Title
Essays on Development Economics and State Capacity

Author
de Gouvea Scot de Arruda, Thiago

Publication Date
2021

Peer reviewed|Thesis/dissertation

Essays on Development Economics and State Capacity

by

Thiago de Gouvea Scot de Arruda

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Business Administration

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ernesto Dal Bó, Chair
Professor Frederico Finan
Professor Aprajit Mahajan
Professor Reed Walker

Spring 2021

Essays on Development Economics and State Capacity

Abstract

Essays on Development Economics and State Capacity

by

Thiago de Gouvea Scot de Arruda

Doctor of Philosophy in Business Administration

University of California, Berkeley

Professor Ernesto Dal Bó, Chair

This dissertation comprises three studies on development economics and state capacity. In the first two chapters, developed through a partnership with the Tax Authority in Honduras, I examine the design and enforcement of tax policies when state capacity is low. State-provided public goods are crucial for economic development, and collecting taxes is necessary to finance the state. The first study evaluates the impact of a policy aimed at increasing tax payments from large corporations and highlights the importance of third-party information on enforcement. The second study reports the result of a large-scale experiment with taxpayers, aimed at estimating the compliance effects of informing taxpayers about the Tax Authority's knowledge on their transactions. The final chapter focuses on another dimension of state capacity and development: the selection of state personnel that are key to the delivery of public services. In that chapter, I study whether the impersonal examinations used to select judges in Brazil are predictive about their performance on the job.

In chapter 1, co-authored with Felipe Lobel and Pedro Zúniga, I study how corporations in Honduras responded to the introduction of a minimum income tax, a provision that taxes firms on revenue when reported profits are low. Minimum taxes are recommended to tax authorities in lower-income countries by the International Monetary Fund and figure prominently in current debates on international tax cooperation. Evidence on their impact on tax collection and behavioral responses of firms is nonetheless still scarce. Using non-linearities in the tax schedule introduced by the policy and tools from the bunching literature, I show that the reaction of corporations reveals substantial evasion under profit taxation and a large elasticity of reported revenue. I also document that the enforcement environment faced by firms is important in determining these responses, highlighting the fact that the trade-offs between policy tools change with enforcement capacity.

In chapter 2, co-authored with Giselle Del Carmen and Edgardo Espinal, I use experimental methods to further explore the use of third-party information to enforce tax compliance. The use of letters and emails to directly influence the behavior of taxpayers is popular among tax authorities and has been shown to positively impact tax compliance. While the marginal cost of communication interventions is very small, targeting the most responsive taxpayers can be important: informing taxpayers about knowledge of their wrongdoing and not following up by punishing non-compliers erodes the credibility of the tax authority. I present results of an experiment that randomized approximately 32,000 taxpayers into receiv-

ing information about the tax authorities' knowledge on their revenues before the income tax filing deadline. Across a range of outcomes, we estimate that our intervention had on average a null effect on compliance. The only subset of taxpayers to consistently react positively were those deemed to be of ex-ante low-risk of noncompliance. Using rich administrative data on past behavior of taxpayers and a causal forest algorithm, I find responsiveness to the intervention and predicted risk are negatively correlated. I conclude the chapter with a discussion of the implications of those findings for targeting of compliance tools.

In chapter 3, co-authored with Ricardo Dahis and Laura Schiavon, I approach another crucial dimension of the link between state capacity and economic development: the selection of public personnel. In the absence of strong incentives, public service delivery crucially depends on bureaucrat selection. Despite being widely adopted by governments, it is unclear whether civil service examinations can predict job performance. I investigate this question focusing on state judges in Brazil. Exploring monthly data on judicial output and cross-court movement, I estimate that judges account for a substantial share of the observed variation in the number of cases disposed. Using a novel data set on admission examinations, we show that judges with higher grades perform better than lower-ranked peers. These results suggest competitive examinations can be an effective way to screen candidates that will better deliver public services.

Contents

List of Figures

List of Tables

Acknowledgments

Coauthor information: Chapter 1 of this dissertation is co-authored with Pedro Zúniga from the Honduras Revenue Administration Service and Felipe Lobel from UC Berkeley. Chapter 2 is coauthored with Giselle Del Carmen from Enodo and Edgardo Espinal from the Honduras Revenue Administration Service. Chapter 3 is coauthored with Professor Laura Schiavon from Universidade Federal de Juiz de Fora and Ricardo Dahis from Northwestern University.

Acknowledgments: I am deeply grateful for the advice, mentorship and support provided by the members of my dissertation committee: Ernesto Dal Bó, Frederico Finan, Reed Walker and Aprajit Mahajan. Throughout the years they pushed me to produce better research and were extremely supportive in my journey as a PhD student. I also benefited greatly from the support of other faculty at Berkeley. Conversations with Guo Xu were particularly important in my research on personnel economics.

Several others were key in my completing of the PhD program. Stuti Khemani took me as a research assistant in the beginning of my program and became a co-author and mentor. My co-authors Ricardo Dahis and Laura Schiavon were excellent partners and friends. Giselle Del Carmen opened the doors to my research in Honduras. Two of the chapters in this dissertation were written as part of a collaboration with the Revenue Administration Service in Honduras (SAR). They were extremely generous with their time and allowed me to do some of the most exciting research. I'd like to thank, in particular, David Pinto, Pedro Zúniga, Edgardo Espinal, Oziel Aron Fernandez, Milton Maldonado, Wilman Nuñez, Roldan Enamorado, Jose Carlo Bermudez and Mario Ramos.

Sharing the experience of pursuing a PhD creates news bonds and strengthens others. Thanks to Petr Martynov, Mehmet Seflek, Abhay Aneja, Yasir Kahn, Andres Gonzalez Lira, Alex Coblin, Tomas Guanziroli, Santiago Garriga, Daniel Valderrama and many others for the conversations throughout the last five years.

Finally, my family has always been extremely supportive in each adventure I've decided to undertake and that has made all the difference. In particular, Jess was by my side celebrating each victory and supporting me in the rough patches, and for that I'm extremely grateful.

Chapter 1

Corporate Taxation and Evasion Responses: Evidence from a Minimum Tax in Honduras

1.1 Introduction

The landscape of corporate taxation has changed significantly in the last few decades. Average statutory corporate tax rates have fallen from over 40% in the 1990s to 30% in low-income countries, and by even more in middle- and high-income countries (International Monetary Fund, 2019b). At the same time, technological changes such as the rise of digital companies and the emergence of tax heavens mean that governments face increasing challenges to assure compliance in corporate tax payments (Zucman, 2014). These trends pose particularly stark threats to the tax base in lower-income countries, which often do not have the institutional capacity to fight tax evasion.

One tool already deployed by several governments to assure tax payments by corporations are minimum taxes. While corporate taxes are usually assessed on declared profits, minimum taxes are assessed on a broader base (possibly gross revenue) when reported profits are very low. The International Monetary Fund (IMF) recommends the use of minimum taxes as part of "simple measures protecting against base erosion" (International Monetary Fund, 2019a). Some form of minimum taxation on corporations is also at the core of recent international tax cooperation initiatives, such as the G20/OECD Inclusive Framework on Base Erosion and Profit Shifting (BEPS)[1].

In this paper we study corporate responses to the introduction of a minimum tax in Honduras between 2014-2017. Despite the prominence of minimum taxes in economic debates, evidence is scarce on their impact on the behavior of firms. Previous studies have

[1]The 2017 US Tax Cuts and Job Act (TCJA) also includes a provision for multinational firms to pay the maximum between corporate profit taxes and taxes on a broader base which does not allow for certain costs usually linked to profit shifting to be deducted. This so called BEAT (base erosion anti-abuse) provision effectively replaced the alternative minimum tax (AMT) for corporations, which was repealed. In the context of the ongoing debate about large corporations not paying federal taxes, President-elect Joe Biden has proposed a minimum tax on corporations with book profits above USD 100 million (Li, Watson, and LaJoie, 2020)

shown that incentives generated by minimum taxes can be used to quantify how firms over report costs under profit taxation to evade taxes (Best, Brockmeyer, Kleven, Spinnewijn, and Waseem, 2015, Mosberger, 2016, Alejos, 2018). In order to understand the implications of broadening the tax base, nonetheless, it is fundamental to also estimate how firms change their reported revenue. The specific design of the minimum tax in Honduras provides us the variation to make progress in that direction. While the elasticity of reported revenue we estimate is the relevant parameter to understand the reaction of firms under the existing enforcement environment, we also provide evidence that stronger oversight by the tax authority dampens that elasticity. The trade-offs involved in broadening the tax base are thus dependent on state capacity to enforce compliance.

Before the introduction of the minimum tax, corporations in Honduras faced a flat 25% tax on reported taxable income (profits), defined as gross revenues minus total claimed deductions. Starting in FY2014, the country introduced a minimum tax provision mandating that taxpayers declaring yearly gross revenue above L10 million (approximately USD 400,000) pay the maximum between their liability under profit taxation and 1.5% of declared gross revenue. The policy effectively introduced a floor on the effective tax rate (tax liability as share of revenues) paid by large corporations, even when reported profits were low. The minimum tax provision was widely debated: it potentially affected the 20% largest corporations in the country, faced strong opposition from the private sector, and was disputed in courts and eventually upheld as constitutional by the Supreme Court.

Using the universe of corporate tax declarations between 2011 and 2018, we start by documenting that taxpayers responded strongly to the incentives created by the minimum tax. Since firms reporting gross revenue below L10 million are exempt from the minimum tax, its introduction created a threshold where tax liability might change discontinuously in response to small changes in declared revenue. As an illustration, a firm declaring L9.99 million in gross revenue and close to zero profits will pay virtually no taxes, but declaring L10 million would create a tax liability of L150,000 (1.5%*L10 million) under the minimum tax. This *notch* (discontinuous change in tax liability) generates strong incentives for firms to strategically locate below the exemption threshold. We show that the distribution of firms declaring gross revenue in the vicinity of the exemption threshold was smooth between 2011 and 2013, but presents a clear and increasing excess mass immediately below the threshold when the minimum tax went into effect in 2014. When the exemption threshold was increased to L300 million in 2018, the excess mass around the previous notch immediately disappeared.

We use tools from the bunching literature (Kleven, 2016, Kleven and Waseem, 2013), adapted to our context, to recover bounds on the elasticity of reported revenue with respect to one-minus the tax rate - a key behavioral parameter to assess the response of firms to a policy taxing revenues. Our estimates suggest that the marginal buncher reduces their reported revenue by 15-30% to avoid being subject to the minimum tax and facing higher tax liability. We estimate revenue elasticities in the range of [0.35,1], considerably higher than previous estimates for similar contexts[2].

The large estimated elasticity emphasizes the limits faced by the tax authority in broadening the tax base: increasing tax rates will lead to a substantially smaller tax base. While

[2]Bachas and Soto (2021), for example, estimate elasticities of reported revenue in the range [0.08 - 0.33] for corporations in Costa Rica.

the revenue response could be entirely driven by real production decisions (firms decreasing sales in order to be exempt) we offer evidence that misreporting revenue is part of the explanation. We construct firm-level measures of *revenue observability*, which we define as the share of self-declared revenue that is independently observed by the tax authority through third-party reporting (VAT withholdings of suppliers being the main source of information). We show that taxpayers are more likely to locate immediately below the exemption threshold when the tax authority has limited ability to independently assess declared revenue: the excess mass below the exemption threshold is 65% larger for firms with below median revenue observability. We also explore different levels of revenue observability across industries and document the same pattern of behavior: firms in high-observability industries are much less likely to bunch below the threshold, implying a lower elasticity of reported revenues. Taken together, we interpret these as evidence that at least part of the observed response of declaring revenue below the exemption threshold is explained by misreporting and thus potentially responsive to the enforcement environment.

While firms that would have declared gross revenue slightly above the exemption threshold might report lower revenue to escape the minimum tax, larger firms will not be exempt. We document that taxpayers with revenues significantly above the threshold reduce their reported costs and increase their reported profit margins, consistent with the fact that under revenue taxation firms cannot decrease their tax liability by inflating costs. We interpret this as clear evidence of evasion under the profit taxation regime. In order to quantify these evasion responses, we explore the fact that a minimum tax creates a *kink* in the tax schedule faced by taxpayers (Best et al., 2015): both the tax rate and the tax base change discontinuously at the profit margin level that separates the two regimes, while the tax liability changes continuously.

We show that corporations in Honduras increased their reported profit margin by 0.9 - 1.1 percentage points when incentives to over report deductions disappeared. Decomposing the profit margin change between real production and cost misreporting components, and using the revenue elasticity obtained using the notch[3], we estimate that under profit taxation corporations increase their reported costs by 13 - 17% of their profits in order to reduce their tax liability. We also explore the rich administrative data to show that not all deduction categories respond in the same manner. We provide both non-parametric evidence and estimate "donut-hole" discontinuity regressions suggesting that, under profit taxation, firms systematically over-report hard-to-trace deductions, like costs linked to the purchase of goods and materials, while no over-reporting is observed in categories that generate paper trail that is easier to verify, like labor or financial costs. This is similar to findings from Mosberger (2016) in Hungary and strongly suggest a focus for tax authorities efforts in assessing the veracity of claimed deductions under profit taxation.

The previous results document strong behavioral responses to the minimum tax and illustrate the main trade-off induced by deviating from profit taxation: a broader tax base reduces tax evasion (Best et al., 2015), at the cost of efficiency loss (Diamond and Mirrlees,

[3]In the minimum tax policy studied in Pakistan, Best et al. (2015) do not have variation that allows them to estimate the elasticity of reported revenue, but show that cost adjustment estimates are robust to a wide range of elasticity values since real production incentives are very small around the kink. We can explore the fact that the exemption threshold for the minimum tax in Honduras introduces a notch in the tax schedule that allows us to estimate the elasticity of revenue and use that to pin down an evasion response.

1971). It also exemplifies the distortions introduced by tax notches: by taxing marginal revenue well above 100% when crossing an arbitrary threshold, notches induce large responses (Kleven and Waseem, 2013, Slemrod, 2013, Sallee and Slemrod, 2012)[4].

In order to quantify the impacts of the minimum tax on government revenue collection and profit of firms, as well as compare these with alternative tax policies, we impose more structure on the profit maximization problem of firms and calibrate a model using behavioral parameters estimated above. We present three main exercises. First, under our parametric assumptions, we quantify the impact of the specific minimum tax policy introduced in Honduras, considering that previously firms were taxed on profits. We estimate that the reform increased tax revenues by up to 30%, but at the cost of reducing aggregate corporate profits by 10% due to larger tax liability and production distortions. We also show the very stark incentives created by the tax notch: firms bunching below the L10 million threshold are able to reduce their tax liabilities by 75%, even though in aggregate the revenue loss from their behavior is less than 1%. We additionally present different scenarios in which the minimum tax rate and/or the revenue eligibility threshold change and assess their impacts on tax revenue and profits.

Our second scenario considers a potentially simpler policy change to increase tax revenue from large taxpayers: an increase in the *average* profit tax rate faced by corporations declaring gross revenue above the L10 million threshold[5]. This policy also creates a notch and incentivizes some firms to bunch below the exemption threshold. Contrary to revenue taxation under a minimum tax, however, here the incentives to misreport are exacerbated: firms will over-report costs even more to reduce taxable income, although production efficiency is preserved. We show that to collect the same amount of revenue as in the minimum tax regime would require an average tax rate of 40%, 15 p.p. higher than the tax rate below the threshold. While production is not distorted under the increased profit tax rate, aggregate profits fall by 20% in this scenario driven by increased evasion related losses.

Given the stark losses faced by owners of capital in the previous scenarios, our third exercise consists in simulating tax systems in which the government varies the share of costs that can be deducted and the tax rate applied to the resulting taxable income base (Bachas and Soto, 2021, Best et al., 2015). Our results highlight the intuition that, starting from a non-distortionary system where only pure profits are taxed, allowing some degree of production distortion might generate welfare gains by decreasing evasion costs incurred by firms[6]. We show that for any deduction level smaller than 85% the tax authority can increase tax revenue between 8-10% without losses in aggregate profits.

Two caveats about our results should be taken into account. First, we do not attempt to estimate who bears the incidence of corporate taxes (Auerbach, 2005, Bastani and Waldenström, 2020). While the classic result of Harberger (1962) is that capital owners

[4]Slemrod (2013) discusses in detail the use of notches and their implications for welfare. Kanbur and Keen (2014), Keen and Mintz (2004), Bigio and Zilberman (2011) discuss optimal enforcement thresholds.

[5]Tax schedules with increasing average rates across the size distribution are not uncommon in the developing world, as exemplified by Bachas and Soto (2021) in Costa Rica and Kleven and Waseem (2013) in Pakistan.

[6]As Bachas and Soto (2021) and Best et al. (2015), we refer to welfare gains considering scenarios in which aggregate profits do not fall but government revenues increase. The assumption that evasion costs are true social costs is crucial to our results (Chetty, 2009).

economy-wide bear the full incidence of corporate taxation in a closed economy, recent empirical evidence suggests that a substantial share of the tax burden is also borne by workers (Suárez Serrato and Zidar, 2016, Fuest et al., 2018). For those reasons we also do not discuss any possible redistribution motives from the minimum tax reform, since such exercises would require attributing incidence. Second, our model of firm optimization and our simulations do not consider general equilibrium effects of a broader tax base. Limiting cost deduction not only distorts firm size directly, but also cascades down production networks and distorts input prices and the size of downstream firms. Best et al. (2015) develop a general equilibrium model and show that introducing some degree of production inefficiency is still optimal when enforcement is imperfect.

This paper provides several contributions to the public finance and development literatures. First, it provides new evidence that complements previous findings about minimum taxes in Pakistan (Best et al., 2015), Hungary (Mosberger, 2016) and Guatemala (Alejos, 2018). Unlike previous studies, the specific design of the minimum tax in Honduras, including a revenue-based exemption threshold, allows us to estimate the elasticity of reported revenue for corporations, a key behavioral parameter to understand the impact of minimum taxes. We also credibly document and quantify tax evasion for large corporations under the profit taxation regime, driven by over-reporting deductions. Estimates of tax evasion for registered taxpayers are particularly absent for lower-income countries where broad, randomized tax audits are rare[7].

Second, our work provides new evidence on tax evasion in developing countries using administrative data. Londoño-Vélez and Ávila Mahecha (2019) document substantial evasion of a wealth tax in Colombia, highlighting the use of offshore accounts and of harder-to-observe wealth components as a relevant mechanism. We construct a measure of revenue observability for each firm, using third-party information available to the tax authority, and show that firms bunch less below the exemption threshold when revenue is more readily observable. Our rich administrative data allows to build on the work of Almunia and Lopez-Rodriguez (2018), which perform a similar exercise using input-output tables at the sectoral level. By documenting that availability of third-party information reduces bunching below the exemption threshold, our paper reinforces the idea that evasion responses are not fundamental primitives that govern firms' behavior, but are to some degree sensitive to the enforcement context (Fack and Landais, 2016, Slemrod and Kopczuk, 2002, Basri et al., 2019). This is consistent with other recent evidence that investment in tax authorities' capacity might generate large gains in revenue by curbing evasion (Congressional Budget Office, 2020, Sarin and Summers, 2020, Johannesen et al., 2020, International Monetary Fund, 2015).

Finally, we contribute to the growing literature on bunching methodologies that use discontinuities in the tax design to identify structural parameters (see Kleven (2016) for a recent review). While there exists extensive research on how individuals react to discontinuities in the tax schedule (Saez, 2010, Bastani and Selin, 2014, Kleven and Waseem, 2013), we contribute to the more limited literature on how corporations respond to these incentives, similarly to the work of Bachas and Soto (2021) in Costa Rica and Devereux et al. (2014)

[7] Trigueros et al. (2012) document that only nine out of eighteen surveyed countries in Latin America have any estimate of evasion available, for any kind of tax. Our estimates for Honduras refer to tax evasion by large corporations filing income tax and do not consider other margins such as non-registration or non-declaration.

in the United Kingdom.

The rest of the paper is organized as follows. In Section 2 we present the context of corporate taxation in Honduras, discuss in detail the minimum tax provisions and describe our sample. In section 3 we present a model of profit maximization by firms that illustrates how we can expect corporations to react when faced with the introduction of a minimum tax. In Section 4 we first present non-parametric evidence of corporate behavior under the minimum tax and then show how this evidence can be used to recover structural parameters of interest. We provide robustness exercises that strengthen our argument that we identify responses to the minimum tax in section 5. In section 6 we present a calibrated model of the decision of firms and simulate the impact of alternative tax systems. We conclude in section 7.

1.2 Institutional Context and Data

We study a reform that introduced a minimum tax on corporations in Honduras, a lower middle-income country in Central America with a population of 9 million and per capita GDP of $5,800 PPP in 2018. The level and composition of government tax revenues in Honduras is comparable to other countries with similar per capita income. First, total tax revenues represent around 18% of GDP, significantly below the average of 25% observed in high income OECD countries[8]. Second, the country is much more reliant on goods and services taxes, representing over 50% of total tax revenue, than on income taxes, which amount to one-third of total tax revenue. Finally, corporate income taxes are equivalent to 4% of GDP, almost twice as much as personal income taxes (International Monetary Fund, 2018a). These last two facts are broadly consistent with the perception that lower income countries face significant informational constraints in assessing more complex tax liabilities and therefore rely more on broader sales taxes and/or taxing large corporations (Gordon and Li, 2009). Recent years have witnessed significant efforts to improve tax collection capacity in the country, including a broad overhaul of the tax authority agency in 2015[9]. Since then the number of income tax filers has doubled (from 74,000 to almost 150,000) and the share of electronic declarations has increased by 16 percentage points to 81%.

Non-incorporated taxpayers (*Personas Naturales*) are approximately 80% of the total

[8]These numbers refer exclusively to taxes and exclude important revenue components such as social security contributions. Considering total revenue, the OECD average revenue-to-GDP ratio is 35% while in Honduras it stays close to 20%, making the gap even starker.

[9]Starting in 2013, the government of Honduras restructured several public institutions under the oversight of the "Centralized and Decentralized Public Administration Reform Commission". The reform of the tax authority (formerly known as DEI, *Dirección Ejecutiva de Impuestos*) was led by the Interamerican Development Bank (IDB). The diagnostic before the reform was that "most administrative and technical staff do not have the basic profiles or training levels required for managing tax administration properly." (Interamerican Development Bank, 2015). Other shortcomings described were similar to what International Monetary Fund (2015) identifies as key challenges to tax administration in many countries: high turnover of senior staff, lack of IT personnel and infrastructure, and lack of professional development. Over 1,500 workers were dismissed and new hires were performed by an international, independent human resources firm.

number of income tax filers and face a progressive tax schedule on labor income[10]. Corporations (*Personas Jurídicas*), on the other hand, face a 25% flat tax rate on taxable income, defined as gross revenues minus standard deductions such as wages, raw materials, depreciation of capital, interests paid and carryover losses[11]. Fiscal years in Honduras run according to the calendar year and taxpayers must file the income tax declaration by April 30th.

The minimum tax studied in this paper was introduced in 2014 as part of the broader "Public Finance Management, Exemptions' Control and Anti-Evasion Measures" tax law[12]. The two main features of the minimum tax are as follows. First, it exempts taxpayers reporting gross revenue below L10 million[13], which are still liable for a 25% rate on declared taxable income. Second, taxpayers reporting gross revenue above L10 million are liable for a minimum of 1.5% of their reported revenue. When filing the yearly income tax declaration, corporations must compute their tax liability under the usual profit regime and the 1.5% regime, and are liable for the largest of the two. Since profits are taxed at 25%, a taxpayer declaring 6% profit margin (reported profits divided by gross revenue) will face a tax liability equivalent to 25%*6% = 1.5% of gross revenues and will be located exactly at the edge between the two regimes.

The immediate objective of the minimum tax was to create a floor to the effective tax rate (tax liability divided by gross revenue) faced by large taxpayers: regardless of declared profits, corporations with revenue above L10 million should pay no less than 1.5% of their declared gross revenues in taxes. In Figure 1.1, panel A, we present evidence that the policy substantially raised the effective rate faced by large corporations.

In the period 2011-2013, before the minimum tax was in place, the median effective rate faced by firms with gross revenue around L10 million was approximately 0.5%. Between 2014 and 2017, when the minimum tax is in place for firms declaring revenue above L10 million, the median effective rate substantially changes around the threshold. Firms declaring gross revenues below that level still face an effective rate close to 0.5%. Corporations with revenue above L10 million, however, are now subject to the minimum tax and the median firm faces an effective rate of exactly 1.5%. The figure also illustrates the *notch* generated by the minimum tax: by declaring gross revenue marginally above L10 million firms face a discontinuous increase in their tax liability. While in panel A we focus on corporations around the exemption threshold, in panel B we document that the policy was effective in increasing the median effective rate for all firms declaring gross revenue well above the threshold.

The increase in effective tax rate for firms above the exemption threshold is driven by firms that declare low profit margins but no longer pay very small tax liabilities. We illustrate that fact in Figure 1.2, where we plot effective tax rates for firms declaring different

[10]The progressive tax schedule is updated yearly to account for inflation and includes four brackets with increasing marginal tax rates. In FY2019, income below L158,995 (approximately USD 6,400) was exempt and amounts above that face increasing marginal rates of 15%, 20% and 25%. Income from other sources such as dividends, interest and capital gains are taxed at a 10% flat rate.

[11]Throughout the paper we use the terms "taxable income" and "profits" interchangeably, always referring to the base taxed at 25%.

[12]The 2014 tax law also increased VAT rates from 12% to 15%, made permanent a surcharge of 5% on taxable income above L1 million and introduced a 10% tax on dividends received by residents.

[13]Approximately USD 400,000 using the average market exchange rate in 2018 (USD 1 = L24.5). This is the exchange rate used throughout the paper when mentioning US dollar amounts.

profit margins. In the period 2011-2013, before the introduction of the minimum tax, the relationship between declared profit margin and tax liability is approximately linear for all profit margin levels. With the introduction of the minimum tax, however, the relationship between profitability and tax liability changes for firms with profit margins below 6%, which face a minimum tax liability equivalent to 1.5% of their gross revenue. For those firms, the incentive to declare lower profits in order to reduce their tax liability disappears. The figure also illustrates that the policy introduces a *kink* in the budget set of taxpayers exactly at the 6% threshold, with a change in the slope of the tax schedule.

Three special provisions of the minimum tax law are worth discussing in more detail. First, taxpayers in certain sectors (cement, state enterprises, pharmaceuticals and bakery) face a 0.75% rate instead of 1.5%. Firms in those sectors are less than 2% of taxpayers, so we exclude them from our main analysis and present separate results showing their behavior is also consistent with predictions from theory. Second, we also exclude from our main analyses firms operating in petroleum-related sectors and those in their first two years of operations, which are exempt from the minimum tax[14]. As discussed below, the number of corporations filing taxes is rapidly increasing in the period of study and "young" firms represent up to 25% of taxpayers in some years. Nonetheless, they are predominantly small, with declared gross revenue well below the exemption threshold. Finally, firms declaring losses are also exempt from the minimum tax. This feature is potentially relevant to our empirical exercises, since that might create strong incentives for low profit firms to report negative results. In practice, however, this behavior is very limited due to the existence of a net asset tax that also applies to firms reporting losses.

Despite being part of a larger tax reform, the minimum tax provision was highly salient and widely debated in the public sphere. A previous attempt to institute a 1% minimum tax in 2011 was ruled unconstitutional by the Supreme Court and never went into effect. The 2014 reform was again challenged in the courts but eventually upheld as constitutional in 2015, and stayed in place until FY2017. In the aftermath of highly contested elections in that year, the government approved a series of policy reforms that included the gradual phasing out of the minimum tax provision. For FY2018, the exemption threshold was raised from L10 million to L300 million. While approximately 20% of corporations declared gross revenue above L10 million before the introduction of the minimum tax, only 1.3% declared revenues above L300 in 2017. The law additionally established further increases in the exemption threshold to L600 million in FY2019 and L1 billion in FY2020, meaning that very few corporations would be affected by the minimum tax at the end of this period (International Monetary Fund, 2018a).

1.2.1 Data and descriptive statistics

The main analyses in this paper are based on administrative data comprising the universe of income tax declarations from corporations in the 2011-2018 period. We supplement this data, in additional exercises, with monthly VAT declarations and third-party information on taxpayers' transactions. Electronic filing by corporations has steadily increased in the period,

[14]Both exemptions in the first years of operation and lower rate for sectors such as pharmaceuticals are common features of minimum tax regimes across the world.

from less than 60% of total declarations in 2011 to almost 85% in 2018[15]. Throughout the paper, we exclude taxpayers in special regimes that exonerate them from paying any income taxes[16]. The resulting dataset is an unbalanced panel of over 180,000 firm-year observations and approximately 41,000 unique firms.

We present basic descriptive statistics of our sample in Table 2.1 for years 2013-2018, highlighting the following facts. First, the number of corporations filing income tax has steadily increased throughout the period, from less than 20,000 in 2013 to approximately 30,000 in 2018. While in our main estimates we use an unbalanced panel of taxpayers, we show that firm's responses to the minimum tax are qualitatively similar in a balanced panel of corporations that file every year. Second, average reported gross revenue was around L30 million (USD 1.2 million) but with wide dispersion: the median corporation in the sample had yearly gross revenues of L1.2 million (USD 48,000) and over 80% reported revenues below L10 million. Third, average pre-tax profit margins steadily increase throughout the period, from less than 2% in 2013 to almost 5% in 2018. As discussed below, part of this increase is likely explained by the introduction of the minimum tax, which induced a decrease in claimed deductions and consequent increase in reported profits for large corporations. Despite that, average profit margins are always well below 6%, meaning that the average tax liability under profit taxation is less than 1.5% of gross revenues. Fourth, even though the minimum tax is not directly aimed at multinational corporations (MNC) operating in the country, these are disproportionately large and thus potentially affected by the policy: even though MNCs represent only 2-4% of corporate filers, they pay approximately 60% of taxes[17]. Finally, even though only a small fraction of firms end up liable for minimum taxes (between 6-8% in 2014-2017), they contribute 20-30% of total corporate tax revenues. Indeed, despite the number of firms liable for minimum taxes falling by an order of magnitude in 2018, when the exemption threshold increased, their contribution to total corporate tax revenues was still close to 15%.

In order to illustrate the relevance of the largest corporations to tax collection, we present in Table 1.2 the share of total revenue and taxes declared by the largest taxpayers. In 2013, before the introduction of the minimum tax provision, the largest twenty corporations in terms of gross declared revenue (top 0.1%) declared almost 30% of total revenues and accounted for 32% of total corporate taxes. Almost 70% of taxes were generated by the top 1% corporations and the top 10% (approximately 2,000 firms) paid more than 90% of taxes[18]. This skewness in the distribution of firm size highlights the potential of the minimum tax to

[15] Our dataset encompasses declarations using three type of forms: DEI-350 was an electronic form discontinued in 2015, when the more detailed SAR-357 was introduced. Throughout the period, taxpayers could also use a paper form (SAR-352) which provides less detailed information on both revenues and deductions.

[16] Approximately 3-5% of corporations in each period, mostly export-oriented manufacturing firms.

[17] Multinational corporations are defined as firms filing transfer price declarations at some point in the period 2014-2017. The potential for the minimum tax to increase tax collection from MNCs depend not only on their gross revenues but also on their profit margin in the absence of minimum taxation.

[18] This is similar to what Devereux et al. (2014) report for corporations in the United Kingdom (top 1% account for 80% of corporate income taxes) and Almunia and Lopez-Rodriguez (2018) report for Spain (top 2% report 80% taxable profits.). In the United States, Auerbach (2005) mentions that the largest 0.04% corporations in terms of assets account for 62% of all corporate income tax in 2001. In a more similar context, Bachas and Soto (2021) document that the largest 20% corporations account for 87% of corporate taxes, which is a substantially smaller share than in Honduras.

significantly increase revenue collection despite exempting approximately 80% of firms.

1.3 Conceptual framework

1.3.1 Model of firm optimization

In this section we present a stylized model of profit maximization by firms in line with the classical approach of Allingham and Sandmo (1972) and adapted by Best et al. (2015) to illustrate the incentives introduced by a minimum tax and motivate the empirical exercises that follow. Firms choose a production level y and the level of costs $\hat{c}$ reported to the tax authority, which might be higher than true costs of production given by an increasing and convex function $c(y)$[19]. Firms face an increasing and convex loss in the amount of cost misreported given by $g\big(\hat{c}-c(y)\big)$, with $g(0)=0$[20]. Since a regime with a minimum tax allows for both profit and revenue taxation, we model the possibility that only a share $\mu \in [0,1]$ of costs can be deducted to obtain the taxable income, taxed at rate τ. Firms then choose the vector $(\hat{c}, y)$ to maximize after-tax profits:

$$\underset{(\hat{c},y)}{Max} \quad \Pi(\hat{c},y) = y - c(y) - \tau\big(y - \mu\hat{c}\big) - g\big(\hat{c} - c(y)\big) \tag{1.1}$$

Under a linear tax schedule, first-order conditions are:

$$g'\big(\hat{c} - c(y)\big) = \tau\mu \tag{1.2}$$

$$c'(y) = \frac{1-\tau}{1-\tau\mu} = 1 - \tau\frac{1-\mu}{1-\tau\mu} = 1 - \tau_E \tag{1.3}$$

When choosing how much costs/deductions to report, firms equalize the marginal cost of misreporting deductions to the marginal benefit $\tau\mu$ (not paying tax rate τ on share μ of marginal reported cost). Similarly, the level of production is obtained by equalizing the marginal benefit of producing one extra unit of output $1-\tau$ to the marginal cost $c'(y)(1-\tau\mu)$, which crucially depends on how much of costs can be deducted to obtain taxable income. We re-write Equation 1.3 so that firms equalize the marginal cost of production to $1-\tau_E$, the net-of-tax benefit of marginally increasing production.

Under a pure profit taxation regime, when all production costs can be deducted ($\mu = 1$), we have that $\tau_E = 0$ and $c'(y^*) = 1$: taxes on pure profits are non-distortionary and firms choose the efficient level of production. In the other extreme, when $\mu = 0$ firms pay taxes on their gross revenue and $\tau_E = \tau$ and $c'(y_r) = 1 - \tau \implies y_r \leq y^*$. That is, firms are sub-optimally small since the marginal benefit of an extra unit of revenue is $1-\tau$. For any interior value of $\mu \in (0,1)$, production levels will be below optimal.

While taxing a broader base than profits induce distortions in productions levels, the opposite is true for evasion levels: under revenue taxation Equation 1.2 becomes $g'\big(\hat{c}-c(y)\big) = 0$ and then $\hat{c} = c(y)$. When costs are not deductible, firms have no incentive to misreport

[19]We assume output prices are fixed and equal to p = 1, so we can express revenue equal to production.

[20]In our stylized model we consider that firms can only misreport costs and not revenues. This is a simplifying assumption we make to illustrate the idea that it is easier to misreport costs than revenue.

and so report truthfully. Increases in costs deductibility μ induce firms to increase their reported costs in order to reduce tax liability, but also produce misreporting losses[21].

1.3.2 Incentives under the minimum tax

Informed by the model, we now discuss the change in incentives faced by firms that were initially subject to a 25% flat rate on profit and face the introduction of a minimum tax. We can write the tax liability faced by firms as

$$T(y,\hat{c}) = \begin{cases} 0.25 * \left(y-\hat{c}\right), & \text{if } y < 10,000,000 \\ Max\{0.25 * \left(y-\hat{c}\right), 0.015 * y\}, & \text{if } y \geq 10,000,000 \end{cases} \tag{1.4}$$

Consider first firms with gross revenue significantly above L10 million and therefore not exempt from the minimum tax. From the expression above, the tax liability under profit and revenue taxation will be the same whenever the declared profit margin $(y-\hat{c})/y$ is equal to $0.015/0.25 = 6\%$. Corporations which in the absence of the minimum tax would have reported profit margins above 6% have no incentive to change their behavior: their liability under profit taxation is still larger than 1.5% of their revenues, so they effectively do not face a different regime. Firms which declare positive profit margins below 6%, on the other hand, now face a tax of 1.5% on their gross revenues instead of 25% on declared taxable income. According to the model discussed, this induces changes in two dimensions. First, production decisions are now distorted (since $\tau_E = 0.015$) and firms will reduce production/revenues. Under the assumption of decreasing returns to scale, that effect will lead to an increase in firms' profit margins (Best et al., 2015). Second, under revenue taxation firms will not over-report costs, since misreporting entails losses but no longer provides the benefit of reducing tax liability. Both effects will cause the pre-tax profit margin distribution to shift right. Since taxpayers reporting profit margins above 6% are not affected, only the distribution below 6% is shifted and we should observe an excess mass around that threshold.

Consider now the incentives faced by firms that, absent the minimum tax, would have declared gross revenue slightly above the L10 million exemption threshold. Just as discussed above, firms that would have declared profit margins above 6% face no change in incentives and will still choose the same revenue and declared cost levels as they would under pure profit taxation. Low profit firms, however, now face a different decision. They might declare gross revenue above L10 million and adjust their production and evasion decisions in response to the 1.5% minimum tax liability. But they might also decide to decrease revenue to slightly below L10 million so that they are exempt from the minimum tax and pay the profit tax. Unlike notches generated by wealth (Londoño-Vélez and Ávila Mahecha, 2019) or gross income taxes (Kleven and Waseem, 2013), where all taxpayers above the notch see their liability discontinuously increase, in our setting only a subset of taxpayers are affected by the notch (Bachas and Soto, 2021). The benefit of declaring revenue below the exemption

[21]Importantly for welfare evaluation, we interpret these evasion losses as social losses, such as the costs of keeping parallel accounting systems or avoiding entering certain economic transactions that might reveal true costs. As discussed by Chetty (2009), implications for welfare analysis differ if evasion costs are actually seen as transfer between agents (fines paid to the government, for example) or if perceived costs are different from actual costs.

threshold, i.e., of bunching, is inversely proportional to the profit margin that would be declared in the absence of the minimum tax.

To see that, consider the profits of a hypothetical taxpayer that must decide between choosing a production level marginally below the exemption threshold (bunching) y^T and reporting cost $\hat{c}$, or producing y_0 above the threshold, reporting true costs $\hat{c}_0 = c(y_0)$ and paying the minimum tax:

$$\Pi(y^T, \hat{c}|Bunch) = y^T - \tau_\pi\big(y^T - \hat{c}\big) - c(y^T) - g\big(\hat{c} - c(y^T)\big) \tag{1.5}$$

$$\Pi(y_0, \hat{c}_0|NotBunch) = y_0 - \tau_y y_0 - c(y_0) - \underbrace{g\big(\hat{c}_0 - c(y_0)\big)}_{=0} \tag{1.6}$$

in which the term of cost misreporting will be zero since staying above the threshold means being taxed on revenue, so there is no incentive to overreport costs.

The gains from deciding to bunch can therefore be written as

$$Bunching\ Gains \approx \underbrace{(y^T - y_0)}_{\leq 0} - \underbrace{\big(c(y^T) - c(y_0)\big)}_{\leq 0} - \underbrace{(\tau_\pi y^T - \tau_y y_0)}_{\geq 0} + \tau_\pi \hat{c} - g\big(\hat{c} - c(y)\big) \tag{1.7}$$

The expression above breaks down the change in profits when deciding to bunch. The first two terms capture the fact that, when bunching, firms will reduce real output, therefore losing revenue, but also reducing costs. The third term captures the fact that bunching means paying a much larger tax rate on gross reported revenues (25% vs. 1.5%), while the fourth term captures the main benefit of bunching: the opportunity to deduct 25% of all reported costs when being taxed on profits instead of revenue. This highlights the fact that the incentive to bunch is directly proportional to costs: for any given level of revenues, firms with higher costs have a stronger incentive to bunch since they will be able to deduct those costs from their tax base when bunching[22]. The fifth term captures the negative effects for the firm in misreporting costs, which is increasing in the distance between true and reported costs.

1.4 Empirical results

We start this section providing non-parametric evidence that taxpayers responded to the introduction of the minimum tax in a manner consistent with the model described above. We then proceed to explore these behavioral responses in order to recover structural parameters of interest.

1.4.1 Evidence of behavioral responses

We start presenting evidence that, consistent with the simple model outlined previously, taxpayers responded to the existence of the exemption threshold by reporting gross revenue immediately below L10 million. In Figure 1.3, we present the empirical densities of reported

[22] As discussed in section 1.2, firms reporting negative profits are not liable for the minimum tax. Incentives to bunch are therefore largest for firms with high costs but positive profits, and turn to zero when firms incur losses.

gross revenues separately for three periods: 2011-2013, before the introduction of the minimum tax; 2014-2017, when the policy was in place with a L10 million exemption threshold; and 2018, when the exemption threshold was increased to L300 million. It is clear that, in the absence of the notch created by the minimum tax, the distribution of reported revenue is smooth throughout the interval. In the period when the minimum schedule creates a notch at L10 million, however, corporations respond by adjusting their reported revenue to slightly below the threshold: there is a clear excess mass of firms in that region, and a more diffuse absence of mass slightly above. Consistent with the theory presented previously, there is no "hole" in the distribution immediately above the L10 million notch, since the minimum 1.5% effective rate is not binding for firms with high enough profit margin[23]. Furthermore, we highlight that the bunching in reported gross revenue might be driven by real production responses, by under reporting real revenue or by a mix of the two. We return to this issue below and provide evidence that at least part of this behavior is driven by misreporting.

While firms immediately to the right of the notch have a strong incentive to bunch at the L10 million threshold, firms that would have reported much larger revenue are infra-marginal to this bunching behavior. As discussed above, the introduction of the minimum tax leads affected firms to decrease evasion through misreporting and decrease scale, increasing reported profit margins. Since only firms otherwise declaring profit margins below 6% are affected we should observe an excess mass of firms exactly at the kink. In practice we often observe a diffuse mass in the vicinity of the kink (Saez, 2010). In Figure 1.4, Panel A, we present the empirical density of reported profit margin for firms declaring revenue above L13 million, and therefore infra-marginal to the bunching behavior at the notch, separately for 2011-2013 and 2014-2017. In the period before the introduction of the minimum tax, we observe a steep negative slope in the density of profits, smoothly distributed around the 6% kink. With the introduction of the minimum taxation in 2014, the distribution becomes starkly different as predicted by theory: there is much less mass around positive but close to zero profit margins and firms bunch around the 6% kink.

While in Panel A of Figure 1.4 we illustrate the change in profit margin density before and after the introduction of the minimum tax, in panel B we present empirical densities for the period 2014-2017, while the minimum tax was in place, separately for firms with reported revenue significantly below and above the L10 million exemption threshold. The pattern is remarkably similar to Panel A: firms declaring revenue below the exemption threshold, and therefore unaffected by the minimum tax, are much more likely to declare low profit margins, while those under the minimum tax regime declare higher profit margins and bunch at the 6% kink. We interpret these differences in reported profit margin as evidence that corporations over-report costs under profit taxation to evade taxes, and adjust their behavior when taxed on revenues.

The previous set of figures are strong evidence that the minimum tax was a highly salient

[23]As discussed by Kleven and Waseem (2013) and Gelber et al. (2020), among others, some firms might not respond to the incentives to bunch due to inattention, high adjustment costs or some combination of other frictions. We discuss below how we interpret the existence of such taxpayers in our elasticity estimates.

policy change that induced behavioral responses from the taxpayers[24]. In the remaining of this section we explore how these responses can be used to identify parameters of interest.

1.4.2 Revenue elasticity at the L10 million notch

In order to translate the observed behavioral responses presented above into estimates of parameters underlying firms' behavior we use tools from the bunching literature. The core insight developed by Saez (2010) is that non-linearities in the tax schedule faced by taxpayers will generate bunching, the amount of which is proportional to the elasticities governing the behavior of taxpayers. Our first step is then estimating the counterfactual distribution that would have prevailed in the absence of these discontinuities, so that we can obtain an estimate of the excess bunching and relate that to underlying behavior.

We first discuss how the bunching in response to the L10 million threshold can be used to estimate the elasticity of reported revenue. As previously shown, the exemption threshold generates a notch, where tax liability discontinuously changes for some taxpayers. According to our model, firms deciding to locate exactly at the notch (bunchers) come from a continuous region $[y^T, y^T + \Delta Y]$, where $y^T = L10$ million.

To recover the counterfactual gross revenue density, we fit a polynomial regression to the empirical density of revenue, including dummies for the "excluded region" - the area around the notch affected by the policy (Saez, 2010, Chetty, Friedman, Olsen, and Pistaferri, 2011). We then predict the counterfactual density for the entire distribution ignoring the dummies, extrapolating the polynomial prediction to the bunching area and assuring a smooth counterfactual distribution around the notch[25].

We first collapse the data in bins of L100,000 (USD 4,080) of revenue and estimate[26]:

$$n_j = \sum_{k=0}^{5} \beta_k y_j^k + \sum_{b=y_L}^{y_H} \gamma_b \mathbb{1}\{y_j = b\} + \epsilon_j \tag{1.8}$$

where n_j is the number of observations in bin j, y_j are the revenue midpoint of bin j, $[y_L, y_H]$ is the excluded region affected by the notch and $\mathbb{1}\{y_j = b\}$ are dummies indicating that bin j belongs to the excluded region.

The predicted counterfactual density is defined as $\hat{n}_j = \sum_{k=0}^{5} \hat{\beta}_k y_j^k$. We can then obtain the excess mass of taxpayers below the threshold as the difference between the empirical and predicted densities $\hat{B} = \sum_{b=y_L}^{y_N} (n_j - \hat{n}_j)$, where y_N is the bin with upper bound equal to the notch.

[24] In Figure 3.5 we present jointly the change in reported revenue and profit margins using heatmaps. With the introduction of the minimum tax (Panel B), an excess mass of firms declare revenue immediately below the L10 million exemption threshold and, for larger firms, increase their reported profit margins up to 6%.

[25] The assumption of a smooth distribution is not a trivial one, as pointed by Blomquist and Newey (2017) and Bertanha et al. (2018). In particular, they show that kinks cannot identify the elasticity of taxable income if we allow for unrestricted heterogeneity of preferences. In our setting, we can partially alleviate concerns about the counterfactual density by showing, as we do in Figure 1.3, that the density was indeed smooth around the threshold before and after the existence of the notch.

[26] Our baseline specification uses a fifth-order polynomial on revenue. We present robustness exercises to that choice in the appendix.

The credible estimation of the counterfactual density requires the excluded region to be correctly determined - all those bins affected by the existence of the notch/kink in the tax schedule should not be used to estimate the counterfactual density. We follow the method pioneered by Kleven and Waseem (2013) when taxpayers face notches: while the lower bound of bunching is visually determined, we use the convergence method to obtain an upper bound for the affected region. We exploit the fact that, according to our model, the excess mass observed immediately below the notch ($\hat{B}$) must be equal to the missing mass above $\left(\hat{M} = \sum_{b=y_N}^{y_u}(n_j - \hat{n}_j)\right)$, so we recursively estimate Equation 1.8 increasing the upper bound y_H until $\hat{B} \approx \hat{M}$[27], at which point we determine that to be the upper bound.

Empirical revenue densities for each year and estimated counterfactual densities are presented in Figure 1.5. In each figure we provide estimates of the total excess number of firms (B), the excess mass of firms as a share of average density in the bunching region (b), the upper bound of the bunching region estimated using the convergence method (y_u) and the number of underlying observations used in each graph (N)[28]. At each year between 2014 - 2017, we estimate an excess between 80 and 150 taxpayers around the cutoff - between four and six times as many firms as the average counterfactual density in the bunching region. The estimated upper bound using the convergence method varies from L11.4 million in 2016/2017 to L13 million in 2015. In Figure 1.6 we provide estimates pooling all corporate filings in the 2014-2017 period, providing us with a larger sample. Our estimates indicate that the excess mass below the notch is equivalent to 5.5 times the predicted counterfactual density and that the marginal buncher would have reported gross revenue of L11.8 million in the absence of the notch, effectively reducing their declared revenue by over 15% in order to avoid the minimum tax. The results for each year and for the pooled sample are presented in columns (1) - (4) of Table 1.3.

In order to recover the elasticity of reported revenue from the behavioral responses estimated above, we adapt the reduced-form approximation developed by Kleven and Waseem (2013). We can show that, for a given revenue response ΔY by the marginal buncher, the elasticity of reported revenue is given by[29]:

$$\epsilon_{y,(1-t)} = \left(\frac{1}{\tau_y\left(2 + \frac{\Delta Y}{Y^T}\right) - 2\tau_\pi \frac{(Y^T - \hat{c})}{Y^T}}\right)\left(\frac{\Delta Y}{Y^T}\right)^2 \tag{1.9}$$

Importantly, the estimated elasticity depends not only on the change in reported revenue, but also on the cost that would have been reported when bunching, since the tax base changes from gross revenue above the notch to reported profits below it.

We will compute lower and upper bounds on the true structure elasticity. The convergence method used to obtain the upper bound of the bunching region provides an estimate of the counterfactual revenue of the marginal buncher. Under the assumption of homogeneous elasticity across all taxpayers, the response of the marginal buncher allows us to recover

[27]Since we estimate the regression using discrete bins, we determine $\hat{B} \approx \hat{M}$ to mean that $|(\hat{B} - \hat{M})/\hat{B}| \leq 0.03$.

[28]Standard errors are obtained by bootstrapping the entire estimating procedure, resampling errors from Equation 1.8 500 times.

[29]This approximation relies on the implicit marginal tax rate between the notch and the counterfactual revenue of the marginal buncher. We present the derivation of the formula in Appendix A.1.

the structural revenue elasticity. If elasticities are heterogeneous, however, the convergence method recovers the response of the taxpayer with higher elasticity (Kleven and Waseem, 2013, Londoño-Vélez and Ávila Mahecha, 2019). For that reason, we consider our estimate using that method as an upper bound on the true structural elasticity.

While the convergence method provides the revenue response of the marginal buncher, we still need the counterfactual cost to estimate the elasticity. Our model indicates the answer: since the marginal buncher is the taxpayer with strongest incentive to bunch and incentives are inversely proportional to the profit margin, the marginal buncher has close to zero profits. That allows us to set $Y^T - \hat{c} = 0$ in Equation 1.9 and write the reported revenue elasticity as a function of known policy parameters and the estimated revenue response of the marginal buncher:

$$\epsilon_{y,(1-\tau)} \approx \left(\frac{1}{\tau_y}\right)\left(\frac{1}{2+\frac{\Delta Y}{Y^T}}\right)\left(\frac{\Delta Y}{Y^T}\right)^2 \tag{1.10}$$

We present results of the estimated upper bound of the elasticities in column (5) of Table 1.3. The key quantity needed to obtain the upper bound estimate is the revenue response of the marginal buncher, estimated using the convergence method and presented in column (4). These estimates yield upper-bound revenue elasticities in the interval of $[0.6, 2.6]$. Estimates are particularly large in 2014 (1.3) and 2015 (2.6), when the upper bound of the bunching region is estimated to be above L12 million. Estimates for 2016 and 2017 are very similar (0.61) and smaller than our preferred estimate using the pooled sample ($\epsilon_y = 0.99$). We also note estimates are noisy, with very large standard errors[30].

We now turn to the estimation of the lower bound of the revenue elasticity. Our approach is similar to the "bunching-hole" method proposed by Kleven and Waseem (2013), but adapted to take into account the fact that bunching incentives depend on firms' profit margins (Bachas and Soto, 2021). We provide a brief description here and save details for Appendix A.2. Since the decision to bunch depends both on counterfactual revenue and costs, we can rewrite Equation 1.9 to find the counterfactual cost that would make a taxpayer indifferent between bunching or not, given a distance ΔY from the threshold and elasticity ϵ_y:

$$\hat{c}^* = Y^T\left(1 - \frac{\tau_y}{\tau_\pi}\right) - \frac{\tau_y}{\tau_\pi}\frac{\Delta Y}{2} + \frac{(\Delta Y)^2}{2\epsilon_y \tau_\pi Y^T} \tag{1.11}$$

Since the incentives to bunch are inversely related to profit margins, we know that if a taxpayer with revenue $Y^T + \Delta Y$ and cost $\hat{c}^*$ is indifferent to bunching, all taxpayers with lower profit margins should also bunch since they face even stronger incentives. If we knew the counterfactual profit margin distribution, we could compute the share of taxpayers bunching for each revenue bin, for a given elasticity, and compare the total amount of predicted bunching to our estimated excess mass below the notch. In order to implement that strategy, we need to make an assumption about the unobserved counterfactual profit

[30]Standard errors are estimated by bootstrap and the empirical distribution of estimated elasticities is highly non-symmetrical: for the pooled sample where the point estimate is 0.99 the empirical 95% confidence interval is $[0.7, 5.7]$, meaning there is significant uncertainty on the upper bound of the estimate, but little on the lower bound.

margin distribution above the threshold. We assume the profit margin distribution for firms reporting revenue in the interval L6 - 8 million, significantly below the notch, is a good approximation for the unobserved distribution (Bachas and Soto, 2021). We then compute the estimated elasticity as the one generating a predicted amount of bunching equal to the excess mass observed below the notch, among a range of elasticity values.

One important caveat of the lower bound methodology is that we consider that all taxpayers that have an incentive to bunch will do so. There is ample evidence, nonetheless, that even when facing strictly dominated regions some taxpayers do not bunch (Kleven and Waseem, 2013, Gelber et al., 2020). While notches often give rise to strictly dominated regions for all taxpayer and allow researchers to estimate optimization frictions, that is not the case with the exemption notch in Honduras. Since the size of the discontinuous change in tax liability depends on counterfactual profit margin, the existence and extent of dominated region also depends on the counterfactual profitability. While it is possible to make stronger assumptions, ruling out extreme preferences in order to estimate optimization frictions (Best, Cloyne, Ilzetzki, and Kleven, 2020), we abstain from doing so and consider our estimates to be lower bounds for the true reported revenue elasticity: the existence of optimization frictions require, all else equal, a larger elasticity to obtain the same amount of predicted bunching mass.

We present lower bound estimates for ϵ_y in column (6) of Table 1.3. Here estimated elasticities are both much lower and more stable across years, and likewise much more precise and statistically different from zero in every period. While the elasticity is lower (0.2) in 2014, when we observe significantly less bunching, for the period 2015-2017 and the pooled sample estimates lie tightly between 0.35 - 0.4.

We take results for the pooled sample as our preferred estimates, where we obtain a range for the reported revenue elasticity of $[0.35, 0.99]$[31]. These are substantially larger than the estimates obtained by Bachas and Soto (2021) for corporations in Costa Rica, for example, where the similar range using lower and upper bound estimates is $[0.08, 0.33]$. They are also much larger than estimates of individual earnings elasticities in Pakistan obtained by Kleven (2018), which mostly fall in the range $[0.05, 0.3]$[32]. Overall, our results suggest that, under the existing enforcement environment while the minimum tax was in place, the reported gross revenue of corporations was highly elastic, limiting to some extent the ability of the tax authority to increase revenues through higher tax rates.

1.4.3 Real or misreporting response at L10 million notch?

The observed response in declared gross revenues under the minimum tax could be due to real production decisions, to under reporting of realized revenues or to a mix of both. In

[31]We perform robustness exercises for the estimated elasticity of reported revenue in Table A2, using different polynomial orders. For the lower bound elasticity, the estimate is unchanged using a higher order polynomial but somewhat larger (0.5 - 0.6) when using a lower order polynomial. Consistent with noisy estimates in our preferred specification, however, estimates for the upper bound vary significantly when using different polynomials, ranging from 1.6 to 6.

[32]The estimates in Kleven (2016) are also noisy when using the convergence method, and for one of the notch points the upper-bound elasticity is above unit. In the different context of wealth taxation, Londoño-Vélez and Ávila Mahecha (2019) reports wealth elasticities as large as 4 using the convergence method.

this section we explore the evidence related to these possibilities.

We investigate whether the amount of bunching is related to the availability of third-party information (TPI) about the sales of taxpayers. Previous studies have documented much less bunching in response to change in marginal tax rates among wage-earners than among the self-employed (Saez, 2010) and also less evasion (measured by audits) for income with third-party information (Kleven, Knudsen, Kreiner, Pedersen, and Saez, 2011, Londoño-Vélez and Ávila Mahecha, 2019). We hypothesize that observing less bunching among taxpayers with high "revenue observability" is evidence in favor of misreporting as opposed to real production decisions.

Several transactions in which firms engage, such as selling to the government or exporting, generate third-party information: these sales are directly reported to the tax authority, allowing them to independently assess part of the revenue declared by taxpayers[33]. The availability of this information, nonetheless, is limited: overall less than 60% of corporations have any third-party information available, and even among larger firms declaring revenue above L5 million more than 15% are not covered at all. We use these reports to create a firm-level measure of revenue observability, defined as the share of self-declared revenue that is independently observed by the tax authority. Conditional on having any third-party information available, the median ratio between third-party informed and self-declared revenue is 25%[34].

In Figure 1.7, panel A, we plot the empirical density of revenue for the period 2015-2017 around the L10 million threshold separately for two groups: corporations for which some third-party information is available and those for which it is not. We observe bunching in both distributions, although there is slightly more mass below the threshold among those firms with no third-party information available. Since for a significant number of taxpayers the amount reported by third-parties is very small, we also repeat the exercise in panel B, now separating the sample in those above and below the (unconditional) median of revenue observability (15%). Here we observe a much sharper bunching behavior for firms with lower revenue observability, although excess mass is still clearly present for firms with higher degree of third-party coverage. We quantify these differences in panel A of Table 1.4. Whereas we estimate the excess mass at the notch for firms with above median revenue observability as four times the counterfactual density, for firms with below median observability we estimate seven times as much mass, and this difference is precisely estimated.

We provide additional evidence that bunching below the exemption threshold is driven by revenue misreporting by evaluating heterogeneity across industries. The availability of TPI varies systematically across industries given the nature of their economic activities. Since the main source of third-party information is withholding through the VAT credit systems,

[33] The tax authority has access to five sources of information on taxpayers' revenues. The most important one are sales to some large companies, which are mandated to report individual purchases as part of the credit system used for VAT. Credit and debit card operators also provide reports on sales as they are VAT withholding agents. All sales to the government and exports are also directly accessible to the tax authority. Finally, some other withholding activities by very large companies also generate information on sales of their suppliers. Data on third-party information is only consistently available since 2015 so we restrict our analysis to the period 2015-2017.

[34] One clear limitation for the availability of third-party information is the rule that determines which firms must provide detailed purchase reports on suppliers as part of the VAT credit system. We discuss that in the conclusion section.

revenues from firms in upstream sectors are more likely to be reported to the tax authority. On one extreme, the median corporation operating in construction or retail sees less than 15% of their total self-declared revenue being reported directly to the tax authority by third-parties. On the other, for the median firm in manufacturing or transportation sectors the revenue reported by third-parties amount to approximately 40% of their self-reported revenue. We then evaluate whether bunching at the sectoral level is systematically correlated with the degree of revenue observability in each industry, in the spirit of the analysis in Almunia and Lopez-Rodriguez (2018) but using firm-level data on revenue observability, allowing for a direct measure of the information set available for the tax authority on the revenue of taxpayers[35].

In panel B of Table 1.4 we present estimates of excess bunching at the notch, normalized by the predicted density at the threshold (column 2). First, we estimate large and precisely estimated excess bunching for firms in all industries. The amount of bunching, however, vary significantly across sectors: the excess mass ranges from 3.5 times the counterfactual density in manufacturing to approximately 8 times in agriculture and construction. To assess whether the amount of bunching is correlated with the availability of TPI, in Figure 1.8 we plot the estimated excess mass below the notch and the median revenue observability in each industry. We observe a strong negative correlation between the two measures: in industries where third-party reporting covers a larger share of firm's revenue much less bunching is observed immediately below the L10 million notch. Consider retail, where the majority of sales are to final customers and a low penetration of debit and credit cards means that only a small fraction of corporations' revenues are reported to the tax authority. The excess mass observed below the notch is seven times the predicted density, indicating a large amount of response to the incentives provided by the minimum tax. Manufacturing firms, on the other extreme, mostly supply to other firms and see a much larger share of their total sales directly informed to the tax authority. Here the excess mass at the notch is only half that observed among retail firms. While other factors might contribute to the observed negative correlation, we interpret this as further evidence that misreporting revenues plays a role in explaining the observed bunching below the exemption threshold.

1.4.4 Estimating evasion under profit taxation

We now turn to firms with gross revenue significantly above L10 million and therefore inframarginal to the bunching behavior below the notch. As documented above, the introduction of the minimum tax led to an increase in the reported profit margins and bunching around the 6% threshold, which separates the profit and revenue taxation regimes.

Let B be the excess mass of taxpayers locating around the threshold. These bunchers are coming from a continuous segment $[\Pi^T - \Delta\Pi, \Pi^T]$ below the kink: these are taxpayers that otherwise would have reported lower profit margins, but under revenue taxation increase their reported profit. The area where these bunchers come from is not empty, however, since the entire distribution shifts to the right as taxpayers declare higher profit margins.

Following a very similar approach as the one used above, our goal is again to estimate

[35] Almunia and Lopez-Rodriguez (2018) rely on input-output tables to compute the share of sales from each sector to final consumers.

the counterfactual distribution and use it to obtain an estimate of excess bunching at the kink. We estimate a counterfactual distribution of profits using a polynomial regression akin to Equation 1.8 and obtain estimates of the excess mass of taxpayers located around the kink[36].

In Figure 1.9 we present the empirical and estimated counterfactual profit margin densities for each year in the period 2014-2017. Between 90 and 210 firms are estimated to bunch around the 6% profit kink each year, an excess mass equivalent to 3-6 times the average density in the interval. In Figure 1.10 we present results for the pooled sample, where we estimate a similar excess mass equivalent to 5.4 times the average counterfactual density around the kink. We present the same results in the first two columns of Table 1.5.

Starting from the estimated excess mass around the kink, we can recover the change in reported profit margin by the marginal buncher noting that the bunching mass B around the threshold can be expressed as:

$$B = \int_{\Pi^T - \Delta\Pi}^{\Pi^T} f_0(\Pi) d\Pi \approx \Delta\Pi f_0\Big(\Pi^T\Big) \implies \Delta\Pi \approx \frac{B}{f_0\Big(\Pi^T\Big)} \tag{1.12}$$

where $f_0(.)$ is the counterfactual profit margin density and the approximation assumes the density is constant on the bunching segment. Empirically, we estimate $f_0(\Pi^T)$ as the average predicted density in the bunching region, and use the estimated excess mass at the kink to obtain $\Delta\hat{\Pi} \approx \frac{\hat{B}}{\hat{f}_0(\Pi^T)}$.

We present results for the estimated change in profit margins in column (3) of Table 1.5. With the exception of 2014, when we observe less bunching, estimates for 2015-2017 and for the pooled sample are very similar: the marginal buncher increased declared profit margin between 0.9 - 1.1 percentage points, a narrow range of precisely estimated responses. To put it differently, the marginal buncher would have declared a profit margin of approximately 5% under profit taxation, when incentives to misreport are strong and production decisions are undistorted.

In order to interpret the magnitude of these changes in reported profit and separate the total effect between cost evasion and production decisions, we use the decomposition of reported profit margin response developed by Best et al. (2015). Totally differentiating the reported profit margin and considering the incentives of a taxpayer around the kink yields:

$$\Delta\hat{\Pi} = \frac{\tau_y^2}{\tau_\pi}\epsilon_{y,(1-\tau)} - \frac{d(\hat{c} - c(y))}{y} \tag{1.13}$$

The main insight provided by the decomposition is that, since the tax rate on revenues is often very small (0.015 in the case of Honduras), even large revenue elasticities will only generate second-order effects on the change in reported profit margins. If we observe large increases in reported profit margin from the marginal buncher, therefore, changes in reported cost due to evasion incentives must be playing a large role. We illustrate that point in column (4) of Table 1.5, where we consider the implied revenue elasticity in a model where there is

[36]We compute the number of taxpayers in bins 0.2 p.p. wide. Following the literature, we determine visually the lower and upper bounds of the bunching region.

no cost evasion[37]. For all years and for the pooled sample, the implied elasticities under no cost evasion are implausibly high: with the exception of 2014 when the estimate is 6.7, the remaining elasticities of 10-12 are four times larger than our largest estimate in Table 1.3 and an order of magnitude higher than our preferred estimates, suggesting that cost evasion must be playing a significant role in explaining the observed response.

We present our estimates of cost misreporting in column (6). We use the upper bound elasticity $\epsilon_y = 0.99$ obtained for the pooled sample, so evasion estimates are a lower bound of the true evasion, and express evasion as a share of reported profits[38]. With the exception of 2014, where bunching is smaller, in the period 2015-2017 and using the pooled data we estimate that cost misreporting is in the range of 13-17% of reported profits[39].

These estimates are very similar to evasion documented by Best et al. (2015) for most corporations in Pakistan, which also fall in the range of $[0.13, 0.17]$, and by Alejos (2018) for corporations in Guatemala that fail to claim an exemption to the local minimum tax. Our results reinforce these previous findings that evasion through cost misreporting in lower income countries is significant even for large corporations, making the use of taxation of broader bases a potential tool to increase tax revenues.

1.4.5 The composition of cost adjustments

In the previous section we document that corporations evade a substantial amount of taxes by over reporting costs under a profit regime, and immediately change their reporting behavior when evasion incentives disappear under the minimum tax. One relevant policy question arising from these evasion responses, that might inform tax authorities' efforts, is whether firms adjust all cost categories similarly between these regimes, or if some cost items seem particularly prone to evasion.

We first present non-parametric evidence, in Figure 1.11, that deduction levels change discontinuously at the L10 million revenue threshold, consistent with the fact that, under the minimum tax, firms above the threshold increase their reported profits. Reassuringly, we observe no discontinuity in claimed costs in the period 2011-2013, before the minimum tax was in place. In order to assess whether specific cost categories are more responsive to the change in incentives, we use detailed cost items claimed in corporate income tax filings to construct five broad cost categories: Labor, Goods and Materials, Operations, Financial and Losses & others[40]. In Figure 1.12, Panel A, we present costs as share of gross revenue for each bin of declared revenue. The figure suggests that costs related to the purchase of goods and materials are the only ones that significantly change at the L10 million threshold. While for firms declaring revenue below L10 million the participation of goods and materials steadily increases, the average share of those costs falls discontinuously by over 5 p.p. at the

[37] That is, we set $d(\hat{c} - c(y)) = 0$ and compute the implied revenue elasticity $\epsilon_y = \frac{\tau_\pi}{\tau_y^2}\Delta\hat{\Pi}$.

[38] We can rewrite Equation 1.13 as $\frac{d(\hat{c}-c(y))}{\pi} = \tau_y \epsilon_y - \frac{\tau_\pi}{\tau_y}\Delta\hat{\Pi}$, where $\pi = y - \hat{c}$ and using the fact that around the 6% profit margin kink $\frac{(y-\hat{c})}{y} = \frac{\tau_y}{\tau_\pi}$.

[39] We present estimates using alternative polynomial orders in Table A1. Results are practically unchanged, with evasions estimates falling between 16-19% for the pooled sample.

[40] The detailed breakdown of cost categories only exists for firms declaring using the electronic SAR-357 form introduced in 2015. In all exercises using detailed cost data, we restrict our sample to the period 2015-2018 and to taxpayers filing electronically (70 - 80% of all corporations).

threshold and remains at a lower level for firms declaring up to L15 million in revenue. We do not observe a similar discontinuous fall in claimed deductions for other categories that generate more paper trail, such as financial or labor costs. In Panel B of the same figure we focus on the goods and materials category, showing that the discontinuous change observed at the notch is not observed in 2018, when the exemption threshold increase to L300 million.

We present a more formal test of whether these discontinuities can be attributed to the minimum tax in Table 1.6. Since we previously presented strong evidence that taxpayers strategically locate below the revenue threshold in order to avoid the minimum tax, we cannot simply estimate a regression discontinuity at the notch. Instead we estimate a linear "donut-hole" discontinuity regression, evaluating whether the level of costs change at the threshold but extrapolating from revenue levels not affected by bunching behavior[41].

In Column (1) we present results from a specification using median deductions by bin as dependent variable. We estimate that the amount of claimed deductions fall by approximately L260,000 at the threshold, consistent with the non-parametric evidence presented. Since the median deduction at the threshold is L9.8 million, the estimated effect implies that the median firm above the threshold decrease deduction claims by 2.7% and doubles the reported profit margin. In Columns (2) through (5) we repeat the same exercise but use the ratio of deductions to revenue as dependent variable. The only estimate statistically different from zero and meaningful in magnitude is goods and material costs: they fall by almost 5 p.p. from an average of 37% below the notch. Mosberger (2016), using a different empirical strategy, also documents a significant change in goods and materials costs by firms facing a minimum tax in Hungary, suggesting this seems to be a deduction category particularly over reported by firms trying to reduce profit tax liabilities and therefore a potential focus for tax authorities.

1.5 Robustness and additional exercises

In this section we provide additional evidence that the empirical patterns discussed previously are indeed the result of corporate responses to the minimum tax.

Our main sample consists of an unbalanced panel of corporations. Since the number of firms filing income tax increases significantly during the period, one might worry that results are purely driven by sample composition. We show that this is not the case by restricting the sample to a subset of firms observed in every year between 2013 and 2018[42]. In panel A of Figure A1 we present empirical revenue densities and in panel B we present profit margin densities for each year. The same pattern observed in the full sample is present in the balanced panel: an excess of firms reporting revenue slightly below L10 million and larger firms bunching around 6% profit margin in 2014-2017, but not before or after the exemption threshold was substantially increased.

[41] Unlike Bachas and Soto (2021), who perform a similar exercise, we cannot use these regressions to recover an estimate of cost elasticity. The reason is that, unlike in their setting where all firms in the bracket above the notch face an incentive to change costs due to a higher average tax rate, in our setting only low-profit firms will have an incentive to change costs, while firms with profit margins above 6% do not change their behavior. The observed change in average costs at the threshold will conflate both behaviors.

[42] There are 12,172 corporations that filed income tax declarations in every year between 2013 and 2018.

We also previously documented that the bunching behavior below the L10 million exemption threshold was observed across industries. We present that evidence graphically in panel A of Figure A2. In panel B we show that the right-shift in the profit margin distribution, accompanied by an excess mass around the 6% kink, is also observed in almost all industries. In Table A3 we present estimates of excess bunching at the 6% profit margin kink and cost evasion for corporations in different industries. With the exception of the small number of firms with undeclared economic sector, we estimate large and significant cost evasion for all sectors, ranging from 10% of reported profit in retail to over 25% in manufacturing, automotive and transportation. These results show that the evasion behavior observed in the full sample is not driven by a few industries, but widespread across corporations in the economy.

We perform two additional exercises that strengthen our case that the shift observed in declared profit margins by firms above the revenue exemption threshold was a response to the specific features of the minimum tax. First, as mentioned in section 1.2, a small number of industries were subject to a reduced minimum tax rate of 0.75% instead of 1.5%. Corporations in those industries therefore face a kink in the tax schedule not at 6% rate of profit margin but at $\frac{0.0075}{0.25} = 3\%$, and according to our model we should observe excess mass around that threshold. In Figure A3 we show that is precisely what happens: between 2014-2017, the distribution of profit margins for firms in these industries is shifted to the left when compared to corporations facing the 1.5% minimum tax and the peak of the distribution is exactly around 3%.

Second, we also investigate whether the increase in declared profit margins is induced by "lazy cost reporting" (Best et al., 2015). If there are fixed-costs in filing different cost line items, taxpayers might respond to revenue taxation by reducing the number of items filed and therefore generating an increase in profit margins, even if they were reporting truthfully under a profit taxation regime. We investigate whether there are significant changes in the share of cost line items reported in Figure A4. Panel A presents the share across the 6% profit margin kink, for firms reporting revenue above L13 million, while panel B reports shares across the L10 million notch. If the observed changes in deductions/profit were being driven by filing costs, we should expect an increase in the share of items reported when firms report profit margins above 6% (Panel A) and a decrease for firms reporting above the exemption threshold (Panel B). Instead, shares are mostly smooth across the thresholds, and no different from the behavior of firms in 2018, when the exemption threshold was much higher and fewer firms are subject to the minimum tax. These results suggest it is unlikely that costly filing drive our results, at least on the extensive margin, and point to the importance of evasion under profit taxation.

1.6 Assessing the impact of counterfactual policies

In order to make progress in quantifying the impacts of the minimum tax and alternative tax policies, we make stronger parametric assumptions about the profit function of firms and calibrate a model. We consider firms with isoelastic production costs and cost misreporting loss functions so we can rewrite Equation 1.1 as follows:

$$\hat{\Pi}(y, \hat{c}) = y - \alpha_i - \frac{\theta_i}{1 + 1/e}\Big(\frac{y}{\theta_i}\Big)^{(1+1/e)} - \tau(y - \mu\hat{c}) - \frac{B_i}{1 + 1/\gamma}\Big(\hat{c} - c(y)\Big)^{(1+1/\gamma)} \tag{1.14}$$

Taxpayer are heterogeneous in three dimensions, characterized by the vector $(\theta_i, \alpha_i, B_i)$ that define productivity, production fixed cost and evasion ability, respectively. Heterogeneity in productivity allows firms to have different optimal production levels, while varying fixed costs generates a distribution of profit margins. We consider the maximization problem of firms under a simple profit taxation regime and calibrate the model using the parameters previously estimated and data from 2013, before the introduction of the minimum tax. We set $e = 0.99$, the upper bound revenue elasticity from our pooled sample, and use the estimates from Best et al. (2015) for evasion cost elasticity $\gamma = 1.5$. We then calibrate the remaining parameters to match the distributions of reported revenue and reported costs, considering that firms evade 17% of profits through cost over-reporting[43].

We perform three main exercises. First, we simulate the actual minimum tax system implemented in Honduras in 2014, with an exemption threshold for firms reporting gross revenue below L10 million and minimum effective tax of 1.5% for larger firms. Second, we consider an alternative to the minimum tax regime where the tax authority increases the average tax rate that large firms pay on profits. Finally, we simulate an alternative tax system in which all firms are taxed not on pure profits but on a broader base, only allowing partial cost deduction.

We present results for our first exercise in Table 1.7. First, consider the actual minimum tax implemented, in which firms reporting gross revenue below L10 million are exempt and those above face a minimum tax liability of 1.5% of gross revenue (first line). We estimate that over 60% of corporations declaring revenue above the exemption threshold are liable for the minimum tax and that total government revenues increase by over 30% when compared to a flat profit tax rate of 25%[44]. This is attained by a 120% increase in the aggregate tax liability of firms paying the minimum tax and a decrease of 10% in aggregate profit for all firms in the economy. The fall in aggregate profits shows that, under the parameters of the actual policy implemented, the potential gains for firms when moving from profit to revenue taxation (decrease in losses from misreporting cost) is dwarfed by the losses from higher tax liability and production distortions.

Our calibrated model also allows us to quantify the strong incentives introduced by the exemption notch: the total tax liability of bunching firms is less than 25% what they would have payed had they stayed above the threshold and paid the minimum tax. Despite that strong reaction at the margin, the increase in taxes paid by infra marginal firms dwarfs this loss: reduction in taxes from bunching firms is only 1% of total revenue from the minimum tax. While in our model bunching below the exemption threshold is exclusively driven by real production decisions, we provided evidence that at least part of this behavior seems to be explained by revenue misreporting. That finding highlights that, despite generating relatively small aggregate losses, notches can generate large horizontal inequities: firms otherwise

[43] Details are presented in Appendix A.3.

[44] In these simulations we exclude taxpayers that were liable for Net Asset tax in 2013, since we do not model firms' asset accumulation and reporting decisions.

similar might be liable for vastly different tax burdens simply due to willingness to misreport revenue.

We also assess the impact of alternative minimum tax specifications, in which we vary both the exemption threshold and the minimum tax rate. We highlight two features of our simulations. First, holding constant the minimum tax rate on gross revenues, increasing the exemption threshold only slowly decreases total revenue gains due to the long right tail of firm size. Doubling the exemption threshold from L10 to L20 million, for example, still leads to 28% revenue gain, and a L50 million exemption threshold still increases tax revenue by 23%. For the same reasons, aggregate profits still fall substantially when considering exemption thresholds at L20 million (-9%) and L50 million (-7.6%). Second, small changes in the minimum tax rate generate large impacts in aggregate tax revenue and firms' profit, given the very broad base (gross revenue). Using the same L10 million exemption threshold and considering a minimum tax rate of 0.5% (implying a minimum profit margin of 2% under profit taxation), for example, generates a tax revenue increase of less than 4% and aggregate profit loss of 0.5%. When comparing these magnitudes with the actual policy implemented, the decrease in tax revenue gain is driven by two forces. First, the minimum "allowable" profit margin is now lower: corporations with a 5% profit margin, for example, are allowed to pay an effective tax rate of 25%*5% = 1.25% when the minimum tax is 0.5%, while they would be liable for the 1.5% minimum tax under the previous regime. Second, firms with very low profit margins now only pay 0.5% in effective tax rate instead of 1.5%. This logic extends to increases in the minimum tax rate: increasing it from 1.5% to 2% leads to a 50% increase in tax revenue but at the cost of a 17% fall in aggregate firms' profits.

Our second exercise considers a progressive tax schedule in which firms declaring gross revenue above L10 million face an increase in *average* tax rates, without a change in the tax base (reported profits). We consider that the average tax rate is still 25% for firms below the exemption threshold, so firms also face a discontinuous change in tax liability when reporting revenue above L10 million and will have a strong incentive to bunch below the threshold. Unlike in our setting where firms with low profits benefit the most from bunching, here firms with high profit margins face the strongest incentives to locate below the notch, since they have the most to lose from higher tax rates. We present results for scenarios that consider an average profit tax rate between 30% and 50% in Table 1.8. Increasing the average tax rate by 5 p.p. to 30%, for example, would increase tax revenues by 12% and reduce aggregate corporate profits by 7%. In order to generate the same amount of tax revenue gains as the minimum tax, average taxes have to increase by 15 p.p. to 40%. While production efficiency is preserved under high tax rate profit taxation, evasion costs are exacerbated in this scenario and lead to large losses in aggregate profits, which fall by 20%.

These simulations suggest that the owners of corporations had strong reasons to oppose the introduction of a minimum tax scheme, at least in the format it was implemented. Following Best et al. (2015) and Bachas and Soto (2021), we consider alternative scenarios that could be more attractive to corporate shareholders. Instead of pure profit taxation and an additional minimum tax on gross revenue, we consider systems that allow only partial deduction for all firms, under the constraint that aggregate firm profit is not reduced when compared to the baseline of pure profit taxation under a 25% rate. Here we explicitly explore the production vs. revenue efficiency trade-off at the heart of the minimum tax discussion of Best et al. (2015): it is only possible to increase both aggregate profits and government revenue because corporations incur in non-deductible misreporting costs under profit taxation. By introducing some production distortion in the form of partial cost deductibility, we can reduce losses from cost misreporting.

Figure 1.13 presents the main results of our simulation. For each level of deduction rate μ, we compute the revenue maximizing tax rate (under the constraint of constant profit) and how aggregate revenues change. For a wide range of deduction levels, we show that aggregate revenues could be increased by 8-10%. Among all possible pairs of (τ, μ), we estimate that allowing 45% of costs to be deducted and taxing the remaining net revenues under a 2.3% rate would increase government corporate tax collection by 9.4%, without reducing aggregate profits. But it is noteworthy that there is little gain to be obtained once we consider deduction rates below 85%: we obtain large revenue increases by introducing small distortions in production starting from firms' optimal production level, but after these initial distortions the government can do little more to raise revenue without decreasing aggregate profits. In particular, it's noteworthy that under a pure revenue tax system ($\mu = 0$), we estimate the optimal tax rate to be 1.3% - not far from the current 1.5% applied under the minimum tax.

1.7 Conclusion

Minimum taxes are seen as effective tools for tax authorities to curb tax evasion in low-income countries and are at the heart of recent debates on global tax cooperation. In this paper we provide new evidence on corporate reaction to minimum taxes in Honduras. The specific design of the minimum tax provision in the country, including an exemption threshold based on declared gross revenues, allow us not only to quantify evasion under profit taxation but also to provide bounds on the elasticity of declared revenue - a key behavioral parameter when considering taxation under less than full deductibility of production costs.

We document meaningful evasion under profit taxation. Corporations liable for a minimum tax declare much larger profit margins when the incentives to over report costs disappears. We quantify that response and estimate that inflated costs allowed these firms to reduce tax liabilities by up to 17%. Curbing evasion through excessive reporting of deductions is costly to tax authorities (Carrillo, Pomeranz, and Singhal, 2017) since it requires time-intensive verification of receipts. The case of Honduras illustrates these limits: in 2018 approximately 150,000 taxpayers filed an income tax declaration, but there were only 45 full and 138 partial audits[45]. We provide evidence that taxpayers exploit these limitations and use hard-to-verify cost categories to reduce tax liability. The introduction of the minimum tax does not lead to changes in the amount of labor or financial deductions claimed - categories that often generate paper trail and are therefore easier to verify - but leads to substantial changes in the amount of costs related to purchase of goods and materials. Improving oversight of claimed deductions of goods and materials seems to be a natural focal point for the efforts of tax authorities.

Using the response of taxpayers to the notch created by the exemption threshold, we

[45]Full audits are in-site visits that review all tax liabilities of a taxpayer, while partial audits are requests for taxpayers to provide more information, such as receipts supporting claimed deductions, focused on specific tax liabilities.

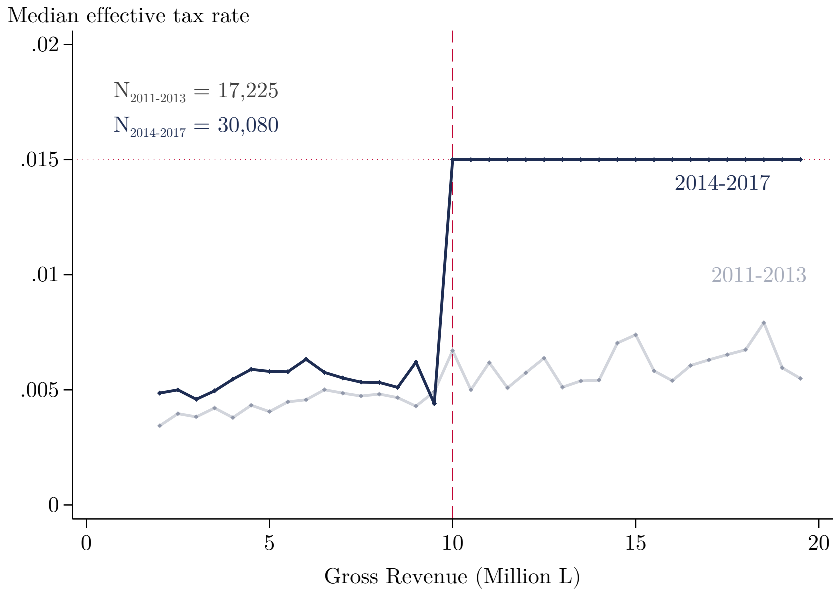
bound the elasticity of reported revenue with respect to the net-of-tax rate at [0.35, 1]. These estimates are substantially higher than previous results for corporate taxpayers in similar settings and illustrate the limits faced by authorities in imposing high tax rates on broader bases. Whereas the elasticity of reported revenue summarizes responses both through real production and reporting decisions, we provide evidence that at least part of the observed response is due to revenue under-reporting. Firms with high revenue observability are less likely to strategically locate below the exemption threshold. The same result holds across industries: firms in upstream industries, for which third-party information is more readily available due to VAT withholding, also present less excess bunching and therefore a lower implied revenue elasticity.

These results highlight the fact that behavioral responses of taxpayers are endogenous to the enforcement environment (Fack and Landais, 2016, Slemrod and Kopczuk, 2002). Building state capacity and properly designing rules to enforce tax compliance, therefore, might substantially change the trade-offs between available instruments. In the case of minimum taxes, improving the ability to assert the veracity of claimed deductions should decrease evasion through cost misreporting, making profit taxation more attractive. Improvements in independent verification of taxpayers' declared revenue, conversely, make broadening the tax base more attractive by reducing the elasticity of reported revenue.

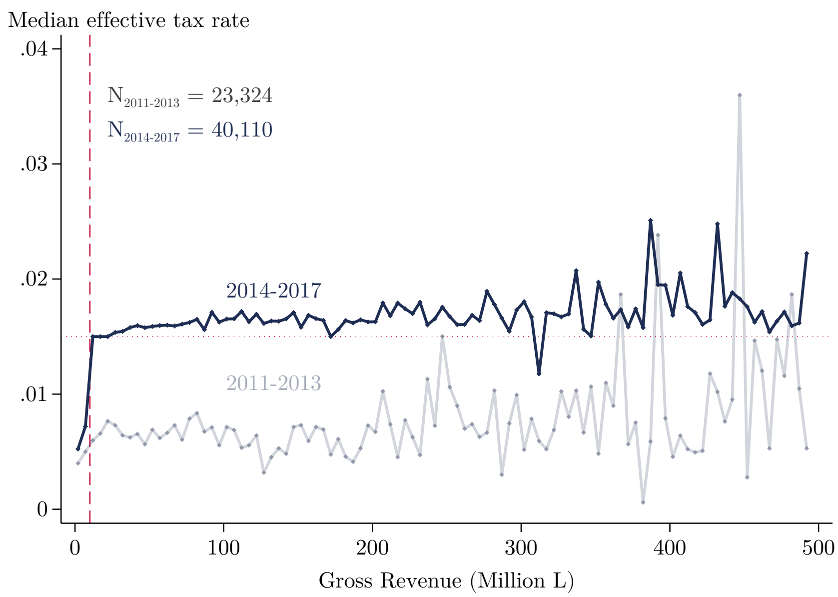
Honduras provides a sharp example of how administrative reporting rules can affect the enforcement environment. The main source of third-party information on the revenue of taxpayers are VAT withholdings. All firms can use VAT paid on their purchases as credit for their own liabilities, but only a subset of firms defined as "medium and large" have the obligation to provide individualized information on the identity of their suppliers, generating independent information on their revenues. While there are good reasons not to require all firms to provide such detailed information, the definition of "medium and large" firms is constantly updated based on revenue or profit, but has been fixed since 2011. In Figure A5 we plot the share of taxpayers in each group that are defined as "medium and large" and therefore required to provide detailed VAT information. Among the top 1% of corporations in terms of revenue, over 90% had to file detailed VAT information in 2011. That share steadily decreased since 2014, reaching less than 80% in 2018. A similar scenario is observed if we focus on the top 0.1% or top 10% of firms: an increasing number of very large firms are not required to file detailed VAT data since the requirements have not been updated since 2011. This example illustrates how seemingly innocuous administrative rules on reporting by taxpayers can have important impacts on tax compliance behavior.

Figures

Figure 1.1: Median effective tax rate across declared revenue distribution



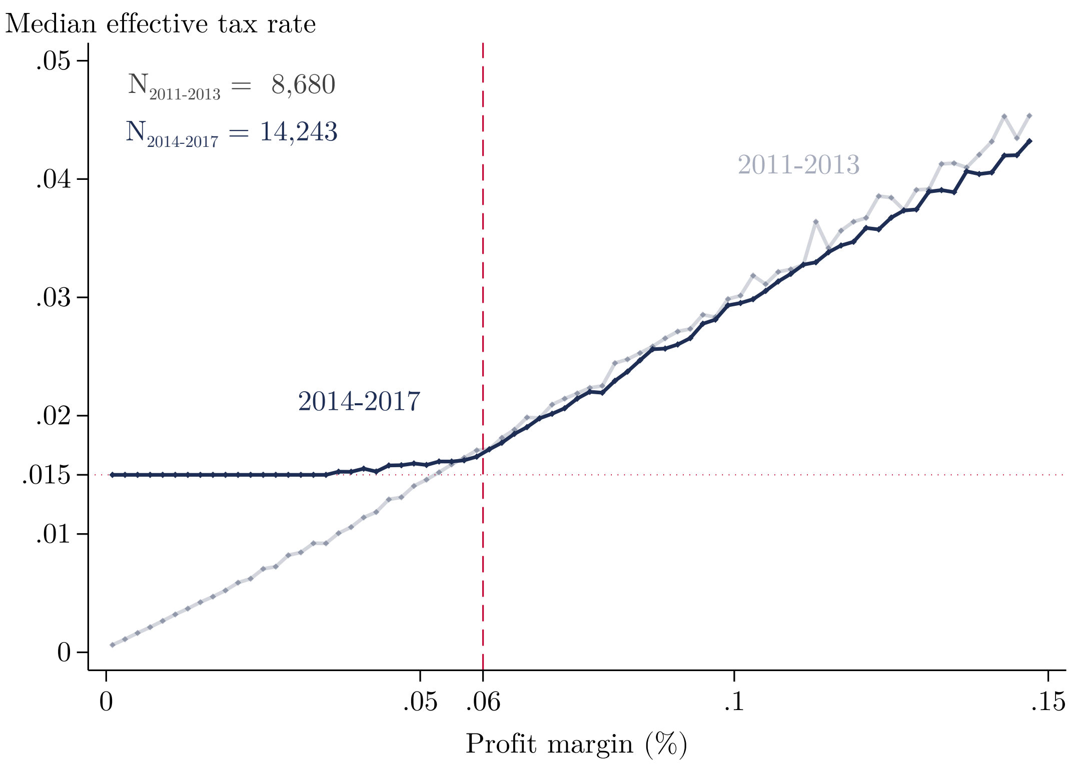
(a) Around L10 million exemption threshold



(b) Across gross revenue distribution

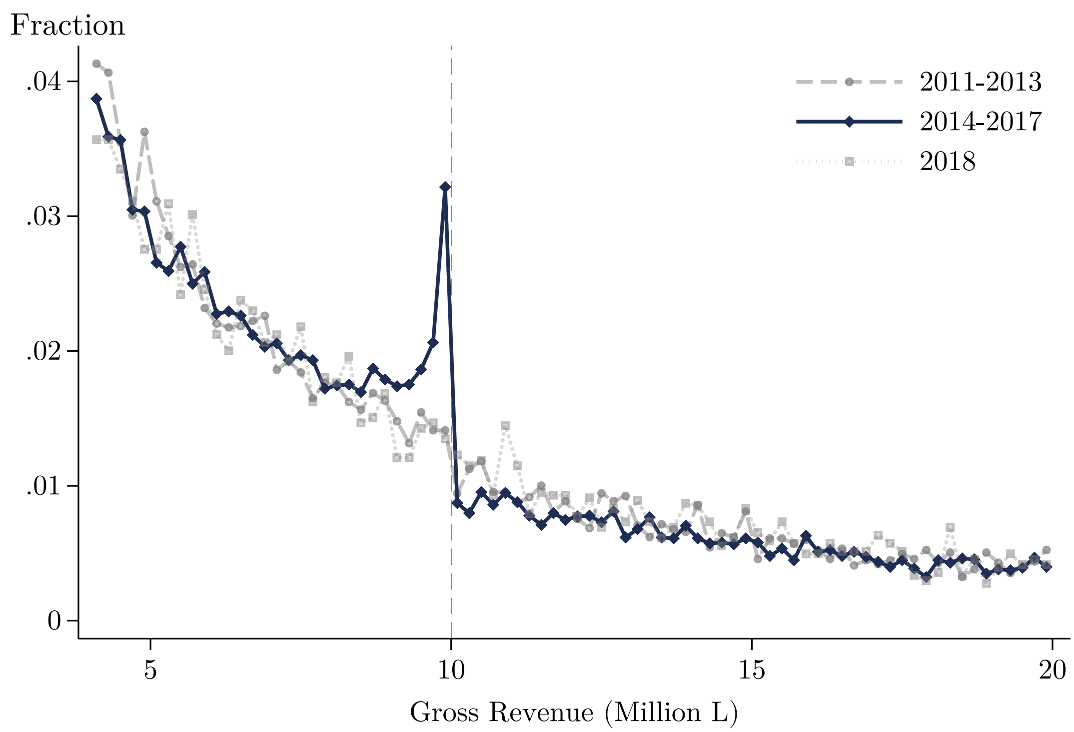
Note: This figure presents median effective tax rates, defined as the ratio between tax liability and gross revenue, for each bin of declared gross revenue. Panel A restricts the sample to taxpayers declaring gross revenue between L2-20 million, while panel B includes taxpayers with gross revenue between L2 - 500 million. Bins are L500,000 wide in Panel A and L5 million in Panel B.

Figure 1.2: Median effective tax rate across declared profit margin distribution



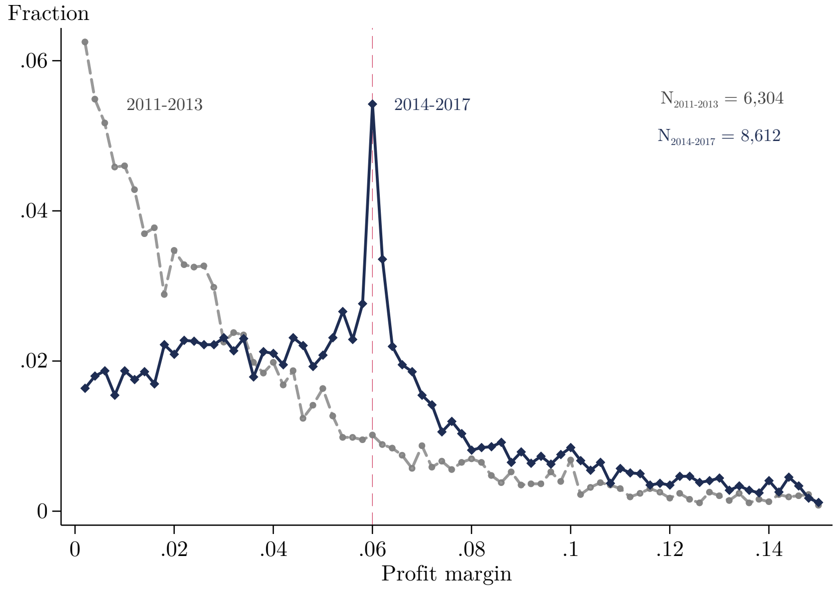
Note: This figure presents median effective tax rates, defined as the ratio between tax liability and gross revenue, for each bin of declared profit margin. The sample is restricted to firms declaring gross revenue above L13 million, and therefore inframarginal to bunching at the L10 million threshold. It documents that the introduction of the minimum tax creates a kink (change in slope of tax schedule) at the 6% threshold. Firms declaring profit margins below that level cannot decrease their tax liability by declaring lower margins, since their liability is at least 1.5% of gross revenues. Bins are 0.2 p.p. wide.

Figure 1.3: Empirical Density of Gross Revenue around L10 Million threshold

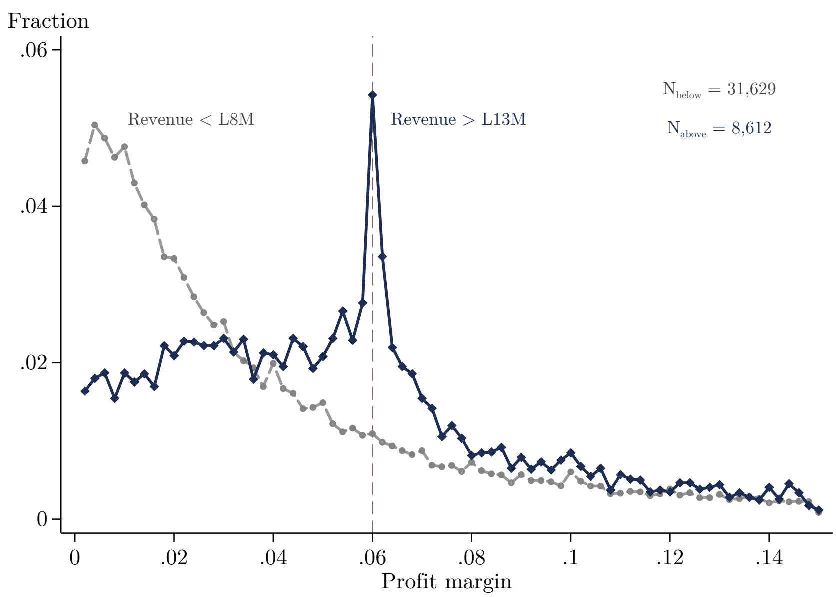


Note: This figure presents the empirical density of gross revenues from firms pooled for three periods: 2011-2013 (before the minimum tax introduction); 2014-2017 (when the exemption threshold was L10 million); and 2018 (after the threshold for eligibility increased to L300 million). It documents that, consistent with theoretical predictions, taxpayers respond to the notch created by the L10 million exemption threshold by bunching below the threshold. Bins are L200,000 wide. The sample is restricted to taxpayers declaring gross revenue between L4-20 million and excludes taxpayers exempt from the minimum tax.

Figure 1.4: Empirical density of profit margins



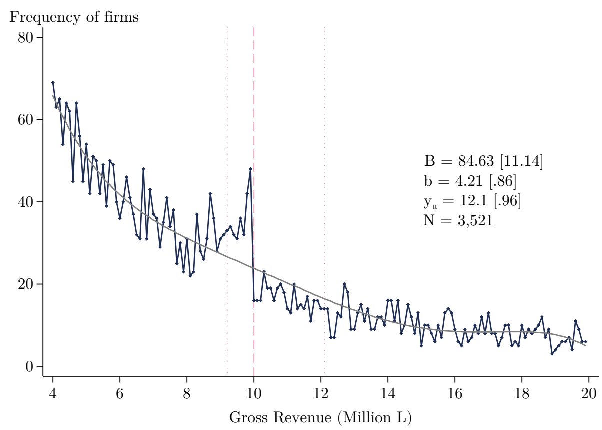
(a) Empirical density of profit margins above L13 million - Pre and Post Minimum Tax



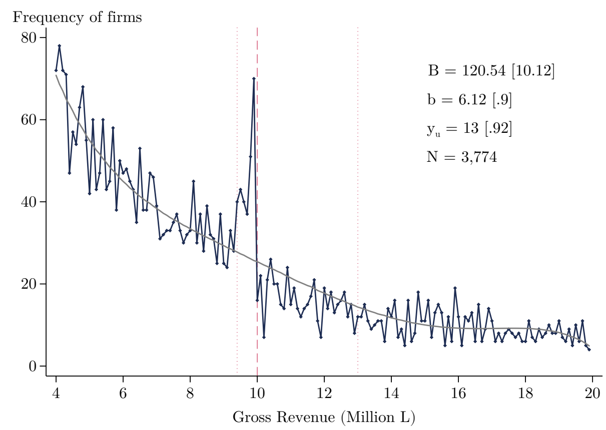
(b) Empirical density of profit margins in 2014-2017 - Below and above L10 million threshold

Note: These figures present the empirical density of positive reported profit margins. Panel A presents densities for firms with gross revenue above L13 million, before (2011-2013) and during (2014-2017) the existence of the minimum tax. Panel B present densities for the period of 2014-2017 of two groups of firms: those reporting gross revenue below L8 million (exempt from minimum tax) and those above L13 million (potentially liable for the minimum tax and infra-marginal to the bunching behavior at L10 million in revenue). Bins are 0.2 percentage points wide and the first bin starts at 0.1%, such that the 6% kink is the midpoint of a bin.

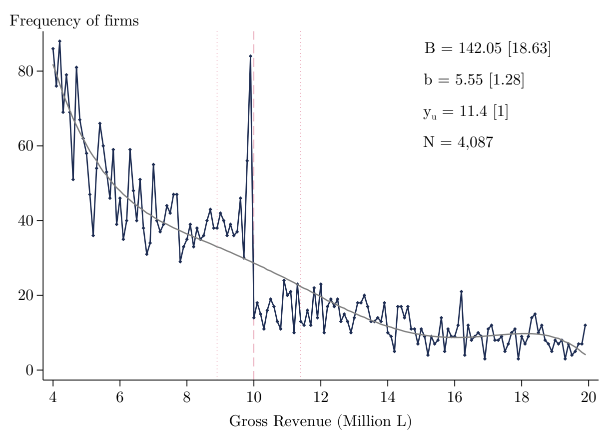
Figure 1.5: Empirical Density of Gross Revenue around L10 million threshold



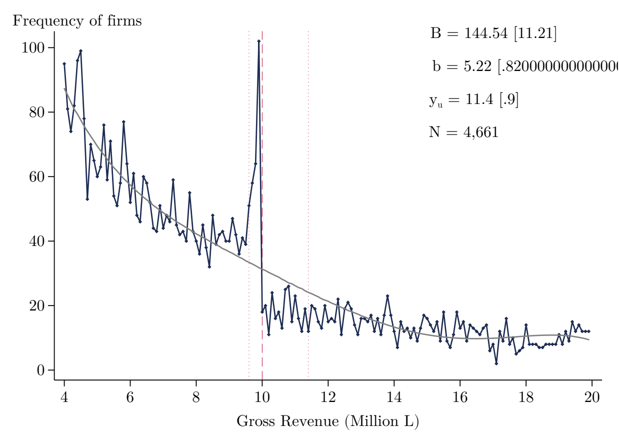
(a) 2014



(b) 2015



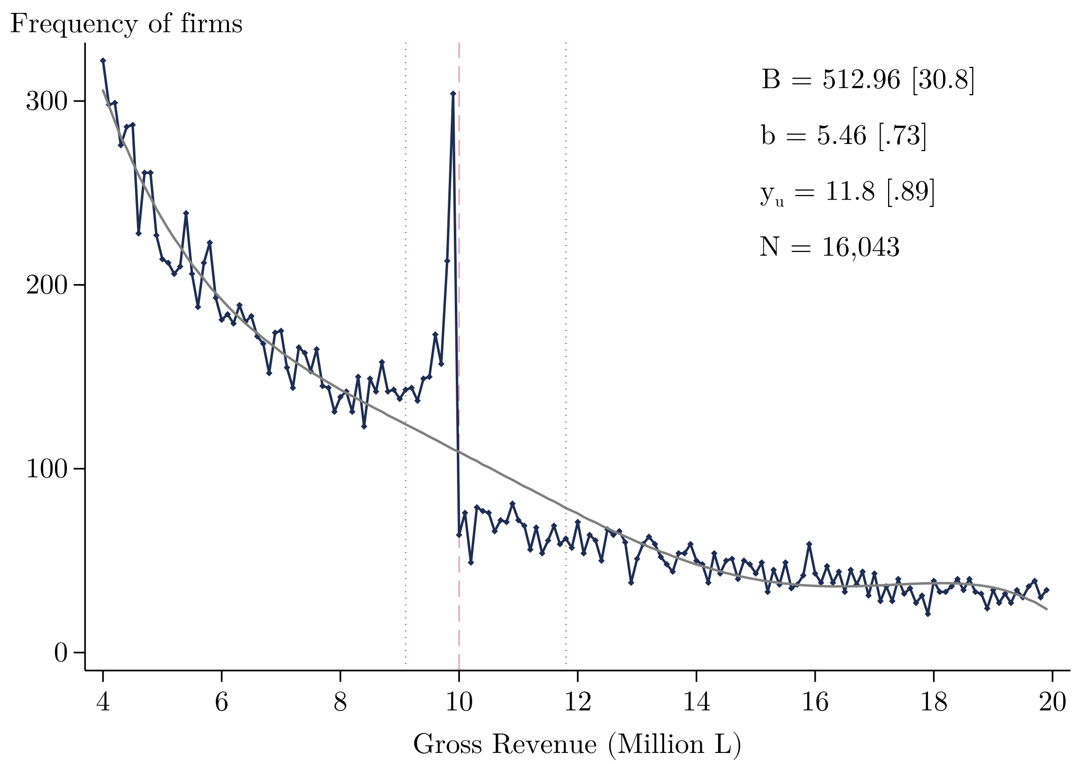
(c) 2016



(d) 2017

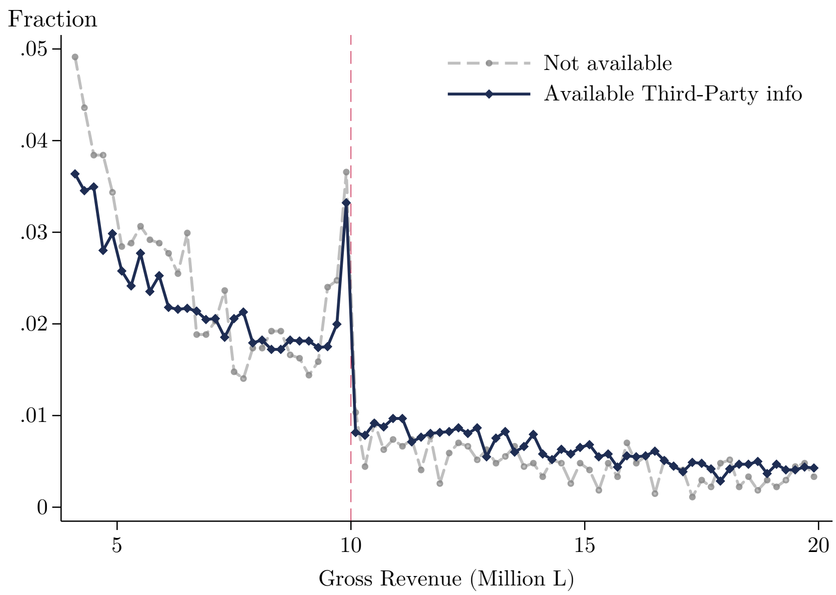
Note: These figures present empirical and counterfactual densities of declared gross revenue for each year in the period 2014-2017. The lower bound of the bunching region is chosen visually while the upper bound is obtained using the convergence method discussed in Section 4.3. The dashed line marks the L10 million notch while the dotted lines mark the lower and upper bounds of the bunching region. For each year we present the excess mass below the notch (B), the excess mass as a share of the predicted mass in the bunching region (b), the upper bound obtained from the convergence method (y_u) and the underlying number of taxpayers in each figure (N). Standard errors in brackets are obtained through bootstrapping. Bins are L100,000 wide.

Figure 1.6: Empirical Density of Gross Revenue around L10 million threshold - Pooled Years (2014-2017)

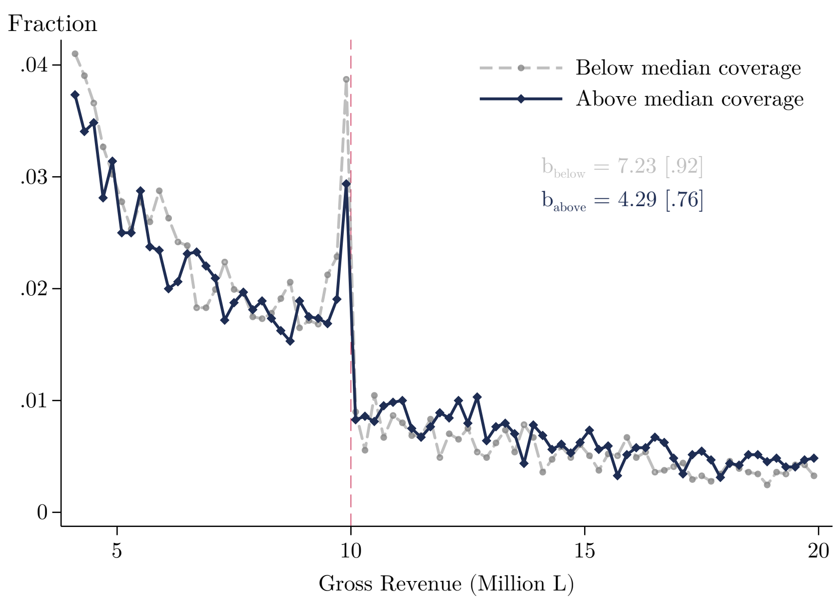


Note: This figure presents empirical and counterfactual densities of declared gross revenue for a pooled sample of firms in the period 2014-2017. The lower bound of the bunching region is chosen visually while the upper bound is obtained using the convergence method discussed in Section 4.3. The dashed line marks the L10 million notch while the dotted lines mark the lower and upper bounds of the bunching region. We present the excess mass below the notch (B), the excess mass as a share of the predicted mass in the bunching region (b), the upper bound obtained from the convergence method (y_u) and the underlying number of taxpayers in each figure (N). Standard errors in brackets are obtained through bootstrapping. Bins are L100,000 wide.

Figure 1.7: Empirical gross revenue density by third-party status - pooled 2015-2017



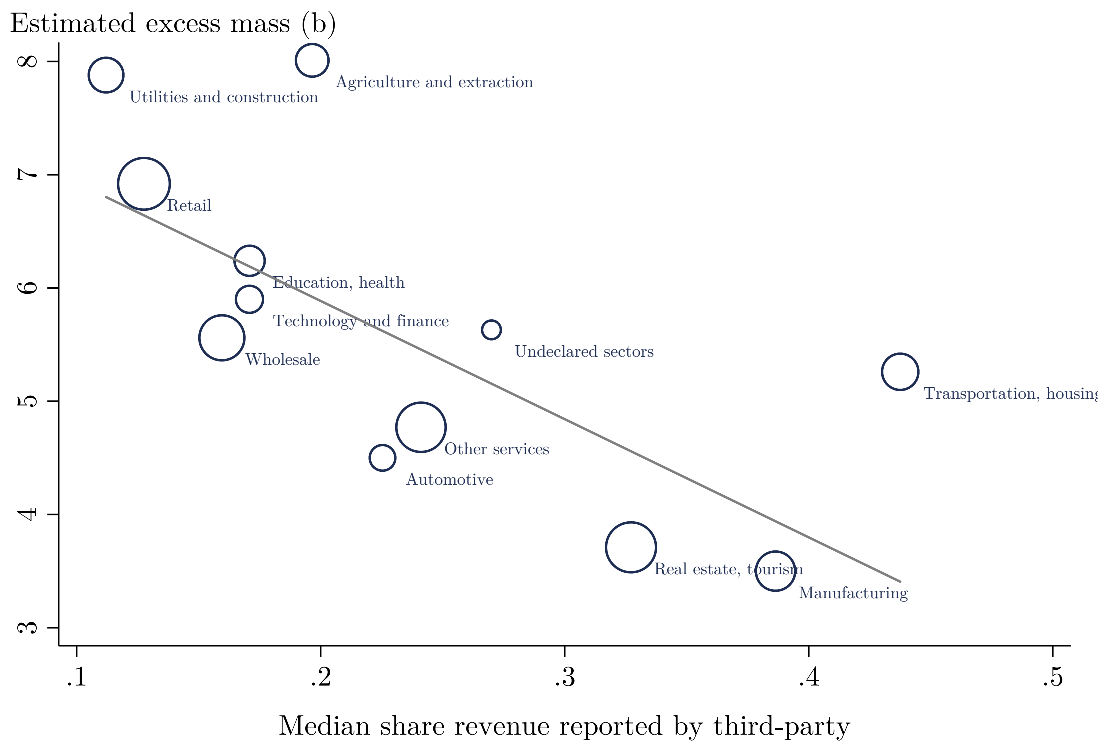
(a) Extensive margin



(b) Intensive margin

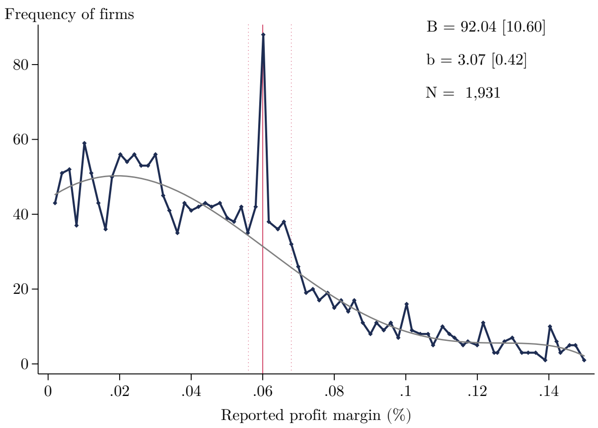
Note: These figure presents the empirical densities of declared gross revenue, pooled for the 2015-2017 period, exploring heterogeneity according to availability of third-party information on revenue. Panel A compares corporations for which no third-party information is available (gray line) with those for which some information is available (blue line). Panel B explores differences in the intensive margin of third-party information: it compares firms with below median (15%) share of declared revenue reported by third parties (gray line) with those above median (blue line). Bins are L200,000 wide.

Figure 1.8: Scatter plot of amount of bunching vs. revenue observability across industries

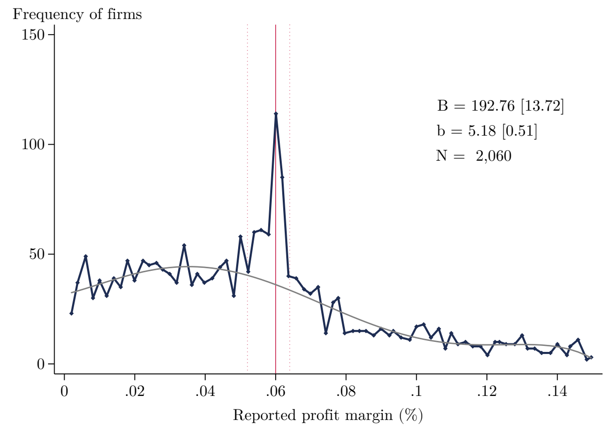


Note: This figure presents a scatter plot of estimated excess mass at the L10 million threshold and the median share of self-reported revenue also informed by third parties in each industry. Results show that in industries with higher third-party reporting we observe less bunching. Excess mass is defined as the excess number of firms bunching at the L10 million notch as a ratio of the predicted mass at the notch. The share of reported revenues is calculated in 2018, for firms declaring gross revenues in the interval L5-15 million. The size of markers is proportional to the reported sales in 2018 by industries.

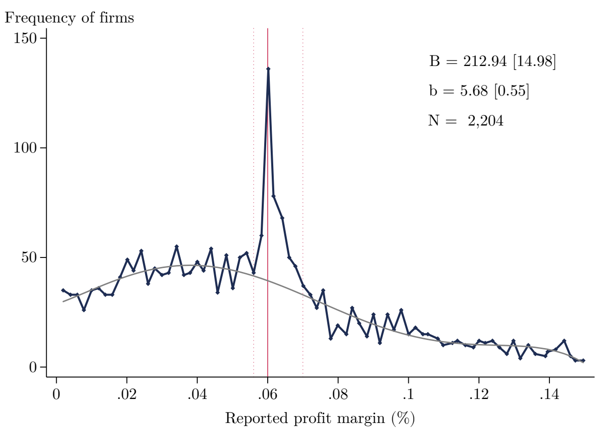
Figure 1.9: Empirical Density of profits around 6% threshold



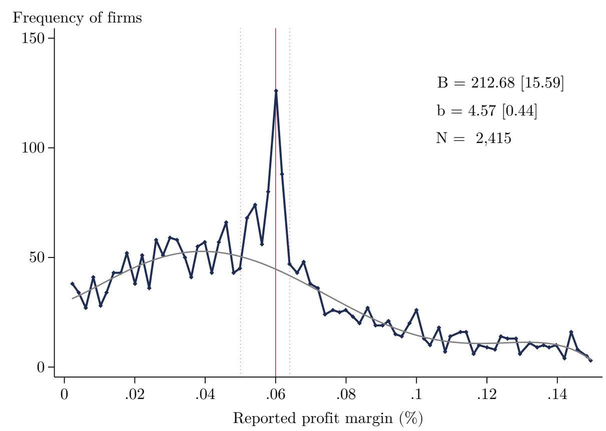
(a) 2014



(b) 2015



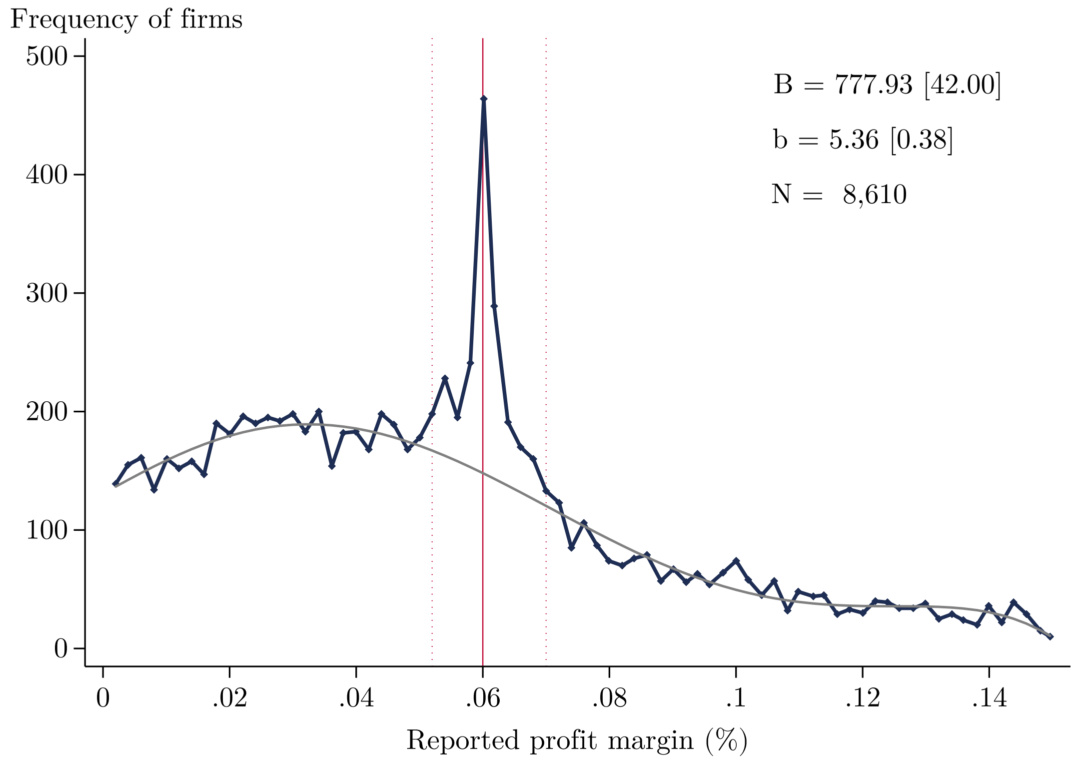
(c) 2016



(d) 2017

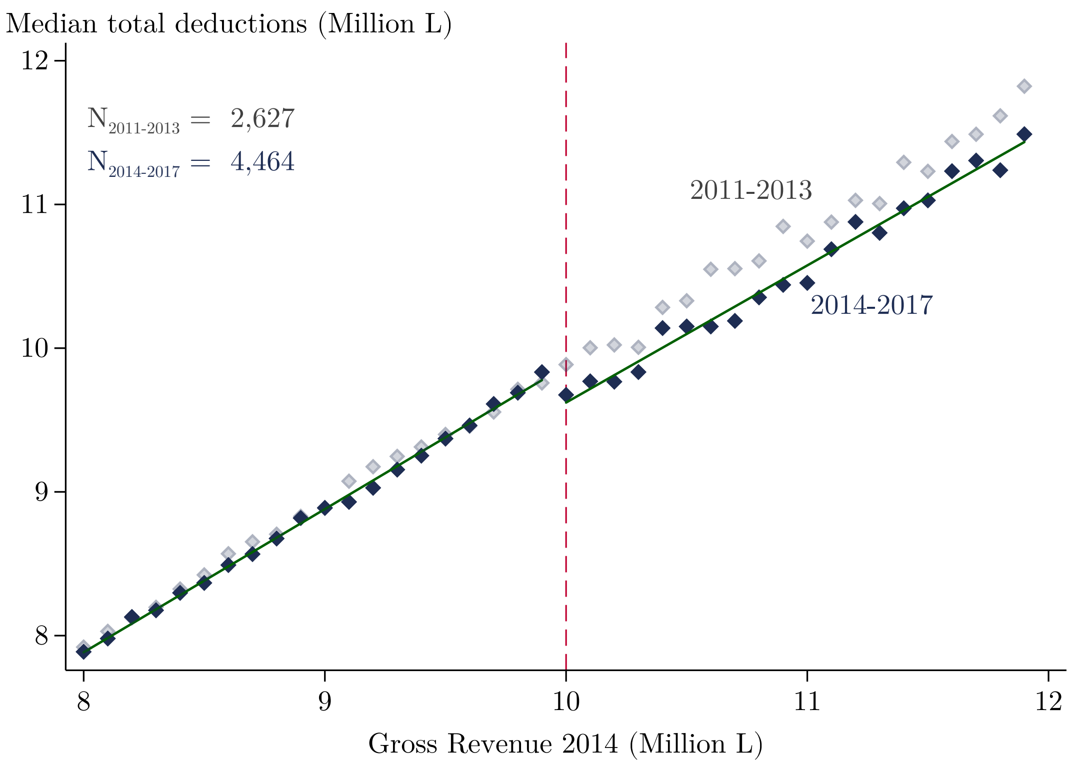
Note: These figures present the empirical and estimated counterfactual distributions of profit margins for each year in the period 2014-2017. The lower and upper bounds of the bunching region are determined visually. The solid red line marks the 6% kink while the dotted lines present the lower and upper bounds of the bunching region. For each year we present the excess mass around the kink (B), the excess mass as a share of predicted density around the kink (b) and the underlying number of taxpayers in each figure (N). Standard errors in brackets are obtained through bootstrapping. Bins are 0.2 percentage points wide and the first bin starts at 0.1%, such that the 6% kink is the midpoint of a bin.

Figure 1.10: Empirical Density around 6% profit margin threshold - Pooled Years (2014-2017)



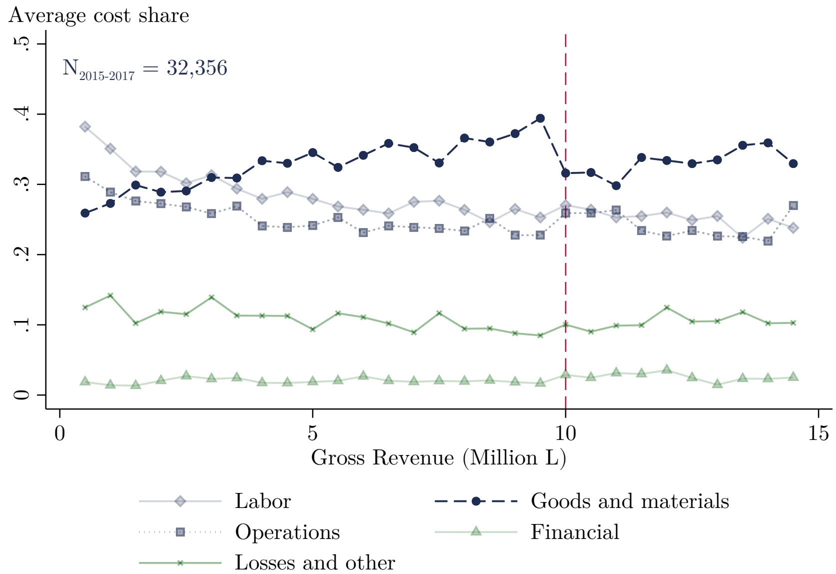
Note: These figures present the empirical and estimated counterfactual distributions of profit margins for a pooled sample of firms in the period period 2014-2017. The lower and upper bounds of the bunching region are determined visually. The solid red line marks the 6% kink while the dotted lines present the lower and upper bounds of the bunching region. We present the excess mass around the kink (B), the excess mass as a share of predicted density around the kink (b) and the underlying number of taxpayers in each figure (N). Standard errors in brackets are obtained through bootstrapping. Bins are 0.2 percentage points wide and the first bin starts at 0.1%, such that the 6% kink is the midpoint of a bin.

Figure 1.11: Median total deductions by gross revenue

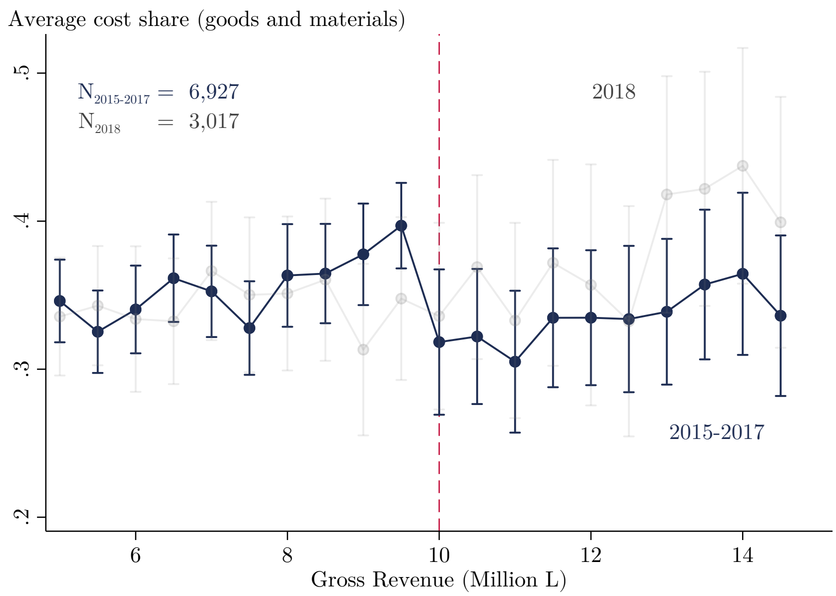


Note: This figure presents median reported total deductions by revenue bin for two groups: taxpayers in 2011-2013, before the introduction of the minimum tax, and 2014-2017, while the minimum tax was in place with a L10 million exemption threshold. The figure documents that claimed deductions fall discontinuously at the exemption threshold during the 2014-2017 period, consistent with the right-shift of the profit distribution observed for taxpayers subject to the minimum tax. No similar discontinuous change is observed in the period before the introduction of the minimum tax. Bins are L100,000 wide.

Figure 1.12: Cost line items as share of revenue



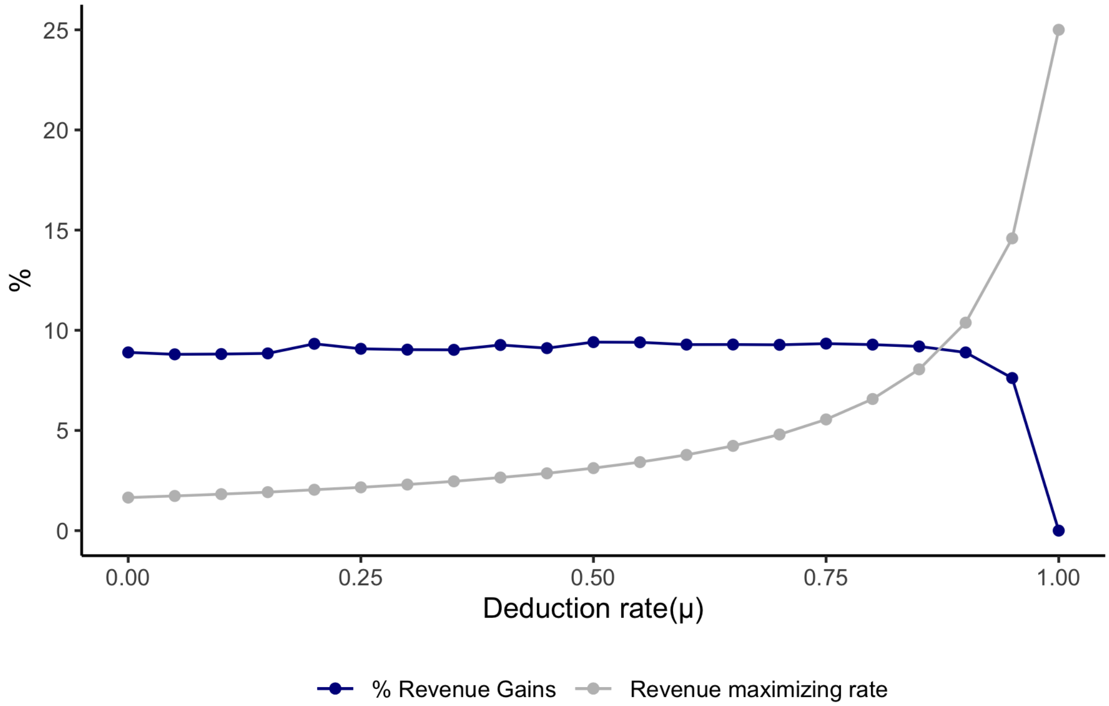
(a) All categories (2015-2017)



(b) Goods and materials (2015-2017 & 2018)

Note: These figures present cost line items as share of revenues in each bin. Panel A presents average shares in 2015-2017 for five cost categories: Labor, Goods and Materials, Operations, Financial, and Losses and other. Panel B focuses on Goods and Materials cost shares, separately for 2015-2017 and 2018. Bins are L500,000 wide in both panels. This sample only includes taxpayers using electronic declaration, for which we have detailed breakdown of cost items (approximately 80% of taxpayers per year) and excludes taxpayers with profit margins above the 99th and below 1st percentile of profit margin distribution.

Figure 1.13: Revenue maximizing tax rate



Note: This figure presents the results of simulations of taxes systems using different sets of tax and deduction rates. The x-axis present different values of μ, the share of costs that can be deducted. The grey line presents, for every level of deduction, the tax rate that maximizes revenue conditional on aggregate profits being no smaller than in baseline, while the blue line presents the revenue gains for each deduction and tax rate pair.

Tables

Table 1.1: Descriptive statistics

	2013	2014	2015	2016	2017	2018
Overall firms' characteristics						
Revenue (Million L)	31.35	30.81	27.99	26.49	28.31	27.47
	(336.33)	(329.80)	(293.49)	(257.53)	(317.50)	(314.64)
Deduction (Million L)	30.54	30.00	26.59	24.85	26.92	26.33
	(347.37)	(342.83)	(281.04)	(235.07)	(311.61)	(299.31)
Pre-tax profits (Million L)	0.83	0.87	1.44	1.68	1.48	1.22
	(63.59)	(65.57)	(40.91)	(33.25)	(54.17)	(57.37)
Pre-tax profit margin (%)	1.94	2.36	3.13	4.19	4.14	4.89
	(20.18)	(21.38)	(22.43)	(22.33)	(22.44)	(24.87)
Tax liability (Million L)	0.54	0.67	0.69	0.68	0.72	0.68
	(10.90)	(10.80)	(11.09)	(9.86)	(11.89)	(12.24)
Exempt from Minimum Tax (%)	.	17.8	24.6	26.3	22.2	21.1
Revenue above L10 Million (%)	18.0	17.4	16.7	17.1	17.1	17.9
Not exempt and above L10 million (%)	.	16.2	14.7	14.1	14.2	16.1
Paid Minimum Tax (%)	.	8.1	6.6	6.1	6.4	0.5
Share taxes from Minimum Tax (%)	.	29.5	21.6	19.5	19.8	14.6
Share of MNC (%)	3.5	3.6	3.2	3.0	2.8	2.6
Share taxes from MNC (%)	66.4	65.4	62.0	60.0	58.7	60.7
N	19,223	20,464	23,658	25,729	27,825	29,944

Note: This table reports descriptive statistics for the sample of corporations filing income taxes in Honduras in the period 2013-2018. Profit margins are defined as the ratio between tax liability and gross revenue and are trimmed below -100% when calculating yearly averages in this table. Exemption from the minimum taxes is defined for taxpayers in first two years of operation and/or by economic sector, and does not include taxpayers declaring revenue below the exemption threshold. Multinational corporations (MNC) are identified as firms presenting a transfer price declaration in the period 2014-2018.

Table 1.2: Share of revenue and taxes across gross revenue distribution

	2013		2017	
	(1) Revenue	(2) Taxes	(3) Revenue	(4) Taxes
Top 0.1%	28.1	32.2	28.5	34.3
Top 1%	63.0	68.6	63.4	67.2
Top 10%	91.0	91.9	90.8	93.2
Top 20%	95.8	96.2	95.6	97.1
Bottom 50%	0.6	0.9	0.5	0.7

Note: This table presents the share of total revenue and total taxes for corporations at the top 0.1%, top 1%, top 10%, top 20% and the bottom 50% of declared yearly gross revenues. Columns (1) and (2) refer to statistics in 2013, while columns (3) and (4) refer to 2017. Corporations exempt from all income taxes are excluded from the sample.

Table 1.3: Estimates by year for L10 million notch

Year	(1) Excess # Firms (B)	(2) Firms % counterfactual (b)	(3) y_u (upper bound)	(4) Δ Revenue (upper bound)	(5) ϵ_y (upper)	(6) ϵ_y (lower)
2014	84.63	4.21	12.10	2.10	1.33	0.20
	(11.14)	*(0.86)*	*(0.96)*	*(0.96)*	*(1.53)*	*(0.06)*
2015	120.54	6.12	13.00	3.00	2.61	0.40
	(10.12)	*(0.90)*	*(0.92)*	*(0.92)*	*(1.53)*	*(0.08)*
2016	142.05	5.55	11.40	1.40	0.61	0.40
	(18.63)	*(1.28)*	*(1.00)*	*(1.00)*	*(1.44)*	*(0.13)*
2017	144.54	5.22	11.40	1.40	0.61	0.35
	(11.21)	*(0.82)*	*(0.90)*	*(0.90)*	*(1.30)*	*(0.06)*
Pooled	512.96	5.46	11.80	1.80	0.99	0.35
	(30.80)	*(0.73)*	*(0.89)*	*(0.89)*	*(1.40)*	*(0.05)*

Note: This table presents estimates of change in reported revenue and elasticities for each year in the period 2014-2017 and also for all years pooled. The first column reports the estimated excess number of firms, defined above as $\sum_{b=y_L}^{y_N}(n_j - \hat{n}_j)$, while column 2 reports the ratio between excess mass and average counterfactual density in the bunching region. Column (3) presents the upper bound estimated using the convergence method and column (4) the change in revenue. Column (5) presents the upper bound estimates of reported revenue elasticity, defined in Equation 1.10, while column (6) presents the lower bound estimates using the methodology presented in section 4.3.

Table 1.4: Bunching at L10 million notch - by TPI and industries

	(1) Excess # Firms (B)	(2) Firms % counterfactual (b)	(3) Number Observations
Third-party information			
Below median TPI	253.33	7.23	6,121
	(20.77)	*(0.92)*	
Above median TPI	166.76	4.29	6,401
	(14.16)	*(0.76)*	
Industries			
Agriculture and extraction	45.75	8.01	865
	(3.62)	*(0.97)*	
Manufacturing	38.09	3.50	1,516
	(7.48)	*(1.29)*	
Utilities and construction	52.20	7.88	1,038
	(6.46)	*(1.90)*	
Automotive	16.70	4.50	650
	(6.08)	*(2.07)*	
Wholesale	65.11	5.56	1,880
	(8.93)	*(0.91)*	
Retail	71.64	6.92	1,884
	(13.01)	*(1.69)*	
Transportation, housing	31.65	5.26	1,174
	(10.09)	*(2.31)*	
Technology and finance	23.70	5.90	757
	(5.39)	*(1.49)*	
Real estate, tourism,other	48.30	3.71	2,530
	(9.40)	*(0.67)*	
Education, health, entertainment	37.00	6.24	1,050
	(10.72)	*(2.15)*	
Other services	62.23	4.77	2,298
	(10.74)	*(1.51)*	
Undeclared sectors	16.33	5.63	401
	(5.20)	*(2.06)*	

Note: This table presents estimates of bunching below the L10 million notch for firms with different levels of third-party information (TPI) (panel A) and in different industries (panel B). Column (1) presents the estimated excess mass of firms while column (2) presents the ratio between excess mass and average counterfactual density in the bunching region. Column (3) presents the number of firms for each exercise. Panel A documents that excess mass is much larger for firms with below median availability of TPI. Panel B shows that, while excess mass is large and precisely estimated across all industries, it is particularly large for sectors such as retail, construction and agriculture, and much less pronounced in manufacturing.

Table 1.5: Estimated responses at the kink

Year	(1) Excess Mass (B)	(2) Bunching (b)	(3) Delta Profit ($\Delta\Pi$)	(4) Implied ϵ_{y} (no evasion)	(5) Estimated evasion ($\epsilon_{\mathrm{y}} = 0.99$)
2014	92.04	3.07	0.60	6.67	-8.52
	(10.60)	*(0.42)*	*(0.10)*	*(0.99)*	*(1.48)*
2015	192.76	5.18	1.00	11.11	-15.18
	(13.72)	*(0.51)*	*(0.10)*	*(1.18)*	*(1.77)*
2016	212.94	5.68	1.10	12.22	-16.85
	(14.98)	*(0.55)*	*(0.10)*	*(1.24)*	*(1.85)*
2017	212.68	4.57	0.90	10.00	-13.52
	(15.59)	*(0.44)*	*(0.10)*	*(1.04)*	*(1.55)*
Pooled	777.93	5.36	1.10	12.22	-16.85
	(42.00)	*(0.38)*	*(0.10)*	*(0.91)*	*(1.36)*

Note: This table presents estimates of change in reported profit margins and evasion estimates for each year in the period 2014-2017 and also for all years pooled. Column (1) reports the estimated excess number of firms while column (2) reports the ratio between excess mass and average counterfactual density in the bunching region. Column (3) presents estimated change in profit margins. Column (4) presents the implied revenue elasticity using the decomposition in Equation 1.13 and considering no cost evasion. Column (5) computes the estimated cost evasion using the same decomposition and $\epsilon_y = 0.99$, our preferred estimate for the revenue elasticity upper bound.

Table 1.6: Deductions discontinuity at the notch

		Deductions components (% of revenue)				
	(1)	(2)	(3)	(4)	(5)	(6)
	Total deductions	Labor	Materials	Operation	Financial	Other
Jump in cost	-0.265***	0.0108	-0.0483**	-0.00268	0.00419	0.0120
	(0.06)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Slope below threshold	0.983***	-0.00573**	0.00793**	-0.00133	0.000893	-0.00281
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Slope change above threshold	-0.0283**	0.00207	-0.00200	0.00162	-0.00137	0.00339
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Intercept	9.764***	0.250***	0.373***	0.233***	0.0205***	0.0931***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)
Observations	160	160	160	160	160	160
R-Squared	0.999	0.214	0.225	0.149	0.266	0.165

Note: This table reports results of "donut-hole" discontinuity regressions using binned data for firms declaring between L4 and L20 million in revenue. The dependent variable is median claimed deductions in column (1) and mean cost as a share of declared revenue, for each cost item, in columns (2) through (6). The sample is restricted to firms with electronic declarations between 2015-2017 and exclude approximately 3% of firms for which the sum of claimed deductions computed from individual cost lines does not match total claimed deductions. We also trim the sample at the first and 99th percentile of declared profit margin distributions. Robust standard errors are presented in parenthesis.

Table 1.7: Simulated impact of counterfactual minimum tax policies

Exemption Threshold (L million)	Minimum tax rate (%)	Share taxpayers owing MT (%)	Tax revenue increase (%)	Tax liability change for MT firms (%)	Change aggregate profits (%)	% tax paid by bunchers	Tax loss from bunchers (%)
10	1.5	62.4	30.3	122.5	-10.0	23.8	1.0
10	0.5	28.0	3.6	94.5	-0.5	17.4	0.5
10	2.0	70.6	49.2	146.8	-17.4	24.0	1.3
20	0.5	16.5	3.3	92.9	-0.5	23.6	0.3
20	1.5	36.4	27.7	120.7	-9.1	30.1	1.2
20	2.0	40.6	45.1	144.8	-16.0	23.7	1.5
50	0.5	7.4	2.6	88.7	-0.4	22.7	1.5
50	1.5	17.1	22.8	117.3	-7.6	30.0	2.4
50	2.0	18.5	36.9	141.2	-13.2	25.6	3.3

Note: This table presents results of counterfactual minimum tax policies using the calibrated model. Columns (1) and (2) present the counterfactual notch above which the firms are subject to the minimum tax and the tax rate applied on gross revenue, respectively. Column (3) presents the share of taxpayers with revenue above the exemption threshold paying minimum taxes; column (4) presents the increase in collected tax revenue; column (5) presents the aggregate increase in tax liability faced by firms paying the minimum tax; column (6) presents the change in aggregate corporate profit; column (7) presents the ratio between total tax liability below and above the exemption threshold for bunchers; and column (8) presents the share of total potential revenue collected under minimum tax that is lost from bunching taxpayers.

Table 1.8: Simulated impact of counterfactual increase in average profit tax

Average profit tax rate (%)	Tax revenue increase (%)	Change aggregate profits (%)
30	12.2	-6.9
35	22.2	-13.5
40	29.9	-20.0
45	35.9	-26.3
50	39.3	-32.5

Note: This table presents results of counterfactual policies where the average profit tax rate is increased for firms declaring gross revenue above L10 million, using the calibrated model. Columns (1) presents the average profit tax rate simulated in each scenario. Column (2) presents the total % increase in tax collection while column (3) presents aggregate profit losses.

Chapter 2

Targeting in Tax Compliance Interventions: Experimental Evidence from Honduras

2.1 Introduction[1]

State capacity is intimately linked to governments' ability to enforce taxation. Tax collection in low-income countries, in that sense, looks very different from what is observed in higher income settings. Tax revenue in OECD countries are equivalent to about 16% of GDP, while that ratio is less than 12% in low-income countries. The composition of collected taxes is also very different: while high-income countries collect over 50% of its total taxes through income taxes, this number is less than 30% for low and middle-income countries, where taxes on goods and services are much more important.

Income taxes can be particularly hard for governments to collect given the complexity involved in correctly assessing liabilities (Gordon and Li, 2009). The capacity to obtain and process the type of information needed to assess income tax liabilities, however, seems to be rapidly increasing in several low- and middle-income countries. International organizations are committed to improve tax capacity: approximately $200 million in Official Development Assistance (ODA) were aimed at improving tax capacity in 2014 (International Monetary Fund et al., 2016). These improvements in "government intelligence", nonetheless, do not translate automatically into higher compliance, but require other actions by the authorities. First, providing taxpayers with knowledge on the new information set of the authorities might trigger increased "voluntary" compliance, if their beliefs about the probability of punishment is updated. Second, authorities need to explore this new information to target their policies: audits, for example, are very expensive and labor-intensive, and should be targeted to maximize revenue recovery.

In this paper we partner with the Tax Authority in Honduras (SAR, for *Servicio de*

[1]The main email experiment in this chapter 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)

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 to assess non-compliance, we conduct an experiment with approximately 32,000 taxpayers considered to be at-risk. Taxpayers in the control group receive a regular reminder about the income tax filing deadline - an usual communication provided by the tax authority. Those in the treatment group, in contrast, are informed that the tax authority observe specific transaction 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 of interest, measured using administrative data on tax filing.

While there is strong evidence that the average effect of these types of intervention is positive, increasing compliance (see Mascagni (2018) for a recent review), this is not inconsistent with negative or null impacts on some subset of taxpayers (De Neve et al., 2020). We argue that measuring this heterogeneity is crucial for authorities to better target this kind of intervention. While the return on investment of these studies is often perceived to be extremely high given the close to zero 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. That 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.

We present some heterogeneity analysis of the treatment effects on key dimensions built into the experimental design, but also explore the rich data on the 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 (Athey and Imbens, 2016, Wager and Athey, 2018) to estimate conditional average treatment effects (CATE) across the sample of taxpayers. With those estimates at hand, we can assess how the treatment effects compare to the perceived risk level of each taxpayer, estimated by the tax authority using non-experimental data from before the intervention. This is informative about future communication interventions: currently those with higher assessed risk are more likely to be targeted, but it's not clear whether these are the most responsive taxpayers.

Our main finding is that, on average, the intervention had a null effect on a range of compliance outcomes. Taxpayers assigned to receive the treatment were no more likely to file income taxes nor declared higher revenues or taxable income. We estimate precise null effects: we can rule out changes in filing probability as small as 1 percentage point and increase 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 nor 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 groups, those receiving the email 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 (+ L103,000[2] 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 them 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 lockdown that depressed economic activity and led the filing deadline for income taxes to be 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: while 86% of taxpayers in our sample filed taxes in 2019, only 62% did in 2020. All of these imply that the external validity of our findings should be taken with caution.

Our study contributes new evidence to the now large literature on communication interventions aimed at increasing taxpayers compliance, reviewed by Mascagni (2018) and Hallsworth (2014). 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, among treated subjects, how we frame the importance of compliance. One third of the treatment group was reminded that non-compliance carries monetary penalties; another third was reminded that SAR can deny documents necessary for their operations; 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 and Singhal,

[2] Approximately USD 4,000.

2014) in these type of letter or email interventions have so far been mixed. De Neve et al. (2019) finds 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 and 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 a threat 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 the effect of the tax morale arm is different from the other interventions.

Finally, our exercise using the causal forest algorithm[3] 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 different setting, Hussam et al. (2020), working with experimental grants to entrepreneurs in India, show 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.

2.2 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, 2018b).

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 a flat rate of 25% on profits (gross revenues minus cost deductions)[4]. Non-incorporated taxpayers face a progressive tax schedule: in FY2019, net income below L159,000[5] was exempt from taxation and higher incomes are taxed with increasing marginal tax rates in three brackets of 15%, 20% and 25%[6]. All taxpayers with commercial activities (both corporations and non-

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

[4]Since 2003, corporations must also assess a 1% tax on its 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.

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

[6]The 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.

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, 2018b). Furthermore, SAR has been working to consolidate these institutional reforms and implement new tools to ensure fiscal compliance, including a sophisticated internal risk model following international best practices.

2.3 Research design

2.3.1 Intervention

Jointly with the tax authority, we sent emails to approximately 32,000 taxpayers seven weeks before the original income tax filing deadline for FY2019[7]. 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 the simple fact of receiving a message from the government[8].

The control group received an email[9], presented in Figure A3, with a reminder about the filing deadline and the importance of truthfully reporting their tax liabilities[10]. 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

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

[8]The 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.

[9]All 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.

[10]The 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.".

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 A4 through Figure A9[11].

Taxpayers for which third-party information was available (Figure A4 through Figure A6) received a message informing that "In the sources of information available in 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 exports[12]. 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 A7 through Figure A9) that "In the sources of information available in 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 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."[13] 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

[11]To 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 policy-makers in Honduras.

[12]Unlike 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.

[13]The 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/).

of "pagos a cuenta", solvency and fiscal documents"[14].

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's 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.

2.3.2 Experimental sample and randomization

The experimental sample is comprised 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 Gestion de Riesgo de Honduras*, MGR-H)[15] 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 risk[16]. 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

[14] "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.

[15] The 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).

[16] The 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).

(*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.

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 sample[17]. Second, power calculations exercises suggested we could significantly increase minimum detectable effects (MDE) by dropping extremely large taxpayers. Figure A1 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 taxpayers[18].

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 and 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 probabilities[19].

---

[17] Approximately 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.

[18] Details about power calculations used in Figure A1 are discussed in Appendix B.1. Taxpayers with declared gross revenue above L19.4 million (approximately USD 780,000) in either FY2017 or FY2018 were excluded from the sample

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

2.3.3 Benchmarking results: expert forecast survey

In order to benchmark the magnitude of our results, we follow DellaVigna and Pope (2018) and DellaVigna et al. (2020) and surveyed academics and policy-makers 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 policy-makers or researchers in international development organizations[20].

The main results of the survey are presented in Table 2.7. In terms of magnitudes of 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 to be more effective than the Tax morale treatment. This is true across the different outcomes and, as seen in Figure 2.7, 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.

[20]We 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.

2.4 Data

2.4.1 Baseline descriptive and balance

We present baseline descriptive statistics of our experimental sample in Table 2.1. 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)[21]. 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 authority[22]. 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 2.2 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 observables[23].

[21]The 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.

[22]These 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.

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

2.5 Results

2.5.1 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:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \gamma_s + \epsilon_i \tag{2.1}$$

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

$$Y_i = \delta_0 + \delta_1 Open_i + \delta_2 X_i + \gamma_s + v_i \tag{2.2}$$

$$Open_i = \lambda_0 + \lambda_1 T_i + \lambda_2 X_i + \gamma_s + \mu_i \tag{2.3}$$

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 2.3, 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 2.2 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 2.1 to estimate ITT as:

$$Y_i = \alpha + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + \beta_2 X_i + \gamma_s + \nu_i \tag{2.4}$$

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 2.4. We again cannot reject the null hypothesis of zero treatment effect. ITT are very small across the treatment arms and never statistically significant[24]. While LATE estimates are somewhat larger in magnitude they are similarly imprecisely estimated.

2.5.2 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 en 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.

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 2.3, 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 2.5 we present regression estimates showing no differential filing probability by each deadline.

[24] In Table A3 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 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 2.5. Both ITT and LATE estimates suggest no statistically significant impact of the intervention on any of the secondary outcomes examined.

2.5.3 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 sample[25]: 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 results are presented in Table 2.6.

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 compliance.[26]. In panel A we estimate the 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

[25]These specific heterogeneity dimensions were pre-specified before the experiment.

[26]While 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 is driven by different email content of the communication or by heterogeneity between the groups.

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 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 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.

2.5.4 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 heterogeneity, we use a causal forest algorithm (Athey and Imbens, 2016, Wager and Athey, 2018) to predict conditional average treatment effects (CATE) on subgroups in our experimental sample (De Neve et al., 2019)[27]. Causal forests predict CATE for each taxpayers, 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 and Imbens (2016) recommend using "honest trees", which use separate sub samples to perform partition and estimate treatment effects.

In Figure 2.4 we present a histogram of predicted treatment effect on the filing probability for our entire sample[28]. 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 2.5, 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 risk.

[27]For a detailed explanation of the causal forest algorithm and application in economics, see Davis and Heller (2017)

[28]We show in Figure A2 tha results are similar if we consider the amount of gross revenue declared as the outcome of interest.

2.5.5 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 authority[29]. 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, 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.

2.6 Conclusion and policy implications

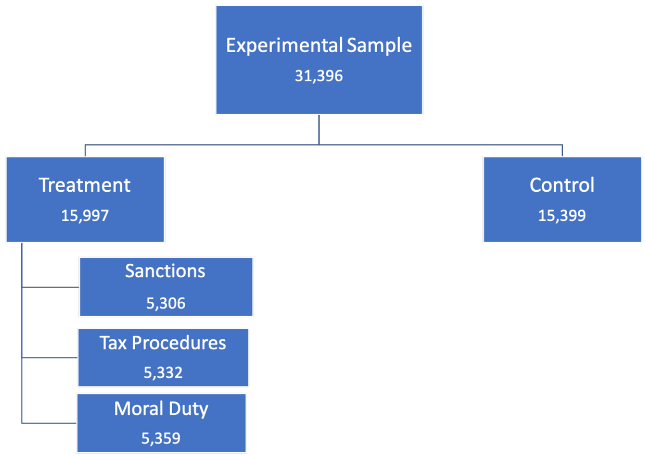
We experimentally informed thousands of taxpayers in Honduras that the tax authority had information about their previous transactions in order to assess whether that increased tax compliance. Overall we found null results: across a range of outcomes, taxpayers in the treatment groups behaved no differently 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, however, 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 reported 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. New methods that combine experimental variation, rich administrative data and machine learning models will be useful in advancing that research agenda.

[29] While 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.

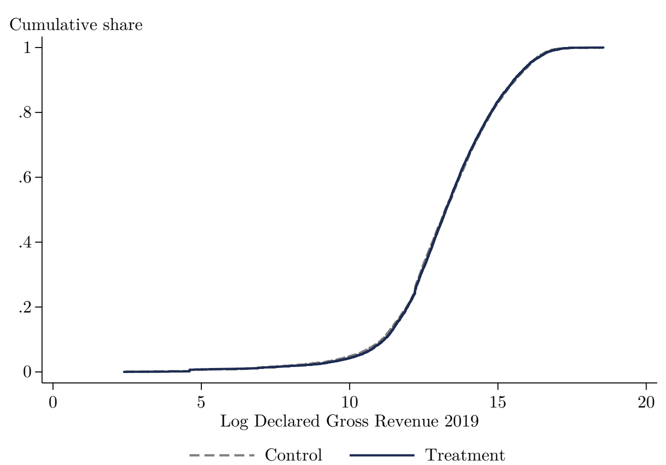
Figures

Figure 2.1: Experimental Design



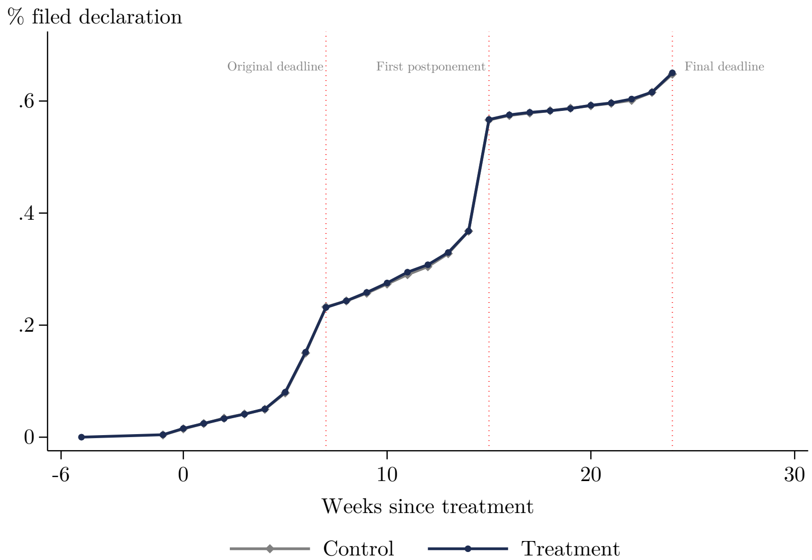
Note: This figure shows the design of the experiment and sample sizes between control and each of the treatment arms. Randomization was stratified in 60 strata defined by availability of third-party information; whether taxpayers is corporation; three geographical regions; and five risk-levels defined by the tax authority (2*2*3*5 = 60 strata).

Figure 2.2: Cumulative distribution function of declared gross revenue - Treatment vs. control groups

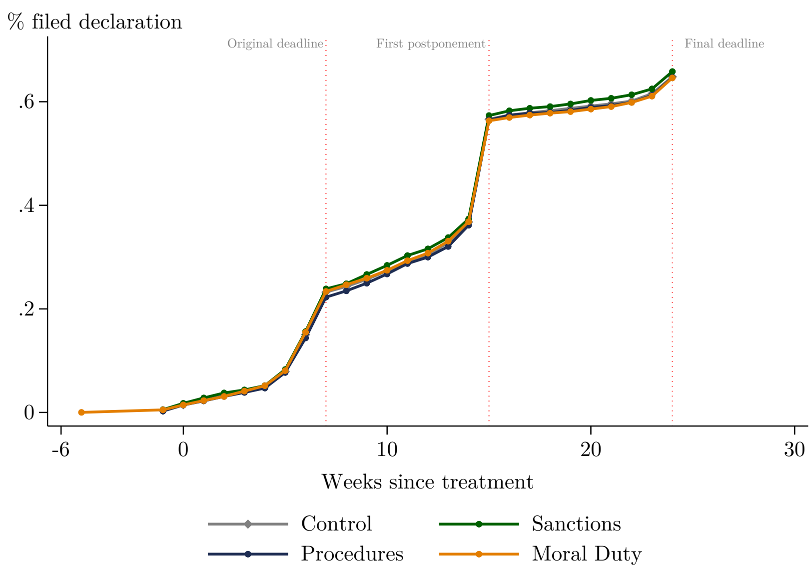


Note: This figure shows the cumulative distribution function of reported gross revenue for FY2019, separately for taxpayers assigned to treatment (blue) and control (grey). We observe no meaningful difference between the two distributions, consistent with the null effect presented in Table 2.3. The sample is restricted to taxpayers declaring at least L10 in gross revenue for better visualization.

Figure 2.3: Cumulative share of taxpayers filing per week



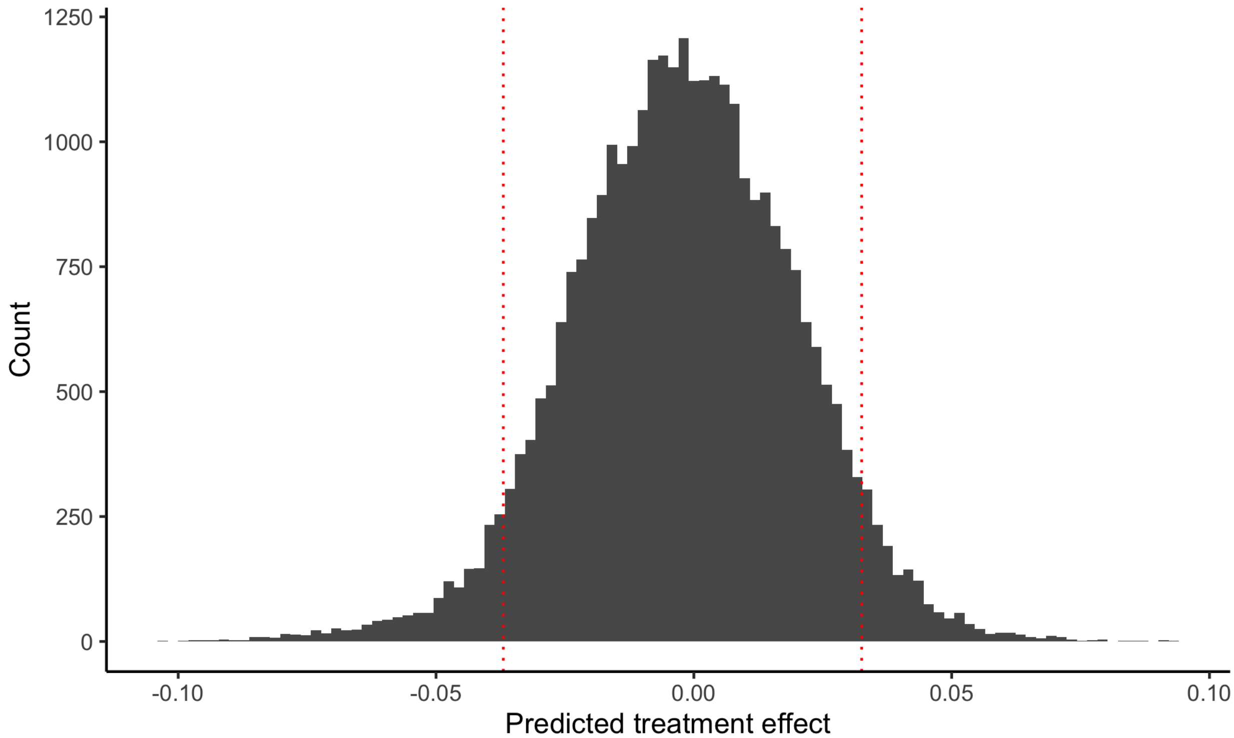
(a) Pooled treatment arms



(b) Separate treatment arms

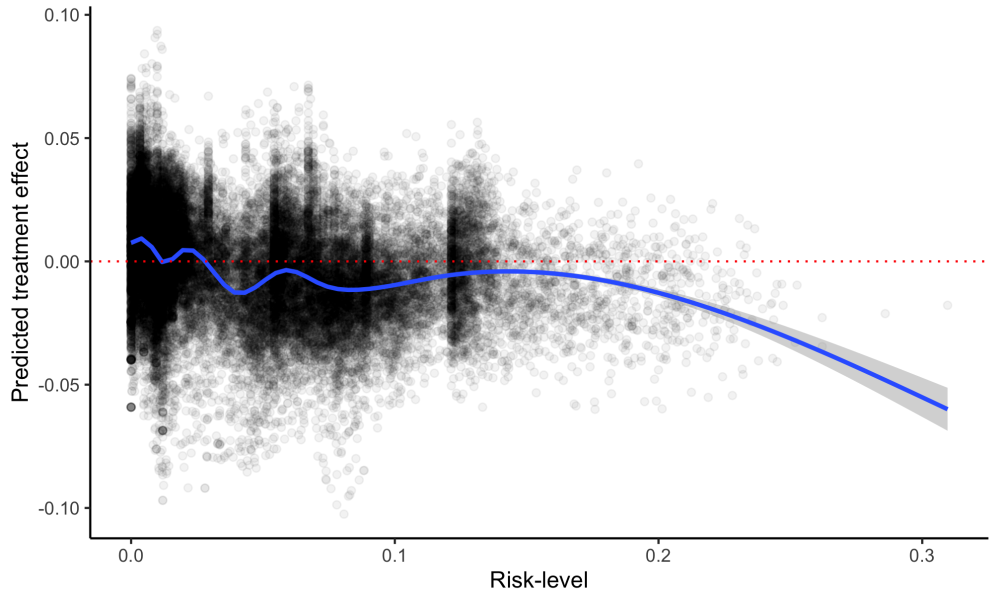
Note: This figure presents the cumulative share of taxpayers filing income tax by each week since the intervention. Panel A presents results for the pooled treatment arms vs. control group whereas Panel B presents results separately for each treatment arm. E-mails were sent on March 11-12th 2020 and 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 2.4: Estimated treatment effects on filing probability (Random Forest)

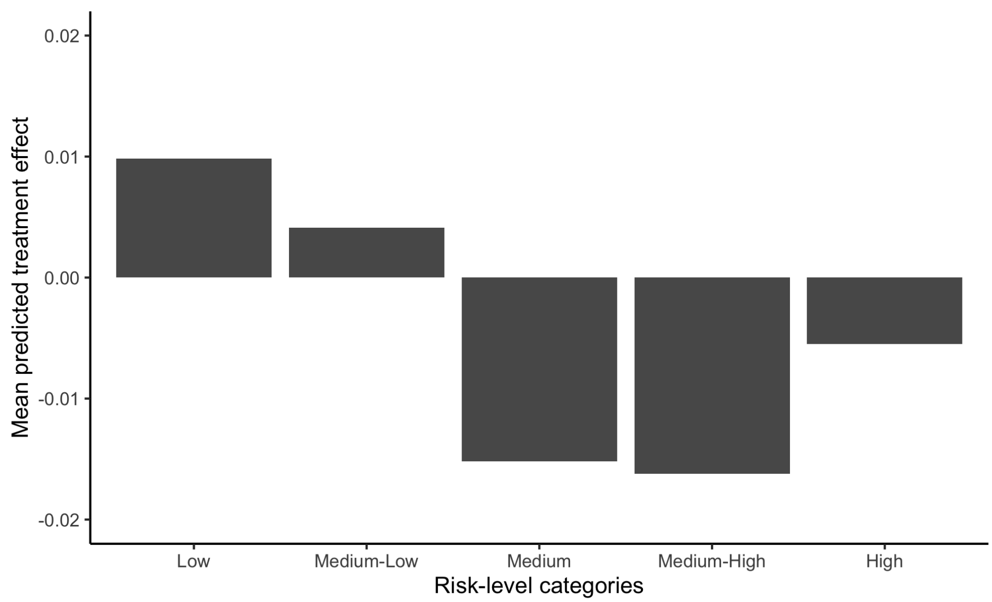


Note: This figure presents the histogram of estimated Conditional Average Treatment Effects (CATE) for the experimental sample, obtained using a random forest model. The dashed lines represent the 5th and 95th percentile of the distribution. CATE are symmetrically distributed around zero and 90% of the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

Figure 2.5: Estimated treatment effects on filing probability (Random Forest) across risk levels



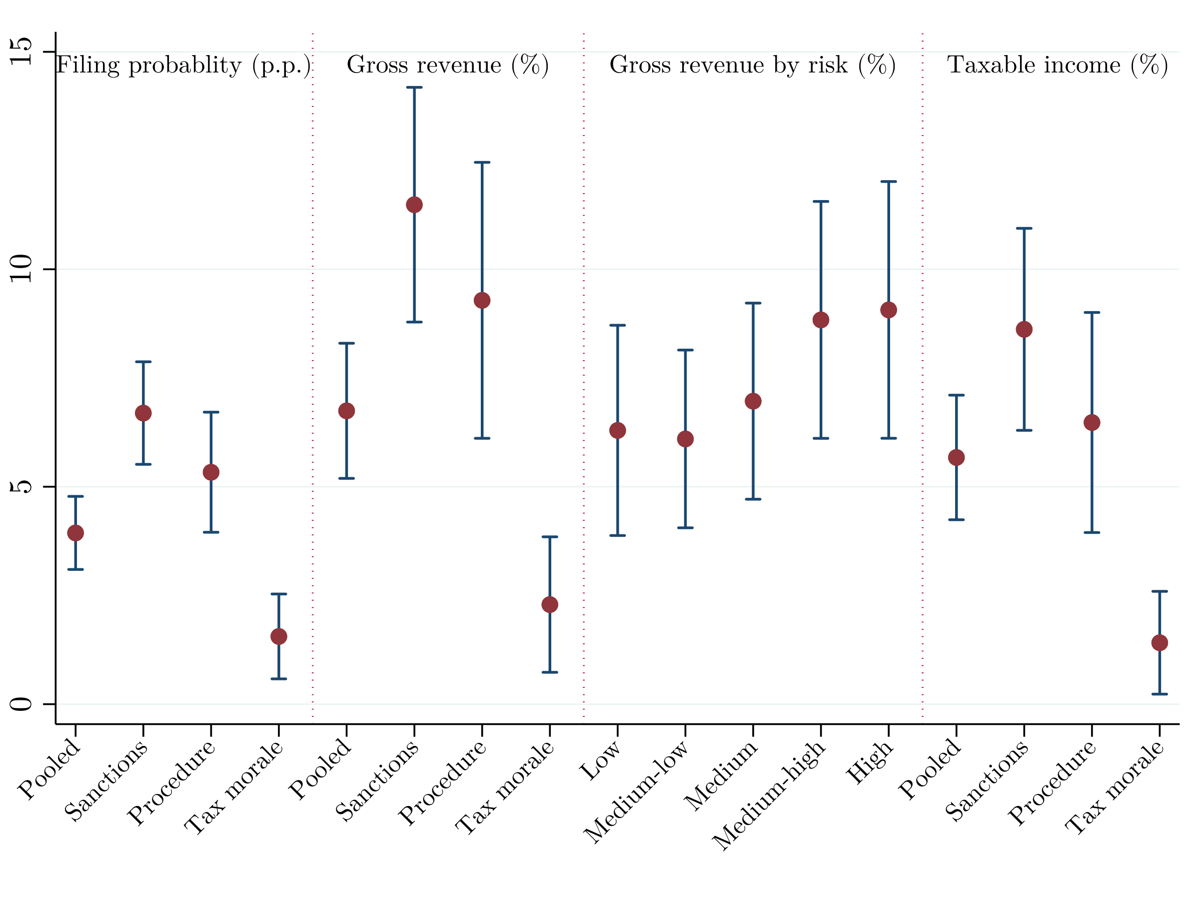
(a) Scatter plot of predicted treatment effects vs. ex-ante risk-level



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

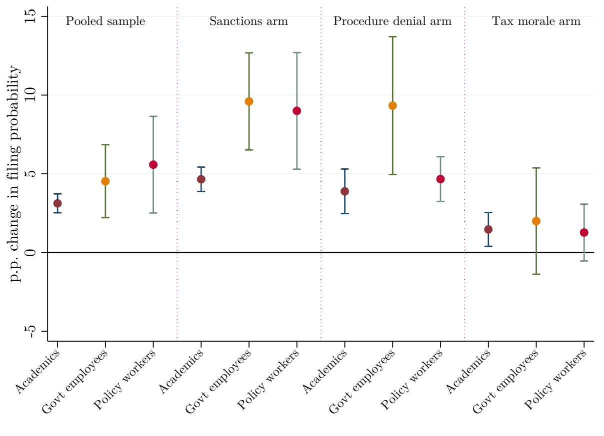
Note: This figure documents how predicted treatment effect varies according to the ex-ante assessed risk level of taxpayers. Panel A presents a scatter plot of predicted treatment effects (vertical axis) and ex-ante assessed risk level (horizontal axis), as well as a local polynomial fit. Panel B presents 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 2.6: Survey estimates

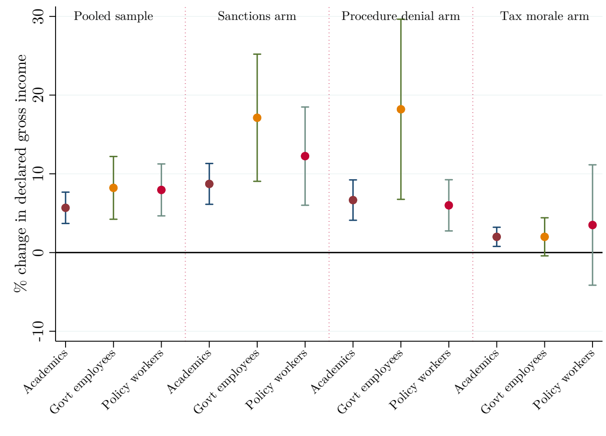


Note: This figure presents means and 95% confidence-intervals for predictions of treatment effect-size on each of the primary outcomes.

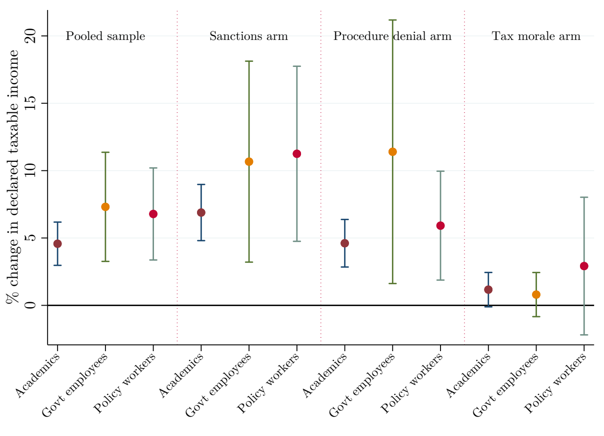
Figure 2.7: Forecasts by groups of respondents



(a) Filing probability



(b) Declared gross income



(c) Declared taxable income

Note: This figure presents means and 95% confidence-intervals for predictions of treatment effect-size on each of the primary outcomes.

Tables

Table 2.1: Descriptive Statistics - Final Sample

	Mean	SD	p50	N
Panel A: Taxpayers' characteristics				
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
Panel B: Past filing behavior				
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
Income Tax 2018				
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
Income Tax 2017				
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
Panel C: Third-party information				
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
Anomalies				
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
Risk assessment				
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 2.2: 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 2.3: Primary Outcomes - Estimating Program Effects

	(1) Filed declaration	(2) Gross Revenue	(3) Deductions	(4) Taxable Income
ITT estimates				
Treatment	-0.001	7.284	2.750	0.383
	[-0.010,0.009]	[-48.146,62.714]	[-51.333,56.833]	[-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
LATE estimates				
Opened email	-0.002	21.277	7.958	1.172
	[-0.030,0.026]	[-140.918,183.472]	[-150.301,166.217]	[-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 2.1. The second panel presents Local Average Treatment Effect (LATE) estimates using the specification in Equation 2.2. 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 2.4: Primary Outcomes - Different Treatment Arms

	(1) Filed declaration	(2) Gross Revenue	(3) Deductions	(4) Taxable Income
ITT estimates				
Sanctions treatment	0.004	24.804	8.857	4.078
	[-0.009,0.018]	[-56.611,106.218]	[-70.397,88.112]	[-10.504,18.659]
Tax procedures treatment	-0.005	23.813	27.387	-1.905
	[-0.019,0.008]	[-54.347,101.974]	[-50.125,104.899]	[-15.422,11.611]
Moral duty treatment	-0.002	-26.499	-27.805	-0.996
	[-0.015,0.012]	[-96.645,43.646]	[-97.125,41.515]	[-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	71.777	25.612	11.803
	[-0.026,0.052]	[-163.608,307.162]	[-203.561,254.786]	[-30.362,53.968]
Opened (Procedures)	-0.015	67.637	77.789	-5.413
	[-0.054,0.023]	[-154.072,289.347]	[-142.066,297.643]	[-43.774,32.948]
Opened (Tax morale)	-0.005	-81.077	-85.083	-3.046
	[-0.046,0.036]	[-295.647,133.492]	[-297.145,126.979]	[-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 2.4. 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 2.5: Secondary Outcomes

	(1) Filed 1st deadline	(2) Filed 2nd deadline	(3) Total taxes paid	(4) Total sales taxes	(5) Revised previous filing
ITT estimates					
Treatment	0.003	-0.002	-0.408	-1024.434	0.000
	[-0.006,0.011]	[-0.012,0.008]	[-2477.610,2476.794]	[-2841.898,793.031]	[-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
LATE estimates					
Opened email	0.008	-0.006	-1.526	-2994.794	0.001
	[-0.016,0.032]	[-0.036,0.025]	[-7251.354,7248.301]	[-8313.479,2323.892]	[-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 2.1. The second panel presents Local Average Treatment Effect (LATE) estimates using the specification in Equation 2.2. 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 2.6: Heterogeneity Analysis

	(1) Filed declaration	(2) Gross Revenue	(3) Deductions	(4) Taxable Income
Treatment	-0.024**	35.906	21.270	11.154*
	(0.009)	(41.373)	(44.146)	(6.515)
Treatment * Third party info available	0.032***	-40.069	-25.926	-15.078
	(0.011)	(55.216)	(56.281)	(9.495)
Third party info available	0.093**	500.667	620.852**	-43.908*
	(0.041)	(312.136)	(298.499)	(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**	103.221**	92.741*	10.740*
	(0.011)	(47.219)	(47.450)	(5.882)
Treatment * Medium-low risk	-0.026*	-112.899*	-120.320*	-11.226
	(0.014)	(65.735)	(66.185)	(9.272)
Treatment * Medium risk	-0.034**	-149.615**	-126.888*	-15.872*
	(0.015)	(71.452)	(72.057)	(8.407)
Treatment * Medium-high risk	-0.041***	-138.890	-107.568	-19.686
	(0.016)	(94.713)	(89.265)	(23.093)
Treatment * High risk	-0.017	-42.755	-73.463	0.771
	(0.020)	(135.415)	(133.553)	(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**	10.678	5.220	2.355
	(0.006)	(23.005)	(22.598)	(3.129)
Treatment * Corporation	-0.043***	-10.381	-7.557	-6.033
	(0.010)	(75.998)	(74.023)	(14.889)
Corporation	0.131***	485.694	611.590**	-48.464*
	(0.041)	(312.441)	(298.872)	(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	23.814	24.246	0.921
	(0.007)	(34.390)	(34.683)	(2.803)
Treatment * Distrito Central	-0.005	-42.821	-44.075	-6.895
	(0.012)	(69.938)	(66.987)	(16.621)
Treatment * San Pedro Sula	-0.020	-22.412	-43.710	6.174
	(0.012)	(74.773)	(74.066)	(8.735)
Distrito Central	-0.106***	-458.701	-585.382*	55.018**
	(0.041)	(313.115)	(299.202)	(27.654)
San Pedro Sula	0.007	-48.828	-68.546	-10.617
	(0.031)	(315.594)	(307.384)	(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 2.1 augmented by interacting the treatment dummy with the characteristics of interest. Robust standard errors presented in parentheses. (* p<0.1, ** p<0.05, *** p <0.01)

Table 2.7: Descriptive Statistics - Forecast survey

	Mean	p50	SD	Min	p25	p75	Max	N
Filing probability (p.p.)								
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
Gross revenue (%)								
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
Taxable income (%)								
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.

Chapter 3

Selecting Top Bureaucrats: Admission Exams and Performance in Brazil

3.1 Introduction

Public employees play a key role in designing and delivering essential public services to development worldwide (Finan, Olken, and Pande, 2017). Recent studies have focused on the role that incentives and monitoring can play to improve the performance of government employees, particularly of frontline providers (Ashraf, Bandiera, and Jack, 2014). However, the impact of such tools is limited for the typical bureaucrat in developing countries whose career is often characterized by tenure benefits, absence of performance pay, and promotion based on seniority (Bertrand, Burgess, Chawla, and Xu, 2019). In the face of such low-powered incentives once hired, the issue of how to select bureaucrats becomes essential.

One widely used selection mechanism, particularly in some of the largest developing countries like Brazil, China and India, is competitive, impersonal examinations. These may reduce corruption and patronage in hiring by political leaders (Brollo, Forquesato, and Gozzi, 2017, Weaver, 2016, Colonnelli, Prem, and Teso, 2020), but potentially at the expense of assessing candidates' soft and noncognitive skills (Hoffman and Tadelis, 2018).[1] Further, it is an open empirical question, and a highly policy-relevant one, as to whether examinations reliably select candidates who are productive on the job.

In this paper we study whether objective examinations predict job performance for a selected group of public sector employees in Brazil: state judges. Similar to the majority of civil servants in the country, judges are selected through highly competitive and mostly impersonal examinations, comprised of written and oral exams. Candidates are ranked based on their grades and top performers are offered jobs based on pre-specified number of available positions. Our estimates suggest that, within selected candidates, those ranking higher in exams are also high performers on the job as judges. In terms of magnitudes, we show that candidates that rank in the top quintile in their admission exam cohort dispose

[1]These characteristics of public sector recruitment differ markedly from what is observed in the private sector, where managers and human resources officers have wide discretion in selecting employees and subjective assessments plays an important role through interviews, for example (Hoffman, Kahn, and Li, 2018).

of approximately 20% more cases on a monthly basis than those in the bottom quintile.

The first step of our analysis is to estimate judge-level measures of performance. To do so, we leverage novel administrative data to construct a panel of judicial productivity at the judge-court-month level, covering the universe of state judges working in Brazil from 2009 through 2015. Across the 76 months encompassed by our data, judges often work in several different courts. This high level of mobility allows us to estimate a two-way fixed-effects model akin to those in the labor literature decomposing wage variation between worker and firm fixed-effects. We separately estimate judge and court fixed-effects, and show that judges are important in explaining the observed variation in output: using bias-adjusted estimates, individuals' fixed effects account for at least 23% of the variation in the number of cases disposed.

We focus on the number of cases disposed for two reasons. First, timely delivery of judicial decisions is critical in developing countries. At the current pace of case disposition, it would take Brazilian courts three years to clear the backlog, assuming no additional cases were initiated (Conselho Nacional de Justiça, 2018). Ponticelli and Alencar (2016) show that judicial timeliness matters for important economic outcomes. They explore differences in court congestion across Brazilian municipalities to show that a bankruptcy reform has larger effects on investment and financial access of firms located in district with more efficient courts. Second, the speed with which judges dispose of cases is considered an important indicator of performance by Brazil's judicial branch and is used, along with other considerations, to define promotions throughout judges' career.

Yet, theory suggests that if the quantity of cases disposed is easily observable but quality is not, judges might divert efforts into the observable dimension of performance (e.g. speed), possibly to the detriment of quality (Holmstrom and Milgrom, 1991). One implication of that hypothesis is that fast judges might sacrifice important inputs in the "production process" of case disposition in order to increase output quantity. We leverage our detailed microdata to show that this is not the case for one important input for court decisions: the number of hearings held by judges. We re-estimate our two-way fixed-effects model using hearings as the dependent variable and show a strong, positive correlation between judges fixed-effects in both models. This show that faster judges are not decreasing the number of hearings, one important input for high-quality case decision.[2]

Next, we examine whether judges highly ranked on entrance exam actually perform better on the job. We collect novel data on admission exams for over 25% of all state judges working in Brazil during the period covered by our productivity dataset, including their final rankings and grades. Our results suggest a positive and strong correlation between admission exam and on-the-job performances: we estimate that, within cohorts, being ranked in the top quintile of one's admission examination is correlated with a 0.2 s.d. increase in estimated FE when compared to the bottom quintile. This result is robust to different measures of performance both in exams and on the job; the results are also robust when excluding the top and bottom candidates in each cohort.

Taken as a whole, our results suggest that admission exams are able to rank candidates

[2]As we will discuss in Section 3.3, one natural measure of quality of case disposition would be the likelihood of case reversals in higher courts. However, there is no such systematized data for Brazil in the time period we study.

in a way consistent with their future performance on the job.[3] In order to make progress in understanding which dimensions of the exams are most relevant for future performance, we restrict our sample to a subset of judges for which we can break-down final grades in each of the recruitment phases and consider whether achievement in any of the specific exams is particularly predictive of performance on the bench. Across different specifications, grades on the Judicial Decision Writing exam, where candidates are given a hypothetical case and asked to produce a decision, are the strongest predictors of performance. While these correlations should be interpreted with caution, they suggest that the use of impersonal examinations to screen candidates might be particularly efficient if focused on "practical" exams that mimic the situations faced by employees on the job.

Our paper makes contributions to three strands of literature. First, we add to the recent literature on the selection of workers in the public sector. The impersonal admission examinations used to select judges is aimed precisely at avoiding the kind of patronage documented in contexts as different as colonial governors in the British Empire (Xu, 2018) and public officials hired at the discretion of newly elected politicians in Brazil (Colonnelli, Prem, and Teso, 2020). In the context of high courts in Pakistan, Mehmood (2020) also documents that politically appointed judges are more likely to rule in favor of the government. However, the use of discretion when selecting officials need not lead to negative selection of providers. In an extreme example, Weaver (2016) shows that the selection of community health workers' supervisors by outright bribery leads to high quality workers being hired, since wealth and performance are strongly positively correlated. To the best of our knowledge, our paper is the first to provide evidence that performance in admission exams is predictive of the timely disposition of cases within the judicial system in Brazil. We contribute to a nascent literature documenting that performance on impersonal admission exams, the dominant screening process for public sector employees in several low and middle income countries, is informative about performance on the job.[4]

Our research also adds to efforts of measuring the role of bureaucrats in determining public sector performance (Finan, Olken, and Pande, 2015). While the relevance of front-line service providers like teachers (Chetty, Friedman, and Rockoff, 2014, Muralidharan and Sundararaman, 2011, Duflo, Dupas, and Kremer, 2015) and community health workers (Deserranno, 2019, Weaver, 2016, Dal Bó, Finan, and Rossi, 2013, Ashraf, Bandiera, Davenport, and Lee, 2020) have been extensively discussed, the role of other decision-makers in the public sector bureaucracy has only recently garnered more attention. Our empirical strategy, exploring the movement of judges between courts to identify individual fixed-effects, is particularly related to the work of Best, Hjort, and Szakonyi (2019) on the role of procurement officers in Russia in explaining price dispersion in public purchases, and of Fenizia (2020) on

[3] The implications of sidelining any subjective assessments of candidates' qualities for job performance are not obvious. If knowledge about objective exam content is the crucial requirement to perform well, or if subjective traits that predict exam performance are also correlated with service delivery capacity, then objective recruitment strategies might be simultaneously effective and impartial. If certain subjective characteristics are very relevant to perform well on the job but hard to capture on objective admission examinations, nonetheless, these recruitment strategies are maintaining impartiality at the expense of accuracy.

[4] Aman-Rana (2020) documents that public officials ranked at the top 10% of their admission cohorts in Punjab, Pakistan, also collect more taxes. Bertrand, Burgess, Chawla, and Xu (2019) documents a positive correlation between admission exam rankings and performance measured by 360 degree evaluations of IAS officers in India.

how managers of Social Security offices in Italy explain variation in productivity.

Lastly, we contribute with new evidence about the determinants of judicial efficiency in the developing world. Research in Pakistan (Chemin, 2009), Senegal (Kondylis and Stein, 2018) and Mexico (Sadka, Seira, and Woodruff, 2018) has shown that judicial reforms aimed at simplifying procedures and speeding up the disposition of cases can be effective. Kondylis and Stein (2018), in particular, collect rich data at the case-level and show that higher speed in commercial courts in Senegal does not seem to affect the quality of decisions. The effects of judicial reforms more broadly also depend on the capacity of courts to deliver timely decisions, as shown in Ponticelli and Alencar (2016) and Rao (2020). To the best of our knowledge, our paper is the first to perform a two-way fixed effects decomposition to quantify the importance of judges to court efficiency and, further, to show that competitive examinations successfully selects the most efficient magistrates.

The remainder of this paper is organized as follows. Section 3.2 describes the structure of Brazilian courts and the admission process for judges. Section 3.3 presents the data used, provides descriptive statistics and explains how we obtain the sample used when estimating the two-way fixed-effects model. Section 3.4 describes our empirical model, identification and estimation procedures. The main results are presented in Section 3.5, while Section 3.6 concludes and discusses avenues for future research.

3.2 Institutional Context

3.2.1 Brazilian Courts

The Brazilian Judiciary is comprised of five branches: State, Federal, Electoral, Labor and Military Courts. This paper uses data exclusively from State courts, which cover all cases that are not specifically under the competency of the other branches (that is, State courts have residual judicial competency). The majority of criminal and civil cases fall under the competence of State courts: in 2017, over 60% of all cases in the Judiciary were allocated to the first instance of these courts (Conselho Nacional de Justiça, 2018).

Each of the 27 Brazilian federative units (26 states plus the federal district) is responsible for establishing and organizing the state courts. Within each state, the main administrative unit of the state justice are the judicial districts (*comarcas*), which encompass one or more municipalities. Judicial districts are mainly divided in three administrative levels, related to the underlying demand for judicial services: *first level* districts are located in rural or less urbanized municipalities and contain a single court of general competency (i.e. it covers all types of cases); *second level* districts are located in municipalities with smaller cities and encompass specialized courts, often separate Civil and Criminal courts; while *third level* districts encompass the state capital and possibly other large cities, and include several specialized courts.

Court congestion is considered a serious impediment to the efficient application of justice in Brazil (Ponticelli and Alencar, 2016): at the state level, there were over 60 million cases allocated to courts in 2017. If no more cases entered the justice system and current levels of productivity were held constant, it would take almost three years to clear the backlog (Conselho Nacional de Justiça, 2018). While overall congestion is very high, there exists a large dispersion among judicial districts not fully explained by simple regional differences: Schiavon (2017) shows that the dispersion of several congestion and performance measures is larger within states than between states, highlighting the relevance of local determinants in explaining variation in performance.

The importance of timely decisions by courts and the challenges faced by the Brazilian Judiciary in that regard have not escaped the attention of policy-makers and legislators. For example, the 2004 Constitutional Amendment that created the National Justice Council, among several other sweeping changes to the organization of the Judiciary, also included specific language requiring that judges' promotion take into account "objective criteria of productivity".[5] During the launch of the Open Justice System, in 2008, a Supreme Court Justice praised the tool as a way to "improve the management of justice and decrease the slowness of decisions".[6]

3.2.2 Selection of judges through competitive examinations

The broad rules for judges' recruitment are determined by Article 93 of the Brazilian Constitution. It states that all judges should be selected through public examinations (*Concursos Públicos*); since 2004, a Constitutional Amendment also institutes the requirement of three years of professional judicial experience. Judgeship admission exams are highly competitive (the ratio of candidates per position often exceeds 100), not only due to the prestige of the position but also likely because it is among the highest paid in the public sector.[7] Until 2009, federal law did not detail the content or structure of these examinations, which were left to the discretion of State courts. Since then, the structure of exams, including minimum content, qualification thresholds in each phase and weights for final ranking were harmonized.[8]

In practice, nonetheless, the overall structure of these examinations was already rather similar across states. Upon deciding to hire new judges, courts publicly announce the beginning of a *Concurso* through a call for applications, informing how many positions are available and details about the timeline, content and structure of examinations. Potential candidates must enroll online and pay a fee[9] in order to be considered eligible for the position.

Most examinations are comprised of four phases: Multiple Choice, Written, Judicial Decision Writing and Oral Exams. The first phase is often a *Multiple Choice Exam* covering a wide range of topics: constitutional, civil, criminal, commercial, administrative and family law are among the themes covered. Like the other three phases, this exam is both qualifying, meaning that candidates with performance below a certain threshold are immediately eliminated, and classifying, since the grade received is a component of the weighted average that

[5] Constitutional Amendment n.45, December 30th 2004.

[6] Available at `https://www.estadao.com.br/noticias/geral,para-stf-criticas-ao-justica-aberta-sao-infundadas,195051`. Accessed 08/10/2020.

[7] While the Constitution establishes that wages in the public sector should not surpass those of Supreme Court Justices, set at R$ 33,763 (approximately USD 8,500) per month until 2018, the vast majority of judges receive total compensation significantly higher than that due to fringe benefits not included in the above mentioned rule.

[8] National Justice Council Resolution 75 05/12/2009

[9] Resolution 75 determines that the fee can be no greater than 1% of the gross monthly salary for the position, which amounts to around R$ 300 or USD 75.

determines the final ranking of candidates. Those approved in the Multiple Choice phase are invited to take a *Written Examination* that encompass the same topics mentioned before and also topics such as the sociology and philosophy of law, and ethics. The following phase is a *Judicial Decision Writing*, also called a "practical exam", where candidates are given a hypothetical case and asked to write a judicial decision. In most cases this phase includes two decisions, one in criminal and another in civil law. The last qualifying phase is the *Oral Exam*. Candidates are randomly assigned a topic from a pre-determined list 24 hours before their examination, and are then expected to answer questions from a committee composed of other judges and attorneys.

Candidates approved in the *Oral Exam* are eligible to be in the final ranking that defines hiring. Other than the grades in each of the previous phases, the final score also includes the so called *Titles Exam* (*Exame de Títulos*), additional points for career and academic achievements, such as previous judgeship, professorship or advanced degree in Law, and publications in Law journals. Since 2009, the weights that define the final score are the following: 10% Multiple Choice, 30% Written Exam, 30% Judicial Decision Writing, 20% Oral Exam and 10% Title Exam. Candidates are ranked according to their final grades and the top performers are offered jobs according to the number of vacancies available.

It is worth briefly mentioning that these recruitment processes are considered transparent and free from undue influence of judges or politicians, unlike the hiring for other public sector positions which are heavily influenced by patronage practices (Brollo, Forquesato, and Gozzi, 2017, Barbosa and Ferreira, 2019, Colonnelli, Prem, and Teso, 2020). First, every step of the process is highly publicized: grades and lists of approved candidates in each phase are made public, as are the content of each exam. The composition of the committee writing exams and participating in the Oral tests is also made public at the beginning of the *Concurso*, and candidates can appeal for the exclusion of members (e.g. due to family ties of members to any candidate).[10] Second, any deviation from the stipulated rules regarding exams often leads candidates to sue and annul specific phases or even the entire recruitment process. In 2014, for example, candidates in the state of Para successfully sued to have their Oral exams annulled after being asked only three questions during the evaluation, while the call for applications determined four questions.[11] In that sense, the selection process of judges is believed to be broadly free from corruption and reflect the performance of candidates.[12]

3.2.3 Judges' careers and allocation of cases

Once hired, judges are considered "substitute judges" for a period of two years, a probational stage before gaining tenure protection. After this period judges can only be dismissed if convicted of crimes or found guilty of administrative infractions. In practice, this is very

[10] Graders are blind to the identity of exam-takers in the Multiple Choice, Written and Judicial Decision Writing phases. In the Oral exam candidates present in front of a committee and therefore identities are known to graders.

[11] Available at: `http://cnj.jus.br/noticias/cnj/61524-cnj-anula-prova-oral-de-concurso-para-ingresso-na-magistratura-do-tjpa`. Accessed 08/10/2020.

[12] Exceptions do exist. In 2010 the Supreme Court ruled in favor of candidates asking for the annulment of a *Concurso* in the state of Minas Gerais, arguing that more candidates were accepted to the second phase of the process than initially announced. Two daughters of an appellate judge from that state were benefited (Available at: `https://www1.folha.uol.com.br/fsp/poder/po2606201029.htm`. Accessed 08/10/2020)

rare: between 2005 and 2017, only 82 judges in the entire Judicial branch were punished by the National Justice Council, and 53 of those received "mandatory retirement", meaning they were excluded from judgeship but kept receiving salaries.[13]

As previously discussed, judicial districts are divide in three levels: first, second and third. This administrative division is directly linked to judges' careers. Substitute judges are often allocated to first level districts, where they work in general courts, dealing with all types of judicial cases. Promotion means being reallocated to a higher level district, which comes with wage increases. After achieving third level status, judges can be promoted to appellate courts, meaning they leave the first instance (and our database).

The allocation of magistrates to judicial districts is governed by the Constitution. One of the core principles considered is that of the *immovability* of judges, meaning that judges cannot be transferred from their assigned district without their consent.[14] This should make clear that in no way we argue that the movement of judges between courts is quasi-random: judges must assent to being transferred between districts.[15] The identification of judges' fixed-effect, therefore, does not rely on exogenous allocation of judges to courts; our model allows for rich patterns of endogenous matching between judges and courts, and as discussed in detail below only rules out specific types of matches.

Finally, it is important to note that the distribution of cases among judges is as good as random. In judicial districts where there is only one court, cases will be randomly assigned to one judge in that court. For larger districts that encompass specialized courts, cases will be assigned to the proper court depending on their topics or, in the case where more than one relevant court exists, randomly assigned to one of the courts and a judge.[16] That should allay concerns that, within courts, different judges will have distinct composition of cases, making it harder to interpret the number of cases disposed.

3.3 Data and descriptive statistics

3.3.1 Data sources

This paper uses three main data sources: information on monthly output of judges and courts provided by the Open Justice System, admission exam's rankings collected from several different sources and administrative data on formal employment (*RAIS*).

All data on judicial performance come from the Open Justice System (*Sistema Justiça Aberta*), an online platform maintained by the National Justice Council (*Conselho Nacional*

[13]Available at: `https://g1.globo.com/politica/noticia/cnj-puniu-82-juizes-no-brasil-desde-2005-53-deles-continuam-recebendo-salario.ghtml`. Accessed 08/10/2020

[14]The principle is supposed to protect the public against the undue influence of politicians who might want to exclude a judge from judging a case in which they have interest, for example, but it is also a clear benefit to judges who are only reassigned if they so decide.

[15]In our empirical exercises we explore the movement of judges between courts, which can occur in courts within a same district or between different districts. The immovability principle applies to the latter.

[16]The exact method used to implement the random allocation varies by states and is not always fully transparent. It is often described as an electronic platform that randomly distributes cases.

de Justiça – CNJ).[17] The Open Justice System provides monthly information, supplied by courts, on a range of quantitative outcomes at both the court and judge levels, including the number of cases disposed, hearings and intermediary decisions.

We construct a panel at the judge-court-month level: each observation is a vector of quantitative outcomes related to a judge working on a given court in a specific month. The dataset covers the universe of state judges working on first instance courts (i.e. excluding appeal level) from January 2009 through April 2015,[18] and we construct unique IDs using judges' full names to track the movement of magistrates between courts over time.

Our preferred measure of judges' performance is the number of *cases disposed on merits* in a given court and month. This refers to the number of cases for which the judge has issued a final decision based on the merits of the process, i.e., it excludes any cases terminated for procedural reasons or by a decision of one of the parts to withdraw. The decision to exclude cases decided for other reason rather than on the merits is an attempt to reduce the possible noise introduced by considering cases that are concluded for reasons unrelated to the judges' efforts.

Figure 3.1 presents preliminary evidence on the dispersion of judges' output. We plot the histogram of average monthly number of cases disposed at the judge level, across the entire panel. There is remarkable dispersion: judges on the 10th percentile of the distribution dispose of 11 cases on the merits on average, while judges on the 90th percentile dispose of 8 times as many. This dispersion reflects several forces, including potentially judges' efforts and capacity to make the court function efficiently, but also levels of demand in different courts.[19] We will attempt to disentangle these determinants with our empirical model.

The data on admission examinations (*Concursos*) was collected from a variety of sources. Results of *Concursos* are mandated to be public and are often published in PDF format either on the website of the State courts hiring or by the private institutions hired by the state to manage and implement the recruitment process. We scraped these document and constructed a database of candidates' exam performance. We have collected data for 79 recruitment waves for the selection of Judges from 24 different states in the period 2000-2013. For all these examinations the final ranking of approved candidates is available; for a subset of them, we also collect the final grade and the individual grades in all phases of the exam.[20] We then match judges' grades with performance using full names and state of judgeship.[21] We are able to match over 2,800 judges observed in the productivity dataset to

[17]The National Justice Council was created in 2004, through a Constitutional Amendment, with the goals of improving the efficiency and transparency of the Brazilian judiciary. Among other tasks, the Council receives complains from citizens against members of the judiciary, promotes tools to improve the efficient functioning of the courts and publishes data on judicial efficiency.

[18]The Open Justice System was extinguished in 2015, and replaced by a new system later that year. The new dataset, nonetheless, is not strictly comparable to the data we use.

[19]Moreover, it is likely not driven by variation in backlogs across courts because the judicial system as a whole faces large excess demand in cases (Ponticelli and Alencar, 2016).

[20]Recent recruitment processes always include results for all the phases of the examinations. As we go back in time, nonetheless, the information available online becomes scanter. The minimal information we require to include an examination in the dataset is the nominal list of approved candidates and their final rankings.

[21]We benefit from the fact that Brazilians often hold several last names, which makes precise matches on names feasible.

their admission examination performance, covering over 25% of all state judges working at some point between 2009 and 2015.

One additional data source used to recover information from judges' careers is administrative matched employer-employee data from *RAIS* (*Relação Anual de Informações Sociais*) for the period 1995-2017. We use unique individual identification numbers (*CPF – Cadastro de Pessoa Física*) to follow individuals over the years, and then match workers at RAIS to the judge productivity database using full names. We are able to uniquely match approximately 9,400 judges between the two datasets, or 80% of all judges observed in the productivity dataset in the period 2009-2015. We use RAIS data to obtain information on judges' gender, education, formal labor market experience, experience as judges and wages (prior to and during judgeship).

3.3.2 Sample Selection and Descriptive Statistics

The complete productivity dataset comprises close to one million observations at the judge-court-month level. Here we briefly describe the steps to obtain the sample used to estimate the two-way fixed effects model.

Despite the efforts by CNJ to assure quality of the performance data reported, there are clear instances of incorrect entries, such as hundreds of thousands of cases disposed by a single judge in a month. We therefore trim all performance measures at the 99th percentile.[22] We also observe a very high frequency of "mobility" in the raw data, as presented in Column (1) of Table 3.1: on average judges work in 11 different courts throughout the period. Yet, a large proportion of these judge-court matches is clearly transitory: for over half of the judge-court pairs the duration of the match is a single month.[23] In our estimates, we drop any judge-court *spells* with a duration of less than three months. Our final sample includes approximately 730,000 observations.

Table 3.1, Column (1) presents descriptive statistics for the full panel, while Column (2) refers to the sample used to estimate the two-way fixed-effects model.[24] There are 10,479 different judges and 9,048 courts in the estimating sample. Unlike other settings where there is limited mobility explored to estimate two-way fixed-effects models, that is clearly not a problem in our context: almost 80% of judges work in at least two different court throughout the period, and only in about 10% of courts we observe a single judge in the entire period.

The first panel of Table 3.1 characterizes judges in the sample. While the panel covers a 76-month period, the median judge is observed working on any court in 56 months. Very few judges work in one single court throughout these five years: on average judges work in four different courts. While judges might work in more than one court on a given month, that is the exception rather than the rule: for over half of judge-month observations, magistrates are working in a single court. Once we drop short-lived judge-court matches, the average number of months for any match is over 16 months and the median 9 months, meaning that we have several repeated observations of output for each pair, reducing the noise inherent in a measure like the number of cases disposed.

[22]For case disposition, the 99th percentile is 350 cases disposed by a judge in a single month.

[23]Informal conversations with judges suggest that it is common for judges work in different courts when colleagues are on vacation or sick leave.

[24]Table A1 in the appendix presents detailed descriptive statistics for the estimating sample.

Details about courts are presented in panel B of Table 3.1. While in any given month most courts are likely to be staffed by a single judge, their rotation means that, throughout the period, the average number of different judges working in a court is almost five, or one per year. We also present the breakdown of courts by category, according to the type of cases they hear. General courts, located in first level districts and handling all types of cases, comprise around 20% of the sample. The remaining courts are specialized on specific cases, such as Civil (22%), Criminal (16%), Small-stakes (18%) and Family Law (10%). As one might expect, courts dealing with different topics present systematic differences in the number of cases disposed on a monthly basis. Figure 3.2 presents the average monthly number of cases disposed by judges, in each type of court. On one extreme, judges in criminal courts typically dispose of only 20 cases per month, while judges in small-stakes courts, which deal exclusively with less severe criminal cases or low-value civil cases, dispose of almost 50 cases. This highlights why simple comparisons of performance between judges working in different courts might be misleading, and the need to condition on court fixed-effects when estimating judge-level performance.

Descriptive statistics on judicial performance are presented in panel C of Table 3.1. The average number of case disposed on the merit per month is 40, but the distribution has a long right tail (maximum number is 350) and a non-negligible number of zeros: in 13% of judge-court-month observations the number of cases disposed was zero. As discussed below, this motivates our main specification using the inverse hyperbolic sine of cases disposed as the main explanatory variable. The Table also shows that the average number of hearings is 35 (median = 17).

The assessment of the predictive power of admission exams about judge performance relies on a smaller subsample of individuals matched between the two datasets. We present descriptive statistics for that matched sample in Column (3) of Table 3.1. We are able to match 2,881 judges in the productivity sample to their admission exam ranking, or 28% of judges observed in the estimation sample. Judges in the matched sample are observed for less months (45 vs. 50 months in non-matched sample), work in more courts (5.9 vs. 4.3) and have slightly lower monthly output of cases disposed on the merit (36 vs. 40). It is important to note that candidates in the matched sample are not a random sample of the universe of judges. In particular, Figure 3.3 highlights the share of judges we are able to match to recruitment exams by state: whereas in some states like Minas Gerais (MG), Paraná (PR) and Acre (AC) we obtain grades for almost half of judges, three states are not represented at all (Amapá (AP), Rio Grande do Norte (RN) and Rondonia (RO)).

3.4 Empirical strategy and identification

3.4.1 Empirical Model

In order to estimate the permanent component of performance for judges, our main challenge is to separate the individual contribution of judges from the effects of courts they work in: courts in larger district might have inherently more demand, or even within districts there might be systematic differences in length of cases between courts, so we cannot simply compare the performance of judges working in different courts. In order to do that, we borrow from the labor literature and estimate a two-way fixed effects model.

We model the number of cases disposed as follows. For a given judge j working on court c on month-year m, we model (the inverse hyperbolic sine of) the number of cases disposed as:

$$y_{jcm} = \theta_j + \gamma_c + \alpha_s + \boldsymbol{X'_{jcm}}\boldsymbol{\beta} + \epsilon_{jcm} \tag{3.1}$$

where θ_j refers to the permanent component of judge effect; γ_c refers to permanent component of court effect; and X_{jcm} is a vector of time-varying controls. In our baseline specification X_{jcm} includes month-year indicators, the number of courts a Judge work in on a single month and the number of judges working in each single court.[25] Note that we also include an intercept for each connected set, α_s. As previously mentioned, the number of cases disposed is zero in approximately 13% of observations in our dataset. To deal with this, we use the inverse hyperbolic sine transformation (Bellemare and Wichman, 2020) of the number of cases disposed, which, unlike the log transformation, does not drop observations with zero cases disposed.

The separate identification of judge and court fixed-effects in the model above, as shown by Abowd, Creecy, and Kramarz (2002) in the context of workers and firms, is only possible within connected sets – groups of individuals and organizations connected by movers, individuals who work on different organizations throughout the period. Formally, within each connected set g with C_g organizations and J_g individuals, we can identify at most $C_g + J_g - 2$ effects.

The vast majority of judges work in several courts during the period, and even in more than one court in the same month, meaning that connected sets *within states* are very large: in the majority of states the largest connected set comprises over 95% of judge-court-month observations, and only one state it comprises less than 90%.[26] Within each state, we lose very few observations by restricting our sample to the largest connected sets, providing us with 27 connected sets in our estimating sample.

As previously discussed in Section 3.2.2, however, judges are selected to work in a specific state, and never work in courts of different states. That means each state is a separate connected set, and we cannot compare court or judge fixed effects across states. While that is not an impediment to our analysis of the predictive power of admission exams, since we only compare individuals in the same exam cohort (and therefore same connected set), adjustments are needed in order to perform the variance decomposition exercise.

We follow Best, Hjort, and Szakonyi (2019) in estimating the variance components with several connected sets. When estimating equation (3.1), we impose the additional restrictions that both court and judge fixed-effects have mean zero in each connected set. If we define $\tilde{\theta}_j$ and $\tilde{\gamma}_c$ to be the true judge and court fixed-effects, respectively, what we can identify in

[25]Both the number of judges working in a court and the number of courts a judge works on are computed in the full sample, and not in the estimating sample. While we do not use the variation coming from short judge-court matches, our estimates take into account that, for any given month, judges might be "moonlighting" in other courts and thus have lower performance.

[26]In the small state of Sergipe (SE), the largest connected set comprises only 65% of observations. Details about the largest connected set in each State and figures illustrating the construction of connected sets are presented in Appendix B.

equation (3.1) are $\theta_j = \tilde{\theta}_j - \overline{\theta}_g$ and $\gamma_c = \tilde{\gamma}_c - \overline{\gamma}_g$, the deviations of the true effects from the connected set means. We can then write the variance of number of cases disposed as:

$$\begin{aligned} Var(y_{jcm}) =& Var(\theta_j) + Var(\gamma_c) + 2Cov(\theta_j, \gamma_c) + Var(\alpha_s) + Var(\boldsymbol{X'_{jcm}\beta})+ \\ & 2Cov(\alpha_s, \boldsymbol{X'_{jcm}\beta}) + 2Cov(\theta_j + \gamma_c, \alpha_s + \boldsymbol{X'_{jcm}\beta}) + Var(\epsilon_{jcm}) \end{aligned} \tag{3.2}$$

Best, Hjort, and Szakonyi (2019) show that, since we can only estimate within connected sets variances, the estimates recovered are lower bounds of the total variance of both judges and courts fixed-effects. The total variance attributable jointly to judges and courts, nonetheless, can be recovered using the variance of the connected sets effects: $Var(\tilde{\theta}_j + \tilde{\gamma}_c) = Var(\theta_j + \gamma_c) + Var(\alpha_s)$.

3.4.2 Identification and estimation

As discussed in detail in Card, Heining, and Kline (2013), Card, Cardoso, and Kline (2016) and Card, Cardoso, Heining, and Kline (2017), identification in the two-way fixed-effects model does not require random allocation of workers (judges) across firms (courts). The structure of the model allows for rich patterns of sorting, including for judges that dispose of more cases to select into better courts, or for judges to specialize in certain courts where their output is higher. In other words, our identification assumption of exogenous mobility is that judges do not sort on the error term in Equation (3.1).

Here we focus on assessing whether two particular issues affect the identification of our model. First, we model judge and court fixed-effects as additive and linearly separable. If that is not the case and there exists a judge-court match effect (i.e. more productive judges are particularly efficient in productive courts), then our estimates of judge effect might be biased. Figure 3.5, panel A, presents a heatmap where we break down residuals of our model by vingtiles of judge and court fixed effects, and graph the average residuals in each cell. To interpret these results, consider Figure 3.5, panel B, where we simulate a model in which there exists judge-court match effects, but we erroneously estimate a linearly separable mode. The residuals then are systematically large/small in cells with extreme fixed-effects, reflecting the incapacity of the model to capture the matching effect. Going back to panel A, the actual heatmap, we do not observe the same pronounced pattern as in the simulation, suggesting that even if match effects are real (our model seems to be unable to match the outcomes at the very top cell in terms of both judge and court fixed-effects), they are not large enough to severely affect our estimates.

The second issue we consider is whether judges are moving into courts systematically due to *trends in court productivity*. While the selection of judges into courts due to levels of productivity does not affect our estimates, the same is not true if judges can select into courts because they are improving/decreasing their performance. To consider whether that seems to be the case, we perform an event study that assess how the number of cases disposed by judges evolve around the time judges make clear transitions between judicial districts (i.e. judges working in a given court for at least three months prior to transition and at least three months after).[27] Figure 3.7 reports the coefficients of the event-study, in which we consider the indicator for 6 months before the transition as the omitted category. Three things stand out from these results. First, productivity starts falling in the last two months before a judge moves: knowing they will change courts, they might put in less effort to dispose of more cases or transfer their cases to other magistrates. Second, the fall in performance persists for at least three to fourth months after the transition, but six months after there is no distinguishable effect on performance. Finally, and most important for the model, there seems to be no selection in trends: judges do not seem to be on a trend to be more or less productive, either before or after the movement between judicial districts. These results suggest that selection on trends do not seem to be a threat to identification in this context.[28]

Consistent estimation of individual fixed-effects require not only that the number of observations in a panel is large enough, but also that the number of periods in the panel grows to infinity. Since our dataset encompass around 70 months, finite sample bias will lead to excess dispersion in our estimates of both judge and court fixed-effects, inflating the estimated share of total variance explained (Best, Hjort, and Szakonyi, 2019, Silver, 2019). We deal with that issue by using a non-parametric, split-sample correction method that shrink our variance estimates (Finkelstein, Gentzkow, and Williams, 2016).

We randomly split our sample in two, stratifying at the judge-court level, so that we preserve the number of judge-court pairs in both samples. We then proceed to estimate the two-way fixed effect model separately in each sample and obtain separate judge and court fixed-effect estimates. While FEs are noisily estimated in each sample, the errors should be uncorrelated due to the random split. Formally, if in each sample $s = \{1, 2\}$ the estimated judge fixed effect can be written as $\hat{\theta}_{(j,s)} = \theta_j + e_{j,s}$, where θ_j is the true FE for individual j and $e_{j,s}$ the error term, with $Cov(e_{(j,1)}, e_{(j,2)}) = 0$, then it holds that $Cov(\hat{\theta}_{(j,1)}, \hat{\theta}_{(j,2)}) = Cov(\theta_j, \theta_j) = Var(\theta_j)$. That is, we can recover the true variance of FEs by separately estimating variances in the random samples and calculating their covariance.

3.5 Results

3.5.1 Judges role in explaining variation in output

Before presenting the results decomposing the variance of total output, we present preliminary evidence that judge fixed-effects matter in explaining courts' output. Table 3.2, Columns (1) and (2), present goodness-of-fit measures when estimating Equation (1) excluding and including judge fixed-effects, respectively. The inclusion of judge fixed-effects increases the adjusted R-squared of the model by 8 p.p. and reduces the residual standard error (RSE) from 1.43 to 1.34. This is evidence that judges matter in explaining the variation in output observed across courts.

[27]It is much harder to create such event-study when judges start working in different courts in the same district, because they often do not clearly leave one court for another, but keep a "connection" to their old appointment. For that reason we restrict our analysis to clear changes of court when judges move from one district to another.

[28]The Figure also shows confidence intervals growing in width after the transition. This happens because we only require judges to reappear in the sample post-transition three times.

We present the results of formal variance decomposition in Table 3.3.[29] Column (1) presents the raw variance estimates, with no finite-sample corrections, while Column (2) present corrected variance estimates using split-sample strategy, and Column (3) presents the share of total variance explained by each component using the split-sample estimates. The finite-sample corrected variance of judges' FE is very similar to the raw estimates, on the range of 0.74-0.80, suggesting that judges explain at least 23% of the total variance of output. To put that magnitude in context, it is significantly larger than the estimate of Fenizia (2020) on the share of social security offices' productivity in Italy explained by managers (9%), but very similar to those of Best, Hjort, and Szakonyi (2019) on the share of public procurement prices explained by procurement officers in Russia. The estimates for share of total variance explained by courts fixed effects, on the other hand, is more sensitive to the correction method, varying from 34-48%. Estimates for the variance explained by the sum of judge and court FEs range from 35-51%: since the sum of explained variance independently explained by judge and courts is much larger than that, it means the covariance of these fixed effects is large and negative, meaning that judges with higher FE are observed matched with courts of low FE, and vice-versa.

While the previous estimates show that judges are important in explaining the quantity of cases disposed and provided individual measures of judge performance, one might worry that judges that dispose of more cases are prioritizing quantity over quality. If that is the case, judges with higher fixed-effects in our model might actually be those that cut back on the inputs necessary to arrive at "good decisions", hastening the process to increase their case disposition number. We test whether this is a plausible explanation in our context by investigating one important input for case decision: the number of hearings that judges hold each month. To assess if "high fixed-effect" judges are conducting systematically less hearings than their peers with lower fixed-effects, we follow Silver (2019) and re-estimate the two-way fixed-effects model using the number of hearings as dependent variable, thus obtaining a new fixed-effect estimate for each judge. If judges are severely cutting back on hearings in order to increase their case disposition, we might expect a weak or even negative correlation between the fixed-effects in both models. Figure 3.6 shows that this is not the case: fixed-effects from the two models are strongly positive correlated, suggesting that judges who dispose of more cases are also those that hold more hearings. While we are not able to assess whether the use of other inputs, including length or quality of hearings, this alleviates concerns that judges who dispose of more cases are systematically sacrificing on quality.

3.5.2 Correlates of judges' and courts' fixed-effects

In this Section we briefly describe whether estimated courts' and judges' fixed-effects are systematically correlated with observable characteristics. We start by presenting results for courts' FEs in Table 3.4. The first panel shows that courts' have higher fixed effects when located in judicial districts outside the state capital, with larger populations and higher urbanization rates. Conditional on time and judges' fixed-effects, this suggests that the

[29]Due to the high dimensionality of fixed-effects, we cannot simply invert matrices to obtain OLS estimates. We then estimate the parameters using the *-lfe-* command in R, also used by Best, Hjort, and Szakonyi (2019).

number of cases disposed is particularly high in poorer, large urban districts outside the largest urban center of states. There are several possible explanations for that finding. If relative demand for judicial services is higher in these poorer areas, relative to supply, courts in those areas might present higher case disposition, possibly in detriment of decision quality. It is also possible that the composition of cases in these areas are different, and the higher number of cases in poorer areas reflect the fact that cases are easier to dispose. All those factors might co-exist, and will be picked up by courts' fixed-effects in our model. The results in Table 3.4 also shed light on how fixed-effects differ by the nature of cases assigned to each court. Similarly to what we observed in the simple descriptive statistics of Figure 3.2, criminal courts and those dealing with other topics such as commercial law (pooled with "others" here) have particular low level of case disposition when compared to general courts.

We now turn to describe how judges' fixed-effects correlate with observable characteristics. Here we rely on the sample matched to RAIS, the employer-employee database of formal workers, in order to construct judges' work history and obtain individual traits such as gender and age. Results are presented in Table 3.5. In Column (1) we present results for all judges that are matched to RAIS, and in Column (2) we restrict to judges that are observed at least once working outside of the judiciary, in order to include wages prior to judgeship as a correlate. All estimates include connected-sets (State) fixed-effects. Results in Column (1) suggest that individual traits explain very little of the estimated effects: gender, education and experience, both in general and in the judiciary, are not significant predictors of judge fixed-effects. Age is correlated with the estimated effect, with a positive and concave relationship: older judges dispose of more cases, but the effect is diminishing in age. These results, however, are not very robust: when we restrict the sample to those observed working outside the judiciary since 1995, we no longer observe age as a significant predictor, but overall experience does seem positively correlated with case disposition. The coefficient on (log) average yearly wage received before joining the judiciary, which we interpret as potential earnings outside of judgeship, is small in magnitude and not statistically different from zero.

3.5.3 Entrance exams are predictive of performance

Results in the previous sections are strong evidence that the identity of judges matters for the timely delivery of justice. While we are unable to explain the reasons why some judges are more effective in disposing of cases than others, the fact that we observe such differences in judge output suggests that the screening of judges might be one tool in improving judicial efficiency. We now turn to the question of how candidates performance in the admission exams is related to their performance on the job. In all the exercises that follow we use the sample for which we can match judges' admission exam performance.

We start by presenting "reduced-form" evidence that entrance exam ranks are correlated with the number of cases disposed on merits, once we control for court and month fixed-effects. That is, in here we do not use estimated judges fixed-effects, but simply present OLS regressions of the form

$$y_{jcm} = \boldsymbol{\beta' ExamRankQuintile_j} + \gamma_c + \delta_{w(j)} + \boldsymbol{X'_{jcm}\theta} + \epsilon_{jcm} \tag{3.3}$$

where y_{jcm} is the IHS transformation of cases disposed, $ExamRankQuintile_j$ are indicators for quintile of exam performance of judge j in their exam cohort and δ_w are indicators for each cohort of candidates, since we can only meaningfully compare ranking among candidates sitting the same examination. Standard-errors are clustered at the judge-level.

Results are presented in Table 3.6, where the omitted category for exam quintile is the bottom 20%. Column (1) presents estimates for a regression that only includes cohort fixed-effects, while in Columns (2) and (3) we add Court and Month fixed-effects, respectively. Focusing on Column (3), the results suggest that, when compared to judges ranking in the bottom quintile of their cohorts, those in the top 20% dispose of approximately 21% more cases. The estimated effect is smaller but statistically significant and economically meaningful for judges with ranks in the second to fourth quintiles, and we can reject that the coefficient for the top 20% is identical to those on the second and third quintiles. In Column (4) we present a much more stringent exercise: we include court-by-month fixed effects, meaning that the only variation used comes from different judges working in the same court on the same month (hence the large drop in sample size, since observations for courts with a single judge in any given month are dropped). The estimated coefficients are slightly larger in absolute value, but broadly consistent with previous estimates suggesting that better ranking in entrance exams are correlated with higher case disposition on the job.[30]

We now present results using the estimated judges' fixed-effects obtained in the previous section. Figure 3.4 presents non-parametric evidence of the correlation between (residualized) ranks in admission exams and standardized FE.[31] The strong positive correlation between performance measures suggests that judges who perform well in the admission exams are also among the ones with highest FE in their cohorts.

In Table 3.7 we present this same evidence in regression form. We estimate simple OLS regressions at the judge-level, using measures of on-the-job performance (standardized fixed-effect) as dependent variables and quintiles of performance in the recruitment exam as the main explanatory variable. Column (1) presents results from an OLS regression including cohort fixed-effects. Consistent with the findings in the reduced form regression, our results suggest that being ranked in the top 20% in the admission exam is correlated with a 0.2 s.d. increase in judge's performance (estimated fixed-effects) in comparison to those in the bottom quintile. Those ranking in lower quintiles are also estimated to perform 0.1-0.15 s.d. higher when compared to those at the very bottom. In Column (2) we replace the quintile ranking in the admission exam with the standardized final grade used to construct ranking.[32] The coefficient on grade is significant and indicates that an increase of 1 standard deviation in final grade is correlated with a 0.07 s.d. increase in performance (measured by judges' fixed-effects). Taken together with the results from the reduced-form model, this suggests that, among the candidates selected in the admissions exam, those that rank higher

[30] Table A3 shows that the results are robust to using the rankings on admission exams as explanatory variables instead of ranking quintiles. Since lower number rankings indicate better performance, the negative coefficients indicate that better ranked judges have higher output.

[31] Since we only compare judges entering in the same cohort, we first regress each rank on cohort indicators and use residuals to construct the binned scatter plot

[32] We could not collect final grades for some of the cohorts, therefore the smaller sample size in Column (2)

do perform better on the job than those ranking lower.

While we believe documenting that the overall ranking is informative about job performance is an important result, it does not shed light on exactly which dimension of the screening process is leading to this positive correlation. It is possible, for example, that the Titles Exam, that takes into account previous work and academic accomplishments, is the most predictive component of the overall ranking. Or that the Oral Exam, in which there exists some degree of discretion by the selection committee, would be more informative. We attempt to provide evidence on that question by re-estimating the previous equation using grades in each exam phase as dependent variables and assessing which of those are more predictive about job performance.[33] As previously discussed, we restrict our analysis to 20 examinations and 619 judges for whom we observe six separate grades: Objective Exam, Written Exam, Civil and Criminal case decisions, Oral Exams and "Titles" Exam.

Results are reported in Columns (3) and (4) of Table 3.7. We first report in Column (3) the results of estimating the equivalent equation of Column (2) in the subsample for which we have detailed grade information. The result is very similar to that obtained in the full sample, suggesting that candidates with higher final grades also perform better on the job. We then estimate the model including standardized grades in each of the exams separately, and report results in Column (4). The only coefficient that is significant, and also the largest in magnitude, is that of grades on the Judicial Decision Writing on civil cases (the coefficient on criminal cases is less than half the size in magnitude and not statistically different from zero). As shown below, this result is robust to other specifications of the estimation equation, suggesting that the civil case admission is indeed the most predictive component of the admission exam.

Recall that, since the harmonization of admission exams in 2009, each of the Judicial Decision Writing exams has weight 15% for the final ranking, so grades in the civil case decision contribute less to the final selection than grades in the written exam (30% weight) or Oral exam (20%). Our results suggest, in contrast, that if the goal is to select candidates who will increase the speed of case disposition, exams should overweight results in the civil case decision.

3.5.4 Results are robust to alternative specifications

We conduct several exercises to assess the robustness of our results. First, Table 3.8 presents regressions in which we drop top and bottom performers in each cohort, evaluating whether results are fully driven by the very best (or very worst) candidates. Column (1) reproduces our main specification, while the remaining Columns restrict the sample by dropping only the top 3 performers in each cohort (Column 2); the top 5% candidates in each cohort (Column 3); the bottom 5% candidates in each cohort (Column 4); and both the top and bottom 5% candidates in each cohort (Column 5). Estimates of the correlation between exam rank and FE are very stable, and we cannot reject they are statistically indistinguishable from our main specification.

In Table 3.9 we re-estimate the results of our main specification but use the rank of

[33] In Table A2 we present pairwise correlations between grades in each of the admission exams phases. Correlations are often insignificant and not very large, suggesting that performing well in one particular phase of the admission process is generally not a good predictor of performance in other phases.

judges' FE as dependent variable instead of the standardized FE. Column (1) presents the coefficients on admission exam quintiles: among judges entering in the same cohort, those in the top quintile rank, on average, four positions higher than those in the bottom 20% (coefficients are negative since a better rank equals a lower rank number). Those in the second quintile rank 2.6 positions higher, on average, and those in the third and fourth quintile between 1.3-2 positions. Overall, the results confirm our main specification findings that candidates ranking better in the admission exam also perform better on the job. Column (2) uses the final grade as explanatory variable, showing that a unit s.d. increase in final grade is correlated with a 1.5 higher position in FE ranking. When we restrict the sample to those observations with detailed grades, in Column (3), the coefficient on Final Grade is very similar in magnitude to that on the full sample. Finally, when we include separate grades by phases as predictors of judge FE we again find that the largest coefficient in magnitude and significant is that associated with civil case exam: an increase in 1 s.d. in the civil case grade is correlated with a 1.4 better ranking in performance.

We also perform a randomization inference exercise to assess the robustness of our findings of the positive correlation between admission exam and job performance ranks. Within each cohort of judges, we randomly assign exam rankings, re-compute quintiles and then estimate the baseline model presented in Column (1) of Table 3.7. Figure 3.8 presents the histogram of these 1,000 simulated beta-coefficients for the top 20% performance indicator, and the solid line marks the true coefficient of 0.227. 95% of estimated coefficients are on the interval [-.118, 0.118], and none of the estimates is larger in magnitude than the true estimate. These results suggest that it is very unlikely that we would obtain a coefficient of this magnitude purely by chance.

3.6 Conclusion

This paper provides evidence that states can effectively design impersonal exams that are able to screen good candidates for top public service positions, even when recruitment practices are constrained by fears of political influence. We explore rich data on judges and courts in Brazil to show that judges are relevant in explaining the observed variation in output, and estimate judge-level measures of performance in case disposition – an important indicator in a judicial system with high levels of court congestion. We then link these measures to judges' performance in the admission exams, and show that within cohorts of hired judges those with higher grades also dispose of more cases. In particular, it seems that not all phases of the admission exams are equally likely to predict job performance: across different specifications, grades on the civil case exam is the only statistically significant predictor.

Our results have meaningful implications for policy makers. First, it adds to recent evidence that not only frontline providers matter for the delivery of public service: managers and other officials working across the state bureaucracy can have significant impact on service provision (Best, Hjort, and Szakonyi, 2019, Fenizia, 2020, Aman-Rana, 2020). Carefully designing systems that select and incentivize these individuals is therefore very important. Secondly, it is also relevant for the debate about rules and discretion in hiring (Hoffman, Kahn, and Li, 2018). We show that an admission process with little discretion by the selecting agency is able to rank individuals in a way that meaningfully predicts job performance. In particular, by breaking down exam performance into its components, we find evidence that an examination that approximates the kind of task faced by candidates on the job (the writing of sentences by judges) is especially predictive about their future performance.

Data limitations do not allow us to further explore three mechanisms we believe are relevant for future research. The first is what makes for an efficient judge. Judges do not work in isolation writing decisions, but, rather, manage complex organizations staffed by several workers and in close contact with other state actors (Pinheiro, 2003, Oliveira Gomes, 2014). A more efficient judge might be one that simply puts longer hours and more effort to increase case disposition, but might just as well be one that is able to put in place an well-oiled machine where every staffer is pulling their weight and ensuring smooth handling of cases.[34] Management practices have shown to be very relevant in explaining productivity in both the private (Bloom and Van Reenen, 2007, Bloom, Eifert, Mahajan, McKenzie, and Roberts, 2013) and the public sector (Rasul and Rogger, 2018, Leaver, Lemos, and Dillenburg Scur, 2019, Bloom, Propper, Seiler, and Van Reenen, 2015), so gaining better understanding of working practices in the judicial sector might shed light on the determinants of judge effectiveness.

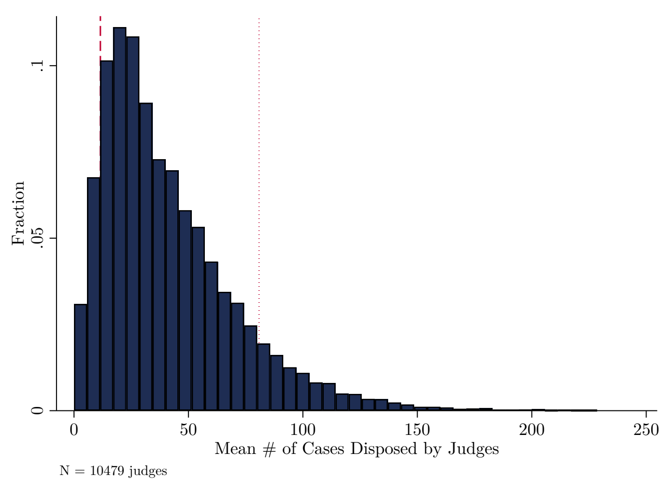
Second, while we find a strong and robust positive correlation between grades in the admission test and performance, and consider this a relevant parameter for policy-makers designing screening processes, it is unclear exactly what is the force driving this correlation. One possibility is that exams are indeed effective in screening candidates with specific knowledge that is also useful for the tasks performed by a judge – the fact that grades in the civil case examination are the only ones with independent predictive power suggest this might be the case. Another possibility, however, is that competitiveness and difficulty of the exams screen candidates with high general ability and/or high motivation to be a judge, which implies that the congruence between test content and requirements of the job is less important. We think this is an relevant distinction, particularly in light of the theory and evidence that highlight the role of intrinsic motivation in driving performance when high-powered incentives are limited (Deserranno, 2019, Ashraf and Bandiera, 2018, Prendergast, 2008).

Finally, given that we only observe judicial productivity for those who pass entrance examinations, we cannot make claims about the remaining pool of candidates. In particular, we cannot directly test if examinations are screening for the most productive candidates overall or not, or if exams should be made more selective or less. Future research with complete characteristics of candidates, their career paths and more examinations could evaluate the overall efficiency of the Brazilian selection mechanism into the public sector.

[34] Fenizia (2020) finds that the mechanism through which managers in social security offices are able to increase output per worker is by letting go of workers while maintaining total output stable.

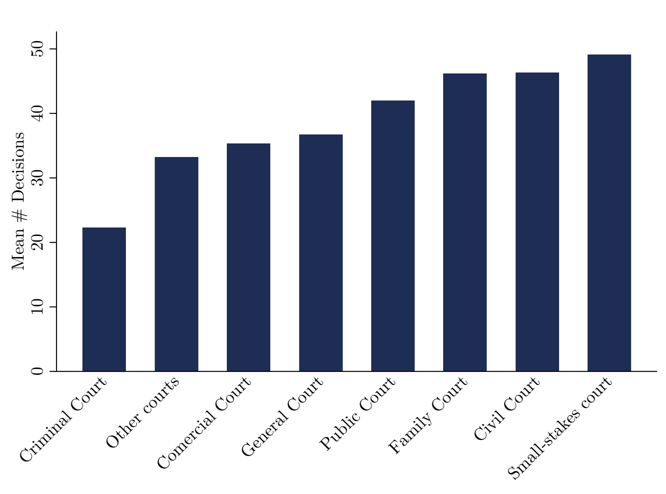
Figures

Figure 3.1: Histogram of mean number of cases disposed on the merits by judge.



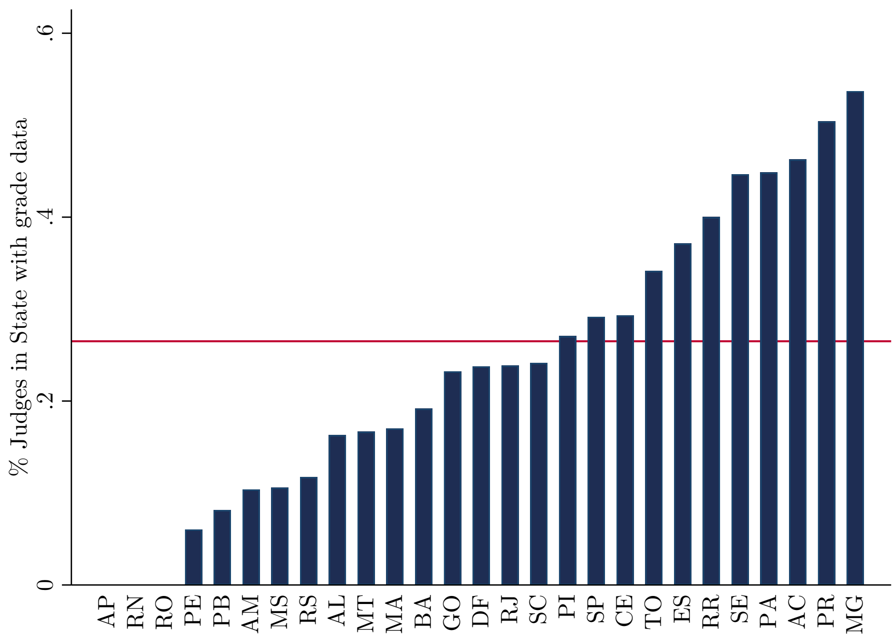
Note: The histogram presents the average monthly number of cases disposed by judges. Average number of cases is calculated in the sample used to estimate the two-way fixed-effects model, where outcome variables are trimmed at the top 1%, judge-court spells shorter than three months are dropped and only observations in the largest connected sets within each state are used. The dashed and dotted lines mark the 10th and 90th percentile of the distribution, respectively.

Figure 3.2: Average number of cases disposed, by type of court



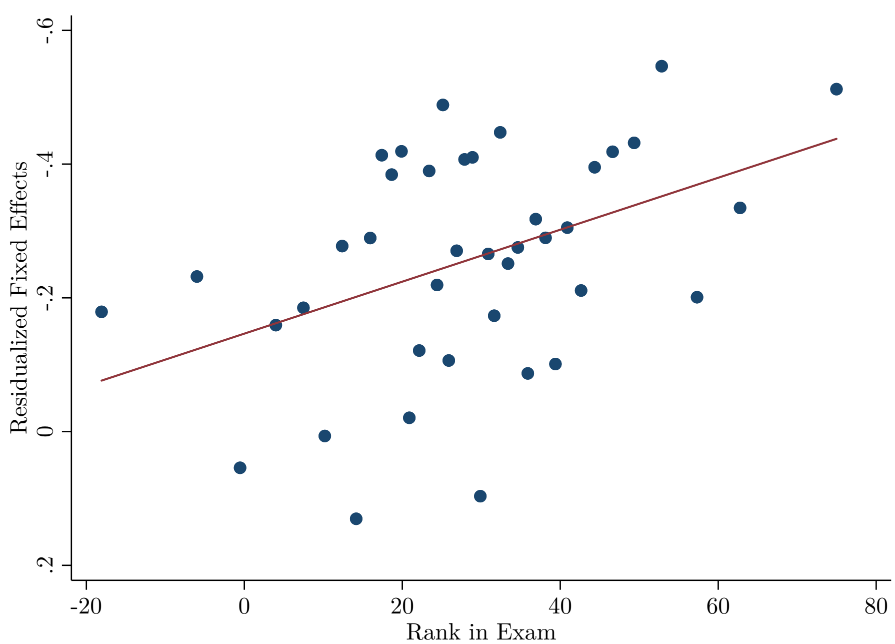
Note: This figure presents the average monthly number of cases disposed by judges, in each type of court. Number of cases is calculated in the sample used to estimate the two-way fixed-effects model, where outcome variables are trimmed at the top 1%, judge-court spells shorter than three months are dropped and only observations in the largest connected sets within each state are used.

Figure 3.3: Share of judges matched by State.



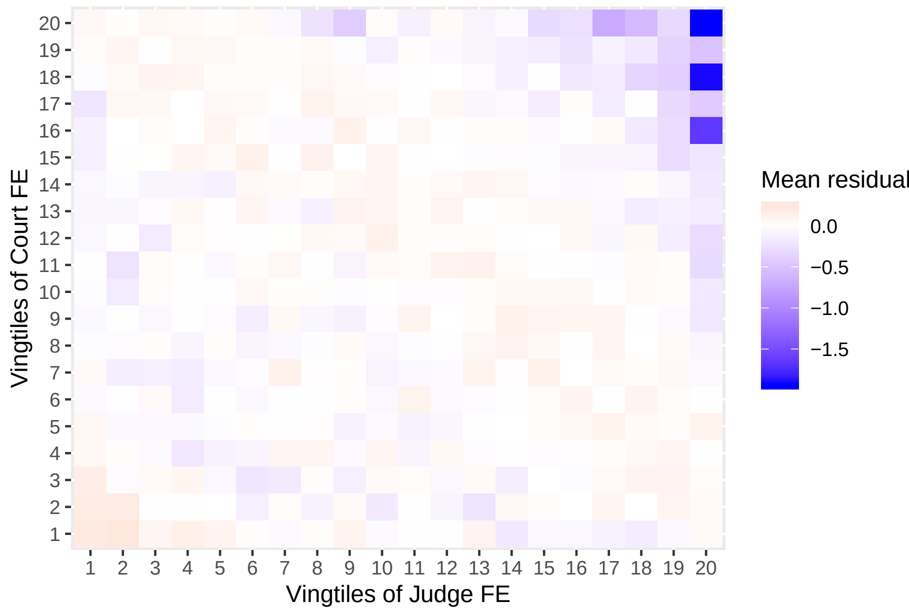
Note: This figure presents the share of judge in the estimation sample that are matched to their admission exam, by State. The red line mark the overall share of judges matched (28%).

Figure 3.4: Binscatter of residual ranks, conditioning on Concurso FE

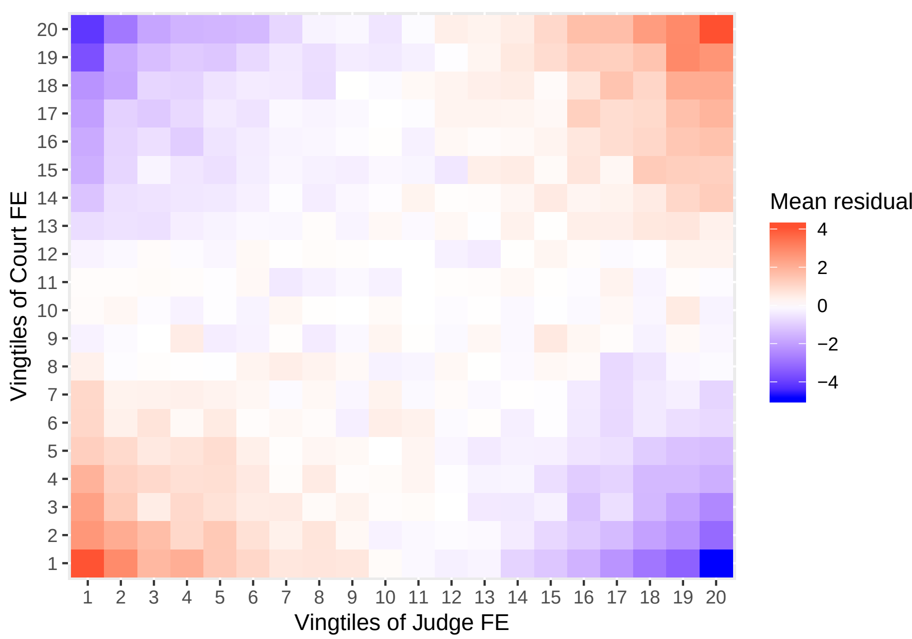


Note: The graph presents a binned scatter plot of residualized rank in fixed-effects obtained by estimating Equation (1) and residualized ranks in admission exams, at the judge level. Residues are obtained by regressing each of the variables on *Concurso* fixed effects.

Figure 3.5: Residuals heatmap from two-way fixed-effects model



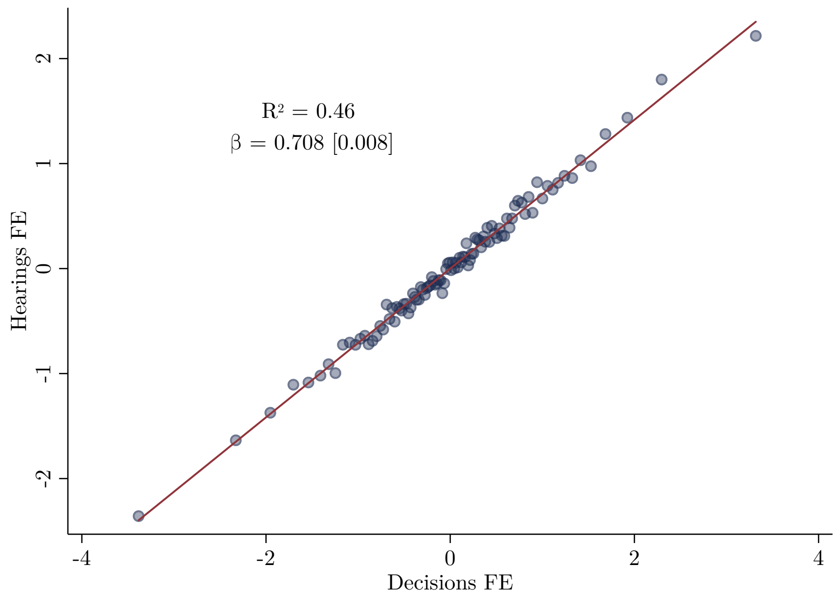
(a) Actual from data



(b) Simulated from misspecified model

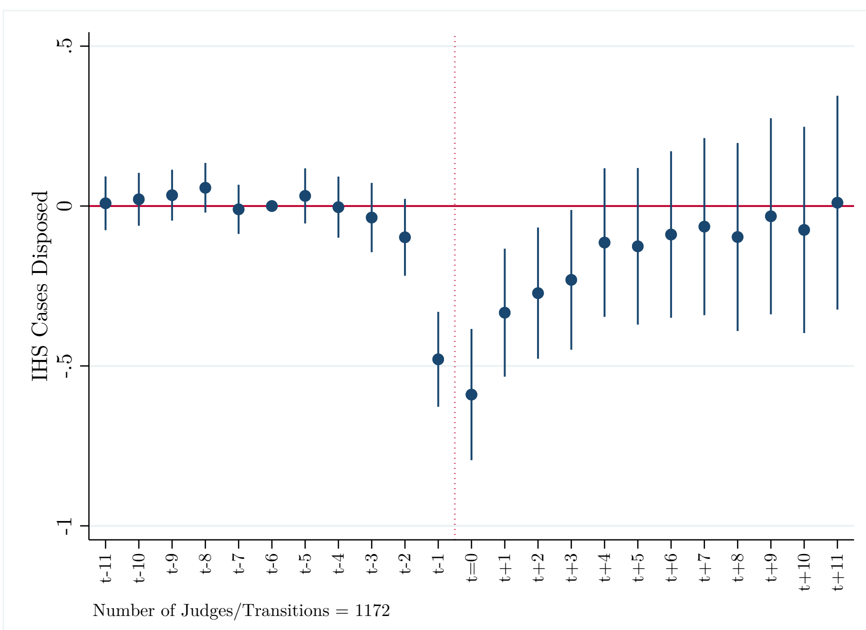
Note: These figures present heatmaps of average residuals from a two-way fixed-effects model. Panel A presents results from actual data estimated using equation (1) while Panel B presents results from a simulated model that contains match effects between judges and courts, but is estimated using a misspecified model in equation (1). Darker blue cells represent large negative residuals, while darker red cells represent large positive residuals. Judges and courts are binned into vingtiles of estimated fixed-effects

Figure 3.6: Binscatter of hearings and case disposition fixed-effects



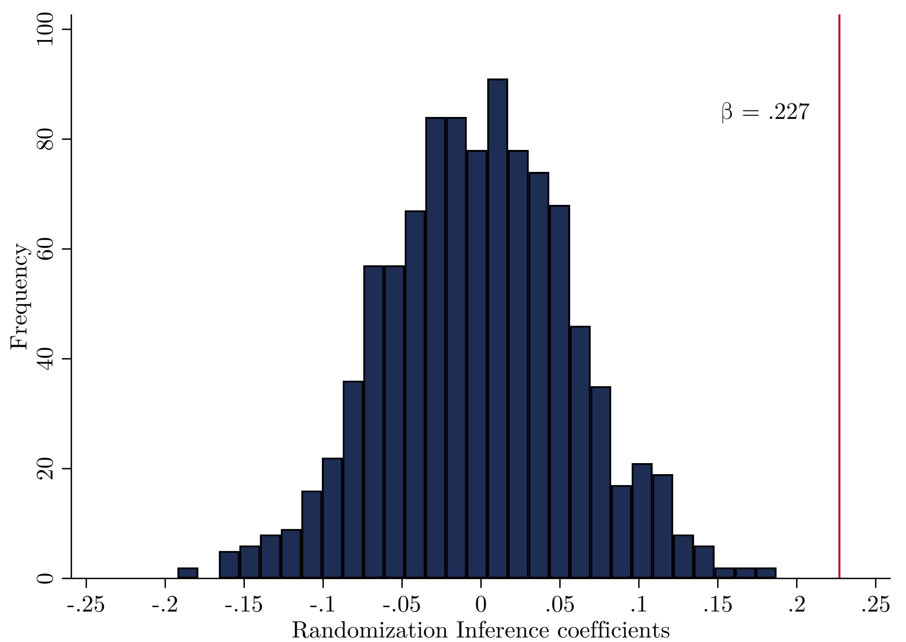
Note: This graph presents a binned scatter plot of residualized judge fixed-effects obtained by estimating equation (1) using hearings and case disposition separately. Residuals are obtained by regressing both FEs on connected set dummies. The R-squared and coefficients presented refer to a regression of hearing FE on case disposition FE including connected set dummies.

Figure 3.7: Event-study around judicial district movement



Note: This figure reports point estimates and 95% CI for coefficients on an event-study regression, where the dependent variable is the IHS of cases disposed and the omitted category is the indicator for six months before the movement. The sample is restricted only to clear moves across judicial districts, as detailed in Section 3.4.2. Standard errors are clustered at the transition-level.

Figure 3.8: Histogram of simulated beta-coefficients



Note: The figure presents the histogram of 1,000 simulated coefficients for the top 20% indicator using our main specification, equivalent to the results presented in Column (1) of Table 3.7, where we randomly assign final admissions ranking within each cohort. The true coefficient is marked by the solid red line

Tables

Table 3.1: Descriptive statistics

	Full Sample	Estimation Sample	Exam matched Sample
Judges			
Share male judges	0.61	0.60	0.61
Mean # courts by judge	16.56	5.54	7.71
Mean # months by judge	61.13	59.09	55.91
Mean # courts at judge-month level	2.85	1.89	2.17
Mean # judicial districts at judge-month level	5.13	2.68	3.94
Mean # months per judge-court pair	30.08	34.35	23.34
Courts			
Mean # of judges by court	19.11	6.36	4.43
Mean # judges at court-month level	2.81	1.92	1.49
Share civil courts	0.22	0.22	0.23
Share general courts	0.20	0.20	0.24
Share small-stakes courts	0.18	0.18	0.16
Share criminal courts	0.16	0.16	0.16
Share family court	0.10	0.10	0.09
Share other courts	0.14	0.13	0.11
Output measures			
Cases Disposed (on merit)	33.82	40.13	36.10
Total Hearings (presided or held)	29.32	34.88	35.85
Number of judges	991,324	727,784	200,212
Number judges ever working in multiple courts	962,395	653,559	196,094
Number of courts	991,324	727,784	200,212
Number of courts with multiple judges	979,015	691,916	176,111
Number of judge-court pairs	991,324	727,784	200,212
Number of judge-court spells	273,074	77,799	24,089
Number of connected sets	68	27	24
Number of judge-court-month observations	991,324	727,784	200,212

Note: This table reports descriptive statistics for key variables. Column (1) refers to the full original panel. Column (2) refers to the sample used to estimate the two-way fixed-effects model, where outcome variables are trimmed at the top 1%, judge-court spells shorter than three months are dropped and only observations in the largest connected sets within each state are used. Column (3) refers to the sample matched to admission exams, i.e., it only retains judge-court-month observations for which judges were matched to their admission exams ranking. This is the sample used in both the "reduced-form" exercises presented in Table 3.6 and the main results on the correlation between admission ranking and performance in Table 3.7.

Table 3.2: Goodness of fit measures

	(1)	(2)	(3)
R-squared	0.379	0.464	0.619
Adjusted R-squared	0.371	0.45	0.593
Residual Standard Error (RSE)	1.434	1.341	1.153
Observations	727784	727784	727784
Judge FE	No	Yes	No
Judge-by-Court FE	No	No	Yes

Note: This table presents goodness-of-fit measures for several different models using the two-way fixed-effects estimation sample. Column (1) presents results from a model that does not include judge fixed-effects; Column (2) is our main specification from equation (1), including judge fixed-effects; while Column (3) includes judge-by-court fixed effects.

Table 3.3: Variance decomposition

	Raw Variance	Split Sample Variance	Split sample Var - % Total
Cases disposed (IHS)	3.27	3.27	1.00
Judge FE	0.80	0.74	0.23
Court FE	1.16	1.10	0.34
Connected Set FE	0.24	0.22	0.07
Judge+Court FE	1.23	1.14	0.35

Note: This table presents the variance decomposition exercise using estimates from the two-way fixed effects model in equation (1). Column (1) presents the variance estimates without adjustment, while Column (2) presents variance estimates corrected for finite-sample bias using the spit-sample technique. Column (3) presents the finite-sample corrected variance estimates as a share of total variance.

Table 3.4: Correlation between courts' fixed-effects and courts' characteristics

	(1)
Judicial district characteristics	
State Capital	-0.0953
	(0.0456)
Log population (2010)	-0.00821
	(0.0154)
Log GDP per capita (2016)	-0.0748
	(0.0285)
Share urban households (2010)	0.0966
	(0.0984)
Second level	0.144
	(0.0414)
Third level	-0.0663
	(0.0524)
Special level	-0.369
	(0.0835)
Type of courts	
Criminal court	0.0343
	(0.0458)
Civil court	-0.390
	(0.0451)
Family court	0.209
	(0.0515)
Small-stakes court	-0.154
	(0.0438)
Other courts	-0.816
	(0.0514)
Observations	9,047
R-Squared	0.114
Number Connected Sets	27
CS fixed-effect?	Yes

Note: This table reports regressions using the estimated courts' FE (standardized to have unit standard deviation within connected sets) as dependent variable. State capital is a dummy variable indicating whether the judicial district where the court is located is a state's capital; Log population is from the 2010 Census and Log GDP per capita is from the 2016 national accounts published by the Brazilian Institute of Geography and Statistics (IBGE). Robust standard errors in parentheses

Table 3.5: Correlation between judges' fixed-effects and individual characteristics

	(1)	(2)
Male	-0.00721	-0.0224
	(0.0220)	(0.0341)
Age in 2015	0.0516	0.0232
	(0.0122)	(0.0203)
Age (squared)	-0.000537	-0.000272
	(0.000119)	(0.000203)
Graduate degree	0.0754	0.121
	(0.0433)	(0.0602)
Formal labor experience in 2015	-0.000124	0.0540
	(0.0142)	(0.0312)
Formal experience (squared)	-0.000161	-0.00194
	(0.000526)	(0.00109)
Formal judicial experience in 2015	-0.00262	-0.00140
	(0.0113)	(0.0161)
Judicial experience (squared)	0.000964	0.00103
	(0.000466)	(0.000743)
Formal experience outside judicial sector	-0.0119	
	(0.0344)	
Log average wage before judiciary (2017 prices)		0.0175
		(0.0158)
Observations	8,597	2,827
R-Squared	0.019	0.046
Number Connected Sets	26	26
CS fixed-effect?	Yes	Yes

Note: This table reports regressions using the estimated judges' FE (standardized to have unit standard deviation within connected sets) as dependent variable. Independent variables are obtained from matching judges' in performance dataset to RAIS, a matched employer-employee administrative dataset. Data from RAIS covers the period 1995-2017, so measures of experience in the formal sector and in the judiciary in 2015 are capped at 20 years. Robust standard errors in parentheses.

Table 3.6: Reduced form regressions: output and admission exam performance

	(1)	(2)	(3)	(4)
Top quintile (β_1)	0.113	0.216	0.207	0.334
	(0.0546)	(0.0403)	(0.0402)	(0.0820)
4th quintile (β_2)	0.102	0.180	0.176	0.199
	(0.0508)	(0.0388)	(0.0387)	(0.0809)
3rd quintile (β_3)	0.0854	0.137	0.133	0.242
	(0.0503)	(0.0371)	(0.0368)	(0.0795)
2nd quintile (β_4)	0.0920	0.143	0.141	0.246
	(0.0511)	(0.0375)	(0.0373)	(0.0783)
Observations	200,206	200,206	200,206	59,795
R-Squared	0.11	0.42	0.43	0.53
Concurso FE	Yes	Yes	Yes	Yes
Court FE	No	Yes	Yes	No
Month FE	No	No	Yes	No
Court-by-Month FE	No	No	No	Yes
$\beta_1 = \beta_2$	0.82	0.35	0.40	0.08
$\beta_1 = \beta_3$	0.55	0.03	0.03	0.20
$\beta_1 = \beta_4$	0.66	0.05	0.07	0.27

Note: This table reports results from estimating equation (3): $y_{jcm} = \boldsymbol{\beta ExamRankQuintile_j} + \gamma_c + \delta_{w(j)} + \boldsymbol{X'_{jcm}\theta} + \epsilon_{jcm}$, where y_{jcm} is the inverse hyperbolic sine of cases disposed. All specifications include examination cohort (*Concurso*) fixed-effects. Columns (1) through (3) use the exam matched sample, observations used in the two-way fixed-effects models for which judge admission exams are available. Column (4) uses a subset of that sample that excludes all observations for which only one judge is working in any given court on a month. Standard-errors are clustered at the Judge level.

Table 3.7: Main results – correlation between admission grades and performance

	(1)	(2)	(3)	(4)
Top quintile	0.227			
	(0.0587)			
4th quintile	0.147			
	(0.0599)			
3rd quintile	0.107			
	(0.0575)			
2nd quintile	0.152			
	(0.0589)			
Final Grade (standardized)		0.0675	0.0692	
		(0.0224)	(0.0389)	
Objective Grade (standardized)				-0.0151
				(0.0417)
Written Exam Grade (standardized)				0.0141
				(0.0392)
Civil case decision (standardized)				0.105
				(0.0408)
Criminal case decision (standardized)				0.0362
				(0.0385)
Oral Grade (standardized)				0.0123
				(0.0422)
Titles Grade (standardized)				-0.0127
				(0.0429)
Observations	2878	2142	619	619
R-Squared	0.253	0.269	0.274	0.280
Number Admission Cohorts	78	65	20	20
Concurso Fixed-Effect	Yes	Yes	Yes	Yes

Note: This table reports results from estimating equations of the form: $JudgeFE_j = ExamOutcome'_j\beta + \delta_{w(j)} + \epsilon_j$, where $\delta_{w(j)}$ are admission cohorts (*Concurso*) fixed-effects and $ExamOutcome_j$ are the the independent variables of interest in each model in Columns (1) though (4). The dependent variable is Judge FEs, standardized to have unitary standard deviation within exam cohorts. All grades are standardized to have unitary standard deviation within each admission cohort. Robust standard errors in parentheses

Table 3.8: Robustness – excluding top and bottom performers

	(1)	(2)	(3)	(4)	(5)
Top quintile	0.227	0.285	0.229	0.221	0.222
	(0.0587)	(0.0678)	(0.0640)	(0.0668)	(0.0713)
4th quintile	0.147	0.155	0.147	0.141	0.138
	(0.0599)	(0.0605)	(0.0600)	(0.0682)	(0.0683)
3rd quintile	0.107	0.107	0.107	0.101	0.0990
	(0.0575)	(0.0575)	(0.0574)	(0.0658)	(0.0657)
2nd quintile	0.152	0.151	0.152	0.146	0.143
	(0.0589)	(0.0589)	(0.0589)	(0.0671)	(0.0671)
Observations	2,878	2,644	2,731	2,653	2,506
R-Squared	0.253	0.255	0.256	0.248	0.251
Number Admission Cohorts	78	77	78	78	78
Drop Top 3	No	Yes	No	No	No
Drop Top 5%	No	No	Yes	No	Yes
Drop Bottom 5%	No	No	No	Yes	Yes

Note: This table reports results from estimating equations of the form: $JudgeFE_j = \boldsymbol{\beta ExamRankQuintile_j} + \delta_{w(j)} + \epsilon_j$, where $\delta_{w(j)}$ are admission cohorts (*Concurso*) fixed-effects. Column (1) reproduces the main result from the Table 3.7, while Columns (2) through (5) re-estimate the model in subsamples that exclude top and/or bottom contenders, as specified above. Robust standard errors in parentheses

Table 3.9: Robustness – correlation between admission grades and performance

	(1)	(2)	(3)	(4)
Top quintile	-4.705			
	(1.082)			
4th quintile	-2.609			
	(1.092)			
3rd quintile	-1.303			
	(1.073)			
2nd quintile	-2.042			
	(1.089)			
Final Grade (standardized)		-1.451	-1.783	
		(0.350)	(0.629)	
Objective Grade (standardized)				-0.366
				(0.678)
Written Exam Grade (standardized)				-0.568
				(0.625)
Civil case decision (standardized)				-1.311
				(0.650)
Criminal case decision (standardized)				-0.845
				(0.656)
Oral Grade (standardized)				-0.626
				(0.667)
Titles Grade (standardized)				-0.227
				(0.756)
Observations	2879	2143	620	620
R-Squared	0.425	0.393	0.421	0.423
Number Admission Cohorts	79	66	21	21
Concurso Fixed-Effect	Yes	Yes	Yes	Yes

Note: This table reports results from estimating equations of the form: $RankFE_j = ExamOutcome_j'\beta + \delta_{w(j)} + \epsilon_j$, where $\delta_{w(j)}$ are admission cohorts (*Concurso*) fixed-effects and $ExamOutcome_j$ are the the independent variables of interest in each model in Columns (1) though (4). All grades are standardized to have unitary standard deviation within each admission cohort. Robust standard errors in parentheses

Bibliography

Abowd, J. M., R. H. Creecy, and F. Kramarz (2002). Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data. *U.S. Census Bureau*.

Alatas, V., A. Banerjee, R. Hanna, B. A. Olken, and J. Tobias (2012). Targeting the Poor: Evidence from a Field Experiment in Indonesia. *American Economic Review 102*(4), 1206–1240.

Alejos, L. (2018). Firms' (Mis)-Reporting under a Minimum Tax: Evidence from Guatemalan Corporate Tax Returns. *Working Paper*.

Allingham, M. G. and A. Sandmo (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics 1*(3), 323–338.

Almunia, M. and D. Lopez-Rodriguez (2018). Under the Radar: The Effects of Monitoring Firms on Tax Compliance. *American Economic Journal: Economic Policy 10*(1), 1–38.

Aman-Rana, S. (2020). In Self Interest? Meritocratic Promotions in a Bureaucracy through Discretion of Seniors. *Working Paper*.

Ashraf, N. and O. Bandiera (2018). Social Incentives in Organizations. *Annual Review of Economics*.

Ashraf, N., O. Bandiera, E. Davenport, and S. S. Lee (2020). Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services. *American Economic Review 110*(5), 1355–94.

Ashraf, N., O. Bandiera, and B. K. Jack (2014). No margin, no mission? A field experiment on incentives for public service delivery. *Journal of Public Economics 120*, 1–17.

Athey, S. and G. Imbens (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.

Auerbach, A. (2005). Who Bears the Corporate Tax? A review of What We Know. Technical Report w11686, National Bureau of Economic Research, Cambridge, MA.

Bachas, P. and M. Soto (2021). Corporate Taxation under Weak Enforcement. *American Economic Journal: Economic Policy Forthcoming.*

Barbosa, K. and F. V. Ferreira (2019). Occupy Government: Democracy and the Dynamics of Personnel Decisions and Public Sector Performance. Technical Report w25501, National Bureau of Economic Research.

Basri, M. C., M. Felix, R. Hanna, and B. A. Olken (2019). Tax Administration vs. Tax Rates: Evidence from Corporate Taxation in Indonesia. Working Paper 26150, National Bureau of Economic Research. Series: Working Paper Series.

Bastani, S. and H. Selin (2014). Bunching and non-bunching at kink points of the Swedish tax schedule. *Journal of Public Economics 109*, 36–49.

Bastani, S. and D. Waldenström (2020). How Should Capital Be Taxed? *Journal of Economic Surveys 34*(4), 812–846. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/joes.12380.

Bellemare, M. F. and C. J. Wichman (2020). Elasticities and the Inverse Hyperbolic Sine Transformation. *Oxford Bulletin of Economics and Statistics 82*(1), 50–61. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/obes.12325.

Bertanha, M., A. H. McCallum, and N. Seegert (2018). Better Bunching, Nicer Notching. *SSRN Electronic Journal*.

Bertrand, M., R. Burgess, A. Chawla, and G. Xu (2019). The Glittering Prizes: Career Incentives and Bureaucrat Performance. *The Review of Economic Studies*.

Bertrand, M., B. Crépon, A. Marguerie, and P. Premand (2017). *Contemporaneous and Post-Program Impacts of a Public Works Program.* Other papers. World Bank.

Best, M. C., A. Brockmeyer, H. J. Kleven, J. Spinnewijn, and M. Waseem (2015). Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan. *Journal of Political Economy 123*(6), 1311–1355.

Best, M. C., J. S. Cloyne, E. Ilzetzki, and H. J. Kleven (2020). Estimating the Elasticity of Intertemporal Substitution Using Mortgage Notches. *The Review of Economic Studies 87*(2), 656–690. Publisher: Oxford Academic.

Best, M. C., J. Hjort, and D. Szakonyi (2019). Individuals and Organizations as Sources of State Effectiveness. Technical report.

Bigio, S. and E. Zilberman (2011). Optimal self-employment income tax enforcement. *Journal of Public Economics 95*(9-10), 1021–1035. Publisher: Elsevier.

Blomquist, S. and W. K. Newey (2017). The Bunching Estimator Cannot Identify the Taxable Income Elasticity. SSRN Scholarly Paper ID 3100040, Social Science Research Network, Rochester, NY.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013). Does management matter? Evidence from india. *Quarterly Journal of Economics*.

Bloom, N., C. Propper, S. Seiler, and J. Van Reenen (2015). The Impact of Competition on Management Quality: Evidence from Public Hospitals. *The Review of Economic Studies 82*(2), 457–489.

Bloom, N. and J. Van Reenen (2007). Measuring and Explaining Management Practices across Firms and Countries. *The Quarterly Journal of Economics 122*(4), 1351–1408. Publisher: Oxford University Press.

Brockmeyer, A., S. Smith, M. Hernandez, and S. Kettle (2019). Casting a Wider Tax Net: Experimental Evidence from Costa Rica. *American Economic Journal: Economic Policy 11*(3), 55–87.

Brollo, F., P. Forquesato, and J. C. Gozzi (2017). To the Victor Belongs the Spoils? Party Membership and Public Sector Employment in Brazil. *SSRN Electronic Journal*.

Bruhn, M. and D. McKenzie (2009). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics 1*(4), 200–232.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2017). Firms and Labor Market Inequality: Evidence and Some Theory. *Journal of Labor Economics*.

Card, D., A. R. Cardoso, and P. Kline (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *Quarterly Journal of Economics*.

Card, D., J. Heining, and P. Kline (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality*. *The Quarterly Journal of Economics*.

Carrillo, P., D. Pomeranz, and M. Singhal (2017). Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement. *American Economic Journal: Applied Economics 9*(2), 144–164.

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

Chemin, M. (2009). The impact of judiciary on entrepreneurship: evaluation of Pakistan's Access to Justice Program. *Journal of Public Economics 93*, 114–125.

Chetty, R. (2009). Is the Taxable Income Elasticity Sufficient to Calculate Deadweight Loss? The Implications of Evasion and Avoidance. *American Economic Journal: Economic Policy 1*(2), 31–52.

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011). Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records. *The Quarterly Journal of Economics 126*(2), 749–804.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review 104*(9), 2593–2632.

Colonnelli, E., M. Prem, and E. Teso (2020). Patronage and Selection in Public Sector Organizations. *American Economic Review 110*(10), 3071–3099.

Congressional Budget Office (2020). Trends in the Internal Revenue Service's Funding and Enforcement.

Conselho Nacional de Justiça (2018). Justiça em Números. Technical report, Brasília.

Dal Bó, E., F. Finan, N. Li, and L. Schechter (2021). Information Technology and Government Decentralization: Experimental Evidence From Paraguay. *Econometrica 89*, 677–701.

Dal Bó, E., F. Finan, and M. A. Rossi (2013). Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service. *The Quarterly Journal of Economics 128*(3), 1169–1218.

Dalton, A. G., L. A. Manning, J. C. Jamison, I. K. Sen, J. G. Karver, J. L. Castaneda Nunez, L. P. Guedes, and S. B. Mujica Estevez (2019). Behavioral Insights for Tax Compliance. Technical Report 144398, The World Bank.

Davis, J. and S. Heller (2017). Using Causal Forests to Predict Treatment Heterogeneity: An Application to Summer Jobs. *American Economic Review 107*, 546–550.

De Neve, J.-E., C. Imbert, J. Spinnewijn, T. Tsankova, and M. Luts (2019). How to Improve Tax Compliance? Evidence from Population-Wide Experiments in Belgium. SSRN Scholarly Paper ID 3389405, Social Science Research Network, Rochester, NY.

De Neve, J.-E., C. Imbert, J. Spinnewijn, T. Tsankova, and M. Luts (2020). 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.

DellaVigna, S., N. Otis, and E. Vivalt (2020). Forecasting the Results of Experiments: Piloting an Elicitation Strategy. Working Paper 26716, National Bureau of Economic Research. Series: Working Paper Series.

DellaVigna, S. and D. Pope (2018). Predicting Experimental Results: Who Knows What? *Journal of Political Economy 126*(6), 2410–2456. Publisher: The University of Chicago Press.

Deserranno, E. (2019). Financial Incentives as Signals: Experimental Evidence from the Recruitment of Village Promoters in Uganda. *American Economic Journal: Applied Economics 11*(1), 277–317.

Devereux, M. P., L. Liu, and S. Loretz (2014). The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records. *American Economic Journal: Economic Policy 6*(2), 19–53.

Diamond, P. A. and J. A. Mirrlees (1971). Optimal Taxation and Public Production I: Production Efficiency. *The American Economic Review 61*(1), 8–27.

Duflo, E., P. Dupas, and M. Kremer (2015). School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools. *Journal of public Economics 123*, 92–110. Publisher: Elsevier.

Fack, G. and C. Landais (2016). The effect of tax enforcement on tax elasticities: Evidence from charitable contributions in France. *Journal of Public Economics 133*(C), 23–40. Publisher: Elsevier.

Fenizia, A. (2020). Managers and Productivity in the Public Sector. *Working Paper*.

Finan, F., B. A. Olken, and R. Pande (2015). The Personnel Economics of the State. Working Paper 21825, National Bureau of Economic Research.

Finan, F., B. A. Olken, and R. Pande (2017). The Personnel Economics of the Developing State. In A. V. Banerjee and E. Duflo (Eds.), *Handbook of Economic Field Experiments*, Volume 2 of *Handbook of Economic Field Experiments*, pp. 467–514. North-Holland.

Finkelstein, A., M. Gentzkow, and H. Williams (2016). Sources of geographic variation in health care: Evidence from patient migration. *Quarterly Journal of Economics*.

Fuest, C., A. Peichl, and S. Siegloch (2018). Do Higher Corporate Taxes Reduce Wages? Micro Evidence from Germany. *American Economic Review 108*(2), 393–418.

Gelber, A. M., D. Jones, and D. W. Sacks (2020). Estimating Adjustment Frictions Using Nonlinear Budget Sets: Method and Evidence from the Earnings Test. *American Economic Journal: Applied Economics 12*(1), 1–31.

Gordon, R. and W. Li (2009). Tax structures in developing countries: Many puzzles and a possible explanation. *Journal of Public Economics 93*(7-8), 855–866. Publisher: Elsevier.

Hallsworth, M. (2014). The use of field experiments to increase tax compliance. *Oxford Review of Economic Policy 30*(4), 658–679. Publisher: Oxford Academic.

Harberger, A. C. (1962). The Incidence of the Corporation Income Tax. *Journal of Political Economy 70*(3), 215–240. Publisher: University of Chicago Press.

Hoffman, M., L. B. Kahn, and D. Li (2018). Discretion in hiring. *Quarterly Journal of Economics 133*(2), 765–800.

Hoffman, M. and S. Tadelis (2018). People Management Skills, Employee Attrition, and Manager Rewards: An Empirical Analysis. Technical Report w24360, National Bureau of Economic Research, Cambridge, MA.

Holmstrom, B. and P. Milgrom (1991). Multitask Principal-Agent Analyses : Incentive. *Journal of Law, Economics, and Organization*.

Hussam, R., N. Rigol, and B. N. Roth (2020). Targeting High Ability Entrepreneurs Using Community Information: Mechanism Design in the Field.

Interamerican Development Bank (2015). Tax Administration Institutional and Operational Strenghtening (HO - L1108).

International Monetary Fund (2015). Current Challenges in Revenue Mobilization - Improving Tax Compliance. *Policy Papers 2015*(5).

International Monetary Fund (2018a). Honduras: 2018 Article IV Consultation. *IMF*.

International Monetary Fund (2018b). Honduras: Staff Report for the 2018 Article IV Consultation. Technical report, International Monetary Fund, Washington, D.C. OCLC: 1048787688.

International Monetary Fund (2019a). *Corporate Taxation in the Global Economy*. Washington, D.C.: International Monetary Fund.

International Monetary Fund (2019b). World Revenue Longitudinal Data.

International Monetary Fund, Organisation for Economic Cooperation and Development, United Nations, and World Bank Group (2016). *Enhancing the Effectiveness of External Support in Building Tax Capacity in Developing Countries*. Other papers. World Bank.

Johannesen, N., P. Langetieg, D. Reck, M. Risch, and J. Slemrod (2020). Taxing Hidden Wealth: The Consequences of US Enforcement Initiatives on Evasive Foreign Accounts. *American Economic Journal: Economic Policy 12*(3), 312–346.

Kanbur, R. and M. Keen (2014). Threshold, Informality, and Partitions of Compliance. *Working Papers 180136, Cornell University, Department of Applied Economics and Management.*. Publisher: Unknown.

Keen, M. and J. Mintz (2004). The optimal threshold for a value-added tax. *Journal of Public Economics 88*(3-4), 559–576. Publisher: Elsevier.

Kettle, S., M. A. Hernandez Hernandez, S. Ruda, and M. Sanders (2016). Behavioral interventions in tax compliance : evidence from Guatemala. Technical Report WPS7690, The World Bank.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics 8*(1), 435–464.

Kleven, H. J. (2018). Calculating Reduced-Form Elasticities Using Notches. *Technical note*.

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. *Econometrica 79*(3), 651–692.

Kleven, H. J. and M. Waseem (2013). Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan. *The Quarterly Journal of Economics 128*(2), 669–723.

Kondylis, F. and M. Stein (2018). The Speed of Justice. *World Bank Policy Research Working Paper No. 8372.*

Leaver, C., R. F. Lemos, and D. Dillenburg Scur (2019). Measuring and Explaining Management in Schools : New Approaches Using Public Data. Technical Report WPS9053, The World Bank.

Li, H., G. Watson, and T. LaJoie (2020). Details and Analysis of Former Vice President Biden's Tax Proposals. *Tax Foundation Fiscal Fact*.

Londoño-Vélez, J. and J. Ávila Mahecha (2019). Can Wealth Taxation Work in Developing Countries? Quasi-Experimental Evidence from Colombia. *Working Paper*.

Luttmer, E. F. P. and M. Singhal (2014). Tax Morale. *Journal of Economic Perspectives 28*(4), 149–168.

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.

Mehmood, S. (2020). Judicial Independence and Development: Evidence from Pakistan. *Working Paper*.

Mosberger, P. (2016). Accounting versus real production responses among firms to tax incentives: bunching evidence from Hungary. Technical Report 2016/3, Magyar Nemzeti Bank (Central Bank of Hungary).

Muralidharan, K. and V. Sundararaman (2011). Teacher Performance Pay: Experimental Evidence from India. *Journal of Political Economy*.

Oliveira Gomes, A. (2014). Estudos sobre desempenho da Justiça Estadual de primeira instância no Brasil. *PhD Thesis*.

Pinheiro, A. C. (2003). Judiciary, Reform and Economics: The Judges' Viewpoint. SSRN Scholarly Paper ID 482801, Social Science Research Network, Rochester, NY.

Pomeranz, D., C. Marshall, and P. Castellón (2014). Randomized tax enforcement massages : 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).

Ponticelli, J. and L. Alencar (2016). Court Enforcement, Bank Loans and Firm Investment: Evidence from a Bankruptcy Reform in Brazil. *The Quarterly Journal of Economics 131 (3)*, 1365–1413.

Prendergast, C. (2008). Intrinsic Motivation and Incentives. *American Economic Review 98*(2), 201–205.

Rao, M. (2020). Judges, Lenders, and the Bottom Line: Court-ing Firm Growth in India. *Working Paper*.

Rasul, I. and D. Rogger (2018). Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service. *The Economic Journal 128*(608), 413–446.

Sadka, J., E. Seira, and C. Woodruff (2018). Information and Bargaining Through Agents: Experimental Evidence from Mexico's Labor Courts. *NBER Working Paper Series 25137*.

Saez, E. (2010). Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy 2*(3), 180–212.

Sallee, J. M. and J. Slemrod (2012). Car notches: Strategic automaker responses to fuel economy policy. *Journal of Public Economics 96*(11-12), 981–999.

Sarin, N. and L. H. Summers (2020). Understanding the Revenue Potential of Tax Compliance Investment. *NBER Working Paper No. 27571*.

Schiavon, L. (2017). *Essays on crime and justice.* PhD Thesis, PUC-Rio.

Silver, D. (2019). Haste or Waste? Peer Pressure and Productivity in the Emergency Department. *Working Paper*.

Slemrod, J. (2013). Buenas notches: lines and notches in tax system design. *eJournal of Tax Research 11*(3).

Slemrod, J. and W. Kopczuk (2002). The optimal elasticity of taxable income. *Journal of Public Economics 84*(1), 91–112. Publisher: Elsevier.

Suárez Serrato, J. C. and O. Zidar (2016). Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms. *American Economic Review 106*(9), 2582–2624.

Trigueros, M. P., F. P. Longinotti, and J. S. Vecorena (2012). Estimación del Incumplimiento Tributario en América Latina: 2000-2010. *Documento de Trabajo n 3 - 2012. Dirección de Estudios y Investigaciones Tributárias - Centro Interamericano de Administraciones Tributárias*, 54.

Wager, S. and S. Athey (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.

Weaver, J. (2016). Jobs for Sale: Corruption and Misallocation in Hiring. *Working paper*.

Xu, G. (2018). The Costs of Patronage: Evidence from the British Empire. *American Economic Review 108*(11), 3170–3198.

Zucman, G. (2014). Taxing across Borders: Tracking Personal Wealth and Corporate Profits. *Journal of Economic Perspectives 28*(4), 121–148.

Appendices

Appendix A

Corporate Taxation and Evasion Responses: Evidence from a Minimum Tax in Honduras

A.1 Approximating the elasticity with notch

In this section we adapt the exercise of Kleven and Waseem (2013) and Kleven (2018) to obtain the elasticity formula when taxpayers face a notch instead of a kink. The intuition behind the derivation is that we try to recover what would have been the kink that would "replicate" the same behavior observed with the notch. We start by considering the average slope of the indifference curve of the marginal buncher: this IC is tangent to the threshold using the hypothetical kink with slope $(1-\tau^*)$ and has slope of $(1-t_0-\Delta t)$ at the point $y^t+\Delta Y$. In our case, $t_0=0$ since the effective marginal rate on revenue is zero below the threshold, and $\Delta t=\tau_y=0.015$. We can write

$$\frac{\int_{y^T}^{y^t+\Delta Y} I'(y)dy}{\Delta Y} \approx \frac{I'(y^T)+I'(y^t+\Delta Y)}{2} = \frac{(1-\tau^*)+(1-t-\Delta t)}{2} = \frac{(1-\tau^*)+(1-\tau_y)}{2}$$

The implicit tax rate faced by corporations is the change in tax liability when we change the reported revenue from above the threshold to exactly at the notch:

$$\begin{aligned} t^* = \frac{T(y^t+\Delta Y)-T(y^T)}{\Delta Y} &= \frac{\tau_y(y^t+\Delta Y)-\tau_\pi(y^T-\hat{c})}{\Delta Y} \\ &= \tau_y + \frac{\tau_y y^T+\tau_\pi(y^T-\hat{c})}{\Delta Y} \end{aligned}$$

Combining the fact that we have these two approximations to the slope of the IC in that region, and that $\Delta t = 0.015 = \tau_y$, we can write:

$$1 - t^* = \frac{(1-\tau^*) + (1-\tau_y)}{2}$$
$$\tau^* = \tau_y + 2\left(\frac{\tau_y Y^T + \tau_\pi\left(y^T - \hat{c}\right)}{\Delta Y}\right)$$

Plugging in the expression for τ^* in the usual expression for obtaining revenue elasticity when facing changes in marginal taxes we obtain:

$$\begin{aligned}\epsilon_{y,(1-t)} = \frac{\frac{\Delta Y}{Y^T}}{\frac{\Delta \tau^*}{(1-\tau^*)}} &= \frac{\Delta Y}{Y^T}\left(\frac{1-\tau^*}{\tau^* - t_0}\right) \\ &= \frac{\Delta Y}{Y^T}\left(\frac{1-\tau^*}{\tau_y + 2\left(\frac{\tau_y Y^T - \tau_\pi (Y_T - \hat{c})}{\Delta Y}\right)}\right) \\ &= \left(\frac{1}{\tau_y\left(2 + \frac{\Delta Y}{Y^T}\right) - 2\tau_\pi \frac{(Y_T - \hat{c})}{Y_T}}\right)\left(\frac{\Delta Y}{Y^T}\right)^2 (1-\tau)\end{aligned}$$

Some things are worth noting from this expression. First, for a firm with zero reported profit at the notch ($y^T = \hat{c}$), than the expression above simplifies to

$$\epsilon_{y,(1-\tau)} \approx \left(\frac{\Delta Y}{Y^T}\right)^2 \left(\frac{(1-\tau)}{\Delta \tau}\right)\left(\frac{1}{2 + \frac{\Delta Y}{Y^T}}\right)$$

which is exactly the same expression in Kleven and Waseem (2013). This is the expression we use to calculate the upper bound of elasticities presented in the text, since the taxpayer with highest incentive to bunch has profits only marginally above zero.

Second, note that if profit margin is exactly 6%, then it's true that

$$\tau_y\left(2 + \frac{\Delta Y}{Y^T}\right) - 2\tau_\pi 0.06 = 0.015\left(2 + \frac{\Delta Y}{Y^T}\right) - 2(0.25)0.06 = 0.015 * \frac{\Delta Y}{Y^T}$$

and the elasticity becomes

$$\begin{aligned}\epsilon_{y,(1-t)} &= \left(\frac{1}{\tau_y(2 + \frac{\Delta Y}{Y^T}) - 2\tau_\pi \frac{(Y_T - c)}{Y_T}}\right)\left(\frac{\Delta Y}{Y^T}\right)^2 (1-\tau) \\ &= \left(\frac{Y^T}{0.015\Delta Y}\right)\left(\frac{\Delta Y}{Y^T}\right)^2 (1-\tau) \\ &= \left(\frac{\Delta Y}{Y^T}\right)\frac{(1-\tau)}{\tau_y} = \epsilon_{kink}\end{aligned}$$

For a taxpayer with 6% reported profit margin, the exemption threshold represents a kink, not a notch, since their tax liability changes continuously around the cutoff.

A.2 Estimation of revenue elasticity lower bound

Following Bachas and Soto (2021), we compute the lower-bound of average revenue elasticity considering that firms with different profit levels (generated by heterogeneity in fixed-costs) will face different incentives to bunch. First, recall that firms with counterfactual profits above 6% or below 0% will not decide to bunch, since they are not affected by the minimum tax. Second, for firms within that profit range, the incentive to bunch is directly proportional to their costs: firms with high costs (low profit margins) will have a strong incentive to bunch since their tax liability at the threshold will be small, while not bunching means a much larger tax liability based on their revenues.

Let $\Psi(y_0, c_0)$ be the joint distribution of revenue and costs. We can then express the amount of bunching taxpayers as

$$
\begin{aligned}
B &= \int_c \int_{Y^T}^{Y^T+\Delta Y} \Psi(y_0, c_0) dydc \\
&= \int_c \int_{Y^T}^{Y^T+\Delta Y} \phi_y(y_0)\phi(c_0) dydc \\
&= \int_{Y^T}^{Y^T+\Delta Y} \phi_y(y_0) \int_{c_0} \phi(c_0) dcdy \\
&= \int_{Y^T}^{Y^T+\Delta Y} \phi_y(y_0) \int_0^{m(y_0)} \phi(m_0) dmdy
\end{aligned}
$$

where in the second line we assume that the cost and revenue distributions are independent; in the third line we make it explicit that, for any given level of revenue, there is a cost region that will induce bunching; and in the last line we re-write the expression as a function of profit levels instead of cost, and make it explicit that, for any given revenue level, only low-profit taxpayers will bunch, the upper threshold of which depends on the revenue level. Intuitively, for taxpayers very close to the notch, all those potentially affected by the minimum tax will decide to bunch, whereas those farther from it will only bunch if the differential tax liability is large due to their low profits.

In order to connect the cost/profit levels that induce bunching at each revenue level, recall that we previously computed that, for the marginal buncher at revenue level $Y^T + \Delta Y$, we can compute the revenue elasticity as

$$
\epsilon_{y,(1-t)} = \left(\frac{1}{\tau_y(2 + \frac{\Delta Y}{Y^T}) - 2\tau_\pi \frac{(Y_T - \hat{c})}{Y_T}} \right) \left(\frac{\Delta Y}{Y^T} \right)^2
$$

We can rewrite this equality putting the reported cost $\hat{c}$ in evidence:

$$\hat{c}^* = Y^T\Big(1 - \frac{\tau_y}{\tau_\pi}\Big) - \frac{\tau_y}{\tau_\pi}\frac{\Delta y}{2} + \frac{(\Delta y)^2}{2\epsilon_y \tau_\pi Y^T}$$

For a given revenue level and elasticity, $\hat{c}^*$ is the cost at the threshold that would make a taxpayer indifferent between bunching and staying above the notch. Any taxpayer with costs above that level, i.e. a lower profit margin, would decide to bunch.

We implement the estimation of the revenue elasticity ϵ_y in the following steps. First, we need to consider the counterfactual profit distribution that would be observed in the absence of the notch. For each period in our sample, we take that to be the observed profit margin density for firms reporting revenue in the interval L6 - 8 million. We then proceed to compute, for each revenue bin (ΔY) and ϵ_y, what is the share of taxpayers with profit margin between 0 and the implied upper bound, and use the counterfactual density to obtain the number of taxpayers that bunch in each revenue bin. This allows us to obtain, for each potential revenue elasticity, the total number of predicted bunchers, which we compare to the estimated number of bunchers. The final elasticity, therefore, is the value that generates the same number of bunchers as the excess mass below the threshold.

A.3 Model calibration details

We modify firms' profit function by making explicit assumption about the cost and misreporting loss functions. Firms have isoelastic costs and also isoelastic loss function from misreporting costs:

$$\hat{\Pi}(y, \hat{c}) = (1-\tau)y + \tau\mu\hat{c} - \alpha_i - \frac{\theta_i}{1+1/e}\Big(\frac{y}{\theta_i}\Big)^{(1+1/e)} - \frac{B_i}{1+1/\gamma}\Big(\hat{c} - c(y)\Big)^{(1+1/\gamma)}$$

Each taxpayer is characterized by the vector $(\theta_i, \alpha_i, B_i)$ that define productivity, fixed cost and evasion ability, respectively. Given our functional forms, optimal vector of output and reported costs $(y^*, \hat{c}^*(y^*))$ are:

$$\begin{aligned} y^* =& \theta(1-\tau_E)^e \\ \hat{c}^*(y^*) =& c(y^*) + B_i\Big(\tau\mu\Big)^\gamma \end{aligned}$$

where $\tau_E = \tau\Big(\frac{1-\mu}{1-\tau\mu}\Big)$. Note that if we have profit taxation then $\mu = 1$ and $\tau_E = 0$, so firm size is undistorted.

In order to calibrate the model, we use data for the 2013, when no notches or kinks were in place. Under profit taxation, we have:

$$\begin{aligned} y^* =& \theta \\ c(y^*) =& \alpha + \frac{\theta}{1+1/e} \\ \hat{c}^*(y^*) =& \alpha + \frac{\theta}{1+1/e} + \left(\frac{\tau}{B_i}\right)^{\gamma} \end{aligned}$$

From the first-order conditions of an interior optimum, θ is simply the vector of reported output, which in this model coincides with real output. We also know the elasticity of output e, which we fix to be $e = 0.99$, the upper bound estimated for the pooled years. By using the upper bound of our elasticity estimate we are conservative in the case for using output taxation, since a higher elasticity will limit the potential benefit of the tax.

While we do not observe $c(y^*)$, the real costs, but only the reported costs $\hat{c}^*(y^*)$, we have estimated evasion as a share of profits using the 6% profit margin kink. Let that quantity be $\epsilon_{\hat{c}}$ .Using the fact that at the profit margin kink $(y - \hat{c})/y = \tau_y/\tau_\pi$ we can write:

$$\frac{(\hat{c}-c)}{y} = \frac{(\hat{c}-c)}{(y-\hat{c})} * \frac{(y-\hat{c})}{y} = \epsilon_{\hat{c}}(\tau_y/\tau_\pi) = \epsilon_{\hat{c}} * 0.06$$

Using the equations above, we have that

$$\frac{(\hat{c}-c)}{y} = \frac{\left(\frac{\tau}{B_i}\right)^{\gamma}}{\theta} = 0.06\epsilon_{\hat{c}}$$

In our setting, we do not have variation to identify γ, the elasticity of misreporting costs. Best et al. (2015) explore different profit tax rates for different subset of firms, while Bachas and Soto (2021) use estimates of cost elasticity in two different thresholds. We calibrate our model using the estimate from Best et al. (2015), which is approximately 1.5, which allows us to recover B_i as $B_i = \frac{\tau}{\left(\theta 0.06\epsilon_{\hat{c}}\right)^{1/\gamma}}$

Finally, given the previous we can just obtain the fixed cost vector α by computing

$$\alpha = \hat{c}^* - \frac{\theta}{1+1/e} + \left(\frac{\tau}{B_i}\right)^{\gamma}$$

A.4 Appendix Figures and Tables

Table A1: Alternative order of polynomial - Profit margin distribution

Year	(1) Excess Mass (B)	(2) Bunching(b)	(3) Delta Profit	(4) Estimated evasion ($\epsilon_y = 0.99$)
Order p = 3	779.64	5.38	1.10	-16.85
	(46.24)	*(0.39)*	*(0.10)*	*(1.44)*
Order p = 4	834.22	6.05	1.20	-18.52
	(40.69)	*(0.38)*	*(0.10)*	*(1.39)*
Order p = 6	788.99	5.49	1.10	-16.85
	(39.83)	*(0.37)*	*(0.10)*	*(1.35)*

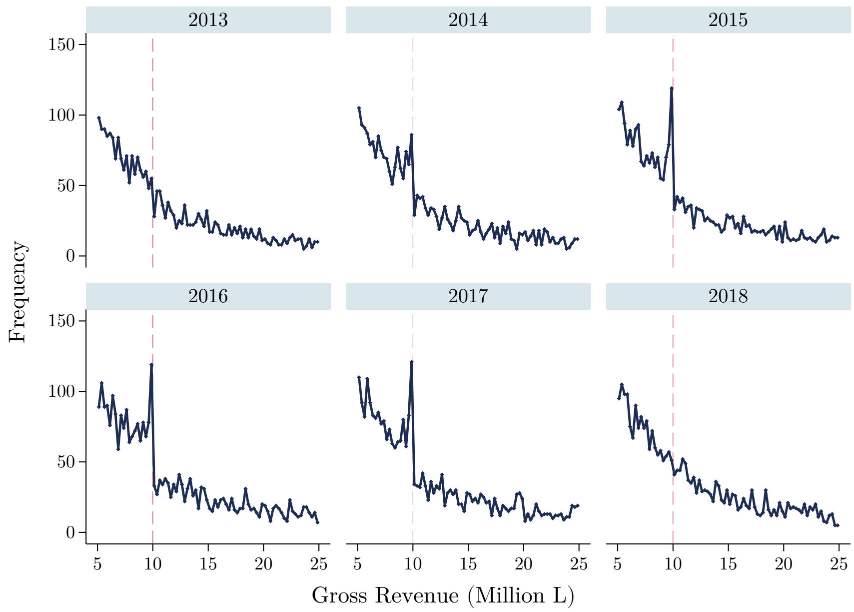
Note: This table presents results from replicating the exercises performed in Table 1.5 using different order of polynomials to estimate the counterfactual distribution of profit margin for the sample of pooled taxpayers in 2014-2017. The baseline specification uses polynomial regression of order five, while in this table we present results using polynomials of order three, four and six.

Table A2: Alternative order of polynomial - gross revenue distribution

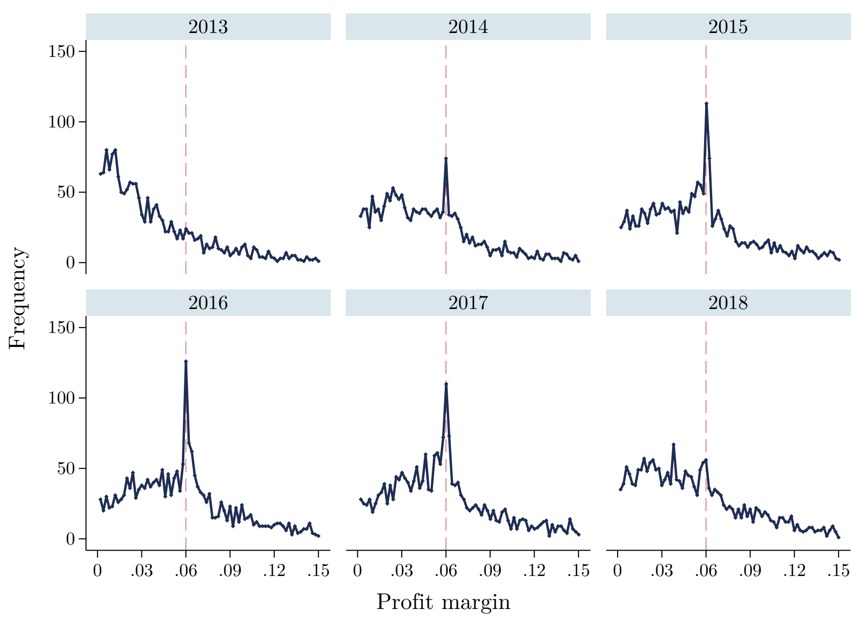
	(1) Excess # Firms (B)	(2) Firms % counterfactual (b)	(3) y_u (upper bound)	(4) Δ Revenue (upper bound)	(5) ϵ_y (upper)	(6) ϵ_y (lower)
Order p = 3	604.30	8.82	14.70	4.70	5.96	0.60
	(34.16)	*(0.70)*	*(0.68)*	*(0.68)*	*(1.36)*	*(0.09)*
Order p = 4	569.91	6.78	12.90	2.90	2.45	0.50
	(30.91)	*(0.60)*	*(0.61)*	*(0.61)*	*(1.09)*	*(0.06)*
Order p = 6	494.55	5.69	12.30	2.30	1.58	0.35
	(25.07)	*(0.63)*	*(0.78)*	*(0.78)*	*(1.29)*	*(0.04)*

Note: This table presents results from replicating the exercises performed in Table 1.3 using different order of polynomials to estimate the counterfactual distribution of gross revenue for the sample of pooled taxpayers in 2014-2017. The baseline specification uses polynomial regression of order five, while in this table we present results using polynomials of order three, four and six.

Figure A1: Robustness: Balanced panel of corporations (2013-2018)



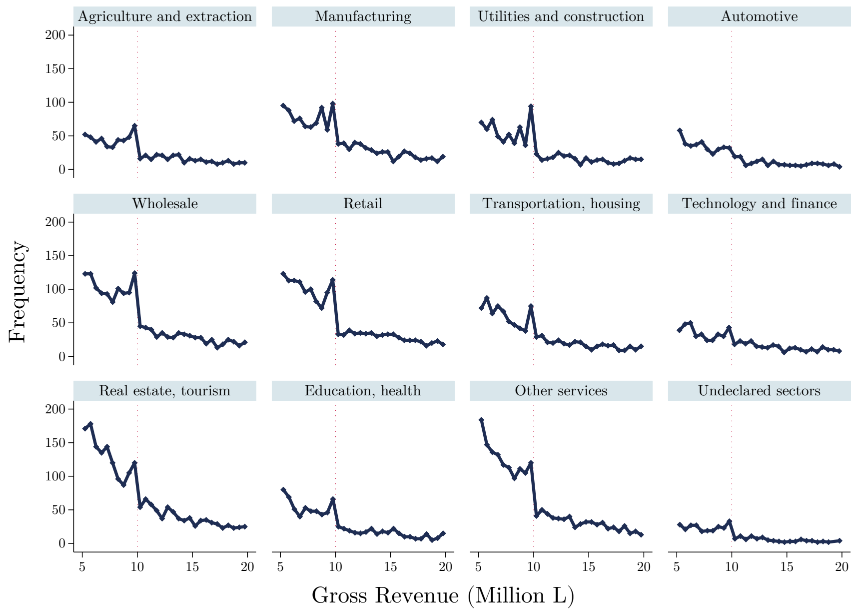
(a) Gross revenue empirical density



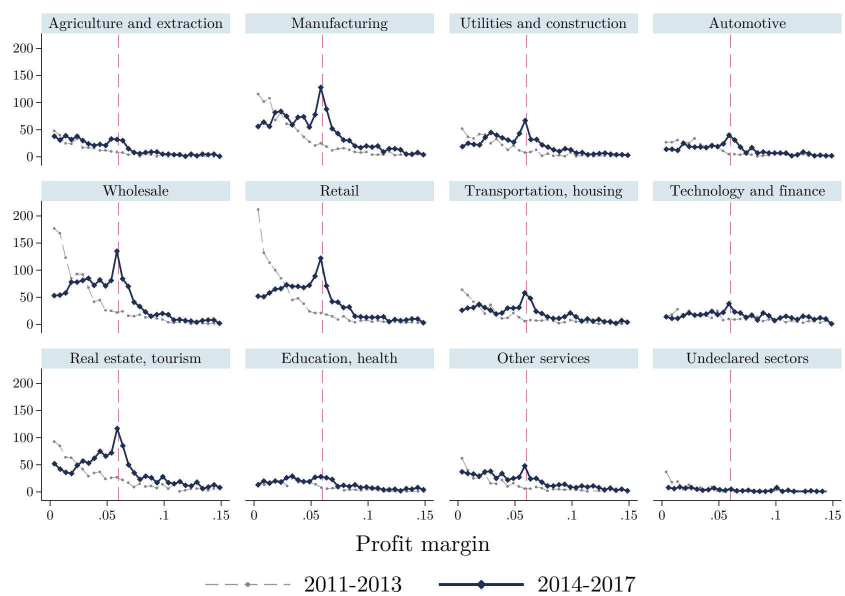
(b) Profit margin empirical density

Note: This figure presents the empirical density of gross revenues (Panel A) and profit margins (Panel B) for a balanced panel of 12,172 firms, for each year in the period 2013-2018. Bins are L250,000 wide in Panel A and 0.2 p.p. wide in Panel B. The sample in Panel B is restricted to firms reporting gross revenue above L13 million in each year.

Figure A2: Robustness: Behavioral responses by economic sector



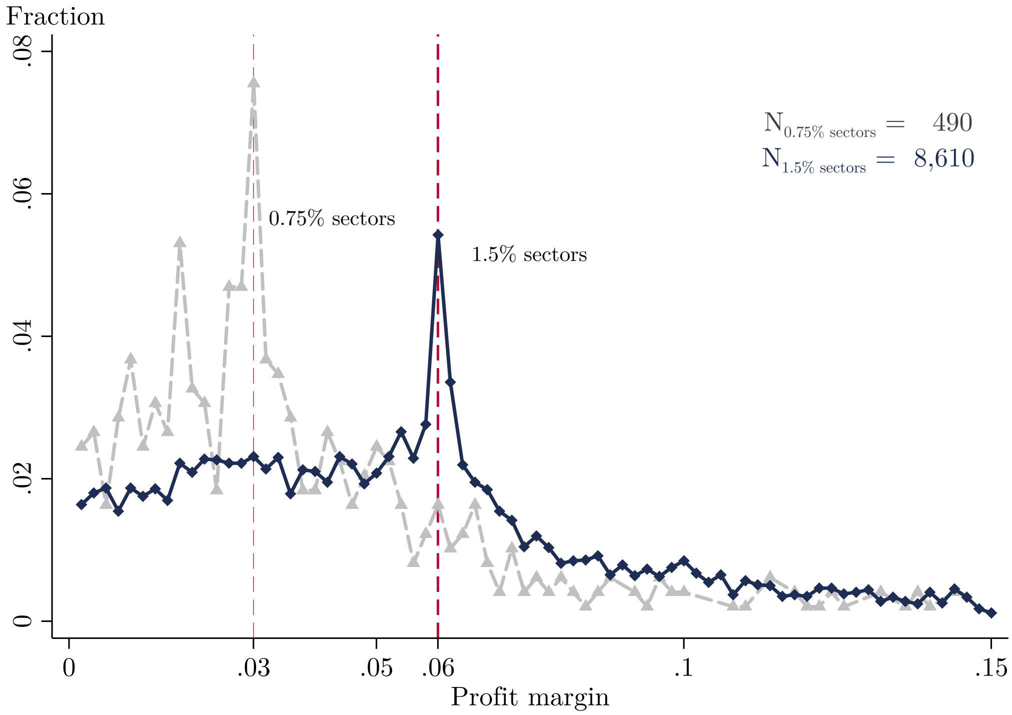
(a) Gross revenue empirical density



(b) Profit margin empirical density

Note: This figure presents the empirical density of gross revenues (panel A) and profit margins (Panel B) for firms in different economic sector for the period 2014-2017 pooled. Bins are L500,000 wide in Panel A and 0.5 p.p. wide in Panel B. The sample in Panel B is restricted to firms reporting revenue above L13 million (infra marginal to the revenue bunching).

Figure A3: Empirical Density around 6% profit margin threshold - 0.75% vs. 1.5% sectors (2014-2017)



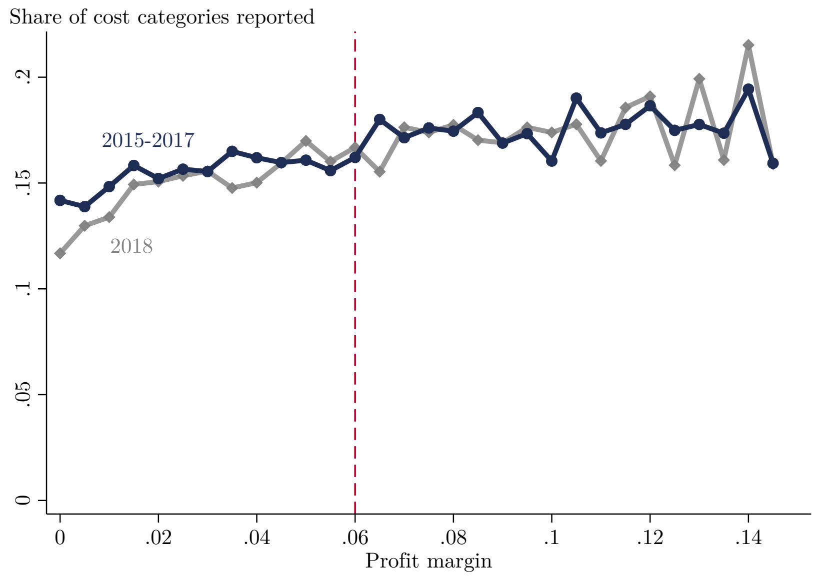
Note: This figure presents the empirical density of reported profit margins for firms subject to the 1.5% minimum tax (in solid blue) and those in sectors subject to the 0.75% rate (in dashed gray) for the period 2014-2017. The sample is restricted to firms reporting revenue above L13 million (infra marginal to revenue bunching). Bins are 0.2 p.p. wide and the first bins starts at 0.1% such that the 6% kink is the midpoint of a bin.

Table A3: Cost evasion responses across economic sectors

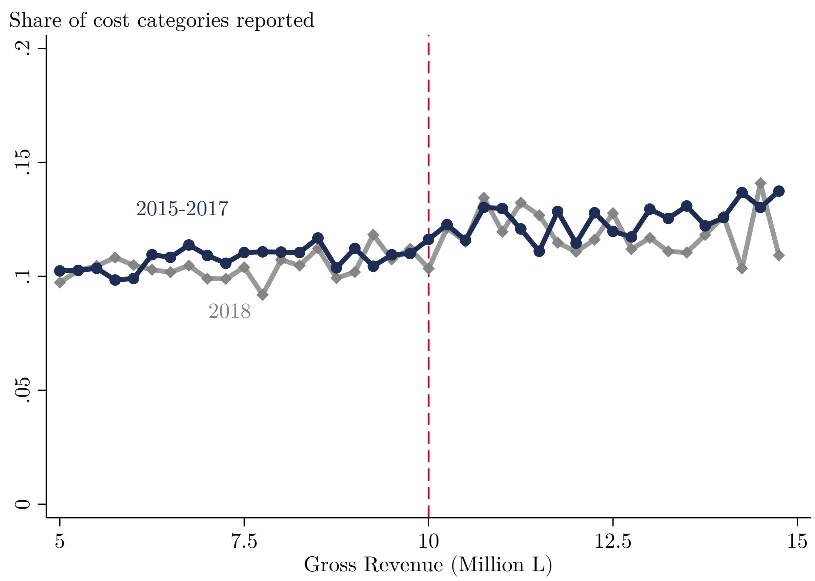
Year	(1) Excess Mass (B)	(2) Bunching(b)	(3) Delta Profit	(4) Estimated evasion ($\epsilon_y = 0.99$)
Agriculture and extraction	38.35	6.06	1.20	-18.52
	(10.02)	*(2.40)*	*(0.50)*	*(7.96)*
Manufacturing	153.10	7.86	1.60	-25.18
	(13.45)	*(1.16)*	*(0.20)*	*(3.87)*
Utilities and construction	61.86	5.55	1.10	-16.85
	(7.52)	*(0.89)*	*(0.20)*	*(3.02)*
Automotive	49.72	7.91	1.60	-25.18
	(6.47)	*(1.54)*	*(0.30)*	*(5.15)*
Wholesale	132.19	5.66	1.10	-16.85
	(13.63)	*(0.81)*	*(0.20)*	*(2.75)*
Retail	85.16	3.71	0.70	-10.18
	(10.82)	*(0.58)*	*(0.10)*	*(2.00)*
Transportation, housing	69.39	8.09	1.60	-25.18
	(8.57)	*(1.76)*	*(0.30)*	*(5.83)*
Technology and finance	28.68	3.80	0.80	-11.85
	(7.22)	*(1.18)*	*(0.20)*	*(4.07)*
Real estate, tourism,other	93.89	4.15	0.80	-11.85
	(12.37)	*(0.67)*	*(0.10)*	*(2.34)*
Education, health, entertainment	31.71	4.59	0.90	-13.52
	(7.82)	*(1.48)*	*(0.30)*	*(4.92)*
Other services	34.21	4.04	0.80	-11.85
	(8.00)	*(1.28)*	*(0.30)*	*(4.26)*
Undeclared sectors	-1.93	-1.11	-0.20	4.82
	(4.98)	*(2.64)*	*(0.50)*	*(8.71)*

Note: This table presents estimates of change in reported profit margins and cost evasion for firms by economic sector, pooled for the 2014-2017 period. The first column reports the estimated excess number of firms (B) while column (2) reports the ratio between excess mass and average counterfactual density in the bunching region (b). Column (3) presents estimated change in profit margin, while column (4) present changes in cost misreporting using the decomposition in Equation 1.13.

Figure A4: Average number of cost categories with positive deduction



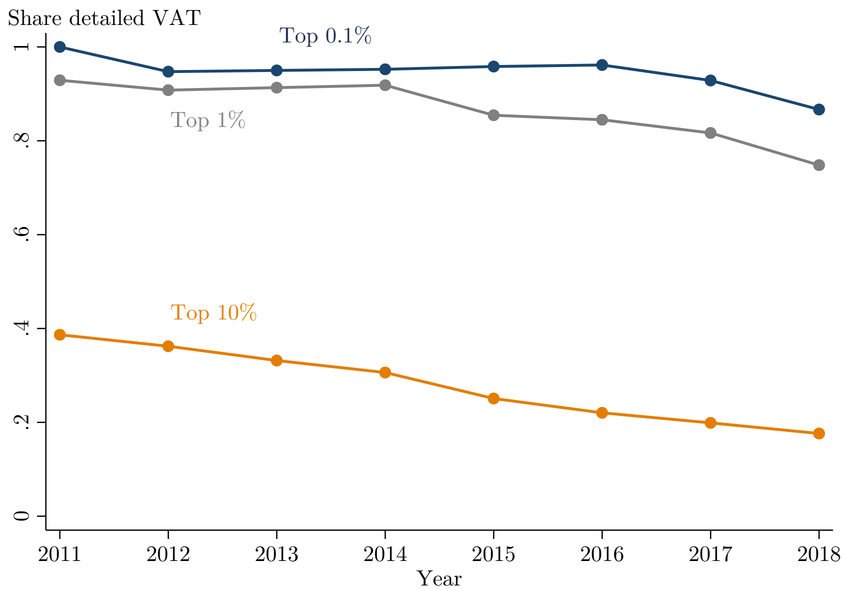
(a) Around 6% profit margin kink



(b) Around L10 million notch

Note: This figure presents the average share of all cost categories reported by taxpayers in each bin. Panel (a) restricts the sample to taxpayers reporting revenue above L12 million and therefore infra-marginal to the revenue bunching behavior. Profit margin bins are 0.5% wide. The blue line represents declarations in the period 2015-2017, when the minimum tax affected a large number of taxpayers, while the gray line refers to declarations in 2018, when only a small subset of corporations were affected by the minimum tax. Panel (b) compares the usage of cost categories across the reported gross revenue distribution, for the period 2015-2017 (blue) and 2018 (gray). Both panels restrict the sample to taxpayers filing electronically, for which detailed cost categories are available.

Figure A5: Share of taxpayers mandated to file detailed VAT purchases



Note: This figure presents, for each year in the period 2011-2018, the share of taxpayers in each revenue group (top 0.1%, top 1% and top 10%) that are defined as medium or large. These are the taxpayers with an obligation to file individualized information on their purchases to claim VAT deductions, generating independent information on suppliers' revenues. The list of medium and large taxpayers was defined in 2011 and has not changed since. Groups are mutually exclusive, so the group defined as top 1% exclude taxpayers in the top 0.1% and the 10% group all those in the top 1% and 0.1%. The sample excludes taxpayers exempt from all income taxes.

Appendix B

Targeting in Tax Compliance Interventions: Experimental Evidence from Honduras

B.1 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:

$$y_{i,2018} = \alpha + \beta' X_i + \epsilon_i \tag{B.1}$$

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 effects[1].

Table A1 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) 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 sent[2].

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 A2, 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

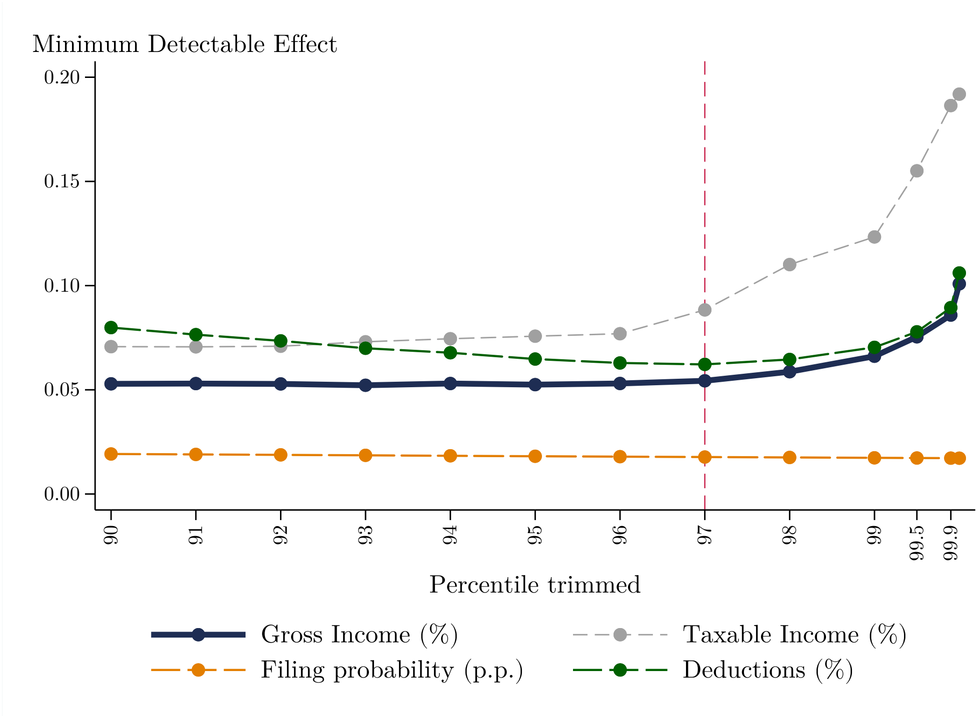
[1]These are strata dummies; dummy for presenting tax declaration in the previous year; amount of declared gross revenue, third-party informed gross revenue and declared taxable income in the previous year; and amount of declared sales tax revenues in the same year.

[2]This figure was informed by a pilot discussed below.

L12,400 for taxable income.

B.2 Appendix Figures and Tables

Figure A1: 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 B.1.

Table A1: Power Calculations - Residual Variance

	2018 primary outcomes		
	(1)	(2)	(3)
	Revenue	Deductions	Taxable Income
Taxable income 2017	-0.005	-0.868***	0.742***
	(0.069)	(0.065)	(0.036)
Reported revenue (Income) (2017) (L1,000s)	-0.306***	-0.284***	-0.014***
	(0.048)	(0.048)	(0.003)
Declared revenue (Income) (2017) (L1,000s)	0.857***	0.870***	0.005***
	(0.015)	(0.015)	(0.002)
Declared income tax in 2017	13.722	26.889	-7.314**
	(15.653)	(16.440)	(3.264)
Reported revenue (Sales) (2018) (L1,000s)	0.395***	0.371***	0.017***
	(0.047)	(0.047)	(0.003)
Constant	-44.950	-54.516	16.175**
	(41.024)	(41.848)	(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 A2: Power Calculations - Final Sample

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

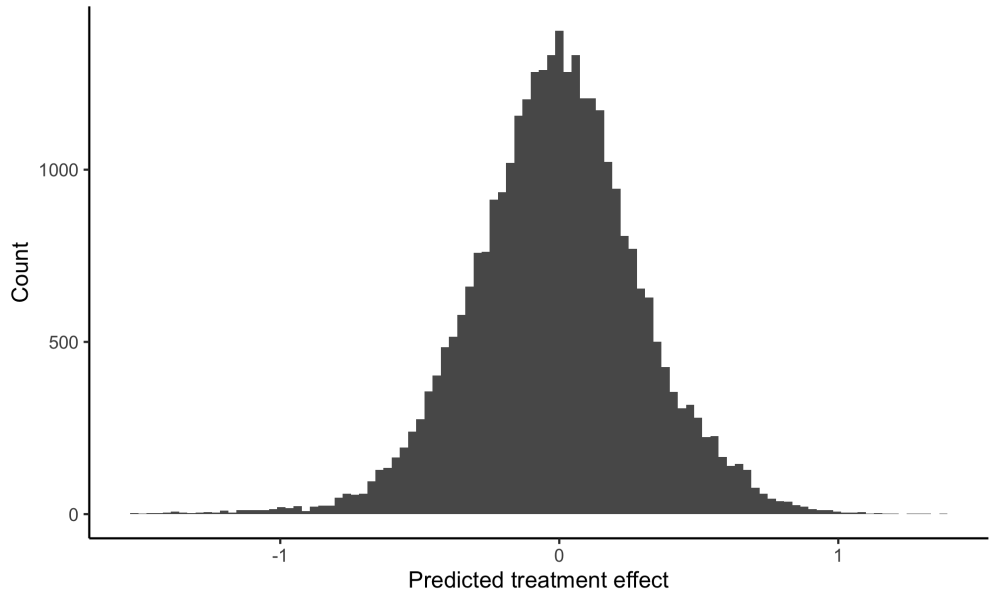
Note: In this table we present Minimum Detectable Effects (MDE) for our primary outcomes. The first panel presents MDEs for the experiment comparing pooled treatment arms with control, while the second panel presents MDE for the experiment comparing a single treatment arm with control. The first column 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.

Table A3: Primary outcomes trimmed at 99th percentile (non pre-specified)

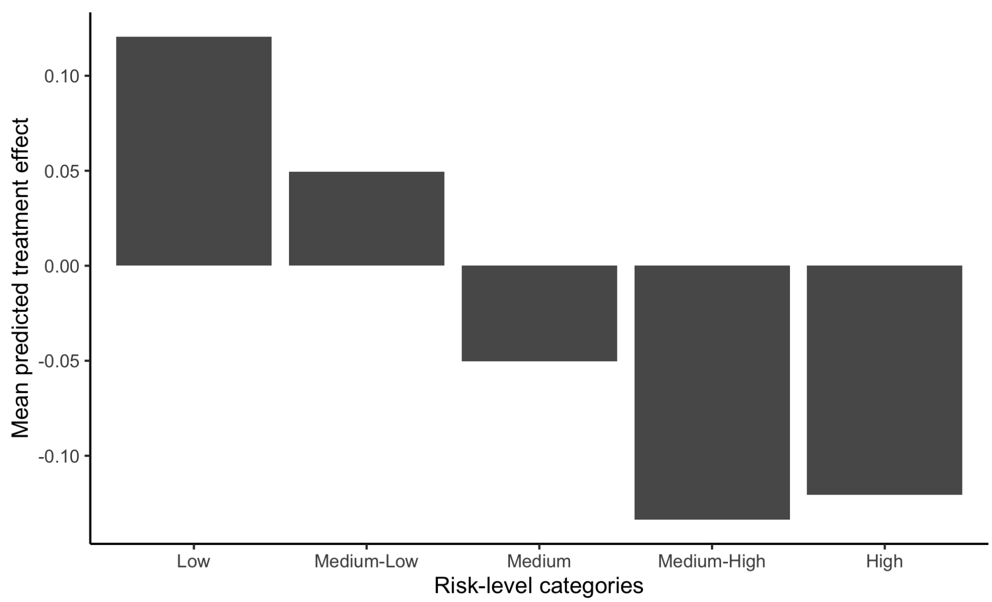
	(1) Gross Revenue	(2) Deductions	(3) Taxable Income
ITT estimates			
Treatment	-5.329	-4.561	2.024
	[-38.917,28.260]	[-38.000,28.878]	[-0.418,4.465]
Observations	31,082	31,082	31,082
R-Squared	0.571	0.564	0.208
Control mean	1,046.67	988.63	66.61
ITT estimates			
Sanctions treatment	-3.100	-5.132	2.079
	[-48.858,42.658]	[-50.835,40.570]	[-1.358,5.516]
Tax procedures treatment	-4.407	1.612	0.523
	[-51.879,43.064]	[-45.994,49.218]	[-2.955,4.002]
Moral duty treatment	-8.447	-10.126	3.460**
	[-56.672,39.779]	[-57.942,37.690]	[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 2.1. The second panel presents ITT estimates using the specification in Equation 2.4 . 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 A2: 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.

Figure A3: Letter 1 - Control group

Señor Obligado Tributario

JUAN PEREZ

RTN: XXXXXXXXXXXXXX

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 A4: Letter 2 - Sanctions treatment arm for available third-party information

Señor Obligado Tributario

JUAN PEREZ

RTN: XXXXXXXXXXXXXX

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 A5: Letter 3 - Procedure denial treatment arm for available third-party information

Señor Obligado Tributario

JUAN PEREZ

RTN: XXXXXXXXXXXXXX

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 A6: Letter 4 - Tax morale treatment arm for available third-party information

Señor Obligado Tributario

JUAN PEREZ

RTN: XXXXXXXXXXXXXX

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 A7: Letter 5 - Sanctions treatment arm for non-available third-party information

Señor Obligado Tributario

JUAN PEREZ

RTN: XXXXXXXXXXXXXX

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 A8: Letter 6 - Procedure denial treatment arm for non-available third-party information

Señor Obligado Tributario

JUAN PEREZ

RTN: XXXXXXXXXXXXXX

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 A9: Letter 7 - Tax morale treatment arm for non-available third-party information

Señor Obligado Tributario

JUAN PEREZ

RTN: XXXXXXXXXXXXXX

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

Appendix C

Selecting Top Bureaucrats: Admission Exams and Performance in Brazil

C.1 Connected sets in the data

Connected sets are defined as as groups of organizations (courts) and individuals (judges) connected by "movers", workers who are observed in more than one organization. In our context, there are two sources of variation that allow us to construct connected sets. First, judges are often observed working in more than one court in the same month, allowing us to create connections even within a single period (month). Figure A1 below illustrates this fact. The top-right figure show three judges observed working in the 10th Civil Court of Porto Velho, in the state of Rondonia, during the month of May 2013. As we can see in the top-right figure, two of these judges also worked in additional courts in that same month – in the 5th Civil Court and the 9th Civil court. These two courts, and all the judges working in them on that same month, are also part of the original connected set – the bottom figure shows that two additional judges were working in these courts in May, and our connected set has expanded.

This within month connections is only one source of variation used to build connected sets. Since we have a panel that covers 76 months, we can build all connections that happened at any point in that period. Figure A2 takes a broader view of these connections and present all connections in the states of Rondonia and Amapa, two small states that allow for better visualization of the judge-court networks. The top two figures and the bottom-left one shows all connections for the states of Rondonia in three periods: 2009, 2009-2010 and 2009-2011. Note that when only connections in 2009 are considered, connected sets are large but multiple: clusters of judges and courts are often not connected to other parts of the network. When we explore judges' movements across several years, on the other hand, the network becomes more densely connected: if we consider the entire 2009-2015 period, all judges and courts *within each state* belong to a single connected set, as shown in bottom-right figure[1]. Since judges are hired to work in a specific state, nonetheless, the figure also shows that each

[1]All graphs represent connections in the sample used to estimate the two-way fixed-effects model, and therefore include a single connected-set by construction. As discussed above, nonetheless, the largest connected set within each state often includes over 95% of all observations.

state is a separate connected set: judges in the state of Rondonia, for example, are never observed matched to courts in Amapa, and vice-versa.

C.2 Appendix Figures and Tables

Table A1: Detailed descriptive statistics in estimation sample

	Mean	SD	Median	N
Panel A - Judges				
Male	0.60	0.49	1	10,218
# Courts by Judge	5.54	4.39	4	727,784
Number of months Judge is observed	59.09	15.34	64	727,784
# of Courts at Judge-Month level	1.89	1.53	1	727,784
Number Municipalities Judge ever works in	2.68	1.99	2	727,784
Unique number of months per judge-court pair	34.35	22.37	31	727,784
Panel B - Courts				
# Judges by Court	6.36	5.13	5	727,784
# of Judges at Court-Month level	1.92	1.88	1	727,784
Civil Court	0.22	0.42	0	9,048
General Court	0.20	0.40	0	9,048
Small-stakes Court	0.18	0.39	0	9,048
Criminal Court	0.16	0.37	0	9,048
Family Court	0.10	0.30	0	9,048
Other Courts	0.13	0.34	0	9,048
Panel C - Output measures				
Cases Disposed (on merit)	40.13	50.09	22	727,784
Total Hearings (presided or held)	34.88	46.39	17	716,736

Note: This table reports descriptive statistics for key variables in the sample used to estimate the two-way fixed-effects model, where outcome variables are trimmed at the top 1%, judge-court spells shorter than three months are dropped and only observations in the largest connected sets within each state are used.

Table A2: Pairwise correlations between performance in admission exam phases

	Objective	Written	Civil case	Criminal case	Oral	Titles
Objective	1					
Written	0.0283	1				
Civil case	0.0127	0.0144	1			
Criminal case	0.0276	0.0886**	0.0604	1		
Oral	0.0352	0.103**	0.112***	0.0656	1	
Titles	-0.0403	0.0825**	0.122***	0.0265	0.154***	1

Note: This table reports pairwise correlations between residualized grades in each one of the six phases of admission examinations. Residues are obtained by regressing grades on admission exam fixed-effects so all grades are represented as deviations from exam average. Sample is restricted to exams with available grades for all exams (N = 619).

Table A3: Regressions using admission ranking as explanatory variable

	(1)	(2)	(3)	(4)
Rank Exam	-0.00114	-0.00315	-0.00305	-0.00355
	(0.000740)	(0.000581)	(0.000576)	(0.00124)
Observations	200,206	200,206	200,206	59,795
R-Squared	0.11	0.42	0.43	0.53
Concurso FE	Yes	Yes	Yes	Yes
Court FE	No	Yes	Yes	No
Month FE	No	No	Yes	No
Court-by-Month FE	No	No	No	Yes

Note: This table reports results from estimating equation (3): $y_{jcm} = \beta ExamRank_j + \gamma_c + \delta_{w(j)} + \boldsymbol{X'_{jcm}\theta} + \epsilon_{jcm}$, where y_{jcm} is the inverse hyperbolic sine of cases disposed. All specifications include examination cohort (*Concurso*) fixed-effects. Columns (1) through (3) use the exam matched sample, observations used in the two-way fixed-effects models for which judge admission exams are available. Column (4) uses a subset of that sample that excludes all observations for which only one judge is working in any given court on a month. Standard-errors are clustered at the Judge level (* p < 0.1, ** p < 0.05, *** p < 0.01).

Figure A1: Construction of connected sets in the data (Rondonia – May 2013)

Note: These graphs present the a selected network of judges (white squares) and courts (blue dots) in the state of Rondonia. Connections between dots and squares represent judges being observed working in a court in the month of May 2013. Starting from the top-left and moving clockwise, the graph expands the connected set by adding courts and judges observed paired in that month. All graphs use data from the sample used to estimate the two-way fixed-effects model.

Figure A2: Visualizing connected sets in the data

Note: These graphs present the networks of judges (white squares) and courts (blue dots) for the states of Rondonia and Amapa. Connections between dots and squares represent judges being observed working in a court in the referred period. The top-left figure presents connections observed in the state of Rondonia in 2009; the top-right includes connection observed in 2009 and 2010, while the bottom left presents connections in the period 2009-2011. The bottom right figure presents the universe of connections observed in in the entire panel for the states of Rondonia and Amapa. It highlights that there are no connection across states, since judges from one state are never observed working in a different state. All graphs use data from the sample used to estimate the two-way fixed-effects model, and therefore within each state all observations are connected by construction.