THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 1 Capitole

Présentée et soutenue par

Alipio FERREIRA DA SILVA FILHO

Le 7 juillet 2022

Essays in Public Economics

Ecole doctorale : **TSE - Toulouse Sciences Economiques**

Spécialité : **Sciences Economiques - Toulouse**

Unité de recherche :
TSE-R - Toulouse School of Economics - Recherche

Thèse dirigée par
Helmuth CREMER et Stefan AMBEC

Jury
Mme Lucie GADENNE, Rapporteure
M. Eduardo SOUZA-RODRIGUES, Rapporteur
M. Helmuth CREMER, Directeur de thèse
M. Stefan AMBEC, Co-directeur de thèse
M. Mathias REYNAERT, Co-directeur de thèse
Mme Hélène OLLIVIER, Présidente

A DISSERTATION

In order to obtain the title of

DOCTOR OF THE UNIVERSITY OF TOULOUSE

Issued by Université Toulouse 1 Capitole

Defended by

Alipio FERREIRA DA SILVA FILHO

On 7 July 2022

Essays in Public Economics

PhD School: **TSE - Toulouse Sciences Economiques**

Department: **Sciences Economiques - Toulouse**

Research Laboratory:
TSE-R - Toulouse School of Economics - Recherche

Supervised by
Helmuth CREMER and Stefan AMBEC

The jury members

Mrs. Lucie GADENNE, Referee
Mr. Eduardo SOUZA-RODRIGUES, Referee
Mr. Helmuth CREMER, PhD Supervisor
Mr. Stefan AMBEC, PhD Co-supervisor
Mr. Mathias REYNAERT, PhD Co-supervisor
Mrs. Hélène OLLIVIER, President

Essays on the Economics of Law Enforcement

Alipio Ferreira

Ph.D. Thesis defended at Toulouse School of Economics

July 7$^{\text{th}}$, 2022

Thesis committee:

External members
Hélène Ollivier, Paris School of Economics, president of the jury
Lucie Gadenne, University of Warwick, report writer
Eduardo Souza-Rodrigues, University of Toronto, report writer

Thesis advisors
Mathias Reynaert, Toulouse School of Economics
Stefan Ambec, Toulouse School of Economics
Helmuth Cremer, Toulouse School of Economics

Acknowledgements

I wouldn't have finished this thesis without the support of people around me. Special thanks to my professors Mathias Reynaert, Stefan Ambec, Helmuth Cremer, who advised me in this journey, and to Nour Meddahi for his firm support. I also owe a lot of gratitude to my co-authors Anne Brockmeyer and Pierre Bachas. I thank my friends and family, in particular Rosa Coutinho, Esther Coutinho, Vanderly Menezes, Vatsala Shreeti, Ivan Salles, Matheus Bueno, Natalia Pacheco, Bruno Barsanetti, Marine Coinon, Jasmin Fliegner, and Alae Baha.

"bonum certamen certavi, cursum consummavi, fidem servavi"

Contents

Satellites and Fines: Using Monitoring to Target Inspections of Deforestation

Alipio Ferreira*

December 31, 2021

(Click here for the latest version)

Abstract

Effectively fighting deforestation requires monitoring of vast areas, which is possible thanks to satellite imagery. However, satellite monitoring can only reduce deforestation if three conditions are met: the monitoring alerts must be informative, the enforcement agency must use them to target inspections, and farmers must respond to enforcement action by doing less deforestation. This paper quantifies the contribution of real-time monitoring in deforestation reduction using detailed satellite and administrative data in the Brazilian Amazon forest. It studies the whole chain of events from the production of a deforestation alert to its effect on deforestation. It first documents an improvement in the monitoring system's ability to detect infractions in real-time. Then it estimates the impact that real-time alerts have on deforestation inspections. Finally, it estimates the impact of inspections on deforestation using an instrumental variable approach and an event study. Overall, the real-time alerts increase by three percentage points the inspection probability for offenders, avoiding approximately 450 square kilometers of deforestation per year.

*alipio.ferreira@tse-fr.eu, Toulouse School of Economics. I would like to thank Mathias Reynaert and Stefan Ambec for their supervision, and Bruno Barsanetti, Anne Brockmeyer, Pierre Bachas, Clara von Bismarck-Osten, Hippolyte Boucher, Sylvain Chabé-Ferret, François Salanié, and Vatsala Shreeti for their support and rich discussions. I also thank Fabiano Morelli (INPE), Ane Alencar (IPAM), Clarissa Gandour (CPI), Jair Schmitt (IBAMA), Luis Eduardo Pinheiro Maurano (INPE), and Francisco Oliveira Filho (IBAMA) for helping me understand the details of environmental law enforcement and satellite data. All errors are mine.

1 Introduction

Fighting tropical deforestation is a problem of paramount importance in environmental policy. More than 12% of global greenhouse gas emissions stem from forests destruction (IPCC 2014). Because standing forests absorb and store carbon, deforestation liberates the carbon back to the atmosphere through forest fires or biomass decomposition. In addition, deforestation destroys biodiversity (Fearnside 2021) and disturbs local rain seasons (Leite-Filho et al. 2021). While deforestation causes collective and diffuse harms, individual farmers reap private benefits from agricultural or timber exploitation of deforested areas. The tension between the individual benefits and the social costs of deforestation calls for governmental action such as forest protection policies. Nevertheless, enforcing these policies over vast forest areas can be daunting for enforcement agencies with limited resources and scarce information about deforestation hotspots.

One way to obtain systematic information about deforestation is by using satellite imagery. Thanks to high resolution satellite images, deforestation worldwide can be computed with a high degree of certainty, usually at a yearly rate (Hansen et al. 2013). Processing and interpreting images may take several months to be concluded (INPE 2019a), and cloudy weather may impair the visibility of forests during large parts of the year. The yearly measurement of deforestation is an invaluable asset to understand patterns and grasp the extent of the phenomenon. However, enforcement action against illegal deforestation may benefit from having real-time information on deforestation to target inspections. In this paper I answer the following question: how much forest does real-time satellite monitoring save?

Monitoring technologies can support enforcement agencies in targeting inspections to fight punishable offenses, such as deforestation. However, technologies' role in reducing the incidence of offenses depends on three factors. First is the quality of the information produced by the technology. For example, real-time data can fail to detect deforestation or produce false-positive signals, thereby misleading inspections. The second factor is the use made by enforcement agency of real-time information in its inspection selection decisions. In a best-case scenario, monitoring technologies provide new and accurate information, thereby changing the behavior of enforcement agents. However, in the worst-case scenario, monitoring information is redundant to other information sources already available to inspectors and does not affect inspection selection. Finally, the third factor is the impact of inspections on offenses. Offenders may be undeterred by inspections and consequently not change their behavior even in the presence of monitoring. In the end, monitoring technologies are only helpful if they affect inspection selection and reduce offenses.

To quantify the effect of a real-time monitoring technology on deforestation, I study three layers of the enforcement problem - monitoring quality, inspection selection, and offenders' behavior - in the context of the Brazilian Amazon forest. In this forest, almost all deforestation is illegal (Valdiones et al. 2021), and a single federal enforcement agency does most law enforcement action, inspecting and punishing offenders. In 2004 the Brazilian government launched a monitoring program to produce real-time deforestation alerts based on satellite images. The system is touted as a breakthrough in Brazilian environmental enforcement. However, it has not been thoroughly investigated in terms of its quality, impact on inspection selection, and deforestation.

I merge the yearly measurement of deforestation, the real-time monitoring alerts, and geo-referenced fines, and other geographical data, creating a balanced panel over the decade 2011 to 2020. As alluded previously, the yearly measurement is an accurate measure of deforestation, computed independently from the real-time monitoring technology or inspections. The ability to observe the degree of offenses is not always the case in other applications in the crime literature, where offenses are only observed if victims report them or if enforcement agents carry out inspections. The fact that deforestation is measured in an accurate way is a crucial asset to understand the quality of monitoring and the behavior of the enforcement agency in this paper.

To assess the quality of the deforestation alerts produced by the monitoring system, I overlay the maps of yearly deforestation with the monitoring alerts to the monitoring system's detection rate and its share of false-positive alerts. To my knowledge, this analysis is the first systematic and independent assessment of the quality of this monitoring system. The results show that the production of deforestation alerts by the monitoring system improved substantially in quality over the years 2011-2020. Furthermore, the comparison of real-time deforestation alerts with the yearly deforestation maps revealed a substantial improvement in detection rates, with a relatively low level of false positives. The increase in detection rate was due to improvements in the satellite image resolution and the technical capacity to monitor the images in real-time by experts.

Next, I study how the real-time alerts impact the behavior of the enforcement agency. I use a monthly-level event study to estimate the causal impact of a real-time deforestation alert on the probability of a fine. The results show that the enforcement agency explicitly uses the deforestation alerts to decide its inspections. Indeed, the inspection probability almost doubles when the agency receives a deforestation alert, while the inspection probability barely changes for other types of satellite alerts, such as fire alerts. Moreover, the share of alerts-driven fines in the enforcement agency's portfolio doubled in the period, reflecting a transformation in the inspection selection strategy with a more significant role for the monitoring system.

Finally, I estimate the impact of inspections on farmers' decisions to deforest. I decompose the behavioral responses of farmers into two parts: the effect on deforestation of changes in the inspection probability (general deterrence) and the effect of punishment over time (specific deterrence). The distinction between general and specific deterrence is well-known in the crime literature, but studies usually estimate either one or another.

To identify the general and specific deterrence effects separately, I use two different identification strategies. First, to identify the general deterrence effect, I exploit variation in the monitoring system's ability to detect deforestation in an instrumental variable approach. Cloud coverage blocks the view from satellites, making it impossible to generate real-time alerts, and therefore less likely to receive an inspection. The exclusion restriction is that cloud coverage only affects the incentives of farmers through its impact on the probability of an inspection. The results show that areas with more cloud coverage have less deforestation fines and show higher deforestation on the extensive (i.e., are more likely to have any deforestation) and intensive margin (i.e., deforest larger areas on average).

Furthermore, I used an event study design to estimate the specific deterrence effect, which enabled me to compute the dynamic effects of punishment several years after it happened. The two effects combined provide a complete picture of the effect of enforcement on deforestation. Inspected areas are 10% less likely to display any level of deforestation even three years after the inspection occurred. In addition, the area of forest fires decreases in line with the reduction in deforestation.

Besides deforestation, I also estimate an additional potential behavioral response of farmers to enforcement: the intensity of use of fire in deforestation. Farmers regularly use fires to clear vegetation, generating forest fires that spread to potentially large areas. Even when fires do not totally destroy the vegetation, they emit large amounts of greenhouse gases (Aragão et al. 2018, Silva et al. 2021) and can generate irreversible damage to the vegetation (Nepstad, Moreira, and Alencar 1999, Balch et al. 2015). I find that areas with higher inspection probability tend to be more intensive in their use of fire. These results suggest that despite the lower overall deforestation caused by inspections, farmers may change their technology of deforestation to try to escape punishment.

In summary, a one percent increase in inspection probability saves almost 150 square kilometers of forest or 2% of average yearly deforestation levels. In a conservative computation, the satellite increased the inspection probability by three percentage points every year for farmers, saving almost 450 square kilometers of forest per year and one thousand square kilometers in a decade. This number is an estimate of the value of the real-time monitoring system in terms of avoided deforestation. Moreover, the monetary value of avoided carbon emissions from deforestation are about 20 times as large as the opportunity costs of agricultural output in the Amazon forest. The benefits also far outweigh the budget of the monitoring and enforcement agencies.

1.1 Related literature

This paper contributes to four strands of literature: i) the effect of monitoring and enforcement on compliance, ii) inspection selection, and iii) tropical forest deforestation. It also adds to the growing literature in economics using geo-referenced satellite data to measure outcomes and identify causal effects (see Donaldson and Storeygard 2016 for a review)[1].

The literature on monitoring and audits has long highlighted the importance of information to induce regulatory compliance. Satellite-based monitoring programs had substantial positive impacts on compliance with air pollution environmental regulation in China (Greenstone et al. 2020) and US (Zou 2021), and in fighting deforestation in Brazil (Assunção, Gandour, and Rocha 2019)[2]. Nevertheless, the availability of information *per se* cannot explain compliance: monitoring can only affect incentives if the information is used to sanction offenders[3]. I contribute to this literature by studying the relationship between monitoring and enforcement, and then its impact on compliance. I perform the analysis at a precise geographical level, where deforestation, alerts and inspections are observed.

The most invaluable aspect of the datasets used is that it allowed me to separately observe deforestation, monitoring alerts and inspections. In several settings, the outcome cannot be observed independently of monitoring or audits, such as tax evasion. The independent measurement of deforestation allowed me to compute monitoring detection rates by overlaying the alerts maps with deforestation maps. As a consequence, it is possible to study with precision what causes detection rates to fail and how detection rates influence audit selection. Inspection selection is an important topic in the enforcement literature, which has been largely studied in the game theoretical literature (see Andreoni, Erard, and Feinstein 1998 for a review in tax compliance) but less so in the empirical literature. (Duflo et al. 2018, Kang and Silveira 2021, and Bachas et al. 2021 have shed light on the value of discretion in inspection selection. Blundell, Gowrisankaran, and Langer 2020 estimate the value of an "escalation" strategy in terms of compliance with environmental regulation.

Furthermore, this paper is unique to estimate both general and specific deterrence effects, the

[1]Examples range from tax compliance (Casaburi and Troiano 2016) to environmental economics, in particular tropical forest deforestation (Burgess, Costa, and Olken 2019, Assunção, Gandour, and Rocha 2019, Souza-Rodrigues 2019, among others) and forest fires (Balboni, Burgess, and Olken 2021). Alix-Garcia and Millimet (2021) provide a discussion of measurement error in the use of satellite data for deforestation studies.

[2]A parallel of monitoring can be also made with tax evasion, where the presence of third-party reporting also bridges the information gap between enforcement agency and taxpayers. Several papers have provided evidence of the role of third-party reporting in inducing tax compliance in Denmark (Kleven et al. 2011), Chile (Pomeranz 2015) and Brazil (Naritomi 2019).

[3]For example, in the issue of CCTV cameras, an extensive review by Welsh and Farrington (2009) has shown mixed evidence on their role in preventing crime. Ashby (2017) shows how the information produced by the cameras is effectively used to solve different types of crime, which helps explain the variety of effects of CCTVs on deterring crime.

latter meaning here the dynamic effects of inspections. The distinction between general deterrence and specific deterrence is well-known in the crime literature (Chalfin and McCrary 2017). General deterrence effects have been in the analytical framework of economists at least since Becker (1968), representing how agents internalize punishment probability in their decision-making. Effects of punishment probability on behavior has been estimated in urban crime (Levitt 1997, McCrary 2002), environmental (Chan and Zhou 2021) and tax compliance settings (Almunia and Lopez-Rodriguez 2018, De Neve et al. 2021), to name a few examples. Specific deterrence was first recognized as "incapacitation" effects of punishments such as imprisonment (Kessler and Levitt 1999, Kuziemko and Levitt 2004), but the concept has been applied to understand the effect of punishment on behavior more generally, also in environmental (Dusek and Traxler 2021) and in tax settings (Advani, Elming, and Shaw 2018).

The paper contributes to the deforestation literature by exploiting the satellite and administrative data in a novel way, and by studying the incentives to use fire in deforestation. This paper is the first to systematically use geo-referenced fines, logging alerts, and fire alerts in a single framework to explain patterns of enforcement and deforestation at a detailed geographical level. Assunção, Gandour, and Rocha (2019) has also studied the role of real-time monitoring on deforestation, using cloud coverage as an instrument for environmental fines at the municipality-year level. Building on that insightful work, I compute the detection probability of the logging monitoring system over time and show how this improvement has affected enforcement strategy and then deforestation patterns. Assunção, Gandour, and Souza-Rodrigues (2019) use logging signals directly as proxies of enforcement and show that they increase the probability of forest regeneration. Other papers have studied the impact of enforcement on deforestation by studying the policy of "priority municipalities" (Assunção and Rocha 2019 and Assunção et al. 2019), and incentive-based approaches to fight deforestation (see Jayachandran et al. 2017 for a study of initiatives in developing countries). Souza-Rodrigues (2019) discusses potential efficiency gains from moving to a more incentive-based approach, using a structural model of deforestation. The use of fire in deforestation has been almost ignored in the economics literature, a gap recently filled by Balboni, Burgess, and Olken (2021), who have shown that farmers take into account the risk of spreading accidents in their decisions to set fires.

2 Background: deforestation in the Amazon forest

The Brazilian Amazon is the world's largest rainforest, with 4 million square kilometers.[4]. As a rich repository for biodiversity, a regulator of local rain seasons, and the carbon concentration in the atmosphere, the forest provides vital local and global environmental services. Starting

[4]The total area of the forest is 6 million square kilometers. Besides Brazil, it spreads over Bolivia, Peru, Ecuador, Colombia, Venezuela, Suriname, Guyana, and French Guyana.

in the late 1980s, awareness about the environmental risks related to the destruction of the forest led to protective legislative action, investments in enforcement activity, and monitoring programs based on satellite data. The 2000s saw several policies centered on monitoring technologies, enforcement capacity, and punishment of offenders (for a historical overview, see Souza-Rodrigues 2014, Nepstad et al. 2014 and Assunção, Gandour, Rocha, et al. 2015). In 2004, the introduction of the satellite-monitoring system called "DETER" represented a breakthrough in the ability to inspect areas using real-time data on deforestation. DETER is the main source of deforestation alerts in this study, and I discuss it in more detail below. Other relevant initiatives that collaborated in reducing deforestation were the Soy Moratorium (Nepstad et al. 2009) and the policy of prioritizing municipalities for enforcement action (Assunção and Rocha 2019, Assunção et al. 2019). Deforestation has consumed approximately 20% of the original forest, although part of it has been recovered as secondary vegetation.

2.1 The process of deforestation

Deforestation is the complete clearing of vegetation from an area. Farmers clear forests to convert the land into agriculture or pasture, with timbering or mining as drivers of small-scale deforestation. While historically soybean culture has been the main driver of deforestation, since the mid-2000s, around 80% of deforested areas were converted to pasture for cattle grazing (Nepstad et al. 2009, Nepstad et al. 2014). Conversion of forest to agriculture or pasture is illegal in the Brazilian Amazon forest for environmental protection reasons. Therefore, the economic rationale for deforestation relies heavily on getting away with illegal deforestation, via lack of enforcement action, unclear property rights, and amnesties.

Deforestation occurs in three steps: selective logging, clearing vegetation, and cleaning remaining biomass (see INPE 2019a for a detailed description). In the first step, selective logging, farmers selectively cut valuable types of timber.[5] After extracting valuable timber, farmers clear trees and other vegetation using mechanized logging and fire. Farmers set fires in forest borders, letting it spread to the forest and damaging the vegetation. Damaged vegetation is easier to clear subsequently via logging activities. The third step usually consists in burning the remaining biomass, which is a technique to fertilize the soil with nutrient-rich ashes (Nepstad et al. 1999).[6]

In this highly humid area, fires do not emerge naturally. Instead, farmers set fires to clear vegetation as a preparation or sequel to logging. Fires aggravate the concerns involving deforestation because they introduce additional environmental risks. First, forest fires inflict irreversible damage on tropical vegetation, which lacks natural defenses against fires (Nepstad

[5] In the Brazilian Amazon forest, some valuable types of wood are *ipê*, *jacarandá*, and *mogno*.

[6] The practice of destroying and then burning vegetation is known as *slash-and-burn*.

et al. 1999, Gillespie 2021). Second, fires severely impair local air quality, with damaging effects on human health (see Reddington et al. 2015 for a study in Brazil and Sheldon and Sankaran 2017, Jayachandran 2009 for Indonesia). Thirdly, fires spread easily to neighboring areas, sometimes getting out of control in catastrophic ways. Forest fires damage vegetation, pollute the air, and emit greenhouse gases even when the areas are not ultimately logged. Official inventories of greenhouse gases often fail to account for forest fires because their methodologies focus on deforestation (Alencar, Nepstad, and Diaz 2006). Controlling fires is costly, consisting of building barriers and monitoring the fire.[7]

2.2 Enforcement by IBAMA

Deforestation is banned in the Brazilian Amazon, except for some particular circumstances. Regarding land tenure, 50% of the Amazonian area is indigenous territory (1.16 million square kilometers) or conservation units (1.2 million square kilometers). At least 13% (roughly half a million square kilometers, according to Azevedo-Ramos et al. (2020)) consists of public forests (also called "undesignated" public forests). It is forbidden to deforest in any of these areas. The remaining areas are privately owned rural properties and are mandated to preserve 80% of their area as forest. Only 2% to 4% of deforestation was legal in 2020, according to estimates by Azevedo et al. (2020) and Valdiones et al. (2021). The main legal instruments regulating deforestation in Brazil are the Criminal Environmental Law of 1998, the Forest Code of 2012, and a Presidential Decree of 2008. The use of fires is also tightly regulated in the region. Under authorization and following safety procedures, the law authorizes fires for agricultural purposes, but all forest fires are illegal. Penalties for farmers caught committing deforestation include high fines (about 1 thousand euros per hectare), seizure of equipment and goods, an economic embargo on the deforested land, and even imprisonment. The use of fire in deforestation is supposed to increase penalties. The law is quite severe against offenders but is not always enforced.

The federal enforcement agency, IBAMA, is the government body in charge of environmental law enforcement in the Amazon forest. Municipal and state authorities may play a subsidiary role in environmental law enforcement. Fighting deforestation is IBAMA's main activity in the Amazon region. In the decade from 2011 to 2020, IBAMA fined 23 thousand deforestation infractions, out of a total of 75 thousand environmental fines imposed by IBAMA in the Amazon region. IBAMA has access to real-time monitoring information on fires and logging and uses it to deploy enforcement personnel on the ground. There are 30 IBAMA units in the Amazon forest, from where enforcement agents leave to perform law enforcement field operations. Operations sometimes require the use of helicopters, as well as support from state police.

[7] In the year 2020 in the Brazilian "Pantanal", fires covered 3.9 million hectares during the months of July and August, which represents 26% of the total area of the biome (Leal Filho et al. 2021).

2.3 Satellite systems

Satellite systems *measure* and *monitor* deforestation in the Brazilian Amazon. The *measurement* of deforestation takes place once a year (see INPE 2019b for a technical description). Using images at a 30m x 30m resolution, the Brazilian National Institute for Spatial Research (INPE) categorizes the land cover entire territory of the Amazon forest as native forest, deforestation, water bodies, or clouds. This system is the source of the official measurement of yearly deforestation in Brazil. The measurement takes place once a year at the end of July, which is when clouds are very dispersed, maximizing visibility. Processing the data takes six to eight months to be concluded (INPE 2008). The result is a complete map of the Amazon forest with the land cover corresponding to late July. The yearly measurement is not a suited for real-time monitoring, since it only measures deforestation once a year and takes several months to be published.

The main tool for monitoring deforestation in the Amazon is the system DETER. Launched in 2004, DETER sends daily deforestation alerts for the enforcement agency.[8] The monitoring program DETER produces deforestation alerts based on rapid degradation of forest ceilings. Degradation can be the result of fires, but the deforestation alerts do not capture active fires. In practice, it captures situations of natural forest degradation, fire-induced degradation, and also active *logging* the forest. DETER has been a major breakthrough in law enforcement in the Brazilian Amazonia, but its ability to detect deforestation with deforestation alerts was relatively low in the early years, with a large number of false positives. In the period used in this paper, the decade of 2011 to 2020, the program progressed substantially in its capacity to flag deforestation areas correctly in real time, as documented later in this paper. DETER also started distinguishing alerts for different types of events on the ground. Today, besides the deforestation alerts, DETER produces alerts for forest degradation, mining, selective logging, and fire scars.

The monitoring system DETER uses essentially the same methodology as the yearly measurement system PRODES (INPE 2019a), but uses higher-frequency, lower resolution images. The production of alerts is made by technicians at the National Institute for Spatial Research, and is not automatized. The technicians use computers to exclude areas covered by clouds and areas that were already previously deforested, based on the measurement system PRODES. From this stage, the technicians monitor the images of the whole Amazon, aided by estimates of land cover at each pixel done via a Linear Spectral Mixing Model (Diniz et al. (2015)). The technicians in charge of monitoring the forest and producing the alert are independent from

[8]DETER initially based on images from the satellite Terra, and since 2017 using images from the satellites CBERS-4 and IRS. Terra is a NASA satellite, CBERS-4 is a Chinese-Brazilian satelliten and IRS stands for Indian Remote Sensing Satellite. To distinguish from its first phase (2004-2017), the program is now named DETER-B

the enforcement agency, and there is no prioritization of monitoring areas in case of shortages of personnel or computing capacity.

Another useful satellite-based system to monitor deforestation produces real-time reports of active fires. The satellites Terra and Aqua produce daily fire alerts at a resolution of 1 square kilometer since the early 2000s.[9] An important improvement in fire measurement was the introduction of the NPP satellite in 2013. This satellite is equipped with the VIIRS sensor, which is able to detect fires at 275m resolution.

2.4 Data

The four main datasets are from three satellite systems managed by the Brazilian Spatial Research Institute (INPE) and the data on fines, namely:

1. the maps from soil coverage system (PRODES), updated yearly
2. the daily geo-referenced fire signals from fire alerts monitoring system ("Queimadas")
3. the maps of deforestation alerts from the monitoring system DETER, published monthly
4. the administrative dataset of fines from IBAMA

I restrict the dataset to fines related to deforestation of native forests in the Amazonian biome using a string search on the free description of the fines typed by inspectors. More details on the classifications of fines can be found in the Online Appendix. I use the geographical coordinates of deforestation fines to locate the enforcement action at a precise area and link it to measured outcomes. These four datasets can be visualized in the set of figures[10] 3a to 3d. Figure 3a shows the categorization of the land coverage by PRODES as forest, old deforestation, and new (i.e., "last-year") deforestation. Figure 3b overlays this the soil coverage with the fire locations. Figure 3c shows the areas of logging signals in yellow. Finally, Figure 3d shows the points where inspectors produced a fine.

To overlay the maps of soil coverage, logging alerts, and fire alerts in the whole Amazon forest, I rasterized the entire area into 300m level squares. I also added more information at this level, such as the administrative divisions of the Amazon into municipalities and the legal status of the land - private property (from the official rural registry CAR), indigenous land, conservation units, or others. Therefore, at a 300m level of precision, there are several layers of merged information. I then aggregate information at the 15km x 15km cell level.

[9]Both Terra and Aqua are satellites owned by NASA. Terra was launched on December 18,1999 and Aqua on May 4, 2002. Both satellites are at an altitude of 705 kilometers above sea level, and complete an orbit of the Earth in 98 minutes, orbiting every area on the planet twice a day.

[10]To produce these figures, I took an area of approximately 30 thousand square kilometers, corresponding to a "scene" of the Landsat satellite, in the northern state of Pará in the year 2016.

The 15km x 15 km cell level is the observational unit used in this study. It roughly corresponds to splitting the Amazon forest into 20 thousand equally-sized squares. I include information on enforcement action at the cell level instead of matching the fines' coordinates with the exact locations of the polygons of deforestation or alert. I also compute some other variables at the (15km x 15km) cell level, such as i) the distance from each cell to each of the three main cities of the Amazon: Manaus, Cuiabá and Belém, ii) the presence of state roads in the cell (binary variable), iii) the presence of federal roads in the cell (binary variable), and iv) the shortest distance from each cell to a federal road, v) the accumulated share of deforested area in that cell-year, and vi) the size of the forest frontier in the cell[11].

3 Assessing the quality of monitoring alerts

In this section I analyze the quality of the information produced by the monitoring system DETER, launched in 2004 by the Brazilian government and run by the Brazilian Institute for Space Research (INPE). The program uses satellite images to produce real-time alerts of rapid vegetation loss in the Amazon forest, called deforestation alerts, shared daily with the enforcement agency IBAMA to support its inspection decisions. However, the potential of DETER deforestation alerts to improve inspection selection depends on how informative they are of actual deforestation.

The production of alerts depends on detected color changes of pixels in the satellite pictures of the Amazon forest. In simple terms, pixels that turn quickly from green to red or brown produce an alert (INPE 2008). Indeed, the destruction of primary vegetation would generally present this pattern. Unfortunately, however, in its early years, the system produced many false-positive alerts and a staggeringly low rate of deforestation detection. In this section, I evaluate the quality of the deforestation alerts in the period of 2011 to 2020. I do that by overlaying the maps of real-time alerts with the yearly maps of measured deforestation.[12]

Thus, I can compute the share of deforested areas that were detected by DETER in the corresponding year (formally $\mathbb{P}(\text{deforestation alert}_t|\text{deforestation}_t)$, or sensitivity rate) as well as the share of deforestation alerts that were truly deforestation areas ($\mathbb{P}(\text{deforestation}_t|\text{deforestation alert}_t)$). A perfect system would have 100% for both these indicators, meaning that the system detects all deforestation in real-time and all alerts are correct. In practice, DETER failed to detect many deforested areas and produced a large share of false positive alerts in non-deforested areas. False positives occur when there is forest degradation but not a complete clearing of vegetation. Over time, the researchers in charge of the monitoring system adjusted alerts'

[11]By forest frontier I meant the border between native forest and already deforested area.

[12]The measurement of deforestation occurs in the program PRODES, once a year, at high resolution. For more details, see section 2

production, improving their detection rate and accuracy.

The constant work of validation and reinterpretation of images substantially improved the quality of the deforestation alerts. Figure 4a depicts the share of an alert in deforested areas, computed by overlaying the maps (or "polygons") of deforestation alerts from DETER with those of deforestation from PRODES. Of all the deforested areas, only about 10% of them received a deforestation alert in 2011, and this share has gradually increased, reaching more than 40% in 2020. The figure also shows that despite switching satellites in 2017, the improvements were not discontinuous. Therefore, the share of false negatives is significant, with most deforestation going undetected by DETER. Figure 4b, on the other hand, shows the probability that a deforestation alert is correct, meaning that it points to an area with deforestation. Like the previously analyzed probability (that is, the probability of detecting deforestation), the probability of the alert being correct is increasing over time and reached 70% in 2019. As mentioned above, the false positives are mainly due to misclassification of natural forest degradation as deforestation.[13]

Failure to detect deforestation is due to three critical technical difficulties: i) the forest is not always visible to satellites due to cloud coverage, ii) small deforestation areas may go undetected due to poor satellite resolution, and iii) prolonged deforestation processes generate slow changes in forest color. The probability of detection is the probability of an alert being produced in the case of an area being deforested. To understand the factors that affect it, I propose the following linear model:

$$\begin{aligned}\mathbb{P}(\text{def. alert}|\text{deforestation})_{it} = \beta_0 + \beta_1 &\underbrace{\text{size\_deforestation}_{it} + \beta_2 \text{share\_fire}_{it}}_{\text{farmer's decision}} \\ &+ \underbrace{\beta_3 \text{cloud}_{it}}_{\text{weather}} + \underbrace{\tau_{it}}_{\text{technology}} + \epsilon_{it}\end{aligned} \tag{1}$$

This model posits that mainly three types of variation affect the probability of an alert being correctly produced. Two of them are exogenous to the farmers' decisions - weather and technology -, whereas the size of deforestation is a choice by farmers. The cloud variable was proposed by Assunção, Gandour, and Rocha (2019) as an approximation of monitoring quality. However, it is only an indirect measure of the monitoring quality, disregarding technological progress.

Table 3 shows the results for the estimation of the detection probability of a polygon of defor-

[13] INPE (2008) did a similar validation exercise by comparing overlaying alerts to deforestation maps, but with a slightly different methodology: they considered an alert as correct if the alert's polygon had some overlap with a deforestation polygon, without taking notice of the size of the area that overlapped versus the area that did not overlap. The result is a lower rate of false negatives than the one I computed.

estation. The regression is run at the polygon level, meaning that each unit is an independent deforestation area in a given year. The left hand side is a binary variable that takes value 1 if there was a deforestation alert that overlapped with the deforestation polygon, even if the overlap was only partial, and 0 if there was no overlap at all. The regression shows that large polygons are much more likely to be detected, and that the use of fire increases substantially the probability of detection. The reason for this latter result is that fires damage the vegetation rapidly, which increases the likelihood that the monitoring system detects a rapid vegetation loss and flags the area as potential deforestation. Finally, as expected the average cloud coverage reduces the detection probability.

The substantial improvement in the accuracy of deforestation alerts is the main source of variation used to study the impact of monitoring on the enforcement system in the Amazon forest. My measure of quality of the deforestation alert is the share of deforested areas that were detected in real time, or $\mathbb{P}(\text{deforestation alert}_t|\text{deforestation}_t)$. That is the indicator that increased four-fold in the years 2011-2020, from 10% to 40%. This increase means that farmers in 2020 had a 40% chance of being observed in real-time when they decided to deforest some areas with logging. This indicator varies across regions and across time, for the reasons mentioned above: differences in cloud coverage, local geographical conditions, and INPE's capacity to classify images correctly. The variation in the detection probability is therefore a function of i) natural phenomena and ii) technological capacity, both of which are exogenous to the farmers' decisions to deforest. However, DETER is also better suited to detect larger areas of deforestation than smaller ones. The improvements in detection probability are visible for all sizes of deforestation. Figure 4c shows the share of deforestation polygons detected by the monitoring system for each decile of polygon size.

4 Monitoring and inspection selection

How does the Brazilian environmental agency use satellite alerts to decide which areas to inspect? I answer this question using geo-referenced data on fines and deforestation alerts at the monthly level. This section aims to quantify the importance of deforestation alerts in Brazilian enforcement action against deforestation. It is unclear to which extent real-time monitoring alerts have an effect on enforcement action since the enforcement agency can also carry out inspections in the absence of satellite inspections, based on helicopter surveillance, denunciations by citizens, regular patrolling, or other types of non-coded information. In Brazil, the real-time monitoring system DETER is touted as a breakthrough in enforcement, and here I assess how much of IBAMA's enforcement action are caused by it.

In this section, I compute the probability of a fine in areas with deforestation, and decompose this probability to how much of it is caused by real-time deforestation alerts. To do that, I use geo-referenced information on fines, true measured deforestation, and the satellite-based deforestation alerts at the 15km x 15km cell level. To estimate inspection selection, I restrict the sample to areas with positive levels of deforestation, that is, non-compliant areas. Restricting the data is necessary to interpret variation in the fines as variation in inspection efforts. Fines only reflect enforcement in areas that are "eligible" for them, that is, areas with positive levels of offenses.[14]. Among non-compliant cells, observed variation in fines can be interpreted as variation in enforcement action. I describe how the overall yearly probability of inspection rises in non-compliant areas, when satellites produce logging or fire alerts in the same areas. Next, I use monthly data to estimate the causal impact of alerts on enforcement action probability, using an event study approach. I then discuss the value of following real-time alerts as opposed to random fines.

4.1 Computing the fine probability

The probability of inspections in the Amazon forest in the 2011-2020 was 13% in the decade from 2011 to 2020. This means that conditional on having a positive level of deforestation in a given year, a 15km x 15km cell had a 13% probability of receiving at least one deforestation fine in the same year. In principle, deforestation can be punished at later dates, even years after the offense has been committed. However, this seems to be rare: more than 80% of deforestation fines by IBAMA happen in areas that have positive new levels of deforestation in the same year (see Appendix figure A1), and yearly additional deforestation seems to be beyond what IBAMA is able to inspect every year, given that only 13% of cells with positive deforestation received a fine.

This probability hides a lot of heterogeneity. Cells that are close to IBAMA's offices, cells that deforest larger areas, or cells that receive real-time deforestation alerts are more likely to be fined. I estimate the following linear probability model to understand the factors which are correlated with fine probability:

$$\begin{aligned} \mathbb{P}(\text{fine}_{it}) = \beta_0 + \beta_1 &\text{deforestation alert}_{imt} + \beta_2 \text{fire signal}_{it} \\ &+ FE_i + \delta_t + \gamma X_{it} + \varepsilon_{it} \end{aligned} \tag{2}$$

where ε_{it} is a cell-year idiosyncratic error term, assumed to have a conditional mean zero. FE_i are cell fixed effects and δ_t are month dummies. X_{it} is a matrix of controls such as distances

[14]Formally, the probability of a fine in any given cell i and period t is $\mathbb{P}(\text{fine}) \equiv \mathbb{P}(\text{inspection \& deforestation}) = \mathbb{P}(\text{inspection}|\text{deforestation})\mathbb{P}(\text{deforestation})$ Using data on fines to infer enforcement action, compliant areas (i.e., areas with zero deforestation) become useless to understand the behavior of the enforcement agency, since fines are trivially equal to zero in these areas.

to three main cities (Manaus, Cuiabá and Belém), distances to the closest IBAMA office, prices of commodities and IBAMA's budget expenditure. Some specifications also include municipality fixed effects and year fixed effects. The regression is estimated with different samples, including a sample with only areas with positive deforestation. All variables are binary, including deforestation (1 if there was positive deforestation) and alerts, except for the controls and unless specified otherwise. The coefficients of interest are β_1 and β_2, which reflect the additional probability of enforcement given the occurrence of a logging or fire alert, relative to no alert. The main specifications are estimated only for the sample of cells with positive deforestation, that is the areas "eligible" for fines.

The table can be analysed for descriptive purposes but is unlikely to yield causal estimates of the different factors on fine probability. The OLS results can be seen in Table 5. The probability of an inspection increases by almost 8 percentage points in areas with positive deforestation (Column 1), and 1.5 percentage point if there is fire. Columns 2 and 3 include a dummy for whether both fire and deforestation alerts are observed in the same year, still conditional on "same year deforestation". Almost all cells with a deforestation alert also presented some degree of forest fires, even though the exact overlap of areas is rare. The interaction coefficient therefore captures almost the full effect of deforestation alerts, and makes the effect of forest fire alerts negative but not statistically significant. The other Columns change the sample in which the model is estimated. In Column 4 only priority municipalities are selected. These are municipalities declared as high-priority by the enforcement agency itself. The effect is strongest for this sample, with deforestation alerts increasing by 14 percentage points the probability of a inspection, although the effect of forest fires is still around 1.5 percentage point. Column 5 has all cells that presented some year of positive deforestation in the 2011-2020 period. The effects if alerts are understandably weaker, since alerts may lead the enforcement to areas where there is no deforestation, such that no inspection would be observed. The same is the case in Columns 6 and 7, which include all data, including cells with no deforestation whatsoever. It cannot be ruled out that "false positives" led to inspections, but these would be unsuccessful and not appear in the dataset. This explains why the effects are attenuated once we account for all alerts, including in areas where no deforestation took place.

Clearly there are several factors which influence fine probability, and the real-time monitoring system that produces deforestation alerts. Below I propose a strategy to estimate the causal effect of the real-time deforestation alerts on fine probability, which allows me to understand the contribution of this technology to the enforcement action in the Amazon forest.

4.2 Estimating the causal effect of alerts on fine probability

Identification

As discussed in previous sections, satellites produce several types of alerts to the enforcement agency, and especially fire and deforestation alerts. While deforestation alerts are observed at a month-cell level, they are not an exogenous event, and it is not possible to infer causal effects immediately from a regression of fines on alerts. The reason is that areas that receive alerts are more easily observed by satellites, particularly because they have less cloud coverage and present larger areas of deforestation. Therefore, I propose a differences-in-differences identification strategy to estimate the causal effect of alerts on fines. This strategy relies on the trends of fines in different areas, and identifies as a causal effect any deviation from parallel trends which follows from an alert.

Exploiting the panel dimension of the data, it is possible to recover the average treatment effect on the treated (i.e., the cells which received alerts) by the evolution of the number of fines before and after alerts with the evolution of fines in the same period for cells that did not receive any alert. This is the differences-in-differences approach. This strategy identifies the average treatment effect on the treated under two mains assumptions. The first one is "parallel trends", meaning that in the absence of alerts, the number of fines would evolve on average the same way for cells with and without alerts. The second one is "non anticipation", which means that the observed outcomes previous to the alert can be interpreted as untreated outcomes.

Estimation

The analysis is done in the form of an event study at the month-cell level. I pool every cell at the monthly level to create a balanced panel of cell i and month t. Furthermore, I only consider cells which have displayed positive levels of deforestation at some point in the decade, because these are areas where a fine could be produced. I then estimate the following regression:

$$P(\text{inspection})_{it} = \sum_{\ell=-6,\ell\neq-1}^{12} \beta_\ell \mathbb{1}\{t - e_i = \ell\} + \delta_t + FE_i + \varepsilon_{it} \tag{3}$$

where ε_{it} is a conditional mean zero error term, and e_i is the month of the an alert event within cell i.[15] $\mathbb{1}\{t - e_i = \ell\}$ is an indicator function that takes value 1 when the period t is ℓ months distant from the event date e_i. The set of all $\mathbb{1}\{t - e_i = \ell\}$ is a matrix containing binary vectors that refer to the lags and leads relative to the alert date.

As mentioned above, only observations which presented positive deforestation were included. Therefore, this estimation captures the effect of an alert in spurring enforcement action in an area which is "eligible" for fines, with or without monitoring alerts. Indeed, many cells had

[15]Deforestation usually spans over several months, and one single area may present several successive alerts. For this reason, I only consider as an "alert event" only the alert that takes place after four months without alerts in the same cell.

positive levels of deforestation but did not have deforestation alerts, making them a group of comparable "never treated" cells. Among treated cells, the treatment date varies from one place to the other, partly because deforestation happens in different moments in time, or because cloud coverage delays detection of deforestation by the monitoring systems.

Estimation of equation 3 by OLS can identify the average effect of alerts on the probability of fines under some strict assumptions. Indeed, omitting the first lag ($\ell = -1$) in the estimating equation means that each β_ℓ is a weighted average of all differences-in-differences parameters. Normally, the differences-in-differences strategy identifies the average treatment effect on the treated under the assumptions of parallel trends and no anticipation (see Wooldridge 2021 for a detailed discussion). However, as highlighted by a recent literature (see Callaway and Sant'Anna 2020, De Chaisemartin and d'Haultfoeuille 2020, Goodman-Bacon 2021, Borusyak, Jaravel, and Spiess 2021, Sun and Abraham 2021), OLS estimation of equation 3 makes potentially invalid differences-in-differences comparisons, in the sense that they subtract values of outcomes that may include treatment effects, even when the parallel trends and no anticipation assumptions are true. This happens in particular when estimating treatment effects in settings in which the treatment date varies across groups, as is the case here. In short, one should be careful not to compare treated observations with other treated observations.

I first estimate the event study in equation 3 using OLS, and then I estimate the model using a method robust to biases stemming from problems of staggered designs. To overcome these problems, I follow the approach suggested by Borusyak, Jaravel, and Spiess (2021), which the authors name an "imputation method". The method consists of three steps. In the first step, I estimate the time and cell fixed effects (i.e., the two-way fixed effect model) only using non-treated observations (the union of "never-treated" and the "not yet treated" observations). This yields cell-month specific estimates of the untreated value of the outcome. Then, I extrapolate these estimates to the remaining part of the sample (the sample of observations after treatment has taken place), which is essentially a prediction of individual counterfactuals. Finally, I compute the average treatment effect as the average difference between the realized values of the outcome (fines in this example) and the imputed counterfactual. This procedure avoids making invalid comparisons with cells that have already been treated in the past, thus yielding a meaningful estimate of the ATT in the sense that it is a convex combination of the individual treatment effects.[16]

[16] Additionally to the problem of negative weights, Borusyak, Jaravel, and Spiess (2021) warn that the absence of groups which are "never treated" in the analysis created an identification problem, in which the time fixed effects cannot be identified from an alternative model in with time trends. The treatment effect estimates with this method happen to be simply the difference between observed outcomes and predicted counterfactuals. This means that there is no "error" term estimated next to the treatment effect, raising the question of how to estimate the variance of the estimator. As shown in Borusyak, Jaravel, and Spiess (2021), the variance of the estimator relied on the variance of these individually computed treatment effects, as well as on the error term o the estimation of the two-way fixed effect model (which is done with only the untreated part of the sample).

Results

The OLS results are shown graphically in Figure 5b. The results for the estimation using the method of Borusyak, Jaravel, and Spiess (2021) are in Figure 5c. In this case, they are qualitatively and quantitatively very similar to the OLS results, suggesting that the problems related to staggered designs are not severe in this particular application. They all show a strong and immediate effect of alerts on the probability of a fine in the cell where the alert was produced. As soon as the alert appears, the cells with the alert become immediately one percentage point more likely to receive an inspection, and then two points more likely in the two months after the alert. The effect fades out over time and disappears after nine months. The fact that the effect is never negative means that the effect of the alerts is not merely an anticipation of fines which would take place anyways later in time. Alerts produce *additional* fines which would not have taken place otherwise.

Placebo tests

Table 8 summarizes the OLS results and includes other specifications testing the effect of other real-time satellite information as placebo tests. Column 1 shows the effect of the occurrence of a deforestation alert on the probability of an inspection. Prior to the occurrence of the alert, there is no difference between the enforcement probability of areas that received an alert and areas that received no deforestation alert, despite having positive amounts of deforestation. Column 2 and Figure 6a show the effect of real-time alerts of forest fires on the probability of fines. They suggest a strong correlation between forest fires and inspection probabilities. However, there are clear differences between areas with or without fires prior to the first alert that the forest is burning, such that the differences between areas with an alert and without an alert cannot be causally attributed to the alert. When one considers any fire alerts, including fires that started outside of forest, as in Column 3 and Figure 6a, there is a more compelling case to suggest that fire alerts lead to enforcement action, but again the differences arise prior to the first alert. These correlations are driven by the fact that fires often happen in the process of deforestation, such that many of the fire alerts are probably happening in the proximity of areas with deforestation alerts, which were shown to have a strong effect on enforcement. In fact, fire alerts bear really no weight in the decision of the enforcement agency: when comparing areas with deforestation and fire versus areas with deforestation with no fire, excluding all areas that also had a deforestation alert, a flat curve appears (Figure 6b)

Other placebo tests can be done using other types of real-time alerts produced by satellites, but which are unrelated to large scale deforestation, such as "selective logging" alerts (*desmatamento seletivo*) and "mining alerts" (*mineração*). Though these activities also encompass destruction of forest, they do so at a smaller scale than deforestation aimed at converting forest to pasture or agriculture. As a consequence, these alerts should not have any effect in altering probability of inspection for deforestation, and can be used as a placebo test to verify whether the effect observed for deforestation alerts is really specific to that kind of information. Indeed, that is clearly what is observed in Columns 4 and 5 of Table 8 and Figures 7a and 7b. In summary, only deforestation alerts have a causal effect on enforcement action, with fire alerts being correlated but not causing increases in inspection probability.

4.3 Decomposing the effect of monitoring in the fine probability

The probability of inspections for offending cells can be decomposed as follows, using the Law of Total Probability:[17]

$$\mathbb{P}(\text{fine}) = \underbrace{\mathbb{P}(\text{fine}|\text{no alert})\mathbb{P}(\text{no alert})}_{\text{probability without alerts}} + \underbrace{\mathbb{P}(\text{fine}|\text{alert})\mathbb{P}(\text{alert})}_{\text{probability with alerts}}$$

The objects in this expression are easily computed from the data, and as already mentioned, the probability of fine in the Amazon forest in the studied period was 13%. The probability of a fine (i.e., a positive number of fines in the year) in areas that receive an alert was 22%, but this value is not the causal effect of alerts. Indeed, this probability is decomposed in a baseline level of fines in areas that receive alerts, which I denote $\mathbb{P}(\text{fine}(0)|\text{alert})$ borrowing from the potential outcomes literature, and the average causal effect of alerts, denoted ATT:

$$\mathbb{P}(\text{fine}|\text{alert}) = \underbrace{\mathbb{P}(\text{fine}(0)|\text{alert})}_{\text{baseline/counterfactual probability}} + \underbrace{ATT}_{\text{causal effect of alerts}}$$

This decomposition allows us to understand how much of the overall enforcement action can be attributed to the alerts. These fines are "additional" to the baseline inspections, which would have occurred regardless of the alerts. To understand the role of real-time monitoring in the overall enforcement risk, I aggregate the monthly causal effects estimated in the event study in the previous section up to the nine-th month after the event date, which is when the average effect seems to disappear. This allows me to understand what is the share of fines in areas with alerts which were effectively caused by the alerts, which is easily computed by dividing the alerts-caused fines by the total fines in areas that had alerts. The share of fines caused by alerts is $s \equiv \frac{\sum_{\ell=0}^{8} \beta_\ell \times \#\text{alerts}}{\#\text{fines in areas with alerts}} \approx 1/3$. This share is computed using the monthly data estimation, and is a useful tool to translate the results into yearly data.

[17] All expressions below are conditional on the cells having positive deforestation, that is, the cell is an "offending cell".

Using yearly information (the level at which deforestation is measured), we know that areas with deforestation *and* alerts had a 22% probability of receiving at least one fine. A fraction $s \approx 1/3$ of the fines is due to real-time monitoring, and I use this fraction to apportion the part of the 22% yearly fine probability to real-time monitoring. This is an approximation, but it seems to be the most natural way to apportion the probability that an areas gets fined in a given year using monthly level estimated ATTs. The consequence is that out of the 22% yearly probability of fine for areas with deforestation and alerts, 7 percentage points are due to the real-time monitoring system, and 15 percentage points are the baseline probability, captured by cell and month fixed effects.

$$\underbrace{\mathbb{P}(\text{fine})}_{\text{13\%}} = \underbrace{\mathbb{P}(\text{fine}|\text{no alert})}_{\text{6\%}}\underbrace{\mathbb{P}(\text{no alert})}_{\text{60\%}} + \left(\underbrace{\mathbb{P}(\text{fine}(0)|\text{alert})}_{\text{15\%}} + \underbrace{ATT}_{\text{7\%}}\right)\underbrace{\mathbb{P}(\text{alert})}_{\text{40\%}}$$

The contribution of the real-time monitoring system to the probability of fine is captured by the last term, which multiplies the ATT by the probability of having deforestation and alert. Notice that not all fines happening in areas with alerts are deemed additional. To a great extent (15%), IBAMA would be able to impose fines on farmers in those areas, even in the absence of alerts. The decomposition of the probability reveals the following: in the decade 2011 to 2020, a cell with positive deforestation had a 13% probability of receiving at least one fine in the same year of deforestation, and the part that is due to the monitoring system is approximately 3 percentage points ($ATT \times \mathbb{P}(\text{alert}) = 7\% \times 40\%$). This represents a substantial amount of the overall fine probability, especially given that it is the part that is due to a single source of information: the real-time monitoring system DETER.

The results of this decomposition exercise can be seen in Figure 9a. In the next section I estimate the impact that this 3 percentage point increase in fine probability has on deforestation reduction.

4.4 Mechanism: why following real-time alerts matters for enforcement

Should IBAMA be concerned with real-time monitoring and quick reactions to alerts, as it seems to be? Deforestation is an offense that endures: once it has taken place, it stays. In any case, the offenses are observed by satellite once a year (via the yearly satellite measurement of PRODES) and become known to the enforcement agency. So why not wait and punish the farmers later? IBAMA can go to the place where deforestation took place and punish agents for exploiting an area economically that was illegally deforested. But in practice such late interventions tend to be less likely to succeed, and in particular less likely to inflict costs on offenders. IBAMA agents must find the offender and establish the link between the offense and its author.

The analysis of the timing of fines allows for a comparison between fines that followed alerts and those that did not. Fines that follow alerts up to three months after the occurrence of an alert, or "timely fines", differ from "random fines" in two important dimensions: timely fines tend to punish much larger areas, and are more likely to seize equipment from offenders. These two differences can be seen by estimating the following simple regression model:

$$\begin{aligned} \text{fine characteristic}_{imt} = \beta + \alpha_0 \text{deforestation alert}_{imt} + \alpha_1 \text{deforestation alert}_{imt-1} \\ + \alpha_2 \text{deforestation alert}_{imt-2} + \alpha_3 \text{deforestation alert}_{imt-3} + FE_m + \delta_t + \varepsilon_{it} \end{aligned} \tag{4}$$

where i is the single fine, m is the municipality and t the month. ε_{it} is a conditional mean zero error term, and FE_i is a fixed effect at the municipality level (a level above the cell level) and δ_t stands for month effects. The model is estimated using two outcomes: the share of fines that ended seizing equipment from the offenders, and the size of the deforestation offense, in hectares. Information from seized equipment is obtained from a separate administrative dataset of IBAMA, and merged with the individual fines. Information about the size of deforestation is extracted from a string description of the fines, and in some cases filled explicitly by inspectors in a separate field. Table 9 shows the α coefficients for these two outcomes. Relative to fines that followed no alert or a alert more than four months old, "timely fines" are different across the two characteristics.

Columns 1 to 3 of Table 9 show that the probability of seizing equipment is increased by 1.5-2 percentage points if the fine takes place in the same month of the alert. The older the alert, the lower this probability, and the effect is even negative if the alert is three months old. Adding month fixed effect (column 2) or restricting the sample to priority municipalities (column 3) do not change the effects. The effect is positive and significant as long as the fine occurs in the same month as the alert. Although the effect may seem small, it represents a 20% increase relative to the baseline probability of 8% of a fine seizing the equipment of the offenders.

Regarding the size of the offense, the fines that follow alerts are larger than other fines by around 40 hectares on average, as can be seen in columns 4-6 of Table 9. If the alert happened more than three months before the fine, the difference is much smaller, 16 to 24 hectares larger than fines that did not follow a deforestation alert. The explanation for this large and persistent effect is that alerts are more likely to be produced for larger infractions. As a result, fines that follow alerts tend to go for the larger offenses as well, which is another potential benefit of using monitoring alerts as a rule for deciding where to deploy enforcement.

In this section I showed that the presence of real-time monitoring alerts for logging leads to increased probability of an inspection in an area. Moreover, the increased quality of these alerts has been followed by an increased reliance by IBAMA on these alerts. Fire alerts, on the other hand, play no substantial role in determining enforcement action. In the next sections I estimate the impact of enforcement on overall compliance (the decision to deforest or not) and then on the choice to use fire in deforestation.

5 Farmers' responses to inspections

Inspections are valuable to the extent that they affect farmers' decisions to deforest. The classical model of crime in economics, first proposed by Becker (1968), posits that agents decide whether to commit a crime based on the probability of punishment. In this model, the credible risk of punishment is enough to deter agents from engaging in unlawful activities. This effect came to be known in the crime literature as "general deterrence" effect. In the context of deforestation, this effect would translate to farmers refraining from deforestation when the probability of being caught is high enough.

Besides the general deterrence effect, another way enforcement can deter crime is by affecting the future behavior of punished agents. One classic example is imprisonment, which incapacitates agents from committing a crime, reducing the future crime incidence. The effect of punishment itself on agents' future behavior is known as "specific deterrence effect". In the context of deforestation, this would be captured by agents' behavior after punishment.

The general and specific deterrence effects are theoretically different and can be thought of as "ex ante" and "ex post" effects of punishment. Understanding the full impact of enforcement on agents' behavior requires accounting for both behavioral responses. To do that, I propose a simple framework to understand how they interact.

Formally, call p_t the probability of inspection in year t, $N(p_t)$ the resulting number of offending farmers, $d_t(p_t, f)$ the average deforestation areas by offending farmers, which is a function of the probability of inspections p_t and the history of inspections f. The variable f codes whether the farmer was inspected in period 0. The share p_t of farmers who have been inspected deforest less up to three years later,[18] whereas those who have not been inspected continue deforesting as before. The four year accumulated deforestation is:

[18] Three years is an arbitrary time horizon.

$$D = \sum_{t=0}^{3} N(p_t) d_t(p_t, f)$$

Suppose there is a marginal increase in p_0 (the inspection probability at period 0), lasting only one period. Then the impact of this marginal increase on a four-year period of deforestation D is:

$$\begin{aligned}
\frac{dD}{dp_0} &= \sum_{t=0}^{3} \frac{dN}{dp_0} d_t(p_t) + N(p_t)\Big(\frac{\partial d_t}{\partial p_0} + \frac{\partial d_t}{\partial f}\frac{df}{dp_0}\Big) \\
&= \frac{\partial N}{\partial p_0} d_0(p_0) + N(p_0)\frac{\partial d_0}{\partial p_0} + \sum_{t=1}^{3} N(p_t)\frac{\partial d_t}{\partial f}\frac{df}{dp_0} \\
&= \underbrace{\frac{\partial N}{\partial p_0} d_0(p_0) + N(p_0)\frac{\partial d_0}{\partial p_0}}_{\text{general deterrence effect}} + \underbrace{\Big(dp_0 N(p_0) + p_0 \frac{\partial N}{\partial p_0} d_0(p_0)\Big) \sum_{t=1}^{3} \frac{\partial d_t}{\partial f}}_{\text{specific deterrence effect}}
\end{aligned} \tag{5}$$

The second equality comes from the fact that the probability of inspection only changes in period 0, and therefore does not affect p_1, p_2, p_3. The third equality comes from the fact that only $f = 1$ only for those that are inspected. Since only $p_0 N(p_0)$ are inspected in period 0, then $dp_0 N(p_0) - p_0 \frac{\partial N}{\partial p_0}$ are inspected in that period as a result of a marginal increase in p_0.

The first part of the decomposition refers to the general deterrence effect, the ex ante reduction in deforestation resulting from an increase in inspection probability. There is an extensive margin response (the change in the number of cells having any level of deforestation) and an intensive margin response (the change in the deforested area within these cells). To understand the magnitude of deforestation reduction in the Amazon forest as a result of an increase in fine probability, we need to estimate the behavioral responses, which are the derivatives in the equation 5.

5.1 General deterrence: effect of inspection probability

Identification and estimation

The general deterrence effect is the effect of changes in fine probability on deforestation. I use data on fines to estimate the fine probability at a cell-year level, using only cells with positive deforestation.[19] It is possible to estimate the probability of a fine (conditional on positive

[19] As explained previously, fines are trivially equal to zero in areas with no deforestation. For that reason, these areas should be excluded, since they convey no meaningful information about the inspection efforts by the enforcement agency.

deforestation) using observable variables in a first stage, and then use the fitted probabilities to estimate the effect of fine probability increases on deforestation in a second stage.

However, it is likely that the probability of fines is correlated with unobserved characteristics of areas where deforestation takes place. As is typically the case in the crime literature (see, for example, Levitt 1997) enforcement efforts tend to be more intense in areas with higher crime incidence because the enforcement agency has knowledge about what are the hotspots of crime and deploys enforcement efforts accordingly. This means that there is a potential endogeneity problem in the fine probability. To overcome this problem, it is necessary to estimate the probability in the first stage using an instrumental variable (or "excluded variable"), that is, a variable that shifts fine probability but does not affect directly the decisions of farmers to deforest areas. A valid instrument then captures exogenous variation in the probability of fines, which can then be used to estimate the impact of fine probability on deforestation in the second stage.

I estimate the following two-equations model:

$$\begin{aligned} y_{it} &= \beta_0 + \beta_1 \pi_{it} + \beta_x X_{it} + \epsilon_{it} \\ \pi_{it} &= \alpha_0 + \alpha_1 Z_{it} + \alpha_x X_{it} + \varepsilon_{it} \end{aligned} \tag{6}$$

where the first equation is the structural equation relating the outcome to the probability π_{it} of fines, and the second equation is a linear probability model of fine probability π_{it} as a function of observables X_{it} and Z_{it}, where Z_{it} is an instrumental variable.

The instrumental variable that provides the exogenous variation on fine probability is *cloud coverage* at the cell level. This instrumental variable was first proposed by Assunção, Gandour, and Rocha (2019), who also used it to estimate the causal impact of fines on deforestation. Cloud coverage blocks temporarily the visibility of a cell, making it impossible for optimal sensors in satellites to produce images of the forest. This feature is a major limitation of the real-time monitoring system DETER, and provide therefore variation in the timing of the deforestation alerts, and consequently on the enforcement agency's ability to inspect and punish offenders on time.

X_{it} are control variables common to both stages. Some of them are time-invariant and at the cell level: dummy variables for deciles of distances to the three main cities in the Amazon forest (Manaus, Belém, and Cuiabá), and dummies for the presence of indigenous territory, conservation units, and roads (federal or state). I also control have cell-year deciles of the share of deforested area. I choose to include the controls as dummies of deciles to allow for potential non-linear relationships between the outcome and these variables. Controlling for the share of deforested area is particularly important because yearly deforestation rates may depend on how much forest is still standing in an area. Finally, in some specifications, I control for commodity prices of soy and ox (aggregate and year-specific) and prices of vegetal coal and wood (state-year specific).

The model is estimated with Two-Stage-Least-Squares using only areas with positive deforestation in the period 2011-2020. The standard errors are clustered at the cell level, thus allowing for autocorrelation of the unobserved error between different years.

Results

Table 10 summarize the results for the regressions of the intensive (d_{it}) and extensive margin ($\mathbb{P}(d_{it} > 0)$). Overall, enforcement probability displays a substantial effect in reducing deforestation along both margins. The table shows, for each outcome, the OLS regression, the 2SLS results, the first stage (using fines as outcome), and the reduced form (the direct effect of the instrument on the outcomes). The samples are different for the two outcomes because I only considered cells with positive deforestation levels for the intensive margin effect. In contrast, for the extensive margin, I considered all cells that had deforestation at some point in the decade from 2011 to 2020.

Columns 1 and 5 show the OLS regressions of deforestation on fines, showing a strong positive correlation both for the intensive and extensive margin, as expected. Columns 2 and 6 show the responses to exogenous increases in the probability of fines. A percentage increase in the probability of fines reduces deforestation areas by 1.9%, and reduces the probability of an area having deforestation by 0.9%. The first stage is strong, as shown in columns 3 and 7, suggesting that intensively cloudy cells-years were less likely to receive fines. Finally, to corroborate the robustness of the results, the direct effect of the instrument on the outcome, on Columns 4 and 8, is positive. This means that cloudier areas have more deforestation than non cloudy areas.

All the results are conditional on a rich set of controls, such as distances from the three main Amazonian cities, distances from the closest IBAMA office, presence of roads, presence of indigenous land, presence of conservation parks, and the percentage of accumulated destroyed forest within the cell-year.[20]

5.2 Specific deterrence: dynamic effects of fines

Identification and estimation

The specific deterrence effect is the effect of punishment on the behavior of farmers. It affects a smaller number of farmers than the general deterrence effect, which is the effect of punishment

[20] The continuous controls (distances and percentage of destroyed forest) were included as dummies of the deciles of the underlying variable. This choice is intended to allow for non-linearities in the relationship between these controls and other variables.

probability. In deforestation, the specific deterrence effect is the change in farmers' behavior *after* an inspection, and caused by the inspection. To identify the causal effect of inspections over time, and in particular distinguish it from time-specific shocks, I use an event study approach like in section 4.

To estimate the specific deterrence effect, I restrict the sample only to those areas that received enforcement action at least once in 2011-2020. Thus subsampling the data, I circumvent the endogeneity issue regarding the inspection decisions since there is no comparison between inspected with non-inspected areas. I estimate the average effect of inspections on inspected (average treatment effect on the treated) using an event study design, where I exploit variation in the timing of the inspections. The assumption that allows this strategy to identify the effect of inspections is a parallel trends assumption. The assumption means that inspected areas would have evolved like non-inspected areas in the absence of an inspection.

I rely on an event study approach, where the "event" is an inspection in a cell. The event occurs in different years for each cell, and the objective is to understand the causal effect of an inspection in several periods relative to the event date, similar to what was done in the analysis of signals and inspections in the previous section. It is possible to see how deforestation evolves relative to the event date. To estimate the specific deterrence effects of inspections, the challenge is, as usual, to find the correct comparison group for the treated cells. The differential timing of inspections gives an opportunity to compare similar areas. Using only the areas that were treated at some point, it is always possible to have some observations that were not yet treated and use them as controls for those that were already treated. Standard practice would lead to an estimation via OLS of an equation like the following:

$$y_{it} = \sum_{\ell=-3}^{5} \beta_\ell \mathbb{1}\{t - e_i = \ell\} + \delta_t + FE_i + \varepsilon_{it} \tag{7}$$

Where the inspection date is denoted by e_i, and which symbolizes the date t in which cell i receives the inspection. As explained in Section 4, estimating this equation via OLS, using only the cells that received an inspection at some point, implies making before and after comparisons between groups. This makes no distinction if the cell used as control has already been treated in the past. As highlighted by recent research, this may be a big problem of the so called two-way fixed effects model for estimating treatment effects: if the control cell has already been treated in the past, its values may be carrying a treatment effect, which is given a negative weight in the estimation of the treatment effect. Therefore I estimate the effects using the imputation method proposed by Borusyak, Jaravel, and Spiess (2021), which avoids invalid comparison between treated groups with other treated groups.

Results

The treatment effects of the deforestation areas is depicted in figure 10b. The treatment effect is negative and increasing in size in periods after the treatment. The reason is that the outcome stabilizes after the inspection, whereas it was accelerating in the years before. The counterfactual scenario is therefore that the outcome would continue accelerating, which yields a growing treatment effect. It is probable that the linear trend is only a good approximation for the counterfactual in the first few years after the inspection occurs, but it does allow for an estimation of the treatment effects in these years. The results show that the treatment effect can only be distinguished from zero from the second year on-wards, when treated cells show 0.3 square kilometer less deforestation than the counterfactual. In the accumulated three years, the treatment effect is approximately 0.7 square kilometer of forest saved on average.

Spatial spillovers

It easy to estimate the event study explained above to capture potential spatial spillovers of fines. Spillovers could be a threat to the estimation of the treatment effects in the preceding sections, since it is assumed that non-treated cells are unaffected by treatment. Two types of spatial spillovers could occur: contagion and leakage.[21] If there is contagion, neighboring farmers may realize that neighboring areas were fined and become more compliant, and this implies that the impact of fines is even greater than the estimated above. On the other hand, if there is leakage, offenders could disperse from areas that suffered enforcement intervention and commit crimes in other areas.[22] In the context of deforestation in the Amazon forest, Assunção, Gandour, and Rocha (2019) and Assunção et al. (2019) have found small contagion effects of enforcement in neighboring municipalities.

I estimate the same event study to look for any effect of a fine on all the neighboring cells, and to second-order neighbors. Figure 11a shows the evolution of deforestation in every year relative to the date of the fine. The blue line shows the evolution for the cells that received the intervention, and the discontinuity in the increasing trend reflects the treatment effects that were discussed above. The neighboring cells (both direct and second-order neighbors) have a lower level of deforestation overall, which is not surprising, since the enforcement agency acts more intensively in areas with large deforestation. Moreover, these neighboring areas show a weak upward trend in deforestation, but no noticeable change in this trend after their neighbor received a fine. Estimation of the treatment effects using the imputation method shows that there is indeed no treatment effect distinguishable from zero, though there the period preceding the fine shows a slight acceleration. Therefore, the results qualitatively confirm previous findings (Assunção,

[21] The terminology is borrowed from Assunção, Gandour, and Rocha (2019)

[22] This effect is also known as "displacement effect" and has been documented recently in the urban crime literature (Blattman et al. 2021)

Gandour, and Rocha (2019) and Assunção et al. (2019)) suggesting that enforcement leads to small reductions in deforestation in neighboring areas.

5.3 Overall effect of inspections on deforestation

After these steps, it is finally possible to compute the impact of increases in fine probability on deforestation by using equation 5. I use the instrumental variable model to obtain estimates of the derivatives of deforestation with respect to the probability of a fine, and then the event study responses for the derivative of deforestation to the realization fine. To complete the computation, it is necessary to use some baseline values of average deforestation and the number of cells with deforestation, which I compute as the averages in the data during the whole period of 2011 to 2020. I thus obtain the following areas of avoided deforestation:

$$
\begin{aligned}
\frac{dD}{dp_0} &= \underbrace{\frac{\partial N}{\partial p_0} d_0(p_0) + N(p_0)\frac{\partial d_0}{\partial p_0}}_{\text{general deterrence effect}} + \underbrace{\left(dp_0 N(p_0) + p_0 \frac{\partial N}{\partial p_0} d_0(p_0)\right) \sum_{t=1}^{3} \frac{\partial d_t}{\partial f}}_{\text{specific deterrence effect}} \\
&= \overbrace{5000 \times (-0.009)}^{\frac{\partial N}{\partial p_0}} \times \overbrace{1}^{d(p_0)} + \overbrace{5000}^{N} \times \overbrace{1 \times (-0.019)}^{\frac{\partial d}{\partial p_0}} \\
&+ \left(\underbrace{0.01 \times 5000}_{dp_0 N(p_0)} + \underbrace{0.13 \times 5000 \times (-0.009)}_{p_0 \frac{\partial N}{\partial p_0} d_0(p_0)}\right) \times \underbrace{3 \times (-0.08)}_{\sum_{t=1}^{3} \frac{\partial d_t}{\partial f}} \\
&= \underbrace{-140}_{\text{gen. deterrence}} \quad \underbrace{-10}_{\text{sp. deterrence}} = -150
\end{aligned}
$$

A one percent increase in inspection probability would reduce yearly deforestation by approximately 150 square kilometers a year. This area represents approximately 1.2% of the deforestation level in 2020 (12 thousand square kilometers) and 2% of the average deforestation in 2011-2020 (7 thousand square kilometers).

As computed in section 4, the treatment effects of the monitoring system represent approximately three percentage points in the overall yearly probability of fines in the Amazon forest. Therefore, the value of this system in terms of reduced deforestation probably lies in the ballpark of 450 square kilometers of saved forest by year, or 6% of the average yearly deforestation in the decade.

5.4 Other behavioral responses: Farmers' use of fire in deforestation

As documented in this section, inspections have a deterrence impact on deforestation, both through the general and specific channels. However, farmers can also change their method of deforestation as a response to inspections, in particular the intensity with which fire is used in deforestation. Farmers can use fire in combination with mechanized logging to deforest areas.

There is no evidence that fire signals cause IBAMA to increase its enforcement effort in an area. The question then becomes whether agents are likely to change the intensity of use of fire as a consequence of the higher enforcement risk. Logging and fire are both techniques used in deforestation, with fire playing a role before and after mechanized logging. If agents are able to choose the intensity of fire and thereby reduce its reliance on logging, they may reduce their enforcement risk. This would be reflected in a higher number of fire signals happening from areas with deforestation, since the quality of these signals has remained stable over time.

For this analysis I use the fire events captures by the satellites Terra and Aqua, which are in operation since the early 2000s. Terra and Aqua are two different NASA satellites which orbit the entire Earth several times a day, going over the Amazon forest twice a day. They are equipped with an image sensor - among other sensors - which produces fire alerts with a precision of roughly 1 square kilometer. To be detectable by the satellites, the fires must be at least 30m long, but unfortunately the satellite cannot tell with precision where in a 1 square kilometer area the fire took place[23].

In a landmark study about fires in the Amazon forest, Nepstad et al. (1999) discuss three different uses of fire in the Amazon region: deforestation fires", "forest surface fires" and "fires on deforested land". Deforestation fires are those that occur in close association with deforestation processes, in particular the "slash and burn" practice of logging and then burning the remaining biomass. Burning biomass disperses nutrients-rich ashes into the soil, boosting crop productivity for one to three years. This ancient practice of clearing forest and burning biomass happens in both primary and secondary vegetation, and typically will repeat itself every year in the so-called "fire season". It is also the reason why most of the fire in the Amazon forest are "fires on deforested land", and are part of an agricultural practice aimed at clearing land from (secondary) vegetation, killing weed, and fertilizing the soil. Slash and burn techniques can also be applied to convert forest into pasture, favoring grass growth. "Forest surface fires", which are fires on standing forest, are treated by the authors as almost entirely accidental, and despite being visually spectacular, they are relatively rare. By accidental fire one should not understand "natural fire", but man-made fires which spread to forest and cause it to burn. To understand the intensity of the use of fire in deforestation, and how it reacts to

[23] See FAQ 9 in the webpage of the Brazilian Institute for Spatial Research: https://queimadas.dgi.inpe.br/queimadas/portal/informacoes/perguntas-frequentes.

improvements in monitoring of logging, I compute the overlaps between newly deforested areas and the areas affected by fire according to the satellites Terra and Aqua. The result is that for each cell and year in the Amazon forest, it is possible to know the percentage of deforestation that has been affected by fire.

I use the exogenous variation in the quality of logging signals to estimate its potential effect on the intensity use of fire in deforestation. For that, I focus on cells with a positive amount of deforestation, and compute the area within the polygons of deforestation that were affected by fires, as measured by the satellites Terra and Aqua. If agents escape react strategically to logging signals by increasing the use of fire, areas with a stronger logging signal would also see a higher share of fire in the deforested polygons. I run the following regression:

$$P(\text{fire}|\text{deforestation}) = \alpha_0 + \alpha_1 \text{detection probability}_{it} + \alpha_2 \text{clouds}_{it} + \beta X_{it} + \varepsilon_{it} \tag{8}$$

Where as usual ε_{it} is a conditional mean zero error term. The equation is estimated using only those cells and years with positive amounts of deforestation, which are the ones for which the share of fire in deforestation and the signal quality can be calculated. Several controls are used, as in previous sections: distances to cities, distances to enforcement agency's neares office and road infrastructure. Moreover, I add the average cloud coverage as a potential confounder for the quality of the signal and the visibility of fires. Fires are indeed covered by clouds. However, the cloud variable used here does not completely cover fires in the Amazon forest, though it completely covers logging signals. Fires are still observed in areas with clouds, even though much less than in areas with no clouds. In any case, clouds may indeed at the same time reduce the quality of the logging signal and block the observation of fires, therefore controlling for it is important.

The results are seen in table 11. It shows that an increase of 1 percentage point in the signal quality leads to an increase in 0.15 percentage point in the intensity of fire, an elasticity of around 15%. The table also shows that areas that receive more intense enforcement, the priority municipalities, also present a much higher share of fire than other areas. This result suggest that farmers substitute towards fire as a result of logging monitoring becoming better. That is certainly an undesirable effect of monitoring, since fires produce other externalities than logging, such as local air pollution and spread accidents.

6 Cost benefit analysis

This paper has so far shown that the enforcement agency has extensively used the monitoring alerts to direct its inspections, which in turn reduced deforestation and forest fires. The use of monitoring alerts has meant a change in the way the enforcement agency targets its inspections. What was the value of this shift in terms of inspection resources saved and welfare gains?

6.1 Inspection costs

I used administrative data on operational expenditures (not including wages) in the Amazon forest to estimate the average cost of a deforestation inspection from 2011-2020. The data is available separately for each of the nine states of the Amazon forest, but it does not distinguish expenditure with deforestation inspections from other operations. I distinguish the deforestation inspections between those that followed a deforestation alert, and those that did not follow an alert, and estimate their costs using a linear regression model, as follows:

$$\text{expenditure}_{it} = \beta_0 + \beta_1 \text{alerts\_inspections}_{it} + \beta_2 \text{no\_alerts\_inspections}_{it} + \delta_t + FE_i + \varepsilon_{it} \quad (9)$$

Where β_1 is an estimate of the average marginal cost of an inspection following an alert and β_2 the average marginal cost of an inspection not following any alert, while δ_t are year fixed effects, and FE_i are state fixed effects. These fixed effects capture other year-specific enforcement activities or state-specific expenditure levels.

The results in table 12 show that the inspections following alerts seem to be considerably less costly than inspections not following alerts. The first column, without any year of state fixed effects, shows that the marginal inspection following alert cost around 7 thousand BRL (1.7 thousand USD), whereas the marginal inspection not following an alert cost 16 thousand BRL (4 thousand USD), more than twice as much. Including year fixed effects, in the second column, does not change much the results and keeps the proportion of the two costs. The third column includes also state fixed effects, which reduces the marginal costs of both inspections by half, but again keeps the proportion between them.

In short, an inspection strategy that follows monitoring alerts seems to be more cost efficient than using other methods to select inspections. The reason is likely to be that inspections without alerts depend on more investigation and attempts before finding an offender to punish. The monitoring system provides the information in real-time, which is almost always correct as shown in section 3. It is therefore cheaper to base inspection selection on them, which is what IBAMA increasingly did. Indeed, since 2014, the operational expenditures of IBAMA in the Amazon forest dropped by 40% (43 million BRL to 25 million BRL), whereas the number of deforestation inspections dropped by 20% (from 935 in 2014 to 730 in 2020).

6.2 Welfare costs and benefits of reducing deforestation

To compute the value of saved forest, I focus solely on the its carbon content, abstracting from the impact of deforestation on biodiversity loss, rain seasons and air quality. On average, deforestation of one hectare in the Amazon forest leads to approximately 560 tonnes of CO2 emissions.[24] The harm caused by these emissions in terms of climate change are estimated from 30 to 100 USD.[25] This substantial benefit accrues globally, whereas some costs and benefits are born locally by Brazilians. In particular, the non-deforested areas have an opportunity cost of economic activities that could be carried out. Using data from the Brazilian Agricultural Survey, I compute that in the Amazon forest, the average value of agricultural output per square kilometer is approximately 100 thousand USD per year, which is approximately 5% of the welfare benefits using 30 USD as the social cost of carbon.

7 Summary and conclusion

This paper has exploited an important improvement in the monitoring of logging in the Amazon forest, and studied its effects on the fights against deforestation. The monitoring system DETER, produces deforestation alerts based on its ability to detect vegetation loss in native forest in the Brazilian Amazonia. DETER is not a system designed to measure deforestation, but to give real-time alerts about where and when deforestation seems to be taking place. By overlaying the maps of yearly deforestation (measured by the system PRODES) with the deforestation alerts issued by DETER, I document an expressive increase in the probability that an area of deforestation produces an alert over the 2011-2020 decade. Overall this means that farmers doing deforestation in the Amazon forest today is three times more likely to be observed in real time than they were ten years ago. Moreover, the number of false positives by DETER also declined sharply, such that almost all deforestation alerts are correct in the sense that they are later verified as deforested areas.

The consequence of this improvement was that the enforcement risk for farmers went up. I show that IBAMA relies on the deforestation alerts to shape its enforcement strategy, but to a great extent the monitoring technology produces redundant information. The average yearly probability of a fine in areas with positive deforestation in the period was 13%, of which three percentage points can be causally attributed to the monitoring system. I then evaluate how farmers respond to increases in fine probability in order to put a value to the monitoring system

[24] This number was computed based on data by the Brazilian Institute for Spatial Research (INPE) over the period 2010-2019. See http://inpe-em.ccst.inpe.br/en/download_en/

[25] The value of 30 USD per ton of CO2 is typically used by authorities such as the US Department of Energy. However, studies may vary regarding the value. Stern (2007) estimates the social cost of carbon at around 85 USD.

in terms of avoided deforestation. I estimate the impact of fines on farmers in two parts. First I estimate the impact of *enforcement risk* on deforestation, and show that a one percentage point in the probability of inspection reduce the probability of farmers engaging in deforestation by 0.9 percentage point, and reduce the deforested area by 1.9 percentage point. Moreover, the *experience of enforcement* has a lasting impact on the areas that are subject to a crackdown. I document lower levels of deforestation in these areas up to three years after the crackdown.

I compute the benefit of these improvements by computing the costs of targeting inspections based on the deforestation alerts. This exercise shows that IBAMA is twice more effective with targeted inspections than with non-targeted ones. This means that it is possible to increase inspection probability for farmers simply by using more extensively the deforestation alerts to guide inspections. Moreover, a one percentage point increase in inspection probability reduces deforestation by 150 square kilometer, or about 2% of average yearly deforestation in the last decade. A three percentage point increase thus represents approximately 450 square kilometers of saved forst (or 6% of average deforestation). The welfare benefits of this reduction are likely to outweigh the opportunity costs by more than twenty times.

References

Advani, Arun, William Elming, and Jonathan Shaw (2018). "The dynamic effects of tax audits". In: *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*. Vol. 111. JSTOR, pp. 1–30.

Alencar, Ane, Daniel Nepstad, and Mariadel Carmen Vera Diaz (2006). "Forest understory fire in the Brazilian Amazon in ENSO and non-ENSO years: area burned and committed carbon emissions". In: *Earth Interactions* 10.6, pp. 1–17.

Alix-Garcia, Jennifer and Daniel L Millimet (2021). "Remotely Incorrect? Accounting for Nonclassical Measurement in Satellite Data on Deforestation". In: *Journal of Applied Econometrics*.

Almunia, Miguel and David Lopez-Rodriguez (2018). "Under the radar: The effects of monitoring firms on tax compliance". In: *American Economic Journal: Economic Policy* 10.1, pp. 1–38.

Andreoni, James, Brian Erard, and Jonathan Feinstein (1998). "Tax compliance". In: *Journal of economic literature* 36.2, pp. 818–860.

Aragão, Luiz EOC, Liana O Anderson, Marisa G Fonseca, Thais M Rosan, Laura B Vedovato, Fabien H Wagner, Camila VJ Silva, Celso HL Silva Junior, Egidio Arai, Ana P Aguiar, et al. (2018). "21st Century drought-related fires counteract the decline of Amazon deforestation carbon emissions". In: *Nature communications* 9.1, pp. 1–12.

Ashby, Matthew PJ (2017). "The value of CCTV surveillance cameras as an investigative tool: An empirical analysis". In: *European Journal on Criminal Policy and Research* 23.3, pp. 441–459.

Assunção, Juliano, Clarissa Gandour, and Romero Rocha (2019). "DETERring deforestation in the Brazilian Amazon: environmental monitoring and law enforcement". In: *Climate Policy Initiative, Rio de Janeiro, Brazil*.

Assunção, Juliano, Clarissa Gandour, Rudi Rocha, et al. (2015). "Deforestation slowdown in the Brazilian Amazon: prices or policies". In: *Environment and Development Economics* 20.6, pp. 697–722.

Assunção, Juliano, Clarissa Gandour, and Eduardo Souza-Rodrigues (2019). "The Forest Awakens: Amazon Regeneration and Policy Spillover". In: *CPI Working Paper*.

Assunção, Juliano, Robert McMillan, Joshua Murphy, and Eduardo Souza-Rodrigues (2019). "Optimal environmental targeting in the amazon rainforest". In: *National Bureau of Economic Research*.

Assunção, Juliano and Romero Rocha (2019). "Getting greener by going black: the effect of blacklisting municipalities on Amazon deforestation". In: *Environment and Development Economics* 24.2, pp. 115–137.

Azevedo, Tasso, Marcos Reis Rosa, Julia Zanin Shimbo, and Magaly Gonzales de Oliveira (2020). “Relatório anual do desmatamento no Brazil”. In.

Azevedo-Ramos, Claudia, Paulo Moutinho, Vera Laísa da S Arruda, Marcelo CC Stabile, Ane Alencar, Isabel Castro, and João Paulo Ribeiro (2020). “Lawless land in no man’s land: The undesignated public forests in the Brazilian Amazon”. In: *Land Use Policy* 99, p. 104863.

Bachas, Pierre, Anne Brockmeyer, Alipio Ferreira, and Bassirou Sarr (2021). “How to Target Enforcement at Scale? Evidence from Tax Audits in Senegal”. In: *mimeo*.

Balboni, Clare, Robin Burgess, and Benjamin A Olken (2021). *The Origins and Control of Forest Fires in the Tropics*. Tech. rep.

Balch, Jennifer K, Paulo M Brando, Daniel C Nepstad, Michael T Coe, Divino Silvério, Tara J Massad, Eric A Davidson, Paul Lefebvre, Claudinei Oliveira-Santos, Wanderley Rocha, et al. (2015). “The susceptibility of southeastern Amazon forests to fire: insights from a large-scale burn experiment”. In: *Bioscience* 65.9, pp. 893–905.

Becker, Gary S (1968). “Crime and punishment: An economic approach”. In: *Journal of Political Economy* 76, pp. 169–217.

Blattman, Christopher, Donald P Green, Daniel Ortega, and Santiago Tobón (2021). “Place-based interventions at scale: The direct and spillover effects of policing and city services on crime”. In: *Journal of the European Economic Association* 19.4, pp. 2022–2051.

Blundell, Wesley, Gautam Gowrisankaran, and Ashley Langer (2020). “Escalation of scrutiny: The gains from dynamic enforcement of environmental regulations”. In: *American Economic Review* 110.8, pp. 2558–85.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (2021). “Revisiting event study designs: Robust and efficient estimation”. In: *Available at SSRN 2826228*.

Burgess, Robin, Francisco Costa, and Benjamin A Olken (2019). “The Brazilian Amazon’s Double Reversal of Fortune”. In.

Callaway, Brantly and Pedro HC Sant’Anna (2020). “Difference-in-differences with multiple time periods”. In: *Journal of Econometrics*.

Casaburi, Lorenzo and Ugo Troiano (2016). “Ghost-house busters: The electoral response to a large anti–tax evasion program”. In: *The Quarterly Journal of Economics* 131.1, pp. 273–314.

Chalfin, Aaron and Justin McCrary (2017). “Criminal deterrence: A review of the literature”. In: *Journal of Economic Literature* 55.1, pp. 5–48.

Chan, H Ron and Yichen Christy Zhou (2021). “Regulatory spillover and climate co-benefits: Evidence from new source review lawsuits”. In: *Journal of Environmental Economics and Management*, p. 102545.

De Chaisemartin, Clément and Xavier d’Haultfoeuille (2020). “Two-way fixed effects estimators with heterogeneous treatment effects”. In: *American Economic Review* 110.9, pp. 2964–96.

De Neve, Jan-Emmanuel, Clement Imbert, Johannes Spinnewijn, Teodora Tsankova, and Maarten Luts (2021). “How to improve tax compliance? Evidence from population-wide experiments in Belgium”. In: *Journal of Political Economy* 129.5, pp. 1425–1463.

Diniz, Cesar Guerreiro, Arleson Antonio de Almeida Souza, Diogo Corrêa Santos, Mirian Correa Dias, Nelton Cavalcante da Luz, Douglas Rafael Vidal de Moraes, Janaina Sant’Ana Maia, Alessandra Rodrigues Gomes, Igor da Silva Narvaes, Dalton M Valeriano, et al. (2015). “DETER-B: The new Amazon near real-time deforestation detection system”. In: *IEEE Journal of selected topics in applied earth observations and remote sensing* 8.7, pp. 3619–3628.

Donaldson, Dave and Adam Storeygard (2016). “The view from above: Applications of satellite data in economics”. In: *Journal of Economic Perspectives* 30.4, pp. 171–98.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan (2018). “The value of regulatory discretion: Estimates from environmental inspections in India”. In: *Econometrica* 86.6, pp. 2123–2160.

Dusek, Libor and Christian Traxler (2021). “Learning from law enforcement”. In: *Journal of the European Economic Association.*

Fearnside, Philip M (2021). “The intrinsic value of Amazon biodiversity”. In: *Biodiversity and Conservation* 30.4, pp. 1199–1202.

Gillespie, Thomas W (2021). *Policy, drought and fires combine to affect biodiversity in the Amazon basin.*

Goodman-Bacon, Andrew (2021). “Difference-in-differences with variation in treatment timing”. In: *Journal of Econometrics.*

Greenstone, Michael, Guojun He, Ruixue Jia, and Tong Liu (2020). *Can Technology Solve the Principal-Agent Problem? Evidence from China’s War on Air Pollution.* Tech. rep. National Bureau of Economic Research.

Hansen, Matthew C, Peter V Potapov, Rebecca Moore, Matt Hancher, Svetlana A Turubanova, Alexandra Tyukavina, David Thau, Stephen V Stehman, Scott J Goetz, Thomas R Loveland, et al. (2013). “High-resolution global maps of 21st-century forest cover change”. In: *science* 342.6160, pp. 850–853.

INPE (2008). “Monitoramento da Cobertura Florestal da Amazônia por Satélites”. In: *INPE report, by Antonio Miguel Vieira Monteiro, Camilo Daleles Rennó, Claudio A Almeida, Dalton de Morisson Valeriano, Joao Vianei Soares, Luis Eduardo Maurano, Maria Isabel Sobral Escada, Silvana Amaral and Taise Farias Pinheiro.*

– (2019a). “Metodologia Utilizada nos Projetos PRODES e DETER”. In: *Instituto Nacional de Pesquisas Espaciais, Brazilian National Institute for Spatial Research.*

– (2019b). “Metodologia Utilizada nos Projetos PRODES e DETER”. In: *Instituto Nacional de Pesquisas Espaciais.*

IPCC (2014). “Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change”. In: *Edenhofer, O., R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel and J.C. Minx (eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.*

Jayachandran, Seema (2009). “Air quality and early-life mortality evidence from Indonesia’s wildfires”. In: *Journal of Human resources* 44.4, pp. 916–954.

Jayachandran, Seema, Joost De Laat, Eric F Lambin, Charlotte Y Stanton, Robin Audy, and Nancy E Thomas (2017). “Cash for carbon: A randomized trial of payments for ecosystem services to reduce deforestation”. In: *Science* 357.6348, pp. 267–273.

Kang, Karam and Bernardo S Silveira (2021). “Understanding disparities in punishment: Regulator preferences and expertise”. In: *Journal of Political Economy* 129.10, pp. 000–000.

Kessler, Daniel and Steven D Levitt (1999). “Using sentence enhancements to distinguish between deterrence and incapacitation”. In: *The Journal of Law and Economics* 42.S1, pp. 343–364.

Kleven, Henrik Jacobsen, Martin B Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez (2011). “Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark”. In: *Econometrica* 79.3, pp. 651–692.

Kuziemko, Ilyana and Steven D Levitt (2004). “An empirical analysis of imprisoning drug offenders”. In: *Journal of Public Economics* 88.9-10, pp. 2043–2066.

Leal Filho, Walter, Ulisses M Azeiteiro, Amanda Lange Salvia, Barbara Fritzen, and Renata Libonati (2021). “Fire in Paradise: Why the Pantanal is burning”. In: *Environmental Science & Policy* 123, pp. 31–34.

Leite-Filho, Argemiro Teixeira, Britaldo Silveira Soares-Filho, Juliana Leroy Davis, Gabriel Medeiros Abrahão, and Jan Börner (2021). “Deforestation reduces rainfall and agricultural revenues in the Brazilian Amazon”. In: *Nature Communications* 12.1, pp. 1–7.

Levitt, Steven D. (1997). “Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime”. In: *The American Economic Review* 87.3, pp. 270–290.

McCrary, Justin (2002). “Using electoral cycles in police hiring to estimate the effect of police on crime: Comment”. In: *American Economic Review* 92.4, pp. 1236–1243.

Naritomi, Joana (2019). “Consumers as tax auditors”. In: *American Economic Review* 109.9, pp. 3031–72.

Nepstad, Daniel, David McGrath, Claudia Stickler, Ane Alencar, Andrea Azevedo, Briana Swette, Tathiana Bezerra, Maria DiGiano, João Shimada, Ronaldo Seroa da Motta, et al. (2014). “Slowing Amazon deforestation through public policy and interventions in beef and soy supply chains”. In: *science* 344.6188, pp. 1118–1123.

Nepstad, Daniel, Adriana Moreira, and Ane Alencar (1999). *Flames in the rain forest: origins, impacts and alternatives to Amazonian fires.* World Bank, Pilot Program to Conserve the Brazilian Rain Forest.

Nepstad, Daniel, Britaldo S Soares-Filho, Frank Merry, André Lima, Paulo Moutinho, John Carter, Maria Bowman, Andrea Cattaneo, Hermann Rodrigues, Stephan Schwartzman, et al. (2009). "The end of deforestation in the Brazilian Amazon". In: *Science* 326.5958, pp. 1350–1351.

Nepstad, Daniel C, Adalberto Verssimo, Ane Alencar, Carlos Nobre, Eirivelthon Lima, Paul Lefebvre, Peter Schlesinger, Christopher Potter, Paulo Moutinho, Elsa Mendoza, et al. (1999). "Large-scale impoverishment of Amazonian forests by logging and fire". In: *Nature* 398.6727, pp. 505–508.

Pomeranz, Dina (2015). "No taxation without information: Deterrence and self-enforcement in the value added tax". In: *American Economic Review* 105.8, pp. 2539–69.

Reddington, CL, EW Butt, DA Ridley, P Artaxo, WT Morgan, H Coe, and DV Spracklen (2015). "Air quality and human health improvements from reductions in deforestation-related fire in Brazil". In: *Nature Geoscience* 8.10, pp. 768–771.

Sheldon, Tamara L and Chandini Sankaran (2017). "The impact of Indonesian forest fires on Singaporean pollution and health". In: *American Economic Review* 107.5, pp. 526–29.

Silva, Claudia Arantes, Giancarlo Santilli, Edson Eyji Sano, and Giovanni Laneve (2021). "Fire occurrences and greenhouse gas emissions from deforestation in the Brazilian Amazon". In: *Remote Sensing* 13.3, p. 376.

Souza-Rodrigues, Eduardo (2014). "Policy interventions in the Amazon Rainforest". In: *GAD Academy: Global Agribusiness Forum* 1.

– (2019). "Deforestation in the Amazon: A unified framework for estimation and policy analysis". In: *The Review of Economic Studies* 86.6, pp. 2713–2744.

Stern, Nicholas (2007). *The economics of climate change: the Stern review.* cambridge University press.

Sun, Liyang and Sarah Abraham (2021). "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects". In: *Journal of Econometrics* 225.2, pp. 175–199.

Valdiones, Ana Paula, Paula Bernasconi, Vinícius Silgueiro, Vinícius Guidotti, Frederico Miranda, Julia Costa, Raoni Rajão, and Bruno Manzolli (2021). "Desmatamento Ilegal na Amazônia e no Matopiba: falta transparência e acesso à informação". In: *https://www.icv.org.br/website/content/uploads/2021/05/icv-relatorio-f.pdf.*

Welsh, Brandon C and David P Farrington (2009). "Public area CCTV and crime prevention: an updated systematic review and meta-analysis". In: *Justice Quarterly* 26.4, pp. 716–745.

Wooldridge, Jeff (2021). "Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators". In: *Available at SSRN 3906345.*

Zou, Eric Yongchen (2021). "Unwatched Pollution: The Effect of Intermittent Monitoring on Air Quality". In: *American Economic Review* 111.7, pp. 2101–26.

FIGURES

F1. Datasets

Figure 1: Timing of deforestation and data

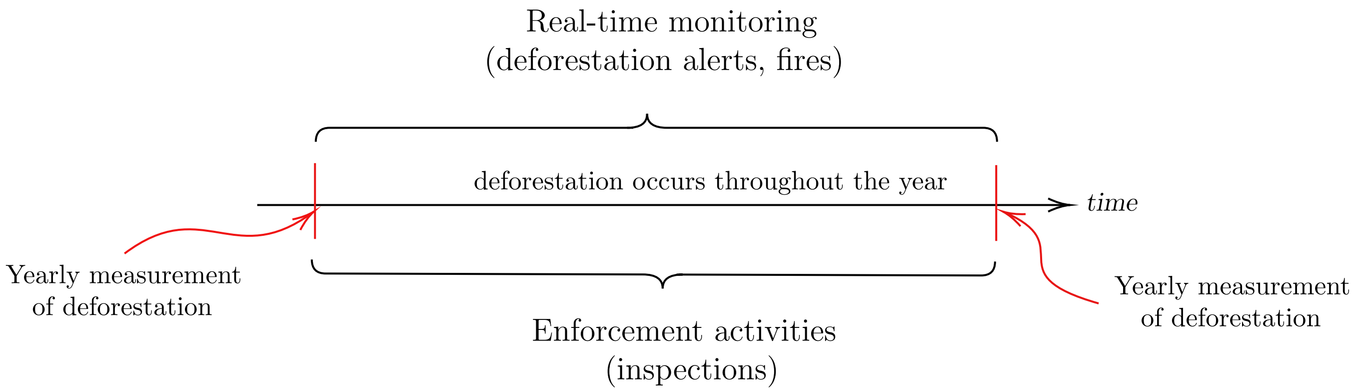
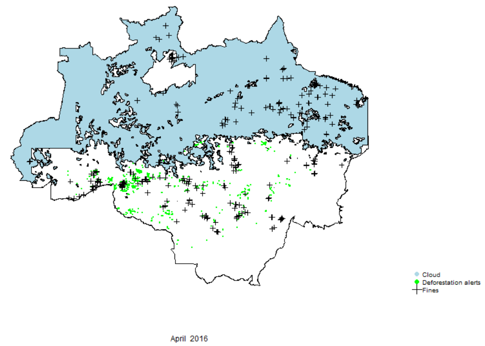
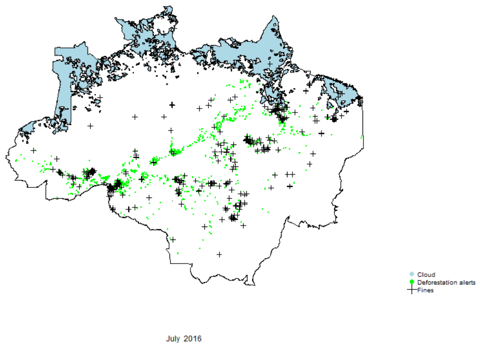


Figure 2: Real time information on the Amazon forest

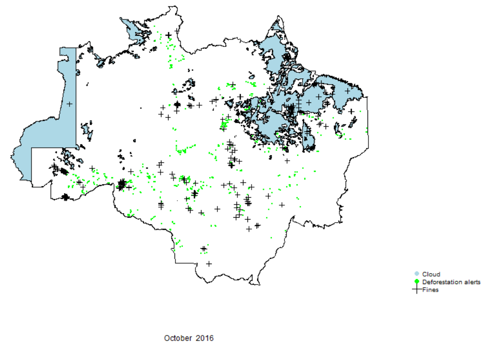
(a) January



(b) April



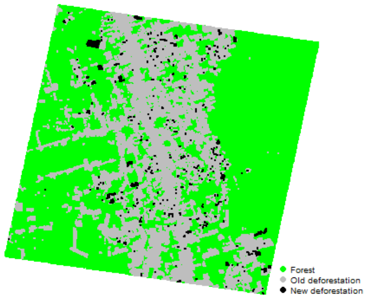
(c) July



(d) October

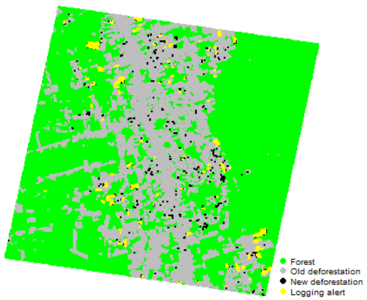
The following figures provide examples of maps used in the analysis. The square represented is an area of approximately 30 thousand square kilometers in the Brazilian Amazon forest, in the state of Pará. The picture corresponds to the year 2016, defined according to the PRODES methodology, that is, from August 2015 to July 2016. The PRODES image represents the state of the soil coverage on August 1st 2016. The logging alers, fire alerts and fines maps represent all the events that took place in the twelve month period from August 2015 to July 2016. The data from PRODES, logging alerts (DETER) and fires were obtained from the Brazilian National Institute for Space Research (INPE). The fines stem from the administrative dataset on environmental infractions of IBAMA, and refer exclusively to "deforestation" fines.

Figure 3: Main datasets

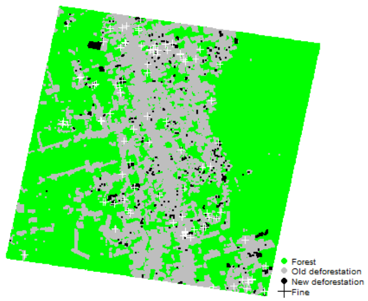


(a) Measurement satellite (PRODES)

(b) Fires (Queimadas)



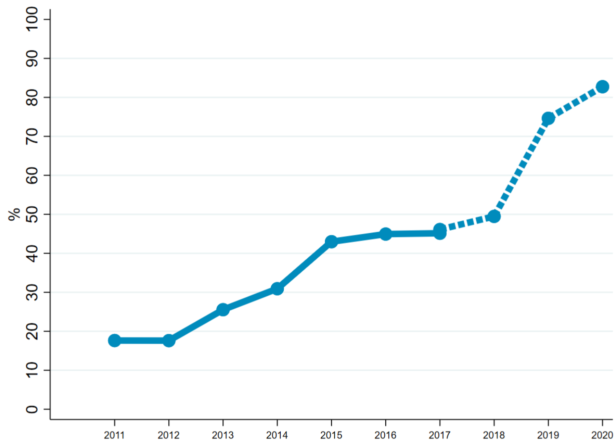
(c) Logging alerts (DETER)



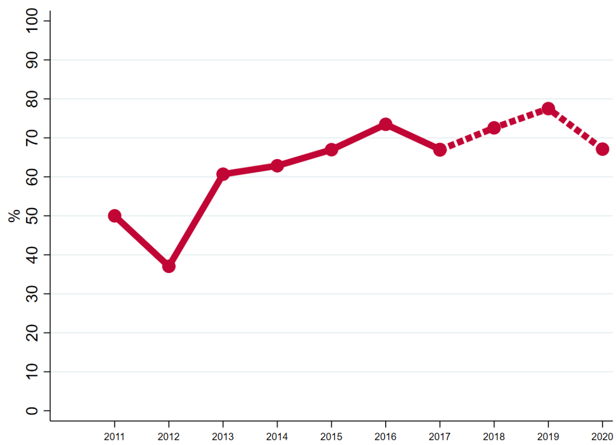
(d) Deforestation fines (IBAMA)

F2. Alert probabilities

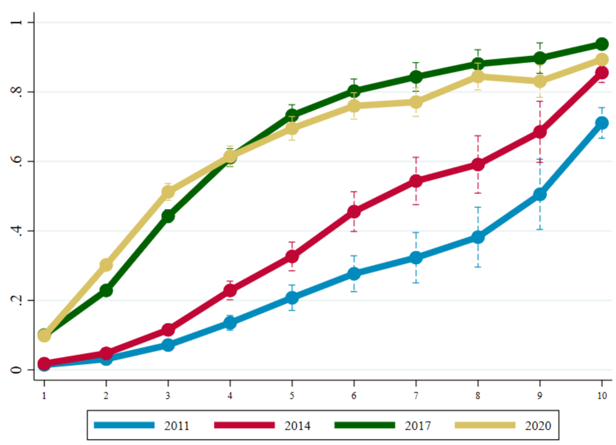
Figure 4: Monitoring quality



(a) Share of detected deforestation by deforestation alerts $\mathbb{P}(\text{alert}|\text{deforestation})$



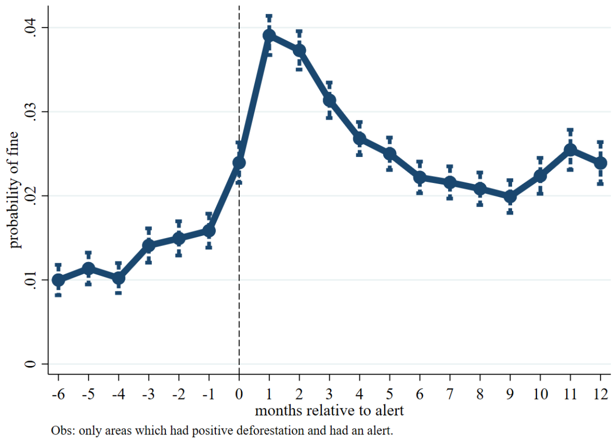
(b) Share of deforestation alerts that were declared deforestation $\mathbb{P}(\text{deforestation}|\text{alert})$



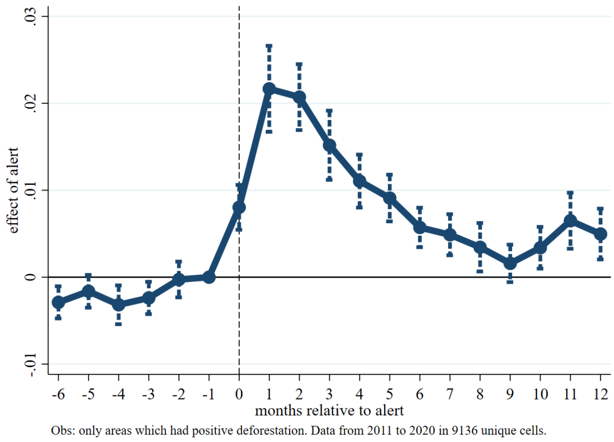
(c) Share of detected deforestation by size decile of deforestation areas and year

F3. Results - inspection probability and alerts - Event study

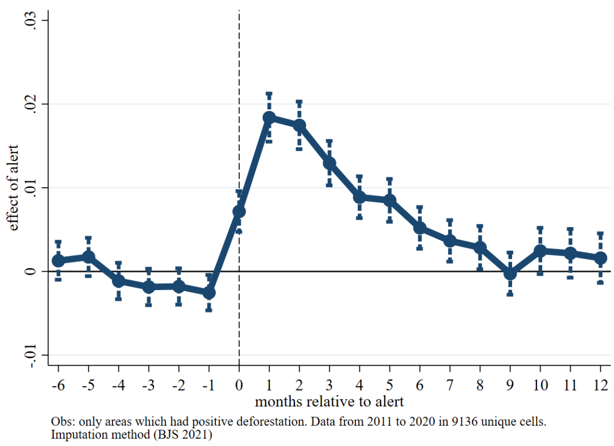
Figure 5: Effects of deforestation alerts on probability of fine



(a) Mean outcome (probability of fine) by months relative to deforestation alert

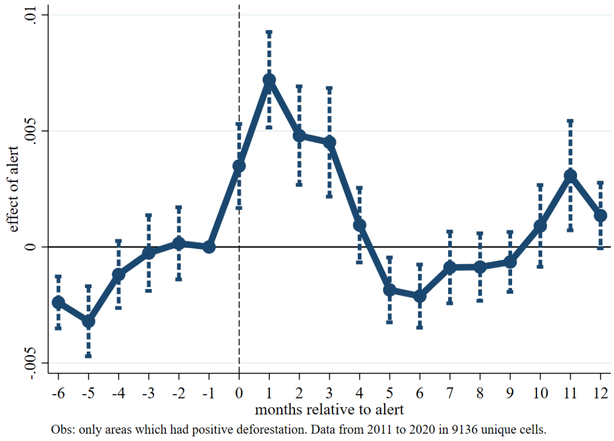


(b) Treatment effects on probability of fine by OLS

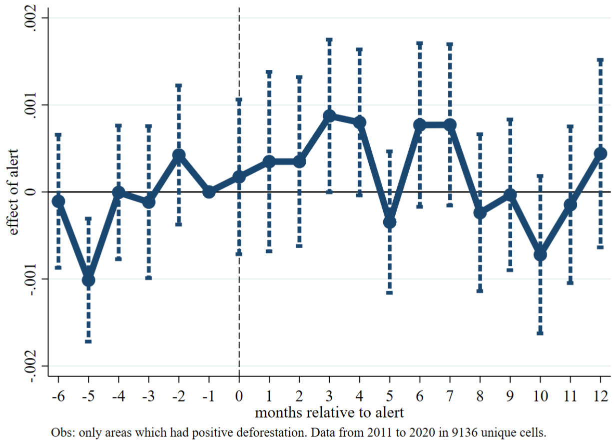


(c) Treatment effects on probability of fine using imputation method (Borusyak, Jaravel, and Spiess 2021)

Figure 6: Fire alerts

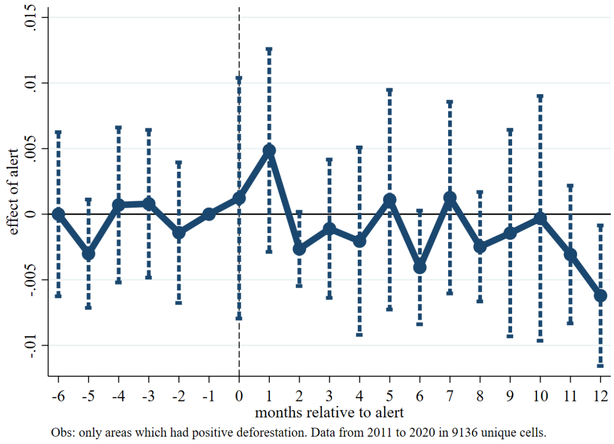


(a) Treatment effects of forest fires alerts on probability of fine

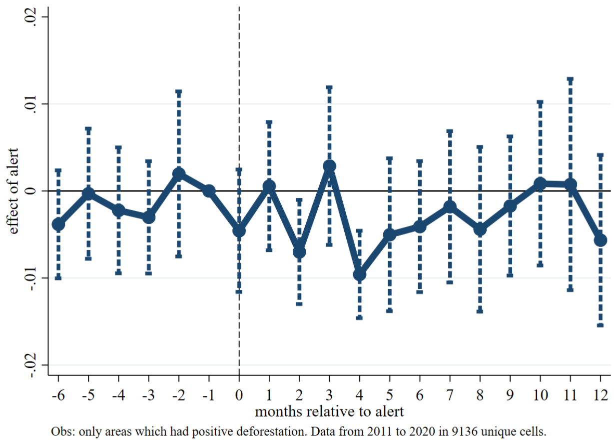


(b) Treatment effects of forest alerts on probability of fine (excluding areas with deforestation alerts)

Figure 7: Placebo tests



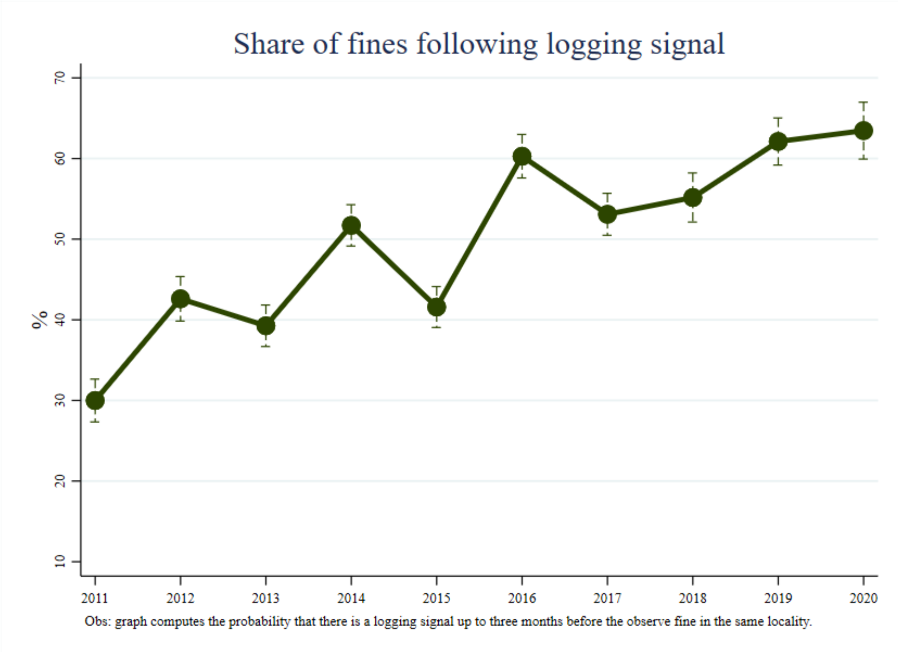
(a) Treatment effects of mining alerts on probability of fine



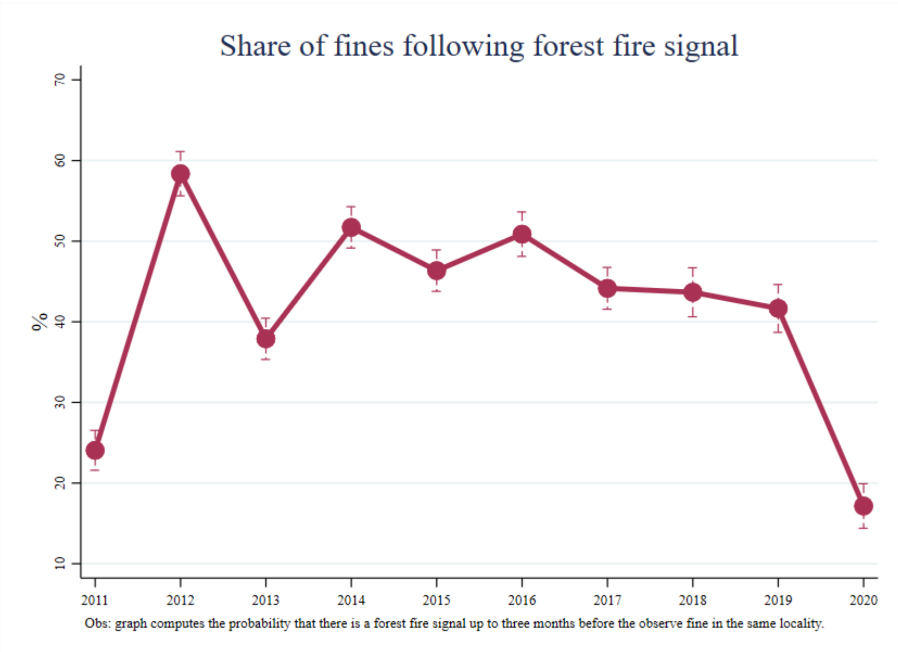
(b) Treatment effects of forest fires alerts on probability of fine

F4. Share of fines following a real-time satellite alert

Figure 8: Targeting using monitoring alerts

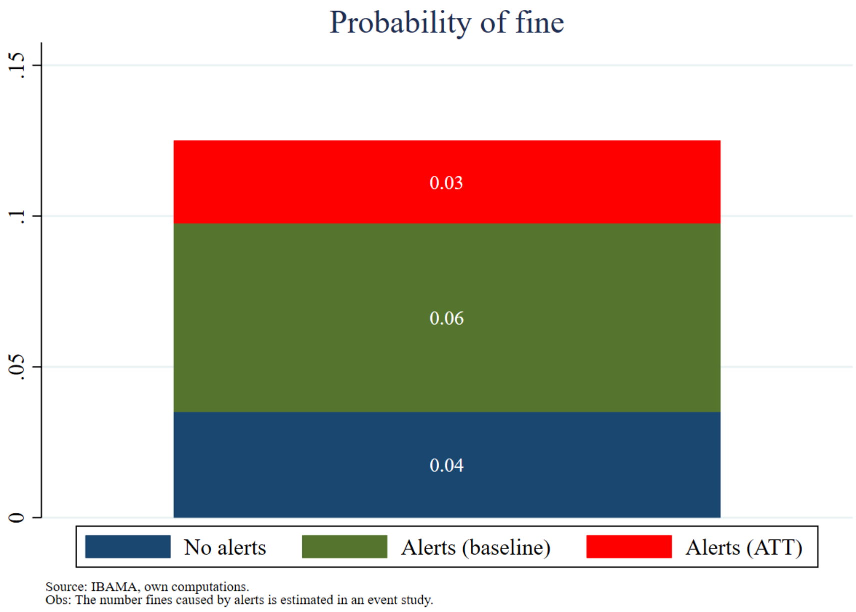


(a) Fines following logging signal



(b) Fines following fire signal

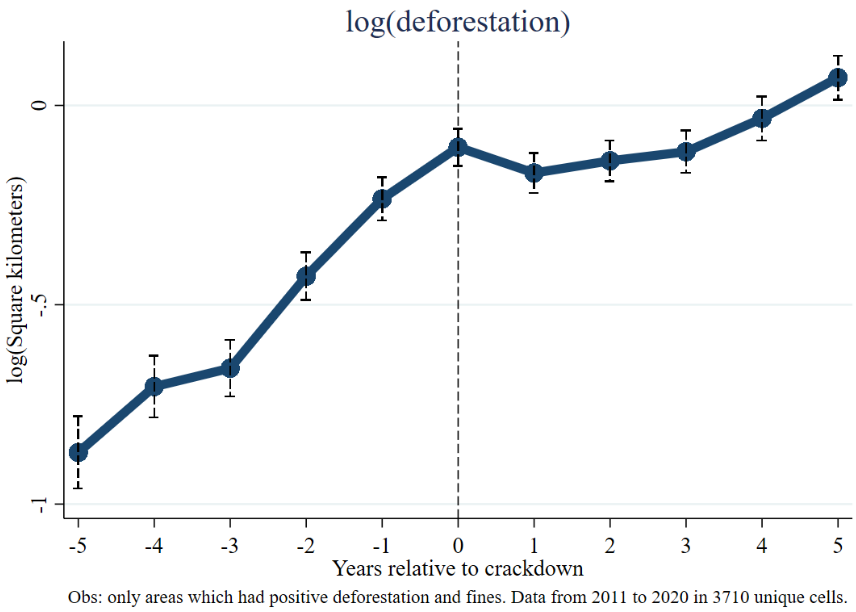
Figure 9: Fines caused by alerts



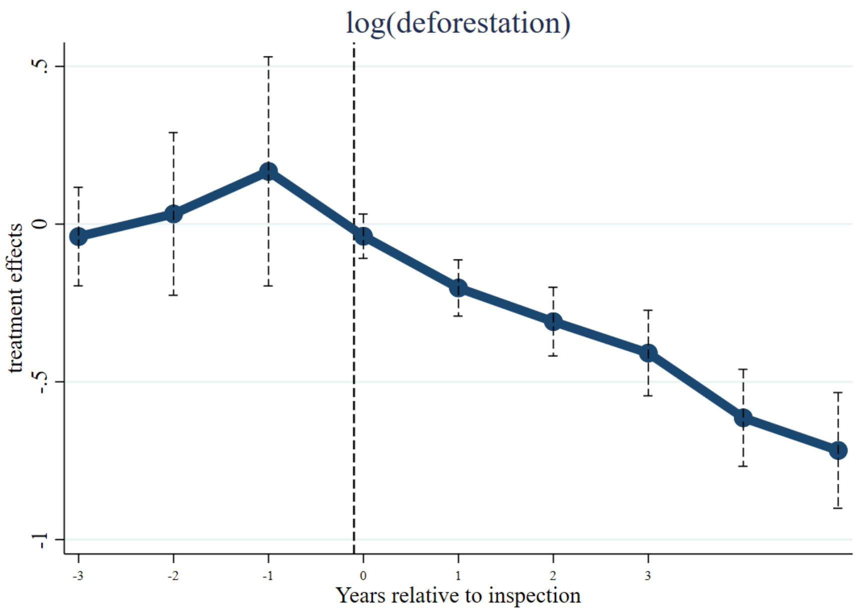
(a) Inspection probability

F5. Event Study of Deforestation and Fines

Figure 10: Deforestation and inspections

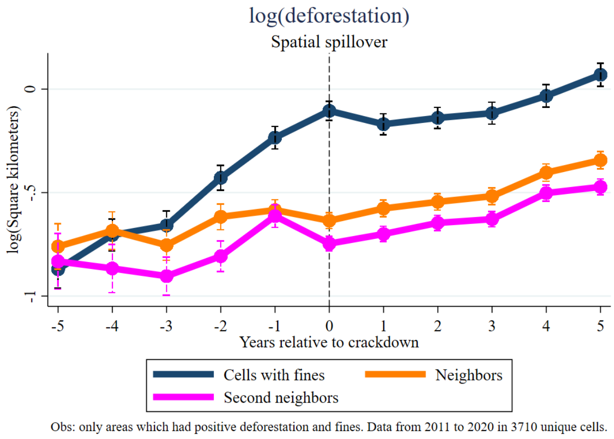


(a) Mean outcome (log deforestation) by year relative to inspection year

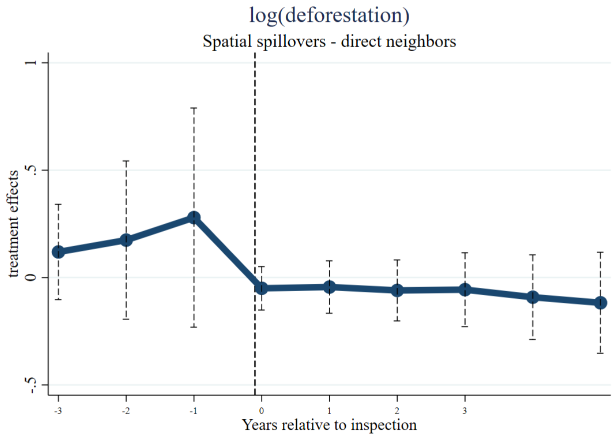


(b) Treatment effects using imputation method (Borusyak, Jaravel, and Spiess 2021)

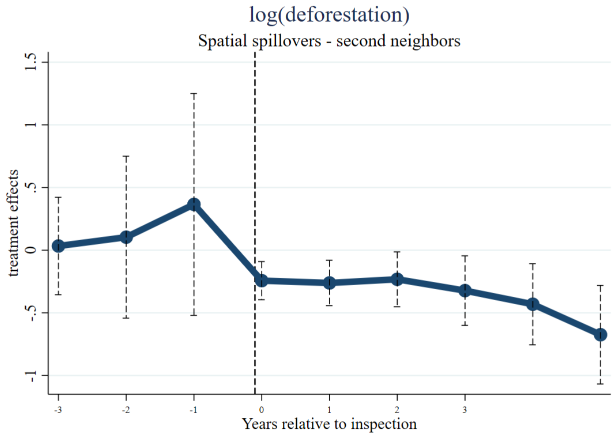
Figure 11: Spatial spillovers



(a) Average log deforestation in neighboring areas



(b) Treatment effect estimation (BJS 2021) for direct neighbors



(c) Treatment effect estimation (BJS 2021) for second neighbors

TABLES

T1. Data sources

Table 1: Main datasets used and sources

NAME OF DATASET	DESCRIPTION	TIME	GEOGRAPHICAL LEVEL	SOURCE
Fire signals (Queimadas)	Fire events in the Amazon forest detected by satellites.	Daily (2011-2020)	Geo-referenced points	Satellites Terra, Aqua and NPP (from 2013), compiled by INPE (Queimadas)
Logging signals (DETER)	Areas with potential deforestation activity, detected by satellite.	Monthly (2011-2020)	Geo-referenced polygons	Satellites Terra (DETER-A, until 2017) and CBERS (DETER-B, from 2017) and compiled and interpreted by INPE
Cloud coverage (DETER)	Areas with cloud coverage, which inhibit satellite monitoring.	Monthly (2011-2017)	Geo-referenced polygons	Satellites Terra (DETER-A, until 2017) compiled by INPE
Soil cover of Amazon forest (PRODES)	Information about soil cover in every area of the Amazon forest, covering in particular the categories: forest, new deforestation and previous deforestation. PRODES is the official program to measure deforestation in Brazil.	Yearly (measured in July of each year)	Geo-referenced polygons	Data from satellite Landsat, interpreted and compiled by INPE.
Environmental fines	Fines issued by the Brazilian federal environmental authority (IBAMA), and specifically the fines for deforestation.	Daily (2011-2020)	Geo-referenced points	Administrative database of IBAMA

Table 2: Auxiliar datasets and sources

Name of dataset	Description	Time	Geographical level	Source
Prices of wood and coal	Prices in Brazilian Real (BRL) of 2020 of wood (*"madeira em tora"*) and vegetal coal (*"carvão vegetal"*).	Yearly (2011-2019)	State-level averages	IBGE, Vegetal Extraction Surveys
Prices of soy and cattle	Prices in Brazilian Real (BRL) of 2020 of 60kg of soy (*"soja industrial"*) and cattle (*"boi em pé arroba"*).	Monthly (2011-2018)	National averages	Agricultural Secretariat of the State of Paraná
Indigenous reserves and Conservation Units	Areas of indigenous reserves (or inhabited traditionally but not officially delimited) and conservation units in the Brazilian Amazon.	Fixed over time.	Geo-referenced polygon.	INPE (TerraBrasilis)
Road infrastructure	State and federal roads in the Brazilian Amazonia.	Fixed over time.	Geo-referenced lines.	MapBiomas
Private rural properties	Areas of private properties in the official public registry (*Cadastro Ambiental Rural*).	Fixed over time.	Geo-referenced polygons.	CAR, Ministry of Agriculture of Brazil
Budget execution of environmental agency	Expenditures in BRL 2020 by the Amazonian units of the federal environmental agency IBAMA.	Yearly (2015-2020)	By state.	Federal Government of Brazil, (http://transparencia.gov.br)
Consumer Price Index	Consumer Price Index used to convert monetary values to values of 2020. The index used was IPCA-IBGE.	Monthly (2011-2020)	National level.	IBGE

T2. Regression tables - Behavior of the enforcement agency

Table 3: Outcome: detection probability

	(1)	(2)	(3)	(4)
DETER B dummy		0.382***		
		(0.0225)		
% year cloud coverage	-0.109			
	(0.0954)			
up to 20% fire	0.00727	0.0387***	0.0387***	0.0144
	(0.0152)	(0.0102)	(0.0102)	(0.00997)
20% to 50% fire	0.0148	0.0358***	0.0358***	0.0144
	(0.0143)	(0.00949)	(0.00949)	(0.00932)
50% to 80% fire	0.0472***	0.0564***	0.0564***	0.0439***
	(0.0130)	(0.00955)	(0.00955)	(0.00947)
80% to 100% fire	0.0698***	0.0673***	0.0673***	0.0669***
	(0.0131)	(0.0105)	(0.0105)	(0.0104)
Size of polygon	0.105***	0.0351***	0.0351***	
	(0.00970)	(0.00467)	(0.00467)	
Size squared	-0.00599***	-0.000855***	-0.000855***	
	(0.00107)	(0.000249)	(0.000249)	
2012.year	-0.104***	-0.0793***	-0.0793***	-0.0818***
	(0.0350)	(0.0268)	(0.0268)	(0.0263)
2013.year	-0.00737	0.0253	0.0253	0.0226
	(0.0375)	(0.0250)	(0.0250)	(0.0246)
2014.year	0.173***	0.191***	0.191***	0.187***
	(0.0284)	(0.0246)	(0.0246)	(0.0242)
2015.year	0.262***	0.288***	0.288***	0.284***
	(0.0288)	(0.0232)	(0.0232)	(0.0229)
2016.year	0.304***	0.349***	0.349***	0.346***
	(0.0379)	(0.0223)	(0.0223)	(0.0220)
2017.year	0.363***	0.0136	0.395***	0.390***
	(0.0295)	(0.0117)	(0.0225)	(0.0222)
2018.year		-0.107***	0.274***	0.271***
		(0.0130)	(0.0227)	(0.0223)
2019.year		-0.0886***	0.293***	0.261***
		(0.00925)	(0.0225)	(0.0222)
2020.year		0	0.382***	0.347***
		(.)	(0.0225)	(0.0222)
Log size				0.120***
				(0.00451)
Intercept	0.444***	0.420***	0.420***	0.480***
	(0.0543)	(0.0204)	(0.0204)	(0.0199)
N	9881	18639	18639	18639
r2	0.423	0.389	0.389	0.404

Obs: ***1% **5% *10% significance levels. Linear probability model of detection probability by monitoring satellite (DETER), done at the polygon level of true deforestation (PRODES). and 2SLS regressions of the size of deforestation (in square kilometers) on the event of a crackdown. Standard errors are clustered at the cell level (15km x 15km).

Table 4: Outcome: probability of inspection

	(1)	(2)	(3)	(4)	(5)	(6)
Positive deforestation				5.715***	6.235***	5.393***
				(7.18)	(7.62)	(5.07)
Positive deforestation X 2012	-4.833***	0	1.475	-3.129***	-3.850***	
	(-5.24)	(.)	(0.92)	(-3.52)	(-4.32)	
Positive deforestation X 2013	2.055	0	14.04***	2.428*	2.690**	
	(1.52)	(.)	(5.67)	(1.96)	(2.09)	
Positive deforestation X 2014	-1.357	1.656	10.69***	-0.722	-0.950	0
	(-1.12)	(0.95)	(4.02)	(-0.64)	(-0.82)	(.)
Positive deforestation X 2015	2.835*	6.400***	18.13***	3.619***	3.824***	4.811***
	(1.87)	(3.66)	(5.53)	(2.71)	(2.67)	(4.08)
Positive deforestation X 2016	-2.602**	0.665	7.323***	-1.671	-1.399	-0.445
	(-2.24)	(0.42)	(3.00)	(-1.59)	(-1.25)	(-0.43)
Positive deforestation X 2017	-0.727	2.979**	9.909***	0.543	0.388	1.432
	(-0.70)	(2.04)	(5.40)	(0.55)	(0.38)	(1.21)
Positive deforestation X 2018	-3.000***	0.839	6.550***	-2.060*	-1.717	-0.797
	(-2.73)	(0.56)	(2.99)	(-1.90)	(-1.56)	(-0.61)
Positive deforestation X 2019	-3.767***	1.261	5.432***	-2.165**	-2.675***	-1.757
	(-3.64)	(1.21)	(3.07)	(-2.10)	(-2.60)	(-1.57)
Positive deforestation X 2020	-6.202***	0	3.794***	-4.660***	-4.896***	-4.014***
	(-7.26)	(.)	(3.25)	(-5.55)	(-6.06)	(-3.74)
State road	1.356	1.357	-0.682	1.040	0.959	0.994
	(1.61)	(1.55)	(-0.60)	(1.56)	(1.62)	(1.51)
Size of forest border	0.349***	0.338***	0.520***	0.321***	0.337***	0.323***
	(9.11)	(8.26)	(7.89)	(9.26)	(9.74)	(8.81)
Indigenous territory	-1.304	-0.690	0.616	-1.322**	-0.961**	-0.884**
	(-1.30)	(-0.66)	(0.47)	(-1.99)	(-2.48)	(-2.12)
Conservation unit	-2.047**	-1.929*	-1.760	-1.768**	-0.671*	-0.714*
	(-2.09)	(-1.82)	(-1.61)	(-2.39)	(-1.74)	(-1.79)
prioritylist	3.357	0.475	0	3.446*	2.237*	1.148
	(1.17)	(0.14)	(.)	(1.69)	(1.85)	(0.81)
expenditureindex		7.730**				1.718
		(2.29)				(1.49)
Sample	Same year deforestation	Same year deforestation	Priority municipalities	Some year deforestation	All data	All data
Mun. Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
N	51657	37139	40263	93624	198819	139310
R2	0.173	0.184	0.207	0.160	0.175	0.181

Obs: ***1% **5% *10% significance levels. Linear regression of punishment (binary) on positive deforestation (binary), with year interactions and controlling for several fixed and varying characteristics of the observations, as well as municipality and year fixed effects. Observational level is a 15km x 15km cell-year in the Amazon forest. Standard errors are clustered at the municipality level.

Table 5: Outcome: inspection probability with real-time signals

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Positive deforestation	0	0	0	4.848***	2.250***	2.301***	1.807***
	(.)	(.)	(.)	(5.47)	(6.85)	(6.90)	(5.22)
Deforestation alert	7.957***	2.562**	2.178**	14.12***	9.152***	10.18***	9.710***
	(11.45)	(2.42)	(2.42)	(13.41)	(12.90)	(12.43)	(10.64)
Forest fire alert	1.482***	-1.132*	-0.0934	1.370***	0.796***	0.393*	0.204
	(2.67)	(-1.80)	(-0.17)	(3.04)	(2.71)	(1.88)	(0.95)
Indigenous territory	-0.967	-0.153	-0.949	0.605	-1.031	-0.774**	-0.646*
	(-0.98)	(-0.15)	(-0.97)	(0.51)	(-1.60)	(-2.12)	(-1.65)
Conservation unit	-2.047**	-1.732*	-2.047**	-1.386	-1.670**	-0.614*	-0.615
	(-2.18)	(-1.69)	(-2.19)	(-1.38)	(-2.40)	(-1.70)	(-1.64)
Fire and Deforestation alerts		6.422***	6.381***				
		(5.39)	(6.26)				
Expenditure index		17.44***					5.294***
		(8.04)					(6.81)
Sample	Same year deforestation	Same year deforestation	Same year deforestation	Priority municipalities	Some year deforestation	All data	All data
Mun. Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Unit cell Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	51657	37139	51657	40263	93624	198819	139310
R2	0.178	0.192	0.178	0.225	0.169	0.187	0.192

Obs: ***1% **5% *10% significance levels. Linear regression of inspection (binary) on positive deforestation (binary), logging signals and fire signals, with year interactions and controlling for several fixed and varying characteristics of the observations, as well as municipality and year fixed effects. Observational level is a 15km x 15km cell-year in the Amazon forest. Standard errors are clustered at the municipality level.

Table 6: Outcome: inspection probability with real-time alerts

	(1)	(2)	(3)	(4)	(5)	(6)
Positive deforestation	0	0	4.616***	0	2.277***	1.887***
	(.)	(.)	(5.32)	(.)	(6.89)	(5.38)
Deforestation alert	12.62***	18.42***	17.56***	12.62***	15.35***	21.63***
	(7.20)	(8.17)	(8.08)	(7.20)	(9.52)	(9.67)
Deforestation alert X 2012	-5.066**		-8.412***	-5.066**	-7.689***	
	(-2.39)		(-3.28)	(-2.39)	(-4.19)	
Deforestation alert X 2013	18.45***		17.15***	18.45***	17.47***	
	(6.23)		(4.44)	(6.23)	(6.18)	
Deforestation alert X 2014	6.473**	0	7.546*	6.473**	6.415**	0
	(2.20)	(.)	(1.83)	(2.20)	(2.44)	(.)
Deforestation alert X 2015	5.370*	-0.960	10.42**	5.370*	6.121**	-0.136
	(1.94)	(-0.36)	(2.58)	(1.94)	(2.42)	(-0.06)
Deforestation alert X 2016	-2.909	-9.000***	-0.711	-2.909	-3.198	-9.254***
	(-1.14)	(-3.50)	(-0.18)	(-1.14)	(-1.38)	(-4.03)
Deforestation alert X 2017	-3.672*	-9.419***	-4.817	-3.672*	-5.217***	-11.18***
	(-1.71)	(-3.88)	(-1.39)	(-1.71)	(-2.81)	(-4.99)
Deforestation alert X 2018	-5.631***	-11.72***	-5.114*	-5.631***	-7.450***	-13.58***
	(-3.03)	(-4.70)	(-1.78)	(-3.03)	(-4.38)	(-5.66)
Deforestation alert X 2019	-7.492***	-13.11***	-6.179**	-7.492***	-8.794***	-14.73***
	(-3.81)	(-5.47)	(-2.18)	(-3.81)	(-4.97)	(-6.66)
Deforestation alert X 2020	-10.98***	-16.86***	-10.79***	-10.98***	-12.27***	-18.32***
	(-5.54)	(-7.70)	(-3.96)	(-5.54)	(-7.19)	(-8.78)
Fire alert	2.346**	2.679**	3.710***	2.346**	1.963***	0.591
	(2.03)	(2.37)	(2.90)	(2.03)	(3.93)	(1.36)
Fire alert X 2012	-3.436**		-4.643***	-3.436**	-2.250***	
	(-2.47)		(-3.12)	(-2.47)	(-3.89)	
Fire alert X 2013	-0.103		0.320	-0.103	-0.771	
	(-0.06)		(0.21)	(-0.06)	(-1.25)	
Fire alert X 2014	0.0585	0	-1.381	0.0585	-1.719***	0
	(0.04)	(.)	(-0.74)	(0.04)	(-2.65)	(.)
Fire alert X 2015	-0.0569	-0.133	-0.745	-0.0569	-0.904	0.790
	(-0.03)	(-0.08)	(-0.35)	(-0.03)	(-1.20)	(1.51)
Fire alert X 2016	0.0114	0.0175	-4.274***	0.0114	-2.157***	-0.527
	(0.01)	(0.01)	(-2.76)	(0.01)	(-3.61)	(-1.08)
Fire alert X 2017	-1.311	-1.533	-1.403	-1.311	-1.645***	-0.0164
	(-0.89)	(-1.00)	(-0.86)	(-0.89)	(-2.62)	(-0.03)
Fire alert X 2018	-3.295**	-3.622**	-4.511***	-3.295**	-2.300***	-0.656
	(-1.97)	(-2.03)	(-2.91)	(-1.97)	(-3.75)	(-1.13)
Fire alert X 2019	-2.381*	-2.853**	-3.992**	-2.381*	-2.425***	-0.903*
	(-1.83)	(-2.20)	(-2.53)	(-1.83)	(-4.22)	(-1.72)
Fire alert X 2020	-1.205	-1.445	-2.964*	-1.205	-2.536***	-0.975*
	(-0.98)	(-1.19)	(-1.84)	(-0.98)	(-4.40)	(-1.87)
Indigenous territory	-0.601	0.0193	0.933	-0.601	-0.643*	-0.540
	(-0.63)	(0.02)	(0.81)	(-0.63)	(-1.84)	(-1.43)
Conservation unit	-1.768*	-1.639	-1.442	-1.768*	-0.557	-0.572
	(-1.96)	(-1.62)	(-1.43)	(-1.96)	(-1.59)	(-1.56)
Expenditure index		5.114				1.046
		(1.46)				(0.89)
Sample	Same year deforestation	Same year deforestation	Priority municipalities	Some year deforestation	All data	All data
Mun. Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Unit cell Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	51657	37139	40263	51657	198819	139310
R2	0.200	0.201	0.245	0.200	0.206	0.205

Obs: ***1% **5% *10% significance levels. Linear regression of inspection (binary) on positive deforestation (binary), deforestation alerts and fire signals, with year interactions and controlling for several fixed and varying characteristics of the observations, as well as municipality and year fixed effects. Observational level is a 15km x 15km cell-year in the Amazon forest. Standard errors are clustered at the municipality level.

Table 7: Outcome: probability of inspection

	(1) Logging	(2) Forest fire	(3) Fire	(4) Sel. logging	(5) Mining
Lag 6	-0.371***	-0.381***	-0.306***	-0.636**	0.103
	(-4.39)	(-5.14)	(-4.20)	(-1.97)	(0.29)
Lag 5	-0.0293	-0.567***	-0.484***	-0.249	-0.190
	(-0.24)	(-5.64)	(-5.65)	(-0.57)	(-0.84)
Lag 4	-0.133	-0.337***	-0.308***	-0.424	0.0676
	(-1.23)	(-3.49)	(-4.35)	(-0.78)	(0.22)
Lag 3	-0.152	-0.215**	-0.297***	-0.523	0.210
	(-1.51)	(-2.15)	(-3.74)	(-1.15)	(0.66)
Lag 2	-0.00385	-0.0757	-0.120	0.236	-0.0495
	(-0.04)	(-0.81)	(-1.63)	(0.36)	(-0.20)
Lag1					
Alert	1.001***	0.447***	0.220***	-0.650	0.0652
	(6.15)	(4.28)	(2.84)	(-1.30)	(0.13)
Lead 1	2.313***	0.844***	0.477***	-0.208	0.700
	(8.50)	(7.13)	(5.04)	(-0.44)	(1.31)
Lead 2	2.381***	0.595***	0.409***	-0.929**	-0.251
	(10.51)	(4.96)	(3.73)	(-2.02)	(-1.42)
Lead 3	1.860***	0.591***	0.501***	0.472	-0.200
	(8.39)	(4.30)	(4.52)	(0.81)	(-0.83)
Lead 4	1.570***	0.191*	0.322***	-1.503***	0.108
	(8.23)	(1.82)	(3.45)	(-4.69)	(0.19)
Lead 5	1.251***	-0.168	0.0152	-0.717	-0.197
	(7.65)	(-1.62)	(0.18)	(-1.21)	(-0.71)
Lead 6	0.906***	-0.163*	0.00515	-0.700*	-0.334
	(5.78)	(-1.84)	(0.06)	(-1.65)	(-1.32)
Month and year dummies	Yes	Yes	Yes	Yes	Yes
Cell fixed effects	Yes	Yes	Yes	Yes	Yes
N	695517	687449	680306	722911	723946
R2	0.0587	0.0567	0.0562	0.0551	0.0548
Share with signal	.03	.06	.07	0	0

Obs: ***1% **5% *10% significance levels. Linear regression of crackdown (binary) periods relative to the earliest signal in a cell-year. Only cells with positive deforestation were used for estimation, since only they can suffer enforcement action. Some cells did not have any signal, and are therefore the reference for all the period dummies. The regression is estimated by OLS, including fixed effects for month, year and cell (15km x 15km). Standard errors are clustered at the municipality level.

Table 8: Outcome: probability of inspection

	(1) Logging	(2) Forest fire	(3) Fire	(4) Sel. logging	(5) Mining
Lag 6	-0.371***	-0.381***	-0.306***	-0.636**	0.103
	(-4.39)	(-5.14)	(-4.20)	(-1.97)	(0.29)
Lag 5	-0.0293	-0.567***	-0.484***	-0.249	-0.190
	(-0.24)	(-5.64)	(-5.65)	(-0.57)	(-0.84)
Lag 4	-0.133	-0.337***	-0.308***	-0.424	0.0676
	(-1.23)	(-3.49)	(-4.35)	(-0.78)	(0.22)
Lag 3	-0.152	-0.215**	-0.297***	-0.523	0.210
	(-1.51)	(-2.15)	(-3.74)	(-1.15)	(0.66)
Lag 2	-0.00385	-0.0757	-0.120	0.236	-0.0495
	(-0.04)	(-0.81)	(-1.63)	(0.36)	(-0.20)
Lag1					
Alert	1.001***	0.447***	0.220***	-0.650	0.0652
	(6.15)	(4.28)	(2.84)	(-1.30)	(0.13)
Lead 1	2.313***	0.844***	0.477***	-0.208	0.700
	(8.50)	(7.13)	(5.04)	(-0.44)	(1.31)
Lead 2	2.381***	0.595***	0.409***	-0.929**	-0.251
	(10.51)	(4.96)	(3.73)	(-2.02)	(-1.42)
Lead 3	1.860***	0.591***	0.501***	0.472	-0.200
	(8.39)	(4.30)	(4.52)	(0.81)	(-0.83)
Lead 4	1.570***	0.191*	0.322***	-1.503***	0.108
	(8.23)	(1.82)	(3.45)	(-4.69)	(0.19)
Lead 5	1.251***	-0.168	0.0152	-0.717	-0.197
	(7.65)	(-1.62)	(0.18)	(-1.21)	(-0.71)
Lead 6	0.906***	-0.163*	0.00515	-0.700*	-0.334
	(5.78)	(-1.84)	(0.06)	(-1.65)	(-1.32)
Month and year dummies	Yes	Yes	Yes	Yes	Yes
Cell fixed effects	Yes	Yes	Yes	Yes	Yes
N	695517	687449	680306	722911	723946
R2	0.0587	0.0567	0.0562	0.0551	0.0548
Share with signal	.03	.06	.07	0	0

Obs: ***1% **5% *10% significance levels. Linear regression of crackdown (binary) periods relative to the earliest signal in a cell-year. Only cells with positive deforestation were used for estimation, since only they can suffer enforcement action. Some cells did not have any signal, and are therefore the reference for all the period dummies. The regression is estimated by OLS, including fixed effects for month, year and cell (15km x 15km). Standard errors are clustered at the municipality level.

T3. Importance of acting quickly

Table 9: Characteristics of fines

	(1) Seized	(2) Seized	(3) Seized	(4) Area	(5) Area	(6) Area
Logging signal same month	0.0185***	0.0133**	0.0135*	38.96***	36.73***	39.89***
	(0.00599)	(0.00614)	(0.00686)	(7.541)	(7.585)	(9.729)
L. signal 1 month before	0.00397	0.00149	0.0125	40.69***	42.30***	43.44***
	(0.00687)	(0.00682)	(0.00987)	(7.260)	(7.439)	(9.526)
L. signal 2 months before	-0.00210	0.00153	-0.0125	43.20***	48.39***	46.82***
	(0.00649)	(0.00654)	(0.00854)	(8.274)	(8.784)	(12.10)
L. signal 3 months before	-0.00910	-0.00100	-0.0109	16.90***	24.07***	20.44**
	(0.00592)	(0.00581)	(0.00740)	(6.400)	(7.639)	(8.377)
Sample	All	All	Priority mun.	All	All	Priority mun.
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Municipality fixed effects	No	Yes	No	No	Yes	No
N	11893	11893	6327	11893	11893	6327
R2	0.0720	0.0752	0.0507	0.104	0.106	0.0889
Mean outcome	.08	.08	.08	124.45	124.45	124.45

Note: * 0.10 ** 0.05 *** 0.01 levels of significance. OLS regression of fine characteristics depending on whether they followed a recent logging signal. The two outcomes are the probability that the fine ended with seized equipment from the offenders, and the area (in hectares) of the inspected deforested area. The unit of observation is a 15km x 15km cell at the monthly level. Standard errors are clustered at the municipality level.

T4. General deterrence

Table 10: General deterrence effect

	$log(d_{it})$				$\mathbb{P}(d_{it} > 0)$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	2SLS	First	Reduced	OLS	2SLS	First	Reduced
Fine	0.722***	-1.915**			0.207***	-0.944*		
	(0.0223)	(0.925)			(0.00607)	(0.485)		
Cloudy			-0.0363***	0.0436**			-0.0125***	0.0238***
			(0.00527)	(0.0172)			(0.00326)	(0.00584)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Semester F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean outcome	1.16	1.16	0.16	1.16	0.53	0.53	0.10	0.53
N	34623	34623	34623	34623	65495	65495	65495	65495
R2	0.248	-0.315	0.0979	0.210	0.212	-0.350	0.100	0.199
F-statistic			20.78				15.06	

Obs: ***1% **5% *10% significance levels. The table shows regression results for two outcomes: log of size of deforestation (restricted to areas with positive deforestation) and probability of positive deforestation (restricted to areas that had some deforestation in 2011-2020). For each group there are four regressions. The first one is an OLS regression of the outcome on the occurrence of a fine. The second one is the 2SLS regression using clouds as an instrument. The third one is the first stage, showing the impact of clouds on the probability of a fine. The fourth one is the reduced form, showing the direct impact of cloudson the outcome. The observational level is a 15km x 15km cell-month in the Amazon forest, and only cells with a some positive level of deforestation in the period 2011-2020 were included. The outcome is measured at a yearly level, and divided by 12 to give a monthly interpretation to the regression coefficients. Standard errors are clustered at the 15km x 15km cell level. The regressions include the following controls at the cell level: federal road, state road, indigenous land, conservation unit, distances from Manaus, Cuiaba, Belem, and the closest IBAMA office, and yearly level of accumulated deforestation. The Cragg-Donaldson F-statistic for a test of instrument weakness is shown in the columns of the first stage.

T5. Share fire signal quality

Table 11

	(1)	(2)	(3)	(4)
Logging signal quality	0.158***	0.158***	0.192***	0.148***
	(11.09)	(10.21)	(12.44)	(7.96)
Indigenous land	0.0675	0.0111	-0.233	0.518
	(0.12)	(0.02)	(-0.36)	(0.61)
Conservation unit	-1.492**	-1.939***	-2.920***	-1.988*
	(-2.50)	(-3.18)	(-3.40)	(-1.82)
Priority municipality	4.376***	3.621**	2.531***	
	(2.65)	(2.06)	(3.00)	
Price ox		0.328***	0.255***	
		(12.33)	(9.53)	
Price soy		-0.120***	-0.107***	
		(-2.73)	(-2.59)	
Price coal		-2.355	-0.358	
		(-1.31)	(-0.69)	
Price wood		15.04***	15.84***	
		(2.65)	(3.53)	
Sample	All	All	All	Priority mun.
Mun. Fixed effects	Yes	Yes	No	No
Year dummies	Yes	No	No	No
Unit cell Controls	Yes	Yes	Yes	Yes
Clouds	Yes	Yes	Yes	Yes
N	34462	31036	31108	10264
R2	0.114	0.106	0.0495	0.135

Obs: ***1% **5% *10% significance levels. Linear regression of crackdown (binary) on positive deforestation (binary), logging signals and fire signals, with year interactions and controlling for several fixed and varying characteristics of the observations, as well as municipality and year fixed effects. Observational level is a 15km x 15km cell-year in the Amazon forest. Standard errors are clustered at the municipality-year level.

T6. Costs estimates

Table 12: Outcome: operational expenditure

	(1)	(2)	(3)
Inspection with alerts	7049.3**	6240.6*	3230.5
	(2983.7)	(3139.1)	(3185.6)
Inspection without alerts	15896.1***	16401.0***	6260.9**
	(2932.8)	(3082.2)	(2433.6)
Year dummies	No	Yes	Yes
State fixed effects	No	No	Yes
N	63	63	63
R2	0.654	0.683	0.935

Obs: ***1% **5% *10% significance levels. Linear regression of operational expenditures of IBAMA on the number of deforestation inspections. The data sources are budget expenditure data for years 2014-2020, in values of Brazilian Real of January 2020 (1 BRL = 4 USD). The deforestation inspections are taken from the administrative dataset on environmental fines, and compared at the month level with the locations of deforestation alerts at a 15km x 15km cell. Inspections are considered to follow an alert if they happen within the same cell at the latest three months after the alert. Standard errors are robust (White) qnd shown in parentheses.

Appendices

A Summary statistics

Table A1: Outcomes

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
1 Deforestation (km2)	.2654262	.2106781	.259974	.2467688	.2988944	.353212	.3383303	.3541823	.5350777	.5193492
	(1.02382)	(.8769656)	(1.102042)	(1.004691)	(1.226216)	(1.392089)	(1.277492)	(1.33321)	(1.92843)	(1.968994)
2 % cells with positive deforestation	26.45886	21.84961	23.22584	23.95471	23.37662	25.67018	25.73879	25.58687	29.11051	29.09091
	(44.11247)	(41.3236)	(42.22833)	(42.68175)	(42.32356)	(43.68241)	(43.72056)	(43.6359)	(45.4283)	(45.41928)
3 Deforestation as % of forest	.3290978	.2383326	.308867	.2878339	.3425139	.4060133	.3835201	.3762334	.6863465	.694053
	(1.603958)	(1.081387)	(1.513847)	(1.054278)	(1.320044)	(1.439758)	(1.449726)	(1.360148)	(2.696546)	(2.897824)
4 % of fire in deforestation	22.93893	13.25001	20.85406	12.65349	19.0664	24.42741	19.68144	24.28621	16.53636	17.55187
	(30.83754)	(24.12336)	(29.05567)	(22.97266)	(27.84682)	(29.68453)	(27.69978)	(29.5878)	(24.52846)	(25.23803)
5 Forest fires (km2)	2.413779	.8827375	1.421161	.8892843	1.357776	2.374192	1.555167	2.053755	1.290429	1.594017
	(7.103305)	(2.474242)	(4.278485)	(2.376738)	(3.741547)	(6.362968)	(4.276577)	(5.42448)	(3.792157)	(4.588439)
N	20307	20307	20404	20401	20405	20405	20405	20405	20405	20405

Obs: Mean and standard deviations of the outcome variables used in the study. The unit of observation is a 15km x 15km cell in a given year.

Table A2: Enforcement variables

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Deforestation crackdown	.0519525	.033683	.053715	.0458311	.058319	.0462142	.0508209	.0447929	.0418525	.0357755
	(.2219367)	(.1804163)	(.2254599)	(.2091237)	(.2343515)	(.2099537)	(.2196372)	(.2068541)	(.2002569)	(.1857346)
Total environmental fines*	42.97717	26.36494	49.14207	37.4781	50.90279	42.97597	38.56233	40.96942	35.23814	25.78182
	(64.99281)	(47.37189)	(91.19592)	(71.79812)	(95.46092)	(99.86462)	(69.03695)	(63.65941)	(49.19647)	(40.55786)
Total flora fines*	29.30239	16.41761	37.63325	29.44408	44.49323	35.18388	31.55098	28.66641	26.40553	20.21024
	(48.63064)	(30.18107)	(78.95736)	(60.75993)	(90.26377)	(88.17945)	(62.20887)	(50.82563)	(41.71788)	(35.25489)
Total deforestation fines*	15.24778	8.477295	29.15738	19.39327	27.19063	18.87518	17.18919	15.58405	14.9108	13.77553
	(26.44976)	(16.67455)	(65.55131)	(44.0056)	(52.62172)	(38.36518)	(33.15072)	(29.77979)	(26.36632)	(24.79145)
Deforestation fines	.1180874	.0651007	.1509508	.116759	.1693212	.1297721	.1321245	.1050723	.1020338	.0706199
	(.7757791)	(.5036035)	(1.283648)	(.9478014)	(1.288702)	(1.000499)	(.8513223)	(.7258064)	(.7548953)	(.5010886)
Inspected area - deforestation fines	535.8382	319.1243	545.0786	461.7437	527.1588	591.9043	677.8204	543.893	549.5421	570.0727
	(6286.231)	(4545.979)	(5681.434)	(5527.99)	(5605.993)	(7661.623)	(7966.827)	(6177.925)	(6042.145)	(6683.198)
Share inspected deforestation	1217.815	607.6077	609.9425	770.8841	624.7853	941.0154	997.3306	760.7041	540.3836	587.7106
	(11465.49)	(7718.209)	(5795.383)	(8565.699)	(5377.092)	(10248.96)	(9308.042)	(8947.663)	(6911.589)	(7715.848)
N	20307	20307	20404	20401	20405	20405	20405	20405	20405	20405

Obs: Mean and standard deviations of the enforcement variables used in the study. The unit of observation is a 15km x 15km cell in a given year.

Table A3: Characteristics (controls)

	Fixed	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Distance to Belém (km)	1405.12										
	(680.693)										
Distance to Manaus (km)	876.8191										
	(369.504)										
Distance to Cuiabá (km)	1384.899										
	(535.18)										
Shortest distance to IBAMA (km)	212.9281										
	(129.2514)										
Distance to closest federal road	99.74416										
	(95.41381)										
% with state road	.2870412										
	(.452382)										
% with federal road	.0667437										
	(.2495782)										
% indigenous land	25.23563										
	(40.88788)										
% conservation park	26.71074										
	(41.25416)										
% deforested		.1788433	.1801693	.1831963	.1845236	.1902239	.1917603	.193361	.1950535	.1967795	.1994097
		(.288144)	(.289156)	(.2908991)	(.2918908)	(.2975602)	(.2986323)	(.2997637)	(.3009289)	(.3020107)	(.303517)
% area as forest frontier		.1406703	.1420399	.1473837	.1463392	.1495253	.1517501	.1524734	.1551945	.156986	.1592723
		(.2355864)	(.2373793)	(.2416018)	(.2406238)	(.2437877)	(.2461842)	(.2465821)	(.2492588)	(.250589)	(.2523324)
N	224156	19833	19833	19930	19927	19931	19931	19931	19931	19931	19931

Obs: Mean and standard deviations of the control variables used in the study. The unit of observation is a 15km x 15km cell in a given year.

Table A4: Instruments

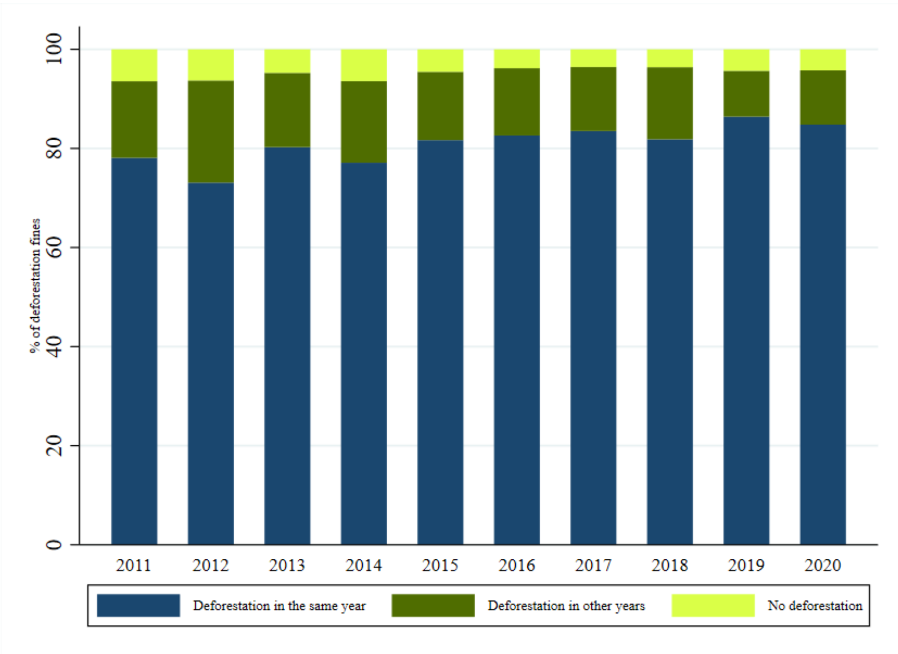
	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
% area covered with cloud	60.57544	39.00893	33.38096	47.95049	45.95313	36.37062	44.94082	.	.	.
	(16.76757)	(17.78106)	(15.89355)	(21.66128)	(19.6307)	(22.13561)	(18.89225)	(.)	(.)	(.)
% cloud January	92.59623	89.24037	75.15724	56.87693	65.22546	63.80798	78.73978	.	.	.
	(19.03889)	(18.76067)	(32.88083)	(35.27754)	(37.33796)	(38.54224)	(31.05746)	(.)	(.)	(.)
% cloud February	97.57204	64.7845	72.61333	79.3718	66.79065	55.36854	88.79403	.	.	.
	(12.24092)	(35.22684)	(33.88139)	(31.64051)	(37.61581)	(40.6135)	(24.2531)	(.)	(.)	(.)
% cloud March	83.00162	60.85564	61.73267	73.49115	72.75709	74.72589	69.96938	.	.	.
	(20.71694)	(35.43112)	(34.88645)	(30.69811)	(31.50097)	(34.60888)	(33.45851)	(.)	(.)	(.)
% cloud April	69.93829	53.37892	47.26916	66.03577	68.80991	56.64101	66.31811	.	.	.
	(37.05235)	(36.4093)	(36.85641)	(35.50767)	(33.89184)	(41.96188)	(36.08657)	(.)	(.)	(.)
% cloud May	37.95744	36.84292	47.87116	57.77042	59.1328	39.27594	43.97832	.	.	.
	(42.06426)	(35.85012)	(38.14238)	(36.92024)	(40.9706)	(38.97552)	(35.59687)	(.)	(.)	(.)
% cloud June	24.86741	18.61005	23.55973	30.85842	23.68315	21.12761	24.02633	.	.	.
	(34.87022)	(28.84492)	(30.04794)	(39.66283)	(33.8683)	(32.24403)	(34.31461)	(.)	(.)	(.)
% cloud July	8.934099	15.79768	14.78544	17.56279	17.44345	14.05419	13.6497	.	.	.
	(19.21792)	(27.12877)	(26.11981)	(28.2397)	(30.44963)	(25.9672)	(28.79048)	(.)	(.)	(.)
% cloud August	40.23689	3.993289	8.517482	15.54531	6.5673	10.09653	4.804969	.	.	.
	(37.19542)	(13.31091)	(18.78924)	(28.16072)	(19.09563)	(21.81281)	(14.29583)	(.)	(.)	(.)
% cloud September	40.16916	6.050032	4.039638	16.48229	9.423742	6.05813	17.66808	.	.	.
	(29.75331)	(15.12971)	(11.67646)	(26.95425)	(20.96731)	(16.55576)	(26.93725)	(.)	(.)	(.)
% cloud October	68.20777	19.41458	11.78465	31.85862	27.55719	10.94551	16.30834	.	.	.
	(25.56567)	(25.89088)	(19.28267)	(33.88339)	(33.36062)	(24.14709)	(27.69273)	(.)	(.)	(.)
% cloud November	80.48111	49.65245	33.26879	65.66872	52.14124	35.36487	43.76608	.	.	.
	(27.86975)	(30.81676)	(27.99752)	(33.72635)	(32.99355)	(35.6319)	(33.53274)	(.)	(.)	(.)
% cloud December	83.28384	49.5625	0	64.18643	82.27178	49.16998	71.43728	.	.	.
	(29.54058)	(32.49636)	(0)	(28.38874)	(26.2782)	(39.17429)	(29.66008)	(.)	(.)	(.)
Alert quality	3.924026	3.759378	4.78773	6.393847	11.61997	12.94871	18.56592	16.43549	16.90809	20.88673
	(13.85829)	(13.66127)	(15.27722)	(17.60343)	(22.49673)	(22.83537)	(26.65604)	(24.67622)	(24.5471)	(25.74542)
N	20303	20300	20392	20392	20401	20398	20401	5221	5940	5936

Obs: Mean and standard deviations of the control variables used in the study. The unit of observation is a 15km x 15km cell in a given year.

A Appendix figures

By restricting the sample to areas that received a deforestation fine, it is possible to compute the share of these areas which had deforestation in the same year as the inspection, some year in the sample, or no deforestation detected by satellites. This analysis shows that over 80% of IBAMA inspections occur in the same year of deforestation, with a tiny minority occurring in areas where no deforestation has been detected by satellites. These may be areas in which deforestation was not completed, and there was still some forest left, such that the area was not declared as "deforested" by satellite systems. This finding provides strong evidence that IBAMA's activity is focused on deterring current crime, as opposed to punishing past offenses.

Figure A1: Deforestation in areas with fines



How to Target Enforcement at Scale? Evidence from Tax Audits in Senegal

Pierre Bachas, Anne Brockmeyer, Alipio Ferreira, Bassirou Sarr*

May 10, 2022

(Click here for the latest version)

Abstract

Developing economies are characterized by limited compliance with government regulation, such as taxation. Resources for enforcement are scarce and audit cases are often selected in a discretionary manner. We study whether the increasing availability of digitized data help improve audit targeting. Leveraging a field experiment at scale in Senegal, we compare tax audits selected by inspectors to audits selected by a risk-scoring algorithm. We find that inspector-selected audits are more likely to be conducted, to uncover tax evasion and to detect larger amounts of evasion. We show, however, that the tax administration invests less manpower in algorithm-selected cases, and that algorithm-selected audits may generate less corruption, based on survey results. In ongoing work, we attempt to unpack the algorithm's (dis)functioning and the relevance of human capital in the audit selection and implementation process.

*Pierre Bachas: World Bank Research, pbachas@worldbank.org; Anne Brockmeyer: Institute for Fiscal Studies, anne.brockmeyer@ifs.org.uk; Alipio Ferreira: Toulouse School of Economics, alipio.ferreira@tse-fr.eu; Bassirou Sarr: Ecole des Hautes Etudes en Sciences Sociales (EHESS). We thank the Senegal Tax Administration (DGID) for an outstanding collaboration, in particular, Bassirou Samba Niasse, Amadou Abdoulaye Badiane, Oumar Diop Diagne, Hady Dieye, Mor Fall, Serigne Mabaye Fall and Mathiam Thioub. We thank Samba Mbaye, Assane Sylla and Medoune Sall from the CRDES, for the excellent partnership in the taxpayer survey. We also thank Denis Cogneau, Laurent Corthay, Janet Jiang, Nicola Limodio, Jan Loeprick, Markus Kitzmuller and Dan Rogger for helpful comments and discussions, and Oumy Thiandoum for excellent research assistance. Finally, we gratefully acknowledge the administrative support of the Paris School of Economics and CEPREMAP, and the financial support from the UK government through the Economic Development and Institutions Initiative (EDI) and the Centre for Tax Analysis in Developing Countries (TaxDev). The findings, interpretations, and conclusions are entirely those of the authors. They do not necessarily represent the views of the World Bank, its affiliated organizations, its Executive Directors or the governments they represent, nor the Senegal Tax Administration.

1 Introduction

Governments set up enforcement systems to ensure regulatory compliance, conducting inspections to punish infractions and impose the law. However, governments have limited resources to spend on enforcement activities. The scarcity of resources is even more severe in low-income countries, where non-compliance with the law is more widespread than in developed nations. Therefore, governments must do their best to allocate enforcement resources in the best possible way and reduce frictions in the enforcement process. One way to improve enforcement efficiency is to automatize decisions over inspections, such as tax inspections. In low-income countries, bureaucrats at enforcement agencies hold a significant degree of discretion over the inspection strategy, whereas developed countries tend to favor more automatized methods with algorithmic selection. How can we improve an enforcement system by moving from discretion to an automatized decision-making process?

In this paper, we study the implementation at scale of an automatized, data-driven selection method for tax audits of firms in Senegal. We created a data-driven algorithm that selected firms based on indicators, following best international practices. We then introduced the algorithm experimentally into the selection process of the tax inspectors in Senegal, with the agreement of the ministry of finance of Senegal. We use this experiment to assess how different selection methods affect audit quality by comparing audits selected by the two different selection methods: algorithm selection and discretionary selection. The experimental nature of the intervention allows us to provide the first evidence of how a shift from discretion to automatization can affect tax enforcement and understand the mechanisms that underlie the performance of an enforcement system. We compare the two selection methods across several dimensions: the probability that inspectors carry out the audit, the amount of recovered taxes in the inspections, and the quality of the inspections.

Efficient administrations are vital to building state capacity (Besley and Persson (2013)), by aligning correctly the incentives of bureaucrats with those of the state (Xu (2019), Bertrand et al. (2018); Finan, Olken, and Pande (2017)) or improving enforcement. The role of enforcement discretion is an open debate in the literature. On the one hand, discretion over inspection choice may be desirable if the bureaucrats' personal experience and soft information enable them to select inspections effectively. On the other hand, discretion can be more prone to human errors and corruption. Moreover, tax audit selection tends to be discretionary in developing countries but automatized in developed nations, where tax compliance is also higher (Khwaja, Awasthi, and Loeprick (2011)). This broad pattern suggests that reducing discretion would improve the enforcement system in developing countries. However, there is little empirical evidence about the benefits of reducing discretion in favor of an automatized, data-driven selection method.

We fill this gap by providing evidence from a field experiment conducted at scale in a developing country using a purely discretionary method to select its tax audits.

The experiment relied on an enormous effort of data digitization within the Senegalese tax authority, partly supervised by the research team. The tax authority in Senegal has been investing for years in digitizing its tax declarations, allowing systematic analysis and cross-referencing of files. Contributing to this data revolution, the research team helped finance and supervise the digitization of audit reports. This effort allowed us to understand evasion patterns and the role of selection methods on the quality of the audit process.

The experiment changed the set of firms selected for audit in several tax offices in the following way: inspectors chose half of the usual number of audits using their standard discretionary method, whereas the research group selected the other half through a risk score algorithm. The algorithm used data available for firms in Senegal to build indicators of potential evasion, using the best international practices. The indicators that compose the algorithm are easily interpretable for the inspectors, and we used them to create a "risk score". Due to the lack of digitized historical data on audits, we could not "train" the algorithm on historical data, using machine-learning tools, for example.

The intervention only targeted the selection methods without changing career or monetary incentives for tax inspectors[1], and tax inspectors could choose to devote efforts asymmetrically across selection methods. Moreover, firms were entirely unaware of the experiment, and we do not expect their tax declaration behavior to have changed due to it. Our goal is to understand how the quality of enforcement changes in response to an attempt to improve the audit selection method.

This study is the first to rigorously evaluate the differences between audit selection methods for tax audits. The IMF and World Bank have long advocated risk-based algorithms for audit selection. However, we know no impact evaluation of the adoption of such algorithms.[2] Moreover, we experimented at scale, intervening in half of the selected audits of the participating tax centers.[3] We also included a number of randomly chosen audits in the selection to use as a benchmark for analysis. Finally, we provided a cross-randomized information treatment to

[1] Changing the monetary incentives was not possible for legal reasons.

[2] According to Khwaja, Awasthi, and Loeprick (2011), for example, in the U.K., 55% of all cases are based on discretionary selection. In contrast, 35% and 10% of cases are respectively selected via a risk-scoring technique and a simple random sample. This approach is closest to the policy reform we introduce in Senegal. In other sub-Saharan African countries, Kenya uses a risk for all large taxpayers and discretionary selection for all others. Tanzania and Lesotho constitute examples on the extreme, respectively relying only on risk-scoring and random selection to audit all taxpayers.

[3] Selected firms represented 24% of corporate tax revenue of the tax centers in the experiment. The total amount of corporate tax liability (VAT and CIT) over the years 2015-2018 for the tax centers used was around 315 billion FCFA, and the selected firms in the 2019 program accounted for 75 billion FCFA.), and implemented directly by the audit planning and intelligence division of the tax administration.

test whether providing data by itself helps inspectors perform better when conducting audits. We submitted the experiment, hypothesis, and specifications to the AEA registry.

We conducted the experiment at scale, intervening in approximately half of the audit program in the participating tax centers.[4] The experiment included the two types of audits in Senegal: in-person full audits, carried out by groups of inspectors, and desk audits, carried out individually by inspectors from their offices. Each tax unit selected half of the cases planned for the *full (in-person) audit* program, and the risk-score assigned the remaining half. Moreover, each inspector selected 45% of her *desk audit* program, 45% came from the risk-score, and the remaining 10% were selected randomly. Moreover, we cross-randomized an *information treatment* for desk audits across the three selection methods. Information-treated cases received information on the most significant compliance risks detected by the risk score and detailed data from third parties regarding that taxpayer. The information treatment facilitates data access and analysis, thus potentially easing inspectors' work. We submitted the experiment, hypothesis, and specifications to the AEA registry.

The two selection methods select different sets of firms but have some overlapping ones. Upon receiving the two lists of selected audits (the one selected by the administration and the one selected by the algorithm), we see that inspectors tend to carry out the audits selected by their hierarchy more often. However, conditional on carrying out the audit, the average uncovered evasion per audit is similar across the two selection methods.

These null results are surprising, given that the sets of selected firms are so different for the two methods. To understand whether these results come from the algorithm or the behavior of the inspectors, we investigate how inspector quality and effort varied between algorithm-selected and inspectors-selected cases. We find strong evidence that the teams investigating algorithm-selected firms were smaller than for inspector-selected cases, but no evidence that education, age, or even time spent on audits differed systematically.

Overall, the results point to no visible improvement in the quality of the selection method. However, results also show that the two methods outperform completely random audits and that the overlapping cases outperform both the inspectors-selected and algorithm-selected cases. These results suggest that discretion yields better results than random audits, an automatized but uninformative method. In addition, the results on overlapping cases suggest that using data can help detect high-yielding cases among the pool of discretionarily selected cases.

This paper contributes to two strands of literature: i) the literature on the selection methods

[4]Selected firms represented 24% of corporate tax revenue of the tax centers in the experiment. The total amount of corporate tax liability (VAT and CIT) over the years 2015-2018 for the tax centers used was around 315 billion FCFA, and the selected firms in the 2019 program accounted for 75 billion FCFA.), and implemented directly by the audit planning and intelligence division of the tax administration.

for audits and ii) the literature on the quality of state bureaucracies (particularly tax administrations).

Our paper contributes to answering the question about the value of discretion in audit selection (Duflo et al. 2018, Kang and Silveira 2021). This question is particularly relevant in developing countries, where the trade-offs of using discretion are clearer. Similar to our paper, Duflo et al. (2018) also propose an experimental approach to estimate the impact of selection methods on an enforcement system. They compare discretionary to a random audits for environmental standards of plants in Gujarat, India. They find that discretion outperforms randomly selected audits. Thus, despite the flaws theoretically associated with discretion, inspectors seem to be able to find infractions and punish them more effectively than under purely random selection. In contrast to that paper, we propose a risk-based algorithm to select audits, but reach similar conclusions: inspectors seem to do a good job at selecting audits and uncovering evasion. Other studies papers have studied how technology can help bureaucracies improve its activities by reducing resource waste (Banerjee et al. 2020), and targeting inspections in a more accurate way (Glaeser et al. 2016, Bullock 2019, Glaeser et al. 2021).

The literature on the quality of bureaucracies tends to focus on human resources aspects, and their role in shaping outcomes such as regulatory compliance or quality of public services. Recent experimental evidence has shown that monetary incentives for tax inspectors improve the quality of inspections (Okunogbe and Pouliquen (2018)) and increase revenues (Khan, Khwaja, and Olken (2015)). Rasul and Rogger (2018) studied how management practices impact the quality of public services supplied by bureaucracts.

2 Institutional Setting: Senegal's Tax Administration

2.1 Taxes in Senegal

Tax revenue represented on average 16.7% of GDP in Senegal between 2013 and 2019. These revenue collection levels are below the West African Economic and Monetary Union (WAEMU) target of 20%, and fall short of goals set in Senegal's own medium term expenditure strategy. Tax gap estimates indicate that 23% of the theoretical VAT revenue is not collected (a shortfall of 2% of GDP) and that close to 63% of theoretical receipts from income taxes are missing (approximately 7% of GDP).

Similar to other developing countries, most taxes in Senegal are remitted by large and medium companies (Slemrod, Blumenthal, and Christian 2001). In particular, firms remit the Value Added Tax (VAT) and income taxes (Corporate income tax, personal income tax and dividend withholding taxes), accounting for 36% and 29% percentage of total tax revenue in 2019. Firms also withhold income taxes on their employees' wages (Pay-as-You-Earn), which is often the only source of reporting on salaried income, given the incompleteness of self-reported personal income taxes. Other significant revenue sources are customs duties (15%) and specific taxes on petroleum, which we do not cover in this study.

The Corporate Income Tax (CIT) is paid annually, at a rate of 30% profits or a 0.5% of turnover, whichever is larger. The Value Added Tax (VAT) is paid on a monthly basis, at a standard rate of 18% and a reduced rate of 10% for tourism businesses and hotels. A small number of financial sector firms pay the financial services tax instead of the VAT, also at a rate of 18%. Small firms with a yearly turnover of less than 50 million CFA Francs (about 100,000 USD) are eligible for a simplified tax (*Contribution globale unique*, CGU), which replaces all other taxes. The CGU is levied on turnover, at rates varying from 1% to 8%, where rates vary across sectors and increase in turnover. As already mentioned, the Pay-As-You-Earn taxes are withheld personal income tax on employees' wages with a formal employment contract.

2.2 Tax audits in Senegal

The Direction Générale des Impôts et des Domaines (DGID) is the administrative body in charge of domestic tax collection and enforcement, and reports to the Ministry of Finance. Figure A1 displays DGID's organizational chart. The large taxpayer directorate oversees firms whose turnover is greater than equal to 3 billion CFA francs (approximately 5.3 million USD) and has four units, which are specialized by economic sectors.[5] The medium taxpayer directorate oversees firms with less than with turnover between 100 million CFA francs and 3 billion CFA francs, and has two units. A third unit is in charge of the regulated liberal professions such as lawyers, notaries and medical practitioners. The remaining taxpayers, mostly small and medium enterprises (SMEs), are assigned to one of 19 regional tax offices.

There are two principal types of audits: desk audits and full audits.[6] Desk audits (or desk audits) are carried out by individual inspectors from within the tax authority's premises, using the firm's tax returns and, eventually, third-party data. Taxpayers are unaware of these audits unless inspectors make information requests, for example, when data is missing or seems inconsistent. Full audits are carried out by a team of inspectors at the taxpayer's premises. Full audits are announced at least five days before the audit starting date with an information request notice to the taxpayer. Tax inspectors may collect information for several weeks at the

[5]Unit 1 is in charge of the mining and energy sectors. Unit 2 deals with financial services and the telecommunications industry. Unit 3 covers real estate and firms. Unit 4 is a generalist one with broad competence covering all other sectors.

[6]There are also surprise audits which can take place either based on information that DGID receives either internally or from whistle-blowers. Surprise audits are similar to full audits, except that they are unannounced, as their name indicates.

taxpayer's premises and continue requesting information for up to 12 months.[7]

The selection method of tax audits in Senegal is essentially discretionary. Inspectors follow some rules of thumb, such as avoiding recently audited firms and firms with low turnover. However, there are no objective rules or formulas to add or drop firms from their selected program. Our study intervenes precisely at this stage of the tax administration's operation by including a machine-based selection in part of the audit program of the tax authority, both for short and full audits.

Figure A2 illustrates the steps in the audit process. After reviewing a case, inspectors list the detected irregularities and penalties and send them to the taxpayer in an "initial notice". They can also request additional information from the taxpayer. Upon receiving the initial notice, taxpayers have 30 days to respond to the inspector's findings.[8] The inspector examines the response has 60 days to prepare and send a "confirmation notice", again with the detected irregularities and penalties. The inspector then creates a revenue order for the tax collection unit, which requires the taxpayer to make a payment within ten business days. Taxpayers can appeal at the Minister of Finance or a judicial court, and the appeal may suspend the payment process temporarily.

3 Data

Our study draws on three sets of administrative data sources and two surveys. The three sets of administrative data are the tax declarations filed by taxpayers, third-party data on transactions, and audit outcomes. We discuss details of the matching process and match rates in Figure E. We complement the administrative datasets with a taxpayer survey and a tax inspector survey, which were designed by the research team and were not available to the Senegalese tax authorities.

Tax Declarations. Table A6, Panel A, provides an overview of the available tax declarations. Our primary sources of information are the tax declarations on Corporate Income Tax, Value Added Tax, and the Pay-As-You-Earn tax (withheld progressive personal income tax), covering the period of 2014-2019. The CIT data covers about 4 thousand firms per year, and the VAT data around 8 thousand firms.[9] Finally, we match these data with monthly Pay-As-You-Earn data, which allows us to calculate the number of employees and the aggregate wage bill for each

[7]For firms with a turnover of less than 1 billion CFA francs (about 2 million USD), full audits can only last up to four months. These maximum limits are general rules. There may be extensions in cases with highly suspicious activity or when there is a delay in the transmittal of the requested information to auditors.

[8]If the taxpayer fails to respond, it means for legal purposes that they agree with the inspector's findings.

[9]Many more firms declare VAT than CIT because self-employed individuals and unincorporated firms file VAT but not CIT.

firm.

Third-Party Data. Table A6, Panel B, describes the third-party data, that is, information about transactions of companies which we obtain from third parties. The third-party datasets are the import-export transactions (customs data), payments from state institutions to firms (procurement data), and in recent years VAT annexes documenting transactions between firms.[10] These datasets are at the transaction level, and we aggregate them at the firm-year level to merge with the tax data. As the last two columns in Table A6 indicate, a non-negligible share of firms captured in the third-party data fail to file taxes in the corresponding year. The share of taxpayers for whom third-party data is available hovers around 28%, with the share increasing over time and in firm size.

Audits data. We collect selected audit programs and audit results data for fiscal years 2018, 2019, and 2020. The selected audit programs are partly produced by the risk-scoring algorithm, and the rest by the inspectors themselves. The audit results contain information on key audit process steps: audit announcement, notification, confirmation, and payment request. The audit results data contains several ad hoc audits which were carried out despite not being initially programmed. The data contain the inspector's name, taxes verified in the audit, infractions detected, evaded amounts, applicable penalties, and the dates of each step. We use this information to compute our outcomes, such as audit yield and evasion rates. Moreover, we asked inspectors to fill in spreadsheets with qualitative information about each audit case, such as the perceived difficulty of the audit, whether the taxpayer was uncooperative, the business activities were complex, or information was unavailable.

Tax Inspector Survey. Prior to our intervention, we conducted a detailed survey among all participating tax inspectors, capturing information about their demographics, employment history, perceptions of the audit function, methods for audit selection, and use of different sources of information. The survey data contain 97 inspectors, which represents approximately 1/3 of inspectors involved in audits in 2018-2020 in the centers under analysis.

Taxpayer Survey. We surveyed approximately 750 firms in the Dakar region, most of which had been audited shortly before. We conducted the taxpayer survey in two waves, from October to December 2020 and March to May 2021. The survey allowed us to elicit taxpayers' perspectives on tax inspections, audit risk, and their opinions on the tax authority.

[10] VAT annexes have become increasingly available in recent years, following efforts by the tax administration to digitize information and require that taxpayers file their VAT annexes electronically.

3.1 Descriptives

Table A8 provides summary statistics of the firms included in the experiment and the population. It is clear from the picture that the sample of firms selected for audits is much larger and profitable than the average firm in the population. This pattern comes essentially from the fact that audits are concentrated at larger centers, which focus on larger firms.

Figure A3 show how firm size is distributed in the population (CDF), and how it is related to the probability of being selected into audit, the conditional probability of the audit being started, and the evasion rate. Larger firms clearly attract much more attention than smaller firms, and though their absolute evasion amounts are typically large, evasion rates fall with firm size.

4 Audit Selection and Experimental Design

4.1 Discretionary Selection

Until 2018, all audit cases in Senegal were selected exclusively with a discretionary procedure. At the beginning of the year, the Director-general of the tax authority requests each unit to propose the annual program of firm audits. Each unit suggests a set of full audits and desk audits, the latter suggested by the inspectors that will conduct them individually.[11]

Tax inspectors use a standardized form to motivate the full audit selection. The form contains information on the identity of the selected firm, past audit history, and a summary of relevant indicators such as tax turnover and profit margin. Once the tax unit's manager approves the form, a selection committee in the Director-general's office finalizes the list of firms for the full audit program. The committee accepts most proposed cases, though the committee may request additional information, reject proposals, or add their proposals based, for example, on denunciations. The committee then returns the names of approved audits to tax units.[12]The selection of desk audits also takes place at the beginning of the year, but the procedure is simpler than for full audits. Individual inspectors propose cases to their tax unit's director without any particular guideline.

[11]Since desk audits are selected individually, different inspectors might select the same taxpayer; in practice, this is rare as inspectors specialize by economic sectors or geographical areas. When this happens, the manager presumably rules which inspector is in charge of the case.

[12]This description is based on interviews with members of the committee.

4.2 Risk-Score Method

In the past decade, the Senegalese tax administration has invested in digitizing its tax data, widening the availability of information about its taxpayers and creating the opportunity to select audits selection in a data-driven way. The cooperation between the researchers and the tax authority started in 2017, first by mapping available data sources and indicators that could be useful to assess compliance risk. We designed a risk-scoring tool based on a set of indicators, drawing on work by the World Bank (tax administration projects in Pakistan and Turkey), SKAT in Denmark, and the IMF's recommendations to Senegal.

We designed an algorithm based on intuitive indicators, which we discussed and explained to the tax authority staff. We preferred this method rather than a machine-learning tool, which would yield a less transparent selection. There are two reasons for preferring an indicator-based parametric algorithm to a nonparametric machine-learning algorithm. First, we needed a simple and transparent tool that would easily convey the identified compliance risks associated with a firm to tax inspectors. Second, the available data on historical digitized audit results was sparse, limiting the scope for model training and prediction of tax evasion.[13] Our proposed risk-score tool is a transparent risk assessment based on international best-practice, designed in cooperation and dialogue with the tax authority taking into account their capacity constraints. The constraints faced by DGID are likely to bind in many low-income countries, especially in West Africa, which often looks at Senegal for administrative innovations.

Table A1 summarizes the seven critical steps in the design of the risk score algorithm. Step (1) corresponded to the construction of a database covering all tax declarations across years and merged with third-party reported sources, as discussed in section 3. Steps (2) and (3) determined the risk indicators based on intra-firm discrepancies across data sources and inter-firm anomalies based on comparisons with similar firms. Step (4) defined the peer-group comparison clusters, defined by economic activity and tax center. Step (5) assigned a numerical value to each risk indicator, depending on the size of the inconsistency or anomaly (with higher scores for larger discrepancies). Step (6) assigned weights to each indicator, reflecting our judgment about their relative importance. Finally, step (7) aggregated the weighted indicators over the past four fiscal years to form a single risk score.

As already mentioned, the risk score relies on two types of risk indicators: discrepancies and anomalies. Discrepancies are intra-firm indicators, which flag taxpayers with inconsistent information across different datasets. For example, a discrepancy arises if the self-reported turnover is inferior to what we can expect from reading customs data, state procurement, and transact-

[13] In the early stages of the design, we implemented a random-forest algorithm to predict evasion, which predicted historical evasion with similar degrees of accuracy as the parametric indicators, but which was far less easy to manipulate and interpret.

ing partners. In contrast, anomalies indicators are inter-firm indicators, which compare a firm to a group of similar peers. An example is a firm with an abnormally low margin of profits relative to its peers. Firms were given a higher risk score for all indicators depending on how severe the irregularity seemed to be. In the last iteration of the algorithm, we included four discrepancy indicators and six anomaly indicators to construct the risk score. We over-weighted the discrepancies compared to anomalies to reflect the higher confidence that discrepancies reflect non-compliance, while anomalies might only reflect temporary economic problems or poor management.

4.3 Study Design

Our experiment changed the set of firms selected for audit by using a different selection method. We implemented the algorithm selection in 2018, 2019, and 2020, each year increasing the number of tax offices included in the experiment. In all offices except the Large Taxpayer Unit, the tax authority agreed to let the algorithm select half of its audits program. In the Large Taxpayers Unit, the administration only allowed the algorithm to pick one-third of cases. The intervention happened both for the selection of full audits and desk audits. Desk audits are shorter, simpler, and cheaper to carry out, and their number is typically twice as large as the number of full audits. In the case of desk audits, we selected 10% of firms completely randomly. Moreover, we provided a cross-randomized information treatment on desk audits. This treatment consisted in attaching a spreadsheet with crucial information about the selected firm in a random draw of cases, regardless of their selection method.

In summary, 50% of full audits were selected by the discretionary method (that is, by inspectors) and 50% by the risk score algorithm (except for the Large Taxpayers Unit). In contrast, 40% of desk audits were selected by the discretionary method, 40% by the algorithm, and 20% at random. The exact number of cases varies by tax center as displayed in Tables A3 and A2 for desk and full audits.

The selection of the two audit programs - full and desk audits - proceeded in three main steps. First, inspectors selected cases at their discretion and submitted them to their hierarchy. Secondly, we ranked firms based on the computed risk scores within each tax center and cluster of economic activity. For each audit selected by the inspectors in a given tax office and cluster, we chose the firm with the highest risk score within the same tax office and cluster. This method allowed for some overlaps between algorithm and inspector selection.[14] As already alluded to above, in the case of desk audits, we also selected some firms at random. Third, a committee within the tax administration reviewed all the selections, excluding some recently

[14] In case of overlap between algorithm and discretionary methods, we add more algorithm cases until we meet the pre-agreed number of "only" algorithm cases.

audited firms or firms that should not be audited for other reasons (e.g. political reasons). The administration then sent the approved lists to the tax centers and individual inspectors, who would then start implementing the audits.[15]

The information treatment implemented on desk audits was cross-randomized across the selection methods: algorithm, inspectors' selection, and random. We provided two types of information treatment: soft and strong. The soft information treatment consisted of attaching the three main flags detected by the algorithm to the selected case, even if the inspectors chose it. The strong treatment consisted of attaching the indicators plus a spreadsheet with the relevant data of that firm. The spreadsheet contained all the tax declarations and third-party information of the firm in the past few years. The goal of the information treatment is to test whether easing informational constraints may have an impact on audit outcomes.

In summary, inspectors received a list containing firms to be inspected. The ones selected by the algorithm contained the mention "selected by new methods". Random cases (only applicable for desk audits) also included this mention. Inspector-selected audits had the comment "selected by the administration" next to them. Therefore, inspectors knew whether the algorithm or their hierarchy (with their inputs) selected each firm on their list. However, they could neither distinguish between random and algorithm cases nor between inspector-selected and overlapping cases.

We took some steps to induce inspectors to comply with the experiment. First, we shared with inspectors a methodological note containing the indicators used in the algorithm. Second, we presented the algorithm at a workshop organized by the intelligence unit of the tax authority in Dakar. Finally, we proposed that the inspectors carry out the algorithm and discretionary audits in alternation.[16]

5 Empirical specification

Changing the selection method is expected to alter the set of selected firms substantially. Therefore, the first part of the empirical analysis estimates the dimensions in which the selection methods diverge. Next, we test how the selection methods differ for three audit outcomes: i) audit completion rates, ii) whether evasion was detected during the audit, and iii) how much

[15]The Director General's office informed inspectors about the experiment, urging them to follow guidelines in carrying out audits at the proposed sequencing and reporting audit results rigorously. This complements presentations by the intelligence unit of DGID and the research team to each center.

[16]We randomly ordered each tax inspector's list of cases. This ensures that the order of audits is uncorrelated with audit quality. However, we were unable to ensure discipline in following the designated sequence for the workload. For instance, inspectors could choose to prioritize cases they select themselves and which they believe could leave to higher yield. Nonetheless, the Director General signed a guideline urging staff to follow the sequence set in their assignments.

evasion was detected. We also test how the information treatment affects these outcomes. Finally, we investigate audit quality and inspectors' effort (section 7 on mechanisms).

We do most of the analysis separately for desk and full audits, since they are carried out in different ways. However, to estimate systematic differences between the selection methods with respect to the characteristics of selected firms, we estimate a slightly modified version of equation 2 which lumps together the observations of short and full audits and includes interactions between selection method and audit type:

$$\begin{aligned} y_{io} = \beta_0 &+ \beta_1 Algorithm_{io} \times DeskAudits_{io} + \beta_2 DeskAudits_{io} \\ &+ \beta_3 Algorithm_{io} \times FullAudits_{io} + \beta_4 FullAudits_{io} + \gamma_o + \varepsilon_{io} \end{aligned} \tag{1}$$

Where β_1 captures the average difference of the outcome between algorithm desk audits and inspector desk audits, β_2 captures the difference between inspector desk audits and the population, including non-selected firms, β_3 captures the difference between algorithm full audits and inspectors full audits, and β_4 captures the difference between inspector full audits and the population.

The main empirical strategy to understand the impact of selection methods on audit outcomes is a linear regression of audit outcomes on selection methods. We estimate the following model:

$$y_{iot} = \beta_0 + \beta_1 Algorithm_{iot} + \beta_2 Overlap_{iot} (+\beta_3 Random_{iot}) + \delta_t + \gamma_o + \varepsilon_{iot} \tag{2}$$

Where y_{iot} is the outcome of an audit for case i, registered in tax office o and selected for audit in year t, and ε_{iot} is a conditional mean zero error term. *Algorithm* is an indicator function that is equal to 1 if the firm was selected for audit by the algorithm, and 0 otherwise. We define in a similar way *Random* for randomly selected cases, which is in parentheses because it applies only for the specifications using desk audits data. We always analyse desk and full audits separately. Finally, δ_t are year fixed effects, and γ_o tax office fixed effects. We estimate the model by ordinary least squares.

We used for the analysis the sample of selected firms over the years 2018-2020. Within this sample, we could tell how these firms were selected (inspector selection, algorithm, or random) and the type of audit for which they were selected (desk or full audit). The specification in equation 2 allows us to compare average outcomes across different methods for centers in which we carried out the experiment. This means that the coefficient β_1 identifies the effect of using an algorithm as opposed to a inspectors-based method on the outcome, that is:

$$\beta_1 = \mathbb{E}[y_{iot}|\text{Algorithm}] - \mathbb{E}[y_{iot}|\text{Inspectors}] \tag{3}$$

The coefficient β_1 identifies the difference in mean outcomes across the selection methods, as in a horse race between them. The experimental design ensures an exogenous treatment of the selection process, in the sense that the tax authority accepted that approximately half its selected audits would be chosen by this alternative method.

Finally, we use a slightly different specification to study the role of the information treatment on outcomes. The information treatment is only relevant for the desk audits, so only the sample of those audits is used. Moreover, the treatment is cross-randomized across the three selection methods: algorithm, random and inspectors. Therefore, we study the effect of providing the treatment by itself and the interactions with each selection method. The specification used is the following:

$$\begin{aligned} y_{iot} = \beta_0 &+ \beta_1 Algorithm_{iot} + \beta_2 Algorithm_{iot} \times Information_{iot} \\ &+ \beta_3 Random_{iot} + \beta_4 Random_{iot} \times Information_{iot} \\ &+ \beta_5 Information_{iot} + \gamma_o + \delta_t + \varepsilon_{iot} \end{aligned} \tag{4}$$

As mentioned earlier, the information treatment was implemented in a weak and a strong form. The weak form only provided a list of indicators linked to the firm, whereas the strong form provided the indicators plus the available data about the firm. To distinguish between differential effects of these treatments, we estimate a version of equation 4 with each of the two treatments plus their interactions with the selection methods.

For all specifications, the standard errors are clustered at the tax office level. Clustering at this level allows for different estimated variances and autocorrelation among firms in within each tax office, but excludes autocorrelation between firms in different tax offices. We believe this may be realistic since tax offices specialize by firm size and sectors of economic activity. For example, the Large Taxpayer Unit is divided in four offices, the medium taxpayer office in two, and firms are categorized based on their economic sector. For SME tax offices, the assignment of firms to tax offices is based on their geography.

6 Results

6.1 Characteristics of algorithm vs inspector-selected firms

We designed the algorithm to improve the inspectors' selection, given their objectives to select firms with high evasion. Moreover, the number of firms selected by the algorithm matched the inspectors' selection by tax centers and sectors of economic activity (the "clusters", as explained above). Therefore, the extent to which algorithmic selection would differ from inspectors' selection was not immediately apparent from the outset of the selection. We estimate equation 1 to illustrate the systematic differences in firm characteristics between the selection method.

The outcomes are characteristics of firms that come from their tax declarations: (log) mean declared sales of the firm averaged over 2014-2020, profit rates, and payroll expenditure. We estimate the equation using the whole sample of firms, but also on later stages of the audit: only for the sample of started audits, then only for the sample of audits with confirmed evasion. For these later stages, the coefficient on Desk Audits is omitted, because there is no "non-selected" group to compare the outcomes with. Table A7 summarizes this difference at each of the three stages.

Upon performing the selection, we assessed that the two methods differ systematically, even within tax centers and sectors of economic activity. The main difference between the two methods is that the algorithm-selected firms seem to declare lower total sales, lower profit rates, and lower payroll expenditure. This difference means that firms with lower tax self-declared quantities were more likely to be flagged by the algorithm's indicators, even conditional on tax office and sector of economic activity. Switching selection methods changes which firms get selected in a clear way, in particular by selecting firms with lower tax declarations. In what follows, we show how the audit outcomes differ depending on the selection method.

6.2 Effect of audit outcomes

6.2.1 Probability of starting audit

We analyze the probability of inspectors' starting case i conditional on the selection method. We observe that inspectors started a case when they filled out at least one key information regarding the audit process, such as an information request to the taxpayer. Inspectors know the selection method of each case and possibly take that into account in deciding to start a case. If discretionary cases are easier to conduct or have a higher expected return, inspectors may be reluctant to open algorithm-selected cases.

The outcome is a dummy variable that takes value 1 if inspectors opened the case and 0 otherwise. We can tell that a case has been started by the fact that inspectors registered some information on the audit report. Usually, this information is a date indicating that they have taken some action, such as starting the audit or demanding some information from the taxpayer. This outcome captures the effort of inspectors to study the cases and allows us to understand whether inspectors were more selective when starting algorithm cases than starting their own cases.

Table A9 summarizes the results for the probability of starting the audit by estimating equation 2 with some additional controls. Each column shows a linear probability model, and the coefficient on *Algorithm* shows the difference between the average probability of starting an algorithm case versus a discretionary case. The left panel refers to full audits, and the right panel to desk audits. The results differ strongly for full or desk audits. On average, algorithm-selected full audits are 16-17% less likely to be opened, but desk audits selected by the algorithm are just as likely to be started as inspector-selected cases or randomly selected ones. Interestingly, overlapping cases were the most likely category to be implemented, even though inspectors were unaware that the algorithm had picked those cases. The coefficient on the overlapping cases is positive and large for short audits, but not significant. The table also shows that the average rate of implementation of the audits program is quite low: 63% of programmed full audits were implemented, and only 37% of desk audits.

The results indicate that inspectors were very selective concerning algorithm cases when it comes to the more costly type of audit, i.e. full audits. However, among the inspector cases, the ones with the largest risk scores (the overlapping cases) were the most likely to be picked from the list.

6.2.2 Audit outcomes

Next, we investigate the impact of selection on audit outcomes after inspectors choose to start an audit. We measure audit yield in two ways: whether inspectors found an infraction (binary variable) and the value of the payment required by the audit report (i.e., the sum of the assessed evasion and penalties). The results discussed below are based on the confirmed amounts of the audit.[17]

We condition the sample on the started cases to study the (confirmed) adjustment. Conditional on starting the audit, the probability that an algorithm case had a confirmed evasion amount was as large as an inspector-selected case. The results are summarized in table Table A10. Table A11 shows the results for the value of the adjustment (in log).

Both tables show a qualitatively similar figure. They show positive point estimates for algorithm cases in full audits and negative point estimates for desk audits. However, these effects are not statistically significant, and they are very small for the probability of having a positive adjustment. For the amounts of evasion, the point estimates are quite large, suggesting that the average algorithm full audit uncovered more evasion than the inspector selected audits. The average algorithm desk audit uncovered much less evasion than their inspector-selected counterparts. The coefficients for overlapping cases are very similar to the coefficients on

[17] The inspectors initially issue a notification to the taxpayer. The taxpayer then clarifies problems, and the inspector issues the confirmation.

algorithm cases.

The more interesting and striking results on Tables A10 and A11 concern the random audits. The probability that these audits found positive evasion was on average 15% weaker than for algorithm or inspector-selected cases, and the adjustment quantities were about 30% of the values found for inspector-selected cases.[18]

In summary, inspectors were careful and "choosy" when deciding to start the algorithm's full audit, but this selectiveness paid off (as shown in Table A9). Meanwhile, they did not discriminate against algorithm desk audits, but their uncovered evasion faltered relative to the inspector cases. Random cases performed worse, showing that discretionary and algorithmic selection are superior ways to find potential tax evaders. This result echoes Duflo et al. (2018), who showed that random environmental audits were less likely to uncover infractions than discretionary audits. In our case, we show that random tax audits also perform poorly, and a data-driven algorithm can perform at least as well as discretionary selection. However, the algorithm does not beat discretion.

6.3 Information treatment

Tables A12, A13, and A14 show the estimated coefficients for two versions of equation 4: one with the detailed information treatment (distinguishing between weak and strong treatment), and one with a dummy lumping together both treatments. The tables are organized as follows. The first column shows the impact of the detailed information treatment by itself. This specification is a valid way to estimate the average treatment effect because the treatment was randomized. The second column shows the interaction between the treatment and the selection method. The third and fourth column are constructed similarly, but use only the lumped treatment (labelled as "Any information").

The results show that the information treatment helped spur inspectors to start audits (Table A12). The coefficients are positive for both treatments, though not always significant. The coefficients on the interactions with random or algorithm selection are never significant. Moreover, the information treatment seems to play no role in the probability of finding evasion or its quantities, except for the case of random audits. Random audits with information on indicators and data found larger evasion quantities.

In summary, providing some information seems to motivate inspectors to start an audit, including for their own selected cases. However, the information conveyed by this treatment had little effect on the outcome of the audits.

[18]The estimated coefficient is around -2.7. Since this value is in logs, this means that the conditional expected value for evasion for inspector selected cases is at least 270% larger than the random cases.

7 Mechanisms

7.1 Effort by inspectors

Changing the selection method can change outcomes simply because the selection method is different. As we showed previously, it is indeed the case that the algorithm picks a very different set of firms from the inspectors. However, since inspectors were aware of which cases were selected by each method, it is possible that they exerted different efforts in carrying them out. Moreover, it is also possible that reducing discretion in inspection selection may increase their diligence and reduce corruption.

To investigate the relationship between selection methods and audit quality, we rely on information collected via a taxpayer survey carried out with a sample of inspected firms.[19] We asked the taxpayers several questions regarding their experience with tax inspections: the length of their most recent full audit, and how they would rate the inspectors in terms of technical knowledge, honesty, and efficiency.

The table suggests that recently audited firms, by the desk and full audits alike, tended to report more extended audits and a more positive view of their inspectors. The algorithm selection does not have a systematic difference with respect to other firms interviewed. Finally, the last row does not allow us to distinguish the quality of algorithm audits from other types of audits. Therefore, there is no evidence that the inspectors exerted different efforts for algorithm cases instead of their own.

7.2 Inspector quality

Besides exerting more or less effort on different cases, it is also possible that agents sorted themselves differently into doing algorithm cases based on their ability. However, this problem is irrelevant for short audits since the inspector assignment is already done at the selection stage. In contrast, the tax administration has some leeway to allocate inspectors and teams to full audits.

We implemented a survey with the inspectors of the tax authority to obtain information about their ability and experience. We can then test whether the inspectors that did algorithm cases differed on dimensions that could be relevant for audit outcomes. Table A19 tests this hypothesis using data for full audits, where we computed statistics on the inspectors that worked in each particular case.

[19] Some firms in the survey sample were not inspected in the period of the experiment, but they constituted a minority.

The table shows that inspectors working on a case selected by the algorithm were just as likely to have a masters or PhD degree as inspectors on other cases. The table does not find any other significant differences for other variables, such as mean age, mean experience, and the share of the team that declared to be supportive of algorithmic selection.[20] On average, algorithm cases had slightly younger agents, but the differences are not statistically significant.

7.3 Manpower

Another way inspectors could tinker with quality is by assigning different manpower to the algorithm or inspector cases. Inspectors carry out full audits in teams, often with four or five agents working on the same case. Short audits are carried out individually most of the time. We constructed an outcome that counts how many agents were involved in each recorded audit, and analyse how the number of agents differs across selection methods. The analysis is done by estimating equation 2 having the number of agents as outcome at the audit level.

Table A16 summarizes the results. We see that full audits have, on average, three agents per audit, whereas the average for desk audits is 1.4. Columns 4 to 7 show no difference between the number of inspectors for desk audits between algorithm and inspector cases. This result is unsurprising given that typically these audits are carried out individually, but they also suggest that algorithm cases were no less likely to be carried out by teams. When it comes to full audits, however, algorithm cases had smaller teams by an average of 0.3 agents. The results are significant and fall to 0.2 after the inclusion of controls for turnover deciles and economic activity fixed effects. However, the size of the coefficient is modest (approximately 10% of the mean size of teams).

In summary, there is evidence that algorithm full audit cases had somewhat less staff allocated to them, while at the same time they were more likely to find evasion. This is the only result suggesting a clear efficiency gain arising from the algorithm selection method.

8 Conclusion

In this paper, we examine the impact of changing the selection method of tax audits from a discretionary to an automatized method. We apply best international practices to create an algorithm that selects firms based on indicators of tax evasion and selected the riskiest firms for audit. Inspectors selected an equal number of audits in the same way. We investigate how likely inspectors are to carry out the audits, how much evasion is discovered, and whether audit quality changes with selection methods.

[20] The exact question in the survey was: "Would you be in favor of automatizing audit selection?"

We find that inspectors tended to treat both types of cases differently along two dimensions only: the probability of opening the case and the human resources allocated to the case. Algorithm audits were less likely to be started and had slightly smaller groups working on them. This result only holds for full audits, whereas desk audits were similarly implemented across the different selection methods.

In terms of audit outcomes, the discretionary method seems to work just as well as the algorithm. However, random audits perform poorly in terms of helping inspectors uncover evasion. A combination of algorithm and discretionary method (the overlap cases) had the largest probability of being started. However these cases did not show significantly larger evasion rates on average than the algorithm or discretionary cases. An information treatment cross-randomized across selection methods seems to play a role in spurring inspectors to open a case, but there is no evidence that it helps them uncover evasion during the investigation.

Overall, our results do not allow any conclusion that moving away from discretion has clear benefits to improving the efficiency of enforcement. The use of data to select audits seems to provide efficiency gains because it requires less manpower to uncover similar levels of evasion. However, the differences between algorithm outcomes and discretionary outcomes are modest an often not statistically significant.

References

Banerjee, Abhijit, Esther Duflo, Clement Imbert, Santhosh Mathew, and Rohini Pande (2020). "E-governance, accountability, and leakage in public programs: Experimental evidence from a financial management reform in india". In: *American Economic Journal: Applied Economics* 12.4, pp. 39–72.

Bertrand, Marianne, Robin Burgess, Arunish Chawla, and Guo Xu (2018). *The Glittering Prizes: Career Incentives and Bureaucrat Performance*. Tech. rep. mimeo.

Besley, Timothy and Torsten Persson (2013). "Taxation and development". In: *Handbook of public economics*. Vol. 5. Elsevier, pp. 51–110.

Best, Michael, Jawad Shah, and Mazhar Waseem (2021). "Detection Without Deterrence: Long-Run Effects of Tax Audit on Firm Behavior". In: *WP*.

Bullock, Justin B (2019). "Artificial intelligence, discretion, and bureaucracy". In: *The American Review of Public Administration* 49.7, pp. 751–761.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan (2018). "The value of regulatory discretion: Estimates from environmental inspections in India". In: *Econometrica* 86.6, pp. 2123–2160.

Finan, Frederico, Benjamin A Olken, and Rohini Pande (2017). "The personnel economics of the developing state". In: *Handbook of Economic Field Experiments*. Vol. 2. Elsevier, pp. 467–514.

Glaeser, Edward L, Andrew Hillis, Hyunjin Kim, Scott Duke Kominers, and Michael Luca (2021). "Decision Authority and the Returns to Algorithms". In.

Glaeser, Edward L, Andrew Hillis, Scott Duke Kominers, and Michael Luca (2016). "Crowdsourcing city government: Using tournaments to improve inspection accuracy". In: *American Economic Review* 106.5, pp. 114–18.

Kang, Karam and Bernardo S Silveira (2021). "Understanding disparities in punishment: Regulator preferences and expertise". In: *Journal of Political Economy* 129.10, pp. 000–000.

Khan, Adnan Q, Asim I Khwaja, and Benjamin A Olken (2015). "Tax farming redux: Experimental evidence on performance pay for tax collectors". In: *The Quarterly Journal of Economics* 131.1, pp. 219–271.

Khwaja, Muwer Sultan, Rajul Awasthi, and Jan Loeprick (2011). *Risk-based tax audits: approaches and country experiences*. The World Bank.

Okunogbe, Oyebola Motunrayo and Victor Pouliquen (2018). "Technology, taxation, and corruption: evidence from the introduction of electronic tax filing". In: *World Bank Policy Research Working Paper* 8452.

Rasul, Imran and Daniel Rogger (2018). "Management of bureaucrats and public service delivery: Evidence from the nigerian civil service". In: *The Economic Journal* 128.608, pp. 413–446.

Slemrod, Joel, Marsha Blumenthal, and Charles Christian (2001). "Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota". In: *Journal of public economics* 79.3, pp. 455–483.

Xu, Guo (2019). "The colonial origins of fiscal capacity: Evidence from patronage governors". In: *Journal of Comparative Economics*.

A Audit procedure

Table A1: Steps of risk-score design

Step	Description
(1) Prepare database	The tax declarations of each taxpayer are merged across type of taxes (VAT, CIT, Payroll) and across years. Data from third parties is then added (customs, procurement, transaction network).
(2) Choose indicators: discrepancies	Discrepancies are situations in which a self-reported tax liability can be considered as misreported or incomplete, by cross checking several data sources together.
(3) Choose indicators: anomalies	Anomalies correspond to abnormal reporting behavior, compared to peers. Anomalies suggest that firms should be monitored, but do not indicate tax evasion behavior with certainty.
(4) Define comparison clusters	Clusters regroup firms in the same economic sector and of comparable size. Peer comparisons are done within clusters
(5) Assign values to indicators	The magnitude of the inconsistency is used to assign a value, ranging from one to ten (using deciles). For anomalies firms within the top decile of a particular indicator receive a value of one.
(6) Assign weights to indicators	Weights are assigned to each indicator reflecting beliefs about their relative importance.
(7) Aggregate indicators and years	The weighted risk indicators are first aggregated across indicators in each year. Then the yearly scores are summed up to form a total risk score covering the past four years of tax declarations. More recent years are slightly over-weighted.

B Program execution

The following sections provide an analysis of the 2019 audit reports, executed in the scope of an experiment in partnership with the Senegalese Internal Revenue Services (DGID in the French acronym, henceforth designated IRS). The experiment consisted in altering the selection method of the audits program of 2019 in some fiscal centers. Part of the audits program was chosen according to the IRS' discretionary method, and part was chosen according to an algorithm, following explicit rules. The tax authority was then asked to carry out the audits on the selected firms. At the end of the year, only part of the initially planned audits had been carried out. The purpose of the analysis is to establish whether the use of the algorithm improved the ability of the tax authority to select firms for audit, especially in terms of verified tax evasion.

The audits program of 2019 consisted of 1298 firms in seven different tax centers: the two centers for middle-sized enterprises (called CME 1 and CME 2 in the French acronym), the center for liberal professionals (CPR) and four location-specific centers for small and medium enterprises, all of them in the region of Dakar, Senegal's capital (the four centers were Dakar Plateau, Grand Dakar, Ngor Almadies and Pikine Guediawaye). Part of the 1298 firms were not initially in the list of selected firms, prepared in the beginning of 2019, but were added at the IRS' discretion during the course of the year. We added them as firms selected by the IRS in our analysis.

Table ?? summarizes the execution of the 2019 progam. Out of the 1298 selected firms, 1068 were chosen to be subject to "short audits" (also called CP in the Senegalese IRS' jargon), and the remaining 230 were supposed to be subject to "full audits" (VG in the IRS' jargon). The execution rate was around 50%, meaning that for half the firms in the list there is no indication that the inspectors audited them. For the remaining half, only 37% of them ended in a request for adjustment and eventual payment of a fine.

Table A2: Summary execution full audits

		Total		Algorithm		Inspectors		
		Selected	Started	Selected	Started	Selected	Started	Ad hoc
All years	All	767	453	375	172	423	281	274
	LTU	295	170	129	48	190	122	112
	MTU	304	217	152	99	155	118	131
	Liberal	95	40	53	11	45	29	20
	SME	73	26	41	14	33	12	11
2018	All	316	201	164	80	167	121	109
	LTU	173	95	92	33	94	62	51
	MTU	113	89	57	42	58	47	51
	Liberal	30	17	15	5	15	12	7
	SME	0	0	0	0	0	0	0
2019	All	287	167	124	58	177	109	109
	LTU	122	75	37	15	96	60	61
	MTU	91	64	45	29	47	35	41
	Liberal	30	15	16	6	15	9	5
	SME	44	13	26	8	19	5	2
2020	All	164	85	87	34	79	51	51
	LTU	0	0	0	0	0	0	0
	MTU	100	64	50	28	50	36	38
	Liberal	35	8	22	0	15	8	8
	SME	29	13	15	6	14	7	5

Obs: This table contain the number of firms selected for audit and the number of audits that were started by the tax authority. Discretionary audits are the audits chosen by the tax authority. Algorithm audits are the ones chosen by the risk-based algorithm. Random audits are selected at random within the tax centers. Ad hoc audits are audits that were not in the initial program but were carried out.

Table A3: Summary execution short audits

		Total		Algorithm		Inspectors		Random		
		Selected	Started	Selected	Started	Selected	Started	Started	Selected	Ad hoc
All years	All	2401	752	1116	265	1094	325	336	111	4124
	LTU	522	117	248	31	299	57	60	23	65
	MTU	756	374	340	130	310	167	135	59	925
	Liberal	544	96	249	33	231	40	85	17	395
	SME	579	165	279	71	254	61	56	12	2739
2018	All	785	303	318	110	298	122	202	71	1434
	LTU	207	81	84	28	75	30	60	23	31
	MTU	341	173	138	65	131	72	83	36	359
	Liberal	237	49	96	17	92	20	59	12	135
	SME	0	0	0	0	0	0	0	0	909
2019	All	808	287	364	107	338	124	134	40	1012
	LTU	0	0	0	0	0	0	0	0	0
	MTU	332	169	156	54	141	80	52	23	210
	Liberal	163	28	71	11	70	11	26	5	108
	SME	313	90	137	42	127	33	56	12	694
2020	All	808	162	434	48	458	79	0	0	1554
	LTU	315	36	164	3	224	27	0	0	34
	MTU	83	32	46	11	38	15	0	0	354
	Liberal	144	19	82	5	69	9	0	0	152
	SME	266	75	142	29	127	28	0	0	1014

Obs: This table contain the number of firms selected for audit and the number of audits that were started by the tax authority. Discretionary audits are the audits chosen by the tax authority. Algorithm audits are the ones chosen by the risk-based algorithm. Random audits are selected at random within the tax centers. Ad hoc audits are audits that were not in the initial program but were carried out.

Table A4: Count of selected firms audits by year, tax office, and audit type

		Short audits	Long audits
DGE	2018	153	193
	2019	0	117
	2020	317	98
CME1	2018	142	55
	2019	118	41
	2020	8	42
CME2	2018	192	64
	2019	214	51
	2020	78	53
CPR	2018	239	29
	2019	164	29
	2020	146	33
DP	2018	0	0
	2019	178	26
	2020	148	13
NGA	2018	0	0
	2019	144	12
	2020	124	11
PKG	2018	0	0
	2019	0	0
	2020	0	0

Note: Number of firms selected for audit within the experiment. The numbers include all types of selection methods, but only for the types of firm for which the algorithm was one of these methods. For example, there was no algorithms selection for short audits in DGE in 2019 (the few cases in the table are reclassification of firms into DGE from other centers), and consequently we did not receive the list of selected firms for that group.

Table A5: Tax audit selection methods in selected countries

Country	Discretionary selection	Risk analysis	Random selection
Kenya	Yes ; For all except large taxpayers	Yes ; Only for large taxpayers	No
Senegal	Yes	Yes, Introduced in FY 2018	Introduced in FY 2018
Zimbabwe	Yes; Inspectors rated on selection.	Yes; based on turnover variances	No
Lesotho	No	No	Yes ; Randomly by managers
Tanzania	Abandonned in 2007	Yes	
United Kingdom	Yes; For 55% of audit cases	Yes; Risk scoring	Yes ; Simple random sample
Switzerland	Yes for all cases	No	Yes, periodically for some taxes
United States	No	Yes	
France	Yes; For intelligence gathering	Yes; statistical techniques, data-mining	No
Bulgaria	Yes ; According to set criteria	Yes; Central risk analysis	No
Turkey	No	Yes; Analysis by tax type	Yes ; to collect unbiased data

Sources; Khwaja, Awasthi, and Loeprick (2011) and Authors' survey of select country tax officials.

B.1 Firms' characteristics

Table A6: Number of firms by data source

		2014	2015	2016	2017	2018	2019
Self reported	VAT	8143	8654	9224	9545	9937	10085
	CIT	3987	4548	4813	5026	4963	5647
	CGU	1209	1363	1454	1441	1609	1551
	WIT	5621	5941	6243	6679	6869	7074
	TAF	63	80	86	93	89	83
Third party	Imports	4997	6716	6951	6387	7231	7326
	Exports	1071	1268	1359	1321	1223	1204
	Treasury	528	471	912	870	1210	1221
	VAT annexes	4	7	19	640	2773	2451

Note: Number of firms for which data was available, according to each data source. There are three main sources of data: self-reported tax declarations (Value Added Tax, Corporate Income Tax, simplified regime CGU, Withtheld Income Tax, financial services tax TAF), third party data (exports, imports, treasury payments and VAT annexes concerning inter-firm transactions) and the data produced by the tax inspectors regarding the audit program of 2019. The data include audit programs 2017-2020 in the following tax centers in Senegal: large taxpayer unit, medium taxpayers 1, medium taxpayers 2.

Table A7: Difference between firms in the audit process

	log(Turnover)			Profit rate			log(Payroll)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Selection	Started	Confirmed	Selection	Started	Confirmed	Selection	Started	Confirmed
Desk audits	1.371***			0.0465*			0.547***		
	(0.313)			(0.0237)			(0.0835)		
Algorithm x Desk audits	-0.458*	-0.559*	-0.379	-0.0382**	-0.0465***	-0.0314**	-0.282*	-0.266	-0.200
	(0.232)	(0.289)	(0.245)	(0.0130)	(0.0125)	(0.0101)	(0.135)	(0.149)	(0.262)
Full audits	1.614***	0.280*	0.225	0.0255	-0.0151*	-0.00861*	1.161***	0.681***	0.630***
	(0.329)	(0.125)	(0.136)	(0.0232)	(0.00782)	(0.00425)	(0.115)	(0.0785)	(0.150)
Algorithm x Full audits	-0.726**	-0.674***	-0.549**	-0.0282	-0.0283	-0.0249*	-0.765***	-0.724***	-0.727**
	(0.236)	(0.176)	(0.214)	(0.0207)	(0.0179)	(0.0129)	(0.113)	(0.165)	(0.225)
Random	-0.632**	-0.520*	-0.414*	-0.0218	-0.0289	-0.0139	-0.272**	-0.0914	-0.246
	(0.267)	(0.282)	(0.211)	(0.0156)	(0.0230)	(0.00970)	(0.117)	(0.104)	(0.243)
Ad hoc	-0.583*	-0.572*	-0.479*	-0.0158	-0.0195*	-0.00671	-0.0350	-0.0171	0.0969
	(0.302)	(0.277)	(0.244)	(0.0119)	(0.00898)	(0.00696)	(0.206)	(0.204)	(0.116)
Tax Center	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Activity group fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control overlaps and replacement	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	14071	3643	1726	6457	2716	1362	7923	2771	1411
R2	0.630	0.626	0.627	0.0704	0.0891	0.0649	0.393	0.392	0.370
Mean outcome	17.22	18.83	19.41	-0.06	-0.03	-0.01	15.53	16.25	16.73

B.2 Outcomes

Table A8: Mean characteristics firms - All firms

		Full Audits		Desk audits	
	Population	Inspectors	Algorithm	Inspectors	Algorithm
Turnover	1139.807	6047.570	4074.537	2131.059	3290.584
	(7894.411)	(17490.326)	(15156.461)	(8558.210)	(14435.059)
Profit	41.430	171.523	80.227	46.372	76.021
	(460.112)	(908.470)	(663.778)	(435.173)	(697.474)
Profit rate	-0.055	-0.014	-0.054	-0.026	-0.052
	(0.207)	(0.131)	(0.226)	(0.155)	(0.197)
Payroll	40.780	216.647	172.073	92.714	103.678
	(269.139)	(577.116)	(591.617)	(368.964)	(458.942)
Tax Liability	256.820	1312.427	1002.583	446.806	844.949
	(2340.720)	(5079.792)	(4854.172)	(2342.851)	(4676.374)
Risk score	-0.089	-0.545	0.597	-0.820	0.133
	(0.534)	(1.014)	(0.408)	(1.041)	(0.447)
N	12088	423	374	1093	1113

Note: Mean characteristics of firms in selection and in the population. Total tax liability includes only self declared tax liability in VAT, CIT, PAYE and CGU for firms. The data includes the following tax centers in Senegal: medium taxpayers 1, medium taxpayers 2, liberal professionals, Dakar Plateau, Grand Dakar, Pikine Guediawaye, Ngor Almadies. Values of turnover, tax liability and profits are expressed in Millions FCFA. Profit rate is in percentage of turnover, computed as the mean profit divided by the mean turnover. Number of employees refers to the number of employees in the PAYE declarations.

B.3 Impact of selection on outcomes

Table A9: Outcome: Probability Of Audit Being Started

	(1) Full audits	(2) Full audits	(3) Full audits	(4) Desk audits	(5) Desk audits	(6) Desk audits	(7) Desk audits
Algorithm selection	-0.171**	-0.171**	-0.162**	0.0476	-0.0319	-0.0244	-0.0212
	(0.0567)	(0.0542)	(0.0559)	(0.0506)	(0.0275)	(0.0266)	(0.0266)
Overlap	0.392***	0.294***	0.291***	0.0568	0.0600	0.0487	0.0512
	(0.0765)	(0.0555)	(0.0612)	(0.0464)	(0.0390)	(0.0337)	(0.0355)
Random				0.0120	-0.0193	-0.0110	-0.0136
				(0.0477)	(0.0361)	(0.0364)	(0.0375)
Tax Centre FE, Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Inspector FE	No	No	No	No	Yes	Yes	Yes
Turnover deciles	No	Yes	Yes	No	No	Yes	Yes
Activity group FE	No	No	Yes	No	No	No	Yes
N	767	767	767	2726	2725	2725	2725
R2	0.125	0.240	0.261	0.225	0.491	0.502	0.506
Mean outcome	0.63	0.63	0.63	0.37	0.37	0.37	0.37

Note: * 0.10 ** 0.05 *** 0.01 levels of significance. OLS regression of the audit outcome on the audit selection method. Different specifications controlling for the type of audit, the deciles of mean turnover (with the information available over years 2015-2018), dummies for sector of economic activity, dummies for the tax centers used (LTU, Medium enterprises 1, Medium enterprises 2, Liberal Professions, Dakar Plateau, Ngor Almadies), and dummies for the year of selection (2018, 2019, 2020). Standard errors are clustered at the tax office level and shown in parentheses.

Table A10: Outcome: Audit Ending In Positive Adjustment (Confirmation)

	(1) Full audits	(2) Full audits	(3) Full audits	(4) Desk audits	(5) Desk audits	(6) Desk audits	(7) Desk audits
Algorithm selection	0.0300	0.0556	0.0235	-0.0579	-0.0713	-0.0600	-0.0490
	(0.0489)	(0.0634)	(0.0761)	(0.0448)	(0.0474)	(0.0505)	(0.0427)
Overlap	0.0915	0.0580	0.0771	-0.0296	-0.0469	-0.0520	-0.0605
	(0.105)	(0.117)	(0.129)	(0.0491)	(0.0651)	(0.0542)	(0.0508)
Random				-0.158**	-0.151**	-0.141**	-0.141**
				(0.0491)	(0.0473)	(0.0493)	(0.0449)
Extras				-0.0245	0.00841	0.0183	0.0409
				(0.0890)	(0.0750)	(0.0744)	(0.0770)
Tax Centre FE, Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Inspector FE	No	No	No	No	Yes	Yes	Yes
Turnover deciles	No	Yes	Yes	No	No	Yes	Yes
Activity group FE	No	No	Yes	No	No	No	Yes
N	484	483	481	1028	1027	1027	1027
R2	0.0376	0.0601	0.0793	0.147	0.299	0.309	0.324
Mean outcome	0.61	0.61	0.61	0.41	0.41	0.41	0.41

Note: * 0.10 ** 0.05 *** 0.01 levels of significance. OLS regression of the audit outcome on the audit selection method. Different specifications controlling for the type of audit, the deciles of mean turnover (with the information available over years 2015-2018), dummies for sector of economic activity, dummies for the tax centers used (LTU, Medium enterprises 1, Medium enterprises 2, Liberal Professions, Dakar Plateau, Ngor Almadies), and dummies for the year of selection (2018, 2019, 2020). Standard errors are clustered at the tax office level and shown in parentheses.

Table A11: Outcome: Log (Final Evaded Tax)

	(1) Full audits	(2) Full audits	(3) Full audits	(4) Desk audits	(5) Desk audits	(6) Desk audits	(7) Desk audits
Algorithm selection	0.519	0.746	0.289	-1.252	-0.995	-1.004	-0.696
	(1.192)	(1.498)	(1.715)	(0.801)	(1.213)	(1.359)	(1.183)
Overlap	1.492	1.079	1.507	-1.064	-1.140	-0.959	-1.379
	(2.338)	(2.476)	(2.581)	(1.188)	(1.499)	(1.390)	(1.372)
Random				-2.733**	-2.701**	-2.602**	-2.779***
				(0.947)	(1.018)	(1.084)	(0.801)
Extras				-0.612	-0.143	-0.103	0.145
				(1.636)	(1.306)	(1.378)	(1.483)
Tax Centre FE, Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Inspector FE	No	No	No	No	Yes	Yes	Yes
Turnover deciles	No	Yes	Yes	No	No	Yes	Yes
Activity group FE	No	No	Yes	No	No	No	Yes
N	437	436	434	739	738	738	738
R2	0.0759	0.0964	0.110	0.0998	0.316	0.325	0.346
Mean outcome	11.82	11.85	11.87	9.30	9.30	9.30	9.30

Note: * 0.10 ** 0.05 *** 0.01 levels of significance. OLS regression of the audit outcome on the audit selection method. Different specifications controlling for the type of audit, the deciles of mean turnover (with the information available over years 2015-2018), dummies for sector of economic activity, dummies for the tax centers used (LTU, Medium enterprises 1, Medium enterprises 2, Liberal Professions, Dakar Plateau, Ngor Almadies), and dummies for the year of selection (2018, 2019, 2020). Standard errors are clustered at the tax office level and shown in parentheses.

B.4 Information treatment

Table A12: Outcome: Probability Of Audit Being Started

	(1) Short audits	(2) Short audits	(3) Short audits	(4) Short audits
Algorithm selection		0.113		0.113
		(0.0751)		(0.0757)
Algorithm X Info: indicators		-0.148*		
		(0.0676)		
Algorithm X Info: indicators and data		-0.111		
		(0.0758)		
Algorithm X Any information				-0.130*
				(0.0651)
Random		0.0299		0.0301
		(0.0796)		(0.0815)
Random x Info: indicators		-0.0414		
		(0.0676)		
Random x Info: indicators and data		-0.0265		
		(0.107)		
Random x Any information				-0.0344
				(0.0836)
Any information			0.0444*	0.0888*
			(0.0208)	(0.0403)
Information: indicators	0.0508	0.108*		
	(0.0278)	(0.0469)		
Information: indicators and data	0.0337*	0.0692		
	(0.0172)	(0.0448)		
Sample	Only selected	Only selected	Only selected	Only selected
Tax Center FE, Year FE	Yes	Yes	Yes	Yes
Turnover deciles	Yes	Yes	Yes	Yes
Activity group fixed effects	Yes	Yes	Yes	Yes
N	2726	2726	2726	2726
R2	0.219	0.253	0.243	0.253
Mean outcome	0.37	0.37	0.37	0.37

Note: * 0.10 ** 0.05 *** 0.01 levels of significance. OLS regression of the audit outcome on the audit selection method and the information treatment. Different specifications controlling for the type of audit, the firm's mean turnover (with the information available over years 2015-2018), and dummies for the tax centers used (LTU, Medium enterprises 1, Medium enterprises 2, Liberal Professions, Dakar Plateau, Ngor Almadies) interacted with the year of selection (2018, 2019, 2020). Standard errors are clustered at the tax office level and shown in parentheses.

Table A13: Outcome: Audit Ending In Positive Adjustment (Confirmation)

	(1) Short audits	(2) Short audits	(3) Short audits	(4) Short audits
Algorithm selection		-0.0377		-0.0363
		(0.0552)		(0.0548)
Algorithm X Info: indicators		0.00370		
		(0.0564)		
Algorithm X Info: indicators and data		0.0197		
		(0.0639)		
Algorithm X Any information				0.0110
				(0.0347)
Random		-0.219***		-0.217***
		(0.0561)		(0.0566)
Random x Info: indicators		0.0483		
		(0.0515)		
Random x Info: indicators and data		0.209*		
		(0.108)		
Random x Any information				0.122
				(0.0663)
Any information			-0.0411	-0.0648
			(0.0309)	(0.0481)
Information: indicators	-0.0217	-0.0377		
	(0.0308)	(0.0282)		
Information: indicators and data	-0.0361	-0.100		
	(0.0317)	(0.0754)		
Sample	Only selected	Only selected	Only selected	Only selected
Tax Center FE, Year FE	Yes	Yes	Yes	Yes
Turnover deciles	Yes	Yes	Yes	Yes
Activity group fixed effects	Yes	Yes	Yes	Yes
N	1028	1028	1028	1028
R2	0.136	0.181	0.168	0.178
Mean outcome	0.41	0.41	0.41	0.41

Note: * 0.10 ** 0.05 *** 0.01 levels of significance. OLS regression of the audit outcome on the audit selection method and the information treatment. Different specifications controlling for the type of audit, the firm's mean turnover (with the information available over years 2015-2018), and dummies for the tax centers used (LTU, Medium enterprises 1, Medium enterprises 2, Liberal Professions, Dakar Plateau, Ngor Almadies) interacted with the year of selection (2018, 2019, 2020). Standard errors are clustered at the tax office level and shown in parentheses.

Table A14: Outcome: Log (Final Evaded Tax)

	(1) Short audits	(2) Short audits	(3) Short audits	(4) Short audits
Algorithm selection		-1.541		-1.448
		(1.383)		(1.414)
Algorithm X Info: indicators		0.862		
		(1.789)		
Algorithm X Info: indicators and data		1.334		
		(1.668)		
Algorithm X Any information				1.068
				(1.452)
Random		-3.895*		-3.800*
		(1.973)		(1.995)
Random x Info: indicators		-0.424		
		(2.855)		
Random x Info: indicators and data		4.650*		
		(2.063)		
Random x Any information				2.025
				(2.201)
Any information			-0.700	-1.408
			(0.684)	(1.041)
Information: indicators	-0.302	-0.711		
	(0.747)	(0.918)		
Information: indicators and data	-0.639	-2.297		
	(0.851)	(1.401)		
Sample	Only selected	Only selected	Only selected	Only selected
Tax Center FE, Year FE	Yes	Yes	Yes	Yes
Turnover deciles	Yes	Yes	Yes	Yes
Activity group fixed effects	Yes	Yes	Yes	Yes
N	739	739	739	739
R2	0.0884	0.141	0.125	0.135
Mean outcome	9.30	9.30	9.30	9.30

Note: * 0.10 ** 0.05 *** 0.01 levels of significance. OLS regression of the audit outcome on the audit selection method and the information treatment. Different specifications controlling for the type of audit, the firm's mean turnover (with the information available over years 2015-2018), and dummies for the tax centers used (LTU, Medium enterprises 1, Medium enterprises 2, Liberal Professions, Dakar Plateau, Ngor Almadies) interacted with the year of selection (2018, 2019, 2020). Standard errors are clustered at the tax office level and shown in parentheses.

B.5 Mechanisms

B.5.1 Audit quality

Table A15: Questions: taxpayers' evaluation of full audits

	(1) Duration	(2) Tech.	(3) Honesty	(4) Effic.
Algorithm selection	0.201	-0.281	-0.132	0.0955
	(0.273)	(0.229)	(0.188)	(0.264)
Desk audit	0.535	0.0819	0.639***	0.238
	(0.412)	(0.208)	(0.192)	(0.270)
Full audit	1.647***	0.291	0.743**	0.247
	(0.302)	(0.644)	(0.259)	(0.182)
Desk audited X Algorithm	0.236	-0.187	-0.392	-0.398
	(0.475)	(0.282)	(0.383)	(0.505)
Full audit X Algorithm	-0.351	-0.190	-0.244	-0.142
	(0.852)	(0.700)	(0.486)	(0.529)
Tax office FE	Yes	Yes	Yes	Yes
N	495	510	504	520
R2	0.0805	0.0314	0.0347	0.0237
Mean outcome	3.03	7.36	6.42	6.33

Note: * 0.10 ** 0.05 *** 0.01 levels of significance. OLS regression of answer to the question in taxpayer survey. In this set of questions, the respondents were asked to evaluate the full audit that they were subjected to. The first question asks them the number of weeks that the full audit lasted. The other questions ask them to grade the inspectors from 0 (bad) to 10 (excellent) the inspectors regarding their technical knowledge (Tech.), honesty and efficiency (Effic.). The regressions control for the 4 tax centers used (Large Taxpayers Unit, Medium enterprises 1 and 2, Liberal Professions and SMEs). Standard errors clustered at the tax office level, and are shown between parentheses.

B.5.2 Manpower

Table A16: Outcome: Number Of Agents

	(1) Full audits	(2) Full audits	(3) Full audits	(4) Desk audits	(5) Desk audits	(6) Desk audits	(7) Desk audits
Algorithm selection	-0.305***	-0.242**	-0.226**	0.124*	-0.0359	-0.0280	-0.0207
	(0.0595)	(0.0785)	(0.0871)	(0.0635)	(0.0241)	(0.0189)	(0.0177)
Overlap	0.268**	0.186	0.150	-0.205	-0.0102	-0.0258	-0.0264
	(0.0997)	(0.104)	(0.127)	(0.153)	(0.0467)	(0.0278)	(0.0257)
Tax Centre FE, Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Inspector FE	No	No	No	No	Yes	Yes	Yes
Turnover deciles	No	Yes	Yes	No	No	Yes	Yes
Activity group FE	No	No	Yes	No	No	No	Yes
N	453	451	451	1028	1027	1027	1027
R2	0.203	0.217	0.249	0.235	0.822	0.829	0.832
Mean outcome	3.00	3.01	3.01	2.01	2.01	2.01	2.01

Note: * 0.10 ** 0.05 *** 0.01 levels of significance. OLS regression of the audit outcome on the audit selection method. Different specifications controlling for the type of audit, the deciles of mean turnover (with the information available over years 2015-2018), dummies for sector of economic activity, dummies for the tax centers used (LTU, Medium enterprises 1, Medium enterprises 2, Liberal Professions, Dakar Plateau, Ngor Almadies), and dummies for the year of selection (2018, 2019, 2020). Standard errors are clustered at the tax office level and shown in parentheses.

B.5.3 Inspector level regressions

Table A17: Short audits - inspector level regressions

	(1) Total assigned cases	(2) Total assigned cases	(3) Total started cases	(4) Total started cases	(5) P(Start assigned case)	(6) P(Start assigned case)	(7) P(Cases yielding adjustment)	(8) P(Cases yielding adjustment)
Education: Bachelors	-2.522	-1.870	-0.617	-0.367	0.225	0.177	0.168	0.0516
	(3.090)	(2.611)	(1.164)	(0.973)	(0.154)	(0.134)	(0.149)	(0.137)
Education: Masters	1.013	0.880	0.693	1.332	0.0146	0.0395	0.0375	-0.00892
	(2.149)	(2.138)	(1.187)	(1.492)	(0.0989)	(0.102)	(0.106)	(0.110)
Education: PhD	6.790**	6.139**	2.258	2.216	-0.0944	-0.0485	0.0678	0.0879
	(3.404)	(3.091)	(1.628)	(1.336)	(0.137)	(0.125)	(0.129)	(0.130)
Age	2.163		-2.344		-0.225		-0.0317	
	(3.084)		(2.611)		(0.150)		(0.150)	
Age sq.	-0.0305		0.0312		0.00312		0.000676	
	(0.0421)		(0.0357)		(0.00215)		(0.00211)	
Years experience	0.596		0.546		-0.0247		-0.0673	
	(1.169)		(0.570)		(0.0437)		(0.0471)	
Years squared	-0.0123		-0.0268		0.000640		0.00398	
	(0.0638)		(0.0337)		(0.00245)		(0.00295)	
Above median age		1.349		-0.677		-0.0747		0.0401
		(1.508)		(1.113)		(0.0782)		(0.0884)
Above median experience		2.320		1.116		-0.0533		0.00803
		(1.427)		(0.967)		(0.0751)		(0.0826)
Tax Center FE, Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	121	121	103	103	121	121	107	107
R2	0.311	0.308	0.171	0.158	0.308	0.272	0.268	0.250
Mean outcome	10.29	10.29	3.68	3.68	0.48	0.48	0.53	0.53

Note: * 0.10 ** 0.05 *** 0.01 levels of significance. OLS regression of probability of audit being started on the selection method for the audits of 2019. The first two columns outcomes are from administrative data, and the third one from survey data.

Table A18: Short audits - inspector level regressions

	(1) Av. log(adjustment)	(2) Av. log(adjustment)	(3) % algorithm cases among assigned	(4) % algorithm cases among assigned	(5) % algorithm cases among started	(6) % algorithm cases among started	(7) In favor of algorithm	(8) In favor of algorithm
Education: Bachelors	2.374	0.705	0.116	0.0244	0.171	-0.0325	0.208	0.276*
	(2.688)	(2.436)	(0.0988)	(0.0878)	(0.154)	(0.136)	(0.154)	(0.145)
Education: Masters	1.227	0.496	-0.0964	-0.0531	-0.185	-0.121	0.190	0.274**
	(1.998)	(2.045)	(0.0765)	(0.0737)	(0.131)	(0.125)	(0.119)	(0.120)
Education: PhD	1.271	1.524	-0.00831	0.0599	-0.129	0.0369	0.223**	0.306***
	(2.394)	(2.371)	(0.107)	(0.0939)	(0.148)	(0.128)	(0.103)	(0.106)
Age	0.483		-0.0639		-0.0936		0.0507	
	(2.675)		(0.109)		(0.138)		(0.123)	
Age sq.	-0.00334		0.000939		0.00161		-0.000863	
	(0.0379)		(0.00150)		(0.00197)		(0.00188)	
Years experience	-1.080		-0.0574		-0.129**		0.0209	
	(0.863)		(0.0382)		(0.0543)		(0.0476)	
Years squared	0.0660		0.00293		0.00590*		-0.00359	
	(0.0539)		(0.00218)		(0.00327)		(0.00335)	
Above median age		0.599		-0.131**		-0.125		-0.145
		(1.682)		(0.0645)		(0.102)		(0.0947)
Above median experience		0.367		0.110*		0.117		-0.104
		(1.628)		(0.0610)		(0.104)		(0.107)
Tax Center FE, Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	107	107	121	121	103	103	120	120
R2	0.283	0.270	0.241	0.240	0.241	0.171	0.287	0.221
Mean outcome	9.58	9.58	0.50	0.50	0.49	0.49	0.80	0.80

Note: * 0.10 ** 0.05 *** 0.01 levels of significance. OLS regression of probability of audit being started on the selection method for the audits of 2019. The first two columns outcomes are from administrative data, and the third one from survey data.

Table A19: Quality of inspectors

	(1) Share with Masters/PhD	(2) Mean age	(3) Mean experience	(4) Share in favor of alg.	(5) Max years experience	(6) Max education
Algorithm selection	0.0462	-0.330	0.0198	-0.0150	-0.369	-0.106*
	(0.0337)	(0.333)	(0.221)	(0.0275)	(0.255)	(0.0539)
Overlap	-0.0165	1.396*	0.293	-0.0228	0.545	0.109
	(0.0566)	(0.755)	(0.402)	(0.0655)	(0.538)	(0.0907)
Tax Centre FE, Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	411	440	411	411	411	411
R2	0.479	0.335	0.234	0.220	0.246	0.262
Mean outcome	0.65	37.04	8.29	0.82	9.53	3.28

Note: * 0.10 ** 0.05 *** 0.01 levels of significance. Only full audits. Standard errors are clustered at the tax office level and shown in parentheses.

C Figures

C.1 Audits in Senegal

Figure A1: DGID's organizational chart

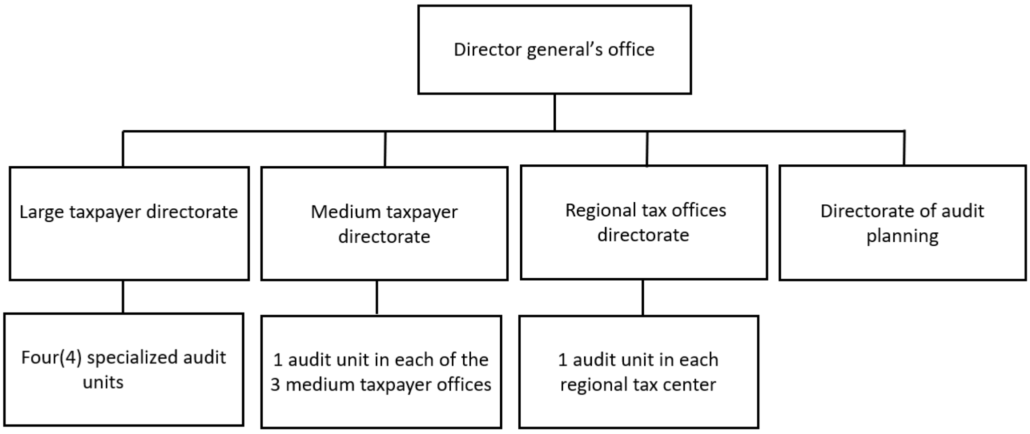
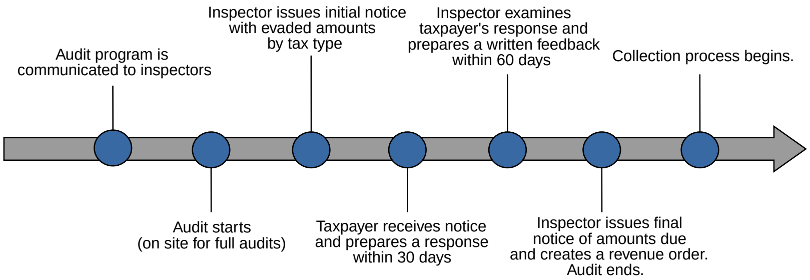
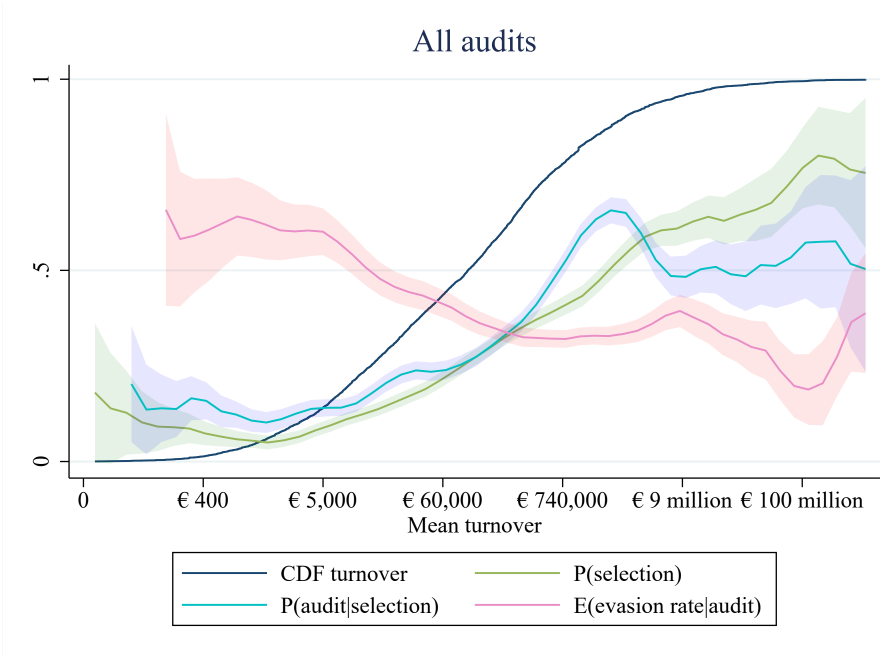


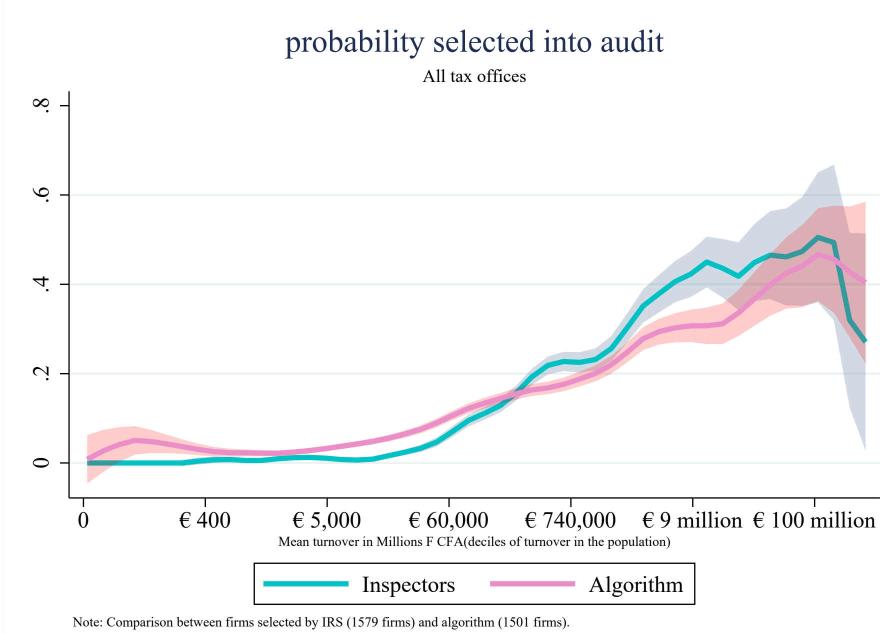
Figure A2: Audit process



C.2 Outcomes and firm size

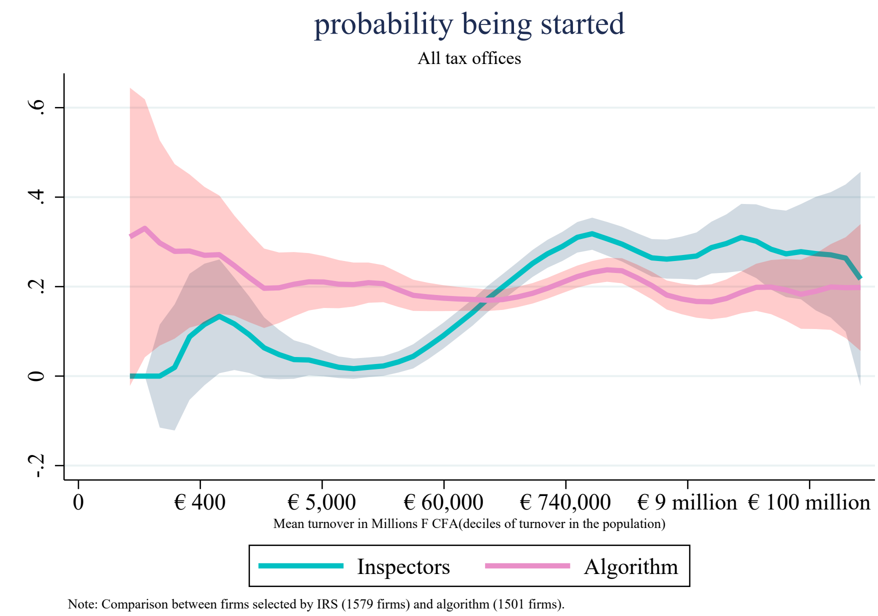
Figure A3: Distribution of turnover by tax center





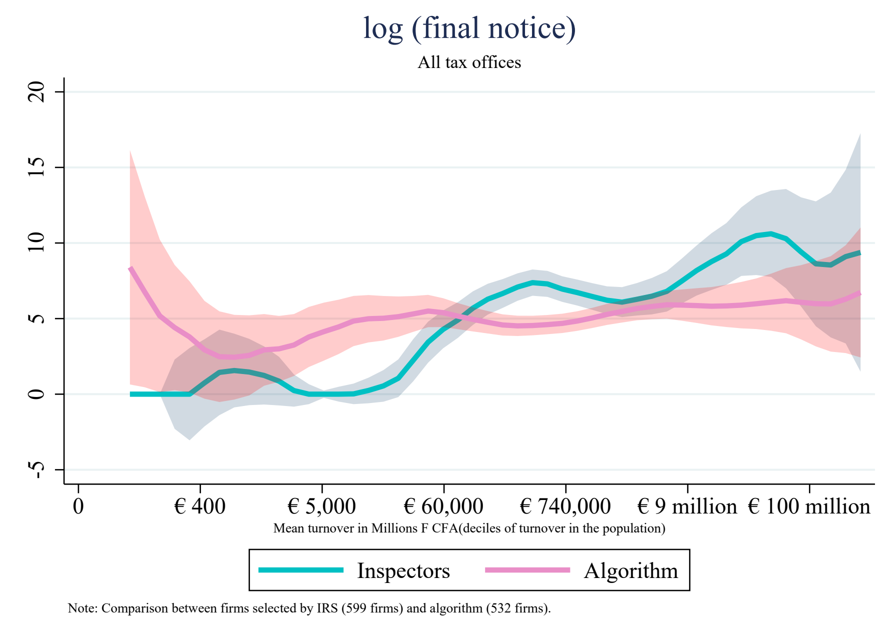
Obs: Non parametric regression of outcome on mean turnover (no controls), using Epanechnikov kernel, bandwidth computed according to the rule-of-tumb method.

Figure A4



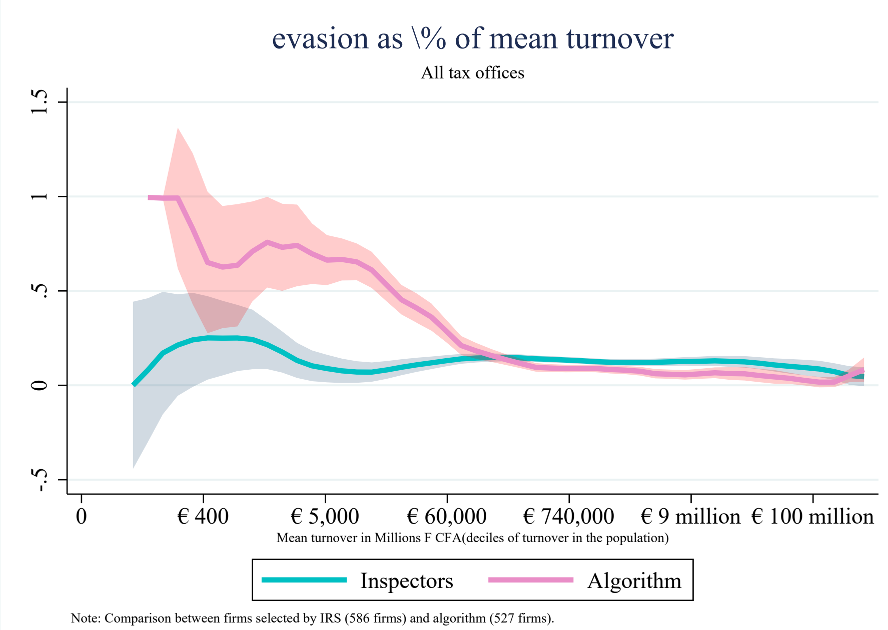
Obs: Non parametric regression of outcome on mean turnover (no controls), using Epanechnikov kernel, bandwidth computed according to the rule-of-tumb method.

Figure A5



Obs: Non parametric regression of outcome on mean turnover (no controls), using Epanechnikov kernel, bandwidth computed according to the rule-of-tumb method.

Figure A6



Obs: Non parametric regression of outcome on mean turnover (no controls), using Epanechnikov kernel, bandwidth computed according to the rule-of-tumb method.

D Appendix figures

E Risk Scoring of Tax Evasion

C.1 Motivation

A key feature of this project is to assist the Senegalese tax administration (DGID) to design a tool which assesses firms' tax evasion risk. Starting in 2017, the team held consultations with DGID leadership and former tax inspectors to map the compliance risks of Senegalese firms and to exploit all available data sources to assess this risk. Moreover, we discussed with experts in the field of taxation and risk management, who worked on tax evasion risk assessment in middle-income countries. With these inputs, we designed a risk-scoring tool, following best international practice, as implemented by the World Bank and its partner institutions.

Although the use of advanced machine-learning tools for prediction has exploded in economic analysis, it was decided together with DGID that the risk-score would be guided by simple variables which logically should predict evasion risk. The simplicity of the design is motivated by several factors, ranked by order of importance. First, the tool needed to be transparent, such that underlying compliance risks could be understood by tax inspectors, and explained to taxpayers when required. Second, the available data on historical audit results was sparse and not digitized, which limited the scope of our model calibration and model selection exercises (further details below). Finally, all cases concluded by 2017 were selected in a discretionary manner.

Thus, one should consider the risk-scoring tool as a transparent best-practice risk assessment, given the administrative capacity, rather than a fined-tool fully optimized algorithm. We note that the constraints faced by DGID are likely to bind in many low income countries, and especially in other West African countries, which often look at Senegal for administrative innovations.

Table XX summarizes the seven key steps in the design of the risk-score. Step (1) corresponds to the construction of a database covering all tax declarations across years and merged with third-party reported sources. Steps (2) and (3) determine specific risk indicators, based on discrepancies across sources or behavioral outliers, examples of which are discussed below. Step (4) defines the peer-group comparison: these clusters regroup firms by economic activity and either size or geographical zones, depending on the structure of each tax center. Step (5) assigns a numerical value to each risk indicator, depending on the size of the deviation (higher scores when larger discrepancies), while step (6) assigns weights to each indicator reflecting beliefs about their relative importance. Finally, step (7) aggregate the weighted indicators in each of the past four fiscal year, and then sums up the yearly scores to form a total risk score.

Table A20: Steps of risk-score design

Step	**Description**
(1) Prepare merged dataset	The tax declarations of each taxpayer are merged across type of taxes (VAT, CIT, Payroll) and across years. Data from third parties is then added (customs, procurement, transaction network).
(2) Choose indicators: discrepancies	Discrepancies are situations in which a self-reported tax liability can be considered as misreported or incomplete, by cross checking several data sources together.
(3) Choose indicators: anomalies	Anomalies correspond to abnormal reporting behavior, compared to peers. Anomalies suggest that firms should be monitored, but do not indicate tax evasion behavior with certainty.
(4) Define comparison clusters	Clusters regroup firms in the same economic sector and of comparable size. Peer comparisons are done within clusters
(5) Assign values to indicators	The magnitude of the inconsistency is used to assign a value, ranging from one to ten (using deciles). For anomalies firms within the top decile of a particular indicator receive a value of one.
(6) Assign weights to indicators	Weights are assigned to each indicator reflecting beliefs about their relative importance.
(7) Aggregate indicators and years	The weighted risk indicators are first aggregated across indicators in each year. Then the yearly scores are summed up to form a total risk score covering the past four years of tax declarations. More recent years are slightly over-weighted.

C.2 Choosing indicators and weights

As explained above, the algorithm computes some ratios from the data of firms (declarations and third party data) and then calculates the value of the indicator based on the distribution of this ratio within a cluster of comparable firms. We tried several combinations of indicators before stabilizing the algorithm in a reduced set of them. The goal was to have a set of indicators that was sensible and correlated with evasion, but at the same time simple and understandable for the tax inspectors.

Table A20 summarizes the steps that we took to conceptualize the algorithm. We tried out several possible indicators that could suggest under-declaration of tax liability. We discarded most based on some analysis of data availability or statistical relevance. In the end, we discarded indicators that required information that was available for a reduced set of firms and indicators that did not seem to have any correlation with evasion, as per past evasion data. We tested these indicators on data from historical audits data. We performed out of sample regressions with LASSO and OLS and computed the out of sample mean squared prediction errors to compare different models. This allowed us to assert that the ranking normalization performed well with respect to alternatives (meaning that it presented a lower prediction error).

We refer to the appendix for an analysis of these indicators using historical audits data. From this analysis we decided to restrict the algorithm to a small list of indicators. Three of them are inconsistencies, plus a flag for inconsistent filing of taxes. On top of that, we have seven anomalies, of which two refer to value added tax, two refer to corporate income tax, one refers to third party data comparisons, one to share of imports from low tax countries and one refers to the financial services tax (only applicable to a reduces set of firms). The final list of indicators that is used in the algorithm, and the respective weights (ω and ξ in equation **??**) is summarized in the following table.

Some details for the calculation of the indicators are worth mentioning. In some cases of anomalies, the top decile within a cluster comprises more than 10% of cases. As long as the value is not zero, we include all these firms. Whenever there is not enough non-zero values that can fill un 10% of the firms, we only flag the non-zero values. We also top code (999 999 999) all values for which the denominator of te underlying ratio of the indicator is zero or missing. Therefore they belong by definition to the top decile. We also top code all values of negative tax liability, to make sure they also get flagged. The idea of the indicators is always that the larger the ratio, the less taxes the firm is paying.

We designed the risk-scoring scheme using best practices, drawing on policy documents from the World Bank (tax administration projects in Pakistan and Turkey), SKAT in Denmark, and the IMF's recommendations to DGID. We provide a high-level description of this process to preserve confidentiality around audit selection processes. We compute risk scores using information sets/tax returns submitted to DGID on corporate income taxes, VAT, personal income tax withholding remittance, as well external data from customs (imports/exports) and public procurement contracts, for the period 2013-2016 [21]. The score relies on two types of risk indicators: discrepancies and anomalies. Discrepancy indicators flag taxpayers whose self-reported information according to their tax returns differs from information in datasets obtained from customs or the government budget department in charge of paying state procurement. For instance, a discrepancy indicator is logged when taxpayers' reported turnover over multiple years is lower than its aggregate costs, that its imports plus its wage bill over the same period. Anomaly indicators use industry/sector benchmarking to flag firms with unusual behavior relative to their peers. An example would be a firm in petroleum retail with low profit rate compared to its peers, which might be associated with evasion. Discrepancies and anomalies are aggregated to produce a risk-score for each taxpayer.

[21]We also attempted to apply predictive analytics from the machine learning literature on these datasets and on previous audit results was conducted to check whether risk indicators could predict DGID audit returns. This exercise was inconclusive because of the selected nature of the sample for whom audit returns are available, the small number of observations and noise in the data.

Optimal (double) taxation with tax evasion and firm growth

Alipio Ferreira

May 10, 2022

Abstract

Tax evasion is in general a nuisance for governments, which must devote resources to fight it to ensure that taxpayers pay their taxes. However, if taxpayers invest avoided taxes in a productive way, governments can also benefit from evasion by taxing the outcome of taxpayers' investments. Moreover, by auditing past tax declarations, governments can still recover avoided taxes from the past while still benefiting from the result of past evasion. This amounts to a form of double taxation. This paper models tax evasion by firms in a dynamic setting where firms have incentives to invest all their assets. It shows that the optimal policy for the government is not to reduce evasion to zero, even when all enforcement parameters are free. In practice, evasion functions as a loan from the government to the taxpayer, where expected fines work as interest rates. The incentives outlined in this paper are likely to hold for small, financially constrained firms with high growth potential.

*alipio.ferreira@tse-fr.eu, Toulouse School of Economics. I would like to thank Helmuth Cremer for the supervision, and the Public Economics group at TSE for insightful comments, in particular Jean-Marie Lozachmeur.

1 Introduction: Firm growth, taxes and financial constraints

In its ability to tax economic activity, the government is a similar to a shareholder of the whole economy: it can collect part of the revenues produced by individuals or firms. Consequently, tax revenues benefit from economic growth, and excessive taxation may be counterproductive for raising revenue. Increasing taxes affects the behavior of agents, encouraging them to produce less or to evade taxes, and at some point a marginal increase in tax rates may reduce tax revenues. This phenomenon is commonly referred to as "Laffer curve", in honor of the American economist Arthur Laffer. There are several different theoretical foundations for the Laffer curve. In this paper, I provide another one: the idea that taxation and enforcement may affect firm growth. I argue that financially constrained firms with growth potential may use tax evasion to alleviate their financial constraint and expand investments. The evade amount is not entirely lost to the government, which can recover it through tax inspections (enforcement action), typically happening after the evaded amount has already been spent. In fact, besides having a "shareholder" claim on economic activity, the government also operates implicitly as an implicit "lender" when agents evade taxes.

Firms may grow faster by evading taxes. The evaded amounts are additional profits, which can be reinvested in the firm's activities and make it grow. If caught by the tax authority, however, the firm must pay a fine on the evaded amount. The returns on the evaded amount invested productively must be weighed against the potential cost of the penalty. But for firms with high growth potential and limited access to financial markets, cheating on taxes may be a way to ease current budget constraints and invest in productive activity. Indeed, as James Andreoni (1992) put it once, evasion in a multi-period setting may function in a similar way to a loan: the firm can raise current revenue by cutting on tax expenditure, but has an expected future payment of a fine. Even if this expected future payment – the "interest" on the loan – is high, firms may find it interesting to take it.

The government may also take advantage of firms evading taxes to grow. First, because firm growth raises the size of the tax base in later periods. A larger firm pays more taxes. Second, because operating as a "lender" also gives it the opportunity to collect "interests" on the amount evaded, by running a tax audit. A tax audit in this setting gives rise to a kind of double taxation. When a firm gets audited, it must pay to the government a fine relative to the evaded amounts in previous period, but it also pays taxes based on its current size, which would be smaller if the firm had not evaded previously. The government thus benefits from evasion, but also forces compliance.

I illustrate this mechanism in a two-period dynamic model where a firm has high growth potential but has limited assets. The firm can evade taxes in both periods, but can be audited only in the second period, which happens with positive probability and implies a fine proportional to the amount evaded in the first period. This is a very standard tax evasion model following on the steps of Allingham and Sandmo (1972) and Yitzhaki (1974), adapted for a risk-neutral decision maker as in Cremer and Gahvari (1993), in a dynamic setting as in Andreoni (1992). The link between tax evasion and financial constraint has been raised in the literature by Andreoni (1992) in his model of personal income tax evasion. Gatti and Honorati (2008) and Alm, Liu, and Zhang (2018) have documented a positive correlation between evasion and lack of access to financial markets in developing economies.

There are good reasons to suppose that firms, and in particular small firms, are financially constrained. Even when they have high expected revenues in future periods, and only need liquidity to reach that stage, financial institutions may hesitate to lend due to asymmetric information problems. Lack of observability of the quality of projects, lack of enforcement of promises or limited contracting capacity lead to the fact that firms cannot borrow freely in financial markets[*] This is particularly true for smaller firms, with little collateral and track record. It is also more likely to hold in developing countries, where financial markets are less developed and enforcement of contracts is weaker.

In the model proposed in this paper, firms have a limited amount of assets that they can invest in a technology with decreasing marginal returns. They have a high growth potential, so that it is in their interest to invest everything they can, and only then distribute dividends. As already mentioned, they can boost their investments by evading taxes. When they do so, they contract a debt with the government, which they may pay with interest if they get audited in the future. The cost of this expected payment will determine the extent to which firms wish to engage in this risky activity of evasion.

In the model, firms have a technology with positive and decreasing marginal returns, so that smaller firms face very high marginal returns and tend to evade more. For them, the cost of paying a penalty on a marginal evaded unit is lower than the marginal benefit of expanding capacity. This leads to the fact that compliance improves for larger firms, a fact that is corroborated in the empirical literature about firm tax evasion, such as Pomeranz (2015) for Chile, Brockmeyer and Hernandez (2017) and Bachas and Soto (2021) for Costa Rica, Naritomi (2019) and Ulyssea (2018) for Brazil.

In the model, firms have strong incentives to evade, including when probability of audit is

[*]There are two maind types of asymmetric information problems: adverse selection and moral hazard. Each of them may lead to credit constraints. Albuquerque and Hopenhayn (2004) and Clementi and Hopenhayn (2006) are models of adverse selection and moral hazard where financial institutions propose a dynamic contract to a firm. Although the two contracts have important differences, in both models the firm is credit constrained and can only borrow up to a certain limit.

extremely high (even 100%). The reason, again, is that evasion eases their financial constraint. Even though expected payment on the evaded amount is greater than 100% of the evaded value, it may be worth doing it. The excessive expected payment is the counterpart of interests in a standard loan. The government, on the other hand, wants the firm to evade. Tax evasion provides the government with cheap finance, since it allows the government to tax firms that are larger in the second period, but still allows them to recover the evaded amounts with penalties. As the model shows, even when the government has a lump-sum tax at its disposal, it still may use distortionary taxation to take advantage of this double taxation opportunity.

2 Model of the behavior of the firm

To illustrate how growth incentives affect tax compliance, I propose a simple dynamic model with two periods. The firm maximizes expected dividends over two periods, and chooses compliance levels at each period, x_1 and x_2 over the firm value $\pi(A_1)$ and $\pi(A_2)$, derived from assets A_1 and A_2. The function $\pi(A_1)$ is increasing and concave. The government is free to set different taxes for each period, τ_1 and τ_2, and audits a proportion p of firms only in the second period. During the audit both periods are verified. If the declared amounts $x_1\pi(A_1)$ and $x_2\pi(A_2)$ are inferior to the truth, the taxpayer must pay the evaded taxes plus a proportional fine ϕ. One key feature of this model is that audits occur only in the second period with probability p, and check tax liability in both periods. This is similar to Andreoni (1992), but in his model is no taxation on income in the second period in his model and therefore also no verification of second period income. Defining as y_1 and y_2 the net value of the firm in each period, the time discount rate β and the share α of distributed dividents in the first period, the firm's problem can be formulated as follows:

$$\begin{aligned} \tilde{\Pi} &\equiv \max_{x_1,x_2,\alpha} \Pi \\ \Pi &= \alpha y_1 + \mathbb{E}[y_2] \end{aligned} \tag{1}$$

$$\begin{aligned} y_1 &= \pi(A_1)(1-\tau_1 x_1) \\ \mathbb{E}[y_2] &= (1-p)\Big(\pi(A_2)(1-\tau_2 x_2)\Big) + p\Big(\pi(A_2)(1-\tau_2(x_2+(1-x_2)\phi) - \pi(A_1)\tau_1(1-x_1)\phi\Big) \\ A_2 &= (1-\alpha)\pi(A_1)(1-\tau_1 x_1) \end{aligned}$$

In this formulation, I abstract from any problem related to time discounting. Moreover, I make the following assumptions:

Assumption 1. ***Taxation of excessive marginal returns***: *Only excessive marginal returns are taxed. This means that* $\pi'(A_t)(1-\tau_t) \geq 1$*, which implies that* $\pi(A_t)(1-\tau_t) \geq A_t$*. The latter can be interpreted as "no wealth taxation". This assumption puts upper bounds on the level of* τ_t *that the government can set at each period* $t \in \{1,2\}$.

The maximization of this problem by the firm implies the following facts. First, assumption 1 implies that the ratio α of dividends distributed in period 1 is equal to 0. This means that the firm uses the first period to accumulated assets and grow, and it is not worth to forgo growth in exchange of first period consumption. The second result regards first period compliance: depending on the firm's initial asset size A_1, firms will be *informal* (compliance $x_1 = 0$), *evaders* ($x_1 \in (0,1)$) or *compliers* ($x_1 = 1$). Larger firms comply more. In the second period, evasion will follow a bang-bang rule: if expected penalty is high, firms comply ($x_2 = 1$), else they will evade totally and run the risk of paying the fine ($x_2 = 0$).

This problem is solved by backwards induction. Determining the compliance level in the second period, x_2 is a static problem:

$$\frac{\partial \Pi}{\partial x_2} = \pi(A_2)\tau_2(p\phi - 1) \tag{2}$$

Since the entrepreneur is risk-neutral, the problem is lineas in x_2 and compliance in second period is either 1 or 0 depending on the values of the enforcement parameters. There is full compliance if $p\phi > 1$ and no compliance (i.e. full evasion) if $p\phi < 1$.

The dynamic problem appears as the entrepreneur chooses the compliance level in the first period, because this affects outcomes in the following period. The first order condition for x_1 yields:

$$\frac{d\Pi}{dx_1} = \pi(A_1)\tau_1 \left\{ \underbrace{\phi p}_{\text{lower expected fine}} \quad \underbrace{-\pi'(A_2)(1 - \tau_2(x_2 + (1-x_2)p\phi)}_{\text{lower expected profits}} \right\} \tag{3}$$

A marginal increase in compliance dx_1 in the first period reduces the expected fines paid in the second period (a gain to the firm) but decreases its profits in the second period, since it can grow less. The optimal compliance level is the one that makes the marginal gains (in terms of lower fines) equal to the marginal costs (in terms of lower second period profits). The optimum is achieved when:

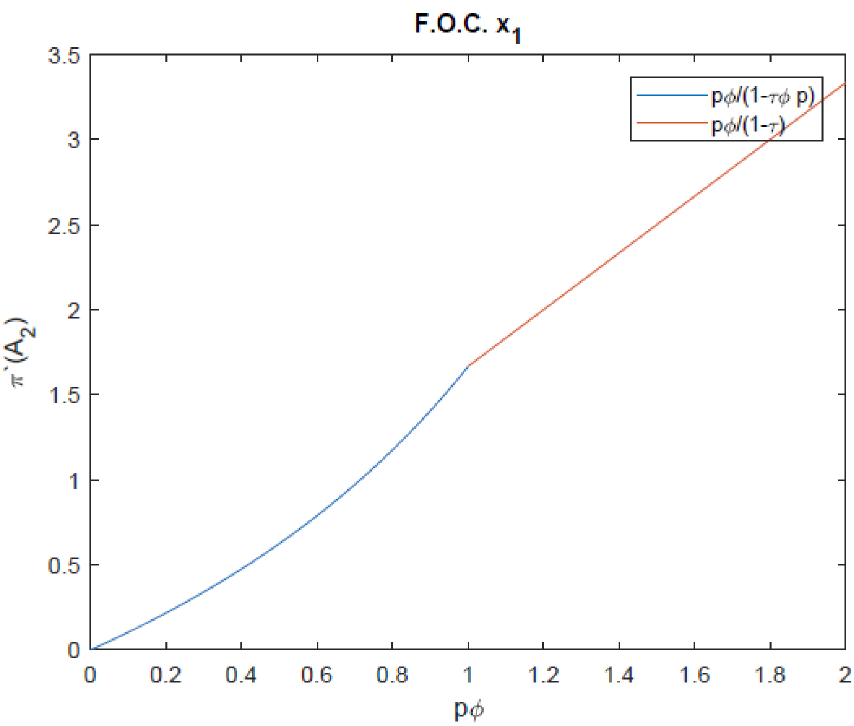
$$\pi'(A_2) = \frac{\phi p}{1 - \tau_2(x_2 + (1-x_2)p\phi)} \tag{4}$$

The right hand side of this equation is a constant that depends solely on the parameters of the tax system. As can be seen from the first order condition with respect to x_2 (compliance in period $t = 2$), if $p\phi > 1$ we have $x_2 = 1$ and if $p\phi < 1$ the optimum is full evasion, $x_2 = 0$. Therefore, the first order condition in equation 4 can be rewritten as:

$$\pi'(A_2) = \begin{cases} \frac{\phi p}{1-\tau_2} & \text{if } p\phi > 1 \\ \frac{\phi p}{1-\tau_2 p\phi} & \text{if } p\phi < 1 \end{cases} \tag{5}$$

This equation maps all possible values of $p\phi$ to the correspondent optimal A_2. The marginal return to capital $\pi'(A_2)$ is monotonically increasing in ϕp, despite the discontinuity that happens at $\phi p = 1$, as illustrated in figure 1. This means that the optimal level of assets A_2 is decreasing in the enforcement parameters.

Figure 1: Mapping of expected fine to marginal benefits at the optimum



There is one unique level of A_2 that is associated with the first order condition, which I call A_2^*. This level determines the behavior of the firm with regard to evasion in the first period, that is, the amount of compliance x_1.

$$A_2^* = \begin{cases} \pi'^{-1}[\frac{\phi p}{1-\tau_2}] & \text{if } p\phi > 1 \\ \pi'^{-1}[\frac{\phi p}{1-\tau_2 p\phi}] & \text{if } p\phi < 1 \end{cases} \tag{6}$$

To achieve A_2^*, the firm can decide the level of compliance x_1. More compliance (higher x_1) means higher costs in the first period and less assets in the second period. The problem is that x_1 is bounded between 0 and 1, and this sets boundaries on the possible range of A_2^* that can be feasible. If a firm is very small for example, and the value for A_2^* is very high relative to the initial size, the firm will not reach it even if it evades fully.

Define $\bar{x_1}$ as the level of x_1 such that the first order condition holds with equality. The compliance rule can be stated as follows:

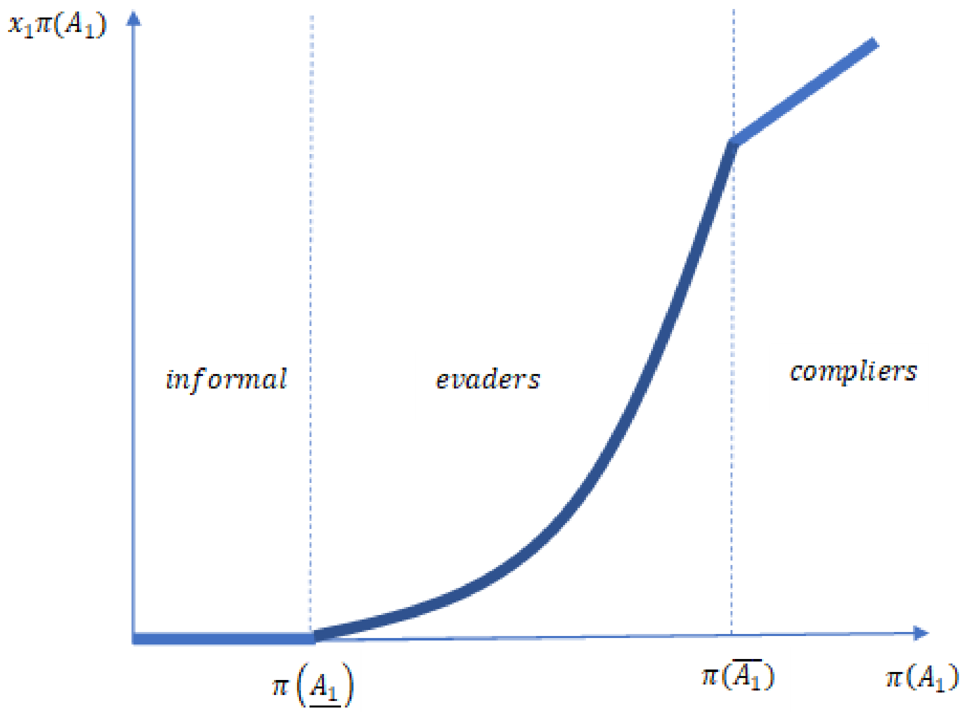
$$x_1 = \begin{cases} 0 & \text{if } \pi(A_1) \leq A_2^* \\ x_1^* & \text{if } \pi(A_1)(1-\tau_1) \leq A_2^* \leq \pi(A_1) \\ 1 & \text{if } \pi(A_1)(1-\tau_1) \geq A_2^* \end{cases} \tag{7}$$

Notice also that since A_2^* is a constant, equation 7 also defines unique thresholds of A_1 that define in which of the three categories the firm belongs: informal, evader or compliant. These thresholds are $\underline{A_1}$ and $\bar{A}_1$, such that:

$$x_1 = \begin{cases} 0 & \text{if } A_1 \leq \underline{A_1} \\ x_1^* & \text{if } \underline{A_1} < A_1 \leq \bar{A}_1 \\ 1 & \text{if } \bar{A}_1 < A_1 \end{cases} \tag{8}$$

In the end the government observes $x_1\pi(A_1)$ declared by the firm, which follows a schedule with respect to the possible values of A_1 as in figure 2:

Figure 2: Optimal evasion rates and initial firm size



Though the analysis in this paper can be done for low of high levels of $p\phi$, the analogy of evasion with a loan becomes more interesting in the case where $p\phi > 1$. In fact, in standard models of tax evasion, this situation tends to lead to full compliance of firms or individuals. However, in the current model this is not necessarily the case, since this cost is weighed against the return of evading taxes and re-investing. For simplicity, the remainder of this paper will make the assumption that $p\phi > 1$.

Assumption 2. ***Positive interest rates****:* $p\phi > 1$*. Costs to evasion are high enough so that the cost of the expected penalty is greater than the value of the initial tax liability.*

Moreover, I will assume that A_1 is such that the firm is an *evader*, that is its compliance level x_1 lies strictly between 0 and 1 and is determined as an interior solution to the maximization problem.

Assumption 3. ***Evader****:* $A_1 \in (\underline{A}_1, \bar{A}_1)$*, such that* $x_1 = x_1^*$*.*

2.1 Comparative statics for the firm's problem

Compliance increases monotonically with size, simply because the benefit from evading is decreasing with size due to the concavity of the profit function. Apart from the two polar cases in which there is no evasion (small firms) or full compliance (large firms), there is partial compliance x_1^*. Partial compliance increases with the firm' initial size and with the probability of being audited, as expected. Full differentiation of the first order condition (equation 4) gives:

$$\frac{dx_1^*}{dA_1} = \frac{\pi'(A_1)}{\pi(A_1)} \frac{1 - \tau_1 x_1^*}{\tau_1} > 0 \tag{9}$$

Unambiguously, larger firms comply more with taxes than smaller firms, for a given technology. This result is compatible with stylized facts documented in the literature of tax evasion by firms, in particular in developing countries, as mentioned in the first section. In this model, this happens because a larger firm needs to evade less than a small firm to achieve the same size in the second period. A large firm would be risking too much downside by growing beyond the target A_2 given by the first order condition. Since they both want the same target A_2 (for a given technology), the smaller firm has to evade more.

Compliance also unambiguously increases with the probability of being audited, as would be expected. Higher probability of penalty increases compliance in the first period, because it reduces the marginal return of evasion.

$$\frac{dx_1^*}{dp} = -\frac{\phi}{1 - \tau_2} \frac{1}{\pi''(A_2)\tau_1 \pi(A_1)} = \frac{1}{\varepsilon_{\pi'}} \frac{1 - x_1 \tau_1}{p \tau_1} > 0 \tag{10}$$

Where $\varepsilon_{\pi'} \equiv -\frac{\pi''(A_2)}{\pi'(A_2)} A_2$ is the elasticity of the marginal returns to assets. The second equality uses the fact that $\pi'(A_2) = p\phi/(1 - \tau 2)$ and $A_2 = \pi(A_1)(1 - x_1 \tau_1)$.

Tax rates also have an impact on compliance levels. However, first period tax rates τ_1 have no impact on the target size A_2 of assets in the second period. Increasing taxes in the first period means indeed that the firm will evade more to achieve that target.

$$\frac{dx_1^*}{d\tau_1} = -\frac{x_1^*}{\tau_1} < 0 \tag{11}$$

Second period taxes τ_2 have no impact on second period decisions, as already discussed, since x_2 depends only on $p\phi$ being greater or smaller than 1. However, they have an impact on the target level A_2. This yields an expression that is very similar to the derivative of x_1^* with respect to p, seen above in equation 10:

$$\frac{dx_1^*}{d\tau_2} = -\frac{\pi'(A_2)}{\pi''(A_2)} \frac{1}{\tau_1 \pi(A_1)(1 - \tau_2)} = \frac{1}{\varepsilon_{\pi'}} \frac{1 - x_1 \tau_1}{(1 - \tau_2)\tau_1} > 0 \tag{12}$$

By reducing net returns on second period profits, τ_2 increases first period compliance unambiguously. Finally, we can check the sensitivity of x_1^* with respect to the penalty rate ϕ.

$$\frac{dx_1^*}{d\phi} = \frac{1}{\varepsilon_{\pi'}}\frac{1 - x_1\tau_1}{\phi\tau_1} > 0 \tag{13}$$

Equations 10, 11, 12 and 13 are useful to solve the government's problem, presented next.

3 The problem of a revenue maximizing government

The government raises revenues over the two periods using taxes and audits. Audits in this setting give the government the chance to tax twice the same tax liability. The reason is that that firms use evaded taxes to increase their size in the second period, which also increases tax liability in the second period. If the firm complies in the second period, the government benefits from the firms' evasion, because it taxes a larger firm. By auditing a firm that grew thanks to evasion, the tax authority makes sure that the full liability of the second period is taxed, and also the full liability in the first period. However, this amounts to double taxation, since the firm would have had another size in the second period if it had paid the full liability in the first period. This is illustrated formally in what follows. Define the government's revenues over two periods for a certain firm of initial size A_1 as $\mathcal{G}(A_1)$:

$$\mathcal{G}(A_1) = \underbrace{\tau_1 x_1 \pi(A_1) + \tau_2 x_2 \pi(A_2)}_{\text{tax revenues}} + \underbrace{p\phi(\tau_1(1 - x_1)\pi(A_1) + \tau_2(1 - x_2)\pi(A_2))}_{\text{audit revenues}} \tag{14}$$

Where x_1, x_2 are defined in the firm's maximization problem. Assume for simplicity that there is no cost of carrying out an audit. This assumption allows us to treat p and ϕ as equivalent. Indeed, what matters in the problem is $p\phi$, which will henceforth be treated as a single parameter. The maximization problem of the government is:

$$\begin{aligned} \max_{\tau_1,\tau_2,p\phi} \; & \mathcal{G}(A_1) = \tau_1 x_1 \pi(A_1) + \tau_2 \pi(A_2) + p\phi(\tau_1(1 - x_1)\pi(A_1) - p\psi \\ \text{s.t.} \quad & \pi'(A_1)(1 - \tau_1) \geq 1 \\ & \pi'(A_2)(1 - \tau_2) \geq 1 \end{aligned} \tag{15}$$

Taking the (unrealistic) assumption that the government knows what is the intial size A_1 of the firms, maximizing $\mathcal{G}$ gives optimal values for all policy parameters: τ_1, τ_2, p and ϕ. Taking first order conditions yields the following expressions:

$$\frac{d\mathcal{G}}{d\tau_1} = x_1\pi(A_1) + \tau_1\frac{dx_1}{d\tau_1}\pi(A_1) + \tau_2 x_2\pi'(A_2)\frac{dA_2}{d\tau_1} + p\phi(1-x_1)\pi(A_1) - p\phi\tau_1\frac{dx_1}{d\tau_1}\pi(A_1) \tag{16}$$

This expression is simplified by differentiating the first order condition of te firm's problem, equation 4, and getting $dA_2/d\tau_1 = 0$ and $dx_1/d\tau_1 = -x_1/\tau_1$.

$$\frac{d\mathcal{G}}{d\tau_1} = \begin{cases} p\phi\pi(A_1) & \text{if } x_1 = x^* \text{ or } x_1 = 0 \\ \pi(A_1) & \text{if } x_1 = 1 \end{cases} \tag{17}$$

This result means that as long as τ_1 respects the assumption of no wealth taxation, increasing first period taxes always raises more revenue, because it does not change the incentives to grow into the second period but increases revenues from penalties (for evaders and informal) or from first period taxation (for compliers). It follows that the revenue-maximizing tax rate in the first period is the highest possible, that is, the one such that $\pi'(A_1)(1-\tau_1) = 1$, or $\tau_1 = 1-\pi'(A_1)^{-1}$. Indeed, τ_1 works as an interest rate in the implicit loan taken by the firm, and as a lender, the government benefits from setting it to the highest level possible (i.e., the higher level at which the firm is willing to borrow money).

As for the second period tax τ_2, the first order condition yields:

$$\begin{aligned} \frac{\partial\mathcal{G}}{\partial\tau_2} &= \tau_1\frac{dx_1}{d\tau_2}\pi(A_1) + \pi(A_2) + \tau_2\pi'(A_2)\frac{dA_2}{d\tau_2} - p\phi\frac{dx_1}{d\tau_2}\tau_1\pi(A_1) \\ &= \underbrace{\pi(A_2)\big(x_2 + (1-x_2)p\phi\big)}_{\text{direct pos. effect}} - \underbrace{\frac{A_2}{(1-\tau_2)}\Big(\frac{\pi'(A_2)-1}{\varepsilon_{\pi'}}\Big)}_{\text{indirect size effect}} \end{aligned} \tag{18}$$

The above equation is equal to zero if and only if:

$$\tau_2 = 1 - \frac{\pi'(A_2)-1}{\varepsilon_{\pi'}}\frac{A_2}{\pi(A_2)\big(x_2+(1-x_2)p\phi\big)} \tag{19}$$

The revenue-maximizing government faces different incentives for the optimization of τ_1 and τ_2. Whereas raising τ_1 is always revenue increasing, this is not the case of τ_2, since higher tax rates in the second period reduce the firms' incentives to grow.

The first order condition of the government's problem with respect to the expected penalty for evasion $p\phi$ is given by:

$$
\begin{aligned}
\frac{\partial \mathcal{G}}{\partial p\phi} &= \tau_1 \pi(A_1) \frac{\partial x_1}{\partial p}\Big(1 - \frac{p\phi}{1-\tau_2} + \frac{\phi(1-x_1)}{dx_1/dp}\Big) \\
&= \underbrace{\pi(A_1)\tau_1(1-x_1) + \pi(A_2)(1-x_2)\tau_2}_{\text{additional audit revenues}} - \quad \underbrace{-\frac{A_2}{p\phi}\Big(\frac{\pi'(A_2)-1}{\varepsilon_{\pi'}}\Big)}_{\text{lower tax revenues in second period}}
\end{aligned} \tag{20}
$$

An increase in the probability of audits increases unambiguously the audit revenues, but decreases the tax revenues in the second period. The decrease in second period taxes comes from the fact that the firm is discouraged from evading in the first period, and therefore achieves a smaller size in the second period. The tax base is lower, yielding less taxes to the government.

This expression is equal to zero if and only if:

$$
p\phi = \frac{A_2 \frac{\pi'(A_2)-1}{\varepsilon_{\pi'}}}{\pi(A_1)\tau_1(1-x_1) + \pi(A_2)(1-x_2)\tau_2} \tag{21}
$$

In this problem, it is not optimal for the government to maximize compliance by setting the punishment ϕ to infinity, for example. In fact, increasing the penalty increases compliance, but discourages the firm from growing. A revenue maximizing government prefers firms to grow before taxing them, and uses the punishment to recover part of the evaded amount used to invest, that is, the part that was implicitly borrowed by the firm.

4 Discussion and conclusion

In this model, a firm evades to ease its financial constraint. The incentives to evade come from the fact that the marginal return on a unit of evaded tax is greater than the marginal penalty that it will have to pay on this amount. A firm evades until the expected marginal profit from evasion is equal to the marginal expected penalty payment, as is common in any classical model of tax evasion. The contribution relative to the literature is that it sheds light on the dynamic incentives that arise from the possibility of investing evaded resources productively.

One striking feature of the model is that firms may have the incentive to evade taxes even if audit probabilities are very high, even if it is equal to 100%, if the marginal return of an investent is high enough. The reason for is that firms use evasion as a loan from the government, where the expected penalty ($p\phi$) take the role of interests on this loan. Firms that have a high revenue potential in the second period find it economically advantageous to take this loan and pay the interest. This result echoes Andreoni (1992), who also found that some financially constrained individuals would evade personal income tax to smooth consumption, even if audit probability in the second period was 100%.

The other point made in this model is that evasion in this setting may provide a form of cheap finance for the government. The government benefits from evasion by taxing second period revenues, but it can still claim first period evaded tax liability. For this reason, the government would like to induce evasion in the first period. By doing that, it spurs firm growth, and still accumulates a credit with the companies, which it can claim by auditing them. The consequence is that governments may use distortionary taxation even in a setting with no information asymmetry and if a lump sum tax is available. Although governments in practice grant tax holidays for some taxes to nascent companies, taxing them more (short of making capital return negative) raises expected government revenues by increasing the government's claim on evaded taxes in the economy.

References

Albuquerque, Rui and Hugo A Hopenhayn (2004). "Optimal lending contracts and firm dynamics". In: *The Review of Economic Studies* 71.2, pp. 285–315.

Allingham, Michael G and Agnar Sandmo (1972). "Income tax evasion: A theoretical analysis". In: *Journal of public economics* 1.3-4, pp. 323–338.

Alm, James, Yongzheng Liu, and Kewei Zhang (2018). "Financial constraints and firm tax evasion". In: *International Tax and Public Finance*, pp. 1–32.

Andreoni, James (1992). "IRS as loan shark tax compliance with borrowing constraints". In: *Journal of Public Economics* 49.1, pp. 35–46.

Bachas, Pierre and Mauricio Soto (2021). "Corporate Taxation under Weak Enforcement". In: *American Economic Journal: Economic Policy* 13.4, pp. 36–71.

Brockmeyer, Anne and Marco Hernandez (2017). "Taxation, information, and withholding: evidence from Costa Rica". In: *World Bank Policy Research Working Paper* 7600.

Clementi, Gian Luca and Hugo A Hopenhayn (2006). "A theory of financing constraints and firm dynamics". In: *The Quarterly Journal of Economics* 121.1, pp. 229–265.

Cremer, Helmuth and Firouz Gahvari (1993). "Tax evasion and optimal commodity taxation". In: *Journal of Public Economics* 50.2, pp. 261–275.

Gatti, Roberta and Maddalena Honorati (2008). "Informality Among Formal Firms: Firm-Level". In: *Cross-Country Evidence On Tax Compliance And Access To CreditWorld Bank Policy Research paper* 4476.

Naritomi, Joana (2019). "Consumers as tax auditors". In: *American Economic Review* 109.9, pp. 3031–72.

Pomeranz, Dina (2015). "No taxation without information: Deterrence and self-enforcement in the value added tax". In: *American Economic Review* 105.8, pp. 2539–69.

Ulyssea, Gabriel (2018). "Firms, informality, and development: Theory and evidence from Brazil". In: *American Economic Review* 108.8, pp. 2015–47.
Yitzhaki, Shlomo (1974). "Income tax evasion: A theoretical analysis". In: *Journal of public economics* 3.2, pp. 201–202.

Appendices

A Derivation of proposition 1

Proof. The results above are derived by using the derivatives of x_1 with respect to each parameter (equations 10, 11, 12 and 13), and the derivatives of $\tilde{\Pi}$ and $\mathcal{G}$ with respect to each parameter. Thanks to the envelope theorem we can simply write the derivatives of $\tilde{\Pi}$ with respect to the parameters as:

$$\frac{d\tilde{\Pi}}{d\tau_1} = \frac{\partial\tilde{\Pi}}{\partial\tau_1} = -p\phi(1-x_1)\pi(A_1) \tag{22}$$

$$\frac{d\tilde{\Pi}}{d\tau_2} = \frac{\partial\tilde{\Pi}}{\partial\tau_2} = -\pi(A_2) \tag{23}$$

$$\frac{d\tilde{\Pi}}{dp} = \frac{\partial\tilde{\Pi}}{\partial p} = \phi(1-x_1)\tau_1\pi(A_1) \tag{24}$$

$$\frac{d\tilde{\Pi}}{d\phi} = \frac{\partial\tilde{\Pi}}{\partial\phi} = p(1-x_1)\tau_1\pi(A_1) \tag{25}$$

The derivatives of $\mathcal{G}$ are:

$$\begin{aligned}\frac{d\mathcal{G}}{d\tau_1} &= x_1\pi(A_1) + \tau_1\frac{dx_1}{d\tau_1}\pi(A_1) + \tau_2 x_2\pi'(A_2)\frac{dA_2}{d\tau_1} + p\phi(1-x_1)\pi(A_1) - p\phi\tau_1\frac{dx_1}{d\tau_1}\pi(A_1)\\ &= p\phi\pi(A_1)\end{aligned} \tag{26}$$

$$\begin{aligned}\frac{\mathcal{G}}{d\tau_2} &= \tau_1\pi(A_1)\frac{dx_1}{d\tau_2}(1-\tau_2\pi'(A_2)-p\phi) + \pi(A_2)\\ &= \frac{A_2}{\varepsilon_{\pi'}(1-\tau_2)}(1-\pi'(A_2)) + \pi(A_2)\end{aligned} \tag{27}$$

$$
\begin{aligned}
\frac{\mathcal{G}}{dp} &= \tau_1\pi(A_1)\frac{dx_1}{dp}(1-\tau_2\pi'(A_2)-p\phi)+\phi(1-x_1)\tau_1\pi(A_1)-\psi \\
&= \frac{A_2}{\varepsilon_{\pi'}p}(1-\pi'(A_2))+\phi(1-x_1)\tau_1\pi(A_1)-\psi
\end{aligned} \tag{28}
$$

$$
\frac{d\mathcal{G}}{d\phi} = \frac{A_2}{\varepsilon_{\pi'}\phi}(1-\pi'(A_2))+p(1-x_1)\tau_1\pi(A_1) \tag{29}
$$

Setting $\lambda = 1$ as a consequence of lump-sum taxes, we get that the expressions for the derivative of the objective function are just the sums of the derivatives of $\tilde{\Pi}$ and $\mathcal{G}$:

$$
\{\tau_1\} \qquad \frac{d\mathcal{L}}{d\tau_2} = -(1-x_1)+1 > 0 \tag{30}
$$

The above expressions shows that it is always advantageous to increase τ_1, since it induces the firm to evade more, raising the possibility of double taxation via audits in the second period. This double taxation is a cheap way to finance the government.

$$
\begin{aligned}
\{\tau_2\} \qquad & \frac{d\mathcal{L}}{d\tau_2} = \frac{A_2}{\varepsilon_{\pi'}(1-\tau_2)}(1-\pi'(A_2)) = 0 \\
& \text{iff } 1-\pi'(A_2) = 0 \\
& \text{iff } 1-\frac{p\phi}{1-\tau_2} = 0 \\
& \text{iff } \tau_2 = \phi p - 1
\end{aligned} \tag{31}
$$

As we will see below, τ_2 and ϕ have interdependent values, but are not determined. Therefore, we can simply set $\tau_2 = 0$ as one possible solution to the problem.

$$
\begin{aligned}
\{\phi\} \qquad & \frac{d\mathcal{L}}{d\phi} == \frac{A_2}{\varepsilon_{\pi'}\phi}(1-\pi'(A_2)) = 0 \\
& \text{iff } 1-\pi'(A_2) = 0 \\
& \text{iff } 1-\frac{p\phi}{1-\tau_2} = 0 \\
& \text{iff } \tau_2 = \phi p - 1
\end{aligned} \tag{32}
$$

This is exactly the same expression for the optimality condition of τ_2. Since we set $\tau_2 = 0$, it follows that $\phi = p^{-1}$. We can now find the value for p.

$$
\begin{aligned}
\{p\} \qquad & \frac{d\mathcal{L}}{dp} = \frac{A_2}{\varepsilon_{\pi'} p}(1 - \pi'(A_2)) - \psi = 0 \\
& \text{iff } p = \frac{A_2}{\varepsilon_{\pi'} \psi}
\end{aligned} \tag{33}
$$

$\square$