NBER WORKING PAPER SERIES

TECHNOLOGY AND TAX CAPACITY: EVIDENCE FROM LOCAL GOVERNMENTS IN GHANA

James Dzansi
Anders Jensen
David Lagakos
Henry Telli

Working Paper 29923
http://www.nber.org/papers/w29923

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2022, Revised June 2024

For helpful comments on this project we thank Augustin Bergeron, Anne Brockmeyer, Ray Fisman, Andrew Foster, Rema Hanna, Gordon Hanson, Asim Khwaja, Gabriel Kreindler, Craig McIntosh, Karthik Muralidharan, Dani Rodrik, Nii Sowa, Chris Udry, Silvia Vannutelli, Mazhar Waseem, Jaya Wen as well as seminar participants at Brown, BU, Dartmouth, Harvard, NYU Abu Dhabi, Peking HSBC Business School, USC, Virginia, Williams and Yale. For outstanding research assistance we thank Manon Delvaux, Soala Ekine, Radhika Goyal, Mary Nyarkpoh, Isaac Otoo and Cynthia Zindam. For help implementing a pilot version of the experiment we thank IPA Ghana. For financial support we thank the International Growth Centre and J-PAL. All potential errors are our own. Harvard IRB approval: IRB17-1310. AEA RCT Registry ID: AEARCTR-0007267. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Technology and Tax Capacity: Evidence from Local Governments in Ghana
James Dzansi, Anders Jensen, David Lagakos, and Henry Telli
NBER Working Paper No. 29923
April 2022, Revised June 2024
JEL No. H2,H71,O12,O33

ABSTRACT

This paper studies the role of technology to improve tax capacity in the developing world. We focus on local property taxation in Ghana, a potentially significant but under-collected source of revenue. We randomize the use of a new technology designed to help revenue collectors locate property owners to deliver tax bills, which is a major challenge in many developing countries with incomplete property addressing. We find that the technology increases bill deliveries by 27 percent and, surprisingly, increases tax collections by 103 percent. To reconcile these experimental findings, we build and estimate a dynamic time-use model in which revenue collectors respond to the new technology by shifting their allocation of time toward learning about households' propensity to pay, which is initially hard to observe, and subsequently collecting from those with highest payment propensity. The model's predictions are consistent with additional experimental results on time allocations, knowledge and collector strategies. Our explanation highlights how a technology that is designed to relax one constraint ultimately relaxes multiple additional constraints once behavioral changes by users of the technology are taken into account.

James Dzansi
International Growth Centre
c/o Institute of Statistical,
Social & Economic Research
University of Ghana
Legon, Accra
Ghana
james.dzansi@theigc.org

Anders Jensen
Harvard Kennedy School
79 JFK Street
Cambridge, MA 02138
and NBER
anders_jensen@hks.harvard.edu

David Lagakos
Department of Economics
Boston University
270 Bay State Road
Boston, MA 02215
and NBER
lagakos@bu.edu

Henry Telli
International Growth Centre
c/o Institute of Statistical, Social
& Economic Research
University of Ghana
Legon, Accra
Ghana
henry.telli@theigc.org

1 Introduction

Developing countries generally struggle to collect taxes and provide public goods. The literature on state capacity argues that the inability to collect taxes is at the heart of why low-income countries are as poor as they are (see e.g. Besley, Ilzetzki, and Persson, 2013; Dincecco and Katz, 2016). This literature suggests that the path to development may begin with investing in the capacity to collect taxes so as to finance productivity-enhancing public goods. Our paper studies the role of technology to improve tax capacity of local governments in Ghana. We focus on property taxes, a potentially significant but under-collected source of revenue (Brockmeyer, Estefan, Arras, and Suárez Serrato, 2023).

We begin by conducting the country's first census of local governments, which measures tax collection practices and outcomes in all 216 districts. The census highlights a key constraint at the beginning of the collection process, namely the limited ability to locate properties. This constraint stems from incomplete addressing: in the average district, only 27 percent of properties are assigned a number on a named street. The location information available to local tax collectors is therefore imprecise, and most report struggling to find taxpayers. As a result, less than half of the tax bills in the average district end up being delivered. To help improve the delivery of property tax bills, a small number of local governments have invested in digital tax registries with exact geographic coordinates of properties. We document that investment in this geographic information system (GIS) technology is associated with more bills delivered and also more taxes collected per bill delivered. This raises the possibility that the GIS-technology may have helped improve collection in ways beyond locating properties to deliver bills.

Motivated by this association in the census, we collaborated with one local government to randomize the use of a new GIS technology and causally investigate how it affects the entire collection process. The experiment randomized the presence of an electronic tablet with GIS capabilities at the level of a tax collector. Both treatment and control collectors were given a stack of bills of similar average value in a randomly assigned work area and were tasked with collecting as much revenue as possible during the tax campaign. Each treatment collector was given the tablet with geo-spatial data that was supposed to increase delivery by improving navigation and taxpayer localization. All other aspects of the bill, collector, and work area were constant across groups.

We find that collectors using the technology delivered 27 percent more bills than control collectors by the end of the campaign. Treatment collectors reported substantially less trouble navigating in their area and locating property owners and spent significantly less time per bill delivered, confirming technology's advantage to deliver bills in an environment with scant addressing. More surprisingly, treatment collectors collected 103 percent more taxes than control collectors, which implies a treatment effect on collections that is almost four times as large as on bill delivery. We frame our investigation of mechanisms as trying to explain how a technology which was primarily designed to improve deliveries caused such a large effect on taxes collected per bill delivered.

One potential explanation is that the technology increased the likelihood that a household pays by improving tax morale or increasing perceptions of enforcement capacity. However, we find that technology has no average or heterogeneous impacts on numerous dimensions of enforcement perceptions and tax morale, including views on government's efforts to improve the efficiency of collection. A second possibility is that the presence of technology discouraged the control collectors by denying them access to a potentially valuable tool to improve an otherwise challenging work environment. However, control collectors report similar job satisfaction, hours of work, and reported work challenges – other than navigation – as treatment collectors. Control collectors that (randomly) differed in their exposure to technology were also comparable in these dimensions. A third possibility is that the presence of technology increased collectors' perception of monitoring or households' perceived bargaining position, leading treatment collectors to substitute away from collecting bribes towards collecting tax payments. If anything, however, we find that bribes increase in treatment areas. Moreover, there are no differences between groups in how collectors report being monitored by supervisors or in the resistance they report facing by households.

Our preferred mechanism, which is novel in the literature, is that technology led revenue collectors to re-optimize their behavior across multiple work domains, focusing less on deliveries and more on collections and learning about property owners. In particular, we argue that collectors allocate more time to collecting overall, and to learning about the hard-to-observe household characteristics that determine payment propensity, such as income, liquidity, and awareness of the tax code. Collectors can then subsequently use this local knowledge to target the households with higher propensity to pay.

We provide two sets of results that support this mechanism, using both a detailed endline household survey and a collector panel survey measured at multiple points during the tax campaign. First, technology has significant impacts on time allocations. Collector surveys reveal that treatment collectors allocate a larger share of their time to non-delivery activities than control collectors. Household surveys show that treatment collectors conduct more follow-up visits with property owners, and spend longer time interacting with property owners during each visit.

Second, treatment collectors appear to learn in the field. Panel collector surveys reveal a significant increase over time in treatment collectors' knowledge about which households have a higher propensity to pay. Treatment collectors also report a greater use over time of collection strategies that focus on households with hard-to-observe characteristics that signal higher propensity to pay. Independent evidence from household surveys show selection patterns that are consistent with these results: property owners with higher actual propensity to pay (e.g. with higher income, liquidity and taxpayer awareness) are more likely to pay taxes in treatment areas. Moreover, though treatment collectors collect more payments in total, a larger fraction of the payments comes from the 'high type' of households with the largest 5 percent values of a propensity to pay index. None of the information on households' propensity to pay was initially provided to collectors, or visible in the GIS-tablet or on the tax bills. Instead, consistent with the more frequent visits of longer duration observed in treatment areas, our interpretation is that the treatment collector learns about the hard-to-observe payment propensity during these visits by interacting with the households and by surveying the local environment.

We formalize our theory in a dynamic model in which forward-looking revenue collectors maximize collections subject to a time constraint each period. Households have a high or low payment probability, and the type is initially unknown to the collectors. Collectors spend their time endowment each period trying to deliver bills, learn about taxpayers, and collect from households of high-, low-, or unknown types. We assume that the collection probability from each household type is the same in treatment and control groups. Technology increases the return to time spent delivering bills, reflecting the navigational improvement. Technology also increases the return to time spent learning. The learning advantage reflects in part the navigational improvement to conduct follow-up visits; moreover, evidence from environmental psychology suggests that improved navigation may encourage the collector to make the effortful and strategic choice to build survey knowledge on households, their properties and the environment.

We estimate the main model to match the 27 percent experimental treatment effect on bill delivery and several other moments representing average outcomes across both groups. The model's predicted treatment effect on collections, which is not targeted, is 77 percent, meaning that the model explains about two-thirds of the experiment's difference between the treatment effects on collections and deliveries. The model also predicts, without targeting, the positive experimental selection result that a significantly higher fraction of collections in the treatment group comes from the high-type households. We show that these predictions arise from endogenous differences in time allocations: Treatment collectors shift from bill delivery to learning and collections earlier in the tax campaign than control collectors, and spend a larger fraction of their time gaining knowledge and a smaller fraction delivering bills; control collectors, continuously hindered by limited navigation, practically avoid learning and switch directly from delivery into collection, though much later on in the campaign.

To disentangle the direct effects of the technology, stemming from the higher returns to delivery and learning, from the indirect ones, arising from the collectors' time re-allocations, we simulate a counterfactual model that endows the collector with the technology but without allowing them to re-optimize their time allocations. This counterfactual yields treatment effects of similar magnitude on deliveries (52 percent) and collections (46 percent) – implying that collector re-optimization is a significant component of the overall experimental impact of technology. In particular, the main model's treatment effects on collections are increased by 67 percent relative to the direct, 'pure' effects of the technology without re-optimization (77/46 = 1.67). Thus, an envelope-theorem logic does not apply well here, meaning that the total effects of technology are not well approximated by their direct effects, where re-optimization of agents is ignored.

We next simulate a counterfactual model in which we allow collectors to re-optimize but where technology only provides a delivery advantage and no learning advantage. In this scenario, the model predicts treatment effects that are twice as large for collections as for deliveries. Thus, the combined effects of a delivery advantage and collector time-reallocation explain around half the gap between the treatment effects on deliveries and collections in our experiment. Moving to the main model adds back the learning advantage, and results in treatment effects that are three times as large for collections, or around three quarters of the gap in the experiment, as explained above. Quantitatively, then, 60 and 40 percent of the model's predicted gap in treatment effects between collection and delivery stem from the delivery and learning advantages, respectively.

Finally, of all the simulated models, only the main model results in significant learning by treatment collectors and, as a result, positive selection. In other words, time re-allocation and a learning advantage are each necessary but not sufficient for learning to emerge as an important activity in the field.

We conclude that technology that helps relax the constraint on bill deliveries not only increases collections directly, by improving delivery, but also indirectly, as collectors devote more time to other activities for which constraints still strongly bind. In this setting, as in other developing countries, tax officials are constrained by limited enforcement capacity, which necessitates multiple follow-up visits with taxpayers after bill delivery, and by limited information about taxpayers, which makes it hard to know which households are worth targeting for collection efforts. By improving the collector's return to time spent delivering, the technology allowed collectors to allocate more of their scarce time to the challenging tasks of learning about taxpayers and attempting to collect from them, particularly those with the highest propensity to pay. Quantitatively, we find that these indirect effects of technology are large, and help explain the bulk of why a technology initially designed to help bill delivery ultimately had such a large impact on collections.

Related literature Our paper contributes to the empirical literature on technology adoption, in particular worker responses to new technologies (Atkin et al., 2017; Prescott and Parente, 1994). The tax collector in Ghana is the archetypal street-level bureaucrat (Lipsky, 2010), who faces limited oversight and exercises significant discretion in deciding how to achieve loosely defined objectives. A related literature studies technologies that monitor the bureaucrat's activities, including to evaluate their performance and potentially curb their discretion: recent studies include Callen, Gulzar, Hasanain, Khan, and Rezaee (2020); Dal Bó, Finan, Li, and Schechter (2021); Dodge, Neggers, Pande, and Moore (2021); Mattsson (2023); Muralidharan, Niehaus, Sukhtankar, and Weaver (2021). Our study instead focuses on how technology can enhance the bureaucrat by alleviating constraints on their field activities. In so doing, we show how workers respond to a technology that helps relax one constraint by productively re-allocating their time toward the constraints that bind more sharply.[1]

Learning valuable local information emerges as a key activity that government officials re-allocate time towards.[2] Previous studies in public finance have shown how *pre-existing* information, from third-parties and local chiefs, can improve taxation and policy (Balan, Bergeron, Tourek, and Weigel, 2022; Kleven, Knudsen, Kreiner, Pedersen, and Saez, 2011; Manara and Regan, 2022; Naritomi, 2019; Pomeranz, 2015). Our paper shows how, in settings where such information-sources are practically non-existent, the state can still strengthen tax capacity by *directly building* locally relevant information.[3]

Our paper studies the impacts on government performance of improving localization in a setting with incomplete addressing. Incomplete addressing is a well-known issue in Ghana, but is not unique to the country: the UN estimates that 4 billion people live in places without an address (link). Moreover, GIS-technologies to overcome challenges

[1]Though it has received less attention in empirical economics, it is well known in public administration research that street-level bureaucrats face multiple constraints that interact with each other. Beyond the public sector context, Suri and Udry (2022) emphasize that no single constraint, but rather the combination of multiple constraints, explains the lack of widespread technology adoption in African agriculture.

[2]Allocation of time use has received limited empirical attention as a dimension of bureaucrat performance (Finan et al., 2017), with recent exceptions (Bandiera, Best, Khan, and Prat, 2021; Kalaj, Rogger, and Somani, 2022). For detailed survey measures of bureaucrats' activities, see also Rasul and Rogger (2018).

[3]The local information gathered by collectors could potentially be used by expenditure officials and allow states to revisit their reliance on non-state actors to target transfers (Basurto, Dupas, and Robinson, 2020). Our results on how officials build local information over time complement previous studies which focus on measuring officials' local information at a point in time (Dal Bó et al., 2021; Duflo et al., 2018).

from incomplete addressing are being implemented by governments around the world (Knebelmann, 2022). Despite its ubiquity, there is little work in economics on the consequences of incomplete addressing. Our experimental results shed light on the value of an address for government's ability to carry out core activities in the field.

Finally, our work relates to studies on technology and tax capacity, including Brockmeyer and Somarriba (2022); Eissa and Zeitlin (2014); Fan, Liu, Qian, and Wen (2021); Mascagni, Mengistu, and Woldeyes (2021); Okunogbe and Pouliquen (2022) and Das et al. (2023); see Hjort and Tian (2024) and Okunogbe and Tourek (2024) for recent reviews. To our knowledge, our paper is the first to directly randomize the presence of technology in the field for tax officials. Moreover, by focusing on the localization benefits of a GIS-based registry, our paper complements work on enhancing tax registries. Okunogbe (2021) studies the impacts on tax compliance of providing information made available by an enhanced registry. In the context of building a digital registry, Knebelmann et al. (2023) study the impacts of providing officials with discretion to determine property valuations. Ferraz, Foremny, and Santini (2024), Gadenne (2017), and Martinez (2023) study the interactions between non-tax revenue flows and registry enhancements, and Casaburi and Troiano (2016) analyze the electoral impacts of an enhanced registry.[4]

2 Census of Tax Capacity in Local Governments

There is limited systematic evidence about the process of tax collection by local governments in developing countries. To better understand this process, and the constraints on collection, we conducted Ghana's first census of all local governments in 2017, focusing on taxation. In this section we summarize the main findings from this census.

In each of the country's 216 local governments, we interviewed three sets of respondents: officials, locally elected assembly members, and citizens. Every local official involved in the tax collection process was surveyed, including: the chief executive (political head); the coordinating director (bureaucratic head); finance officers; physical planning officers; revenue supervisors; and, field collectors. Survey modules for officials and assembly members captured information on the taxation process and demographics. Surveys of citizens measured tax knowledge and morale and demand for public goods.

The census contains responses from 5,375 citizens (approximately 25 per district) and 2,785 local government officials and assembly members (13 per district). We also digitized and harmonized administrative records to measure all sources of local tax collection and all types of local public expenditure across the 216 districts.

[4]Our work also relates to the literature on experiments with tax collectors, including Bergeron, Bessone, Kabeya, Tourek, and Weigel (2022), Khan, Khwaja, and Olken (2015, 2019), and Weigel (2020).

Table 1: Characteristics of Local Tax Capacity in Ghana

	Mean	Median
Taxes collected per capita (GHC)	4	3
Share of bills delivered (%)	43	43
Share of delivered bills that are paid (%)	30	29
Share of properties with address (%)	27	21
Common not to locate properties	78	100
Take tax defaulters to court	22	0
Citizen has tax awareness	7	0
Citizen has public goods awareness	35	29
Cost of collection (% taxes collected)	64	48
Has electronic property database	15	0
Number of local governments	216	216

Note: All variables are calculated at the local government level using unweighted averages. For a detailed description of all variables, see the Data Appendix. B.1.

Limited local taxation Table 1 shows that local governments on average collected only around 4 Ghanaian cedis (GHC, equal to 0.32USD) per capita in the year of the census. This is a small amount relative to per capita income, and relative to the expectations of the Ghanaian central government, which views current property tax levels as being far below their potential (Government of Ghana, 2014).

Tax collections are determined by the probability of bill delivery (the delivery margin) and the amount paid conditional on delivery (the payment margin).[5] On the delivery margin, we find that in the average local government only 43 percent of property tax bills are delivered (Table 1). The delivery margin is thus an important determinant of low tax collections. Many studies in public finance and development abstract from bill delivery and focus instead on the payment margin. The payment margin is also an important factor in this setting: in the average district, officials report that the likelihood a property owner pays their property taxes after receiving a bill is just 30 percent.

Constraint: Incomplete addressing What are the most important factors limiting local tax collection? When presented with this question, the absence of data on properties was

[5]The denominator in the bill delivery measure is the set of properties that are registered; this delivery margin is different from the registration margin (the share of existing properties that are registered).

the most frequently cited constraint by local officials and assembly members (Figure A1). The absence of property valuations was second. When asked in turn about the constraints on bill deliveries, the absence of data on property owners and challenges in locating them are cited as two of the three most important constraints (Figure A1).[6] These responses highlight how limited information on property owners hinders revenue collections right from the start of the taxation process by making bill delivery difficult.

The lack of information starts with the simple absence of precise street addressing. Table 1 shows that in the average district, only 27 percent of properties have an official address, meaning a property number on a named street. The property tax registry inherits this limited address information, and hence the location listed on most property tax bills is imprecise. Typically, the main reference is a nearby landmark. For example, Figure A2 provides an illustration of an actual tax bill where the location of the property is listed as "Opposite Presec School" (a secondary boarding school). In the absence of precise addressing, collectors typically navigate to the landmark and then attempt to locate the targeted property (Figure A3). Unfortunately, these attempts often result in failure. Across all of Ghana's districts, 78 percent of local revenue collectors reported that it is common not to be able to locate the intended property for delivery (Table 1).[7]

Incomplete addressing is a well-known issue in Ghana[8] but is not unique to the country. The United Nations estimates that 4 billion people live in places without an official address (link), while Farvacque et al. (2005) estimate that half of the global urban population does not have a named street and property address. The existence of incomplete addressing has received little attention in economics. While the policy target may seem simple, a multi-disciplinary literature has identified several factors that limit broad addressing coverage. These factors, discussed in Appendix C.1, often reflect deeper constraints on a country's administrative, legal, and social governance capacities.[9]

Other constraints: Enforcement and local information The census highlights several additional constraints on the tax collection process. Even if a collector can find a property and deliver the bill to the owner, they typically must conduct follow-up visits with

[6] The absence of precise information on taxpayers has been documented in national tax authorities around the world: of 61 assessments in lower income countries conducted by the World Bank, only 5% received a score of good or better for the accuracy of information in the taxpayer registry (Nyanga, 2021).

[7] Table A1 also shows a positive association across the country's districts between the share of properties with addressing and the share of bills that are delivered, suggesting that limited addressing is ultimately a binding constraint on local governments' activities. These observations are consistent with the central government's message that local governments "have no accurate spatial reference" and that "the importance of street and property addressing cannot be overstated" (Government of Ghana, 2011).

[8] Some examples of discussions in the media include link #1, link #2 and link #3.

[9] Some of these challenges may be particularly pronounced in slums (Marx, Stoker, and Suri, 2013).

the owner to collect payment. Even after multiple follow-up visits, the likelihood of collecting any payment remains low. The lack of enforcement capacity by local governments undoubtedly plays a role in the low payment rates conditional on receiving a bill. The main enforcement action that local governments can take is to summon the delinquent taxpayer to court. However, just 22 percent of local governments resorted to this for any delinquent property owner during the past year (Table 1).[10] In high-capacity tax systems, enforcement is supported by the existence of third-party information coverage (Gordon and Li, 2009; Kleven et al., 2016; Pomeranz, 2015); such sources of 'hard information' are virtually non-existent at the local government level in Ghana, however.

It is precisely in settings where hard information is scarce and enforcement capacity is limited that 'soft' information on property owners' propensity to pay can be helpful (Luttmer and Singhal, 2014). Balan et al. (2022) show, for example, that traditional leaders in the DRC have relevant soft information on property owners that can be leveraged to collect taxes. The challenge is that propensity to pay may vary significantly across households and be hard for officials to initially observe. For example, knowledge of the tax code and local public good provisions are potential determinants of payment propensity. In our citizen surveys, however, we find that only 7 percent of citizens in the average district were aware of the "fee fixing resolution" that underlies the official tax rates and regulations, and only 35 percent could name any project undertaken by their local government in the past two years (Table 1). Knowing which citizens have higher propensity to pay taxes could help revenue officials target their collection efforts more effectively. Unfortunately, few are likely to have much soft information on taxpayers.

Reflecting the many challenges in the collection process, most districts have a high cost of collecting taxes. We proxy a district's collection cost as the average monthly salary of its revenue collectors as a percent of their average monthly collections. This cost is 64 percent in the average district in Ghana. By contrast, the IRS in the United States estimates that their average cost of collecting taxes is just 0.3 percent (link).

Potential role of technology Local governments can invest in technology to alleviate constraints on tax capacity. One technology that seems promising in this setting is a GIS-enhanced tax registry, which can help increase bill deliveries by improving localization (World Bank, 2020). Constructing this registry requires digitally recording the geographic coordinates of all relevant parcels using GPS coordinates and aerial data. A GIS-based registry can provide the local government with precise location information

[10] Survey responses by officials indicate that the reasons for limited court action lie outside the tax administration's immediate scope, and are due to legal constraints and political costs.

for delivery purposes without needing to have an official addressing system in place.[11]

The census reveals that only 15 percent of local governments have adopted a GIS-enhanced tax registry. Adoption of technology is at the discretion of each local government. High costs surely play a role in the limited adoption, as building a digital tax registry requires significant manpower and funding; limited local evidence on the cost-effectiveness of this type of technology likely also increases the local government's reluctance to invest. Moreover, technical failures to maintain the digital registry can bring field activities to a halt, and it is possible that the technology may be subverted by officials in the field. Table A2 provides cross-district correlates of adoption choices. Local governments are more likely to adopt technology when their district has: a larger share of properties with official valuations; stronger legal capacity; a larger population size; and, a more urban composition. These results suggest that technology investments are complementary to other determinants of tax collection (Besley and Persson, 2011).

Across local governments, Table A3 shows that investment in technology is robustly associated with higher tax collection and a higher share of bills delivered. These associations hold when controlling for the district covariates that predict adoption (Table A2), the share of neighboring local governments with technology, and region fixed effects. In the specification with all controls, adoption of a GIS-tax registry is associated with an 18 percent improvement in the bill delivery margin. To support the interpretation that this impact reflects improved localization due to incomplete addressing, Table A1 shows that the association between delivery and technology is larger when the district's addressing coverage is smaller.[12] Finally, Table A3 shows that adoption of the GIS-technology is also associated with more taxes collected per bill delivered.[13] This result suggests that a technology that is initially meant to alleviate constraints on delivery may have helped alleviate additional constraints on collections that bind once deliveries have been made.

3 Experiment and Main Tax Results

The associations in the census suggest a plausible role for GIS-technology to improve tax capacity, both by directly relieving the intended delivery margin and by indirectly helping to alleviate other constraints on taxation. Motivated by these patterns across

[11]We review non-technology initiatives to overcome addressing constraints on taxation in Appendix C.2, which faced implementation issues in Ghana and elsewhere (Abebrese, 2019; Bigon and Njoh, 2012).

[12]This result suggests that GIS-technology and addressing are substitutes and is consistent with the observation that property tax registries are not systematically associated with the use of GIS technologies in high capacity settings where broad addressing is sufficient to support tax collection (Knebelmann, 2022).

[13]These results hold when we control for whether the district made additional investments in technologies that automate the creation of tax bills and that assist with payment recording and enforcement.

local governments throughout the country, we implement an experiment within one local government to causally establish the impacts of a GIS-based technology and to precisely investigate mechanism impacts on different tax constraints.

3.1 Setting, Design and Data

Setting We conducted the experiment in 2021 in La Nkwantanang Madina Municipal Assembly (henceforth, Madina). Madina is more affluent and urban than the average district. We collaborated with the municipal government and a private Ghanaian firm which developed a new technology to help increase property taxes. The technology is based on a digital property tax registry with precise geo-coordinates of each structure that was constructed from high-resolution aerial photographs and in-person visits. The digital registry is accessed using a hand-held electronic tablet with GIS capabilities. The tablet provides the user's live location on a digital map as well as the location of a designated property (Figure A3), but does not automate navigation.[14] The tablet aims to assist collectors by helping them navigate to specific properties in the field. As we detail below, what we vary across treatment and control groups is the presence of the tablet.[15]

During a fiscal year, the local government assigns collectors to designated geographical areas for approximately six weeks at a time (a 'campaign'). The designated areas are called 'collection units' and are defined with geographical boundaries that create a cluster of physically adjacent properties (Figure A4). During each six-week campaign, collectors are responsible for delivering bills and collecting payments from assigned property taxpayers in their collection unit. After each campaign, the collector is assigned to a new collection unit. Each area of Madina is covered only once during a fiscal year, due to the large number of properties relative to the limited number of collectors. Property owners are legally required to pay within four weeks of receiving the tax bill. Pay stations do exist, but virtually all payments are made directly to the collector and most often in cash.

Our experiment was embedded in the six-week campaign between March and April 2021. Before the campaign, collectors received training. The main training sessions, common to all collectors, described the rules for property tax collection in Madina and the protocols to follow during interactions with property owners. In addition, the collectors assigned to the treatment group received training in how to use the handheld tablets. The compensation scheme, an 8 percent commission rate on taxes collected from

[14] Such as by calculating the most efficient route or providing turn-by-turn directions. Indeed, based on its features, the tablet scores only 2 out of 10 on the system automation scale of Parasuraman (2000).

[15] The GIS-data was also integrated into a software that compiles tax bills, records payments and issues enforcement notices, but these technology components are not randomized in the experiment.

assigned bills, was chosen by the local government and held constant across groups. Collectors also received a daily transportation allowance and a base salary.

Experimental design Out of 56 collectors that were trained, 28 were randomly assigned to the treatment group and 28 to the control group. Of the 56 collectors, 39 had previously worked with the firm and local government and 17 were hired shortly before the experiment. Of the 39 collectors with previous experience, 11 were rated as 'high performing' by the firm and local government. Collectors worked individually in their assigned collection unit and were assigned to approximately 145 bills each. Each collector had a supervisor randomly assigned to them during the campaign. All supervisors were randomly assigned to both treatment and control collectors, and were in charge of monitoring the revenue collectors and assisting them with challenges in the field.[16]

At the beginning of the experiment, all collectors in the treatment group were given the tablet for use during the tax campaign. Other than the tablet, the treatment group was not provided with any other advantages. Both groups were provided with the printed bills for all properties in their respective collection units, which contain information on the property's location (though imprecise), the amount of taxes due, and the amount of past taxes due, if any (see Figure A2 for an example).

Both groups were provided with physical maps that provide limited, aggregated spatial information by delineating the collection unit's boundaries relative to a small set of main roads. Qualitative work from our pilot revealed that control collectors hardly made use of the maps due to their lack of detail on specific property locations, however (Figure A3). We therefore consider that the control group represents a reasonable approximation to the status quo where collectors do not rely on technology in their field work.[17]

Our randomization proceeded in two steps. First, we randomly assigned each collector to a collection unit. Second, we randomly assigned the collector-collection unit pair to the treatment or control group. We stratified on the share of properties in the collection unit that were businesses (rather than residential). To avoid chance imbalances, we ran the full randomization 100 times and selected the run with the minimum t-statistic from

[16] The implemented experiment differs in two ways from the protocol described in the pre-analysis plan. First, 56, rather than 60, collectors were included because 4 dropped out before assignment to treatment. Second, the pre-analysis plan indicates that the navigational tablet: records the payment status of all assigned properties; is monitored in real time; and, issues a digital receipt upon payment. For logistical reasons, these features were not active or available in the tablet for the experiment.

[17] The physical map was created based on the GIS-enhanced registry but, as noted, control collectors in practice did not make much use of this map. Moreover, apart from a potential improvement in the accuracy of the delineations, the aggregate spatial information conveyed in the map is similar before and after the GIS-enhanced registry. Finally, no other technological tools related to the GIS-registry were used during the experiment, including for supervisors.

balance checks on six variables (as in Banerjee, Chassang, Montero, and Snowberg, 2020). Two of these variables were specific to collectors: a dummy for previous work experience with Madina and the firm, and a dummy for high-performance rating. The other four variables were specific to the collection unit: total bills to deliver; total taxes (current due and arrears); average current amount due per bill; and average previous pay status per bill (unpaid, partially paid, fully paid). Table A4 summarizes balance checks for a large set of characteristics at the tax bill level (Panel A), the collector-unit level (Panel B), and the household-level (Panel C). None of the individual characteristics are statistically significantly different between groups at 10%. Moreover, we reject to fail the null that the difference in all characteristics are all zero at the tax bill-level ($F = 0.71, p = 0.66$), collector-unit level ($F = 0.16, p = 0.95$) and household-level ($F = 1.07, p = 0.38$).

Experimental data and estimation In this section, we describe the data sources used in our analyses. We use administrative data at the property level, covering 8,120 residential and business properties, which contain information on owner names, property location, current tax due and arrears. This data set served to create the collection units for all collectors and to issue all the bills that were to be delivered during the tax campaign.

Our research team supported the local government's gathering of daily data from each collector on the number of bills delivered and the amount of revenue collected. These data, gathered in a uniform manner from all collectors, allow us to study the activities of both groups at a high frequency. Collecting these data is part of the government's routine campaign process, which helps explain the very low attrition (uncorrelated with treatment). To minimize the role of idiosyncratic measurement error at the daily level, our main results winsorize outcomes at the 95^{th} percentile by group and day.

In addition to the daily data, enumerators working for the research team conducted three rounds of detailed surveys with the 56 collectors. The first round was conducted during the initial week of the campaign; the mid-line during the third and fourth weeks; and, the end-line at the end of the sixth week. There is 17% attrition in the collector surveys: Table A5 shows this is uncorrelated with treatment. The main tables report treatment effects based on the unbalanced sample, though Table A6 shows that the results are similar in the balanced sample. This implies that any characteristic which predicts attrition is not a significant source of treatment effect heterogeneity. Notwithstanding, we use the balanced sample in all graphs that show levels of variables by group and survey round, to remove any influence from compositional effects between rounds.

Finally, the enumerators administered end-line surveys with 4,334 randomly selected households in May 2021. A random sample of equal size was drawn from each of the 56 collection units. Whenever an initially selected property could not be contacted, an adjacent property within the same collection unit would be randomly chosen. All outcomes created with these data are described in detail in Data Appendix B.

Given the random treatment assignment, we use OLS to estimate the causal impacts of technology. The econometric specification varies slightly depending on the unit of observation. For outcomes that vary at the day and collector level, we estimate:

$$y_{cd} = \beta_d \cdot \mathbf{1}(Tech)_c + \theta_d + \Omega \cdot X_c + \epsilon_{cd}, \tag{1}$$

where y_{cd} is the outcome for collector-collection unit c on day d, θ_d are campaign-day fixed effects, and X_c contains time-invariant controls. In the main analysis, X_c only includes strata fixed effects for the share of businesses in total properties. In robustness checks, we include additional controls for previous work experience in Madina, a dummy for high quality collector rating, total number of bills to deliver, and the average tax due per bill. The dummy $\mathbf{1}(Tech)_c$ takes a value of 1 for all collector-units randomly assigned to treatment, and 0 for collector-units assigned to control. The treatment coefficient, β_d, is indexed by day because we estimate dynamic treatment effects by interacting the treatment dummy with the individual campaign-day fixed effects. In a robustness check, we leverage the panel-structure and include fixed effects for each collector-collection unit. In this case, the identifying variation is the treatment effect that varies within a collector-unit over time, relative to the initial impact on day 1, β_1 (the omitted category). Standard errors are clustered at the collector-unit level. We also use (1) to estimate impacts in the collector surveys, replacing day d by survey round s.

For outcomes at the household level, we estimate:

$$y_{hc} = \beta \cdot \mathbf{1}(Tech)_c + \Omega \cdot X_{hc} + \epsilon_{hc}, \tag{2}$$

where h indexes households and c collector-units. Standard errors are clustered by collector-unit. X_{hc} always includes strata fixed effects. In robustness checks, we also include the controls at the collector-unit level from (1), as well as the household's property category and previous pay status (fully paid, partly paid, not paid).

3.2 Main Experimental Effects on Tax Outcomes

We begin by studying the impacts of technology on bill delivery and tax collection using the collector daily reports. Figure 1 shows the average bills delivered (Panel A) and taxes collected (Panel B), by group and campaign-day. Figures A5 and A6 show the corresponding daily treatment coefficients β_d (equation 1). Panel A shows that the treatment group delivers more bills than the control group. This difference initially builds up and peaks by the 24^{th} day, where treatment collectors have delivered 34 more bills than the control group (a 58 percent increase). The gap narrows in the second half of the campaign, where the stock of bills delivered in the treatment group steadies while control collectors continue to deliver bills. The treatment coefficients are statistically significant at the 5 percent level in all campaign-days beyond the 10^{th} day (Figure A5). At the end of the campaign, the treatment collectors have delivered 21.5 more bills on average, representing a 27 percent increase over the 80.7 bills in the control group.

Panel B of Figure 1 shows that technology causes a large increase in total taxes collected. There are no differences in collection performance during the first week, when most collectors focus on bill delivery. However, from the second week onward, the treatment group collects at a higher rate; the treatment effect is statistically significant at the 5 percent level on all subsequent days and grows over time (Figure A6). In the end, the treatment group has collected an additional 856 GHC on average, representing a 103 percent increase over the 829 GHC collected on average in the control group.[18]

We can infer from these results that the treatment group collects more taxes per bill delivered. Appendix Figure A7 shows that this outcome grows over time; at the end of the campaign, the treatment group has collected substantially more taxes per bill delivered than the control group. This result implies that the tax collection impact is not only driven (mechanically) by the increase in bills delivered. This result mirrors the census cross-sectional regression, where we also found a higher collection rate per bill delivered in local governments with GIS-technology than without (Table A3).

Robustness Figures A5 and A6 explore the sensitivity of the results to three alternative specifications. First, the estimates are similar when using non-winsorized outcomes.[19] Second, results are similar, but more precisely estimated, upon including the additional covariates contained in X_c (equation 1). Third, the results are comparable when we include collector-unit fixed effects; in this case, β_d reflects the treatment effect based on changes within collector-unit over time (relative to the initial impact β_1 on day 1).[20]

An important concern is whether the context of the COVID-19 pandemic impacted the results. We conducted a pilot experiment in early 2019 in the same location, using the same technology and the same research protocol (though with a smaller sample of

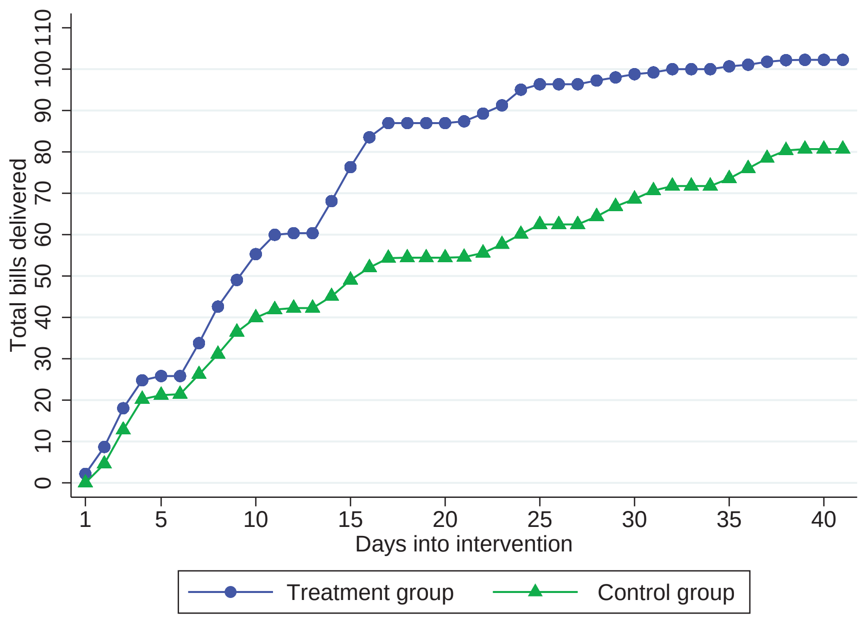
[18]Against its running costs, technology's cost-benefit analysis is positive (Table A7).

[19]Figure A8 shows that the results are almost identical across all sub-samples which leave out one collector at a time, alleviating concerns over undue influence by one outlier performing collector.

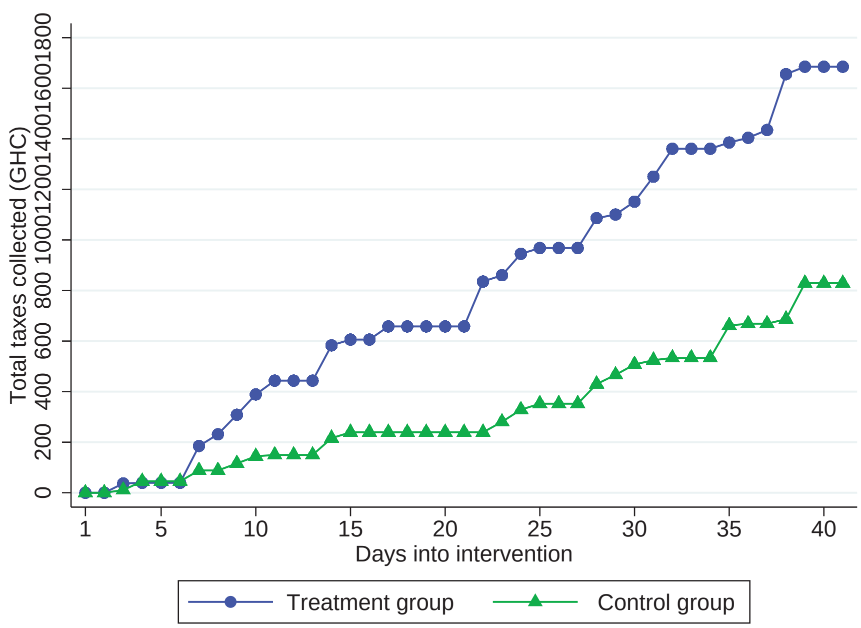
[20]The fixed effect technically captures variation within each collector-collection unit. However, we interpret it as reflecting a treatment effect over time within collector, since we found no evidence suggesting there are time-varying effects within collection units unrelated to changes in collectors' behavior.

Figure 1: Impact of Technology on Bills Delivered and Taxes Collected

(a) Bills Delivered



(b) Taxes Collected



Notes: This figure shows the impacts of technology on the number of property tax bills delivered and taxes collected. Panel A shows the average number of bills delivered by group and by day of the tax campaign. Panel B displays taxes collected on average by group and day. Figures A5 and A6 show the corresponding daily treatment coefficients (β_d in equation 1) for bills delivered and taxes collected, respectively. The analysis is based on the daily collector data, described in Section 3.1.

collectors). We found similar qualitative and quantitative effects in the pilot as in the main experiment (Figure A9). This suggests that the results of the main experiment were not somehow an artifact of abnormal conditions during the pandemic.[21]

Complementary evidence from household surveys Table A8 reports the treatment effects on key tax outcomes based on estimating equation (2) and using the independent household survey. Households in the treatment group are more likely to report having been visited by a tax collector, and having received a tax bill.[22] The impact on visits is statistically significant at conventional levels, but not the impact on bill delivery. At the same time, we cannot reject the null that the effect-size on bill delivery is the same in the household survey and in the daily collector reports (p-value 0.30).[23]

Households in treatment areas are more likely to report making a tax payment and report higher payments than in control areas, and the magnitudes imply a higher payment amount conditional on bill delivery. While the magnitudes of the effects on collections differ between the household surveys and the daily collector reports, we fail to reject the null hypothesis that the effects are similar across these data-sources (p-value 0.27).

Heterogeneity by collector We can leverage the random assignment of collectors to collection units to estimate the fixed effects for each collector-unit. Using the tax outcomes from the household survey, Figure A11 shows that while there is significant variation in performance between control collectors, the technology intervention appears to have increased performance at most parts of the control group distribution. For policy, this suggests that the effectiveness of technology does not seem to hinge on a particularly high or low initial level of collector capacity. For our further investigation, this result motivates our focus on mechanism channels for the average collector.

4 Potential Mechanisms For Experimental Tax Results

In this section and the following one, we frame our investigation of mechanisms as trying to explain how the technology caused a treatment effect on collections that was 4 times larger than on delivery – despite being a navigational tablet with the main purpose of improving delivery. In this section, we investigate three potential mechanisms that are

[21]Qualitatively, the pilot and main experiment both produce an effect on bill delivery that is larger in the middle of the campaign than at the end. On the quantitative side, at the end of the interventions, the impact on bills delivered was 32 percent in the pilot versus 27 percent in the main experiment; the impact on taxes collected was 79 percent in the pilot versus 103 percent in the main experiment.

[22]Table A8 also shows the household-level results are robust to removing all controls X_{hc} and to including more extensive controls (specifically income, liquidity and taxpayer awareness from Table A4).

[23]We also find that the treatment effect on deliveries increases significantly with household income (Figure A10). This is consistent with the collector learning mechanism we investigate in Section 5.

plausible based on findings from the literature: increases in tax morale among treated households; decreases in bribe payments for treated households; and, decreases in motivation among control collectors. In the following section, we focus on a mechanism that is new in the literature, which combines time use re-allocation and learning.

4.1 Tax Morale and Perceived Enforcement Capabilities

The first mechanism we consider is that technology improved households' tax morale or increased households' perceived enforcement capabilities of local government. Tax morale is broadly defined as the non-pecuniary motivations for tax compliance (Luttmer and Singhal, 2014). For instance, the presence of technology may improve households' views that the government is making efforts to collect taxes more efficiently and equitably. Household perceptions of government enforcement may also change if seeing a collector with a new technology raises their expected pecuniary costs of non-compliance.

We use our household survey to create three indices for tax morale: government efforts to collect taxes in equitable and efficient ways; satisfaction with government services; government integrity and governance capacity. We also create an index for information and enforcement, which measures households' perceptions of government informational capacity and enforcement strength. Each index is based on several individual questions, which are detailed in Data Appendix B.3.

In Panel A of Table 2, we find null effects of the technology on all four indices of household tax morale and enforcement perceptions. In Table A9, we find null effects on 12 of the 13 individual underlying questions used to build the morale and enforcement indices. For example, there is no treatment impact on the households' perception that a non-complier will end up paying, or on the share of households agreeing that government will use tax revenues wisely. The only statistically significant impact is a decrease in the perception that everyone pays their fair share of taxes, which, if anything, should be a sign of lower tax morale in the treatment group. The average null effects may mask heterogeneity along the income distribution if, for example, morale is boosted only among the well off that previously paid taxes. Yet we find no significant heterogeneous effects on any of the indices at any point in the income distribution (Figure A12).

In the learning mechanism described below, we find that households with hard-to-observe propensity to pay (measured by income, liquidity and taxpayer awareness) are more likely to pay in treatment than in control areas. Importantly, Table A10 shows there are no heterogeneous effects on the four indices by propensity to pay – suggesting that the higher payment rate in treatment areas amongst those with higher propensity to pay is not driven by an increase in morale or enforcement specifically for those households.

Table 2: Mechanism Results Based on Household Survey

Panel A: Beliefs & Morale	Satisfaction w. gov't services (1)	Integrity of gov't (2)	Equity & efficiency (3)	Enforcement & information (4)
1(Technology)	-0.007	0.062	-0.014	-0.053
	(0.070)	(0.072)	(0.060)	(0.057)
Mean in CG	0.045	-0.039	-0.033	0.004
Observations	4334	4334	4334	4334
Panel B: Bribes	**Any bribe (5)**	**Total bribe (in %) (6)**	**Coercive bribe (in %) (7)**	**Collusive bribe (in GHC) (8)**
1(Technology)	0.116***	0.025**	0.011*	6.160**
	(0.039)	(0.011)	(0.006)	(3.071)
Mean in CG	0.139	0.117	0.039	11.612
Observations	4334	4334	4334	4334
Panel C: Collector visits	**1(visit) (9)**	**Total visits (in %) (10)**	**Time per visit (in %) (11)**	
1(Technology)	0.087**	0.097**	0.242**	
	(0.033)	(0.045)	(0.116)	
Mean in CG	0.549	0.652	0.214	
Observations	4334	4334	4334	

Notes: This table presents the impacts of technology on mechanism outcomes in the household survey based on estimating (2). Panel A focuses on beliefs and morale: satisfaction with government's delivery of services; perceived integrity and competency of local government; perceived government efforts to collect taxes in an equitable and efficient manner; perceived enforcement capacity and informational knowledge of local government. Panel B focuses on bribe outcomes: a dummy which takes a value of 1 if the household estimates that the tax collector asks for an unofficial payment during visits to property owners (collusive bribe) or pockets some of the money collected from property owners (coercive bribe), and 0 otherwise; total bribe amount (in %), which is the average of the coercive bribe amount and the collusive bribe amount; coercive bribe amount (% of a hypothetical 1000 GHC); collusive bribe amount (in GHC). Panel C focuses on visit outcomes: a dummy for any visit received by a tax collector; the total number of visits (expressed in %); time spent per visit (in %). Standard errors clustered at the collector-unit are reported in parentheses. See Data Appendix B.2-B.4 for details. * p<0.10 ** p<0.05 *** p<0.01.

Finally, we leverage the fact that a subset of households were included for the first time in the tax registry for the experiment campaign. The average null effects may mask heterogeneity by extent of prior exposure to tax collectors; for example, positive impacts of technology on morale and enforcement may be concentrated amongst those that have no prior interaction with the taxation process. In Table A11, we find no significant treatment heterogeneity between newly and previously registered property owners. Finally, the null results may be due to a lack of salience: the treatment group collectors were not instructed to show the tablet to property owners to convey changes to the tax collection process; moreover, the tablet mainly helps with navigation and the issuance of a receipt for payment still has to be done manually in both treatment and control areas.[24]

4.2 Bribes

The second mechanism we consider is that technology may have improved the payment margin, conditional on delivery, by reducing bribe activity. Bribes can take the form of a "collusive bribe," where the household and collector agree on a payment made to the collector in exchange for a cessation of follow-up visits. They can also take the form of a "coercive bribe," in which the collector pockets tax payments made by the household in combination with a threat of retaliation against whistle-blowing.

The effect of technology on these two types of bribe activities is ambiguous ex-ante. On one hand, technology can reduce these bribe activities through better monitoring by supervisors, or easier reporting of bribe taking by households. On the other hand, technology may increase households' perception of collectors' enforcement capacity and raise collectors' bargaining power, which could increase bribe taking. Technology could also free up time for the collectors, which they may use to attempt to collect bribes.

We use the household survey to capture various measures of bribes which, due to their illegal and culturally sensitive nature, come from indirect questions – for example, we ask if it is likely that collectors *in the household's area* will ask for bribes. In panel B of Table 2, we find positive and statistically significant treatment effects on bribes.[25] While technology causes a meaningful increase in the likelihood of coercive or collusive bribes (column 5), the treatment effects on bribe amounts are smaller (columns 6-8) – for example, the treatment effect is over 4 times smaller on collusive bribe amount in GHC (column 8) than on tax amount paid in GHC (Table A8).[26]

[24]Our survey asked households in treatment areas if they had seen the tablet. In a household level OLS regression with collector fixed effects, having seen the tablet is not strongly associated with morale and enforcement perceptions (results not shown). Our null effects capture short run impacts; it is possible that tax morale or enforcement views may be shaped in the longer run, for example if repeated interactions with technology cause households to feel they are becoming more legible to the state (Okunogbe, 2021).

[25]One possible explanation for the positive bribe impact could be that the supervisors monitored the treatment group less than the control group. However, we find a null effect on monitoring as perceived by the collectors (Table 3). Rather, Figure A14 shows that treated households who report bribes also report higher awareness of the tax system and enforcement. Households that have witnessed government's enforcement actions may more strongly perceive collectors' threat of retaliation and therefore be amenable to paying the bribe. Consistent with the story of learning (Section 5), collectors in the treatment group may have spent time discovering which households have higher enforcement awareness, and subsequently target those households in attempts to collect bribes. We lack data on bribe incidence outside of our experiment, which limits our ability to construct predictors of bribe and pursue this hypothesis rigorously.

[26]The results are robust to different specifications (Table A8) and measures of bribe (Figure A13).

These positive impacts suggest that the larger treatment effect on tax collection than on bill delivery does not operate through a substitution away from bribe activities.

4.3 Collector Effort and Motivation

An important concern is that control collectors may have put in less effort, or felt less motivated, since they were not given access to the new technology.[27] This could explain the larger treatment effect on collection relative to delivery if control collectors' drop in activity occurred in the later parts of the campaign where more time was spent on collection rather than delivery. Four pieces of evidence suggest that this is not a significant concern in our setting.

First, Table 3 shows there are no treatment effects on collectors' self-reported job satisfaction and hours worked. The job satisfaction variable is an index, and Table A12 shows there are no treatment effects on the underlying individual questions used to create the index. Figure A15 shows that these null average effects do not mask significant heterogeneity over time. Importantly, the control group does not see a drop off in hours worked or job satisfaction over time. If anything, both hours worked and job satisfaction increase, though by small amounts, towards the end of the campaign in the control group (similar to the treatment group). These results support the interpretation that control collectors were not specifically discouraged over the course of the campaign.

Second, we leverage the fact that all collectors were randomly assigned to collection units to investigate if control collectors that worked in closer proximity to treatment collectors performed differently. For each control collector unit, we calculate the share of geographically adjacent collection units that are populated by treatment collectors. In the sample of control collectors, Table A13 shows that this variable does not cause any impact on a broad set of outcomes from the collector surveys (including job satisfaction, hours worked, strategies used, knowledge) and the daily administrative data (bills delivered and taxes collected). With the caveat that this variable is an imperfect measure for knowledge about the technology, these null results support the interpretation that the technology did not induce major artificial changes in control collectors' behavior.

Third, in addition to navigational issues that the tablet was specifically designed to alleviate, collectors face other challenges – such as wrong information printed on the bills or resistance from households to acknowledge receipt of the bill delivery. Tables 3 and A12 show there are no significant differences between treatment and control collectors in these additional challenges, which could otherwise have discouraged the control group.

[27]In the training sessions, neither group was made aware of the other group's activities, nor was it suggested that the continued implementation of technology depended on the performance during the tax campaign. Control collectors may, however, have learned about the technology in other ways.

Table 3: Collector Performance and Challenges Reported in the Field

Panel A: Performance	# unsuccessful visits per successful visit (1)	Total hours worked (2)	Hours per bill delivered (3)	Hours on non-delivery activities (4)	Fieldwork is prepared and narrowly focused (5)	Overall satisfaction in job (6)
1(Technology)	-0.933	-0.744	-0.772***	8.594*	0.041	0.103
	(1.345)	(1.848)	(0.198)	(4.721)	(0.077)	(0.158)
	[0.420]	[0.632]	[0.001]	[0.092]	[0.482]	[0.524]
Mean in CG	8.028	19.057	1.230	3.601	0.536	-0.065
Observations	141	141	141	141	141	141

Panel B: Challenges	Wrong information printed on bills (7)	Resistance from property to accept bill (8)	Supervisors do not monitor activities in the field (9)	Supervisors unavailable for support if needed (10)	Supervisors check for mistakes in the field (11)
1(Technology)	-0.265	-0.060	-0.141	-0.213	-0.078
	(0.186)	(0.163)	(0.214)	(0.213)	(0.216)
	[0.084]	[0.628]	[0.420]	[0.278]	[0.668]
Mean in CG	0.131	0.04	0.076	0.158	0.032
Observations	141	141	141	141	141

Notes: This table estimates the impacts of technology on collector outcomes, based on equation (1). All regressions include collector-unit controls and survey round fixed effects (Section 3). Panel A focuses on performance measures: # of unsuccessful visits per successful one; total hours worked; hours per bill delivered; hours spent on non-delivery activities; whether fieldwork is prepared ex ante and focused on narrow areas; satisfaction in the job. Panel B measures the extent to which the collector agrees that a certain challenge characterizes their field work: wrong information on the bills; resistance from property owners to accept the bill; and, supervisors do not monitor activities, are unavailable for support, and check for mistakes. Standard errors clustered at the collector-unit are reported in parentheses. The randomization inference based p-value is reported in brackets. Results are based on the full sample of collector surveys (results from the balanced sample are in Table A6). For details on the outcomes, see Data Appendix B.6. * p<0.10 ** p<0.05 *** p<0.01.

Fourth, supervisors were randomly assigned to groups. Supervisors may have sought to disproportionately help treatment collectors or restrain control collectors to artificially create positive impacts of technology. While it is hard to fully rule out this concern, we reassuringly find no differences between groups in collectors' reports (based on confidential surveys with our independent enumerators) of how often supervisors were available to collectors, monitored them or checked for mistakes in their work (Table 3).

5 Collector Time Allocation and Learning

This section investigates the hypothesis that collectors respond to the new technology by shifting their allocation of time away from deliveries and toward learning about and collecting from households, particularly those with higher propensity to pay. As this mechanism is novel in the literature, we begin with background observations to motivate it. We then present evidence on time allocations, learning, and collector strategies.

5.1 Navigational Constraints and Household Types

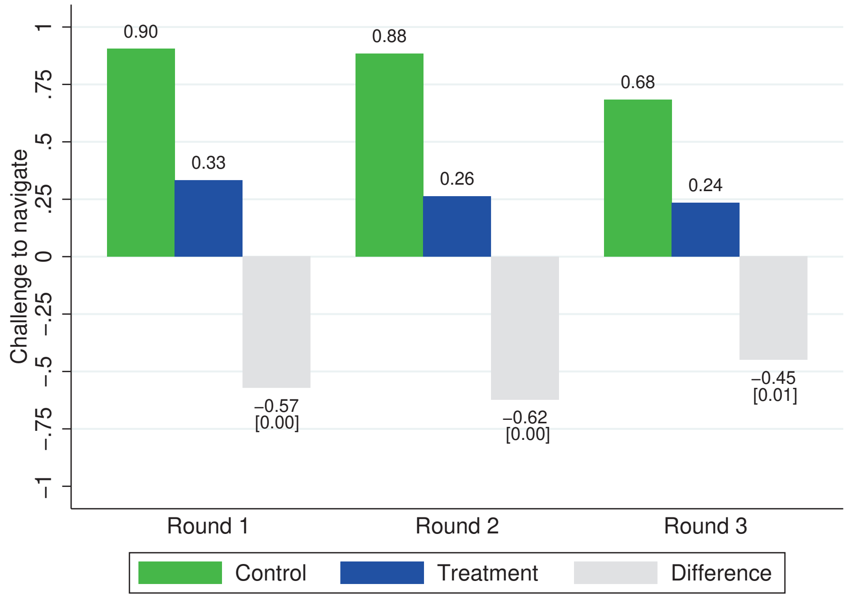
For this mechanism to help explain the results of Section 3, it has to be that collectors were initially constrained in their ability to locate property owners and in their knowledge of households' propensity. Several observations are consistent with such a setting.

Self-reported time-use data show that the average control collector would require 9 weeks to deliver all 145 assigned bills, let alone conduct follow up visits to collect payment, while the campaign lasts 6 weeks (based on Table 3 and Figure A15). Navigational challenges seem to be at the heart of the constraint on delivery: at baseline, 90 percent of control collectors find it challenging to navigate in the field, and 86 percent find it hard to locate the assigned taxpayer (Figure 2). In the control group, the size of the collection unit, measured as the time it would take to travel to all assigned properties once for delivery, is negatively associated with the actual number of bills delivered (Figure A16). These observations suggest that eased navigation may improve the delivery margin.

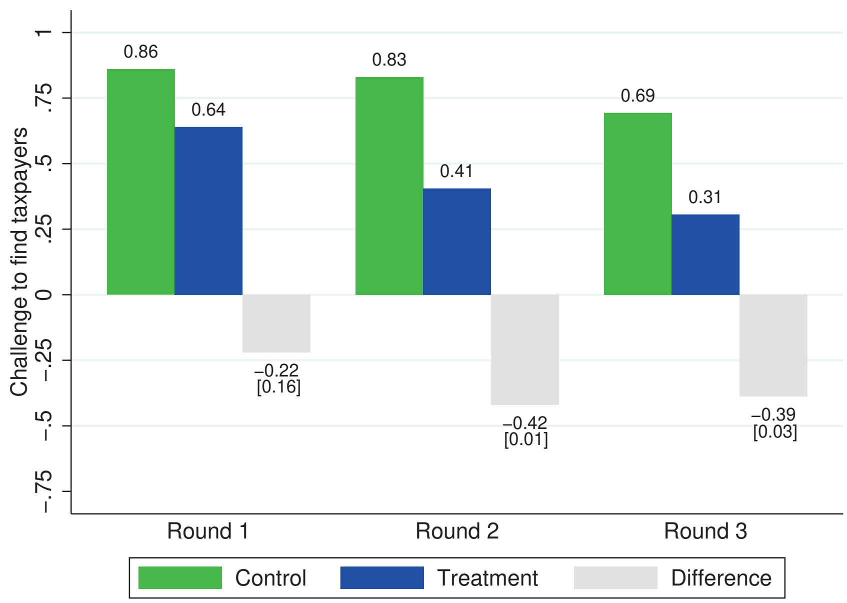
Household propensity to pay is an important determinant of tax payment when enforcement is limited. This observation was revealed in qualitative focus groups with local government officials and experienced tax collectors in our setting; it is also consistent with findings from the literature (Luttmer and Singhal, 2014; Balan et al., 2022). In our setting, propensity is determined primarily by a household's income, its liquidity and its awareness of the tax system. We use the household survey to measure propensity to pay, by combining proxies for income, liquidity and awareness into an index at the household level (Data Appendix B.5). We label propensity to pay the 'hard to observe' index because these household characteristics are hard to directly observe. This index strongly predicts actual tax payment in a sample outside of our experiment (Table A14).

Figure 2: Challenges to Navigate and Locate Taxpayers

(a) Challenges to Navigate in the Field



(b) Challenges to Locate Taxpayers



Notes: This figure shows navigation and localization challenges, based on equation (1). Round 1, 2 and 3 correspond to the baseline, mid-line and endline collector survey rounds, respectively. In panel A, the outcome equals 1 if a collector reports finding it challenging or very challenging to navigate in the field. In panel B, the outcome equals 1 if a collector reports finding it challenging or very challenging to locate an assigned taxpayer, and 0 otherwise. The grey bar measures the difference in outcome between treatment and control; the number in brackets is the randomization inference-based p-value on the statistical significance of the difference. For a description of the challenge measures, see Section 5 and Data Appendix B.6. The analysis is based on the balanced sample of collector surveys (Section 3.1). Results based on the unbalanced sample are in Table A6.

Household propensity to pay is poorly correlated with more easily observable characteristics. We construct an 'easy to observe' index, which combines the following characteristics: amount of taxes due; taxes paid in the past; easily observable assets (e.g. ownership of a car); distance to main roads and markets. The first two are directly observable on the tax bill (Figure A2); the last two can arguably be inferred easily in the field. Importantly, the easy to observe index *is* positively associated with compliance outside of our experiment, but it is a less strong predictor than the hard to observe index (Table A14). Within a collection unit, the amount of taxes due and the observable index account for only 1 percent and 4 percent, respectively, of the variation across households in the hard to observe index.[28] While it is ultimately a strong predictor of tax payment, propensity to pay is hard to observe, heterogeneous across households and weakly correlated with more easily observable characteristics. Reflecting these features, at baseline 88 percent of control collectors report not having a good understanding of which households in their assigned area have a higher propensity to pay (Figure 3).

5.2 Treatment Effects on Collector Time Allocations

From the outset of the campaign, treatment collectors report less challenges than control collectors to navigate and to locate the assigned taxpayer (Figure 2). Both groups experience reductions in challenges over time but technology's treatment effects are substantial throughout the campaign and, despite the small sample size, significant in almost all survey rounds. These improvements are not mechanical, as the tablet does not automate navigation (Section 3.1). Instead, the collector has to actively apply effort to make use of the tablet's self-localization and assigned property localization features; by continuously switching back and forth between the visual information in the field and the tablet's map, the collector makes effortful and deliberate choices that improve navigation.

The positive effect of technology on bill delivery shows up in significant reductions in time spent per bill delivered. In particular, the treatment group spends 62 percent less time per bill delivered than the control group (Table 3). This reflects technology's intended impact, namely to enhance delivery in a setting with incomplete addressing.

Technology leads to significant changes in collector time allocation. In the average

[28] For taxes due, the weak correlation arises because the property tax in Madina is based on a presumptive schedule rather than property valuation. In the presumptive tax, coarse proxies for capital value (e.g. number of floors and area-size) are used to calculate taxes owed, which weakens the link between property tax due and 'true' property value, let alone household wealth or income. Our census data reveals that less than 20% of local governments have any property valuations in place. Feasibly implementing property valuation methods requires continuously updated property and market information from third-parties such as banks and mortgage providers, which have limited coverage in Ghana.

week, collector surveys reveal that the treatment group spends two and a half times more hours on non-delivery activities than the control group, even though treatment collectors work the same total number of hours as control collectors (Table 3). Thus, treatment collectors allocate a larger share of their time away from delivery to other activities than control collectors.[29]

Household surveys show that treatment collectors spend more time engaging with households. Indeed, households in treatment areas report receiving significantly more visits by collectors (Table 2). Moreover, the duration of each visit lasts longer, on average, in treatment areas.

5.3 Treatment Effects on Learning and Collection Strategies

On which activities do the treatment collectors spend more time, as they devote less time to delivering bills? In this subsection we present evidence that they spend time learning about households' hard-to-observe type and using collection strategies that leverage the new information. By learning, we mean that the household's propensity to pay is initially not observed but the collector can discover the household's type through a time-costly process of engaging with the household.

Collectors can learn in various ways based on interactions with the household (which Table 2 revealed were more frequent and longer in treatment areas). The collector can ask questions that directly or indirectly relate to the property owner's liquidity, income and taxpayer awareness. Moreover, the collector can infer propensity to pay based on the property owner's actions during repeated interactions (e.g. willingness to schedule a follow-up appointment may signal higher propensity). Finally, the collector can spend time 'surveying' the property and local area to notice less obvious clues, and talking to neighbors and community members. While a household may understandably be reluctant to reveal their type directly to a government official, collectors can learn in multiple ways based on interacting with the household and the environment.

This form of learning by interacting takes place prior to attempting to collect, and would be referred to as 'active learning' in the literature (Thompson, 2010). In principle, learning could also occur 'passively' if the collector, through repeated attempts to collect, uncovers the spatial clustering of types and forms beliefs about the household's type based on the immediate area's payment rate. We explore this possibility in Appendix

[29] Measuring the number of failed attempts per successful visit, Panel A in Figure A17 shows that the treatment group switches over time into more demanding tasks with a higher failure rate. The switch over time into more demanding tasks may reflect that, as technology eases the delivery constraint, collectors in the treatment group re-allocate time to activities with relatively more binding constraints.

D.1 and find limited limited evidence that passive learning is important in our setting.[30]

Evidence from collector surveys The collector surveys show an increase over time in treatment collectors' self-reported knowledge about households' propensity to pay. Panel A of Figure 3 shows that there were initially no differences across groups in how well collectors knew which households had higher propensity to pay.[31] Over time, a positive knowledge gap opens up, which is statistically significant at 5% by the end of the campaign. The knowledge gap is driven by the six-fold, statistically significant increase in knowledge in the treatment group; in comparison, the control group sees a muted and statistically insignificant increase. Table 4 provides regression results. In column 2, we leverage the panel nature of the collector surveys and include collector fixed effects, finding an even stronger impact of technology on knowledge. As the fixed effects isolate the treatment impact that varies within collector over time (relative to any initial impact at the beginning of the experiment), this result is consistent with knowledge that builds over time as the collector discovers households' types through repeated interactions.

Mirroring the knowledge result in Panel A, Panel B of Figure 3 shows there were initially no significant differences between groups in the extent to which collectors report making use of collection strategies that focus on hard to observe characteristics, but over time, treatment collectors increasingly make use of strategies to visit specific households where they have identified higher propensity to pay.[32] Indeed, at the end of the campaign, the treatment group is more likely than the control group to: go to areas on specific days where they know property owners are more likely to be able to pay; go to properties where they know taxpayers are aware of their duty to pay; and, go to properties where owners are more satisfied with public services and willing to pay.

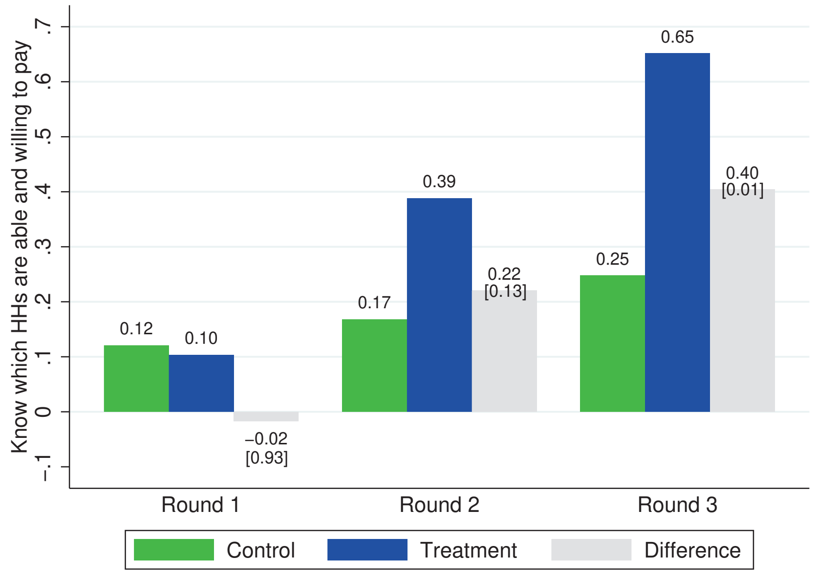
[30] Specifically, the *level* of payment rate in cold spots is lower in treatment areas than in control areas (Figure D2). Cold spots are geographical clusters of households with a high concentration of low values of the hard to observe index. Since passive learning is based on inferring ex post based on observed payment rates and treatment collectors achieve a higher payment rate in general, the lower payment rate in cold spots is inconsistent with this form of learning. The lower payment rate in cold spots is strongly consistent with active, ex ante learning about households' types which precedes any attempt to collect.

[31] Knowledge is a dummy variable which takes a value of 1 if the collector chooses the statement "I think I have a good understanding of which properties are more able and willing to pay" rather than the statement "I put a lot of effort to get my job done, but it remains unclear to me which exact properties are more likely or more willing to pay their property rates". See Data Appendix B.6 for more details.

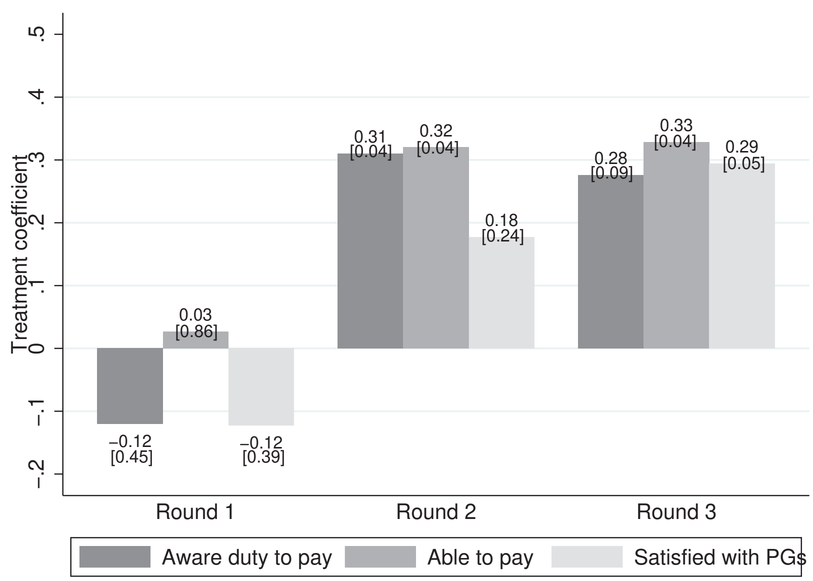
[32] Strategy use takes a value of 1 if the collector reports using 'all the time' or 'often' the collection strategy which focuses on this household characteristic, and 0 otherwise (Data Appendix B.6).

Figure 3: Collector Knowledge and Collection Focus

(a) Knowledge About Which Households are Able and Willing to Pay?



(b) Collection Focus on Households That are Able and Willing to Pay?



Notes: These panels show collector knowledge (panel A) and strategies (panel B), based on equation (1). Round 1, 2 and 3 correspond to the baseline, mid-line and endline collector survey rounds, respectively. In panel A, the outcome takes a value of 1 if the collector reports having a good understanding of which properties are more able and willing to pay (and 0 otherwise). Panel B shows the estimated difference between groups in three collection strategies: focus on property owners that are more aware of their duty to pay; focus on properties on specific days where the owners are more likely to be able to pay; focus on property owners that are more satisfied with the local public goods. Collection strategy takes a value of 1 if the collector uses the specific strategy all the time or often (and 0 otherwise). The grey bar measures the difference in outcome between treatment and control, with the the randomization inference-based p-value in brackets. See Data Appendix B.6 for details on the variables.

One important question is the extent to which this increase in collector strategy reflects the fact that the treatment group allocates more time to collection activities in general. To investigate this, Table 4 studies the impacts of technology on how frequently both hard-to-observe strategies and easy-to-observe strategies are used. The hard to observe collection strategies focus on the characteristics that determine payment propensity (Panel B of Figure 3) and which make up the hard to observe index (Section 5.1); the easy to observe strategies focus on the characteristics that make up the easy to observe index (Section 5.1).[33] The table shows no initial differences in strategy and, over time, treatment collectors make more frequent use than the control group of both types of strategy. At the same time, the treatment group makes disproportionately more use of hard to observe strategies relative to easy to observe strategies. The disproportionate reliance on hard to observe strategies holds with collector fixed effects – consistent with a time-varying change in strategy within collector, such as the use of gradually acquired knowledge that informs over time who to target for collection.

Evidence from household surveys These results are consistent with learning in the field, but they are based on self-reports by collectors. We therefore undertake additional analyses based on the independent household surveys. Our first exercise investigates how characteristics of paying households differ from non-paying ones within a collection unit and how technology causes this to differ between treatment and control areas:

$$y_{hc} = \theta \cdot \mathbf{1}(Pay)_h + \beta \cdot [\mathbf{1}(Pay)_h * \mathbf{1}(Tech)_c] + \Omega \cdot X_h + \mu_c + \epsilon_{hc} \tag{3}$$

where y_{hc} is a fixed household/property characteristic and $\mathbf{1}(Pay)_h$ is a dummy for making any positive tax payment. Since $\mathbf{1}(Pay)_h$ is endogenous, θ indicates whether there is a statistical (non-identified) difference in a fixed characteristic between paying and non-paying households in the control group. The treatment coefficient β shows how the difference in characteristic between paying and non-paying households causally changes in treatment versus control areas; any non-zero β would indicate that technology causes selection in the characteristics of who pays. We can include collection area fixed effects (μ_c) since we focus on differences in characteristics between households within collection areas. Standard errors are clustered by collector-unit.

[33] There are small differences in the characteristics included in the strategy variables versus in the indices. Specifically, the hard-to-observe strategy variable but not the index includes 'satisfaction with public goods' as a determinant of payment propensity. The easy-to-observe index but not the strategy variable includes assets as a characteristic. These differences arise from variation in content between the household and the collector surveys; results are unchanged if we limit the characteristics to be exactly the same in the indices and strategy variables.

Table 4: Impacts of Technology on Collector Knowledge and Strategies

	Knowledge of hard-to-observe household characteristics		Focus on collections, hard-to-observe household characteristics		Focus on collections, easy-to-observe household characteristics		Difference in strategies: Hard versus easy to observe characteristics	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Average effect								
1(Technology)	0.187**	0.610**	0.140*	0.667***	0.088	0.445**	0.051	0.223**
	(0.091)	(0.252)	(0.078)	(0.252)	(0.058)	(0.218)	(0.045)	(0.093)
	[0.030]	[0.003]	[0.070]	[0.000]	[0.068]	[0.049]	[0.238]	[0.001]
Panel B: Dynamic effects								
1(Technology) × **1**(Round 1)	-0.001	–	-0.082	–	0.008	–	-0.091	–
	(0.131)	–	(0.104)	–	(0.069)	–	(0.079)	–
1(Technology) × **1**(Round 2)	0.211*	0.273*	0.264**	0.364**	0.123	0.137	0.140**	0.227**
	(0.121)	(0.156)	(0.126)	(0.153)	(0.093)	(0.101)	(0.066)	(0.095)
1(Technology) × **1**(Round 3)	0.369***	0.401**	0.253**	0.349**	0.138	0.130	0.114*	0.219**
	(0.135)	(0.167)	(0.119)	(0.151)	(0.089)	(0.106)	(0.064)	(0.093)
Collector-unit controls	X		X		X		X	
Survey round FE	X	X	X	X	X	X	X	X
Collector-unit FE		X		X		X		X
Mean in CG	0.195	0.195	0.280	0.280	0.239	0.239	0.041	0.041
Observations	141	141	141	141	141	141	141	141

Notes: This table presents the impacts of technology on collector knowledge and strategies, based on equation (1). In columns (1)-(2), the outcome takes a value of 1 if the collector reports having a good understanding of which properties are more able and willing to pay and 0 otherwise. In columns (3)-(4), the outcome is the likelihood that a collector makes uses all the time or often of collection strategies which focus on hard-to-observe household characteristics (taxpayer awareness, ability to pay, satisfaction with public goods). In columns (5)-(6), the outcome is the likelihood that a collector makes use often or all the time of collection strategies which focus on more easily observable household characteristics (value of tax bill, past tax payment, geographical location). In column (7)-(8), the outcome is the difference between the reliance on hard-to-observe versus easy-to-observe strategies. For more details, see Data Appendix B.6. Standard errors clustered at the collector-unit are reported in parentheses. In Panel A, the randomization inference based p-value is reported in brackets. * p<0.10 ** p<0.05 *** p<0.01.

Figure 4 investigates selection in payment patterns for the (fixed) household characteristics of the hard to observe index (income, liquidity, and tax awareness) and easy to observe index (bill value, previous tax payment, assets, distance to main road and markets) constructed in Section 5.1.[34] We find that technology causes a large, positive selection on all components of the hard to observe index. While the index is positively associated with payment in all areas, technology causes the association to be significantly larger in treatment areas. This positive selection on the hard to observe characteristics is consistent with the independent result from the collector surveys where treatment collectors reported focusing more on households with these hard to observe characteristics.[35] These results are robust to various definitions of tax payment (Figure A21).

Data in the control group shows that the likelihood of tax payment and amount paid hardly vary with values of the hard-to-observe index until the top percentiles where they both spike (Figure A22). We capture this by creating a 'high type' indicator, which equals 1 for values of the hard-to-observe index above the 95^{th} percentile and 0 otherwise. The selection results hold with this discrete formulation of the index (Figure A23).

One challenge with the selection exercise in (3) is that it investigates differential payment rates by household characteristics in a setting where the household payment rate is larger in general in treatment areas. We therefore provide a second selection exercise that measures the composition of taxpayers in the collector's payments (and controls for the payment rate). We use the discrete version of the hard-to-observe index and measure, for each collector at the end of the campaign, the share of collected payments from high-types. Panel A of Figure 5 shows that this selection measure is over twice as large in the treatment group (28.7 percent) than in the control group (12.6 percent) and the difference is significant. The result is robust to different payment measures (Panel B). Technology's positive impact on this second selection measure is consistent with learning, as treatment collectors discover and focus their collection efforts on the high-types.[36]

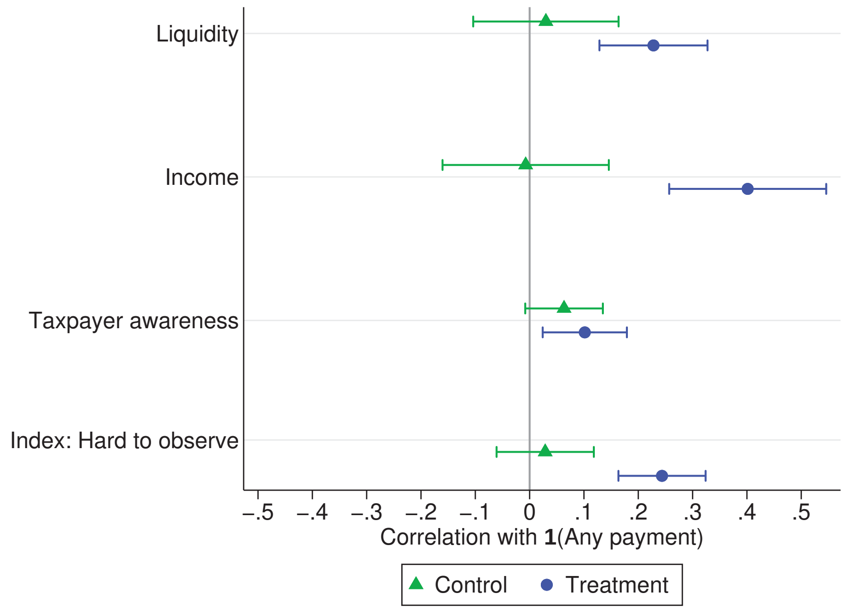
[34]See Data Appendix B.5. Even though these proxies are based on end-line household surveys, we think they are plausibly not impacted by the treatment. It is unlikely that technology-induced payment of taxes affects households' income choices within the six-week span of the tax campaign. The questions on liquidity refer to a 'typical' month rather than the specific past month during the campaign. Finally, no property owner from the areas of the experiment was neither summoned to court nor had their property confiscated during the tax campaign, which could otherwise have raised taxation awareness.

[35]Due to the increased focus on high income households, Figure A18 shows that the tax system with technology becomes more progressive. However, bribe payments also become more regressive in treatment areas. Re-designing the technology to minimize the bribe impact is an important next step in research.

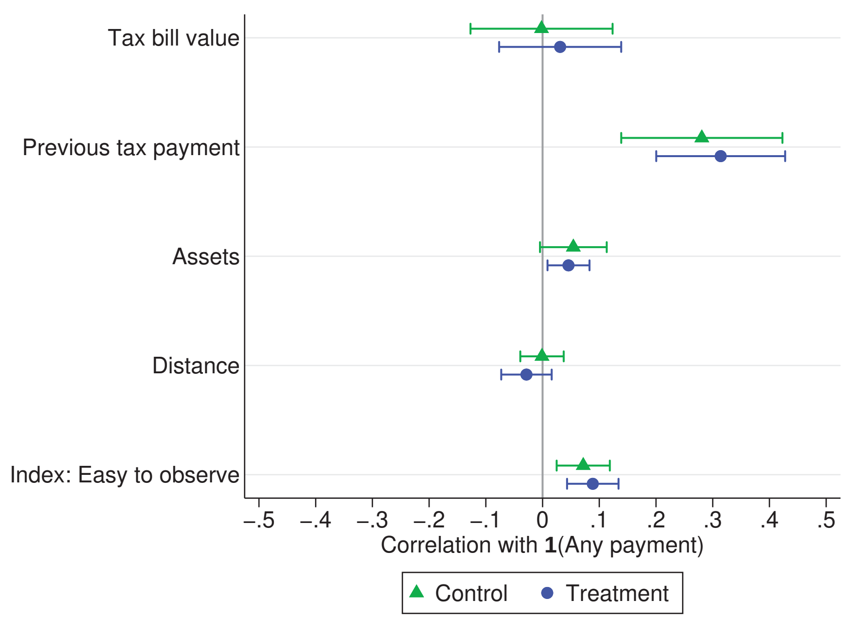
[36]We have focused on a setting where the household type is hard-to-observe but discoverable. This is consistent with the finding of limited initial knowledge that grows over time (Figure 3), though that result is based on collectors' self-reports. In Appendix D.2, we investigate the possibility that the household type is either perfectly unobservable or perfectly observable to all collectors. We find no support to suggest that the full set of household survey results can be reconciled with either of these two alternative settings.

Figure 4: Characteristics of Households That Made a Tax Payment by Treatment Status

(a) Hard-to-Observe Characteristics



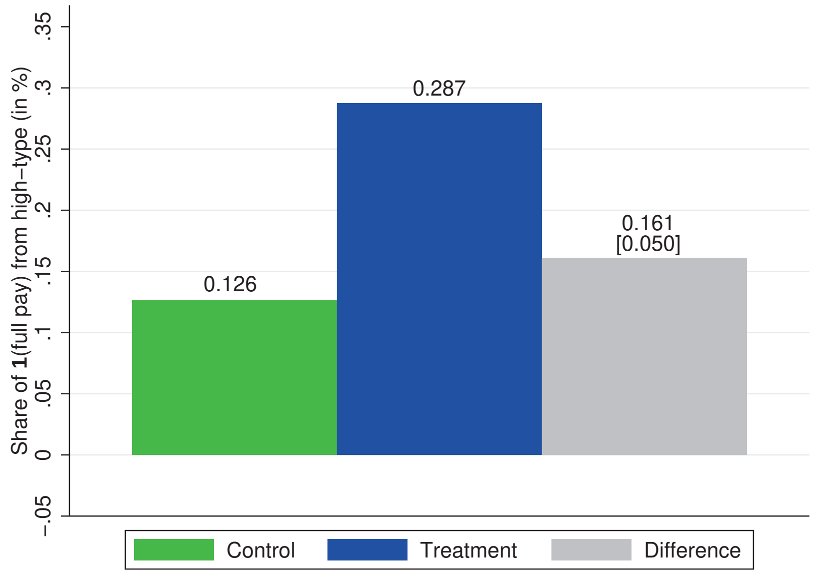
(b) Easy-to-Observe Characteristics



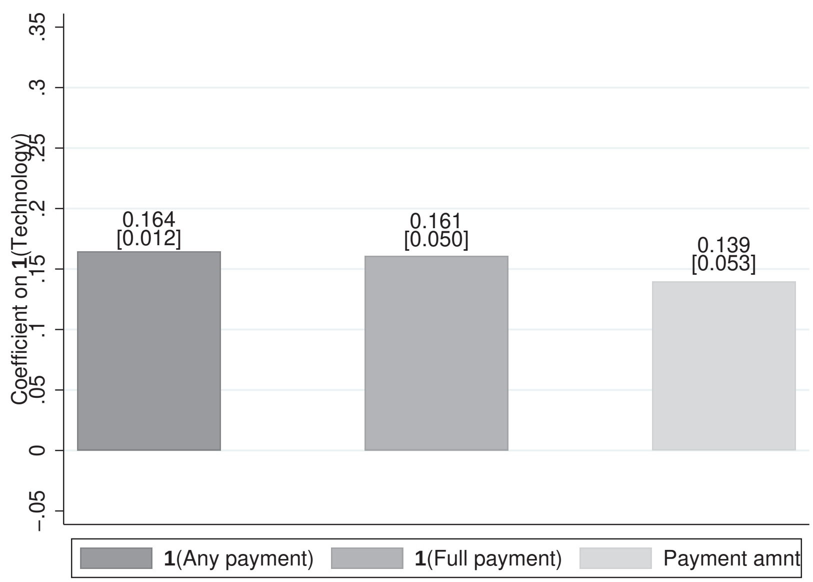
Notes: These figures show targeting of property owner characteristics for tax payment, based on estimating (3). Coefficients together with the 95% confidence intervals are displayed. Standard errors are clustered at the collector-unit level. Characteristics vary by row and the bottom row of each graph is an index which is the unweighted average of the characteristics. The characteristics in Panel A are harder to observe, while the characteristics in Panel B are easier to observe (see Section 5 and Data Appendix B.5).

Figure 5: Composition of Payments Collected

(a) Share of Payments From Hard-to-Observe High-Type



(b) Robustness: Different Measures of Tax Payment



Notes: These graphs show the impact of technology on the composition of tax payments, based on equation (1) estimated in a single cross-section of collectors at the end of the campaign. In all graphs, the outcome variable is the share of all payments in a collection unit that are collected from the top 5% of the hard-to-observe index distribution. In panel A, the payment variable is a dummy which takes a value of 1 when the household makes a full payment (and 0 otherwise); panel A shows the level of this variable in the treatment and control groups, as well as the difference between groups. Panel B shows the difference between groups for three payment variables: a dummy for making any payment; a dummy for making a full payment; total amount paid. The number in brackets is the randomization inference-based p-value on the statistical significance of each difference. See Data Appendix B.2 for details on the variables.

6 Model

In this section, we formalize our theory of why a technology designed to assist in navigation for the purposes of bill delivery ended up having a substantially larger effect on collections than on deliveries. We do so in a dynamic Beckerian model of time use by forward looking revenue collectors. We then estimate the model to match moments of the experimental data, evaluate its predictions for non-targeted moments, and illustrate our theory by simulating counterfactual scenarios.

6.1 Model Environment

Collectors are endowed with one unit of time each period, and the campaign lasts a total of T periods. Collectors spend their time delivering bills, learning about households to whom they've delivered bills, and trying to collect payments. Each collector is endowed with a large number of bills in the initial period, and every bill has a face value of one local currency unit. Neither the exact initial number of bills or variation in value across bills play an important role in the experiment, so we abstract from these in the model.

Collectors come in two types: treatment (T) and control (C). For each unit of time devoted to delivering bills, treatment collectors can distribute θ_T bills, and control collectors can distribute θ_C bills. We assume that $\theta_T \geq \theta_C$, motivated by the technology's reduction in navigational challenges (Figure 2) and time spent to deliver a bill (Table 3).

Households also come in two types: "high" and "low," referring to their propensity to pay conditional on bill delivery. A fraction μ of households are high types, and μ is known to the collectors. Collectors do not know which type each household is at the outset of the campaign. To discover a household's type, we assume that collectors must spend additional time learning about them after their bill has been delivered. This learning time is meant to capture the follow-up visits needed to interact with the household and become informed about their ability and willingness to pay (see Section 5.1 and 5.3). Our discrete formulation of the household hard-to-observe type is motivated by the empirical finding that the returns to learning about hard to observe characteristics appear to be limited to the discovery of a small set of high-types (Figure A22).

For each unit of time devoted to learning, treatment collectors have a probability η_T of discovering the household's true type, and control collectors have learning probability η_C. We assume that $\eta_T \geq \eta_C$. There are several ways to motivate this learning advantage based on the GIS-tablet. For the purpose of our benchmark set-up, perhaps the simplest way to interpret the learning advantage is to consider that the navigational advantages of the tablet also apply when trying to locate households for the follow-up interactions that are required to learn. Consistent with this premise, in our benchmark we choose to fix the learning advantage to be equal to the delivery advantage; we revisit this choice, which may be conservative, at the end of the section. Conditional on learning the household's type, the collector learns that the household is a high type with probability μ and a low type with probability $1 - \mu$. By modeling learning as a choice, we allow for the possibility that treatment collectors spend more or less time learning about household types than the control group. It is not obvious ex ante whether technology increases or decreases time spent learning in the model.

The collection technology in the model is exactly the same for treatment and control collectors. Each period, collectors can devote time to collecting from households to whom they have delivered a bill. Collectors can target their time toward households whose type is unknown, households known to be high types, and households known to be low types. For each unit of time spent trying to collect from high type households, the probability of collection is π_h. The collection probability per unit of time spent collecting from low type households is π_ℓ, where $\pi_h > \pi_\ell$. We let $\pi = \mu\pi + (1 - \mu)(1 - \pi)$ denote the probability of collection from unknown types.

Our assumption that the technology does not offer direct advantages in collection is based on two observations. The first is that the GIS-technology we randomized did not offer any direct or specific assistance in collecting from households once their bills had been delivered. The second is that the experiment showed a null treatment impact on household tax morale and enforcement perceptions, suggesting that households did not change their payment propensities directly by being visited by a collector with the tablet. Still, the assumption of no collection advantages may be restrictive in the sense that the same navigational advantages offered by the technology may well increase the return to time spent trying to collect as well. Our approach is to be as conservative as possible, and to see whether the model can generate a much larger treatment effect on collection than on delivery without any direct advantage in collections.

A collector's state variables each period are: b, the number of bills delivered to households of unknown type; b_h, the number of bills delivered to known high-types; and b_ℓ, the number of bills delivered to known low-types. Their choice variables are time spent distributing bills, d, time spent learning about household types, x, and time spent trying to collect from unknown, high-type and low-type households: c, c_h, and c_ℓ.

The collector's dynamic problem is to maximize total tax collections. Their dynamic program can be written as:

$$V(b_h, b, b_\ell, t) = \max_{\{d,x,c_h,c,c_\ell\}} \{c\pi + c_h\pi_h + c_\ell\pi_\ell + V(b'_h, b', b'_\ell, t+1)\}$$

subject to:

$d + x + c + c_h + c_\ell \leq 1$ (the time constraint),
$c \leq b$, $c_h \leq b_h$, $c_\ell \leq b_\ell$ (the collection constraints),
$b' = b + \theta_j d - x\eta_j - c\pi$ (the law of motion for unknown-type bills delivered).
$b'_h = b_h + x\mu\eta_j - c_h\pi_h$ (the law of motion for high-type bills delivered), and
$b'_\ell = b_\ell + x(1-\mu)\eta_j - c_\ell\pi_\ell$ (the law of motion for low-type bills delivered).

The time constraint requires that the total time spent on deliveries, learning, and collection does not exceed the time endowment. The collection constraints ensure that collectors cannot collect from more than their current stock of bills. The laws of motion for unknown-type bills govern how the stock evolves given inflows of new bills delivered and outflows of bills for which the collector learns the true type or collects upon. The laws of motion for high-type and low-type bills feature inflows from learning and outflows from collection. There is also a terminal period – omitted for brevity – in which collectors use all their time endowment trying to collect from their current stock of bills.

The dynamic tradeoffs for a collector can be summarized as follows. The benefit of time spent on deliveries is that it increases the stock of unknown-type delivered bills that can later be collected upon. The opportunity cost of additional delivery time is not collecting from the current stock of bills already delivered. Learning time also carries this opportunity cost. But it can help identify which bills are the high-type.

Conditional on time being allocated to collections, collectors have to decide how much time to devote to collecting from unknown-, high-, and low-types. It is easy to see that the optimal collection strategy involves first trying to collect from the high-type households, since they have the highest chances of paying. Any remaining time gets spent trying to collect from the unknown types, followed by the low types, which have the lowest propensity to pay. One can see then that learning time is valuable because it can help favorably shape the stock of bills, allowing collectors to focus more on households with higher propensities to pay.

6.2 Model Estimation

We estimate the model assuming that each period represents two days, meaning that $T = 21$. This choice allows the model to capture the rich dynamics of the collectors'

Table 5: Model Estimation

Panel A: Moments Targeted in the Estimation

Moments	Target	Model
Treatment Effect on Bills Delivered (%)	27.0	27.3
Ratio of Learning Time to Collection Time (Average)	15.0	15.8
Probability of Full Payment | Delivery (Average)	13.7	13.6
Fraction of Collections from High Type (Average)	21.9	22.1

Panel B: Estimated Parameters and Confidence Intervals

θ_T	η_C	π_ℓ	π_h
0.104	0.383	0.008	0.032
(0.077, 0.206)	(0.313, 0.501)	(0.007, 0.012)	(0.013, 0.046)

Note: Panel A reports the moments targeted in the estimation and their values in the data and in the model. Panel B reports the estimated parameter values and their bootstrapped 95-percent confidence intervals.

problem while still allowing us to estimate the model in a reasonable amount of time. We set θ_C, the delivery parameter in the control group, to be 0.05. This choice amounts to a normalization on the units of bills in the model, and has no substantive impact on any of our results. We set the fraction of high types, μ, to be 0.05, consistent with the empirical observation that the higher likelihood of payment is concentrated amongst the households with the highest 5 percent of the hard-to-observe index (Figure A22).

We then estimate the model to match four target moments, which we list in Table 5, Panel A. The first is the treatment effect on deliveries of 27 percent. This is the only experimental outcome we target directly. The remaining moments represent averages across all collectors in the experiment. These are: a ratio of learning time to collection time of 15 percent; the probability of a collector getting a full payment during the experiment conditional on bill delivery, which is only 13.7 percent; and, the fraction of collections coming from the high type, which is 21.9 percent. The learning time target is not directly observed, but consistent with a small but positive amount of learning on average (Figure 3). We have experimented with alternative targets but find that our results are not particularly sensitive to other values. The probability of getting a full payment is observed directly in our data, and implies low overall payment probabilities by the households. The fraction of collections coming from the high-type households is much larger than the fraction of households that are high types, which the model will attribute to collector learning plus a higher payment probability by the high type.

Our estimation strategy formally solves for the parameter vector $\{\theta_T, \eta_C, \pi_h, \pi_\ell\}$ that minimizes the sum of squared differences between these four moments and their counterparts in the model. We impose in the estimation that $\eta_T/\eta_C = \theta_T/\theta_C$, which implies that the treatment group's learning advantage is proportional to its delivery advantage as we explained above.

The estimated parameter values are reported in Table 5, Panel B, along with bootstrapped 95 percent confidence intervals. The value of θ_T is 0.104 implying that bill delivery probability per unit of time is just over twice as high with the technology as without it. Notice that this is substantially larger than what one might naively assume about the delivery advantages offered by the technology based on the experiment's 27 percent treatment effect on deliveries. The intuition is that collectors in the model respond to the technology by substantially reducing their time spent on deliveries, consistent with the empirical evidence in Section 5. Though the confidence interval for θ_T is fairly wide (stemming from the confidence interval around the delivery treatment effect), it does not include the value of return on time spent delivering in the control group $\theta_C = 0.05$. In magnitudes, the estimate of θ_T is consistent with the experimental evidence that the treatment collectors spend less than half as much time delivering each bill than control collectors (Table 3).

The estimated value of η_C is 0.383, and implies that a control collector spending all of their time in a period trying to learn about the household types for a given set of delivered bills results in them learning about just under two in five of them. The resulting value of η_T is 0.801, meaning a treatment collector trying to learn about unknown type bills will do so for around four in five. The estimated values of π_ℓ and π_h are 0.008 and 0.032. Overall, these low values imply that efforts to collect from households for whom a bill has been delivered are quite unlikely to lead to a payment. This has to be the case in order for the model to match the observed overall low payment rate of 13.7 percent conditional on delivery. The confidence intervals for the two probabilities do not overlap, indicating that the estimated model features two distinct household types. The probability of a payment conditional on a per unit of collection time is about four times as high for high-types as low-types. This highlights the potential returns to time spent learning in the model: if a collector knows the households' types, they can focus on the high-types and enjoy a substantially higher chance of collecting payment.

6.3 Quantitative Predictions of Model

The top two panels of Figure 6 plot the model's predicted deliveries and collections by group and by day. The treatment effect on deliveries of 27 percent, which we target in the estimation, is represented by the difference in deliveries between the two groups on the last day of the campaign. Before that, the model matches well the concave pattern of deliveries in the experiment and the peak effect on deliveries in the middle of the campaign (Figure 1). The model also matches well the convex pattern of collections in the experiment. By the end of the campaign, the model predicts – without targeting – a 77 percent increase in collections, which is around three times as high as the predicted effect on deliveries (compared to four times in the experiment). In differences, the model predicts treatment effects that are 50 percentage points higher for collections than for deliveries, compared to 76 percentage points in the experiment, meaning that the model can explain around two-thirds of the experimental difference (50/76).

The middle two panels of Figure 6 plot the allocations of time for the control and treatment collectors in the model, which are useful for understanding how the behavior of the two groups differs. In both groups, deliveries (the dark shaded area) starts out as the main activity and declines during the campaign. Yet delivery time declines more quickly in the treatment group, meaning an earlier and steeper drop in the dark shaded region. This highlights a key behavioral change resulting from technology. Because the treatment collectors can deliver more bills in the same amount of time, they choose to spend less time delivering. What do the treatment collectors do instead of bill deliveries? As the panels show, they spend more time collecting (lightest shaded area) and also more time learning (medium shade areas). The control group concentrates the little amount of learning it chooses to do late in the campaign, around day 30, whereas the treatment group learns earlier, starting around day 7, and more often throughout the campaign.

The bottom two panels of Figure 6 plot the stocks of bills by type in each period in the control and treatment groups. In both groups, the stock of bills delivered grows over time and initially consists only of unknown types. In the control group, once the collectors engage in some learning, the stock of unknown-type bills drops and are replaced with known low- and high-types. In the treatment group, the earlier and more consistent learning results in a much greater share of known types in the stock of bills delivered. At the end of the campaign, treatment collectors know the type of 61 percent of households in their unit, while control collectors have only learned the type for 19

Figure 6: Predictions of Estimated Model

(a) Bills delivered

(b) Revenue Collections

(c) Time Allocation (Control)

(d) Time Allocation (Treatment)

(e) Composition of Bills Delivered (Control)

(f) Composition of Bills Delivered (Treatment)

percent.[37] This mimics our empirical result that treatment collectors report being far more knowledgeable about their assigned households' type by the end of the actual experiment (Figure 3).

The last two columns of Table 6 report the model's predictions for the fraction of

[37]We calculate this by supposing that the measure of bills delivered in each group by the end of the model corresponds to the actual average number of bills delivered in each group by the end of the experiment (Figure 1). This presumption is reasonable as the model targets the experimental effect on delivery.

Table 6: Model Predictions versus Experimental Data

	Treatment Effects (%)		Collections, High Type (%)	
	Deliveries	Collections	Control	Treatment
Experimental data	27	103	13	28
Estimated main model	27	77	19	26
No re-optimization	52	46	19	18
Easier collection (π_h, π_ℓ x 10)	58	76	15	23
No learning advantage ($\eta_T = \eta_C$)	33	64	19	19
No delivery advantage ($\theta_T = \theta_C$)	-2	8	19	24

Note: This table reports the treatment effects on deliveries and collections in the data and in the model, and the fraction of collections that comes from the high type in the control and treatment groups.

collected payments coming from high-type households. This corresponds to the second selection moment from the experiment, described in Section 5 and plotted in Figure 5. The estimation strategy targets the selection on average but not by group, and can thus serve as a validation exercise on our model. The first row reproduces the positive experimental selection result: only 13 percent of collections in the control group came from the high-types, compared to 28 percent in the treatment group. As in the experiment, our benchmark model generates positive selection with a larger fraction of high-type collections in the treatment group (26 percent) than the control group (19 percent). In the model, this positive selection arises from the knowledge gained by treatment collectors.

6.4 Insights Based on Counterfactual Simulations

Direct and indirect effects of technology One value of the model is to quantify the direct effects of the new technology, putting aside any indirect effects stemming from changes in collector behavior in response to technology. To do so, we conduct a counterfactual exercise in which the treatment group has the same advantages in delivery and learning as in the main estimation, but which keeps the treatment's groups *decision rules* the same as the control group. In other words, this counterfactual simulates an "envelope theorem" type scenario where the technology is introduced but collector re-optimization is ignored. The third row of Table 6 shows that this counterfactual predicts treatment effects on deliveries and collections of similar magnitude (52 percent and 46 percent). Since the treatment effect on collection in the main estimation is 77 percent, we can conclude that collector re-optimization in response to the technology amplifies the direct effects of the technology by 67 percent (77/46). Similarly, collector re-optimization scales down the impact on deliveries roughly in half (27/52). Thus, an envelope theorem logic, where the technology simply scales up the performance of control collectors, does not give an accurate representation of the technology's actual experimental impacts. This model insight is consistent with additional results from the experiment which suggest that the treatment group engages in a set of activities that differ from a 'scaled up version' of the control collector's path of activities.[38]

In this counterfactual, despite being endowed with a learning advantage, the treatment collector only learns the type for 23 percent of households - a muted knowledge impact compared to the control collector (19 percent), and a much smaller impact compared to the benchmark treatment collector (61 percent). This result reflects the fact that learning is not a fundamental activity in the field to begin with, given the control collector's high opportunity cost of learning imposed by the delivery constraint – hence a direct, technology-induced enhancement of the return to time spent learning has little impact on actual learning. Table 6 shows that, in this counterfactual without a significant difference in knowledge between groups, there is also no differential selection.

To help understand why the indirect effects of technology are so important, we simulate another counterfactual scenario in which we relax the collection constraint by making collections far easier than they are in reality. We do so by multiplying the collection probabilities, π_h and π_ℓ, by 10. We then introduce the technology as in the main model, and simulate its impact. This time, the treatment effect on deliveries is much higher, at 58 percent, and the treatment effect on collections is similar to the main model, at 76 percent (fourth row of Table 6). Collector time use is much more focused on deliveries than in the main model (Appendix Figure A25). The reason is that collections are now not such a binding a constraint as before, so the collectors respond to the better delivery technology by continuing to focus on deliveries. The lesson from this counterfactual is that it is not an artifact of the main estimated model, but a result, that collectors respond to the new technology by shifting their allocation of time towards activities with more binding constraints.

A related insight is that, in this counterfactual with stronger enforcement, collectors choose to learn very little: treatment and control collectors end up knowing only about

[38] In particular, panel B of Figure A17 shows that the treatment group overhauls the organization of its field activities over time while the control group effectively makes no changes to its field organization.

18 percent and 9 percent of households' types, respectively.[39] This result highlights how it is particularly in settings where broad and strong enforcement is limited, such as local taxation in Ghana (Section 2), that learning about households' types is a relevant strategy, given the subsequent targeted collection that it permits.

Relative importance of delivery and learning advantages To help illustrate the individual importance of the delivery and learning advantages of technology, we simulate two counterfactual scenarios in which we shut each of these advantages down one at a time. In the "no learning advantage" counterfactual we set $\eta_T = \eta_C$, and leave all other parameters as in the benchmark estimation. Note this leaves in place the treatment group's advantage in delivering bills, but gives them no additional advantage in learning the households' type. In the "no delivery advantage" counterfactual we set $\theta_T = \theta_C$ but keep other parameters the same, leaving the treatment group with the learning advantage. The bottom two rows of Table 6 summarize the model's predicted treatment effects for deliveries and collections in these counterfactual simulations.

With no learning advantage, the treatment group now delivers 33 percent more bills and collects 64 percent more revenues. Relative to the main model, the treatment collectors choose to spend more time both delivering and collecting, but less time learning (Figure A26). Since the treatment effects in this counterfactual are 31 percentage points larger for collections than for deliveries, we can say that about sixty percent (31/50) of the difference predicted by the estimated main model stems from the delivery advantage. The remaining forty percent then comes from the learning advantage.

This counterfactual reveals that, with time that can be re-allocated to learning but no advantage in learning, treatment collectors effectively choose to not gain more knowledge than control collectors (they discover 22 percent of households' types compared to control collectors' 19 percent). Reflecting the absence of any additional learning, there is no differential selection: treatment and control collectors, having approximately the same amount of knowledge, collect the same fraction of payments from the high-type. In other words, our model shows how the large positive selection observed in the actual experiment (Figure 5) is consistent with a setting where treatment collectors learn over time; the positive selection is inconsistent with settings without learning where

[39] This counterfactual generates positive selection despite limited learning by treatment collectors. However, the magnitude of selection is amplified when π is scaled upwards, so it is not directly comparable to the magnitude of selection in the other counterfactuals.

treatment and control collectors have the same knowledge at the end of the campaign.[40]

With no delivery advantage, the model's predictions look markedly different from the benchmark model and actual experiment. Now, the treatment effect on deliveries is -2 percent, and the treatment effect on collections is 8 percent. Due to the initial delivery constraint, in this case the treatment collector spends much more time on delivery than in the benchmark (Figure A27). The resulting opportunity cost of delivery never becomes large enough that a significant amount of time is re-allocated to learning and, consequently, the treatment collector also devotes less time in total to learn than in the benchmark. Due both to this, and to the mechanically smaller pool of households that received a bill, the treatment collector learns about 31% of households' types (compared to 61% in the benchmark). At the same time, this treatment collector spends a larger share of its non-delivery time on learning versus collecting than in the benchmark: the resulting (infinitely) large ratio of collection versus delivery effects highlights the potential for learning to ultimately cause a large impact on collection. Notwithstanding, this counterfactual shows how, without an initial delivery advantage, the model's treatment group bears little resemblance to the actual experiment's treatment group.

Discussion of learning When combined, the counterfactual exercises reveal how the different parts of the model interact for learning to emerge as a quantitatively important activity – whereby the collector learns the household type for a significant share of properties in their assigned area. Specifically, learning emerges in our setting in two sequential steps: by easing the delivery margin, technology first causes a re-optimization of time-allocations which can potentially be devoted to learning; by enhancing the return to time spent learning, technology in turn ensures that some of this re-allocated time is spent on learning. That is, both technology's direct impact, through the learning advantage, and technology's indirect impact, through time re-allocation, are each necessary but not sufficient for learning to emerge as a significant activity in the field.

Finally, we discuss the factors that determine the learning advantage. In the model, we assumed that the learning advantage is proportional to the delivery advantage, since the navigational improvement increases the return to time spent locating a property and this applies both for the initial delivery and for follow-up visits to learn. However, it is possible that, beyond the improved return on time spent localizing a property, the GIS-

[40]Our benchmark model and counterfactuals assume that the household type is hard to observe but learnable. In alternative settings, the type may instead either be perfectly observable or perfectly unobservable. In Appendix D.2 we discuss how, based on plausible assumptions for the functional form of the cumulative likelihood that a household pays as a function of time spent interacting with the collector, technology's impact on the selection measure from Figure 5 will be *negative* in these alternative settings. The assumptions on the cumulative payment likelihood are plausible given the results in Section 5.

tablet may have inadvertently created a favorable environment that enhances the return to time spent trying to learn. This observation is motivated by findings in environmental psychology which show that improved navigation encourages agents to build survey knowledge – a detailed understanding of the environment and the spatial relationship between locations (Munzer et al., 2006). In Appendix D.3, we discuss how the specific features of the GIS-tablet, in contrast to the poor spatial information available to control collectors, could encourage treatment collectors to make the active, effortful decision to acquire survey knowledge. In addition, we discuss how survey knowledge can enhance the return to time spent trying to learn – by improving spatial orientation, and by increasing the agent's ability to engage with the environment and their willingness to pay attention to details.[41]

This discussion, while suggestive, highlights plausible ways in which the technology may have created an environment that is conducive for the treatment collector to make the strategic and cognitively effortful choice to learn. Future research could seek to integrate measures of survey knowledge from environmental psychology into an enriched economic model of navigation and learning.

7 Conclusion

This paper studies the role of technology in improving tax capacity, by focusing on the local property tax in Ghana. The experiment randomized the presence of an electronic GIS-tablet at the level of a tax collector. The technology was designed to help collectors locate property owners to deliver tax bills – an otherwise challenging task in an environment with incomplete addressing, as is common in many developing countries. Technology caused bill deliveries to increase significantly, as intended. Interestingly, however, technology caused tax collections to increase by a disproportionately larger amount. Experimental results and a dynamic model support our theory that technology, by alleviating the constraint on delivery, led collectors to significantly re-optimize their time allocations across all activities, so as to focus more on activities for which constraints still strongly bind. Specifically, treatment collectors allocated more of their scarce time to the challenging tasks of learning about taxpayers and attempting to collect from them, particularly those with the highest propensity to pay.

Technology improved the collector's ability to locate taxpayers in a setting with scant addressing, but did not otherwise automate any tasks or directly provide other technical

[41] Loosely speaking, the increased return to time spent localizing taxpayers and the survey knowledge may respectively be related to the larger number of collector visits and the longer duration of each visit observed in treatment areas (Table 2). See Appendix D.3 for a detailed discussion.

or organizational enhancements in the field. It is in this sense that we interpret our results as providing a first empirical step towards understanding what is the value of an address for government performance. More work is needed to establish the longer run impacts of improved localization, including the potential persistence of collectors' learning and adaptation of citizens to increased legibility (Okunogbe, 2021; Scott, 1998).

Investments in GIS-technologies for taxation are limited but growing in Africa and other areas of the world (Knebelmann, 2022). In the context of this increasing investment rate, more work is needed to help ensure that the taxes collected with technology adopted at scale are ultimately used to fund locally beneficial public goods.

References

Abebrese, K. (2019): "Implementing street addressing system in an evolving urban center. A case study of the Kumasi metropolitan area in Ghana," Working Paper.

Atkin, D., A. Chaudhry, S. Chaudry, A. Khandelwal, and E. Verhoogen (2017): "Organizational Barriers to Technology Adoption: Evidence from Soccer-Ball Producers in Pakistan," *Quarterly Journal of Economics*, 132.

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

Bandiera, O., M. Best, A. Khan, and A. Prat (2021): "The Allocation of Authority in Organizations," *Quarterly Journal of Economics*, 135.

Banerjee, A. V., S. Chassang, S. Montero, and E. Snowberg (2020): "A Theory of Experimenters," *American Economic Review*, 110.

Basurto, M., P. Dupas, and J. Robinson (2020): "Decentralization and Efficiency of Subsidy Targeting," *Journal of Public Economics*, 185.

Bergeron, A., P. Bessone, J. Kabeya, G. Tourek, and J. Weigel (2022): "Optimal Assignment of Bureaucrats: Evidence from Randomly Assigned Tax Collectors in the DRC," Working Paper.

Besley, T., E. Ilzetzki, and T. Persson (2013): "Weak States and Steady States: The Dynamics of Fiscal Capacity," *American Economic Journal: Macroeconomics*, 5, 203–235.

Besley, T. and T. Persson (2011): *Pillars of Prosperity*, Princeton University Press.

Bigon, L. and A. Njoh (2012): "The Toponymic Inscription Problematic in Urban Sub-Saharan Africa," *Journal of Asian and African Studies*, 50.

Brockmeyer, A., A. Estefan, K. Arras, and J. C. Suárez Serrato (2023): "Taxing Property in Developing Countries: Theory and Evidence from Mexico," *Working Paper*.

Brockmeyer, A. and M. S. Somarriba (2022): "Electronic Payment Technology and Tax Compliance: Evidence from Uruguay's Financial Inclusion Reform," *World Bank Policy Research Working Paper*.

Callen, M., S. Gulzar, S. A. Hasanain, M. Y. Khan, and A. Rezaee (2020): "Data and Policy Decisions: Experimental Evidence from Pakistan," *Journal of Development Economics*, 146.

CASABURI, L. AND U. TROIANO (2016): "Ghost-House Busters: The Electoral Response to a Large Anti-Tax Evasion Program," *Quarterly Journal of Economics*, 131.

DAL BÓ, E., F. FINAN, N. LI, AND L. SCHECHTER (2021): "Information Technology and Government Decentralization," *Econometrica*, 89.

DAS, S., L. GADENNE, T. NANDI, AND R. WARWICK (2023): "Does going cashless make you tax-rich?" *Journal of Public Economics*, 224.

DINCECCO, M. AND G. KATZ (2016): "State Capacity and Long-run Economic Performance," *The Economic Journal*, 126.

DODGE, E., Y. NEGGERS, R. PANDE, AND C. MOORE (2021): "Updating the State: Information Acquisition Costs and Social Protection Delivery," *Working Paper*.

DUFLO, E., M. GREENSTONE, R. PANDE, AND N. RYAN (2018): "The value of regulatory discretion: Estimates from environmental inspections in India," *Econometrica*, 86.

EISSA, N. AND A. ZEITLIN (2014): "Using Mobile Technologies to Increase VAT Compliance in Rwanda," Working Paper.

FAN, H., Y. LIU, N. QIAN, AND J. WEN (2021): "Computerizing VAT Invoices in China," *NBER Working Paper Series*, 24414.

FARVACQUE, C., L. GODIN, H. LEROUX, F. VERDET, AND R. CHAVEZ (2005): "Street addressing and the Management of Cities," *World Bank Directions in Development*, 32923.

FERRAZ, C., D. FOREMNY, AND J. F. SANTINI (2024): "Revenue Slumps and Fiscal Capacity: Evidence from Brazil," *NBER Working Paper*, 32440.

FINAN, F., B. OLKEN, AND R. PANDE (2017): "The Personnel Economics of the Developing State," *Handbook of Economic Field Experiments*, 6.

GADENNE, L. (2017): "Tax Me, But Spend Wisely? Sources of Public Finance and Government Accountability," *American Economic Journal: Applied Economics*, 9.

GORDON, R. AND W. LI (2009): "Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation," *Journal of Public Economics*, 93.

GOVERNMENT OF GHANA (2011): "National Street Naming and Property Addressing Policy," Ghana Ministry of Local Government and Rural Development.

——— (2014): "Internally Generated Revenue Strategy and Guidelines," Ghana Ministry of Finance.

HJORT, J. AND L. TIAN (2024): "The Economic Impact of Internet Connectivity in Developing Countries," *Annual Review of Economics*, Forthcoming.

KALAJ, J., D. ROGGER, AND R. SOMANI (2022): "Bureaucrat time-use: Evidence from a survey experiment," *World Development*, 152.

KHAN, A., A. KHWAJA, AND B. OLKEN (2015): "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors," *Quarterly Journal of Economics*, 131.

——— (2019): "Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats through Performance-Based Postings," *American Economic Review*, 109.

KLEVEN, H. J., M. B. KNUDSEN, C. T. KREINER, S. PEDERSEN, AND E. SAEZ (2011): "Unwilling or Unable to Cheat?" *Econometrica*, 79.

KLEVEN, H. J., C. T. KREINER, AND E. SAEZ (2016): "Why can Modern Governments Tax so Much? An Agency Model of Firms as Fiscal Intermediaries," *Economica*, 83.

KNEBELMANN, J. (2022): "Digitalisation of property taxation in developing countries: Recent advances and remaining challenges," *ODI Report Series*.

KNEBELMANN, J., V. POULIQUEN, AND B. SARR (2023): "Discretion versus Algorithms:

Bureaucrats and Tax Equity in Senegal," Working Paper.
LIPSKY, M. (2010): *Street-Level Bureaucracy*, Russell Sage Foundation.
LUTTMER, E. AND M. SINGHAL (2014): "Tax Morale," *Journal of Economic Perspectives*, 28.
MANARA, M. AND T. REGAN (2022): "Ask a Local: Improving the Public Pricing of Land Titles in Urban Tanzania," *Review of Economics and Statistics*.
MARTINEZ, L. (2023): "Natural resource rents, local taxes, and government performance: Evidence from Colombia," *Review of Economics and Statistics*.
MARX, B., T. STOKER, AND T. SURI (2013): "The Economics of Slums in the Developing World," *Journal of Economic Perspectives*, 27.
MASCAGNI, G., A. MENGISTU, AND F. WOLDEYES (2021): "Can ICTs Increase Tax Compliance," *Journal of Economic Behavior and Organization*, 189.
MATTSSON, M. (2023): "Information Systems, Service Delivery, and Corruption: Evidence from the Bangladesh Civil Service," *Working Paper*.
MUNZER, S., H. ZIMMER, M. SCHWALM, J. BAUS, AND I. ASLAN (2006): "Computer-assisted navigation and the acquisition of route and survey knowledge," *Journal of Environment Psychology*, 26.
MURALIDHARAN, K., P. NIEHAUS, S. SUKHTANKAR, AND J. WEAVER (2021): "Improving Last-Mile Service Delivery Using Phone-Based Monitoring," *American Economic Journal: Applied Economics*, 13.
NARITOMI, J. (2019): "Consumers as Tax Auditors," *American Economic Review*, 109.
NYANGA, M. (2021): "Performance Outcome Area 1 Integrity of the Registered Taxpayer Base: Essentials of a High Integrity Taxpayer Register," *TADAT Insights*.
OKUNOGBE, O. (2021): "Becoming Legible to the State: The Role of Detection and Enforcement Capacity in Tax Compliance," *World Bank Policy Research Working Paper Series*.
OKUNOGBE, O. AND V. POULIQUEN (2022): "Technology, Taxation and Corruption," *American Economic Journal: Economic Policy*, 14.
OKUNOGBE, O. AND G. TOUREK (2024): "How Can Lower-Income Countries Collect More Taxes ? The Role of Technology, Tax Agents, and Politics," *Journal of Economic Perspectives*, 38.
PARASURAMAN, R. (2000): "Designing automation for human use: empirical studies and quantitative models," *Ergonomics*, 43.
POMERANZ, D. (2015): "No Taxation without Information: Deterrence and Self-enforcement in the Value Added Tax," *American Economic Review*, 105, 2539–2569.
PRESCOTT, E. C. AND S. L. PARENTE (1994): "Barriers to Technology Adoption and Development," *Journal of Political Economy*, 102.
RASUL, I. AND D. ROGGER (2018): "Management of Bureaucrats and Public Service Delivery," *Economic Journal*, 128.
SCOTT, J. (1998): *Seeing Like a State*, Yale University Press.
SURI, T. AND C. UDRY (2022): "Agricultural Technology in Africa," *Journal of Economic Perspectives*, 36.
THOMPSON, P. (2010): "Learning by Doing," in *Handbook of the Economics of Innovation*, Oxford University Press.
WEIGEL, J. L. (2020): "The Participation Dividend of Taxation," *Quarterly Journal of Economics*, 135.
WORLD BANK (2020): "Property Tax Diagnostic Manual," World Bank Group.

Appendix
Technology and Tax Capacity: Evidence From Local Governments in Ghana

A Additional Figures and Tables

Table A1: Address Coverage, Technology and Bill Delivery

Outcome: Share of bills delivered (%)	(1)	(2)	(3)	(4)	(5)
Panel A					
Share of properties with address (%)	0.305**	0.170*	0.250**	0.236**	0.193**
	(0.120)	(0.089)	(0.107)	(0.082)	(0.061)
Panel B					
Share of properties with address (%)	0.270*	0.210*	0.263*	0.251**	0.246**
	(0.127)	(0.110)	(0.120)	(0.088)	(0.084)
1(Technology)	0.358***	0.322***	0.296**	0.195**	0.199***
	(0.095)	(0.078)	(0.107)	(0.077)	(0.041)
(Share w address) × **1**(Technology)	-0.353	-0.481**	-0.385	-0.272	-0.402**
	(0.203)	(0.163)	(0.201)	(0.180)	(0.145)
Technology impact at level of share with address					
1^{st} percentile (0%)	0.358***	0.322***	0.296**	0.195**	0.199***
	(0.095)	(0.078)	(0.107)	(0.077)	(0.041)
50^{th} percentile (20%)	0.287***	0.226***	0.219**	0.140**	0.118***
	(0.064)	(0.056)	(0.070)	(0.043)	(0.022)
99^{th} percentile (94.5%)	0.024	-0.132	-0.067	-0.062	-0.181
	(0.122)	(0.105)	(0.095)	(0.096)	(0.104)
District controls		x			x
Share neighbors with tech			x		x
Region FE				x	x
Mean outcome variable	0.430	0.430	0.430	0.430	0.430
Observations	216	216	216	216	216
Clusters	10	10	10	10	10

Notes: The regression model is a cross-sectional regression of all 216 districts in Ghana, with one local government per district. In all regressions, the outcome is the share of bills delivered. In panel A, the outcome is regressed on the share of properties with addresses. Panel B adds as regressors **1**(Technology), a dummy variable taking a value of 1 if the local government has an electronic tax registry of properties and 0 otherwise, and its interaction with the address-share variable. The bottom of panel B shows the estimated impact of technology at different percentiles of the address-share variable: 1^{st} percentile (districts where 0% of properties have addresses), 50^{th} percentile (20%); and 99^{th} percentile (94.5%). Across columns, the same controls are used as in Table A3. Standard errors, clustered at the region level, are shown in parentheses. For details on the variables, see Data Appendix B.1. * p<0.10 ** p<0.05 *** p<0.01.

Table A2: Associations with Technology Adoption

	1(Technology exists)	
	(1)	(2)
Total population	0.103***	0.065**
	(0.026)	(0.024)
Income per capita	0.083**	0.015
	(0.031)	(0.020)
Urban share of population	0.118***	0.073*
	(0.025)	(0.032)
Share of properties with address	0.125**	0.094**
	(0.054)	(0.040)
Share of properties with valuation	0.177***	0.134***
	(0.028)	(0.025)
Legal capacity to enforce taxes	0.083*	0.049*
	(0.041)	(0.022)
Tax-delinquents taken to court	0.017	0.012
	(0.031)	(0.021)
Officials' years of work experience	0.046*	0.047
	(0.024)	(0.033)
Officials' years of education	0.029	0.017
	(0.030)	(0.025)
Trust in officials	-0.002	-0.001
	(0.014)	(0.015)
Citizen tax awareness	-0.010	-0.011
	(0.012)	(0.019)
Citizen compliance attitude	-0.009	0.016
	(0.019)	(0.010)
Region FE		X
Observations	216	216
Clusters	10	10

Notes: Each cell represents the β coefficient from a separate cross-district regression, based on the model $\mathbf{1}(Technology)_{dr} = \beta \cdot X_d + \mu_r + \epsilon_{dr}$ where $1(Technology)_{dr}$ is a dummy equal to 1 if the local government in district d in region r has an electronic tax registry of properties (see Section 2). X_d is the district characteristic which varies between rows; in column (2), region fixed effects (μ_r) are included. All district characteristics are standardized, for ease of comparison across rows. Standard errors are clustered at the regional level. For a description of all the district characteristics, see Data Appendix B.1. * p<0.10 ** p<0.05 *** p<0.01.

Table A3: Tax Outcomes and Technology Across Local Governments

	(1)	(2)	(3)	(4)	(5)
Panel A: Taxes collected per Capita					
1(Technology)	5.525**	2.482***	3.116***	2.919***	2.667*
	(1.742)	(0.768)	(0.796)	(0.835)	(1.372)
Mean outcome variable	4.153	4.153	4.153	4.153	4.153
Panel B: Share of Bills Delivered (%)					
1(Technology)	0.239***	0.097*	0.177***	0.134***	0.079**
	(0.052)	(0.051)	(0.034)	(0.015)	(0.035)
Mean outcome variable	0.430	0.430	0.430	0.430	0.430
Panel C: Taxes per Bill Delivered					
1(Technology)	6.868*	6.685**	4.629**	3.593***	2.911*
	(3.617)	(2.475)	(1.952)	(0.809)	(1.501)
Mean outcome variable	11.453	11.453	11.453	11.453	11.453
District controls		x			x
Share neighbors with tech			x		x
Region FE				x	x
Observations	216	216	216	216	216
Clusters	10	10	10	10	10

Notes: The regression model is a cross-sectional regression of all 216 districts in Ghana, with one local government per district. The variable **1**(Technology) is a dummy variable taking a value of 1 if the local government has an electronic tax registry of properties. Across panels, the outcome is: local taxes collected per capita (Panel A); the share of bills that are delivered (Panel B); local taxes collected per bill delivered (Panel C). Across columns, the specifications are: no controls in column (1); district controls (log per capita income, log population, urban share of population, share of properties with valuations, share of properties with official addresses, legal capacity, officials' years of work experience) in column (2); the share of each district's geographically adjacent neighbor governments with technology in column (3); region fixed effects in column (4); all three sets of controls in column (5). Standard errors are clustered at the region level. See Data Appendix B.1 for more details on the variables. * p<0.10 ** p<0.05 *** p<0.01.

Table A4: Randomization Balance

	N	Control mean	Treatment coefficient
	(1)	(2)	(3)
Panel A: Tax bill characteristics			
Current tax amount	8120	322.8	-9.0
			(16.4)
Total tax amount	8120	692.5	-5.5
			(29.1)
Previous pay status	8120	1.2	0.0
			(0.0)
Previous tax payment	8120	59.7	-6.6
			(9.4)
Residential	8120	0.5	0.0
			(0.0)
Property quality	8120	0.5	0.0
			(0.1)
F-test joint significance [F, p]			[0.71,0.66]
Panel B: Collector-unit characteristics			
Experience in Madina	56	0.7	-0.1
			(0.1)
Performance rating	56	0.2	-0.1
			(0.1)
Total bills to deliver	56	135.2	1.7
			(4.7)
Average amount per tax bill	56	322.6	-7.4
			(16.5)
F-test joint significance [F, p]			[0.16,0.95]
Panel C: Household characteristics			
Income index	4334	-0.014	0.003
			(0.106)
Liquidity index	4334	0.051	-0.177
			(0.119)
Taxpayer awareness index	4334	0.011	-0.01
			(0.039)
F-test joint significance [F, p]			[1.07,0.38]

Notes: This table presents balance checks of the randomization assignment for characteristics at the bill level (Panel A), the collector-unit level (Panel B), and the household level (Panel C). Details on the variables used in this table are provided in Section 3.1. The treatment coefficient in column (3) is the coefficient on technology in a cross-sectional regression with strata fixed effects. Standard errors are clustered at the collector-unit level. At the bottom of each panel, the F-test on the joint significance of all characteristics is reported along with the p-value. For more information on the household characteristics, see Data Appendix B.5. * p<0.10 ** p<0.05 *** p<0.01.

Table A5: Attrition

	1(Survey response exists)			
	(1)	(2)	(3)	(4)
1(Technology)	-0.017	0.005		
	(0.086)	(0.079)		
	[0.720]	[0.940]		
1(Technology) × **1**(Round 1)			-0.021	-0.014
			(0.096)	(0.092)
1(Technology) × **1**(Round 2)			0.011	0.034
			(0.125)	(0.124)
1(Technology) × **1**(Round 3)			-0.022	-0.011
			(0.096)	(0.084)
Collector-unit controls		X		X
Survey round FE	X	X	X	X
Observations	168	168	168	168

Notes: This table investigates attrition in the collector surveys. The sample is the fully balanced set of 3 collector surveys for all 56 collectors ($N = 168$). The outcome is a dummy variable which takes a value of 1 if there is a survey response collected by the enumerators, and 0 otherwise. Attrition is estimated based on (1) – with the average treatment effect in columns (1)-(2), and the dynamic effects (by survey round) in columns (3)-(4). All regressions include survey round fixed effects; even columns include the collector-unit controls described in Section 3. Standard errors clustered at the collector-unit are reported in parentheses. In columns (1)-(2), the randomization inference p-value is reported in brackets. * p<0.10 ** p<0.05 *** p<0.01.

Table A6: Change in Sample Balance for Collector Survey Outcomes

Outcome	Chall. navigate	Chall. locate	Know HH-type	Focus aware	Focus able	Focus PGs	Focus hard obs.	Focus easy obs.	Focus hard-easy	# failed/ per success	Hours worked	Hours per bill	Hours non-deliv	Fieldwork prepare	Content in job	Wrong info	Resistance from HH	Superv. monitor	Superv. unavai.	Superv. mistakes
Panel A: Unbalanced sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
1(Technology)	-0.55***	-0.31***	0.18**	0.13	0.19**	0.08	0.14*	0.08	0.05	-0.93	-0.74	-0.77***	8.59*	0.04	0.10	-0.26	-0.06	-0.14	-0.21	-0.07
	(0.07)	(0.06)	(0.09)	(0.09)	(0.09)	(0.09)	(0.07)	(0.05)	(0.04)	(1.34)	(1.84)	(0.19)	(4.72)	(0.07)	(0.15)	(0.18)	(0.16)	(0.21)	(0.21)	(0.21)
Observations	141	141	141	141	141	141	141	141	141	141	141	111	141	141	141	141	141	141	141	141
Panel B: Balanced sample	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)	(33)	(34)	(35)	(36)	(37)	(38)	(39)	(40)
1(Technology)	-0.54***	-0.30***	0.20*	0.15	0.22**	0.11	0.16*	0.08	0.07	-1.57	-1.09	-0.80***	8.64*	0.08	0.15	-0.22	-0.10	-0.08	-0.25	-0.03
	(0.07)	(0.07)	(0.10)	(0.09)	(0.10)	(0.10)	(0.08)	(0.06)	(0.04)	(2.18)	(2.00)	(0.20)	(4.69)	(0.08)	(0.17)	(0.19)	(0.16)	(0.22)	(0.23)	(0.23)
Observations	123	123	123	123	123	123	123	123	123	123	123	92	123	123	123	123	123	123	123	123
Collector-unit controls	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Survey round FE	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Notes: This table estimates the average impact of technology on collector outcomes, based on equation (1). Panel A presents results in the unbalanced sample, panel B in the balanced sample. The outcomes are the same as in Table 3, Table 4, and Figure 2. Standard errors clustered at the collector-unit are reported in parentheses. For details on the variables, see Data Appendix B.6. * p<0.10 ** p<0.05 *** p<0.01.

Table A7: Cost-Benefit Analysis per Collector

			Control	Treatment
Cost items	Unit price	Total units		
Daily allowance	10	42	420	420
Commission rate	8%	42	66.3	134.8
Tablet rental	10	42	0	420
Network connection	40	1	0	40
Total cost			486.3	1014.8
Totals				
Total taxes collected			829	1685
Total taxes net of cost			342.7	670.2

Notes: This table presents a cost-benefit analysis for the running costs of the average collector in the treatment and control groups during the 42 days of the tax experiment campaign. Some cost items are common to all collectors. Each collector receives 10 GHC in daily allowance. Moreover, each collector in both groups receives an 8% commission for taxes collected – which corresponds to 66.3 GHC ($0.08 \times 829 = 66.3$) for the average collector in the control group and 134.8 GHC ($0.08 \times 1685 = 134.8$) for the average collector in the treatment group (based on Figure 1). Some cost items are specific to the treatment group. In particular, the private firm pays a 10 GHC daily rental price to the tablet provider; moreover, the tablet requires network connection. The top panel reports the total costs for the average collector over the 6-week campaign. In the bottom panel are reported the average taxes collected at the end of the campaign, as well as the taxes collected net of total cost. It is important to note that the cost items in the treatment group refer to the running cost of using the tablet on a daily basis – they do not account for any cost of building the GIS-registry database which serves as the input to the tablet (see Section 3).

Table A8: Robustness Checks for Technology Impacts on Tax and Bribe Outcomes

	Any visit by tax collector	Total visits (%)	Time per visit (%)	Bill delivered	Any positive tax payment	Full tax payment	Total payment amount (in GHC)	Any bribe (coercive or collusive)	Total bribe amount (in %)	Coercive bribe amount (in %)	Collusive bribe amount (in %)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Benchmark											
1(Technology)	0.087**	0.097**	0.242**	0.054	0.043**	0.023**	25.910**	0.116***	0.025**	0.011*	0.040*
	(0.033)	(0.045)	(0.116)	(0.036)	(0.021)	(0.010)	(10.901)	(0.039)	(0.011)	(0.006)	(0.021)
Panel B: No Controls											
1(Technology)	0.082**	0.090*	0.232*	0.049	0.039**	0.021*	24.931**	0.120***	0.027**	0.012**	0.042*
	(0.034)	(0.046)	(0.127)	(0.036)	(0.017)	(0.012)	(10.891)	(0.038)	(0.012)	(0.005)	(0.023)
Panel C: Extensive Controls											
1(Technology)	0.086**	0.092**	0.240*	0.055	0.047**	0.027**	27.211**	0.113***	0.025**	0.014**	0.036*
	(0.032)	(0.044)	(0.128)	(0.034)	(0.020)	(0.012)	(11.181)	(0.037)	(0.010)	(0.005)	(0.019)
Strata FE	X	X	X	X	X	X	X	X	X	X	X
Mean in CG	0.549	0.652	0.214	0.506	0.162	0.061	40.951	0.139	0.107	0.022	0.190
Observations	4334	4334	4334	4334	4334	4334	4334	4334	4334	4334	4334
Clusters	56	56	56	56	56	56	56	56	56	56	56

Notes: This table presents technology impacts on the main set of tax and bribe outcomes. Panel A uses the benchmark specification in equation (2). In Panel B, the estimation model is the same except that all household and collector controls are removed. Panel C augments (2) with additional controls, which are the set of fixed, hard-to-observe characteristics used in the targeting analysis, specifically income, liquidity and taxpayer awareness from Table A4. See also Section 5 and Figure 4. Standard errors clustered at the collector-unit are reported in parentheses. For a description of the variables, see Data Appendix B.2-B.4. * p<0.10 ** p<0.05 *** p<0.01.

Table A9: Beliefs about Enforcement and Tax Morale

	Technology coefficient ($\hat{\beta}$) (1)	Mean in CG (2)	*N* (3)
Panel A: Enforcement & Information Capacity			
Share of HHs that comply with taxes	0.80	60.32	4334
	(2.38)		
Likelihood non-complier will end up paying	-0.07	3.08	4334
	(0.07)		
Likelihood Gov't has info about my tax status	-0.13	2.95	4334
	(0.13)		
Likelihood Gov't has info about my job	0.03	2.52	4334
	(0.09)		
Panel B: Equity & Efficiency of Tax Collection			
Agree efforts to collect taxes efficiently	0.01	3.58	4334
	(0.07)		
Agree efforts to ensure fair share paid	-0.18***	3.42	4334
	(0.07)		
Agree efforts to collect for useful purposes	0.08	3.04	4334
	(0.11)		
Panel C: Government Integrity and Competency			
% of taxes wastefully spent	-3.48	55.81	4330
	(4.64)		
Agree Gov't has capacity to improve roads	0.04	3.94	4334
	(0.11)		
Overall Gov't competency rating	0.07	2.41	4334
	(0.07)		
Panel D: Satisfaction with Government Services			
Quality of tax collector services	-0.003	2.31	4334
	(0.05)		
Quality of tax authority services	-0.02	2.31	4334
	(0.05)		
Quality of overall Gov't services	-0.01	2.20	4334
	(0.05)		

Notes: This table presents technology impacts on beliefs and tax morale. Each row presents the technology treatment coefficient (in column 1) from estimating equation (2) on different outcomes (which are described to the left). Standard errors are clustered at the collector-unit. Column (2) presents the mean of the outcome variable in control areas, while column (3) shows the sample size. For a description of all the outcomes, see Data Appendix B.3. All regressions include household and collector controls (Section 3). * p<0.10 ** p<0.05 *** p<0.01.

Table A10: Heterogeneity in Beliefs about Enforcement and Tax Morale

	Technology coefficient ($\hat{\beta}$)	Heterogeneity coefficient ($\beta \hat{\times} H$)
Outcome: Enforcement and Information Capacity		
Heterogeneity H: Liquidity index	-0.050	-0.016
	(0.056)	(0.053)
Heterogeneity H: Income index	-0.051	0.002
	(0.057)	(0.042)
Heterogeneity H: Taxpayer awareness index	-0.050	-0.021
	(0.056)	(0.057)
F-test joint significance of interaction terms [F, p]		[0.09, 0.96]
Outcome: Equity & Efficiency of Tax Collection		
Heterogeneity H: Liquidity index	-0.016	0.048
	(0.059)	(0.055)
Heterogeneity H: Income index	-0.010	0.059
	(0.060)	(0.039)
Heterogeneity H: Taxpayer awareness index	-0.012	0.068
	(0.061)	(0.069)
F-test joint significance of interaction terms [F, p]		[1.27, 0.29]
Outcome: Government Integrity and Competency		
Heterogeneity H: Liquidity index	0.048	0.039
	(0.070)	(0.063)
Heterogeneity H: Income index	0.060	0.012
	(0.073)	(0.040)
Heterogeneity H: Taxpayer awareness index	0.064	-0.036
	(0.072)	(0.048)
F-test joint significance of interaction terms [F, p]	[0.32, 0.81]	
Outcome: Satisfaction with Gov't Services		
Heterogeneity H: Liquidity index	-0.018	0.042
	(0.069)	(0.059)
Heterogeneity H: Income index	-0.009	0.011
	(0.069)	(0.032)
Heterogeneity H: Taxpayer awareness index	-0.017	0.041
	(0.070)	(0.064)
F-test joint significance of interaction terms [F, p]		[0.45,0.72]

Notes: This table investigates heterogeneous technology impacts on beliefs and tax morale. Each row presents the technology treatment coefficient and the interaction coefficient, from estimating equation (2) augmented with the interaction between technology and the heterogeneity dimension H. Rows differ in the interaction (liquidity, income or taxpayer awareness), and panels differ in the outcome. The F-test at the bottom of each panel tests the joint significance of the three interaction coefficients for a given outcome. The outcomes are described in Data Appendix B.3. The heterogeneity dimensions are described in Data Appendix B.5. Standard errors clustered at the collector-unit are reported in parentheses. * p<0.10 ** p<0.05 *** p<0.01.

Table A11: Heterogeneity in Tax Outcomes, Beliefs and Morale by Registration Status

	Technology coefficient $\hat{\beta}$	Heterogeneity coefficient $\widehat{\beta \times 1(\text{New register})}$
Panel A: Tax payment outcomes		
Outcome: 1(Any payment)	0.043**	-0.003
	(0.020)	(0.040)
Outcome: 1(Full payment)	0.024*	0.012
	(0.012)	(0.026)
Outcome: Amount paid (in GHC)	25.796**	0.901
	(11.608)	(22.241)
Panel B: Beliefs and tax morale outcomes		
Outcome: Satisfaction with gov't services	-0.030	0.094
	(0.066)	(0.140)
Outcome: Integrity of gov't	0.063	-0.044
	(0.074)	(0.134)
Outcome: Equity & efficiency	-0.021	0.016
	(0.059)	(0.124)
Outcome: Enforcement & information	-0.056	0.019
	(0.055)	(0.131)

Notes: This table investigates heterogeneous technology impacts by registration status. A subset of the properties were registered in the tax registry for the first time during the experiment's campaign; 1(New register) takes a value of 1 if the property is newly registered, and 0 otherwise. Each row corresponds to a different outcome. In each row, we present the technology treatment coefficient and the interaction coefficient, from estimating equation 2 augmented with the interaction between technology and 1(New register). Panel A focuses on tax outcomes: a dummy for any tax payment made; a dummy for a full tax payment made; and, the amount of taxes paid (in GHC). Panel B focuses on household beliefs and tax morale outcomes: an index for satisfaction with government services; an index for the perceived integrity of local government; an index for the perceived equity and efficiency of the tax collection process; and, an index for the perceived enforcement and information capacity of the local government. The outcomes are the same as in Table 2 and Table A8. Standard errors clustered at the collector-unit are reported in parentheses. For more details on the variables used, see Data Appendix B.2-B.3. * p<0.10 ** p<0.05 *** p<0.01.

Table A12: Collector Survey Outcomes

	Technology coefficient ($\hat{\beta}$) (1)	Mean in CG (2)	*N* (3)
Panel A: Contentment in job			
This was a productive week for me	0.239	3.391	141
	(0.184)		
I was content while working this week	0.024	3.260	141
	(0.175)		
I am satisfied with my job	0.054	3.391	141
	(0.194)		
Panel B: Collection challenges faced this week			
Some of my bills had the wrong address	-0.304	3.449	141
	(0.279)		
Some of my bills had the wrong amount to be paid	0.004	3.594	141
	(0.300)		
Bill recipients preferred not to pay in cash	0.015	3.115	141
	(0.233)		
Bill recipients did not trust me to collect their payment	-0.200	3.507	141
	(0.322)		
Bill recipients felt the amount to pay was too high	-0.052	3.782	141
	(0.335)		
Panel C: Supervisors "..."			
Spent a lot of time monitoring my work	0.163	3.362	141
	(0.247)		
Were available when I needed them	0.335	3.347	141
	(0.228)		
Checked on me regularly to prevent mistakes	0.089	3.405	141
	(0.248)		

Notes: This table shows the average impacts of technology on collectors' job satisfaction (panel A), work challenges (panel B), and perceptions of supervisors (panel C). Each row presents the technology treatment coefficient (in column 1) from estimating equation (1) on different outcomes (which are described to the left). All questions take on a value from 1 (strongly disagree) to 5 (strongly agree). Column (2) presents the mean of the outcome variable in control areas, while column (3) shows the sample size. For the outcomes in Panel C, note that they are reverse coded in the other tables and figures of the paper (e.g. Table 3) to indicate challenges. Standard errors clustered at the collector-unit are reported in parentheses. For a description of all the outcomes, see Data Appendix B.6. All regressions include collector-unit controls (Section 3). * p<0.10 ** p<0.05 *** p<0.01.

Table A13: Exposure of Control Group Collectors to Technology

	Data: Collector Surveys							Data: Collector Daily Information	
	Challenge locate (1)	Knowledge of HH type (2)	Coll. focus hard vs easy (3)	Content in job (4)	Total hours worked (5)	# failed per successful visit (6)	# hours per bill delivered (7)	Total bills delivered (8)	Total taxes collected (GHC) (9)
TG-share in adjacent units (%)	0.125	0.039	0.009	-0.018	-0.252	0.202	0.413	-1.744	-27.577
	(0.108)	(0.174)	(0.080)	(0.221)	(4.390)	(1.515)	(0.522)	(13.130)	(191.255)
Mean in CG	0.794	0.197	0.041	-0.065	19.057	8.028	1.662	52.515	329.206
Observations	71	71	71	71	71	71	71	1164	1164
Collector-unit controls	X	X	X	X	X	X	X	X	X
Time FE	X	X	X	X	X	X	X	X	X

Notes: The analysis in this table is restricted to the control collectors. The estimation is based on equation (1) and is estimated at the collector-survey round level in columns (1) to (7), and at the collector-day level in columns (8)-(9). The estimating equation replaces the treatment assignment, $\mathbf{1}(Tech)_c$, with the variable $(TG - share)_c$ which measures for each collector c the share of geographically adjacent collection units which are assigned to the treatment group. In columns (1) to (7), the outcome variables are from the collector surveys: a dummy equal to 1 if the collector reports finding it challenging or very challenging to locate assigned taxpayers (and 0 otherwise); a dummy equal to 1 if the collector reports having a good understanding of which properties are more willing and able to pay (and 0 otherwise); the difference in strategy focus between hard to observe characteristics and easy to observe characteristics; contentment in job; hours worked; # of failed visits per succesful visit; and # of hours worked per bill delivered. In columns (8)-(9), the outcome variables are from the collector daily administrative data (Section 3): total bills delivered; total taxes collected (in GHC). All regressions include the collector-unit controls (Section 3). All regressions include time fixed effects, which are survey round fixed effects in columns (1)-(7) and campaign-day fixed effects in columns (8)-(9). Standard errors clustered at the collector-unit are reported in parentheses. For details on the variables from the collector surveys, see Data Appendix B.6.For details on the variables from the collector daily data, see Section 3. * p<0.10 ** p<0.05 *** p<0.01.

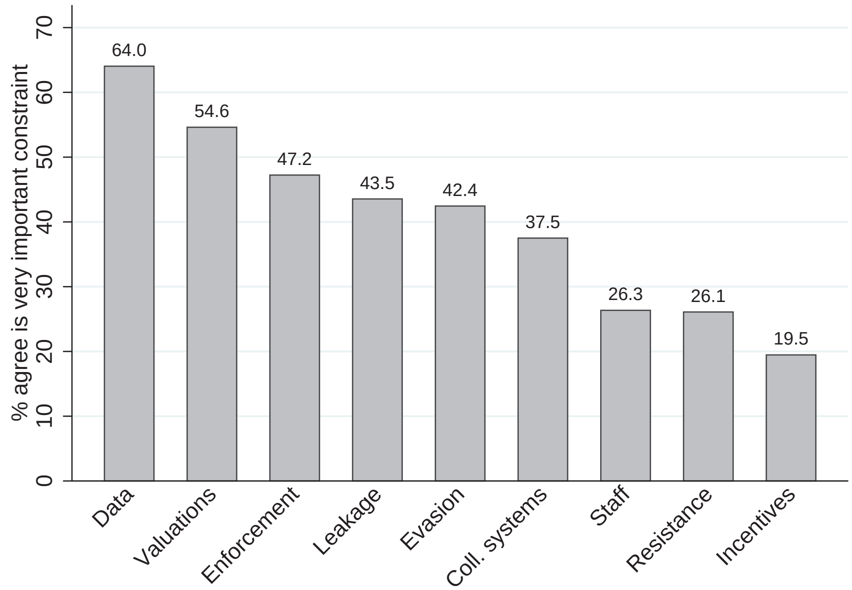
Table A14: Hard to Observe and Easy to Observe Indices as Predictors of Tax Payment

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Outcome is 1(Any payment)						
Income index	0.016**					
	(0.008)					
Liquidity index		0.017**				
		(0.007)				
Awareness index			0.011			
			(0.0139)			
Hard to observe index				0.035**		0.029**
				(0.014)		(0.014)
Easy to observe index					0.021***	0.017**
					(0.006)	(0.008)
Panel B: Outcome is 1(Full payment)						
Income index	0.012**					
	(0.005)					
Liquidity index		0.012**				
		(0.005)				
Awareness index			0.005			
			(0.009)			
Hard to observe index				0.024***		0.025***
				(0.009)		(0.009)
Easy to observe index					-0.003	-0.007
					(0.009)	(0.010)
Block FEs	X	X	X	X	X	X
Observations	4334	4334	4334	4334	4334	4334
Clusters	56	56	56	56	56	56

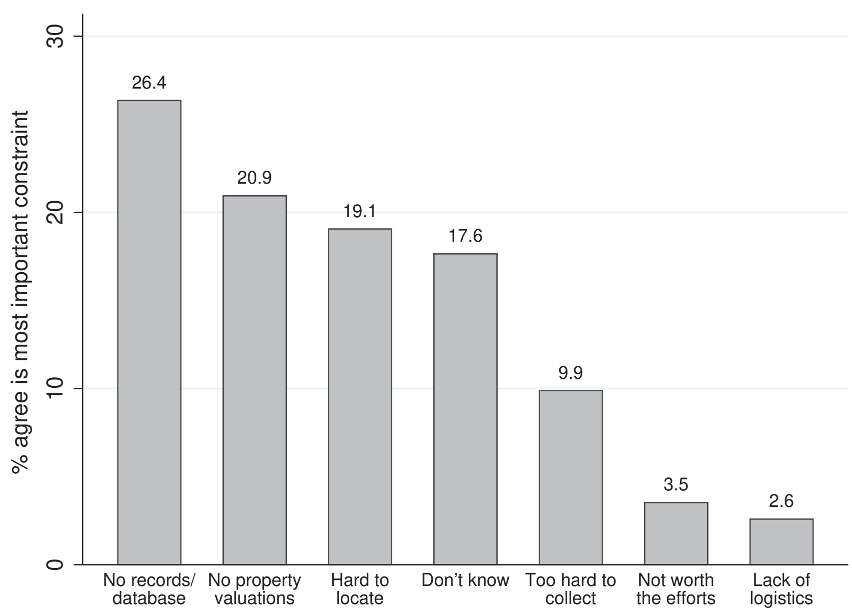
Notes: This table estimates how the hard to observe index, its components, and the easy to observe index are predictors of tax payment outside of the experimental sample. Specifically, the tax outcomes are measured at the property level in the tax campaign prior to the experiment's campaign. In panel A, the outcome is a dummy which takes a value of 1 if the household makes any tax payment (and 0 otherwise). In panel B, the outcome is a dummy which takes a value of 1 if the household makes a tax payment equal to the entire tax liability due. Across columns, the outcome is regressed on different explanatory variables: the income index; the liquidity index; the taxpayer awareness index; the hard to observe index (which combines income, liquidity and awareness); the easy to observe index; and, the hard and easy to observe indices. For a description of the indices, see Data Appendix B.5. All regressions include block fixed effects (7-8 properties per block, on average). The outcome is not defined for the subset of property owners that were newly registered for the experiment's campaign (Section 4.1). We assign an arbitrary value of the outcome for those property owners, and include a dummy for new registration. Standard errors are clustered at the collector-unit level. * p<0.10 ** p<0.05 *** p<0.01

Figure A1: Constraints on Tax Collection and Bill Delivery

(a) Perceived Importance of Different Constraints on Tax Collection

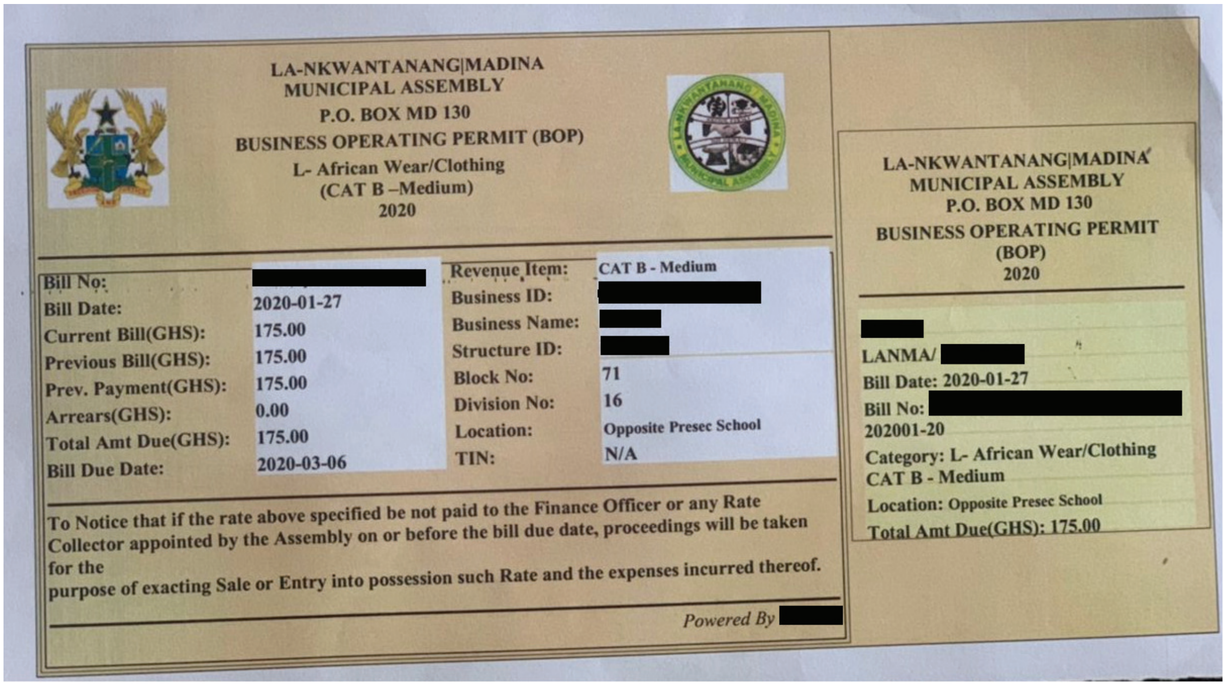


(b) Most Important Perceived Constraint on Bill Delivery



Notes: These panels show the perceived constraints on tax collection and bill delivery as reported by local government officials and politicians. In Panel A, the bars show the percent of all respondents that consider a particular constraint to be 'most important', on a five-choice scale from 'least important' to 'most important'. In Panel B, the bars show the percent of all respondents who consider a particular constraint to be the most important constraint (mutually exclusive choices). Responses are pooled across all respondents in all 216 local governments.

Figure A2: Illustration of Tax Bill



LA-NKWANTANANG|MADINA
MUNICIPAL ASSEMBLY
P.O. BOX MD 130
BUSINESS OPERATING PERMIT (BOP)
L- African Wear/Clothing
(CAT B –Medium)
2020

Bill No:	[redacted]	Revenue Item:	CAT B - Medium
Bill Date:	2020-01-27	Business ID:	[redacted]
Current Bill(GHS):	175.00	Business Name:	[redacted]
Previous Bill(GHS):	175.00	Structure ID:	[redacted]
Prev. Payment(GHS):	175.00	Block No:	71
Arrears(GHS):	0.00	Division No:	16
Total Amt Due(GHS):	175.00	Location:	Opposite Presec School
Bill Due Date:	2020-03-06	TIN:	N/A

To Notice that if the rate above specified be not paid to the Finance Officer or any Rate Collector appointed by the Assembly on or before the bill due date, proceedings will be taken for the
purpose of exacting Sale or Entry into possession such Rate and the expenses incurred thereof.

Powered By [redacted]

LA-NKWANTANANG|MADINA
MUNICIPAL ASSEMBLY
P.O. BOX MD 130
BUSINESS OPERATING PERMIT
(BOP)
2020

[redacted]
LANMA/ [redacted]
Bill Date: 2020-01-27
Bill No: [redacted]
202001-20
Category: L- African Wear/Clothing
CAT B - Medium
Location: Opposite Presec School
Total Amt Due(GHS): 175.00

Notes: This figure provides an illustration of a typical business property tax bill used by the local government of Madina, where the experiment takes place. Instead of a property number and street name, the location information refers to a local landmark 'Opposite Presec School' (a private school).

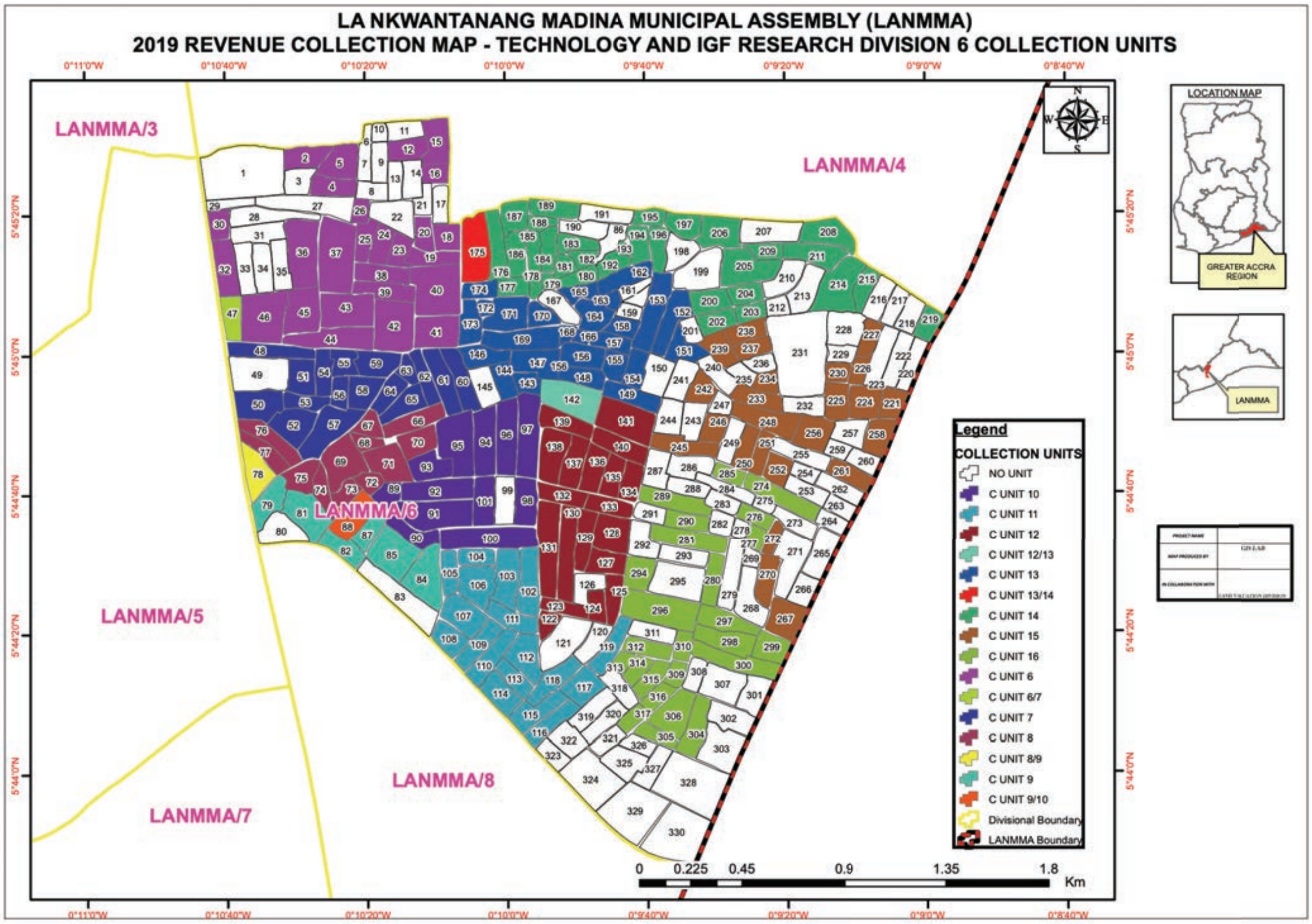
Figure A3: Illustrations of Navigation With and Without Technology

(a) Navigation with Tablet in Treatment Group

(b) Navigation without Tablet in Control Group

Notes: Panel A illustrates the navigational assistance provided by the GIS-tablet, which provides the user's live location on the digital map and the location of a designated property. Panel B illustrates navigation in the control group, where collectors sometimes ask local residents for assistance to navigate and locate a property.

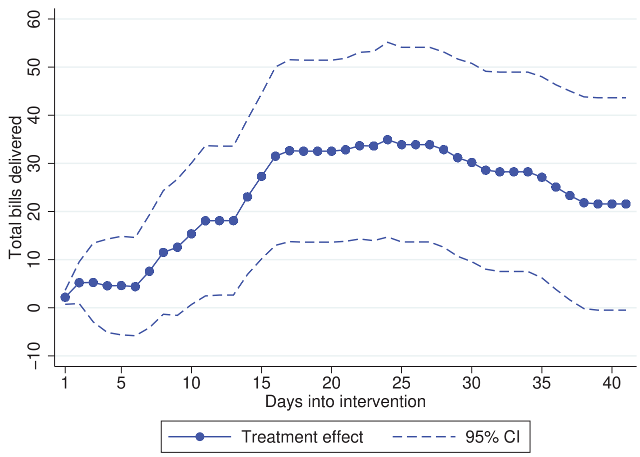
Figure A4: Illustration of Map with Tax Collection Units



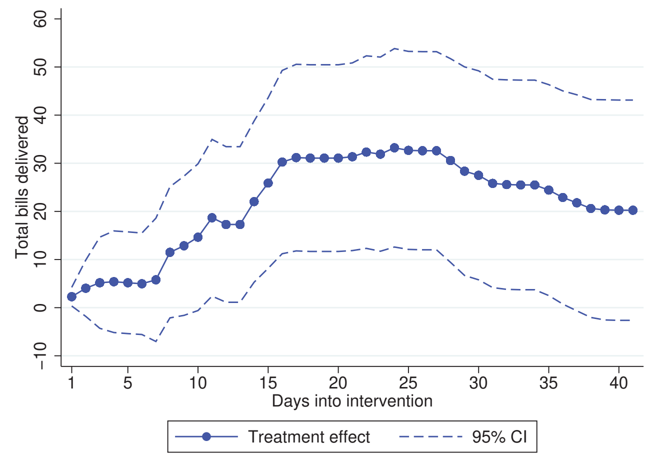
Notes: This graph provides an illustration of a map which shows some of the collection units that exist in the district of Madina. Due to confidentiality, these collection units are not necessarily included in the experimental sample.

Figure A5: Treatment Effect for Bills Delivered

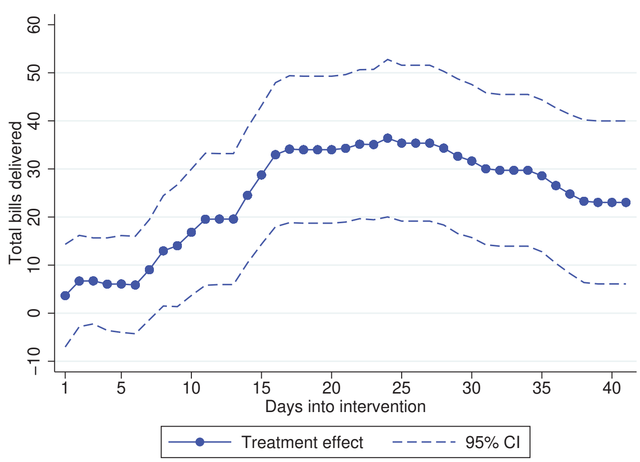
(a) Benchmark



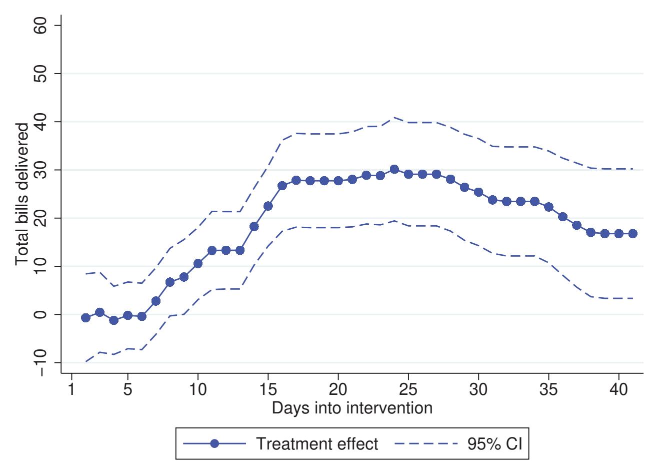
(b) Non-Winsorized Outcome



(c) Inclusion of Covariates



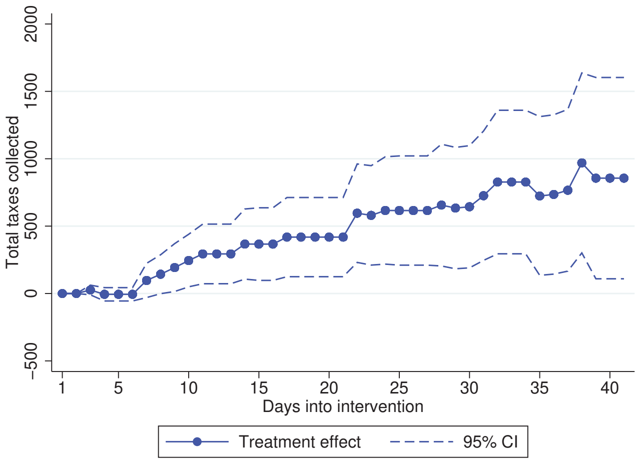
(d) Inclusion of Collector-Unit Fixed Effects



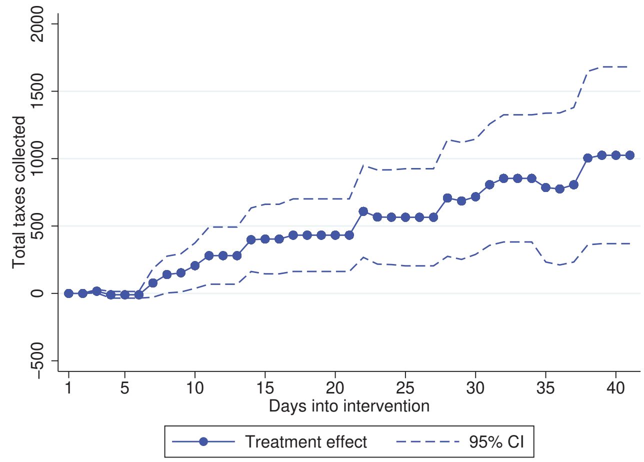
Notes: These panels show robustness for the estimated treatment impact of technology on the number of property tax bills delivered. Panel A shows the treatment effect from the benchmark result in Figure 1, based on estimating equation (1). In Panel B, the benchmark is changed by using the non-winsorized outcome. In Panel C, the benchmark is changed by including control variables: a dummy for whether the collector has previously worked in Madina; a dummy for whether the collector is assessed to be high performing; the total number of bills assigned to the collector; and, the average tax bill value per bill assigned. In Panel D, the benchmark estimation is augmented with collector-unit fixed effects – in this case we omit β_1, the treatment category in day 1 (see equation 1). Coefficients together with the 95% confidence intervals are displayed. Standard errors are clustered at the collector-unit level. The analysis is based on the daily collector data, described in Section 3.1.

Figure A6: Treatment Effect for Taxes Collected

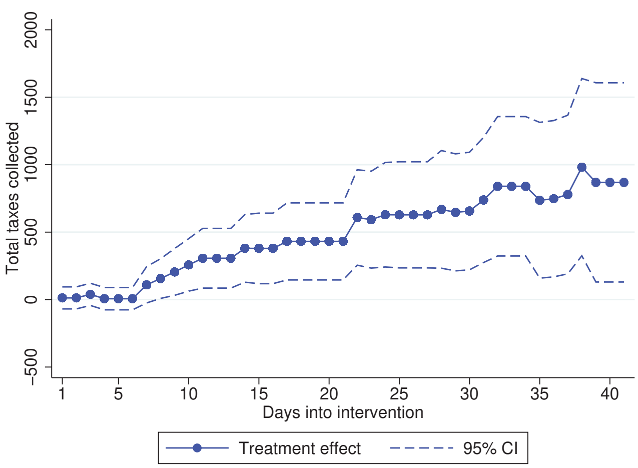
(a) Benchmark



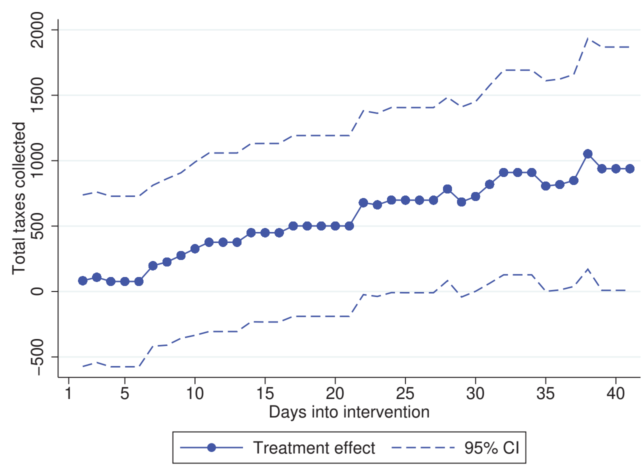
(b) Non-Winsorized Outcome



(c) Inclusion of Covariates



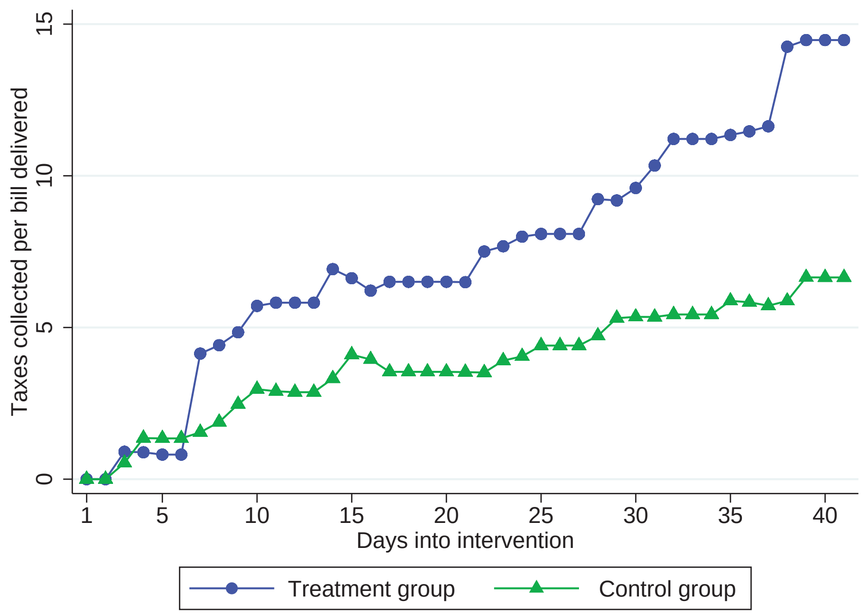
(d) Inclusion of Collector-Unit Fixed Effects



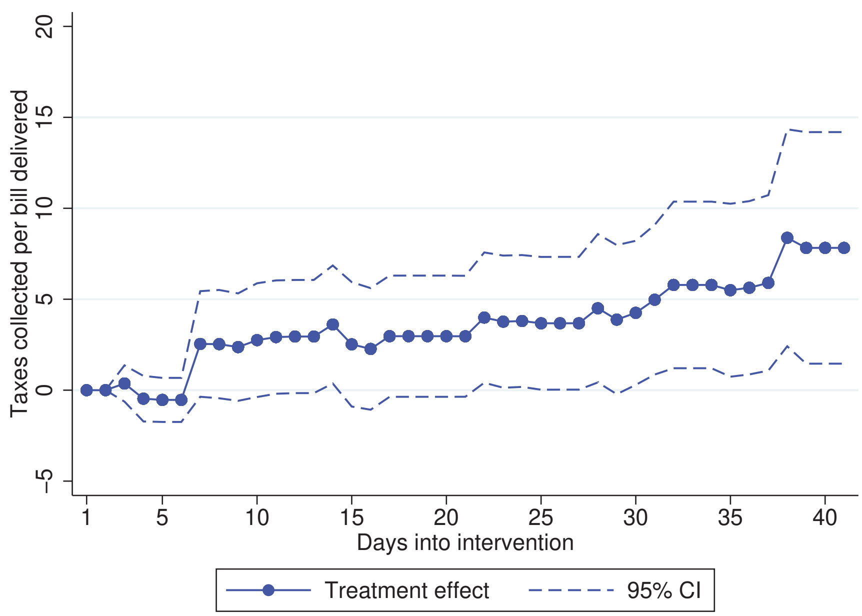
Notes: These panels show robustness for the estimated treatment impact of technology on total taxes collected. Panel A estimates the treatment effect for the benchmark result in Figure 1, based on estimating equation (1). In Panel B, the benchmark is changed by using the non-winsorized outcome. In Panel C, the benchmark is changed by including control variables: a dummy for whether the collector has previously worked in Madina; a dummy for whether the collector is assessed to be high performing; the total number of bills assigned to the collector; and, the average tax bill value per bill assigned. In Panel D, the benchmark estimation is augmented with collector-unit fixed effects – in this case we omit β_1, the treatment category on day 1 (see equation 1). Coefficients together with the 95% confidence intervals are displayed. Standard errors are clustered at the collector-unit level. The analysis is based on the daily collector data, described in Section 3.1.

Figure A7: Impacts of Technology on Taxes Collected per Bill Delivered

(a) Taxes Collected per Bill Delivered by Group



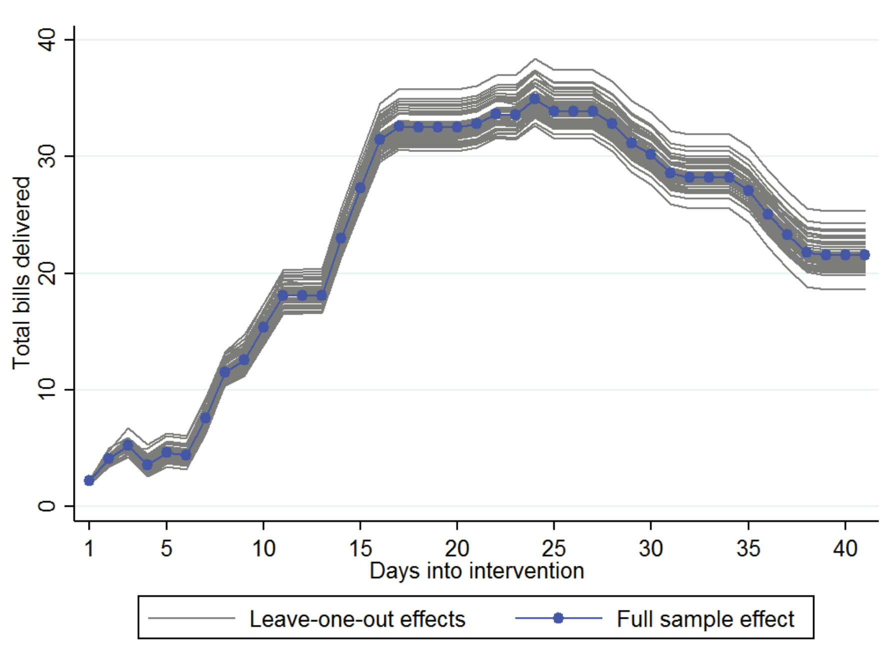
(b) Treatment Effect



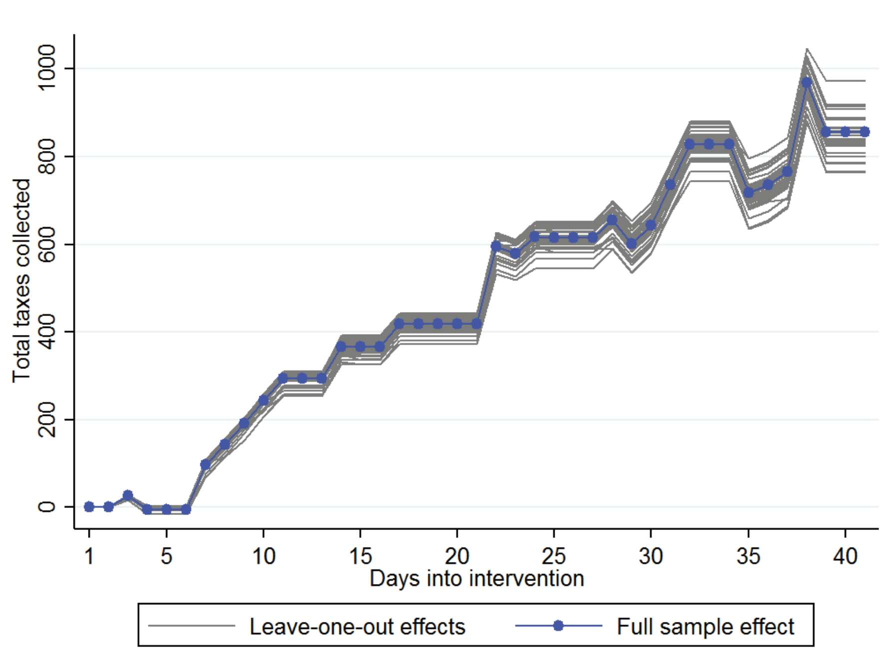
Notes: These panels show the impact of technology on the amount of taxes collected per bill delivered. Panel A shows the average amount of taxes collected per bill delivered by group (treatment, control) and by day of the intervention. Panel B displays the treatment effect coefficients on technology, separately by day, based on estimating equation (1). Coefficients together with the 95% confidence intervals are displayed. Standard errors are clustered at the collector-unit level. The analysis is based on the daily collector data, described in Section 3.1.

Figure A8: Robustness of Impacts to Leave-one-out Sample Restrictions

(a) Bills Delivered



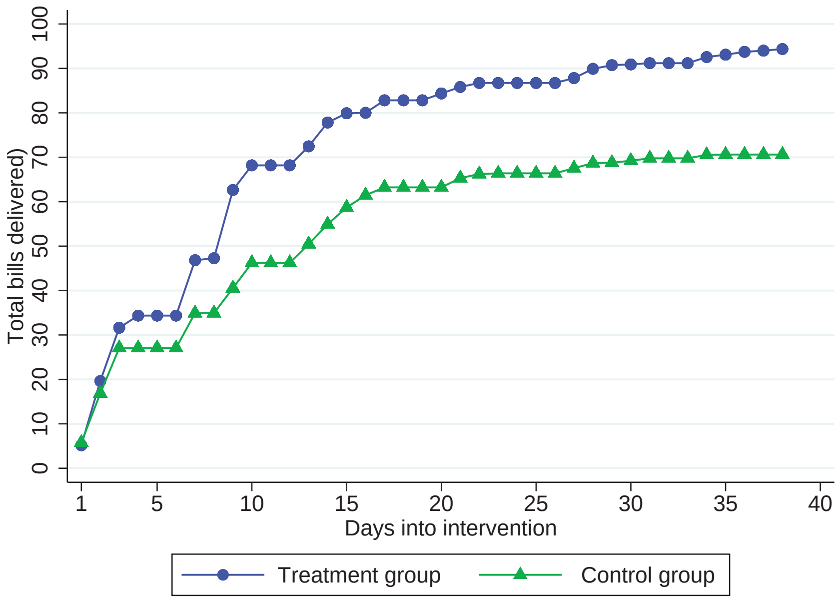
(b) Taxes Collected



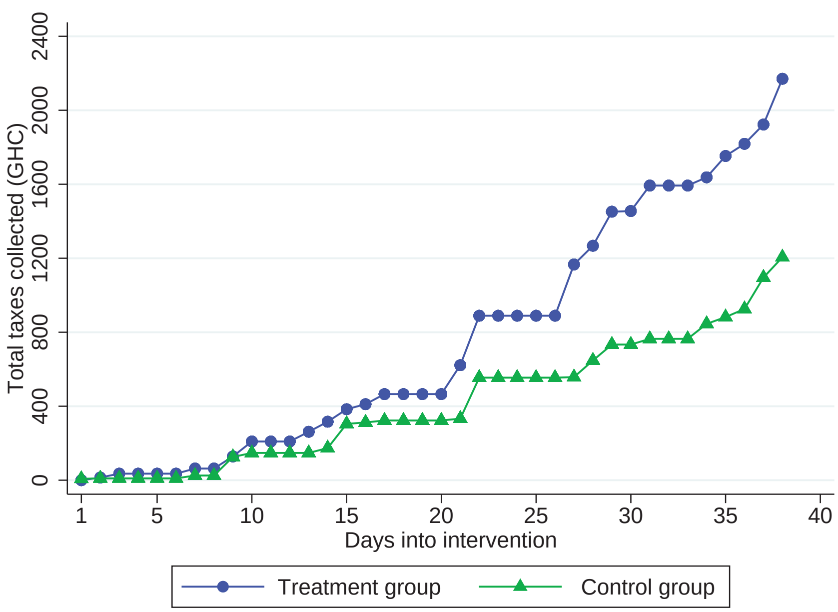
Notes: These panels show the robustness of technology impacts on total bills delivered (Panel A) and total taxes collected (Panel B). In both panels, the blue dotted line represents the dynamic treatment effect estimated in the full sample (Panel A of Figure A5 and Figure A6 for bills delivered and taxes collected, respectively). Each dark-gray line represents the dynamic treatment effects from estimating the same econometric model, but in individual sub-samples which remove one collector at a time. The analysis is based on the daily collector data, described in Section 3.1.

Figure A9: Results from Pilot Experiment

(a) Average Number of Bills Delivered



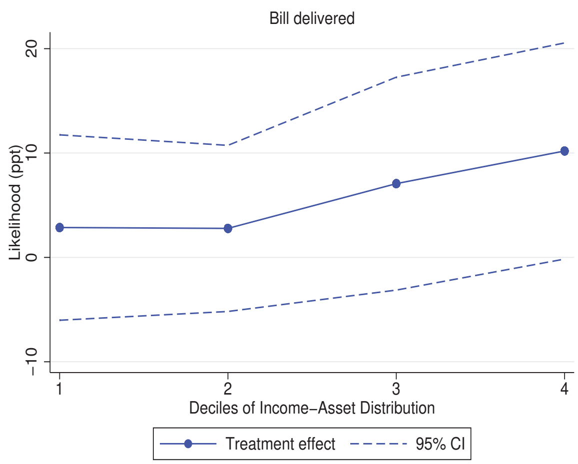
(b) Average Total Taxes Collected



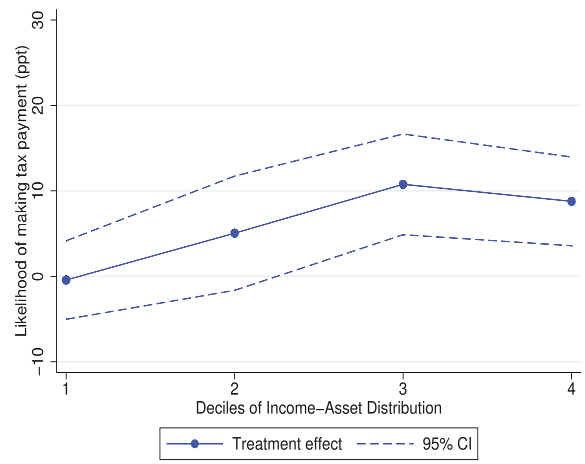
Notes: These panels show the impacts of technology on bills delivered and taxes collected based on the pilot experiment conducted in early 2019. The pilot was implemented in the same location as the main experiment, using the same technology, and following the same protocol for randomization and data-collection (see Section 3 for details). The pilot involved only 24 collectors and lasted 5 weeks, while the main experiment involves 56 collectors and lasts 6 weeks. These panels are constructed in the same way as in Figure 1. The treatment collectors had delivered 32% more bills at the end of the pilot experiment (compared to 27% at the end of the main experiment) and collected 79% more taxes (103%).

Figure A10: Robustness of Distributional Impacts to Additional Tax Measures

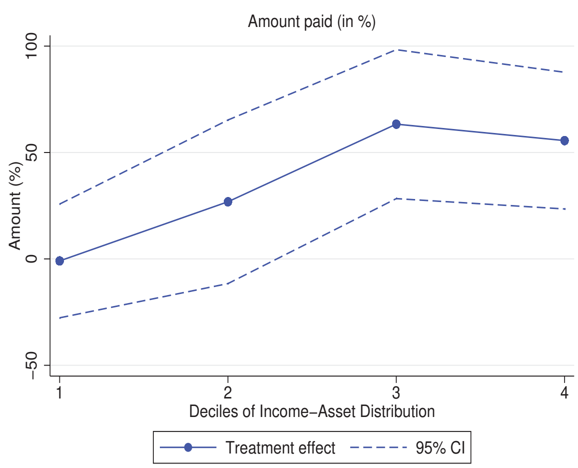
(a) **1**(Bill Delivered)



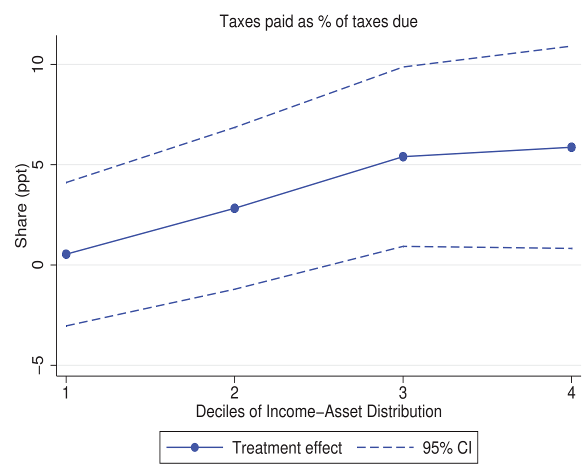
(b) **1**(Any Taxes Paid)



(c) Total Amount Paid (%)



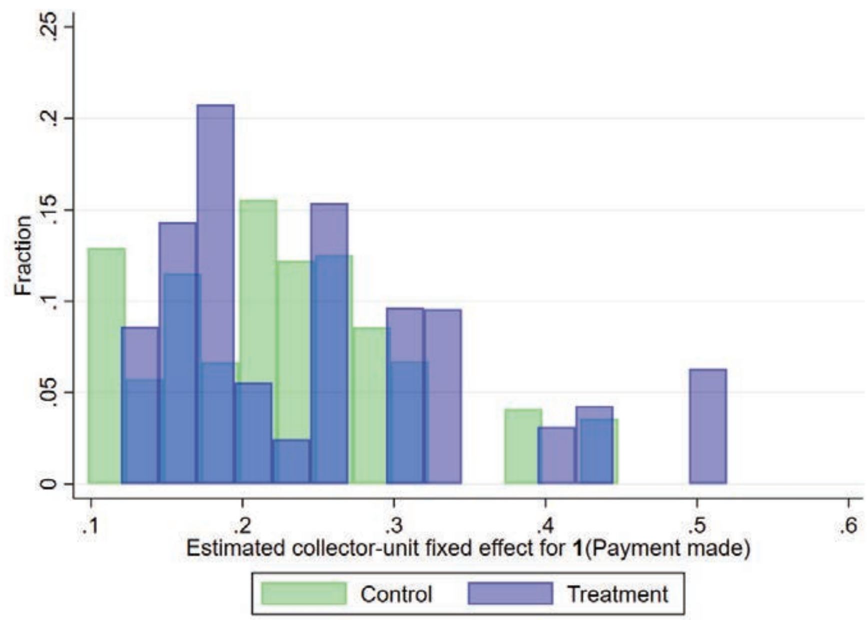
(d) Total Taxes Paid (% of Taxes Due)



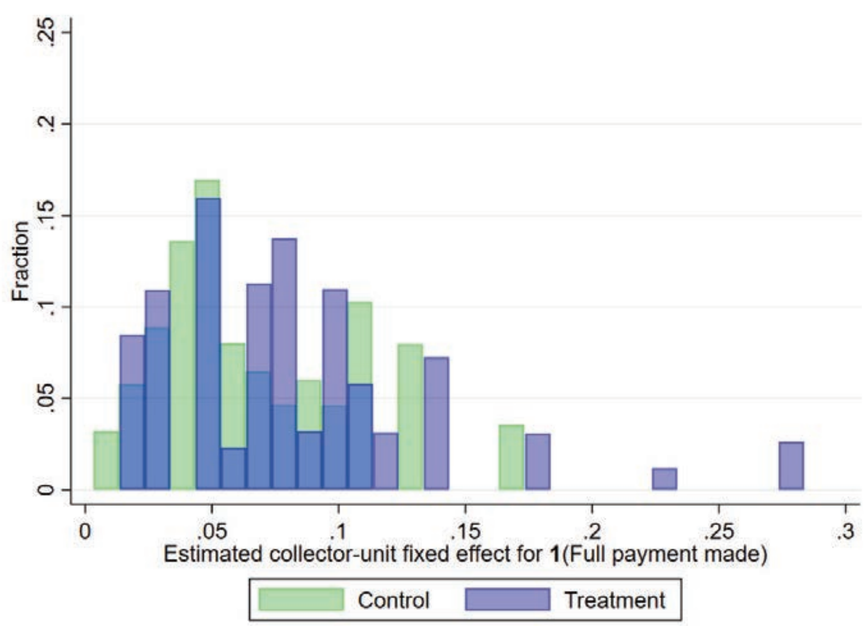
Notes: These panels show robustness for the distributional impact of technology on tax outcomes. The econometric model is the same as Figure A18, but the outcome varies across panels: a dummy for a bill delivered (Panel A); a dummy for any taxes paid (Panel B); amount of taxes paid, expressed as a percent using the inverse hyperbolic sine (Panel C); and, the amount of taxes paid, expressed as a percent of taxes due (Panel D). Coefficients together with the 95% confidence intervals are displayed. Standard errors are clustered at the collector-unit level. See Data Appendix B.2 for details on the variables.

Figure A11: Estimated Collector Fixed effects

(a) **1**(Any tax payment)



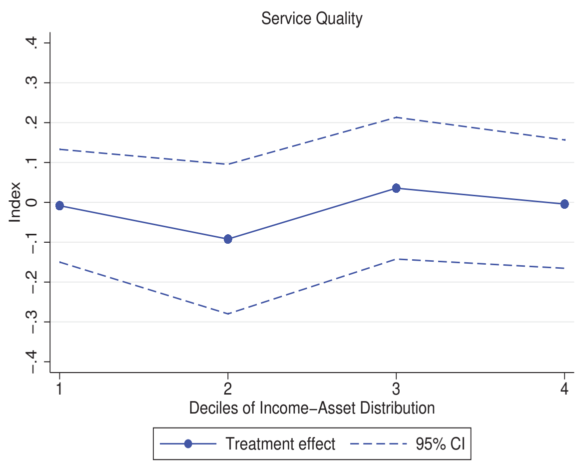
(b) **1**(Full tax payment)



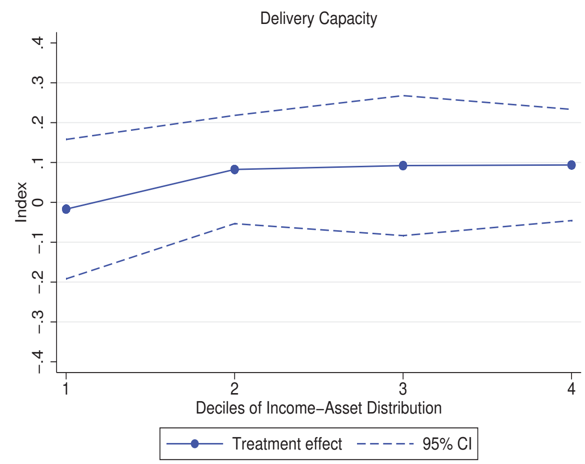
Notes: This figure is based on the household survey sample. The outcome in panel A is a dummy for any tax payment, and the outcome in panel B is a dummy for full tax payment. In panel A, we regress the outcome on the set of collector-unit fixed effects. In turn, we plot the density distribution of these estimated fixed effects, distinguishing between collector-units in the treatment group versus the control group. The same process is repeated in panel B for the full tax payment outcome. Collectors were randomly assigned to collection units, and each collector-unit pair was subsequently randomly assigned to the treatment or control group. See Data Appendix B.2 for details on the variables.

Figure A12: Distributional Effects on Beliefs about Government Capacity and Tax Morale

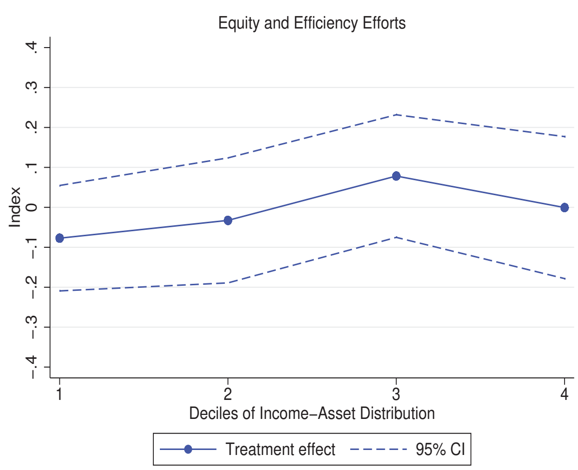
(a) Satisfaction with Gov't Services



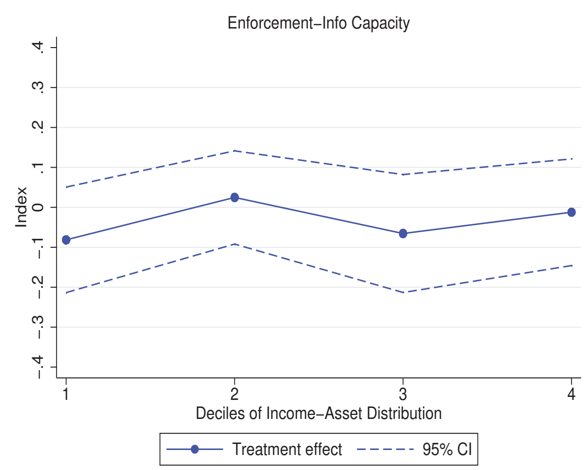
(b) Gov't Integrity and Competency



(c) Equity and Efficiency of Tax Collection



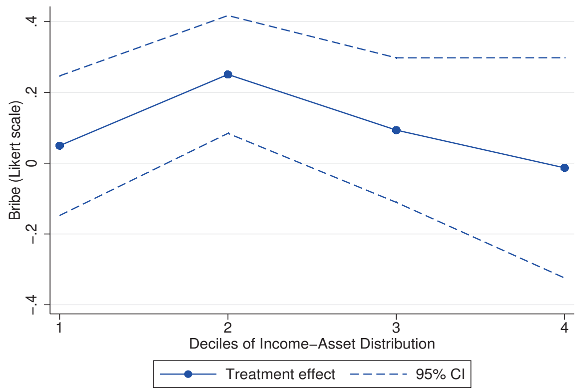
(d) Enforcement and Information Capacity



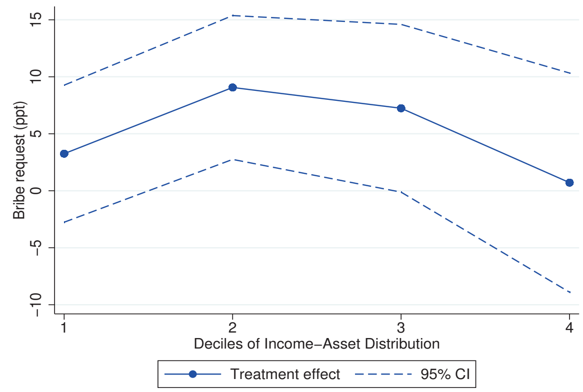
Notes: These panels investigate distributional impacts of technology on household beliefs about government capacity and tax morale. The econometric model is the same as in Figure A18, but the outcome varies across panels. The four panels study four indices: satisfaction with the quality of government services (Panel A); competency and integrity of local government (Panel B); government efforts to improve the efficiency and equity of the collection process (Panel C); the enforcement and information capacity of the local government (Panel D). These indices are the outcomes in Table 2. The income-asset distribution is calculated as the unweighted average, by household, of the income index and the assets index. Coefficients together with the 95% confidence intervals are displayed. Standard errors are clustered at the collector-unit level. For a detailed description of the different indices, see Data Appendix B.3-B.5.

Figure A13: Robustness of Distributional Impacts to Additional Bribe Measures

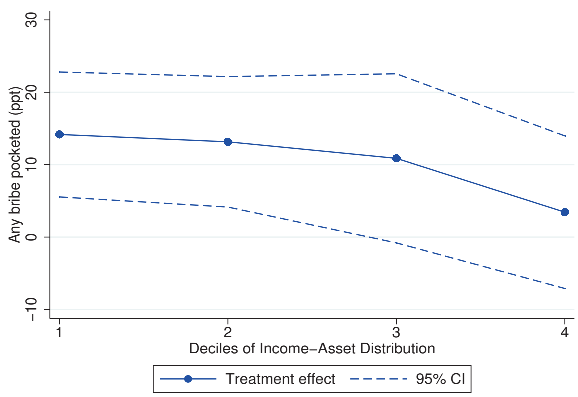
(a) Collusive Bribe (Likert Scale)



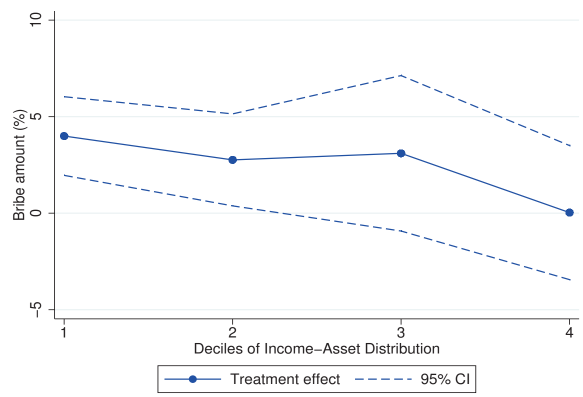
(b) **1**(Any Collusive Bribe)



(c) **1**(Any Coercive Bribe)

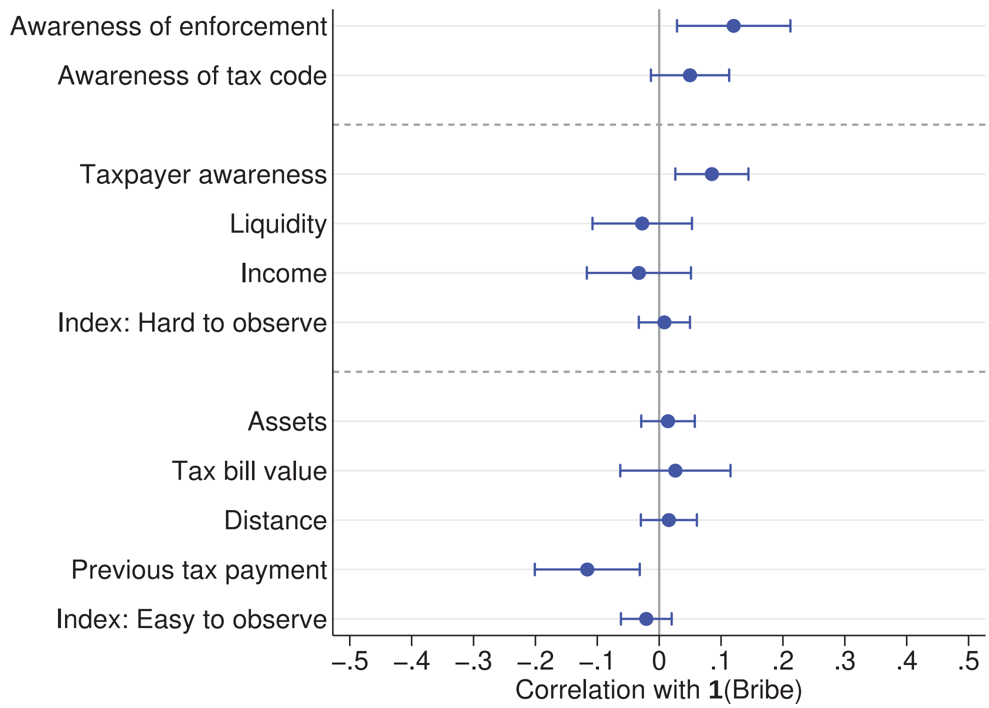


(d) Total Bribe Amount (%)



Notes: The econometric model is the same as in Figure A18, but the outcome varies across panels. In Panel A, the outcome is the likelihood (on a scale from 1 to 5, where 1 is 'very unlikely' and 5 is 'very likely') estimated by the household that a local collector will solicit any unofficial payment while conducting visits with property owners. In Panel B, the outcome is a dummy variable taking a value of 1 if the answer to the question in Panel A is 'maybe', 'somewhat likely' or 'very likely', and 0 if the answer is 'not very likely' or 'very unlikely'. In Panel C, the outcome is a dummy variable which takes a value of 1 if the household estimates that the local tax collector will pocket any strictly positive amount out of a hypothetical 1000 GHC collected from property owners. In Panel D, the outcome is the total bribe amount, in percent. This is calculated at the household level as the average of the coercive bribe amount, expressed as a percent of a hypothetical 1000 GHC collected by the tax collector, and the collusive bribe amount, expressed as a percent of the household's true tax liability. Coefficients together with the 95% confidence intervals are displayed. Standard errors are clustered at the collector-unit level. For additional description of the outcomes, see Data Appendix B.4.

Figure A14: Characteristics of Households Targeted for Bribes in the Treatment Group



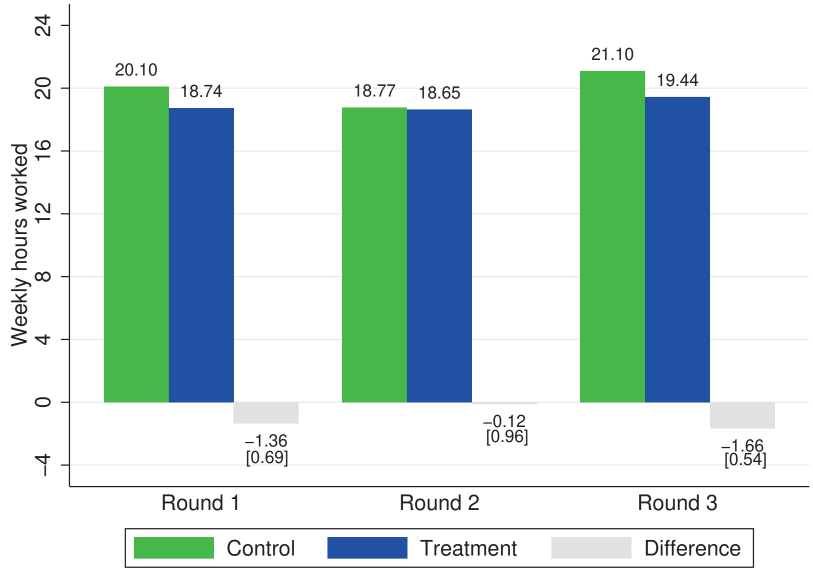
Notes: This panel shows the selection on bribe incidence for fixed household characteristics. The econometric model is similar to equation (3), except that the dummy for tax payment is replaced with a dummy for any bribe incidence (same as in Panel B of Figure A18). Moreover, the analysis is limited to treatment areas, where there was an overall increase in bribe incidence (Table 2). Formally, we estimate

$$y_{hc} = \theta \cdot \mathbf{1}(Bribe)_h + \Omega \cdot X_h + \mu_c + \epsilon_{hc}$$

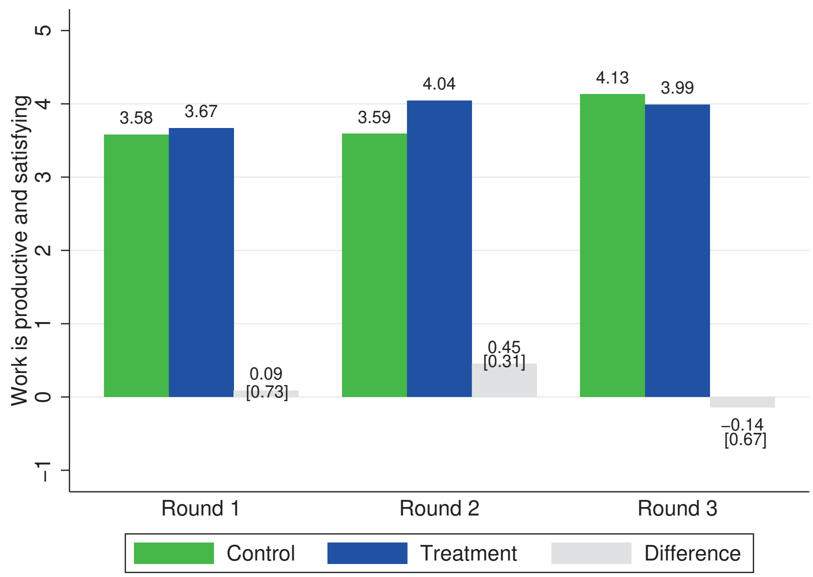
The fixed household characteristics y_{hc} are the same as those described in Figure 4. In addition, the top panel reports two additional characteristics which measure awareness about enforcement and awareness about the tax code. Coefficients together with the 95% confidence intervals are displayed. Standard errors are clustered at the collector-unit level. For a detailed description of the variables, see Data Appendix B.4-B.5.

Figure A15: Collector Hours Worked and Job Satisfaction

(a) Hours Worked per Week



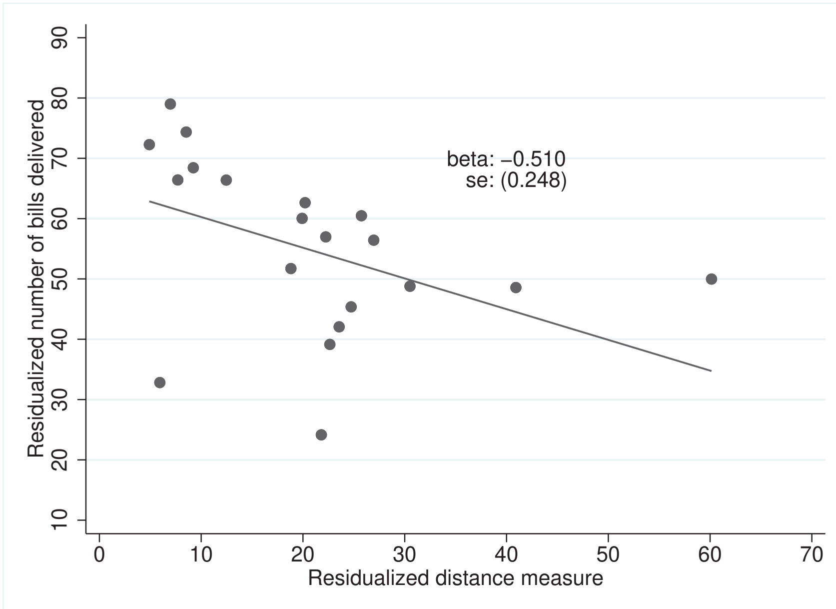
(b) Job Satisfaction



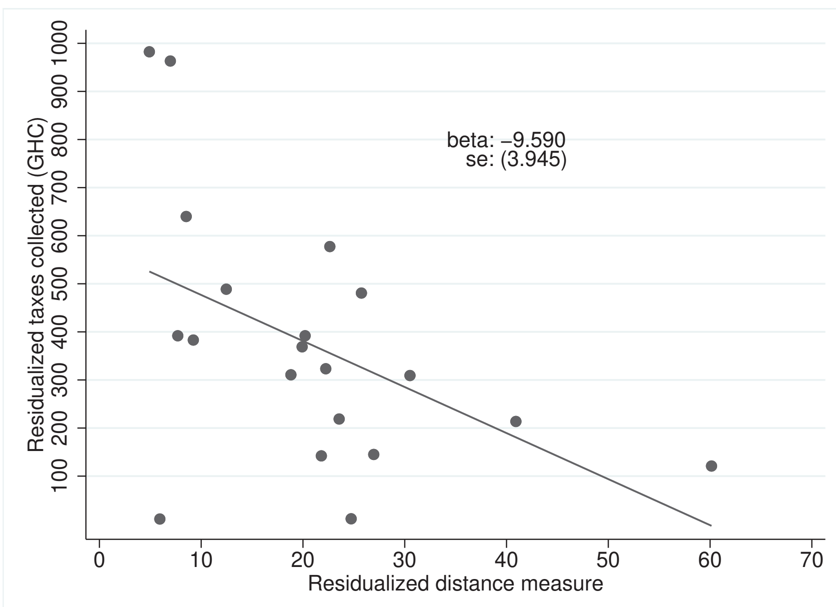
Notes: This figure shows results on collectors' hours worked and job satisfaction, based on equation (1). Round 1, 2 and 3 correspond to the baseline, mid-line and endline collector survey rounds, respectively. In panel A, the outcome is the collector's self-reported number of hours worked per week. The grey bar measures the difference in outcome between the treatment and control groups; the number in brackets is the randomization inference-based p-value on the statistical significance of the difference. In panel B, the outcome is the average job satisfaction, which combines measures of how much the collector, during the past week, agrees that: their work was productive; they were content while working; and, they were satisfied with their job. The indexed transformation of the job satisfaction variable is used in other parts of the analysis (e.g. Table 3). For a detailed description of the variables, see Data Appendix B.6. The analysis is based on the balanced sample of collector surveys (Section 3.1); results based on the unbalanced sample are in Table A6.

Figure A16: Size of Collection Unit and Tax Outcomes in the Control Group

(a) Bills Delivered



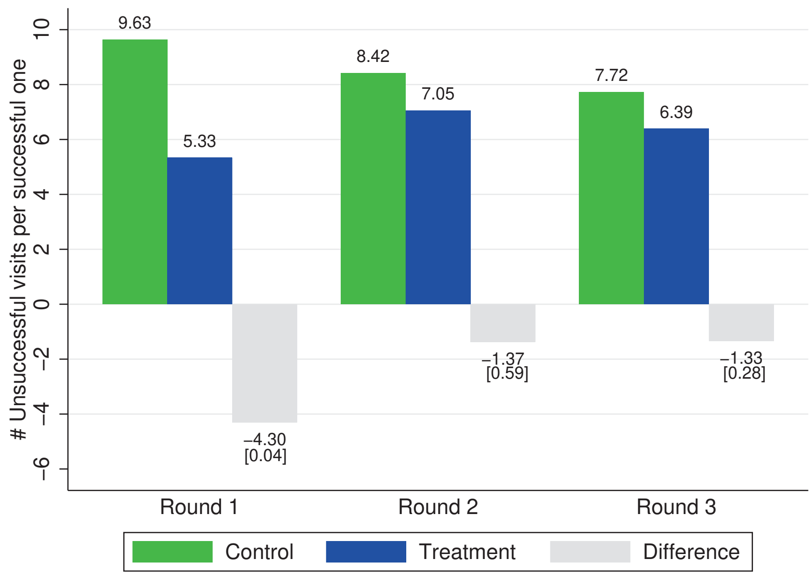
(b) Taxes Collected (GHC)



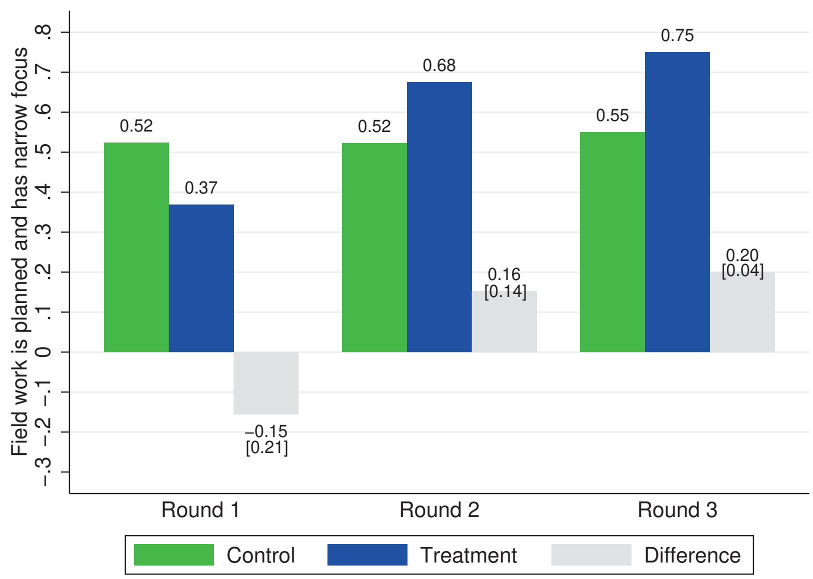
Notes: These figures show associations between the size of the collection unit and tax outcomes in the control group. In each collection unit, the distance is measured as the shortest total distance that is required to visit every property (once) that the collector is assigned to. The starting point is the location of the local government office (results are similar if we instead use the geographical center of each collection unit as the starting point). Panel A and panel B show the conditional association between the distance measure and bills delivered and taxes collected, respectively. Bills delivered and taxes collected are measured in the daily collector data. All variables are first regressed on the total number of bills assigned to the collection unit and campaign-day fixed effects. In turn, twenty equal sized bins of the residualized distance measure are created and the dots represent the average residualized bills delivered (panel A) or taxes collected (panel B) by bin. In each panel, the line of best linear fit based on the underlying collection unit-day data is shown, with the corresponding OLS slope coefficient and standard error reported.

Figure A17: Collector Field Work Organization

(a) # of Failed Attempts per Successful Visit



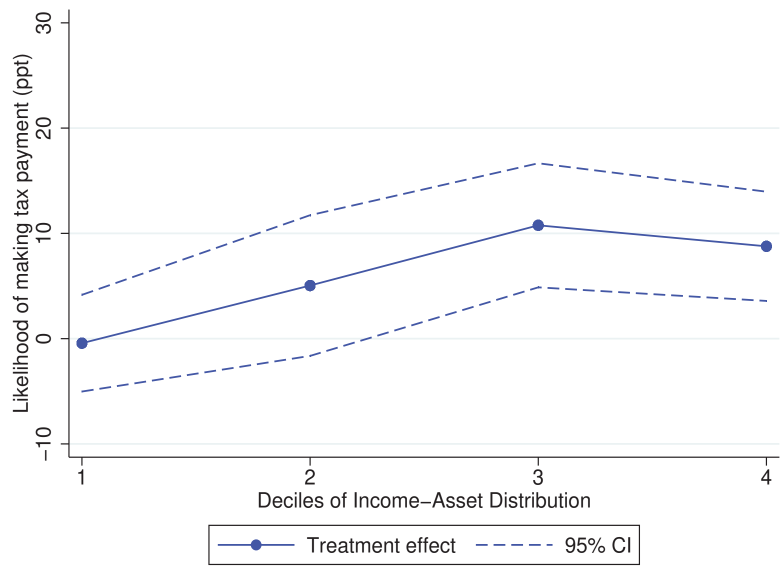
(b) Field Work is Organized and Focus is Narrow



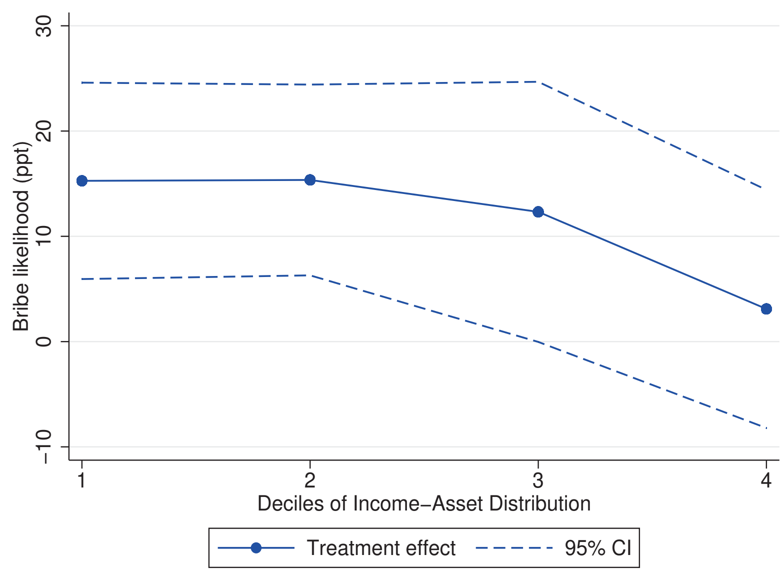
Notes: This figure shows results on collectors' field work, based on equation (1). Round 1, 2 and 3 correspond to the baseline, mid-line and endline collector survey rounds, respectively. In panel A, the outcome is the collector's self-reported number of failed attempts per successful visit. The grey bar measures the difference in outcome between the treatment and control groups; the number in parentheses is the randomization inference-based p-value on the statistical significance of the difference. In panel B, the outcome is the average of two dummy variables. The first dummy takes a value of 1 if the collector reports that they focus on specific properties in their unit or 0 if they instead report working in broad areas throughout their unit. The second dummy takes a value of 1 if the collector reports that they organize properties to visit before going into the field each day or 0 if they instead report going into the field directly. For a detailed description of the variables, see Data Appendix B.6. The analysis is based on the balanced sample of collector surveys (Section 3.1); results based on the unbalanced sample are in Table A6.

Figure A18: Distributional Effects of Technology on Taxes and Bribes

(a) Treatment Effect on Tax Payment



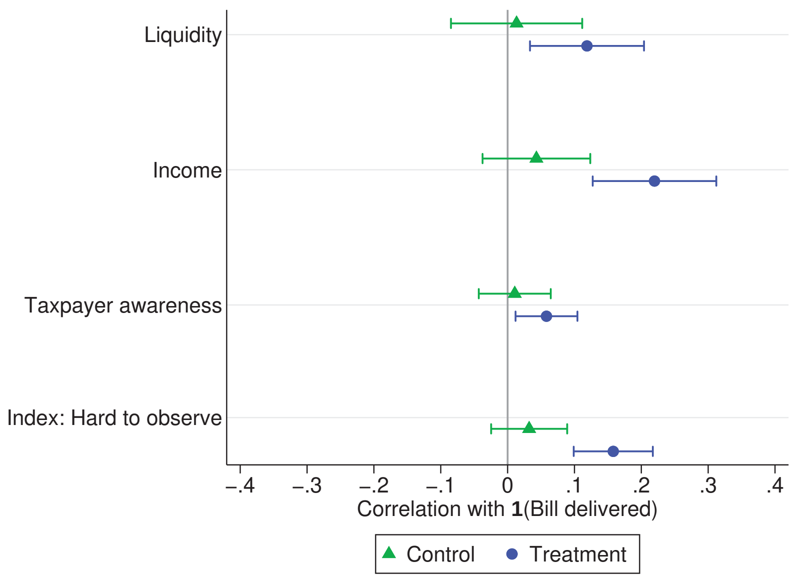
(b) Treatment Effect on Bribe



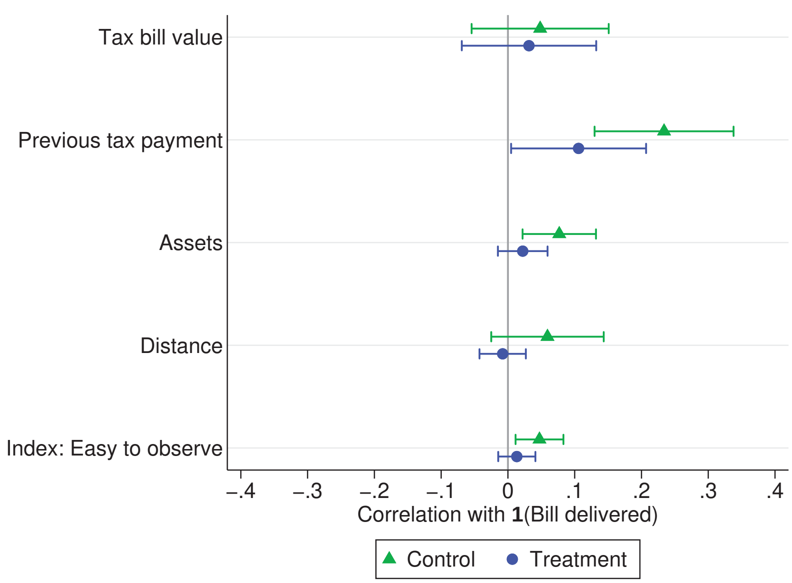
Notes: These panels show the impact of technology on the likelihood of making any positive tax payment (Panel A) and on the likelihood of bribes (Panel B). The bribe dummy takes a value of 1 if the household estimates that the tax collector will ask for any positive amount of unofficial payment during interactions with property owners (collusive bribe); or, if the household estimates that the collector will keep for themselves any positive amount of money collected from property owners (coercive bribe). The bribe variable takes a value of 0 only if the estimated amounts of collusive bribe and coercive bribe are both equal be 0. Both panels display the treatment effect coefficient on technology, separately by quartile of the income-asset distribution, based on estimating equation (2) augmented with a set of interactions between the treatment assignment $\mathbf{1}(Tech)_c$ and dummies for quartiles of the household income-asset distribution. For more details on the measures and the bribe and tax payment measures, see Data Appendix B.2-B.4. Coefficients together with the 95% confidence intervals are displayed. Standard errors are clustered at the collector-unit level. Results are robust to alternative tax and bribe measures (Figure A10 and A13).

Figure A19: Characteristics of Households that Received a Bill by Treatment Status

(a) Hard-to-Observe Characteristics



(b) Easy-to-Observe Characteristics

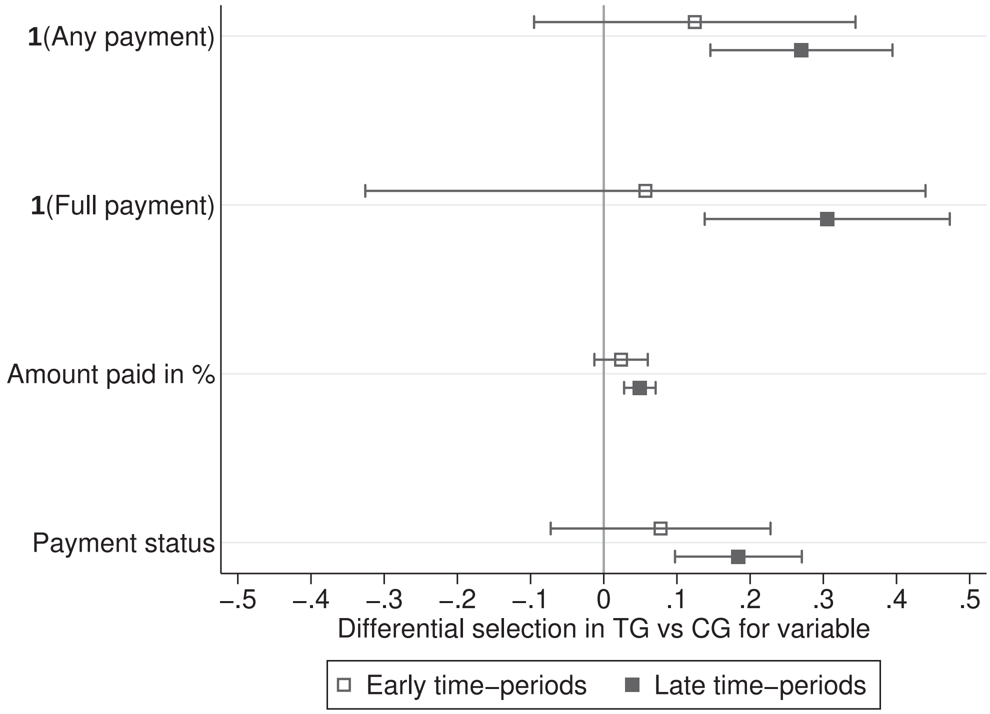


Notes: These panels show the selection on bill delivery for fixed household characteristics. The characteristics are the same as in Figure 4. The econometric model is the same as equation (3), except that the dummy for tax payment is replaced with a dummy for bill delivery. Formally, we estimate

$$y_{hc} = \theta \cdot \mathbf{1}(Billdelivered)_h + \beta \cdot [\mathbf{1}(Billdelivered)_h * \mathbf{1}(Tech)_c] + \Omega \cdot X_h + \mu_c + \epsilon_{hc}$$

For a detailed description of the household characteristics and the indices, see Data Appendix B.5. Coefficients together with the 95% confidence intervals are displayed. Standard errors are clustered at the collector-unit level.

Figure A20: Selection in Early versus Late Periods of Tax Campaign



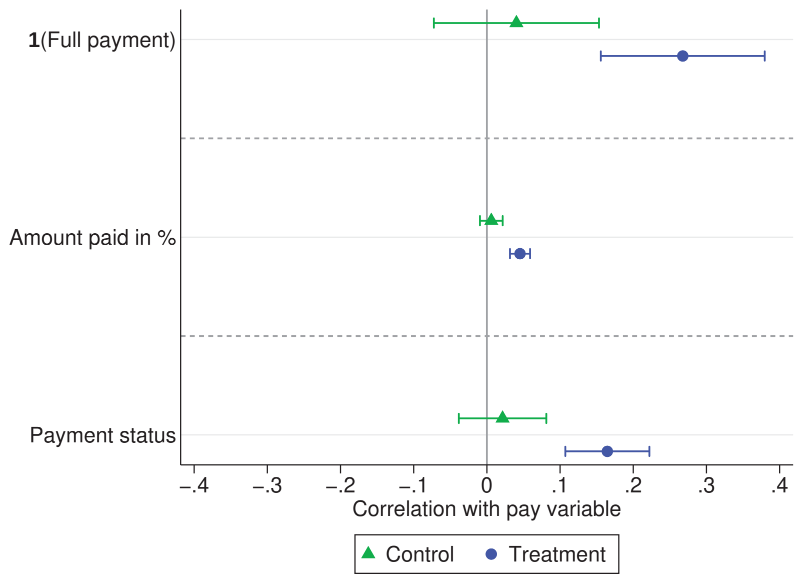
Notes: This figure shows targeting of the household hard-to-observe index for tax outcomes, based on estimating an augmented version of equation (3) which distinguishes between payments made early versus late in the tax campaign. Specifically, we estimate:

$$y_{hc} = \theta^E \cdot \mathbf{1}(Pay:e)_h + \theta^L \cdot \mathbf{1}(Pay:l)_h + \beta^E \cdot [\mathbf{1}(Pay:e)_h \cdot \mathbf{1}(Tech)_c] + \beta^L \cdot [\mathbf{1}(Pay:l)_h \cdot \mathbf{1}(Tech)_c] + \Omega \cdot X_h + \mu_c + \epsilon_{hc}$$

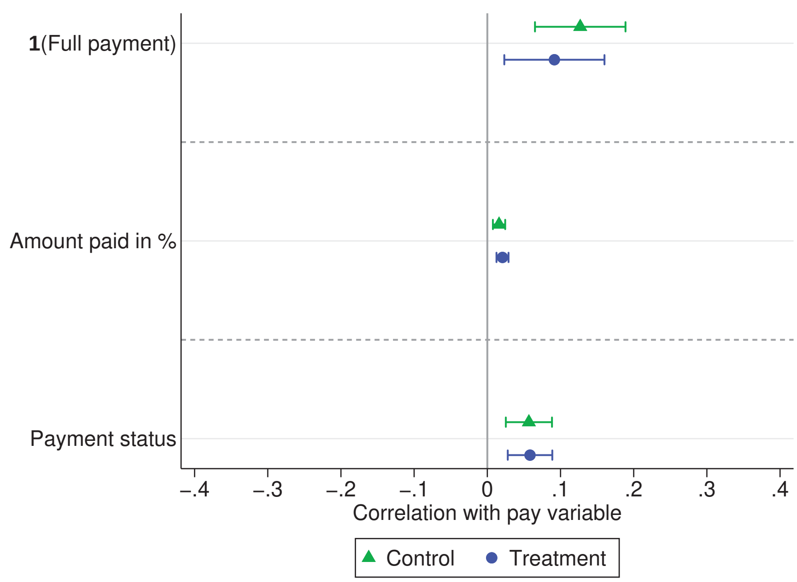
where $\mathbf{1}(Pay:e)_h$ takes a value of 1 if the household made a payment in the early campaign days (and 0 otherwise), and $\mathbf{1}(Pay:l)_h$ takes a value of 1 if the household made a payment in the later campaign days (and 0 otherwise). The figures display the coefficients β^E and β^L, with the 95% confidence intervals. The outcome y_{hc} is the hard-to-observe index. Across rows, the tax payment variable differs: a dummy equal to 1 if there is any tax payment, and 0 otherwise; a dummy equal to 1 if there is a tax payment in full, and 0 otherwise; total amount paid in %; payment status. The payment status variable equals: 1 if the household made no payment; 2 if the household made a partial payment; 3 if the household made a full payment. Coefficients together with the 95% confidence intervals are displayed. Standard errors are clustered at the collector-unit level. See Data Appendix B.2 for details on the variables.

Figure A21: Selection: Robustness of Tax Outcome

(a) Hard-to-Observe Index



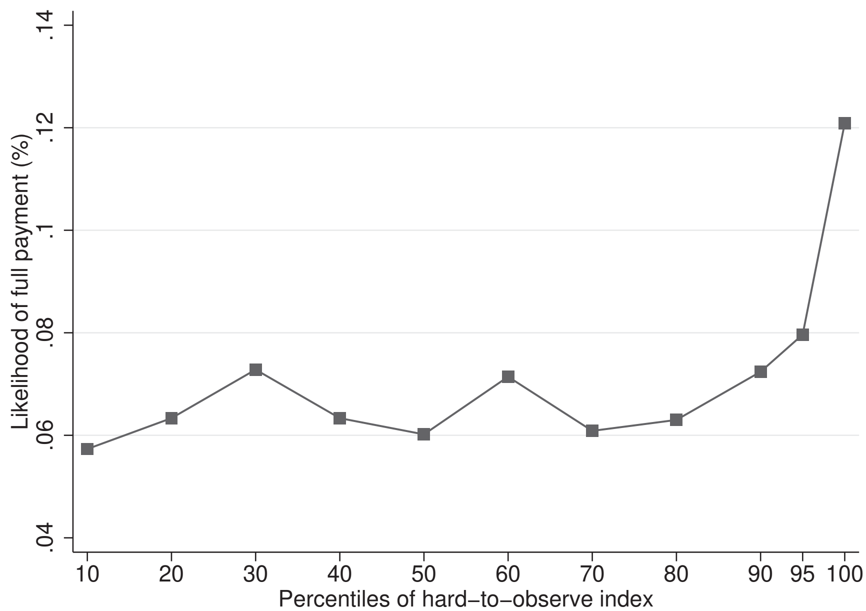
(b) Easy-to-Observe Index



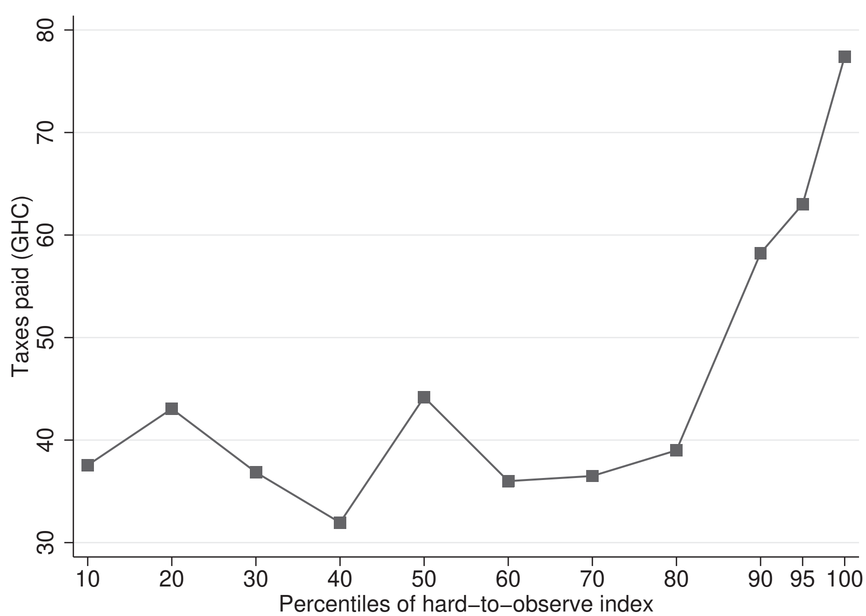
Notes: These figures show targeting of property owner characteristics for tax outcomes, based on estimating equation (3). The tax payment variable varies by row: a dummy which equals 1 if the household pays the full amount due; total amount paid in %, using the inverse hyperbolic sine transformation; payment status. The payment status variable equals: 1 if the household made no payment; 2 if the household made a partial payment; 3 if the household made a full payment. Panel A shows the associations between these pay variables and the hard to observe index, separately in the treatment and control groups. Panel B shows the associations between these pay variables and the easy to observe index, separately in the treatment and control groups. Coefficients together with the 95% confidence intervals are displayed. Standard errors are clustered at the collector-unit level. See Data Appendix B.2-B.5 for details on the variables.

Figure A22: Tax Outcomes by Values of the Hard-to-Observe Index

(a) Likelihood of Tax Payment

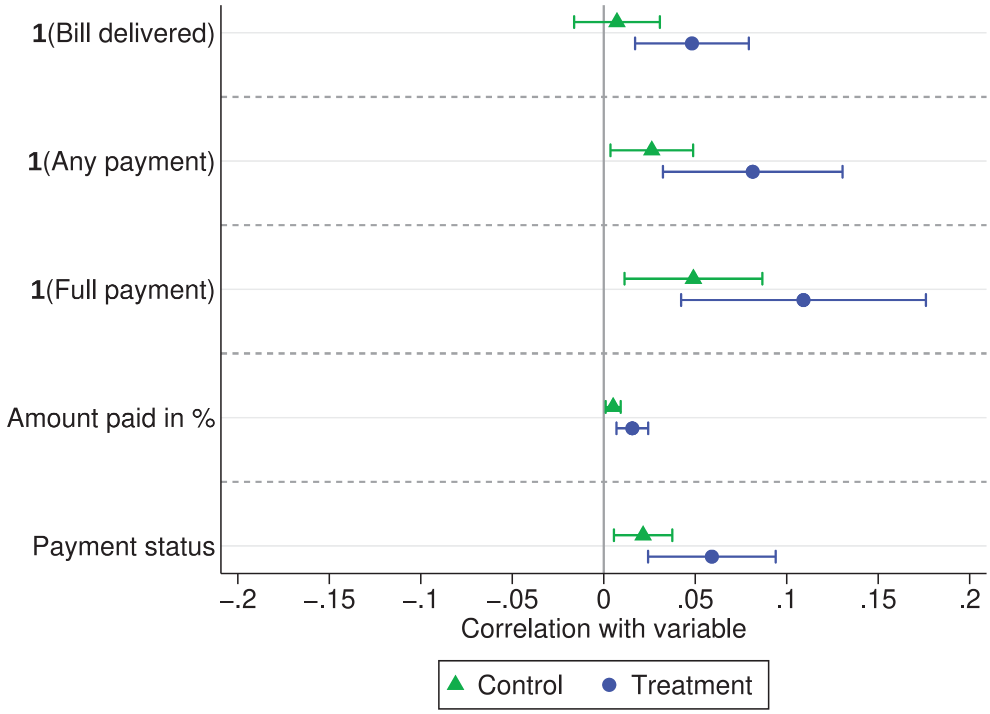


(b) Amount of Taxes Paid (GHC)



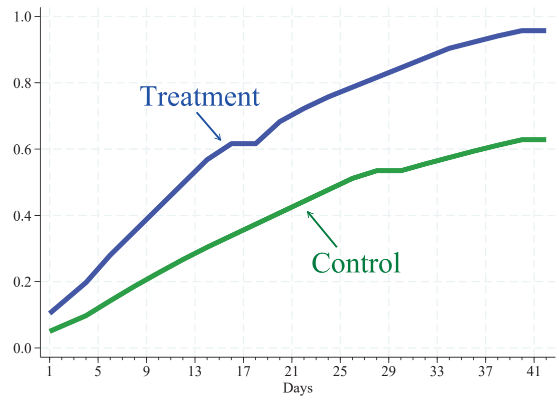
Notes: These figures are constructed using the household survey sample in control collection units. In panel A, the outcome variable is a dummy which takes a value of 1 if the household made any tax payment, and 0 otherwise. The hard-to-observe index is partitioned into 100 percentiles; in turn, the figure shows the average value of the dummy variable in each decile, as well as the in $90-95^{th}$ percentile range and the $95-100^{th}$ percentile range. In panel B, the figure is created in the same way, except that the outcome is the amount of taxes paid (in GHC). See Data Appendix B.2 for details on the variables.

Figure A23: Selection: Robustness with Discrete Measure of Hard-to-Observe Index

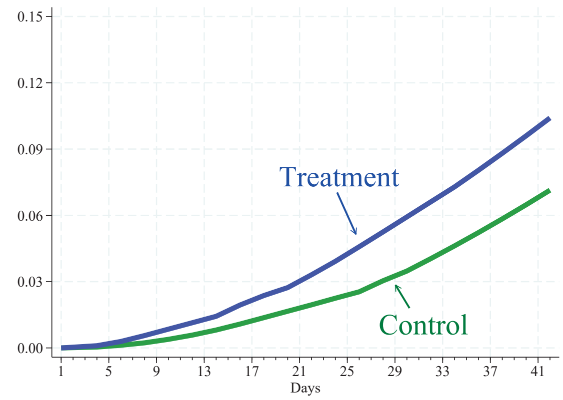


Notes: This figure shows targeting of the household hard-to-observe index for tax outcomes, based on estimating equation (3). The outcome is a dummy variable which equals 1 if the value of the hard-to-observe index is in the top 5 percentile of the index distribution, and 0 otherwise. The tax variable varies by row: a dummy equal to 1 if the tax bill is delivered, and 0 otherwise; a dummy equal to 1 if there is any tax payment, and 0 otherwise; a dummy equal to 1 if there is a tax payment in full, and 0 otherwise; total amount paid in %; payment status. The payment status variable equals: 1 if the household made no payment; 2 if the household made a partial payment; 3 if the household made a full payment. The figure shows the associations between these tax variables and the high-type dummy variable, separately in the treatment and control groups. Coefficients together with the 95% confidence intervals are displayed. Standard errors are clustered at the collector-unit level. See Data Appendix B.2-B.5 for details on the variables.

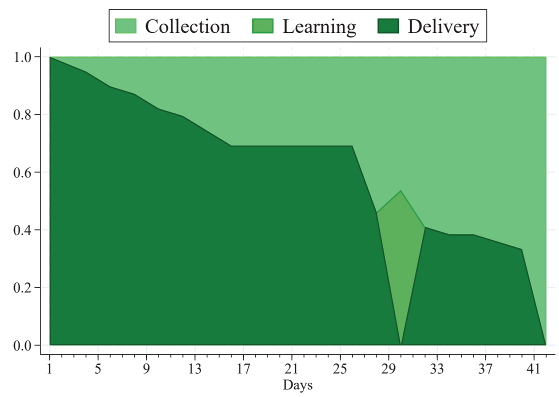
Figure A24: Predictions of Model Counterfactual with No Re-optimization



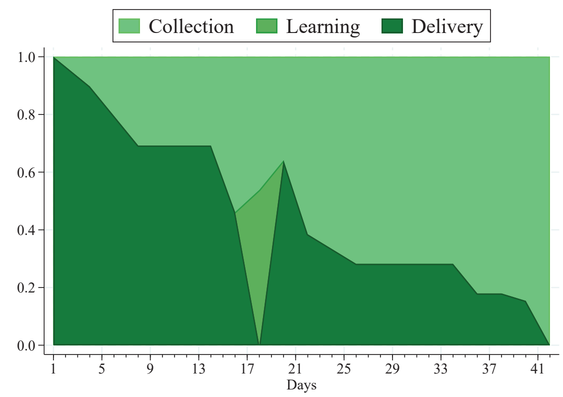
(a) Bills delivered



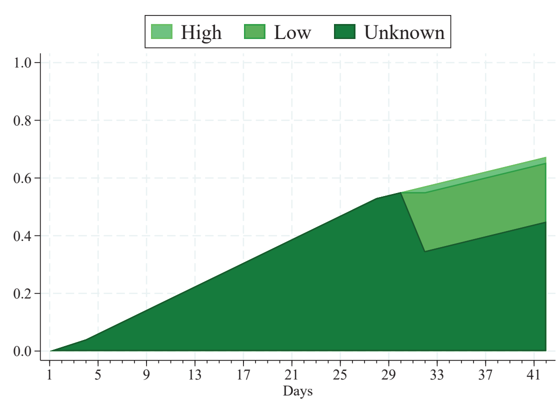
(b) Revenue Collections



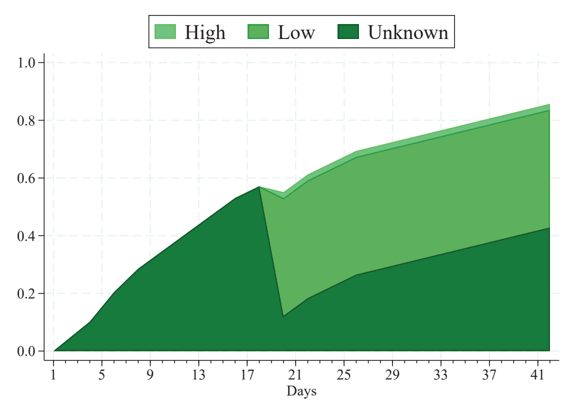
(c) Time Allocation (Control)



(d) Time Allocation (Treatment)

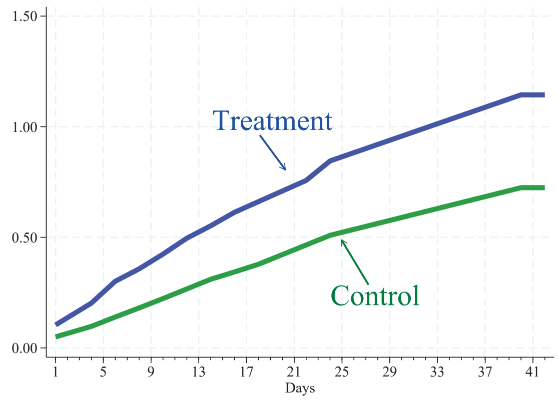


(e) Composition of Bills Delivered (Control)

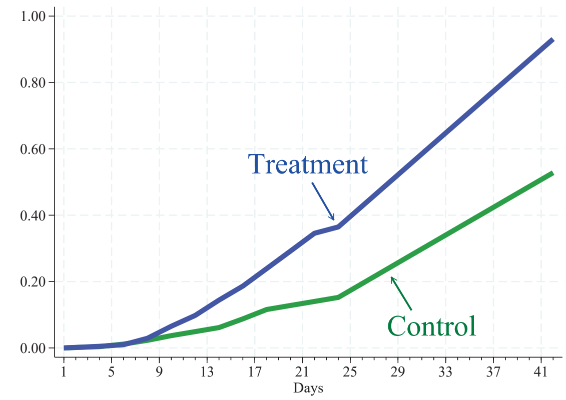


(f) Composition of Bills Delivered (Treatment)

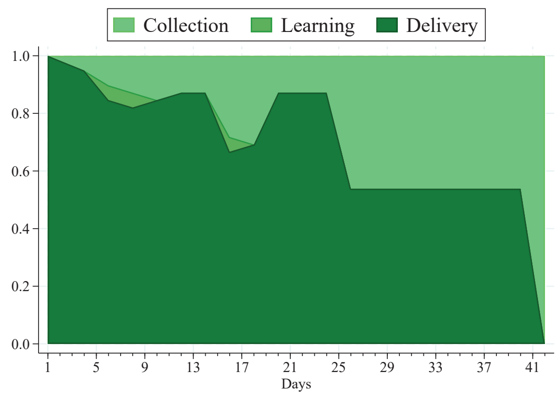
Figure A25: Predictions of Model Counterfactual with Higher Payment Probabilities



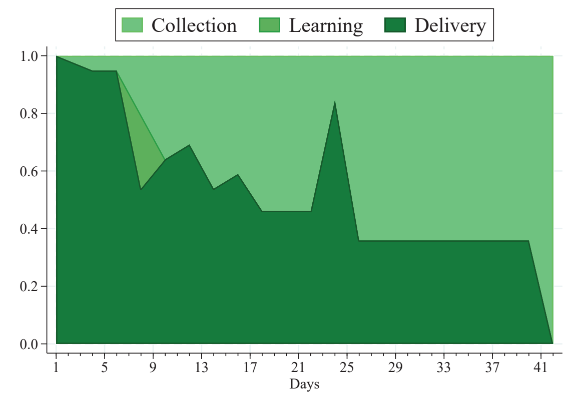
(a) Bills delivered



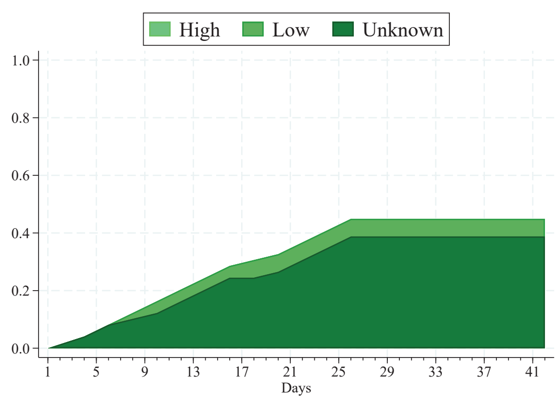
(b) Revenue Collections



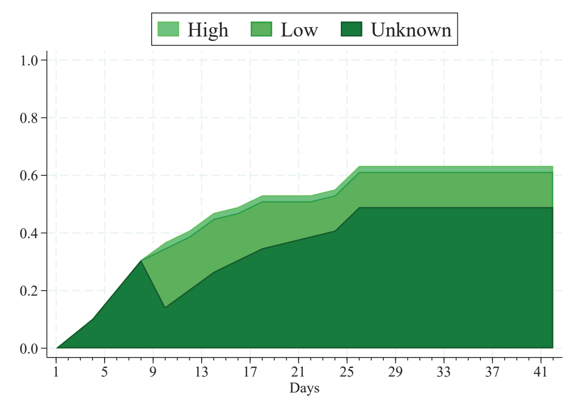
(c) Time Allocation (Control)



(d) Time Allocation (Treatment)

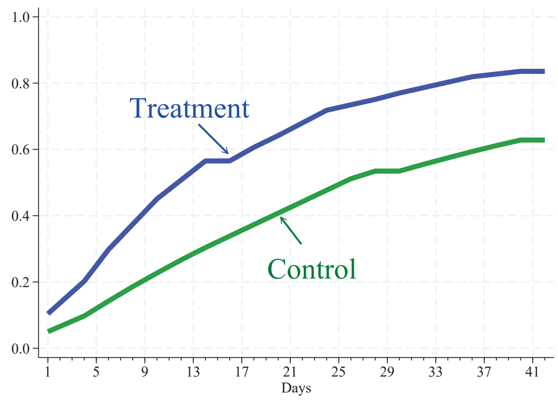


(e) Composition of Bills Delivered (Control)

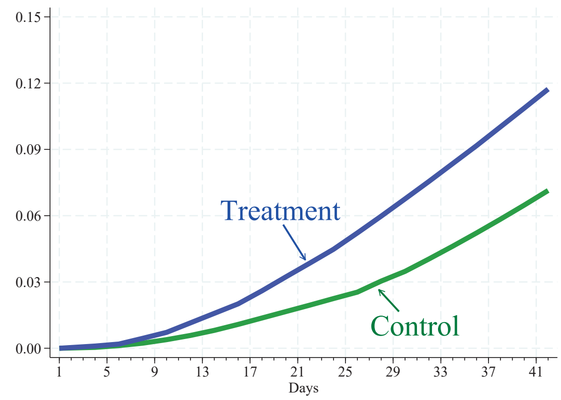


(f) Composition of Bills Delivered (Treatment)

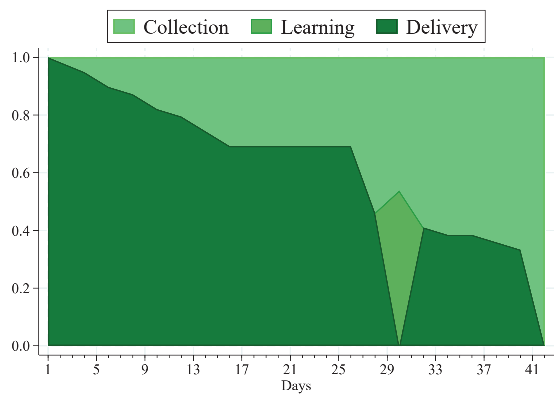
Figure A26: Predictions of Model Counterfactual with No Learning Advantage



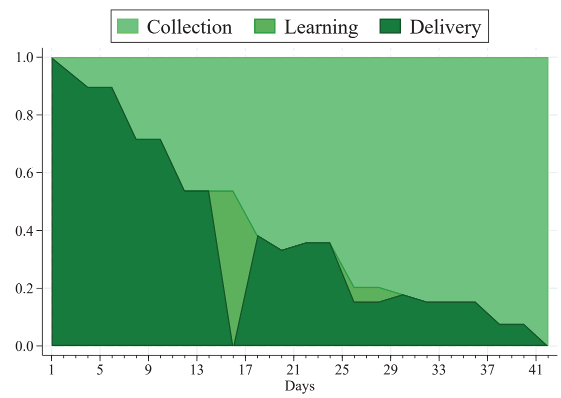
(a) Bills delivered



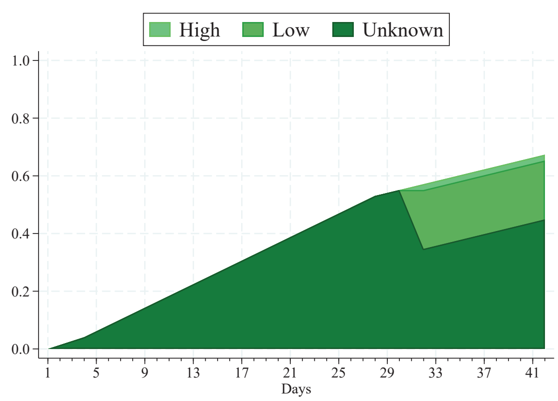
(b) Revenue Collections



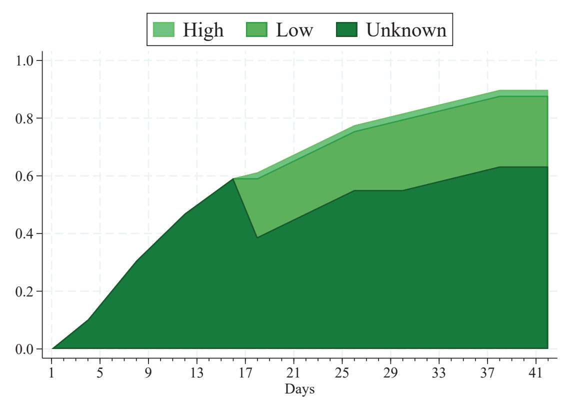
(c) Time Allocation (Control)



(d) Time Allocation (Treatment)

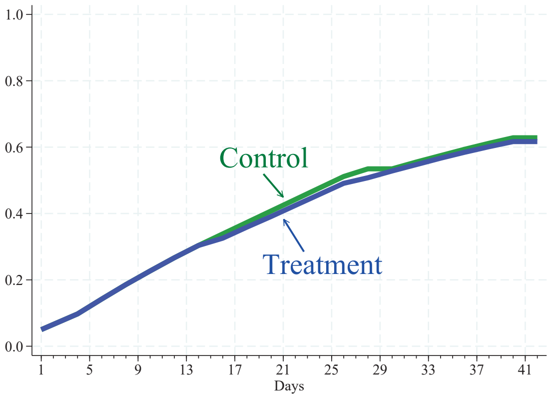


(e) Composition of Bills Delivered (Control)

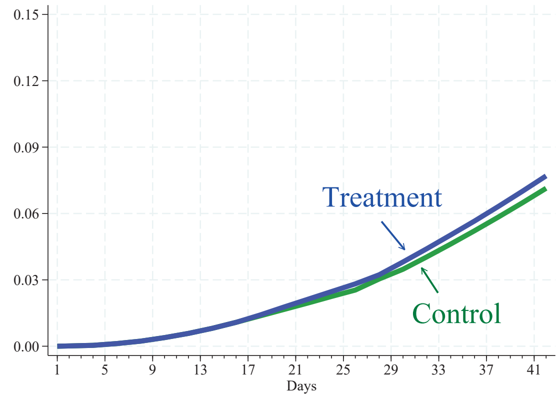


(f) Composition of Bills Delivered (Treatment)

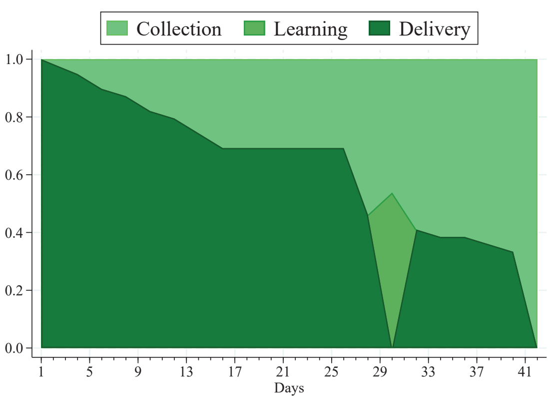
Figure A27: Predictions of Model Counterfactual with No Delivery Advantage



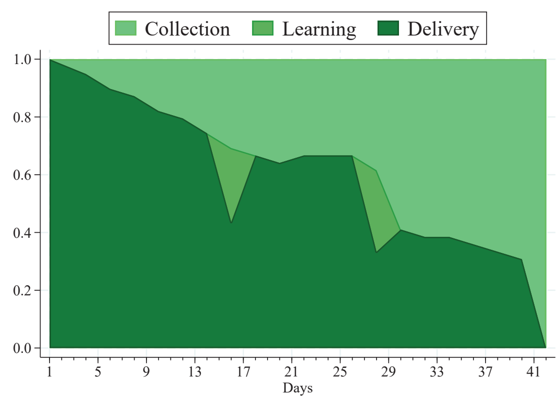
(a) Bills delivered



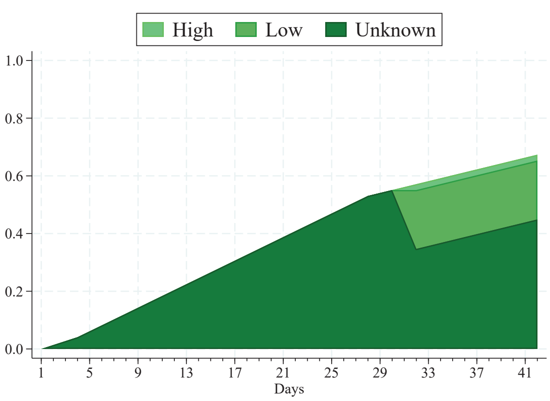
(b) Revenue Collections



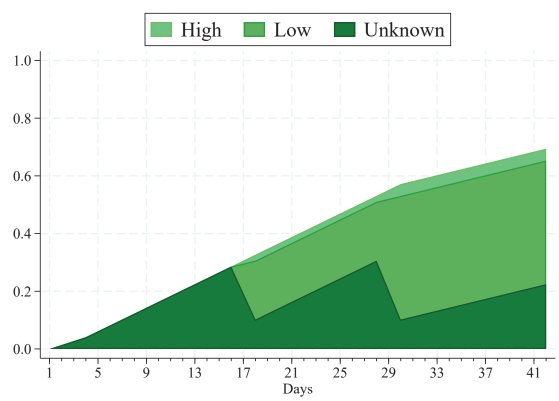
(c) Time Allocation (Control)



(d) Time Allocation (Treatment)



(e) Composition of Bills Delivered (Control)



(f) Composition of Bills Delivered (Treatment)

B Data Appendix

B.1 Variables from Census of Local Governments

- *Share of bills delivered (%)* This variable is the answer to the question "Considering all the properties in your district, approximately what percent were sent a bill this year?" The answer ranges from 0% to 100%.

- *Taxes collected per bill delivered (GHC)* This variable divides the total taxes collected per capita (in Ghanaian Cedi) by the variable *share of bills delivered*.

- *Share of bills that are paid (%)* This variable is based on the answer to the question "Cumulatively, what share of bills are paid by the end of the year?". This answer is asked separately for business property taxes and for resident property taxes. We construct the district-level variable as the unweighted average over the responses for businesses and residents.

- *Share of properties with address (%)* This variable is the answer to the question "Approximately what percent of the properties in your assembly have an official address assigned to them?". The answer ranges from 0% to 100%.

- *Common to not locate property* This variable takes a value of 1 (0) if the respondent answers 'Yes' ('No') to the question "When delivering bills, it is common that you cannot locate the property/business for the bill to be delivered?"

- *Common to not locate owner* This variable takes a value of 1 (0) if the respondent answers 'Yes' ('No') to the question "When delivering bills, it is common that you locate the property/business but cannot locate the owner?"

- *Share of properties with valuation (%)* This variable is the answer to the question "Approximately what percent of the properties in the district are currently assessed by the Lands Valuation Board?". The answer ranges from 0% to 100%.

- *Share of tax payments made in cash (%)* This variable is the answer to the question "Approximately what percent of property rates are paid in cash?". The answer ranges from 0% to 100%.

- *Cost of collection (% of taxes collected)*. This variable is based on two questions asked to collectors in the census. The first question asks the collector what is their salary in a typical month. The second question asks the collector what is total revenue collected in a typical month. The variable is the ratio of salary to revenue collected, expressed as a a percent.

- *Officials with post-secondary education*. This variable is a dummy variable equal to 1 (0) if the local official has completed any form of post-secondary education (has completed secondary education or less). In turn, we calculate the unweighted share of officials with post-secondary education in each district.

- *Officials' average years of work experience*. This variable is the answer to the question "For how many years and months have you worked in local government?". Note that this variable includes working in the local official's current district as well as other districts in the past. In turn, we calculate the unweighted average years of work experience in each district.

- *Legal capacity to enforce taxes*. This variable is a dummy variable which takes a value of 1 if the local assembly has gazetted the fee fixing resolution for the fiscal year 2017-2018, and zero otherwise.

- *Take tax defaulters to court*. This variable is a dummy variable equal to 1 (0) if the respondent answers 'Yes' ('No') to the question "Does the assembly normally take ratepayers/business owners to court for non-payment of property rates".

- *Main reason for no court: Legal*. This variable is a dummy variable equal to 1 if the respondent answers 'Legal constraints' in answer to the question "Why does your district not take more ratepayers/business property owners to court for non-payment?" and zero otherwise. The other possible answers are 'Not worth it'; 'Politically sensitive'; and, 'Yet to implement/prefer non-enforcement'.

- *Main reason for no court: Political*. This variable is a dummy variable equal to 1 if the respondent answers 'Politically sensitive' in answer to the question "Why does your district not take more ratepayers/business property owners to court for non-payment?" and zero otherwise. The other possible answers are 'Not worth it'; 'Legal constraints'; and, 'Yet to implement/prefer non-enforcement'.

- *Citizen tax awareness*. This variable is a dummy variable equal to 1 if the respondent answers 'Yes' to the question "Have you heard about the fee fixing resolution?" and 0 if the respondent answers 'No'.

- *Citizen public good awareness* This variable is a dummy variable which takes a value of 1 if the respondent could name or describe a project that their local government has undertaken in the past two years in any one of these areas: road building; schools; waste management; health care; water provision. The dummy takes a value of 0 if the respondent could not name or describe any project in these areas.

B.2 Variables from Household Survey Related to Tax Outcomes

- *1(visit)* This is a dummy variable which takes a value of 1 if the household says that it was visited by a local government tax collector during the past 6 weeks (which corresponds to the time of the campaign when the experiment was implemented), and 0 otherwise.

- *Total visits (in %)* This is a continuous variable, which corresponds to the total number of visits that the household reports it was visited by the tax collector during the past 6 weeks. The variable is expressed in percent, using the inverse hyperbolic sine transformation.

- *Time per visit (in %)* This variable is calculated in multiple steps. First, we calculate the time per bill delivered for each collector. Second, we multiply time per bill delivered with the total number of bills delivered by each collector at the end of the campaign. Third, we subtract this estimate of the total time spent delivering bills from the total hours worked throughout the campaign for each collector to obtain a measure of time spent on non-delivery. Fourth, we divide this time by the total number of households that reported receiving any visit in each collector's unit (based on *1(visit)*), as a measure of the total time spent per visited household. Finally, we divide the time spent per visited household by the total number of visits reported by each household (based on *Total visits (in %)*). We express this variable in percent, using the inverse hyperbolic sine transformation. Note that this variable is only defined for households with at least one visit; to maintain the full sample size for estimation, we assign an arbitrary constant value to this variable for all remaining households and include a dummy variable in the estimating equation to capture this assignment. In the first step, we prefer to use the time per bill delivered that is calculated in the first survey round. The first round is when both groups of collectors spend most of their time on delivery (Figure 1), and this provides a clear setting to measure delivery time use. The difference between treatment and control groups in time spent per bill delivered is unlikely to be smaller in subsequent survey rounds. In fact, based on Figure 2, the unweighted average of the treatment effects for navigation challenges and taxpayer localization challenges is larger in magnitude in rounds 2 and 3 than in round 1. Results are similar if we use a measure of time per bill delivered in each round.

- *1(Bill delivered)* This variable takes a value of 1 if the household reports that they received a property tax bill from the tax collector in the past 6 weeks, and 0 otherwise.

- *Any tax payment* This variable takes a value of 1 if the household reports that they made a positive payment for the property tax liability that was due in the past 6 weeks, and 0 otherwise.

- *Full tax payment* This variable takes a value of 1 if the household reports that they made a positive payment for the property tax which corresponds to the full amount that was due in the past 6 weeks, and 0 otherwise.

- *Amount paid (in GHC)* This variable is the total amount in Ghanaian Cedi that the household reports having paid in property taxes during the past 6 weeks.

- *High hard-to-observe type* This variable takes a value of 1 if the value of the hard-to-observe index is in the top 5% of the index distribution, and 0 otherwise.

- *Share of payments from high type* This variable measures, in each collection unit, the share of payments that are from the *high hard-to-observe type*. The variable is created based on the household survey data.

B.3 Variables from Household Survey Related to Tax Morale and Enforcement

- *Satisfaction with government services index* This is an index variable, which is based on the average responses of households to three questions related to satisfaction with services. Possible responses are 'very satisfied', 'somewhat satisfied', 'neutral', somewhat unsatisfied', and 'very unsatisfied'. For each of the three questions, the answer is reverse coded such that higher values imply more satisfaction and all answers are standardized. The index variable is the unweighted average across the three standardized satisfaction questions outlined below

  1. "In your personal dealings with tax collectors in Madina, how satisfied are you with the outcomes?"
  2. "What has been your level of satisfaction with the overall quality of services offered by the local tax department of Madina"
  3. "What has been your level of satisfaction with the overall quality of services offered by the local government of Madina?"

- *Integrity of government index* This is an index variable, which is created as the unweighted average over the standardized responses to the different questions outlined below. Questions are reverse coded where relevant such that higher answers always indicate more positive view on integrity and competency of the local government

  1. "In your opinion, approximately what percent of the collections by the Madina Assembly will be put to good use for the benefit of the community?"
  2. "If the Madina Assembly wants to improve all the roads, it will do this efficiently and without problems". There are five answers, ranging from 'strongly agree' to 'strongly disagree'.
  3. "If the Madina Assembly wants to improve access to water for most citizens, it will be able to do so efficiently and without problems". There are five answers, ranging from 'strongly agree' to 'strongly disagree'.
  4. "If the Madina Assembly needed to improve waste management, it would be able to do so efficiently and without problems". There are five possible answers, ranging from 'strongly agree' to 'strongly disagree'.
  5. "Overall, how would you rate the Madina Assembly?". There are four possible answers, ranging from 'very competent' to 'not competent at all'.

- *Tax equity and efficiency efforts by government index* This is an index variable, based on the respondent's strength of agreement with three statements. Possible answers to each question are 'agree strongly', 'agree somewhat', 'neither agree nor disagree', 'disagree somewhat', 'strongly disagree'. Answers are reverse coded such that higher values reflect stronger agreement, and standardized. The index is the average across the respondent's agreement with the statements below

1. "Madina is making efforts to collect taxes in an efficient way"
2. "Madina is making efforts to ensure everyone in their community pays their fair share of taxes"
3. "Madina is making efforts to collect taxes that will be useful for local development of the community"

- *Enforcement and information capacity of the government index* This is an index variable, which is created as the unweighted average over the standardized responses to the different questions outlined below. Questions are reverse coded where relevant such that higher answers always indicate stronger perceptions of enforcement and informational capacity

  1. "What share of households and businesses in the Madina Assembly do you think usually pay their taxes?" Answers range from 0% to 100%
  2. "Imagine a tax collector comes to your neighborhood, and someone refuses to pay. How likely do you think that the local government will pursue and enforce sanctions?". There are four answers, ranging from 'very likely' to 'very unlikely'.
  3. "Do you think the local government knows the precise address of your residence?". There are four answers, ranging from 'very likely' to 'very unlikely'.
  4. "Do you think the local government knows which of your neighbors did not pay property or business tax in 2020?". There are four answers, ranging from 'very likely' to 'very unlikely'.
  5. "Do you think the local government knows what you do for a living?". There are four answers, ranging from 'very likely' to 'very unlikely'.

B.4 Variables from Household Survey Related to Bribes

- *Any bribe (coercive or collusive)* This variable is based on two dummy variables. The first dummy variable takes a value of 1 if the household estimates that tax collectors will ask for any strictly positive unofficial payments when they are working in the field, and zero otherwise. This variable proxies for the likelihood of collusive bribes. The exact question is: "Do you think it is likely that a local revenue collector will offer to take an unofficial payment from property owners/businesses in order to not make any return visits to their property?" The possible answers were: "very likely"; "somewhat likely"; "maybe"; "not very likely"; "very unlikely". If a respondent answered "very likely", "somewhat likely" or "maybe", then the follow up question was: "what is the amount in GHC that is typically asked for?". We replace this answer with zero if the respondent's first answer was "not very likely" or "very unlikely", and use this modified answer to construct the coercive bribe dummy. The second dummy variable takes a value of 1 if the household reports that the tax collector will pocket any positive amount out of a hypothetical 1000 Ghanaian Cedi collected from households (coercive bribe). The exact question is:

"Suppose a collector comes to a typical neighborhood in Madina and collects 1000 Ghanaian Cedi. How much of this money do you think the collector will submit to LANMA's tax finance office account? And, how much will they put in their pockets?". The variable used in the analysis takes a value of 1 if either the coercive dummy or the collusive dummy is equal to 1, and takes a value of 0 otherwise.

- *Any bribe (coercive)* This variable is a dummy variable which takes a value of 1 if the household reports that the tax collector will pocket any positive amount out of a hypothetical 1000 Ghanaian Cedi collected from households. The exact question is: "Suppose a collector comes to a typical neighborhood in Madina and collects 1000 Ghanaian Cedi. How much of this money do you think the collector will submit to LANMA's tax finance office account? And, how much will they put in their pockets?".

- *Any bribe (collusive)* This variable is a dummy variable which takes a value of 1 if the household estimates that tax collectors will ask for any strictly positive unofficial payments when they are working in the field, and zero otherwise. The exact question is: "Do you think it is likely that a local revenue collector will offer to take an unofficial payment from property owners/businesses in order to not make any return visits to their property?" The possible answers were: "very likely"; "somewhat likely"; "maybe"; "not very likely"; "very unlikely". If a respondent answered "very likely", "somewhat likely" or "maybe", then the follow up question was: "what is the amount in GHC that is typically asked for?". We replace this answer with zero if the respondent's first answer was "not very likely" or "very unlikely", and use this modified answer to construct the collusive bribe dummy.

- *Collusive bribe (Likert scale)* This variable is the answer to the question "Do you think it is likely that a local revenue collector will offer to take an unofficial payment from property owners/businesses in order not make any return visits to their property/business?". The 5 possible answers range from 'very unlikely' to 'very likely'. We assign numerical values from 1 to 5 which increase in the likelihood.

- *Total bribe amount (in %)* This variable is constructed at the household level as the unweighted average of the variable *'Collusive bribe amount (% of tax due)'* and the variable *'Coercive bribe amount (% of payment collected)'*. Both of these variables are described below.

- *Collusive bribe amount (% of tax due)* The collusive amount is the amount that the household estimates will be asked by the tax collector as unofficial payment while conducting visits to the household, expressed as a percent of the household's actual property tax due. The exact question is: "Do you think it is likely that a local revenue collector will offer to take an unofficial payment from property owners/businesses in order to not make any return visits to their property?" The possible answers were: "very likely"; "somewhat likely"; "maybe"; "not very likely"; "very unlikely". If a respondent answered "very likely", "somewhat likely" or "maybe", then the follow up question was: "what is the amount in GHC that is typically asked for?". We replace this answer with zero if the respondent's first answer was "not very likely" or "very unlikely", and express this modified answer relative to the value of household's actual amount of property tax due.

- *Coercive bribe amount (% of payment collected)* The coercive amount is the percent that the household estimates will be pocketed by the tax collector out of a hypothetical 1000 Ghanaian Cedi that the official has collected as payments from households while working in the field. The exact question is: "Suppose a collector comes to a typical neighborhood in Madina and collects 1000 Ghanaian Cedi. How much of this money do you think the collector will submit to LANMA's tax finance office account? And, how much will they put in their pockets?" We use the answer to the latter question to construct this variable.

- *Collusive bribe amount (in Ghanaian Cedi)* The collusive amount is the amount that the household estimates will be asked by the official as unofficial payment while conducting visits to the household. The exact question is: "Do you think it is likely that a local revenue collector will offer to take an unofficial payment from property owners/businesses in order to not make any return visits to their property?" The possible answers were: "very likely"; "somewhat likely"; "maybe"; "not very likely"; "very unlikely". If a respondent answered "very likely", "somewhat likely" or "maybe", then the follow up question was: "what is the amount that is typically asked for?". We replace this answer with zero if the respondent's first answer was "not very likely" or "very unlikely", and use this modified answer as the variable.

B.5 Variables from Household Survey Related to Learning and Targeting

- *Liquidity* This variable is created as the unweighted average over two household survey questions, which are outlined below. The survey questions are reverse coded such that higher values reflect lower liquidity constraints. Answers to both survey questions are standardized, and the liquidity index is in turn the unweighted average over these two standardized survey variables. The two variables are

  1. "Think of a typical month. On how many days did you find yourself short of cash for basic expenditures for your house?". The answer can range from 0 to 30 days
  2. "In a typical month, imagine that one day you learn you need to pay an additional 300 Cedi fee in order to remain in your house. Could you find this money in the next 4 days?". The possible answers are 'Yes, with a little difficulty'; 'Yes, with great difficulty'; 'Very unlikely'; 'I could never pay this fee'

- *Income* This variable is based on the answer to the household question "What was the household's total earnings this past month?". The answer is given in Ghanaian Cedi. The income index is the standardized answer.

- *Taxpayer awareness* This variable is the unweighted average of six dummy variables which each take a value of 1 if the person answers 'Yes' to the individual questions outlined below, and take a value of 0 if the respondent answers 'No'. In turn, the unweighted average across the six variables is standardized to create the awareness index.

  1. "Do you know of someone who received a letter from their MMDA summoning them to appear in court for non-payment of property rates"
  2. "Do you know of someone who was actually taken to court for non-payment of property rates?"
  3. Have you heard of any instance where a property owner had their property confiscated for non-payment of property rates?"
  4. "As best as you can remember, did you receive any text message earlier this year from your MMDA about paying the property rate?"
  5. "As far as you know, do the MMDAs have the legal authority to collect property rates?"
  6. "Have you heard of the fee-fixing resolution?"

- *Taxpayer awareness – Enforcement* This variable is the unweighted average of three dummy variables which each take a value of 1 if the person answers 'Yes' to the individual questions outlined below, and take a value of 0 if the respondent answers 'No'. In turn, the unweighted average across the three variables is standardized to create the index variable.

  1. "Do you know of someone who received a letter from their MMDA summoning them to appear in court for non-payment of property rates"
  2. "Do you know of someone who was actually taken to court for non-payment of property rates?"
  3. Have you heard of any instance where a property owner had their property confiscated for non-payment of property rates?"

- *Taxpayer awareness – Tax code* This variable is the unweighted average of three dummy variables which each take a value of 1 if the person answers 'Yes' to the individual questions outlined below, and take a value of 0 if the respondent answers 'No'. In turn, the unweighted average across the three variables is standardized to create the index variable.

  1. "As best as you can remember, did you receive any text message earlier this year from your MMDA about paying the property rate?"
  2. "As far as you know, do the MMDA's have the legal authority to collect property rates?"
  3. "Have you heard of the fee-fixing resolution?"

- *Propensity to pay index/hard to observe index* This variable is the unweighted average of the three index variables *Liquidity*, *Income* and *Taxpayer awareness*

- *Tax bill value* This variable is based on the administrative data and measures the total amount of taxes that are owed. The total amount owed is the sum of the current year's property taxes and outstanding arrears due to less than full payment of the past year's property taxes. The variable is standardized.

- *Previous tax payment* This variable is based on the administrative data and measures the payment status from the previous fiscal year. It takes a value of 1/2/3 if the past year's property taxes were not paid at all/partially paid/fully paid. The variable is standardized.

- *Assets* This variable is the sum over how many of the following assets the household currently possesses: motorbike; car or truck; television; electric generator; sewing machine; radio. In turn, the variable is standardized.

- *Distance* This variable is the average distance measured for each household to three locations: main roads, markets, and the local government headquarters. There are several main roads; the distance variable calculates the shortest distance from the household's property to any of the main roads. Similarly, the shortest distance is calculated from the household's property to any of the main markets. Finally, the shortest distance is calculated from the property to the local government's headquarters. The shortest distance is calculated using the existing set of streets and information on walking time as calculated in Google Maps. Each of the three distance variables are standardized; in turn, the *Distance* variable is the average of the three standardized variables.

- *Easy to observe index* This variable is the unweighted average of the four standardized variables *tax bill value*, *previous tax payment*, *assets* and *distance*.

B.6 Variables from Collector Surveys

- *Challenge to navigate in the field* This variable is a dummy variable which takes a value of 1 if the respondent 'strongly agrees' or 'agrees' with the statement "Finding my way around my collection unit was a challenge for me this week"; the dummy variable takes a value of 0 if the respondent answers 'neither agree nor disagree', 'disagree' or 'strongly disagree.

- *Challenge to locate taxpayers* This variable is a dummy variable which takes a value of 1 if the respondent 'strongly agrees' or 'agrees' with the statement "Locating bill recipients was challenging for me this week"; the dummy variable takes a value of 0 if the respondent answers 'neither agree nor disagree', 'disagree' or 'strongly disagree.

- *Knowledge about households which are willing and able to pay* This variable takes a value of 1 if the respondents chooses statement A "I think I have a good understanding of which properties are more able and willing to pay and am able to focus my efforts on them" rather than statement B "I put a lot of effort to get my job done, but it remains unclear to me which exact properties are more likely or willing to pay their property rates". The variable takes a value of 0 if the respondent picks statement B. Respondents had to pick the statement which "you would say best characterizes your work in the field over the past weeks".

- *Focus on households that are able to pay* This variable takes a value of 1 if the respondent uses ’all the time’ or ’often’ the collection strategy "Go to areas on specific days where I know property owners are more likely to be able to pay"; the variable takes a value of 0 if the respondent uses this strategy ’only from time to time’, ’not much’ or ’never’.

- *Focus on households that are aware of tax payment duty* This variable takes a value of 1 if the respondent uses ’all the time’ or ’often’ the collection strategy "Go to areas where I know most taxpayers are aware of their duty to pay property rates"; the variable takes a value of 0 if the respondents uses this strategy ’only from time to time’, ’not much’ or ’never’.

- *Focus on households that are satisfied with public goods* This variable takes a value of 1 if the respondent uses ’all the time’ or ’often’ the collection strategy "Go to areas where I know owners are more satisfied with the delivery of public services and are more likely to pay"; the variable takes a value of 0 if the respondents uses this strategy ’only from time to time’, ’not much’ or ’never’.

- *Focus on collections with hard-to-observe household characteristics* This variable measures the frequency with which collectors make use of the three strategies that target hard-to-observe household characteristics: *focus on households that are aware of tax payment duty*, *focus on households that are able to pay*, and *focus on households that are satisfied with public goods*. The variable is the average across those three strategy use variables, and takes a value between 0 and 1.

- *Focus on collections with easy-to-observe household characteristics* This variable measures the frequency with which collectors make use of six strategies that target easy-to-observe household characteristics. For each strategy, outlined below, we measure use with a value of 1 if that collection strategy is used ’often’ or ’all the time’ and 0 if it is used ’only from time to time’, ’not much’ or ’never’. In turn, the variable is the average use across these six strategies, and takes a value between 0 and 1. The six strategies considered are

  1. "Go to areas where I know most taxpayers have paid property rates in the past year"
  2. "Go to areas where I know there are many properties with high property rates"

3. "Go to areas where I know there are many property rate payers that have not yet paid this year's rates"
4. "Go to areas which are close to the main road/center of activity"
5. "Go to areas which are close to my home"
6. "Go to areas which are closer to the Madina headquarters"

- *Difference in strategies: Hard versus easy to observe* This variable is the difference between the variable *'Focus on collections with hard-to-observe household characteristics'* and the variable *'Focus on collections with easy-to-observe household characteristics'*

- *Unable to locate properties and owners* This variable measures the collectors' extent of agreement with two statements: "Finding my way around my collection unit was challenging"; "Locating bill recipients was challenging". For each statement, the respondent can answer 'strongly disagree', 'disagree', 'neither agree nor disagree', 'agree', 'strongly agree'. We assign numerical values from 1 to 5, with larger values indicating stronger agreement. The answer to each statement is standardized, and the variable is the average over the two standardized answers.

- *Wrong information printed on bills* This variable measures the collectors' extent of agreement with the two statements: "Some of the bills I tried to deliver this week had the wrong addresses"; "Some of the bills I tried to deliver this week had the wrong amounts". For each statement, the respondent can answer 'strongly disagree', 'disagree', 'neither agree nor disagree', 'agree', 'strongly agree'. We assign numerical values from 1 to 5, with larger values indicating stronger agreement. The answer to each statement is standardized, and the variable is the average over the two standardized answers.

- *Resistance from property to accept bill* This variable measures the collectors' extent of agreement with three statements: "Collection was challenging this week because bill recipients preferred not to pay in cash"; "Collection was challenging this week because bill recipients preferred mobile payments, but I was not able to accept mobile payments"; "Collection was challenging this week because bill recipients said that they did not trust me to collect their payment". For each statement, the respondent can answer 'strongly disagree', 'disagree', 'neither agree nor disagree', 'agree', 'strongly agree'. We assign numerical values from 1 to 5, with larger values indicating stronger agreement. The answer to each statement is standardized, and the variable is the average over the three standardized answers.

- *Supervisors do not monitor field activities* This variable measures the extent to which collectors perceive that their supervisors are not monitoring their work. Specifically, we ask the collector's extent of agreement with the statement: "My supervisors spent a lot of time monitoring my work this week". For each statement, the respondent can answer 'strongly disagree', 'disagree', 'neither agree nor disagree', 'agree', 'strongly agree'. We assign numerical values from 1 to 5, with larger values indicating stronger *disagreement*. Values are standardized to be comparable with other outcomes.

- *Supervisors do not check mistakes made in the field* This variable measures the extent to which collectors perceive that their supervisors are not checking mistakes made by collectors in the field. Specifically, we ask the collector's extent of agreement with the statement: "My supervisors checked on me regularly this week to make sure I was not making mistakes". For each statement, the respondent can answer 'strongly disagree', 'disagree', 'neither agree nor disagree', 'agree', 'strongly agree'. We assign numerical values from 1 to 5, with larger values indicating stronger *disagreement*. Values are standardized to be comparable with other outcomes.

- *Supervisors are unavailable for support* This variable measures the extent to which collectors perceive that their supervisors are not available to support the collectors in the field. Specifically, we ask the collector's extent of agreement with the statement: "My supervisors were available to help me this week when I needed them". For each statement, the respondent can answer 'strongly disagree', 'disagree', 'neither agree nor disagree', 'agree', 'strongly agree'. We assign numerical values from 1 to 5, with larger values indicating stronger *disagreement*. Values are standardized to be comparable with other outcomes.

- *# Unsuccessful visits per successful visit* This variable is the answer to the question "There are many challenges to getting things done in the field. Looking back at this past week, let us think about the unsuccessful visits you made to properties. A successful visit is a visit to a property where you were able to complete the task you had planned. For every successful visit, how many unsuccessful visits would you say that there were, for the typical property?"

- *Total hours worked per week* This variable is the product of the following two questions: "How many days did you work this week?"; and, "During the days where you did work this week, what would you say is approximately the number of hours you worked?".

- *Average # hours spent to deliver one bill* This variable is the ratio of the total weekly hours spent on delivery divided by the total weekly bills delivered. Both variables are based on the collector's self-reports.

- *Satisfaction and happiness on job* This variable measures the collectors' extent of agreement with three statements: "Overall, this was a productive week for me"; "Overall, I was content while working this week"; "Overall, I am satisfied with my job". For each statement, the respondent can answer 'strongly disagree', 'disagree', 'neither agree nor disagree', 'agree', 'strongly agree'. We assign numerical values from 1 to 5, with larger values indicating stronger agreement. The answer to each statement is standardized, and the variable is the average over the three standardized answers.

- *Field work is organized and narrow* This variable is the unweighted average of two dummy variables, by collector and survey round. The first dummy takes a value of 1 if the collector reports that their field-work this week is best characterized by a focus on specific properties in their unit or 0 if their work is best characterized by working in broad areas throughout their unit. The second dummy takes a value of 1 if the collector reports that their field-work this week is best characterized by organizing which properties to visit before going into the field each day or 0 if their work is best characterized by going into the field immediately and making the most of it.

- *Hours on non-delivery activities* This variable is calculated in several steps. First, we calculate the time per bill delivered for each collector. Second, we measure the total number of bills newly delivered by each collector separately in round 1 (first two campaign weeks), round 2 (third and fourth campaign weeks) and round 3 (final two campaign weeks). Third, we multiply the total number of bills delivered by round with the time per bill delivered. Fourth, we subtract this estimate from the total hours worked during each round. In the first step, we prefer to use the time per bill delivered that is calculated in the first survey round. The first round is when both groups of collectors spend most of their time on delivery (Figure 1), and this provides a clear setting to measure delivery time use. The difference between treatment and control groups in time spent per bill delivered is unlikely to be smaller in subsequent survey rounds. In fact, based on Figure 2, the unweighted average of the treatment effects for navigation challenges and taxpayer localization challenges is larger in magnitude in rounds 2 and 3 than in round 1. Results are similar if we use a measure of time per bill delivered in each round.

C Addressing, Technology and Tax Collection: Discussion

Incomplete addressing is a feature in many settings around the world, but it has received limited attention in economics. In this section, we review the findings from other literatures on the origins of incomplete addressing. We then discuss non-technology initiatives to support tax collection when there is limited addressing. We review initiatives both from Ghana and from other relevant settings.

C.1 Factors That Limit Addressing Coverage

Based on a multi-disciplinary literature in several regions, general constraints on achieving broad addressing coverage have been identified, which we summarize below. We then discuss the historical experience of developing addressing infrastructure in the US.[42]

Spatial planning Studies in urban planning highlight how governments struggle to organize, execute and enforce spatial planning: see United Postal Union (2011) and Farvarcque et al (2005) for reviews of spatial planning practices around the world; see Fuseini and Kemp (2015) for a review of challenges in Ghana's context. In the face of uncertain,

[42]This discussion is related to the broad field of toponomy, the study of street and place naming. See also Adarkwa (2013), Agyeman (2006), Azaryahu (1996), Coetzee and Cooper (2007), Fyfe (1995), Rose-Redwood (2008), Rutgers (2000), Pred (1990) and Twumasi-Fofye (2015).

delayed and unenforced procedures for building approval and permit issuance, property development often proceeds outside of government processes and are built without addresses (Yeboah and Obeng-Odoom, 2010). Property development sometimes also takes place without laid-out access to the pre-existing set of streets. This process is further complicated by the fragmentation and legal uncertainty surrounding land ownership and the underdeveloped official land and property transaction markets (Njoh, 2007). Due to these uncertainties, many property owners and developers intentionally avoid interacting with official government processes. Of course, property owners may also avoid having their property and street officially named because they perceive the link to (property) taxation (Bowles, 2023).

These factors result in the inability of governments to monitor property ownership and development for the purpose of creating street and property addresses, let alone regulating and organizing local spatial structure. The challenge is compounded when there is an increasing rate of urbanization or rapid shifts in population size, which occur in many developing countries (UN Habitat, 2022). Spatial planning may be particularly challenging in urban slums, where over 800 million people live and which are expanding globally (Marx, Stoker and Suri, 2013).

Social identity and formalization Studies in social psychology and political science have emphasized how individuals and local communities resist the process of addressing if and when they perceive that, by officially numbering properties and naming streets, government is formalizing their identities in ways that do not align with their individually and socially constructed realities (Adebanwi, 2012; Bigon and Arrous, 2022; Oto-Peralias, 2018). A concrete example is when formalizing a property through addressing and assigning it to an individual is at odds with the local legal and social reality of property ownership.

More generally, addressing has historically been used both as a governmental tool to influence identity and local autonomy and as a battle-ground of resistance by communities (Scott, 1998). Some scholars view the challenges faced in attempting to create a comprehensive map of citizens' properties as reflecting the inherent resistance towards building a 'modern state' that is schematic, precise and uniform. Analyzing historical episodes of state building, Kain and Bagent (1992) write that "maps and modernization went hand in hand. The state made the maps, and the maps upheld the state."

In Ghana, officially naming a street requires extensive community consultation, including with local organizations and traditional leaders (Government of Ghana, 2011). The ultimate approval may be stalled because of deeper power dynamics (Kasanga and Kotey, 2001). Indeed, historical accounts of achieving and sustaining a high street and property addressing coverage associate this success with weaker civil society and fewer groups with local influence (Arcy and Nistotskaya, 2017).

Administrative decentralization and fragmentation Studies in governance and public administration highlight that administrative centralization is a determinant of comprehensive addressing (Knebelmann, 2022; Scott, 1998). Conversely, incomplete addressing arises in decentralized settings where processes are fragmented and the power to impose and sustain official naming is limited (Farvacque et al., 2005).

In Ghana, each local government is individually responsible for naming all its streets and properties. While granting this mandate to the local level has been justified to deepen devolution, it has also contributed to the proliferation of technically inadequate addressing systems across the country (Government of Ghana, 2011). Of the local governments that do not currently use technology for addressing in Ghana, our census reveals that only 17% have in the past 10 years implemented policies to increase addressing coverage. Moreover, 12 distinct service providers were identified by these local governments – including collaborations with private domestic firms, technical units in central government and international donors. The proliferation of providers highlights the fragmented processes across the country.

Developing addressing infrastructure in the United States The historical achievement of high addressing coverage in the United States without resorting to technology is indicative of the factors described above (Feirstein, 2001). First, spatial planning was initially strong and local authorities executed convenient planning strategies, such as the grid system, prior to large population influx. Second, most streets would initially be numbered rather than named (the five most popular street names today in the US are still "Second", "Third", "First", "Fourth" and "Park": link). Processes to modify the street names took place subsequently, but they largely did not hinder the initial implementation of an addressing system with comprehensive coverage. These processes are more active today than in the past: by some accounts, 40% of the local law changes passed in New York City in recent years have been street name changes. Third, while the addressing mandate is assigned to sub-federal authorities, federal agencies (including the Census Bureau and the Postal Service) exerted significant influence early on, including to disseminate the spatial planning models of "address success stories" like New York City and Philadelphia.

C.2 Non-Technology Initiatives to Support Taxation with Limited Addressing

Our paper focuses on a GIS-based technology to overcome constraints on tax collection that arise from incomplete addressing, but non-technology alternatives do exist.

One alternative initiative is to physically and visibly mark each house with an individual number that is recorded in the tax registry. In Ghana, this type of unofficial designation is not permitted by the local tax collectors, since they are part of the local government whose mandate it is to designate properties through official processes that are approved by the community (Section C.1). It is also possible that this type of designation by the government would be resisted and overturned by property owners. Finally, this type of initiative has been found to function poorly in more populous areas and in areas with rapid population growth (Bigon and Njoh, 2013; Abebrese, 2019).

A second possibility is to draw on location databases created by other government agencies and third-party institutions. This would, however, require significant and continuous coordination. Moreover, the location information may be recorded in non-standard ways and be of limited use to tax collectors without further training. Finally, there is no guarantee that the coverage in those datasets is complete. Behr et al. (2023)

describe how third party institutions in developing countries, including land registries, real estate agencies, financial institutions and utilities providers do not collect comprehensive information, nor do they organize it in a harmonized manner.

A third possibility is to bundle bill delivery and tax collection together with the initial discovery and registration phase. This 'snowball' approach alleviates navigation issues that arise by decoupling registration from collection. However, this approach also limits revenue collectors from being able to conduct follow-up visits, which may ultimately matter for compliance when ability to pay is limited and temporary. More generally, this approach constrains the government's ability to predict, control and alter the collection process, including for redistributive purposes.

Finally, both the delivery of the tax bill and payment of the taxes due could entirely be done electronically. Such a technology initiative may ultimately be preferred to the GIS-enhanced tax registry, but it is plausibly still beyond the technical capacity of citizens and local governments in many settings.

Relative to these alternatives, investing in a GIS-based technology may be a realistic way forward in many settings. Consistent with this observation, virtually all of the 'best case practices' in the World Bank's manual for property taxation (2020) involve transitioning to the use of a GIS-enhanced tax registry. GIS-technologies are being implemented in many countries around the world: see Knebelmann (2022) for a comprehensive review and Okunogbe (2021) for a detailed discussion in Liberia.

D Learning Mechanism: Additional Results

D.1 Active versus Passive Learning

In this appendix section, we explore the possibility of passive learning – whereby the collector, through repeated attempts to collect, uncovers the spatial clustering of types and forms beliefs about the household's type based on the immediate area's payment rate. Untangling active from passive learning may ultimately be challenging, since passive learning is a by-product of collection and the treatment group collects more than the control group.[43] However, the distinction is potentially important for policy design if the knowledge gathered by one set of officials (through active learning) is transferred for use by other officials to target individual households.

In any case, the passive spatial form of learning depends on the extent of geographical clustering in the hard-to-observe index. We leverage the spatial coordinates of all households and use Moran's I statistic and the nearest-neighbor spatial weight matrix to measure auto-correlation. We calculate Moran's I statistic as a function of the number of nearest neighbors. The formula for Moran's I is given by

$$I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \omega_{ij} z_i z_j}{\sum_{i=1}^{n} z_i^2}$$

[43] For example, a collector with strong passive knowledge of the spatial distribution of types may make educated guesses and score highly on a quiz about individual property owners' propensity to pay.

where n is the number of household observations, z_i is the residualized value of the hard-to-observe index for household i, and ω_{ij} is the ij^{th} element of the spatial weight matrix. z_i is obtained as the residual value of the index after regressing the index on collection unit fixed effects. In so doing, we are calculating the Moran's I which is relevant at the level where the collector's activities are defined (i.e. within a collection unit). Longitude and latitude coordinates for all households are used to calculate the k^{th} nearest neighbor for each household, where $k = 1, ..., 50$. In turn, the k^{th}-nearest neighbor spatial weight matrix is used to calculate Moran's I statistic, separately for each value of k from 1 to 50.

Figure D1 shows the resulting set of Moran I statistics as a function of k when z_i is the hard-to-observe index (square line) or the easy-to-observe index (cross line). The figure shows that there is positive global spatial auto-correlation in the full sample for the hard-to-observe index; in other words, neighbors to a household with a higher index value are more likely to also have a higher index value. However, the correlation is economically small and it decreases rapidly in magnitude once k moves beyond the immediately closest 3 to 5 neighbors. Interestingly, spatial auto-correlation for the easy-to-observe index is both much larger in magnitude and also less localized – yet we find no differential selection on the easy-to-observe index (Figure 4). This observation suggests spatial learning may not be the most prevalent collector strategy in the field.

With this global correlation structure in mind, we measure local 'hot spots' and 'cold spots' for the hard-to-observe index. We use the Getis-Ord statistic to measure hot and cold spots, with a 1% significance level. The hard to observe cold spots are calculated in two steps. First, the Getis-Ord G_i^* statistic is calculated for the hard-to-observe index z_j

$$G_i^* = \frac{\sum_{j=1}^{n} \omega_{ij} z_i}{\sum_{j=1}^{n} z_j}$$

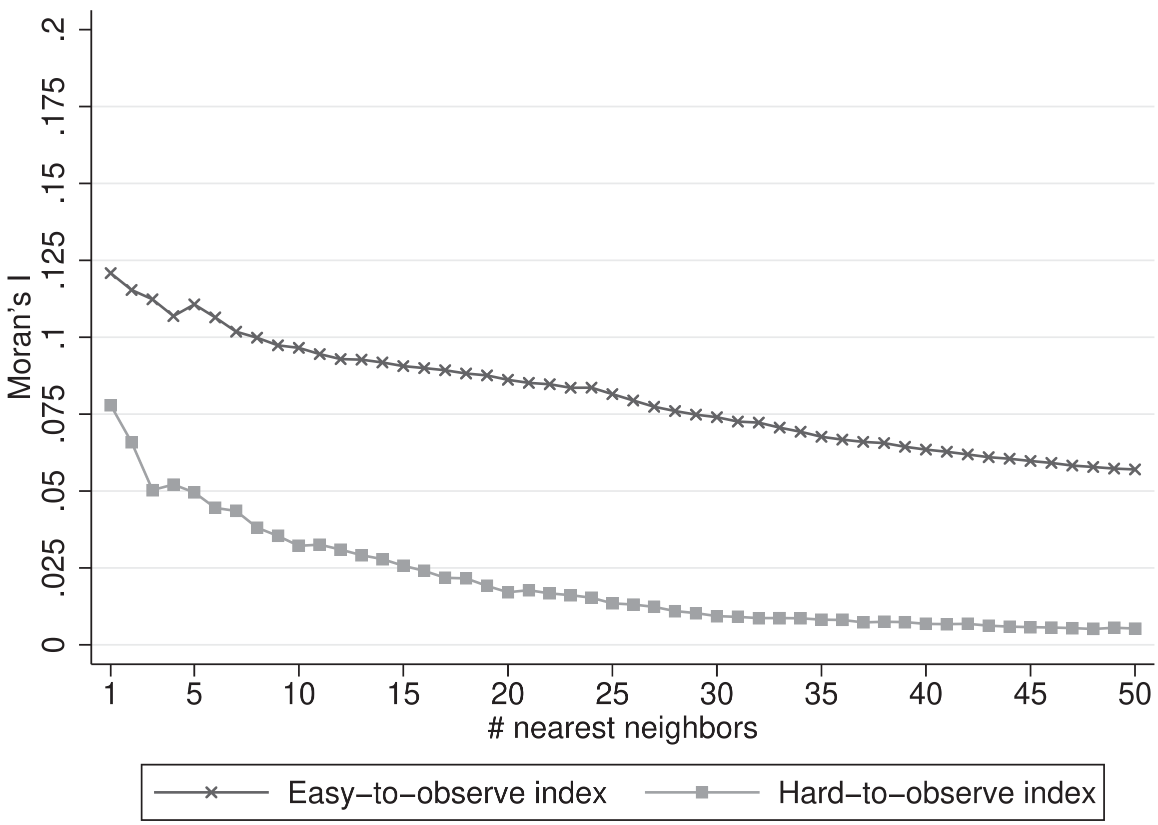
where n is the total number of households and ω_{ij} denotes the ij^{th} element of the binary spatial weight matrix with a threshold distance of 0.1km. z_j is the residual index value after regressing it on collection-unit fixed effects. Second, G_i^* is standardized and a household i is in a cold spot if its standardized $G_i^* \leq -2.58$ (critical value at 1% significance). Hard to observe hot spots have standardized $G_i^* \geq 2.58$.

This exercise reveals that 1.80% of observations are in 'hot spots', with a particularly strong local concentration of high hard-to-observe index values, and 1.98% are in cold spots, with a strong concentration of low values of the hard-to-observe index. The existence of both hot and cold spots are balanced across treatment and control areas.

In turn, we estimate if the collection rate in hot and cold spots differs between treatment and control areas. Specifically, we restrict the household survey sample to the hard-to-observe cold spots and estimate equation (2) when the outcome is a dummy for whether the household made a tax payment. The first row of Panel A in Figure D2 plots the treatment coefficient along with the 95% CI. The next two rows in Panel A repeat the exercise, but estimates the treatment coefficients in the sample of hard-to-observe hot spots and easy-to-observe cold spots, respectively. Panel B of Figure D2 is constructed similarly to Panel A, except the outcome is a dummy for whether the household was delivered a bill.

Panel A reveals that the *level* of the payment rate is larger in treated than in control areas for hard-to-observe hot spots, but smaller for hard-to-observe cold spots. Since treatment collectors achieve a higher payment rate in general and passive learning is based on inferring ex post based on observed payment rates, this form of learning is inconsistent with the lower payment rate in treatment areas for cold spots. The lower payment rate in cold spots is strongly consistent with active, 'ex ante' learning about households' types which precedes any attempt to collect. The negative difference in payment rates is not confounded by some attribute of cold spots in general, since there is no difference in payment rate for easy-to-observe cold spots. These results suggest that passive spatial learning may have limited relevance in our setting. Moreover, Panel B reveals that the same patterns hold for bill delivery. These patterns for bill delivery rates are *a priori* inconsistent with passive learning, since passive learning occurs only later when attempts to collect are being made.

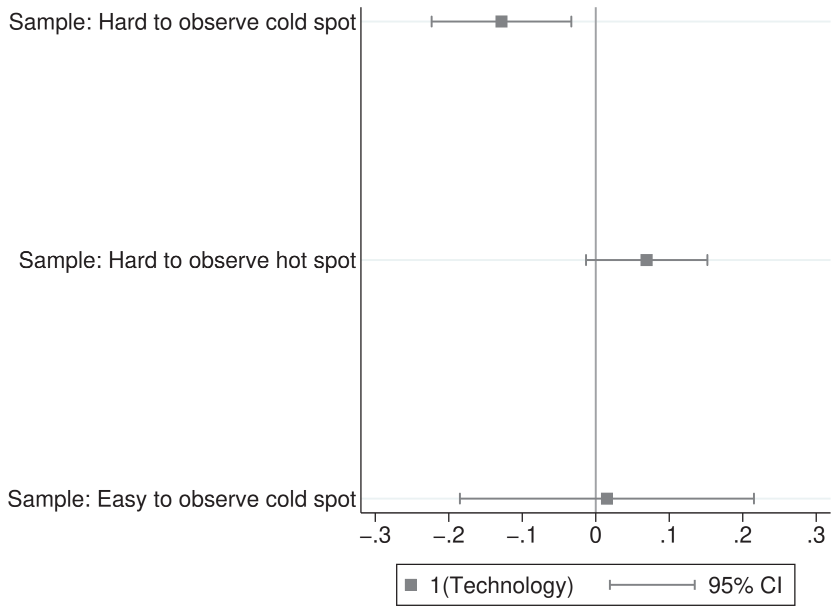
Figure D1: Moran's I for the Easy-to-Observe Index and Hard-to-Observe Index



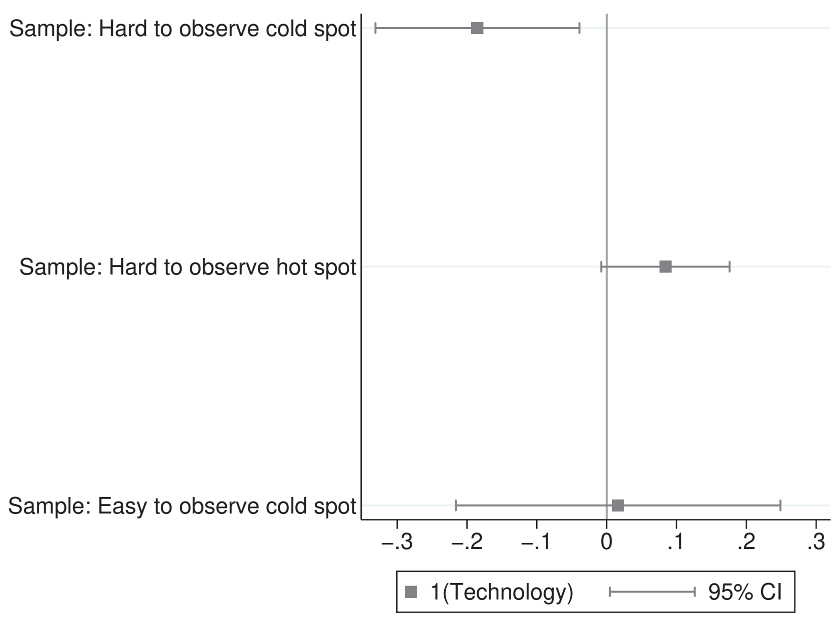
Notes: This figure shows the Moran's I statistic calculated as a function of the number of nearest neighbors, and separately for the hard-to-observe index (square line) and the easy-to-observe index (cross line). See Section D.1 for more details on the method. See Data Appendix B.5 for details on the indices.

Figure D2: Impacts of Technology in Sub-Samples

(a) Outcome: Tax Payment



(b) Outcome: Bill Delivered



Notes: These panels show the impacts of technology on the likelihood of making a payment (Panel A) and receiving a bill (Panel B) in separate samples. In each panel, the β coefficients on $\mathbf{1}(Tech)_c$ from estimating equation (2), along with the 95% CI, are displayed based on separate estimation in three samples: hard-to-observe cold spots; hard-to-observe hot spots; easy-to-observe cold spots. Standard errors are clustered at the collector-unit level. See Section D.1 for more details on the method to determine the samples, and Data Appendix B.2 for details on the outcomes.

D.2 Observability of Household Types: Discussion

In the main text, we have focused on a setting where types are hard to observe – in the sense that the household's propensity to pay is initially not observed but the collector can discover it through a time-costly process of engaging with the household. The collector can then subsequently target those households with higher propensity to pay during follow up visits and collect a payment with a fixed probability. This set-up is consistent with the treatment collectors' limited initial knowledge that grows over time (Figure 3). However, that result is based on the collectors' self-reports. It is worth considering that this self-report may simply reflect that treatment collectors have a limited initial understanding of the work environment and that, as they collect more taxes over time, they report better knowledge of households' types as a 'by-product' of improved performance but without having actually learned.

In this section, we therefore investigate whether the mechanism results based on the household survey (Section 5) can be made consistent with alternative settings that differ in the assumption of observability for household's propensity to pay. In the alternative setting where each household's type is perfectly unobservable to all collectors, in the sense that the type can never be discovered, the collector may allocate time to indiscriminately make follow-up visits to all properties they delivered a bill to and collect a payment with a fixed probability. In the alternative setting where types are perfectly observable to all collectors, the collector may allocate time to conduct follow-up visits with any targeted household and collect a payment with a fixed probability.

1st selection result: Figure 4 We begin with the results on the first selection measure in Figure 4. In Panel A, the positive differential payment (β in equation 3) for the hard to observe index is consistent with learning: treatment collectors allocate time to gradually learn which households have higher values of the hard to observe index, and they subsequently target those with higher values during follow-up visits. This generates a positive difference in payment patterns for the hard to observe index in the treatment versus control group ($\beta > 0$).

Positive differential payment $\beta > 0$ for the hard to observe index is potentially consistent with the setting where the household type is perfectly unobservable. To see why, consider a simplified setting where there are only two types and where only the high-type makes a payment with some positive probability while the low-type never pays; the intuition extends to the general case with a continuous type. In this case, the positive differential payment $\beta > 0$ reflects the fact that the treatment collector allocated more time to indiscriminate follow-up collection visits, which leads to a higher payment rate from the high type than in the control areas. However, in this setting we would also expect to observe some $\beta > 0$ between treatment and control areas on the easy to observe index – since this index is a predictor of payment (Table A14). However, we observe practically no differential payment rate on the easy to observe index (Panel B in Figure 4). Moreover, when the objective is to maximize the number of indiscriminate repeat-visits, it is not clear why the treatment collector would spend more time per visit than the control collector (which we observe in the experiment – see Table 2).

Finally, $\beta > 0$ for the hard to observe index is also potentially consistent with settings where the household type is in fact perfectly observable to all collectors. In particular, when all collectors begin by focusing on households with the highest observable type-value, $\beta > 0$ arises only if the treatment collector, despite delivering more bills and collecting more payments, does not exhaust the number of highest types to collect from. In this specific case, the $\beta > 0$ reflects the higher collection rate from the highest type-value achieved in the treatment areas versus control areas. If the treatment collector does exhaust the top type-value to collect from and starts moving 'down the curve', this would generate a negative differential payment pattern $\beta < 0$, since the average type-value would be smaller in the treatment group than in the control group. In the case of no exhaustion at the top, the treatment group should collect 100% of its payments from the highest type-value. However, we observe that the treatment group collects much less than 100% of its payments from the 'hard to observe' high type, defined as households with a hard to observe index value in the top 5% (Figure 5).

In Figure A19, we find almost exactly the same β selection results for bill delivery as for payment. This result is consistent with the 'learning' setting, if collectors anticipate they will not be able to deliver all bills and therefore start to learn about properties while delivering. This result is inconsistent with the 'perfectly unobservable' setting, since there should be no specific association in this setting between receiving a bill and household characteristics, let alone a differential association in treatment versus control areas. The bill delivery result is consistent with the 'perfectly observable' setting under the same condition of 'no exhaustion at the top' as for payments.

In Figure A20, we find that the selection result for payment patterns grows over time. Since the treatment group devotes more time to non-delivery activities over time, this dynamic result is broadly consistent with the three settings. It is consistent with the learning setting, where the returns to learning during earlier parts of the campaign pay off in terms of more precise targeting in later parts of the campaign. It is consistent with the 'perfectly unobservable' setting, so long as the difference in the number of repeat visits made in the treatment versus control group is larger in the later part of the campaign than in the early part. It is consistent with the 'perfectly observable' setting, so long as the 'no exhaustion at the top' condition holds in the late parts of the campaign.

2^{nd} **selection result: Figure 5** In Figure 5 we find that the treatment group collects a larger share of its payments from the 'hard to observe' high-type. This positive selection on the composition of payers is consistent with a learning setting.

A strictly positive selection on composition is inconsistent with any setting where types are perfectly unobservable and where the cumulative likelihood of collecting a payment as a function of the number of repeat visits is not strictly convex. For example, if the likelihood of making a payment is constant for each visit (as in our model), then the likelihood that the household has made a payment by the n^{th} visit is concave. In this 'war of attrition' where the treatment group allocates more time to repeat-visits than the control group, the concavity of the payment function ensures that high-types will make up a smaller share of payments in the treatment group than in the control group – that is, there will be negative selection on the composition of payers (contrary to Figure 5). More generally, if the high-type has a higher probability of payment than the low-type, for any number of repeat visits, and if the cumulative probability of having collected a payment by the n^{th} visit is non-convex, for both types, then the high-type's share in total payments will necessarily (weakly) decrease in the total number of repeat visits. This would generate negative selection on the composition of payers if the treatment group only differs from the control group in that it conducts more repeat visits. This logic holds regardless of the high-type's share in the population of taxpayers.

In the perfectly observable setting, under the assumption of a non-convex cumulative payment likelihood, the selection measure will be null if the treatment group does not exhaust collections from the high-types, or negative if it does. A strictly positive selection on composition when types are observable only occurs with a specific setup. In particular, despite the high-types being observable to all collectors, the treatment collectors must, purely as a function of allocating more time, conduct a larger share of their interactions with high-types. This would occur if the return per unit of time is higher for the low-type at lower levels of time, and higher for the high-type at higher levels of time. For example, attempting to collect from a high-type may require a fixed time cost, beneath which the pay-off is zero, while no such fixed cost exists for the low-type. Some parameter settings within this convex payment structure will deliver a positive treatment effect on composition; but this payment structure does not per se guarantee it, and some parameter settings for the difference in time-allocations and the threshold where returns by type intersect would generate a negative selection effect. Moreover, the parameter settings have to be specific to match the positive selection in Figure 5 of the experiment: the threshold can be neither too large, since the control group collects a strictly positive amount from high-types, nor too small, since both treatment and control groups collect much less than 100% of payments from high-types.

It is also not clear what would generate this convex structure, given our full set of results. A fixed time-cost for the high-type may reflect some characteristics that are specific to the high-type. For example, all the high-types may be located far away from the collector's daily starting point and it might require a large time-allocation to travel there to collect payment. However, recall in Figure 4 that there was no differential payment selection β between treatment and control groups for a large set of observable characteristics, including the property's distance to main roads, markets and the local government office. Moreover, there were also no differential selection patterns in terms of bill delivery for these observable characteristics (Figure A19). The convex returns to time for the high-type may alternatively reflect that trust, morale or enforcement perceptions are only activated for the high-type beyond a certain amount of time spent interacting with the collector. This would be consistent with the higher time spent per visit in the treatment group (Table 2). However, recall that we found no treatment effects on tax morale and enforcement perceptions, neither in general (Table 2) nor as a function of the level of the hard-to-observe index (Table A10).

In summary, the positive selection on composition result in Figure 5 is only consistent with the alternative settings where household types are either perfectly observable or perfectly unobservable if the cumulative likelihood that a household makes a payment as a function of time spent interacting with the collector takes a (specific) convex form. However, this convexity must be determined by factors other than those that are captured in our numerous null results for selection on observable characteristics and for perceptions of enforcement and tax morale (Sections 4 and 5).

Hot and cold spots: Figure D2 In Panel A of Figure D2, we find that the payment rate is larger in treated than in control areas for hard to observe hot spots, but smaller for hard to observe cold spots. Panel B of Figure D2 shows that the same patterns hold for bill delivery rate as for payment rate.

A lower level of payment rate in treatment areas versus control areas with specific household characteristics is inconsistent with the 'perfectly unobservable' setting, since the indiscriminate use of extra time to collect in the treatment group necessarily raises the payment rate (even if weakly) in all areas – including hot and cold spots. Similarly, the fact that the same patterns occur for bill delivery as for payment is also inconsistent with the 'perfectly unobservable' setting.

The lower payment rate in hard to observe cold spots is difficult to reconcile with the 'perfectly observable' setting. Both the hot spots and cold spots each make up approximately 2% of the sample of properties (Section D.1). While the control group may have limited access to the small set of 'hot spots' with a large concentration of high types, for reasons outlined above, it is not clear why it would spend more time than the treatment group collecting from the observable 2% out of the remaining 98% of locations that have particularly low values. This logic extends to the result for bill delivery, where properties in cold spots are less likely to receive a bill in treatment areas than in control areas.

D.3 Learning Based on Navigation: Discussion

Technology and survey knowledge The literature on spatial learning shows that there are two steps to learn in large-scale environments based on navigation (Munzer et al., 2006). Learning means paying attention to the environment, acquiring spatial knowledge and retaining relevant information from the environment. First, people acquire 'route knowledge' – the knowledge of reference places (e.g. landmarks) and the main routes that connect them. Route knowledge is based on the (ego-centric) individual perspective. Second, people can acquire 'survey knowledge' – the understanding of the spatial relationships between locations, based on an extrinsic frame of reference (that is, based on an extrinsic 'map-like' perspective that is decoupled from the individual's perspective). Research shows that survey knowledge develops from a basis of route knowledge, but acquiring survey knowledge is a strategic, time-costly and cognitively effortful choice (Pazzaglia and De Beni, 2001).

The navigational tablet encourages the building of survey knowledge for two reasons. First, the technology provides advantages over the control group which the literature has found to improve acquisition of survey knowledge. The tablet represents information through a bird's eye view, in the form of a complete and accurate aerial map that has the appropriate amount of detail at the street-level where the agent operates (Zimmer, 2004). In contrast, the control group is provided with an aggregate map that provides information on the boundaries of collection units relative to a small number of main streets – this spatial information is 'distorted' in that it is not provided at the level of detail where the agent is navigating.

Moreover, the tablet provides the agent's updated self-location and the localization of the targeted property, while no such information is provided in the control group. In a setting with limited street and property addressing, these differences are crucial: acquiring survey knowledge requires the agent to be able to relate the visual information in the field to the information provided in the navigational tool (Ludwig et al., 2014). In the absence of physical street and property addressing, self-localization is important to provide such contextual relevance and it helps the agent create context-adaptive strategies for navigation (Brugger et al., 2019). The contextual relevance also supports the agent's ability to form a 'mental map' because of self-localization; this in turn helps with retention of spatial information. By contrast, in the control group the aggregate level of detail on the map provides limited contextual relevance: almost nothing of what the official can see in the field is designated on the map, and very little of the information on the map can be referenced in the field. It is true that the location information on the individual tax bills features references to landmarks (Figure A2). The ability to see the same landmark in the field and on the bill helps, but since the landmarks are not referenced on the map, this constrains spatial learning in the control group. Moreover, studies have shown that building survey knowledge based on landmarks is possible but challenging (Jabbari et al., 2022). This is because landmarks have limited spatial utility: the view of the space changes across local areas, such that obtaining extrinsic 'map-like' information based on multiple ego-centric perspectives is difficult. In other words, the control group could also build survey knowledge, but it is more challenging without the encouragements provided by the 'complete map' in the treatment group.

Second, the tablet does not automate navigation (e.g. by calculating the most efficient route or providing step by step navigational instructions). Indeed, the navigational tablet used in the experiment scores only 2 out of 10 on the system automation scale (Parasuraman, 2010). Automated navigation has been found to increase navigation performance but decrease spatial learning (Ruginski et al., 2022). Through self-localization and localization of the targeted property, the tablet forces the collector to switch back and forth between the environment and the digital map – which is one of the most effective ways to stimulate survey knowledge (Richter and Fabrikant, 2017).

Survey knowledge and learning How does survey knowledge help towards learning about property owners' hard-to-observe payment propensity? First, survey knowledge improves collectors' spatial orientation and their understanding of the true spatial map (Taylor and Tversky, 1992). This can help treatment collectors organize their routes better as they plan follow-up visits to interact and learn. This can also help collectors recognize spatial patterns of tax compliance (either passively or actively) and uncover clusters of types with high or low propensity to pay (Sections 4 and D.1).

Second, studies in environmental psychology have found that building survey knowledge increases the agent's ability to engage with the environment and their willingness to pay attention to details in the field (Aginsky and Rensink, 1997; Kozlowsky and Bryant, 1977). By building survey knowledge, the treatment group may be more inclined to inspect the household's property and the immediate surrounding area, even as doing so is costly time-wise and cognitively effortful. As treatment collectors are stimulated to pay attention, they are more likely to find hard-to-observe clues about propensity to pay and more wiling to directly engage with the property owner.

Technology's potential impact on survey knowledge also arises because of what the control group does (not) do. Studies have shown that when map information is too incomplete or too complex, spatial learning requires too many cognitive processes and too much effort. In such settings, agents face 'spatial anxiety' and may abandon the use of any map (Cash and Gaulin, 2016). Instead, they rely on 'response strategies', where the emphasis is on route knowledge and following the well known routes that consist of a series of landmarks, specific turns, or intersections from a first-person (egocentric) perspective (He and Hegarty, 2020). Cognitive research shows that one of the most decisive elements for building survey knowledge is the individual psychological level (Ishikawa and Montello, 2006). Studies have found that primary reliance on 'route knowledge' is associated with high spatial anxiety and lower spatial knowledge effort; the building of survey knowledge has been associated with lower spatial anxiety and a strong willingness to engage in experimentation to learn in the environment (Weisberg and Newcombe, 2018).

In summary, due both to the features of the GIS-tablet and the spatial environment without it, technology may have created an environment that is conducive for the treatment collector to make the strategic and cognitively effortful choice to learn. This advantage may create a distinct difference in behaviour between groups – where treatment collectors leverage the advantage to build survey knowledge and engage in learning strategies about taxpayer types, while the control group largely avoids trying to build survey knowledge and instead remains focused on route knowledge. If survey knowledge is a requisite for substantial learning in the field, the small and statistically insignificant change over time in the control group's knowledge about household types (Figure 3) would be consistent with this interpretation.

Future research could further investigate the different components inside the learning mechanism – including the distinction between passive and active learning, and between route knowledge and survey knowledge. Doing so may ultimately require equipping the control group with some form of technology; for example, recent papers in economics that provide detailed measures of agents' navigation patterns rely on GPS-tracking devices (see, for example, Miyauchi, Nakajima and Redding (2022) and Tang (2024).