clicked the email.

Table A2: Pilot Results

	Presented declaration		Tax liability (L)	
	(1)	(2)	(3)	(4)
<i>ITT estimates</i>				
Treatment	0.0598*** (0.01)	0.0596*** (0.01)	484.5 (413.77)	500.0 (405.89)
Constant	0.0410*** (0.01)	0.0241*** (0.01)	576.7** (280.02)	190.4 (159.09)
Observations	2544	2544	2544	2544
R-Squared	0.0142	0.0212	0.000504	0.00433
Controls?	No	Yes	No	Yes
Control average	0.041	0.041	576.69	576.69
<i>LATE estimates</i>				
Clicked on email	0.183*** (0.03)	0.181*** (0.03)	1482.3 (1260.95)	1523.0 (1231.04)
Constant	0.0410*** (0.01)	0.0250*** (0.01)	576.7** (279.91)	197.9 (159.71)
Observations	2544	2544	2544	2544
R-Squared	0.0618	0.0680	0.00453	0.00797
Controls?	No	Yes	No	Yes
Control average	0.041	0.041	576.69	576.69

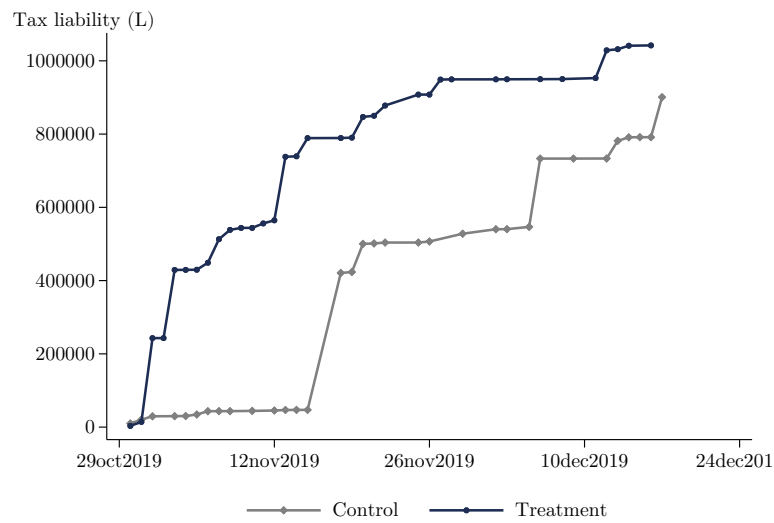
Note: This table reports ITT (first panel) and LATE (second panel) results of our pilot experiment. Controls include a dummy for corporations, whether the taxpayer presented income tax declaration for FY2017 and the amount of gross revenue declared in 2017. Robust standard errors in parentheses (* p<0.1, ** p<0.05, *** p <0.01)

Figure A2: Probability of filing income taxes by treatment status



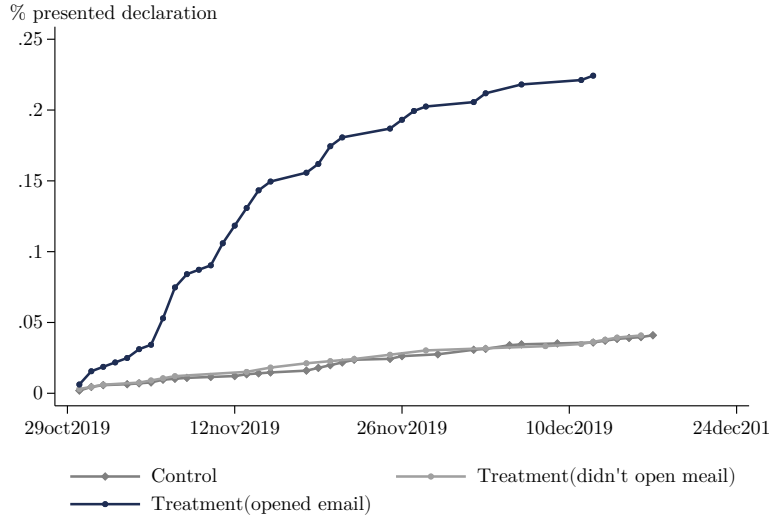
Note: This figure presents the cumulative share of taxpayers in treatment and control groups that have filed a late income tax declaration by each date.

Figure A3: Declared tax liability by treatment status



Note: This figure presents the cumulative amount of tax liability declared by taxpayers in treatment and control groups by each date.

Figure A4: Probability of filing income taxes by status of email opening



Note: This figure presents the cumulative share of taxpayers in treatment and control groups that have filed a late income tax declaration by each date, distinguishing between those in treatment that clicked on the email sent and those who did not.

C Minimum detectable effects

We calculate Minimum Detectable Effects (MDE) of an experiment with 80% power and 5% significance level for our four primary outcomes: probability of filing income taxes; amount of declared gross revenues; amount of declared deductions; and amount of declared taxable income. Using data from FY2018 and FY2017 filings, we first calculate the residual variance of outcomes of interest after estimating regressions of the form:

$$y_{i,2018} = \alpha + \beta' X_i + \epsilon_i \quad (5)$$

where $y_{i,2018}$ are one of the three continuous primary outcomes described above and X_i are the same covariates used as controls when estimating treatment effects³⁵.

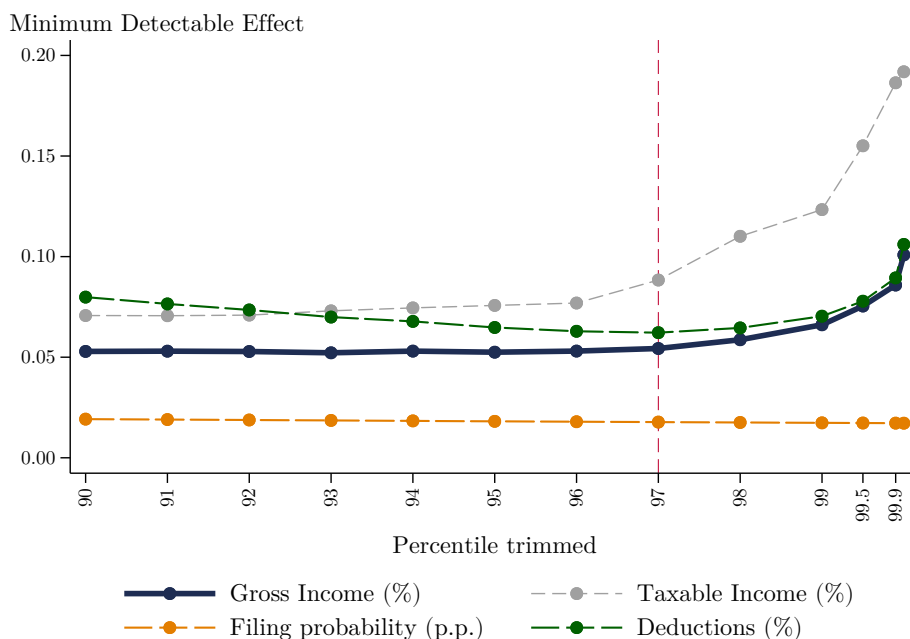
Table A3 presents the results of these estimates for each of the three continuous primary outcomes in 2018. We are able to explain between 55% (for taxable income)

³⁵These are strata dummies; dummy for presenting tax declaration in the previous year; amount of declared gross revenue, third-party informed gross revenue and declared taxable income in the previous year; and amount of declared sales tax revenues in the same year.

and 75% (for gross revenues) of total variance by using these controls, highlighting the importance of using past filing information to increase the precision of our estimates. In our power calculations, we also assume that compliance with treatment is 60%, under the assumption that 60% of taxpayers assigned to the treatment group will open the email sent³⁶.

Under these hypotheses and using our sample of 31,396 taxpayers, our MDE for the pooled treatment vs. control comparison, presented in the first panel of Table A4, is 2 percentage points for the probability of filing declaration; L71,000 of gross revenues (5.5% of the baseline mean); L78,000 of deductions (6.2% of the mean); and L8,800 of taxable income (8.8% of the mean). For each of the individual treatment arms (second panel) our MDEs are 2.5 p.p. for filing probability; L100,000 of gross revenues; L110,000 for deductions and L12,400 for taxable income.

Figure A5: Minimum Detectable Effect - Final Sample



Note: This figure presents Minimum Detectable Effects (MDE) of experiments with 80% power and 5% significance level for each of our primary outcomes. Each point is the MDE if we trim the experimental sample at the Xth percentile of the FY2018 declared gross revenue distribution. MDEs are calculated considering the residual variance of primary outcomes obtained from the estimation of Equation 5.

³⁶This figure was informed by a pilot discussed below.

Table A3: Power Calculations - Residual Variance

	2018 primary outcomes		
	(1)	(2)	(3)
	Revenue	Deductions	Taxable Income
Taxable income 2017	-0.005 (0.069)	-0.868*** (0.065)	0.742*** (0.036)
Reported revenue (Income) (2017) (L1,000s)	-0.306*** (0.048)	-0.284*** (0.048)	-0.014*** (0.003)
Declared revenue (Income) (2017) (L1,000s)	0.857*** (0.015)	0.870*** (0.015)	0.005*** (0.002)
Declared income tax in 2017	13.722 (15.653)	26.889 (16.440)	-7.314** (3.264)
Reported revenue (Sales) (2018) (L1,000s)	0.395*** (0.047)	0.371*** (0.047)	0.017*** (0.003)
Constant	-44.950 (41.024)	-54.516 (41.848)	16.175** (7.056)
Observations	31,396	31,396	31,396
R-Squared	0.742	0.701	0.554
Strata FE	Yes	Yes	Yes

Note: In this table we present the result of regressions used to obtain residual variance in the power calculations. The dependent variables are Gross Revenue, Deductions and Taxable Income in FY2018 in Columns (1), (2) and (3), respectively. All regressions include strata fixed-effects. Robust standard errors in parentheses. (* p<0.1, ** p<0.05, *** p <0.01)

Table A4: Power Calculations - Final Sample

Outcome	Levels	Percent
<i>MDE of pooled treatment</i>		
Gross Income	71.47	0.055
Deductions	77.94	0.062
Taxable Income	8.81	0.088
Filing probability	0.02	0.021
<i>MDE of treatment arms</i>		
Gross Income	100.79	0.077
Deductions	109.92	0.088
Taxable Income	12.42	0.010
Filing probability	0.02	0.029

Note: In this table we present Minimum Detectable Effects (MDE) for our primary outcomes. The first panel presents MDEs for the experiment comparing pooled treatment arms with control, while the second panel presents MDE for the experiment comparing a single treatment arm with control. The first column presents the MDE in levels, the units are thousand Lempiras for Gross Income, Deductions and Taxable Income; and percentage points for filing probability. The second column presents the MDE in levels as percentage of the baseline mean.

D A model on the value of targeting

In this section we present a highly simplified model that we believe illustrates the value of experimentally acquiring information on taxpayers when targeting interventions. Consider a model in which agents must decide whether to pay their taxes (Alingham and Sandmo, 1978). In our setting agents only choose the extensive margin, whether paying or not, and not the intensive margin - conditional on paying they truthfully report their tax liability.

Agents are characterized by a vector (θ_i, γ_i) , with some joint distribution $F(\theta, \gamma)$. θ_i measures how much disutility an agent gets from not reporting their taxes. It can be interpreted as a psychological cost of non-compliance, or as how costly it is for some agents to go through the hurdles of misreporting instead of simply reporting their book numbers. γ_i captures how intensely agents update their beliefs about being caught cheating upon receiving a letter/email from the tax authority (noted by the treatment indicator variable T_i). It might capture other traits such as fear (Bergolo et al., 2019) or just knowledge about how credible the threat it. Regardless of interpretation, it reflects the fact that the treatment will have heterogeneous effects depending on agents' type (Dal Bó et al., 2021).

If agents file their taxes, they do so truthfully and pay a tax rate τ on their profits π , receiving payoff $(1 - \tau)\pi$. If they decide not to file, they pay the cost θ_i regardless of being caught or not, and with an exogenous probability p they are caught and must pay their taxes plus a penalty f , and with probability $(1-p)$ they simply do not pay any taxes.

Given this setting, agents will pay their taxes whenever

$$\begin{aligned} \text{Payoff}_{\text{pay}} &\geq \text{Payoff}_{\text{notpay}} \\ (1 - \tau)\pi &\geq (1 - p - T_i\gamma_i)\pi + (p + T_i\gamma_i)((1 - \tau)\pi - f) - \theta_i \\ \theta_i &\geq \tau\pi - (p + T_i\gamma_i)C \end{aligned}$$

where $C = \tau\pi + f$ is the cost incurred by the taxpayer if caught in non-compliance. Consider first what happens in the absence of treatment, or with the control group.

Since $T_i = 0$, the expression above simplifies to

$$\theta_i \geq \tau\pi - pC$$

There is a minimum value for θ_i such that all individuals with values above that threshold will comply, since their non-compliance cost $pC + \theta_i$ is larger than their benefit of non-filing $\tau\pi$; and all with θ_i below that threshold will not file taxes.

The parameter θ_i , in this model, fully characterizes the compliance behavior in the absence of treatment. Suppose the tax authority needs to deploy audits, a costly investment that once deployed fully reveals whether the taxpayer cheated or not (some taxpayers do not need to file, so their not filing is not cheating). The only thing the TA needs to know is θ_i , and it can target those taxpayers who should have filed, but did not.

Now consider that audits are not available to the tax authorities, which needs to rely on sending letters that hopefully will encourage some taxpayers to comply. Taxpayers in the treatment group will comply if

$$\theta_i \geq \tau\pi - (p + T_i\gamma_i)C$$

For taxpayers with $\gamma_i = 0$, the condition remains the same as in the treatment group: they will not update their beliefs upon receiving the letter, and the threshold on θ_i is the same as in the treatment group. For any positive value of γ_i , on the other hand, the threshold is now lower: now agents with a somewhat lower cost θ_i will start to comply since they increased their belief on the probability of punishment. For a large enough γ_i (and fine), potentially all agents could comply regardless of their fixed cost, since they see punishment as certain.

In Figure A6 below we present a summary of taxpayer behavior arising from the model. In the absence of treatment, the only thing that matters is taxpayer's cost θ_i : if it is above the threshold, they will file, otherwise not. When the treatment is introduced, however, taxpayers' heterogeneity in how they react to the letter start to play a role: even for those with low cost θ_i , if the update in beliefs is large enough the taxpayer will

file their taxes.

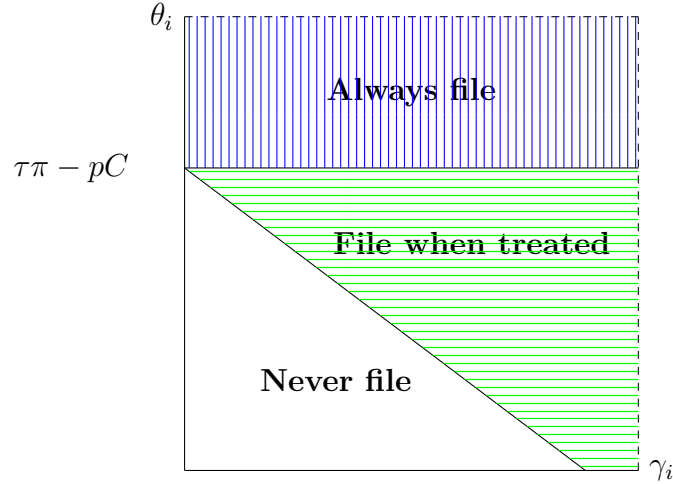


Figure A6: Taxpayer behavior according to types

Now suppose that the TA must choose which taxpayers to target with emails/letters. Since this intervention is virtually costless, it could be argued that the TA does not face a targeting problem: they should send emails to all taxpayers and, under the assumptions of this model, are guaranteed to have a positive return, the magnitude of which depends on the distribution of (θ_i, γ_i) in the population, as well as on other parameters.

But consider the case in which contacting taxpayers which will not file is costly for the TA for credibility reasons: if taxpayers are informed of non-compliance and still are not punished they might (correctly) update their beliefs about the capacity of the authority, and be even less compliant in the future. Whatever the underlying reason, the TA now faces a targeting problem: it would like to contact only those taxpayers in the green region of the graph, who will change their behavior in response to the letter, and not contact those in the white region³⁷. Under incomplete information, however, the TA does not observe γ_i so cannot perfectly target.

Let $Z = \theta_i + \gamma_i C$ and consider the distribution $G(Z)$ with support $[Z_l, Z_H]$. If the TA sends emails to all individuals, the total effect of this intervention will be

$$\int_{\tau\pi - pC}^{Z_H} BG(Z)dz - \int_{Z_l}^{\tau\pi - pC} \delta G(Z)dz$$

³⁷That does not mean the TA should not take some action regarding those taxpayers in the white area above, but simply that emails/letters are not the right tool for those taxpayers.

where δ is the per taxpayer reputation cost when they are treated and do not file, and B is the benefit received when the taxpayer files.

Under perfect information the TA would simply target those with $Z \in [\tau\pi - pC, Z_h]$. Absent that, they can do better than universal intervention by acquiring some information that is predictive about taxpayers Z . Our goal of using the experimental variation in letters with the causal forest algorithm is precisely to approximate this information using conditional average treatment effects (CATE), or treatment effects for groups of taxpayers that share some (observable) characteristics.

E Forecast survey

Our survey was designed using Qualtrics and invitations were sent to approximately 300 individuals, as well as the Development listserv of PhD students at UC Berkeley. We received 111 partial or complete survey answers.

We initially dropped 35 surveys mostly empty. We then performed a (mild) consistency check: since respondents were asked to estimate the overall treatment effect for our three primary outcomes using the pooled sample and individual treatment arms, we checked whether the overall effect was larger than the minimum and smaller than the maximum effect of the individual arms. We then dropped 11 surveys for which the responses were inconsistent *in all tests*. That left us with our main final sample of 65 responses. In order to avoid outliers that severely distort means, we also trimmed each of the questions where the response was larger than the mean plus three standard deviations. That explains the small differences in sample for different questions.

Targeting in Tax Compliance Interventions: Experimental Evidence from Honduras

Start of Block: Consent page

Targeting in Tax Compliance Interventions: Experimental Evidence from Honduras (Del Carmen, Espinal Hernandez and Scot; 2020)

Welcome!

This survey aims to collect predictions on the results of a randomized control trial (RCT) studying how taxpayers change their compliance behavior when notified about the information available to the tax authority regarding their transactions. We will then compare these forecasts to actual experimental results. You can learn more about our experiment [here](#) [this link will open in a new tab].

The survey should take around 10 minutes to complete. We appreciate your time and help, and thank you in advance for participating!

You might find more information on consent in participating in this survey on the [consent form](#). Please click the consent button below to proceed to the survey.

End of Block: Consent page

Start of Block: Survey introduction

About you

Before you start we'd like to know what's your profession. Choose the option below that best fits your career:

- Academic economist (Faculty, PhD student, Researcher in academic institution) (1)
 - Public sector employee in Honduras (2)
 - Researcher or officer in policy-oriented organization (IDB, World Bank, IMF, OECD, other) (3)
 - Other (4)
-

Based on your knowledge about other studies with taxpayers and/or your knowledge about tax administration, how confident are you about your ability to predict the results of this experiment?

- Not confident at all (1)
- Not very confident (2)
- Somewhat confident (3)
- Confident (4)

Page Break

Study overview

We are implementing a randomized control trial to study the impact of providing personalized messages to taxpayers about the information available to the Tax Authority (TA) regarding their transactions. Our experimental sample consists of **31,396 taxpayers** in Honduras considered to be at-risk of non-compliance according to the TA's risk model. The distinguishing feature of RCTs is the **random assignment** of subjects to different groups, which allow us to compare behavior between groups and determine the **causal effect of policy interventions**.

Approximately 7 weeks before the deadline for taxpayers to file their FY2019 income tax declaration, we send emails to all taxpayers in our sample. Subjects assigned to the **control group** receive a message with a reminder about the filing deadline and the importance of truthfully reporting their tax liabilities.

While this is all information provided to subjects randomly assigned to the control group, emails to those in the **treatment group** additionally include information available to the Tax Authority (TA) regarding their transactions. The emails are personalized for each taxpayer and include either the sources of third-party information the authority possess on their revenue (sales to other taxpayers, debit/credit card sales, exports or sales to the government) or indicators about their operations flagged by the authority (repeated reported losses, financial transactions inconsistent with reported revenue or low declared revenues compared to peers) for taxpayers with no third-party information available. For legal reasons, no specific amounts or partners are mentioned - only the knowledge of specific categories of transactions.

We will measure the intervention's impact using administrative data on taxpayers' filing, including whether they filed their income tax declaration or not, and amount of revenue and taxable income declared.

We are aware that the results of the experiment may be affected by the impact of COVID-19 in the country, particularly if the period for filing taxes is extended. In that case we plan to re-send the emails to make the intervention more salient. We are monitoring the situation to adjust the experiment going forward.

Page Break

Baseline measures

We have information on baseline outcomes of interest using taxpayers' **FY2018 filings** which might be useful to illustrate magnitudes. Among the 31,396 taxpayers in our experimental sample:

- 86% filed income taxes in 2018.
- The average declared gross revenue was 1.3 million Lempiras (s.e. L 15,000), or approximately USD 52,000 using 2018 average exchange rate (1 USD = L 25).
- The average taxable income (gross revenue net of deductions) was L 100,000 (s.e. L 1,400), or approximately USD 4,000.

End of Block: Survey introduction

Start of Block: Survey questions

In this section we ask you to predict **how** opening an email from the Tax Authority mentioning specific information about your past transactions affects **taxpayers'** compliance with tax obligations.

Since not all taxpayers assigned to treatment actually open the emails sent (due to possibly incorrect email addresses, full mailboxes or simply not clicking on it), we ask you to predict what's the **causal effect of opening the email** (i.e. we estimate an instrumental variable (IV) regression where clicking on the email is instrumented by the random treatment assignment).

Question 1

What do you predict will be the effect of opening the email on the probability of filing income taxes, in **percentage points**?

As a reminder, 86% of our sample filed income taxes in 2018.

Page Break

You answered $\${Q8/ChoiceTextEntryValue}$ percentage points in the last question. As an illustration, if 86% of taxpayers in the control group file their taxes (as in FY2018), the intervention would change the filing rate to $\$e\{86+\${Q8/ChoiceTextEntryValue}\}\%$.

If this is correct, please proceed to the next question. Otherwise feel free to go back and adjust your answer.

Page Break

You answered $\{Q8/ChoiceTextEntryValue\}$ percentage points in the last question. As an illustration, if 86% of taxpayers in the control group file their taxes (as in FY2018), the intervention would be (more than) enough to induce 100% of taxpayers opening the email to file.

If this is correct, please proceed to the next question. Otherwise feel free to go back and adjust your answer.

Page Break

Question 2

What do you predict will be the **percentage change** effect of opening the email on (unconditional) declared gross revenue? [If you believe it will increase revenue by X%, please type X].

As a reminder, in 2018, the average declared gross revenue was 1.3 million Lempiras (s.e. L 15,000), or approximately USD 52,000.

Page Break

You answered $\{Q11/ChoiceTextEntryValue\}$ % in the last question. As an illustration, if the average declared gross revenue among the control group is L 1,300,000 (as in FY2018), the intervention would change average gross revenue to L mynumber .

If this is correct, please proceed to the next question. Otherwise feel free to go back and adjust your answer.

Page Break

Question 3

What do you predict will be the percentage change effect of opening the email on declared taxable income (gross revenue net of deductions)? [If you believe it will increase revenue by X%, please type X].

As a reminder, in 2018, the average taxable income (gross revenue net of deductions) was L 100,000 (s.e. L 1,400), or approximately USD 4,000.

Page Break

You answered $\${Q13/ChoiceTextEntryValue}\%$ in the last question. As an illustration, if the average declared taxable income among the control group is L 100,000 (as in FY2018), the intervention would change average gross revenue to L mynumber.

If this is correct, please proceed to the next question. Otherwise feel free to go back and adjust your answer.

Page Break

Effects of different treatments

Among the treated sample, taxpayers were further randomized into receiving **three slightly different messages** in addition to the information available to the TA.

One-third of the treatment group will see a message reminding them that non-compliance makes them **subject to fines and sanctions according to the law ("Sanctions message")**. One-third will see a message reminding them about other **administrative sanctions ("Procedure denial")**, such as the denial of documents necessary for firms' operations. Finally a last group will receive a call to "moral duty", reminding them that **good citizens pay taxes that finance public goods for the kids ("Tax morale message")**.

Question 4

Using the sliders below, please predict the effect of each of these treatments on the **probability of filing**.

As a reminder, you previously stated that the effect for the pooled treated sample (i.e. pooling all treatments below together) would be $\${Q8/ChoiceTextEntryValue}$ p.p..

- _____ Sanctions message (1)
- _____ Procedure denial message (2)
- _____ Tax morale message (3)

Question 5

Using the sliders below, please predict the effect of each of these treatments on the (unconditional) declared gross revenue?

As a reminder, you previously stated that the effect for the pooled treated sample would be

$\{Q11/ChoiceTextEntryValue\}\%$.

- _____ Sanctions message (4)
 - _____ Procedure denial message (5)
 - _____ Tax morale message (6)
-

Question 6

Using the sliders below, please predict the effect of each of these treatments on the (unconditional) declared taxable income?

As a reminder, you previously stated that the effect for the pooled treated sample would be $\{Q13/ChoiceTextEntryValue\}\%$.

- _____ Sanctions message (4)
 - _____ Procedure denial message (5)
 - _____ Tax morale message (6)
-

Page Break

Treatment heterogeneity - how different taxpayers respond?

One specific question we are interested in is how different taxpayers will respond to the intervention. As the main dimension of heterogeneity, we ask you to predict **how taxpayers of different risk levels will respond**. The tax authority categorizes taxpayers in five different risk levels, according to i) discrepancies between self-reported and third-party informed data and; ii) anomalies in self-reported information, such as repeated losses and revenue inconsistent with volume of financial transactions.

In the table below we provide some descriptive statistics on each of those five groups, and then ask you to predict what will be the treatment effect in each of them for one outcome: **declared gross revenues**.

<i>Risk-level</i>	<i>Number of taxpayers</i>	<i>Corporations (%)</i>	<i>Filed income taxes (2018) (%)</i>	<i>Third-party information available (%)</i>	<i>Declared gross revenue (2018) (L1,000s)</i>
Low	6,460	10%	78%	74%	708.87
Medium-low	8,658	19%	80%	67%	1,480.60
Medium	7,270	40%	91%	72%	1,342.67
Medium-high	5,839	61%	93%	66%	1,777.21
High	3,169	48%	99%	87%	1,085.41

Question 7

Using the sliders below, please predict the **effect of being assigned to the treatment group** (i.e. we ask you to predict the Intention-to-Treat effect), in percentage change, on the (unconditional) declared gross revenue for each risk-level.

As a reminder, you previously stated that the effect for the full treated sample would be **\${Q11/ChoiceTextEntryValue}%**.

- _____ Low risk group (1)
- _____ Medium-low risk group (2)
- _____ Medium risk group (3)
- _____ Medium-high risk group (4)
- _____ High risk group (5)

End of Block: Survey questions

Start of Block: Block 3

You have reached the end of the survey.

If you want to change any of your answers you might click the back button below and review your forecasts.

If you have any comments about this survey or about the experiment, we would very much appreciate you leaving your thoughts below. Otherwise, please click the "Submit" button to submit your final answers.

End of Block: Block 3