The Deterrence Value of Tax Audit: Estimates from a Randomized Audit Program

Michael Best, Jawad Shah, and Mazhar Waseem*

March 2021

Abstract

In modern tax systems audit is the sole instrument through which the tax authority can detect noncompliance and create deterrence. We exploit a national program of randomized audits covering the entire population of VAT filers from Pakistan to study how much evasion audit uncovers and how much evasion it prevents by changing behavior. While audit uncovers a substantial amount of evasion (the evasion rate among firms in the bottom three size quartiles is more than 100%), it does not deter future cheating. Examining more than ten intensive and extensive margin outcomes, we detect no effect of audit on proximate or distant firm behavior. Our results suggest audits are suboptimally utilized in checking mechanical violations of law instead of creating deterrence against evasion.

Keywords: VAT, Tax Evasion, Firm Behavior

JEL Classification: H25, H26, H32

.

*Email Addresses: Michael Best mcb2270@columbia.edu; Jawad Shah jawad.shah@uky.edu; Mazhar Waseem mazhar.waseem@manchester.ac.uk.

I Introduction

Modern tax system are based on the principle of self-assessment. Taxpayers assess their tax liability without interference from the revenue authority and report it through the tax return. The returns are considered final unless they are selected for audit. Typically, audit is the only point of contact between a taxpayer and the revenue authority and therefore the sole instrument through which the authority can punish noncompliance and create deterrence. How effectively audit does this is critical to how much revenue a country collects. Sarin & Summers (2019) estimate that in the US around $1 trillion of additional revenue can be generated by improving IRS's audit capacity. Notwithstanding its importance to tax collection, audit has received little attention from public finance researchers. Importantly, we still do not understand fully how effective audits are in uncovering tax evasion and preventing it in future.

The central difficulty in identifying audit's role in the tax evasion decision of a taxpayer is its endogeneity. Modern tax administrations use sophisticated, risk-based algorithms to target audits toward more egregious tax evaders. While such targeting helps the authority deploy its scarce audit resources optimally, it prevents researchers from estimating audit impacts cleanly. In this paper, we overcome this central identification challenge by exploiting a national program of randomized audits from Pakistan. The program covers the entire population of tax filers in the country, and we have access to three waves of such randomized audits, leveraging which we estimate tax evasion at the baseline and audit's role in preventing it in future.

The randomized audit program began in 2013. Before that Pakistan's revenue authority (FBR) used to pick cases for audit using parametric, risk-based criteria. This practice, however, was challenged before the superior courts of the country inter alia on the grounds that the criteria were confidential and likely discriminatory against some taxpayers. While these challenges were pending, the FBR could not use parametric selection and was constrained to pick audit cases using random computer ballots. It is important to emphasize that randomized audits in our setting are not a subset of audits but for three consecutive years the entire audit program of the country was randomized. We focus on VAT audits conducted under the program. The VAT return is filed every month. The high-frequency VAT data allow us to identify both immediate and distant impacts of audit on behavior cleanly.

In the standard tax compliance model, a taxpayer reports it tax liability to the government trading off the benefit and cost of tax evasion (Allingham & Sandmo, 1972). The cost of evasion here is that with some probability the government would discover evasion and would recover the evaded amount along with a penalty. The probability this event occurs with is a composite term comprising the probability of audit and the probability of detection conditional on audit. In general, these two probabilities are unknown to taxpayers, although they may have formed beliefs on these based on their past interactions with the government. In our setting, the first of these probabilities is public knowledge. Before each random ballot, the FBR informed taxpayers the fraction of population to be picked for audit. The program thus creates a clean experiment whereby only the latter component of detection probability is manipulated: a random sample of firms are exposed to audit; they learn its ability to uncover evasion and update their priors accordingly. Based on the direction of such updating, they may start paying less or more revenue.

Random audits are commonly used to estimate the extent and anatomy of tax evasion in the economy. Our aim in this paper extends beyond that. We are also interested to see if audit changes the perceived likelihood of detection, thereby causing a permanent change in behavior. We do so using a long panel of administrative tax records spanning 120 months (July 2008 – June 2018), comprising the entire population of tax filers and covering both audit findings and tax returns.

We first document the results of audit. Of the 3,482 firms audited in the first wave, a positive unpaid amount was found against 986 (28.3%). In terms of volume, the unpaid amount roughly equals 8% of the aggregate baseline tax liability of *all* audited firms. For a developing country like Pakistan the evasion rate of 8% does not seem too high but its distribution is extremely unequal. The evasion rate is only around 6% for large firms (top 25%) but more than 100% for the rest. A related finding is that the former group contributes more than 99% of the revenue remitted by audited firms at the baseline. In combination, we therefore find an extreme right-skewed distribution of tax payment and a bimodal distribution of tax evasion. There roughly are two types of firms: evaders who contribute little to revenue and nonevaders who evade little and contribute roughly the entire revenue collected in the country. We obtain similar results from later audit waves.

We next look at the effects of audit on firm behavior. We have access to multiple waves of randomized audits, and our rich dataset lets us examine both proximate and distant impacts on a variety of firm outcomes. None of these impacts, however, is significantly different from zero. We examine ten intensive margin outcomes, including reported sales, costs, and revenue and one extensive margin outcome but find no effect for any of the audit waves and at any post-audit tenure. Audit seems to have no effect on firm behavior. Nor is there any heterogeneity in this result. We use two non-parametric approaches to explore heterogeneity: (1) the standard approach of adding the treatment and firm characteristic interactions into the model, and (2) the more flexible, machine-learning based approach developed in Athey *et al.* (2019) using Generalized Random Forests. We divide firms on the basis of more than ten characteristics measured at the baseline including size, age, industry, location, and position in the supply chain, but find null effect in almost every subgroup we look at. Nor do we find any variation in results if we divide the sample on the basis of audit outcomes, comparing firms audit found positive liabilities against with the others or firms audited earlier with those audited later.

Pakistan's revenue authority could not audit all firms picked through random ballots. In addition, a few firms were audited by local tax offices on their own. To account for these violations of the experimental protocol, we also estimate the LATE parameters using initial random assignment as instrument. When the treatment effect is heterogeneous and there is selection into treatment on the unobserved gain, the LATE is informative only about the average effect on compliers (Imbens & Angrist, 1994). To show our estimates apply to a much wider population, we use the marginal treat effects (MTEs) framework (Heckman & Vytlacil, 2005, 2007), identifying a linear version of the model (Brinch *et al.*, 2017; Kowalski, 2016). The MTE functions we estimate are flat, showing that treatment heterogeneity and selection on unobserved gains are not important in our setting so that our LATE estimates have global external validity.

That audit produces no behavioral response means it does not reveal any new information to firms. Audit is a rare event. Only around 5% of firms in Pakistan undergo audit in a given year, meaning a typical firm experiences it once every twenty years. It is therefore surprising that audit does not register any change in firm priors in either directions. Reading this result together with the baseline distribution of tax evasion we uncover, we propose a simple explanation. Given the peculiar nature of VAT, the cost of hiding a transaction varies a lot depending on who the other party to the transaction is. If the other party is (1) a consumer, or (2) an unregistered firm, or (3) a firm willing to collude, the cost is typically low as such transactions do not produce third-party information. The cost of hiding a transaction, on the other hand, is typically high if the other party is an uncooperative firm. This results in an S-shaped detection probability function first suggested by Kleven *et al.* (2011) and later confirmed in other setting including the Pakistan's (Waseem, 2020a). In this world, the easy-to-detect component of the tax base is reported and the hard-to-detect component is not. Audit would change firm priors only if it goes after the latter component. Our personal interviews with auditors suggest it is usually not the case. During an audit, auditors go through returns filed by a firm line by line, verifying if each line adheres to the tax code. They, for example, see that the correct tax rate has been applied, no inadmissible input tax has been claimed, no unlawful exemption has been availed, and the tax liability has been correctly calculated. While these activities are important and are likely to result in additional revenue, they are unlikely to move firm priors on the detection probability outward.

In the existing literature, no consensus exists on the sign or magnitude of the deterrence value of audit. Earlier contributions to this line of literature are lab studies some of which do find a positive effect (see Kirchler, 2007 for a survey). But in others tax evasion increases after audit (for example Maciejovsky *et al.*, 2007). This occurs either because audit forces a downward revision of the perceived detection probability or because taxpayers irrationally believe current audit makes them less likely to face future audit, a phenomenon known as the gambler's fallacy (Gilovich, 1983) or the bomb crater effect (Mittone, 2006). Another strand of this literature manipulates one or both components of the detection probability, sending deterrence messages to a random sample of taxpayers. To maximize power, these studies usually target more noncompliant sections of the population and their results are thus not directly comparable to ours. In a recent meta analysis covering 45 such studies, done largely in rich economies, Antinyan & Asatryan (2020) find that on average the effects of such interventions are modest, increasing the probability of compliance by only 1.5-2.5 percentage points.

Another set of studies exploit random audits to estimate their effects on future behavior. Examples include Gemmell & Ratto (2012), DeBacker *et al.* (2013), DeBacker *et al.* (2018), and Advani *et al.* (2019). Of these, the latter two, based in the US and the UK, find significant dynamic effects of audit: the audited taxpayers continue to pay more in years after the audit. In contrast, the former two, looking at the UK taxpayers and US corporations, report a null effect. Random audits are in general not an optimal way to allocate resources by the tax authority and these audits therefore are usually a small subset of audits done in a year. This is not the case in our setting. Our sample frame is the universe of VAT filers and our randomized sample includes all audits done in a year. Our results therefore apply to a typical firm in the VAT net with the audit done under conditions (managerial oversight, intensity of audit, political economy, etc.) a typical audit would be done under. The scale of the intervention also means our estimates are robust to external validity concerns randomized studies face commonly, arising for example from ignoring the general equilibrium effects (Muralidharan & Niehaus, 2017; Deaton & Cartwright, 2018).

Tax evasion has received renewed research interest in recent years. This revival is driven by the strong link between the economic development and fiscal capacity of a state (Besley & Persson, 2013). In part, it is also driven by the economist-as-plumber approach emphasized recently by Duflo (2017), which requires researchers to be mindful of how economic policies work in the real world. One important contribution of the paper is to use randomized audits to uncover the contours of tax evasion in a representative emerging economy. In this effort, the paper is similar to Kleven *et al.* (2011); Waseem (2020b,a) who do so in other contexts. We find substantial evasion with an extremely skewed distribution. This reinforces the point in Best *et al.* (2015) that both economic theory and public policy must take into account enforcement constraints developing countries face more seriously than is the case now.

II Conceptual Framework

In this section, we present a simple model of firm behavior to a VAT under imperfect enforcement. The primary aim of the model is to showcase the channels through which tax audit may influence future firm behavior. The model is based on a version of the Allingham & Sandmo (1972) framework presented in Kleven *et al.* (2011). We adapt the model to the VAT setting of our empirical application.

II.A Firm Behavior to Taxation

Consider a firm that uses taxable inputs valuing $c(s)$ and nontaxable inputs valuing $\psi(s)$ to produce an amount s of output. Nontaxable inputs include inputs that are not contractible, such as effort, and inputs that are not taxable, such as labor. The firm is subject to the standard VAT whereby it is required to charge tax at the rate τ of its sales and is allowed to adjust tax paid on inputs, facing a tax liability of $T(\tau) = \tau\ (s - c)$. We assume that the enforcement is imperfect so the firm can underreport its sales $\hat{s} < s$ and overreport input costs $\hat{c} > c$, evading an amount e of its tax liability $e = \hat{T} - T$, where $\hat{T} = \tau\ (\hat{s} - \hat{c})$.

The government runs an audit program to detect any tax evaded by the firm. In case of positive detection, the evaded amount is recovered along with a proportional penalty θ imposed on the evaded tax. The probability the government detects evasion with is $p(e)$ with $p'(e) > 0$ and $p''(e) > 0$. The true value of $p(e)$ is unknown to the firm, and its belief on the true value $\tilde{p}(e)$ may be biased, meaning it may overestimate $\tilde{p}(e) > p(e)$ or underestimate $\tilde{p}(e) < p(e)$ the detection probability. The risk-neutral firm reports its activity to the government taking its belief $\tilde{p}(e)$ and other parameters of the tax system as given, solving the following program

(1)
$$\max_{e}\ \tilde{p}(e).\pi^{A} + (1 - \tilde{p}(e)).\pi^{NA}.$$

Here $\pi^{A} = s - c(s) - \psi(s) - \theta\tau e$ and $\pi^{NA} = s - c(s) - \psi(s) + \tau e$ denote the after-tax profits of the firm in the detected and undetected states. The FOC of the problem

(2)
$$\left[\tilde{p}(e) + e.\tilde{p}'(e)\right](1 + \theta) = 1$$

implicitly defines the evaded amount of tax $e\,(\tilde{p}, \theta)$. The comparative statics of the problem with respect to $\tilde{p}(e)$ are unambiguous: the evaded amount decreases as the expected detection probability increases $\frac{de}{d\tilde{p}} < 0$.[1]

In simple versions of the AS model, the detection probability $\tilde{p}(e)$ is introduced in a reduced-form way. It is, however, important to note that it is a composite term, comprising two components: the perceived audit probability $\tilde{p}_a(e)$ and the perceived probability of detection conditional on audit $\tilde{p}_d(e)$. This distinction is particularly important in our setting as the randomized audit program we exploit affects both. Pakistan's revenue authority, before each wave of audit, explicitly notified the fraction of the population to be audited. The program, thus, reveals the exact audit probability (p_a) faced by firms, anchoring their expectations on the true audit probability $\mathbb{E}(\tilde{p}_a) = p_a$.

The second component of deterrence $\tilde{p}_d(e)$ is also affected by the randomized audit program. Tax audit is a rare event. It thus provide firms a rare, and to some extent the exclusive, opportunity to learn how effective government processes are in uncovering and recovering tax evasion. Based on this exchange of information,

[1]See, for example, (Kleven *et al.*, 2011).

firms would update their priors on $\tilde{p}_d(e)$, revising them upward or downward. Such updating would affect their future behavior according to equation (2).

Of the two components of the detection probability $\tilde{p}(e)$, the existing empirical literature primarily focuses on the first. A number of studies manipulate the audit probability through randomized interventions, studying its effect on both current and future tax payments.[2] By contrast, much less is known about the second component. Besley & Persson (2014) argue that weak administrative capacity of the state is perhaps the key reason developing countries tax so little. In the standard deterrence model, the administrative capacity of the state is encapsulated in $\tilde{p}_d(e)$, capturing the ability of the government to detect and recover tax evasion. Of course, this ability to a first-order depends upon the information the government receives on the tax base: whether the tax base is self-reported or third-party reported (Kleven, 2014). But even for the third-party reported bases, recent empirical literature finds that the administrative capacity of the state matters critically (e.g. Carrillo *et al.*, 2017). The parameter $\tilde{p}_d(e)$ therefore lies at the heart of what distinguishes the tax systems of developed and developing economies, holding the key to curtailing tax evasion and building stronger fiscal capacity.

In our setup, all firms learn the true audit probability $p_a(e)$ once the program begins, but only the audited firms learn the true detection probability conditional on audit $p_d(e)$. The randomized audit program, thus, creates clean variation, allowing us to unpack the black box of $\tilde{p}(e)$ and look specifically on how firms update their priors on $\tilde{p}_d(e)$ as a result of audit and how such updating affects their future tax payments. We quantify the direction and size of these movements by defining the deterrence value of audit (DV) as follows

$$DV = \frac{e(\tilde{p}_{t'}, \theta) - e(\tilde{p}_t, \theta)}{e(\tilde{p}_t, \theta)}. \tag{3}$$

The DV measures the proportional reduction in evasion caused by a marginal audit. The subscripts t' and t denote the firm's posterior and prior beliefs on the detection probability. If a firm revises its belief upward, the evaded amount shrinks $e \to 0$ and vice versa. In our setting, such revision is entirely driven by the second component of the detection probability. The sign and magnitude of the DV therefore informs if a marginal tax audit fosters or worsens tax compliance through the channel of $\tilde{p}_d(e)$ alone.

[2] See for example Kleven *et al.*, 2011; Pomeranz, 2015; Bérgolo *et al.*, 2017.

III Institutional Background

In this section, we document institutional features of the Pakistani environment that are important for our empirical analysis.

III.A Randomized Audit Program

Like all tax authorities, the FBR conducts audit of a proportion of firms each year. Before 2010, the selection for audit used to take place at the local level. Each regional tax office used to pick a proportion of firms within its jurisdiction for audit. In 2010, the FBR centralized the audit selection at the federal level. The new provision empowered the FBR to select audits on behalf of all regional offices using a computer ballot, which could be random or parametric. Exercising these powers, the FBR made its first selection based on parametric criteria in 2012. The selection, however, was challenged before the superior courts of the country inter alia on the grounds that the FBR did not reveal its selection criteria and that the selection was discriminatory to a section of taxpayers. While these challenges were pending, the FBR could not use parametric criteria for selection and was constrained to pick cases using random computer ballots. The legal complications preventing parametric selection continued till the end of 2015, and thus for the next three years audits were assigned randomly. It is important to emphasize that random audits in our setting are not a small subsample of total audits, but for three consecutive years (2013-2015) the entire audit program of the country was randomized.

The FBR issued an audit policy each year before the random ballot, setting out among other things the proportion of the population to be audited and the eligible sample for the draw. The first piece of information, as we not above, anchors firms' expectations on the true audit probability $\mathbb{E}(\tilde{p}_a) = p_a$. The eligible sample for each draw consisted of all firms other than those expressly excluded. Such exclusions were fairly minor in the first two audit waves, as only government departments and firms already under audit were excluded. The set of exclusions, however, was expanded in the third wave, and firms paying the VAT under fixed and withholding type regimes were also excluded. The required number of cases were picked randomly from the eligible sample after stratifying it on the corporate and non-corporate dimension (please see **?**FBR, 2015 for details on the randomization procedure and the set of exclusions). The ballots were held in public in the presence of taxpayers representatives, and the list of drawn cases was placed on the FBR portal. The list contained only the 9-digit firm identifier and did not include any publicly known information about the firm such as its name or address. Both audit policies and lists of drawn cases are public information and have been available on the FBR portal for view and download.

The drawn cases were simultaneously communicated to local tax offices for initiating audits. While these audits were conducted by the regional offices, the FBR maintained central oversight through the newly developed Taxpayers' Audit Monitoring System (TAMS).[3] In addition to the centrally assigned audits, regional tax offices could also initiate audits in special circumstances, such as when they received specific information on tax evasion. Before initiating such an audit, it was, however, mandatory upon the Commissioner to inform the taxpayer in writing the grounds for initiating the audit.

Table I reports descriptive statistics of the five audit waves in our sample. For our empirical analysis we use only the first three, where audit was assigned through a random ballot. The fraction of population picked for audit (p_a) varied across waves, ranging between 5% and 12%. The FBR, however, did not have the capacity to take up audits of all selected cases, and the actual audit rate remained below 100% in all years (70% for the first wave and significantly lower in the later). As we not above, regional tax offices conducted a small number of audits initiated on the basis of their own information. These audits are listed in the last column of the table. Our empirical framework takes into account these two violations of the experimental protocol namely that the audit rate was less than 100% and that some audits not assigned through random ballots were carried out.

Table II shows audits were initiated soon after assignment. Almost 65% of audits assigned in the first ballot were initiated within one month of the draw. This ratio was even higher for the later waves. A significant amount of nonpayment was detected by audits. The distribution of the detected amount, however, was strongly skewed rightward, and the median detection in all three waves was zero. We present a more detailed analysis of the audit findings in section V of the paper.

[3]TAMS was the new audit portal of the FBR. All processes related to audit, including all communications to taxpayers, were to be handled through it. This meant the FBR could monitor the progress of audits, compare it across regional offices, and take action in case of delinquency.

III.B Pakistani VAT System

Pakistani VAT largely follows the standard design. Firms with annual turnover above the exemption threshold are required to register with the tax administration.[4] Firms not required to register can do so voluntarily. While registered, whether voluntarily or otherwise, firms are required to charge VAT on their sales and are allowed to adjust the tax paid on inputs. In case the adjustment exceeds the output tax, they can carry forward or obtain the refund of the balance amount. The tax is destination-based: imports into the country are taxed at the standard rate and exports are zero-rated.

Firms are required to file a return and remit the tax due every month.[5] The filing is based on the principle of self-assessment and there is no nonaudit contact between taxpayers and tax collectors. Filed returns are considered final unless selected for audit. Audit, thus, is the sole instrument through which the revenue authority can detect and deter noncompliance.

Pakistan's revenue authority, FBR, is composed of a head office, located in Islamabad, and multiple regional office located throughout the country. These regional offices include four Large Taxpayers Units, two Corporate Regional Tax Offices and twenty Regional Tax Offices. Random audits in our sample were assigned by the head office and were completed at the regional offices. An audit team typically consists of two auditors who report to the local hierarchy. The central audit office, located at the FBR headquarter, exercises overall oversight through the online monitoring system (TAMS). Importantly, all written communication with taxpayers has to be routed through it and is considered invalid unless it contains bar code issued by the TAMS (FBR, 2015).

Revenue authorities conduct multiple types of audits, which vary in terms of the degree of engagement with the taxpayer, such as desk audits or comprehensive audits. All random audits in our sample are comprehensive audits. In each case, the taxpayer was notified that their audit has been selected through the random ballot. The records were called and examined, and the results were entered into the TAMS.

Like other developing economies, tax evasion is a major issue in Pakistan. In a recent paper, Waseem (2020b) estimates an evasion rate of 35-40% among the VAT

[4] Exemption threshold is applicable to manufacturers and retailers only. For manufacturers, it was PKR 1 million in 1998, and was increased to 2.5 million in 1999 and to 5 million in 2004. For retailers, it remained at PKR 5 million throughout the sample period.

[5] Some small firms in some of the periods included in our sample were allowed to file on a quarterly rather than monthly frequency.

filers of the country. The tax evasion occurs through both undeclared sales and over-claimed tax credits. Given a nontrivial amount is evaded in the country, tax audits indeed have the potential to shift firms' priors on the likelihood of detection outward, creating abiding deterrence against future noncompliance.

In terms of tax evasion and quality of its institutions, Pakistan is not different from other developing and emerging economies. Gómez Sabaini & Jiménez (2012), for example, estimate the VAT evasion rate among a host of Latin American economies. These evasion rates are quite similar to the Pakistan's.[6] Similarly, Pakistan's score on the Ease of Doing Business of 59.51 is virtually the same as the average score of 59.06 among all countries excluding the High Income ones (World Bank, 2020).[7] Nor is Pakistan an atypical country in terms of its tax morale. In fact, Pakistan's score on the tax morale question in the World Value Survey is better than the world's average (Haerpfer *et al.*, 2020).[8]

III.C Data

We use administrative data from Pakistan that include the universe of VAT returns filed in the country between July 2008 and June 2018. The VAT return consists of three main sections. In the first section, firms report the aggregate value of their sales, breaking it down into export and domestic sales. In the second section, the aggregate value of inputs purchased are reported, divided likewise into the two components. In the final section, firms calculate their tax liability, indicating the tax charged on sales, the tax credited on inputs, and the final tax payable.

Each firm in the VAT net is assigned a unique registration number and is expected to file every tax period. The data, therefore, have a panel structure. In addition to the return data, we use information on firm characteristics from the tax register. This information includes the tax office whose jurisdiction the firm falls under, the business organization of the firm (corporate vs. noncorporate etc.), its date of registration and other variables we exploit in our heterogeneity analysis. Appendix A.1 provides the

[6] For example, the VAT evasion rates of Guatemala, Nicaragua, Panama, and Peru are 37.5%, 38.1%, 33.8%, and 37.7%. These are withing the range for the Pakistan's estimate.

[7] The Ease of Doing Business score is widely used as a measure for the quality of institutions of a country (see for example Besley & Persson, 2014).

[8] We refer to the Question 180 on the World Value Survey 2017-2021. The question asks respondents if "Cheating on taxes if you have a chance" is justified, with responses varying from 1 (never justifiable to 10 always justifiable). Pakistan's average score on the question is 1.967, which is better than the world's average of 2.197.

detail of these variables.

Finally, we use audit data available on the FBR portal and the TAMS. As we note above, the list of cases drawn in each computer ballot is public information. We download it from the FBR portal and merge it with the VAT return data using the unique firm identifier. We are able to merge 43,465 out of 43,625 audits in our sample. For the remaining 218 cases, the firm identifier indicated in the list is incorrect. We add the audit information from the TAMS to this dataset. This information includes the date the audit was initiated, the type of audit (randomly assigned vs. locally assigned), and the amount detected.

IV Empirical Strategy

Our primary goal in this paper is to estimate the deterrence value of audit defined in equation (3). Since the VAT can be evaded by underreporting sales ($\hat{s} < s$) or overreporting input costs ($\hat{c} > c$), the DV in our setup takes the following form

$$DV = \frac{\hat{s}(\tilde{p}_{t'}, \theta) - \hat{s}(\tilde{p}_t, \theta)}{\hat{s}(\tilde{p}_t, \theta)} - \frac{\hat{c}(\tilde{p}_{t'}, \theta) - \hat{c}(\tilde{p}_t, \theta)}{\hat{c}(\tilde{p}_t, \theta)}. \tag{4}$$

We can compute the two terms on the RHS by estimating how reported sales and input costs respond to a tax audit, running regressions of the following type

$$y_i = \alpha + \beta\, assign_i + corporate_i + \epsilon_i, \tag{5}$$

where y_i is the log of reported sales or input costs, $assign_i$ denotes that firm i's audit was assigned through a random ballot, and $corporate_i$ is a dummy indicating that the firm is a corporation. For space consideration, we sometimes denote the $assign_i$ dummy simply as Z_i. Since audits in our sample are assigned randomly on stratified corporate and noncorporate samples, $\hat{\beta}$ from these regressions identifies the causal effect of audit. For most of our results, however, we exploit the long panel of records we have access to and estimate the following difference-in-differences model

$$y_{it} = \mu_i + \gamma\, assign_i \times after_t + \lambda_t + \varepsilon_{it}, \tag{6}$$

where μ_i and λ_t are the firm and tax period (month) fixed effects, and $after_t$ denotes tax periods following the assignment of audit. Note we do not need to add the corporate dummy here as the business organization of a firm does not change in our sample.[9] Relative to the baseline specification (5), the DD model offers us greater transparency (we are able to show audit impacts through visual event-study charts spanning multiple pre- and post-assignment periods) and precision (reduced standard errors). Standard errors in our baseline results are clustered at the firm level, but we also run specifications where we cluster at the tax office level (Abadie *et al.*, 2017).

Since we have imperfect compliance with the experimental protocol in our setup, $\hat{\gamma}$ from above regressions identifies the intention-to-treat effect (ITT) of audit. We also estimate the corresponding LATE parameter by instrumenting audit with initial random assignment. When the treatment effect is heterogeneous and there is selection into treatment on the unobserved gain, the LATE is informative only about the average effect on compliers (Imbens & Angrist, 1994). Compliers are an interesting population in our setup. They are firms the tax authority would like to audit whenever they have spare audit capacity available. Notwithstanding the policy-relevance of LATE, it is important to know the average effect of audit among the population of VAT filers. For this purpose, we also estimate the marginal treat effect (MTE) of audit following the framework developed in Heckman & Vytlacil (2005, 2007). Because we have access to a binary instrument only, we cannot identify the MTE nonparametrically and do so assuming linearity of MTE (Brinch *et al.*, 2017; Kowalski, 2016).

In addition to the average, we are interested in exploring heterogeneity in audit effects. We do so using two nonparametric methods. First, we estimate triple-difference versions of model (6), interacting the DD term with dummies for predetermined firm characteristics. We supplement this method with a machine learning based approach developed in Athey *et al.* (2019). Our date is large in terms of both its length and width. The richness of the data allows us to explore heterogeneity more flexibly using Generalized Random Forests.

Table III runs balance tests on our baseline data. We compare ten VAT outcomes and ten firm characteristics at the baseline across firms drawn in a given random ballot ($Z_i = 1$) with others using model (5). The compared groups are very similar for the first two waves: the difference in means is almost always insignificant or trivial. This, however, is not true for the third wave as the drawn firms are on average different from the others. They, for example, are larger and more likely to be man-

[9] The tax code requires a firm that changes its business organization from non-corporate to corporate and vice versa to re-register. Upon re-registration, a new identifier is issued to the firm.

ufacturers. These differences are unlikely to have arisen by chance. We have noted in section III.A that the set of exclusions from the eligible sample was significantly expanded for the third wave. Importantly, firms paying VAT under fixed and withholding regimes were excluded from the eligible sample. We do not identify these firms in our data and are thus unable to replicate the eligible sample for the third wave. For this reason, we focus solely on the first two waves for our empirical results. Nevertheless, for the sake of completeness we always present our main results for the third wave as well.

V Tax Evasion at the Baseline

We begin our empirical analysis by describing audit findings, focusing in particular on unpaid VAT detected by them. Because audits we consider were randomly assigned, the detected amount potentially represents an unbiased estimate of the average evasion rate at the baseline.

Table IV presents the results. All amounts in this table are in PKR billions. The top row shows that 3,482 firms picked through the first the first random ballot were audited. These firms reported aggregate turnover of nearly 500 billion in the baseline year. The audits detected 2.15 billion of short payment against them, which constitutes 0.45% of their turnover. The last two columns report VAT paid by these firms at the baseline, showing they paid 28.16 billion with an average effective tax rate of 5.65%. The unpaid revenue therefore amounts to nearly 8% of the reported tax liability (column 7).

This average evasion rate masks considerable heterogeneity across firms. The next five rows explore this heterogeneity. In the second row, we reduce the sample to the 28% firms against whom a positive amount was detected. The evasion rate among this group is 67%. The next four rows divide firms into four quartiles based on their annual turnover in the baseline year. The evasion rate is remarkably heterogeneous across the size dimension. It is strictly greater than 100% in the bottom three quartiles but only 6% in the top quartile. The top-quartile firms also contribute disproportionately to the tax revenue. Of the 28.16 billion VAT paid by the audited firms at the baseline, more than 99% (27.91 billion) was paid by them. We find qualitatively similar results for the second audit wave, although the evasion rate among the top-quartile firms is even lower for this wave.

The distribution of revenue and evasion we report here is typical of a developing country VAT (Bird & Gendron 2007; International Tax Dialogue 2013). Large firms tend to have transparent accounting mechanisms within the firm. These mechanisms let them operate at their economically optimal scale, but make commonly used strategies to achieve evasion such as cash payments and keeping double books of account infeasible.[10] As a result tax evasion is lower among large firms, who end up paying a disproportionately large chunk of revenue. To this extent, our results are consistent with recent economic theory (see for example Kleven *et al.*, 2016; Gordon & Li, 2009; Kopczuk & Slemrod, 2006).

In the audit data, the detected amount is reported in six heads. Table A.I presents breakdown of the detected amount into its major heads. Less than 2% of the detected amount is recovered at the time of audit either by direct payment (column 2) or by curtailing the taxpayer's refund claim (column 7). The rest of the amount, categorized recoverable, is contested by the taxpayer. It is therefore subject to quasi-judicial adjudication and appeal processes and can be recovered only after these processes have ended. We do not have data on the outcomes of these processes but anecdotal evidence suggests they are long, leaky, and cumbersome so that only a fraction of the recoverable amount materializes in the end.[11]

Although audits in our sample were randomly assigned, the audit rate for both waves remained below 100%. If audits were targeted toward specific types of firms within the randomly assigned sample, it could bias the evasion rates we report above. Figure A.I explores such selection, seeing if firms audited early were systematically different from those audited later. We find no systematic correlation between the amount detected and the order in which audits were taken up. Nor is the order correlated with other firm observables (see Table A.II). A much detailed analysis of selection in audit appears later in the paper. We find no evidence of such selection: within the randomly assigned sample, audits do not appear to target any specific group. To this extent, our estimates represent unbiased estimates of noncompliance at the baseline.

Tax audits are unlikely to uncover all tax evasion. For this reason, revenue author-

[10] Without strong internal controls, firms cannot grow beyond a given scale as they may worry about pilferage and stealing by local managers.

[11] According to a recent press report a total of 76,700 cases involving a recoverable amount of PKR 1.77 trillion are stuck in litigation. Nearly two-thirds of the litigated amount (PKR 1.1 trillion) is pending internally (at the two appeal fora available within the FBR) and the rest with the superior courts of the country. For details of these numbers see here.

ities that use random audits to estimate the tax gap in the country use a multiplying factor for converting the detected evasion into their official estimate. Internal Revenue Service of the US, for example, blows up the detected amount by a factor of 3.28 to arrive at its official estimate of the tax gap. The multiplying factor is derived from a direct survey of taxpayers on tax compliance (see IRS, 1996; Kleven *et al.*, 2011 for details). We do not have access to such a multiplying factor for the VAT in Pakistan. Nor are audits in our sample *extensive* audits, done for the express purpose of measuring noncompliance in the country. They rather are routine audits revenue authorities conduct during the course of their normal operation. Our estimates therefore likely represent a conservative lower bound on the actual evasion rate in the country.

VI Impacts of Audit on Firm Behavior

We now turn to the effects of audit on future firm behavior, assessing in particular if they managed to prevent or reduce tax evasion in future periods.

VI.A ITT Estimates

We begin by presenting nonparametric evidence. Figure I plots the coefficients δ_js from the following regression

$$y_{it} = \mu_i + \sum_{j=2}^{N} \delta_j .\, 1.(tax\ period{=}j)_t + u_{it}, \tag{7}$$

where y denotes the log of variable indicated in the title of each panel. The regression is run separately for firms drawn in the random ballot ($assign_i = 1$) and other eligible firms in the sample ($assign_i = 0$).[12] We drop the dummy for the first period (July 2008) and plot coefficients on other time dummies (up to June 2018). Figure II illustrates the DD version of these plots, where we add interactions of the time and $assign$ dummies into (7) and plot the coefficients on these interactions along with the 95% confidence intervals around them. Vertical, dashed lines indicate the time the random draw was made. Given the drawn firms are a random subsample of eligible

[12] Eligible firms here include all firms other than government departments and firms already under audit. Both categories of excluded firms together constitute a small (<5%) fraction of the $assign = 0$ sample.

firms, it is unsurprising that all four outcomes evolve on a common trend for a fairly long pre-draw period. Table A.III confirms this formally by estimating model (6) on the baseline data.

Strikingly, however, the outcomes continue to evolve on the same common trend even in the post-draw period. The DD coefficients are firmly centered on zero and the 95% confidence intervals around them almost always include zero. Figures III and IV replicate this analysis for the draw held in 2014, showing the second wave of randomized audits also did not have any effect on firm behavior. The event-study results thus suggest firms on average do not update their priors on the detection probability in either directions after audit. To the extent both reported sales and input costs are utterly unresponsive, it seems audit creates no meaningful deterrence in our setting. We probe this result further by running formal, regression-based tests next.

Table V reports regression results for the first wave of randomized audits. In the top panel, we report our intention-to-treat estimates from model (6). We estimate the impact of getting picked for audit on five VAT outcomes reported by firms in their VAT returns. We look for both immediate (one-year) and medium-term (three-year) impacts. Unsurprisingly, the results are consistent with the visual evidence. None of the ten coefficients is statistically significant at the conventional level. Nor is there any systematic difference between the short- and medium-term responses. Table VI repeats the exercise for the second wave, and Table VII explores the impacts on five other VAT outcomes, focusing on the first wave only. We also examine if audits had any effect along the extensive margin, influencing the firm's propensity to file a VAT return. Table VIII conducts this exercise. We estimate model (6) using an indicator the firm filed its return for period t as the outcome variable. Finally, Table A.IV replicates the baseline analysis in Table V, clustering the standard errors at the tax office level. Results from all these specifications, covering two audit waves and looking at ten intensive margin and one extensive margin outcome, portray a consistent story: audit has no meaningful impact on future firm behavior, either in the short or in the long run.

VI.B LATE Estimates

Since the FBR did not conduct audit of all cases drawn in the random ballots, the above estimates capture the average effect of getting *picked* for audit rather than the average effect of audit. To compute the latter parameter, we estimate the 2SLS models corresponding to (6), instrumenting the endogenous variable $audit$ by the initial random assignment.[13] Table A.V reports the first stage of these regressions, illustrating that a strong first stage exists in our setting. The bottom panels of Table V through Table VIII report the LATE estimates corresponding to each specification we run. Given the reduced form of these specifications returned trivial or insignificant coefficients, it is unsurprising that the corresponding LATE parameters are also similar. The majority of these coefficients are of negative sign, statistically insignificant, and economically trivial.

We produce the parallel analysis, containing both the ITT and LATE estimates, for the third wave of audits in Figures A.II-A.III and Table A.VI. Recall that the balance tests reveal significant differences between the compared groups for this wave of randomization (see Table III). We therefore do not draw any conclusion from these results and produce them only for the sake of completeness.

VI.C ATE Estimates

When treatment effects are heterogeneous and there is selection into treatment on the unobserved gain, the LATE is informative on the average effect of the treatment on compliers only (Imbens & Angrist, 1994; Abadie, 2003). Compliers, in our setting, are firms pushed into audit by the instrument (being drawn in the random computer ballot). The LATE we identify therefore may not reflect the average effect in the population unless the impact of audit does not vary across firms or auditors do not target specific firms, using information we do not observe.

We first explore the latter of these points, asking if audits were selectively targeted toward specific types of firms. Table IX compares the characteristics of audited and unaudited firms.[14] Audited firms here include both whose audit was assigned by a random draw ($Z_i = 1$) and whom the local tax office picked based on their own information ($Z_i = 0$). Tables X-XI separate the analysis for the two subgroups. A typical audited firm indeed differs from the unaudited in terms of observables we look at (Table IX). But these differences are almost entirely driven by the small group of firms local tax offices picked for audit ($Z_i = 0$) using their own information. Within the random-assignment group ($Z_i = 1$), audits do not seem to be targeted toward any

[13] For brevity, we sometimes denote $audit$ variable simply as D in the subsequent sections.

[14] Since audits were done at the local tax office, we need to compare audited and unaudited firms within a tax office to rule out selection. We therefore include tax office fixed effects into these regressions.

specific subgroup. Figures V-VI compare audited and unaudited firms in our event study framework (7). Since the specification includes firm fixed effects, the results capture any residual selection into audit not explained by the firm's fixed characteristics, such as size or industry. There does not appear to be any such residual selection as the reporting histories of the audited and unaudited groups are indistinguishable from each other. Table A.VII establishes this running formal tests on the baseline data. Parallel trends for a long preaudit period mean our DD estimator remains internally valid and applies to all audited firms.

That selection on the unobserved gain may not be too important in our setting is also demonstrated by two pieces of evidence we presented earlier. First, the compliance rate varies from 70% in the first wave to 30% in the second, yet we see no difference between the corresponding LATE estimates (compare Tables V and VI). This suggests the marginal firm pushed into audit may not be significantly different from others within the randomly assigned ($Z_i = 1$) sample. Second, Figure A.I shows the amount detected is not systematically correlated with the order in which audits were taken up. This suggests audits are not systematically targeted toward firms with potentially large gains for the revenue authority in terms of the detected or recovered amount. Auditors do not seem to have any information or inclination for such selection.

Continuing our effort to go beyond LATE, we next exploit the marginal treatment effect (MTE) framework popularized by Heckman & Vytlacil (1999). The MTE, in our setting, captures the audit effect on the marginal firm the revenue authority is indifferent between auditing and not auditing. The approach has the useful feature that all conventional treatment parameters can be expressed as different weighted averages of the MTEs. Since we have access to a binary instrument only, we cannot identify the MTE function nonparametrically. Instead, we follow Brinch *et al.* (2017) and Kowalski (2016) to identify a linear version of the model that allows for both treatment heterogeneity and selection based on the unobserved gain from treatment.

Figure VII shows the $MTE(p)$ functions we estimate using the first randomization wave as instrument. The blue curve shows the potential outcomes of unaudited firms. Along the horizontal axis the potential fraction audited (p) increases from no audit to full audit. Given that unaudited firms include never-takers and unaudited compliers, this curve is identified at two points, indicated by square markers in the plots. The green curve, on the other hand, shows the potential outcomes of audited firms. Because audited firms include audited compliers and always-takers, this curve is also identified at two points, indicated by circular markers in the plots. We extrapolate the two curves using the linearity assumption. The $MTE(p)$ is the difference between the two curves. Since in our setting all three curves lie above each other, we lift both $MTO(p)$ and $MUO(p)$ up by adding the constant from the corresponding regression to distinguish them from the $MTE(p)$. Figure VIII repeats the analysis for the second randomization wave. The technical details of the estimations are in Appendix A.2.

Unsurprisingly, the MTE functions we estimate are flat. The change in the unaudited outcomes as the potential fraction audited increases reflects selection. The gradient in the audited outcomes, on the other hand, reflects selection and audit effect heterogeneity. That both these curves are flat rules out these factors in our setting. Note the functional form assumption we make to identify the MTEs is not too restrictive. Because we have access to two waves of randomized audits, we have more information than is typically available in an RCT. Specifically, as the compliance rate varies between the two waves, both audited and unaudited outcomes in our setting are identified at four rather than two points. That the $MTE(p)$ is flat across these two waves and four outcomes we explore provides strong evidence that the LATE has global external validity in our setting.

VI.D Heterogeneity in Audit Effects

To strengthen the above conclusion, we also look at audit effect heterogeneity directly. We do so using two nonparametric approaches. First, we estimate triple-difference versions of model (6), interacting the DD term with firm traits. We explore eight traits introduced into the model as dummies indicating (i) firm size; (ii) firm age; (iii) firm location; (iv) local tax office having jurisdiction over the firm; (v) the type of local tax office (LTU vs. RTO etc.); (vi) firm's position in the supply chain (manufacturer vs. wholesaler etc.); (vii) firm's business organization; and (viii) industry the firm operates in. All these traits are measured at the baseline before the announcement of ballot results, and we estimate the model separately for the two audit waves. Figures A.IV-A.XI display the results. We do not find any systematic treatment effect heterogeneity across the subgroups we compare. The 95% confidence interval almost always includes zero, showing that the response of each subgroup is statistically indistinguishable from that of the omitted category.

In addition to the predetermined firm traits, we also explore heterogeneity by the timing and outcome of audit. Figure A.XII divides audited firms into ten groups, depending upon the time lag between the assignment and initiation of audit. If auditors have hidden information they use to target specific subgroups, it would be reflected in the order they took up the assigned audits in. We, however, do not see any differences along this dimension. Audited firms in all deciles appear to be very similar. Table A.VIII stratifies the audited sample by the detected amount, looking for any differential effect upon firms auditors did find an underpaid amount against. Here also we do not find any differential effect.

Finally, we explore heterogeneity in treatment effects using a more flexible, machine learning based approach. We ask if the audit effect varies with the firm's predetermined traits using the Generalized Random Forest algorithm developed in Athey *et al.* (2019).[15] To reduce the computational demands of the algorithm, we use the simple difference-in-means model (5) as the baseline rather than the difference-in-differences model (6) we have been using so far. The results are in Figures A.XIII-A.XXII. The first four of these figures show the audit effect does not vary with firm size or age. The rest of the figures explore binary traits. Again, we do not find any systematic heterogeneity in the audit effect along any of the eight traits we look at.

VII Why Does Audit Not Influence Behavior?

We present extensive evidence showing audit has no effect on future firm behavior. Not only does this finding hold on average but also within every subgroup we define based on firm observables. It means either that audit reveals no new information to firms or that information revealed by it is consistent with their priors. Audit is a rare event. During the period we consider, the FBR could not audit more than 5% of the population in any year. At this rate a typical firm is likely to experience audit once every twenty years.[16] Nor is there any significant nonaudit contact between firms and the revenue authority.[17] No updating in either directions is therefore puzzling,

[15]In the approach, individual trees are grown by greedy recursive partitioning of the sample space, with each split chosen to improve the model fit. The trees are then randomized using bootstrap aggregation, whereby each tree is grown on a different random subset of the training data, and random split selection that restricts the variable available at each step of the algorithm.

[16]The likelihood of a firm facing the audit is endogenous to firm behavior if the authority runs a parametric, risk-based system of audit selection. The raw audit probability is for illustrative purpose only, showing that on average the authority can only audit one-twentieth of the population each year.

[17]Some tax authorities such as Indonesia's assign a tax manager to each firm. The responsibility of the manager is to keep a tab on the firm's tax affairs and contact it if some discrepancy is observed

suggesting that even before audit firms know the detection probability they face with certainty $\tilde{p}_d(e) = p_d(e)$. In this section, we make sense of this result.

We begin by emphasizing two features of evasion we uncovered in section V. First, while a substantial amount of evasion exists in the country, its distribution is highly skewed. The evasion rate is more than 100% in the bottom three size quartiles but only 6% in the top. Second, the top quartile firms remit virtually all the VAT collected in the country. These features suggest roughly a bimodal distribution of tax evasion. There are essentially two types of firms in terms of their propensity to cheat: evaders who evade everything and nonevaders who evade little and contribute nearly the entire tax revenue.

Our starting point in explaining these results is equation (2). The equation pins down the evasion choice of firms, showing they will evade up to the point the marginal cost of evasion (the LHS) equals its benefit (the RHS). We make one change to the baseline model to fit it to our empirical findings. We recast the model in terms of transactions, making evasion a discrete choice.[18] This allows us to generate the bimodal distribution of evasion we note above. The firm engages in L transactions, which we index by $l = 1...L$. For each transaction, the firm faces the binary choice of whether to report it or not. If the firm decides to report, it will have to remit the tax due. Alternatively, it may hide the transaction, evading the tax due amounting to $e_l = \tau(s_l - c_l)$. Hiding the transaction is optimal if the following condition is satisfied

$$\left[\tilde{p}_l(e_l) + e_l.\tilde{p}_l^{'}(e_l)\right](1+\theta) < 1. \tag{8}$$

The condition is a discrete choice version of equation (2). Ordering transactions in terms of the detection probability $\tilde{p}_l(e_l)$ they entail, we define L^* as the first transaction for which the above inequality fails. It means the firm will report transactions $L^*...L$, remitting revenue amounting to $\int_{L^*}^{L} \tau(s_l - c_l)d(l)$. Note that L^* could be the first transaction, in which case the firm does not evade at all, or it could be the last, in which case the firm evades all the tax due. In general, L^* would be idiosyncratic to firms, depending on factors such as their scale, industry, and trading network. It,

(Basri *et al.*, 2019). It is not the case in Pakistan. The Pakistani revenue authority is organized on functional lines with no single official responsible for the affairs of a firm. In general, a firm is not contacted other than for the purpose of audit.

[18]Note this does not make the model restrictive. We can always define transactions narrowly to cover any situation. If, for example, the firm hides a transaction partially, we can divide it into two, making evasion a binary choice in each part. We follow Saez (2002) and Basri *et al.* (2019) in recasting the choice of a continuous variable as a discrete choice.

for example, is likely to be small for large firms given their internal accounting mechanism make them less able to hide transactions. The model can thus fit the bimodal distribution of evasion we noted above.

A critical element of this model is how $\tilde{p}_l(e_l)$ varies within firms, across transactions. Note hiding a transaction would be easier for the firm if the other party to it is (1) a consumer, (2) an unregistered firm, or (3) a firm willing to collude. In these cases, the firm can cover its tracks, making it difficult for the government to uncover evasion. On the other hand, hiding a transaction would be harder if the other party is unwilling to collude, such as a firm that cannot handle unaccounted cash and therefore cannot keep a transaction out of books.[19] The $\tilde{p}_l(e_l)$ faced by the firm therefore typically has the shape shown in Figure IX. The probability of detection is low for the former type of transactions but turns sharply once the latter type begins. Such an S-shape function was first proposed by Kleven *et al.* (2011) and has since then confirmed in other empirical settings (see Waseem, 2020a for one such example). The shape reflects the fact that the likelihood of detection to a first order depends on the third-party information an economic transaction generates for the government.

The discrete choice model predicts simple behavioral rules. Firms would report transactions entailing high detection probability $[L^*, L]$, hiding the rest. In this world audit would not cause updating of firm priors unless auditors go after hidden transactions $[1, L^*)$. But going after such transactions is likely to cost more resource as they do not leave any information trails. This dilemma lies at the heart of the government's audit conundrum. By moving the perceived detection probability up, uncovering hidden transaction would increase revenue in all future periods, but the net current period rewards from doing so are likely to be small. It therefore may not be optimal for policy makers to gear audits toward uncovering hidden transactions.

In our data, we do not observe activities auditors perform during an audit, but personal interviews with them reveal that audits indeed are geared more toward checking mechanical violations of law. During an audit, auditors go through returns filed by a firm line by line, verifying if each line adheres to the tax code. They, for example, see that the correct tax rate has been applied, no inadmissible input tax has been claimed, no unlawful exemption has been availed, and the tax liability has been correctly calculated. While these activities are important and are likely to result in

[19]These consideration can lead to segmentation of firms into good and bad VAT chains with compliant firms dealing with compliant firms only and vice versa. See de Paula & Scheinkman (2010); Gadenne *et al.* (2019); Gerard *et al.* (2019) for empirical evidence on market segmentation caused by a VAT.

additional revenue, they are unlikely to move firm priors on the detection probability outward.

A testable prediction of our conjecture that auditors devote little attention to transactions not reported by firms is that the detected amount will fall as the proportion of such transactions in a firm's sales go up. Table A.IX tests this prediction. We divide firms into four groups based on the share of final sales reported by them at the baseline. Final sales are transactions where the other party does not possess a national tax number: they are either final consumers or informal firms. Theory predicts that the incidence of evasion will be higher on such transactions, and audit therefore must detect a greater amount payable against firms with a higher share of such transactions. But it is not what we find. The amount detected in fact falls as the share of final transactions in a firm's sales rises. This finding holds as we add important controls to the model including firm size. The evidence thus supports our explanation that much of the audit effort goes into reconciling reported transactions rather than uncovering the unreported ones.

That audit has little deterrence value has profound tax policy implications. Importantly, it reinforces the point made among others by Gordon & Li (2009) and Best *et al.* (2015) that enforcement environments of developing and developed countries are fundamentally different. These differences place important restrictions on admissible tax structures, especially on the choice of tax instruments. Otherwise optimal tax instruments may not remain optimal once one factors in these enforcement restrictions. Both economic theory and tax policy must tax these departures from the text book setting more seriously than is the case now.

VIII Conclusion

In modern tax systems, audit is to some extent the sole instrument through which the revenue authority can detect and deter cheating. We exploit a national program of randomized audits from Pakistan to examine how much evasion audit detects and how much evasion it prevents by changing post-audit behavior. Combining VAT returns and audit outcomes data, we find audit detects a substantial amount of evasion: the detected amount is 8% of the aggregate annual turnover of audited firms. The evasion rate, however, varies substantially across firms. It is more than 100% among firms in the bottom three size quartiles but only 6% among the rest. Despite detecting such a large amount of evasion, audit does not create any deterrence against it. Examining more than ten intensive and extensive margin outcomes, we find no significant impact of audit on immediate or distant behavior for any of the randomization wave we consider. This result is robust to a number of specification checks, and we do not find any heterogeneity in audit effects across any subpopulation.

That audit does not affect behavior is puzzling. Audit is a rare event, with a typical firm likely to experience it once every twenty years. Lack of response to it means audit does not reveal any new information to firms. We suggest a simple explanation of this result. Transactions carried out by a firm can be roughly divided into two types. Transactions with consumers, unregistered firms, or colluding firms can be hidden easily, while those with uncooperative firms cannot. In this world, profit-maximizing firms report easy-to-detect transaction but hide the rest, and audit would change firm priors only if it goes after the hidden transactions. Our interviews with auditors reveal it is usually not the case. Instead, auditors scrutinize reported transactions only, looking for any mechanical violations of law. Insufficient focus on uncovering hidden transactions means audit does not change firm priors on the detection probability and thus does not induce a permanent change in behavior.

References

ABADIE, ALBERTO. 2003. Semiparametric instrumental variable estimation of treatment response models. *Journal of econometrics*, **113**(2), 231–263.

ABADIE, ALBERTO, ATHEY, SUSAN, IMBENS, GUIDO W, & WOOLDRIDGE, JEFFREY. 2017. *When should you adjust standard errors for clustering?* Tech. rept. National Bureau of Economic Research.

ADVANI, ARUN, ELMING, WILLIAM, & SHAW, JONATHAN. 2019. The dynamic effects of tax audits. IFS working paper.

ALLINGHAM, MICHAEL G., & SANDMO, AGNAR. 1972. Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics*, **1**, 323–338.

ANTINYAN, ARMENAK, & ASATRYAN, ZAREH. 2020. Nudging for tax compliance: A meta-analysis.

ATHEY, SUSAN, TIBSHIRANI, JULIE, WAGER, STEFAN, *et al.* 2019. Generalized random forests. *The Annals of Statistics*, **47**(2), 1148–1178.

BASRI, M CHATIB, FELIX, MAYARA, HANNA, REMA, & OLKEN, BENJAMIN A. 2019. *Tax Administration vs. Tax Rates: Evidence from Corporate Taxation in Indonesia*. Tech. rept. National Bureau of Economic Research.

BÉRGOLO, MARCELO L, CENI, RODRIGO, CRUCES, GUILLERMO, GIACCOBASSO, MATIAS, & PEREZ-TRUGLIA, RICARDO. 2017. *Tax audits as scarecrows: Evidence from a large-scale field experiment*. Tech. rept. National Bureau of Economic Research.

BESLEY, TIMOTHY, & PERSSON, TORSTEN. 2013. Taxation and Development. *In:* ALAN J. AUERBACH, RAJ CHETTY, MARTIN FELDSTEIN, & SAEZ, EMMANUEL (eds), *handbook of public economics, vol. 5*.

BESLEY, TIMOTHY, & PERSSON, TORSTEN. 2014. Why Do Developing Countries Tax So Little? *Journal of Economic Perspectives*, **28**(4), 99–120.

BEST, MICHAEL CARLOS, BROCKMEYER, ANNE, KLEVEN, HENRIK JACOBSEN, SPINNEWIJN, JOHANNES, & WASEEM, MAZHAR. 2015. Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan. *Journal of Political Economy*, **123**(6), 1311–1355.

BIRD, RICHARD, & GENDRON, PIERRE-PASCAL. 2007. *The VAT in Developing and Transitional Countries*. Cambridge University Press.

BRINCH, CHRISTIAN N, MOGSTAD, MAGNE, & WISWALL, MATTHEW. 2017. Beyond LATE with a discrete instrument. *Journal of Political Economy*, **125**(4), 985–1039.

CARRILLO, PAUL, POMERANZ, DINA, & SINGHAL, MONICA. 2017. Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement. *American Economic Journal: Applied Economics*, **9**(2), 144–64.

DE PAULA, AUREO, & SCHEINKMAN, JOSE A. 2010. Value-Added Taxes, Chain Effects, and Informality. *American Economic Journal: Macroeconomics*, **2**(4), 195–221.

DEATON, ANGUS, & CARTWRIGHT, NANCY. 2018. Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, **210**, 2–21.

DEBACKER, JASON, HEIM, BRADLEY T, TRAN, ANH, & YUSKAVAGE, ALEXANDER. 2013. The Impact of Legal Enforcement: An Analysis of Corporate Tax Aggressiveness after an Audit. *Middle Tennessee State, Indiana University, and US Department of Treasury working paper*.

DEBACKER, JASON, HEIM, BRADLEY T, TRAN, ANH, & YUSKAVAGE, ALEXANDER. 2018. Once bitten, twice shy? The lasting impact of IRS audits on individual tax reporting. *Journal of Law and Economics*, **61**, 1–35.

DUFLO, ESTHER. 2017. The Economist as Plumber. *American Economic Review*, **107**(5), 1–26.

FBR. 2015. Audit Policy 2015. Federal Board of Revenue, Pakistan, Taxpayer's Audit Wing.

GADENNE, LUCIE, NANDI, TUSHAR K., & RATHELOT, ROLAND. 2019. *Taxation and Supplier Networks: Evidence from India*. Mimeo.

GEMMELL, NORMAN, & RATTO, MARISA. 2012. Behavioral responses to taxpayer audits: evidence from random taxpayer inquiries. *National Tax Journal*, **65**(1), 33.

GERARD, FRANÇOIS, NARITOMI, JOANA, SEIBOLD, ARTHUR, & ZULIAN, BRUNO. 2019. Two-Tier Tax Systems and Firms: Evidence from Brazil.

GILOVICH, THOMAS. 1983. Biased evaluation and persistence in gambling. *Journal of personality and social psychology*, **44**(6), 1110.

GÓMEZ SABAINI, JUAN CARLOS, & JIMÉNEZ, JUAN PABLO. 2012. Tax structure and tax evasion in Latin America. Macroeconomics of Development Series 118.

GORDON, ROGER, & LI, WEI. 2009. Tax structures in developing countries: Many puzzles and a possible explanation. *Journal of Public Economics*, **93**(7-8), 855–866.

HAERPFER, C, INGLEHART, R, MORENO, A, WELZEL, C, KIZILOVA, K, DIEZ-MEDRANO, J, LAGOS, M, NORRIS, P, PONARIN, E, PURANEN, B, *et al.* 2020. World Values Survey: Round Seven–Country-Pooled Datafile. *Madrid, Spain & Vienna, Austria: JD Systems Institute & WVSA Secretariat*.

HECKMAN, JAMES J, & VYTLACIL, EDWARD. 2005. Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica*, **73**(3), 669–738.

HECKMAN, JAMES J, & VYTLACIL, EDWARD J. 1999. Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the national Academy of Sciences*, **96**(8), 4730–4734.

HECKMAN, JAMES J, & VYTLACIL, EDWARD J. 2007. Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. *Handbook of econometrics*, **6**, 4779–4874.

IMBENS, GUIDO W., & ANGRIST, JOSHUA D. 1994. Identification and Estimation of Local Average Treatment Effects. *Econometrica*, **62**(2), 467–475.

INTERNATIONAL TAX DIALOGUE. 2013. International Tax Dialogue, Key Issues and Debates in VAT, SME Taxation and the Tax Treatment of the Financial Sector. *International Tax Dialogue*.

IRS. 1996. Federal Tax Compliance Research, Individual Income Tax Gap Estimates for 1985, 1988, and 1992. *Department of the Treasury Internal Revenue Service*. IRS Publications 1415 (Rev. 4-96), Washington DC.

KIRCHLER, ERICH. 2007. *The economic psychology of tax behaviour*. Cambridge University Press.

KLEVEN, HENRIK J. 2014. How Can Scandinavians Tax So Much? *Journal of Economic Perspectives*, **28**(4), 77–98.

KLEVEN, HENRIK J., KNUDSEN, MARTIN B., KREINER, CLAUS THUSTRUP, PEDERSEN, SØREN, & SAEZ, EMMANUEL. 2011. Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. *Econometrica*, **79**(3), 651–692.

KLEVEN, HENRIK JACOBSEN, KREINER, CLAUS THUSTRUP, & SAEZ, EMMANUEL. 2016. Why can modern governments tax so much? An agency model of firms as fiscal intermediaries. *Economica*, **83**(330), 219–246.

KOPCZUK, WOJCIECH, & SLEMROD, JOEL. 2006. Putting Firms into Optimal Tax Theory. *American Economic Review Papers and Proceedings*, **96**(2), 130–134.

KOWALSKI, AMANDA E. 2016. *Doing more when you're running LATE: Applying marginal treatment effect methods to examine treatment effect heterogeneity in experiments*. Tech. rept. National Bureau of Economic Research.

MACIEJOVSKY, BORIS, KIRCHLER, ERICH, & SCHWARZENBERGER, HERBERT. 2007. Misperception of chance and loss repair: On the dynamics of tax compliance. *Journal of Economic Psychology*, **28**(6), 678–691.

MITTONE, LUIGI. 2006. Dynamic behaviour in tax evasion: An experimental approach. *The Journal of Socio-Economics*, **35**(5), 813 – 835.

MURALIDHARAN, KARTHIK, & NIEHAUS, PAUL. 2017. Experimentation at Scale. *Journal of Economic Perspectives*, **31**(4), 103–24.

POMERANZ, DINA. 2015. No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax. *American Economic Review*, **105**(8), 2539–2569.

SAEZ, EMMANUEL. 2002. Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses*. *The Quarterly Journal of Economics*, **117**(3), 1039–1073.

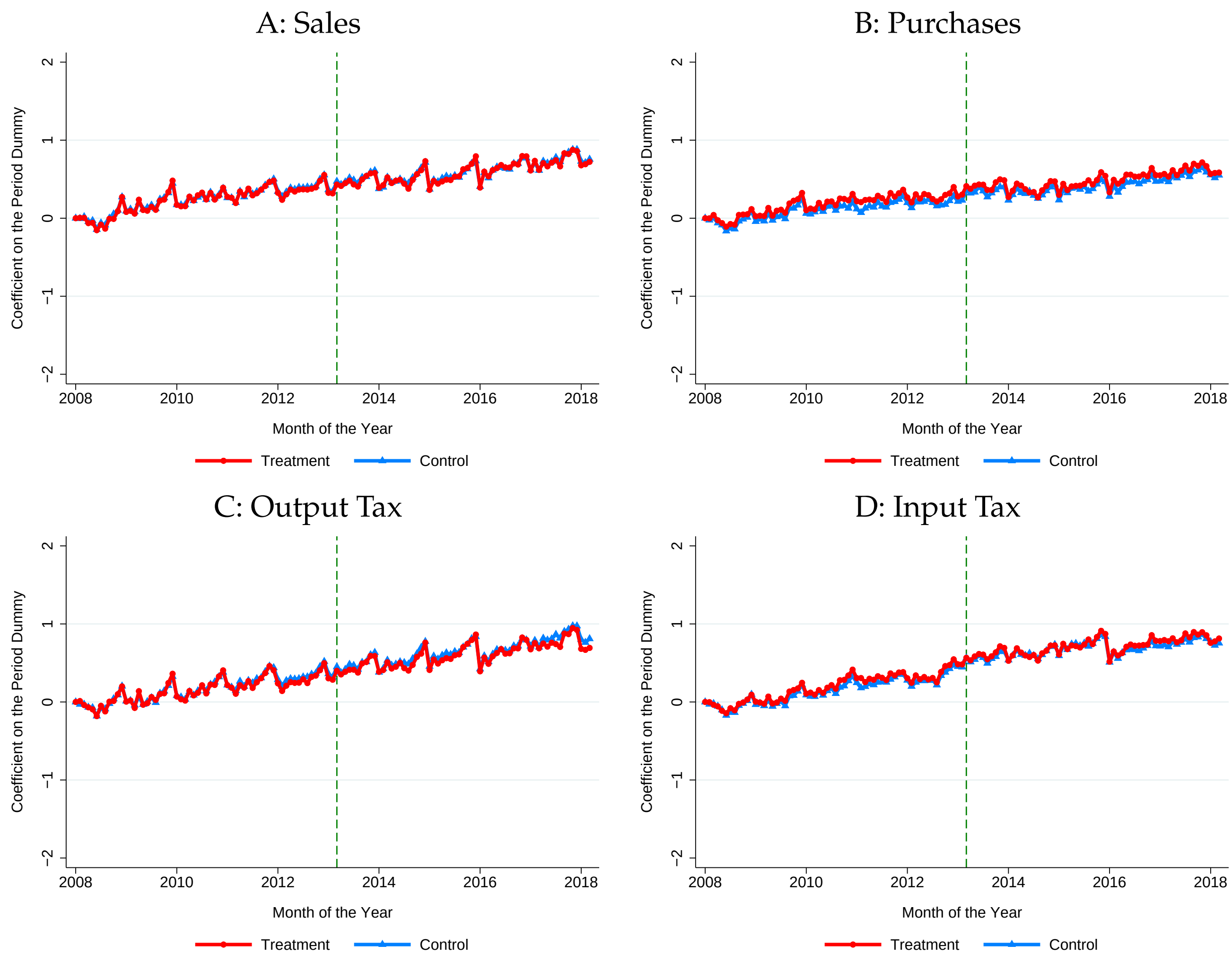
SARIN, NATASHA, & SUMMERS, LAWRENCE H. 2019. *Shrinking the Tax Gap: Approaches and Revenue Potential*. Tech. rept. National Bureau of Economic Research.

WASEEM, MAZHAR. 2020a. Does Cutting the Tax Rate to Zero Induce Behavior Different from Other Tax Cuts? Evidence from Pakistan. *The Review of Economics and Statistics*, **102**(3), 426–441.

WASEEM, MAZHAR. 2020b. Overclaimed Refunds, Undeclared Sales, and Invoice Mills: Nature and Extent of Noncompliance in a Value-Added Tax. CEPR. Discussion Paper No. 14601.

WORLD BANK. 2020. *Doing Business 2020 : Comparing Business Regulation in 190 Economies*. Tech. rept. World Bank, Washington DC.

FIGURE I: **INTENTION TO TREAT EFFECTS OF AUDIT – FIRST WAVE**



Notes: The figure explores the impacts of audit on future firm behavior. We compare the evolution of four VAT outcomes across the treatment and control groups. The treatment groups consists of firms whose audit was assigned through the first random ballot held on September 13, 2013. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. To construct these charts, we regress the log of the outcome variable shown in the title of each panel on the full set of firm and month fixed effects, dropping the dummy for July 2008. We then plot the coefficients on the time dummies of these regressions. The sample includes all tax periods from July 2008 to June 2018. The regressions are run separately for the two groups of firms. Year t on the horizontal axis indicates July of the corresponding year. Vertical dashed lines demarcate the date the random computer ballot was held on.

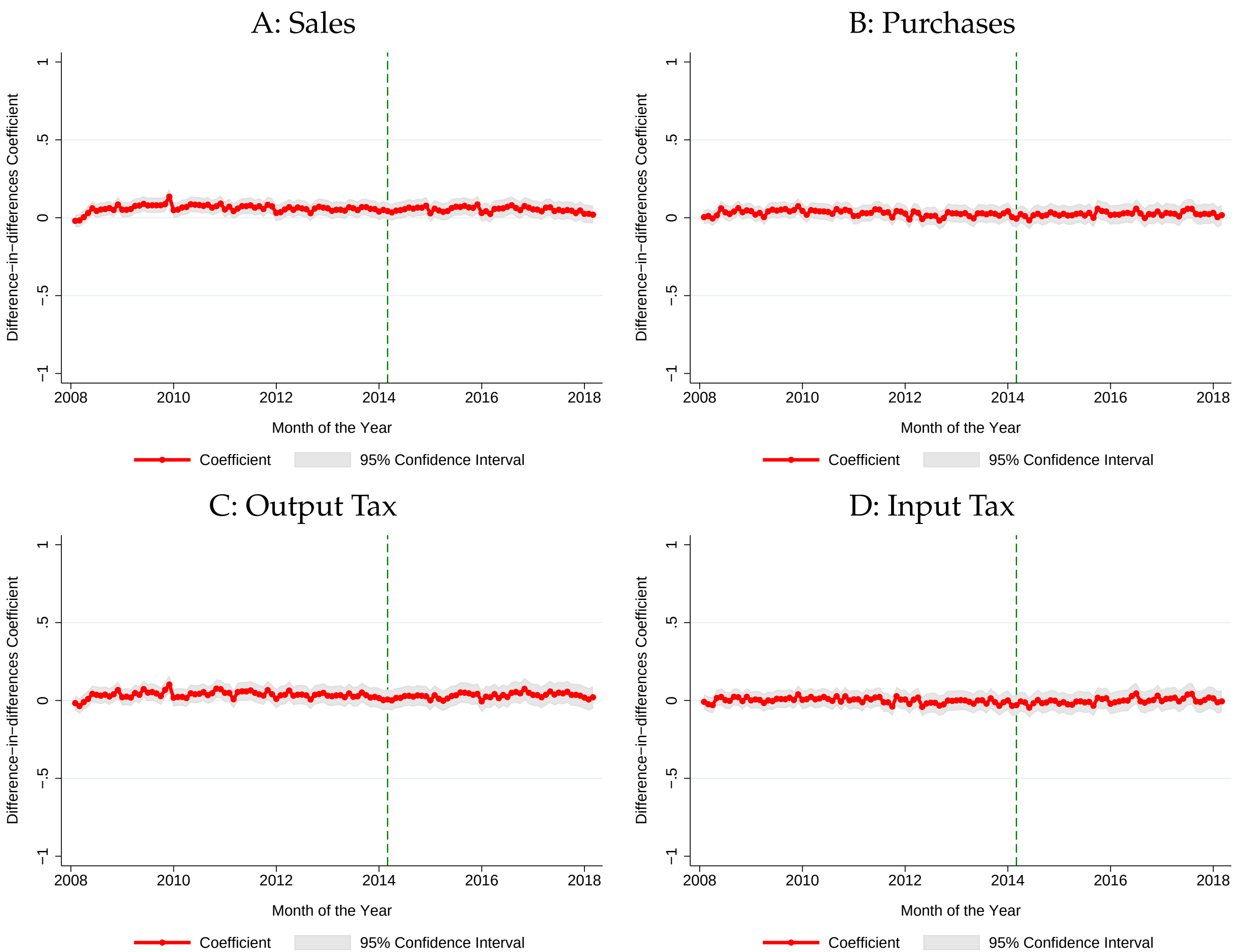
FIGURE II: **INTENTION TO TREAT EFFECTS OF AUDIT – FIRST WAVE**

Notes: The figure shows the difference-in-differences version of the plots in Figure I. To construct these charts, we regress the log of the outcome variable shown in the title of each panel on the full set of firm, month, and month×treat dummies, dropping the dummies for July 2008. We then plot the coefficients on the month×treat dummies from these regressions. The gray surface plot shows the 95% confidence interval around the coefficient. The treatment groups consists of firms whose audit was assigned through the first random ballot held on September 13, 2013. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. We cluster standard errors at the firm level. Year t on the horizontal axis indicates July of the corresponding year. Vertical dashed lines demarcate the date the random computer ballot was held on.

FIGURE III: **INTENTION TO TREAT EFFECTS OF AUDIT – SECOND WAVE**

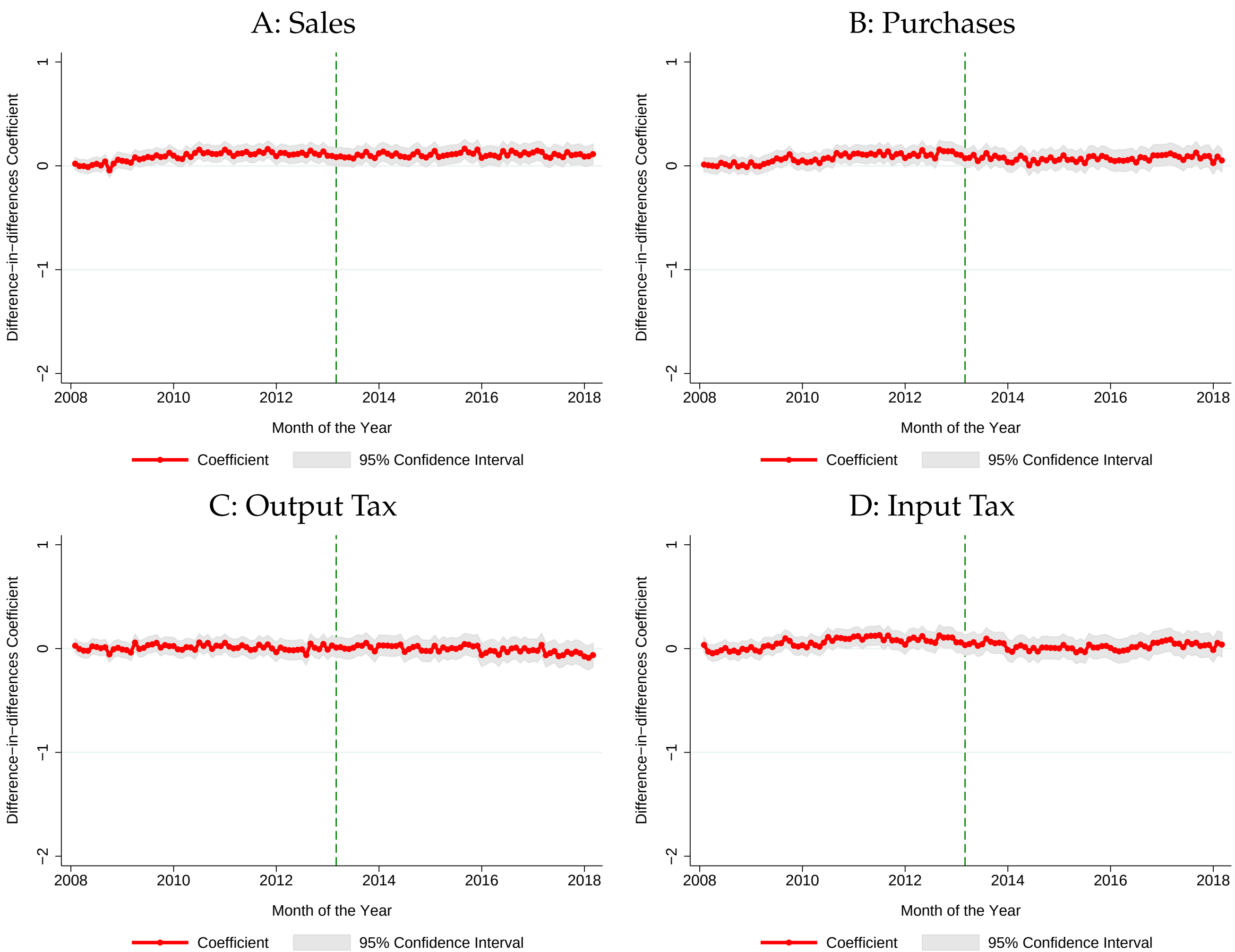
Notes: The figure explores the impacts of audit on future firm behavior. We compare the evolution of four VAT outcomes across the treatment and control groups. The treatment groups consists of firms whose audit was assigned through the first random ballot held on September 25, 2014. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. To construct these charts, we regress the log of the outcome variable shown in the title of each panel on the full set of firm and month fixed effects, dropping the dummy for July 2008. We then plot the coefficients on the time dummies of these regressions. The sample includes all tax periods from July 2008 to June 2018. The regressions are run separately for the two groups of firms. Year t on the horizontal axis indicates July of the corresponding year. Vertical dashed lines demarcate the date the random computer ballot was held on.

FIGURE IV: **INTENTION TO TREAT EFFECTS OF AUDIT – SECOND WAVE**



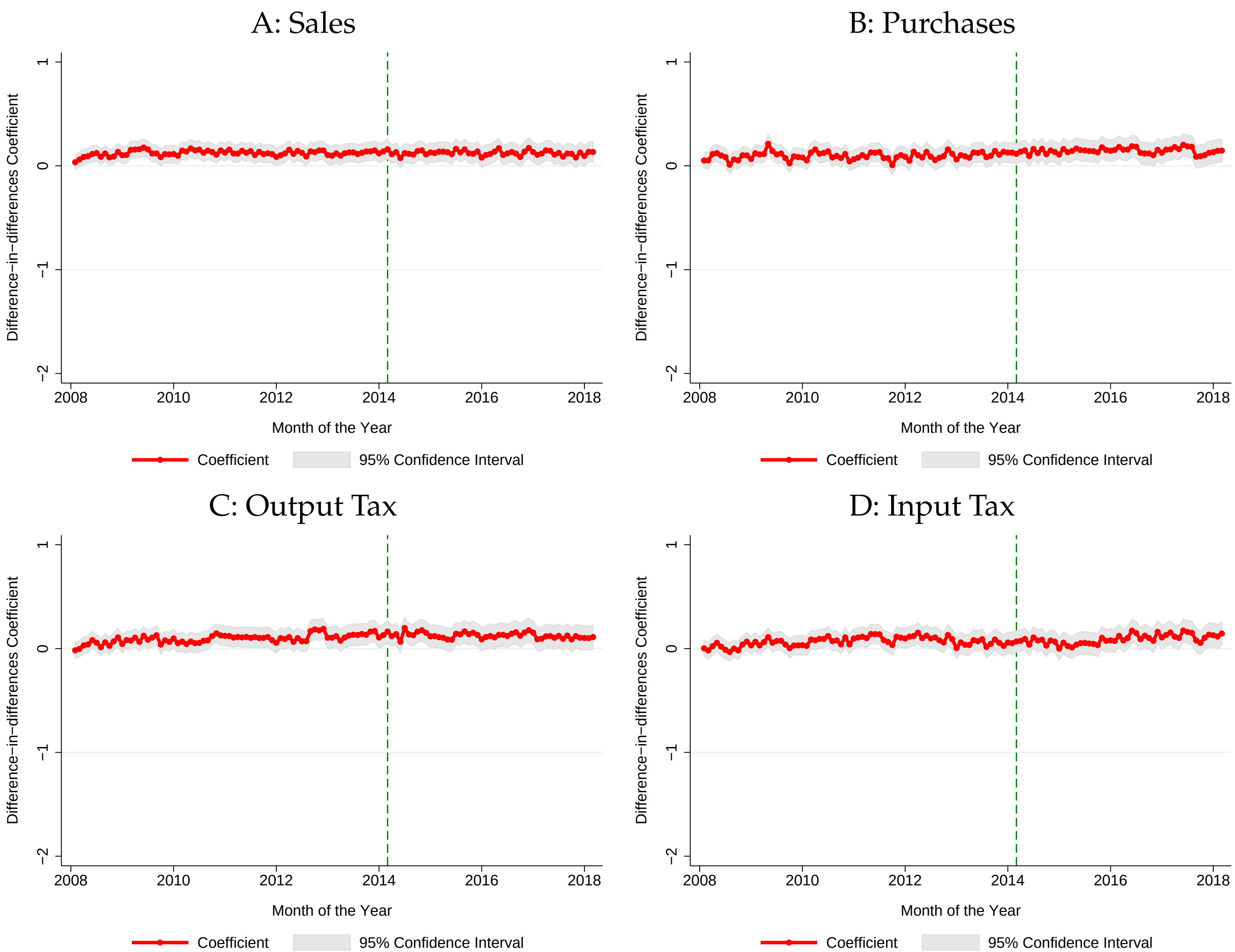
Notes: The figure shows the difference-in-differences version of the plots in Figure III. To construct these charts, we regress the log of the outcome variable shown in the title of each panel on the full set of firm, month, and month×treat dummies, dropping the dummies for July 2008. We then plot the coefficients on the month×treat dummies from these regressions. The gray surface plot shows the 95% confidence interval around the coefficient. The treatment groups consists of firms whose audit was assigned through the first random ballot held on September 25, 2014. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. We cluster standard errors at the firm level. Year t on the horizontal axis indicates July of the corresponding year. Vertical dashed lines demarcate the date the random computer ballot was held on.

FIGURE V: **AUDITED VS. UNAUDITED FIRMS – FIRST AUDIT WAVE**



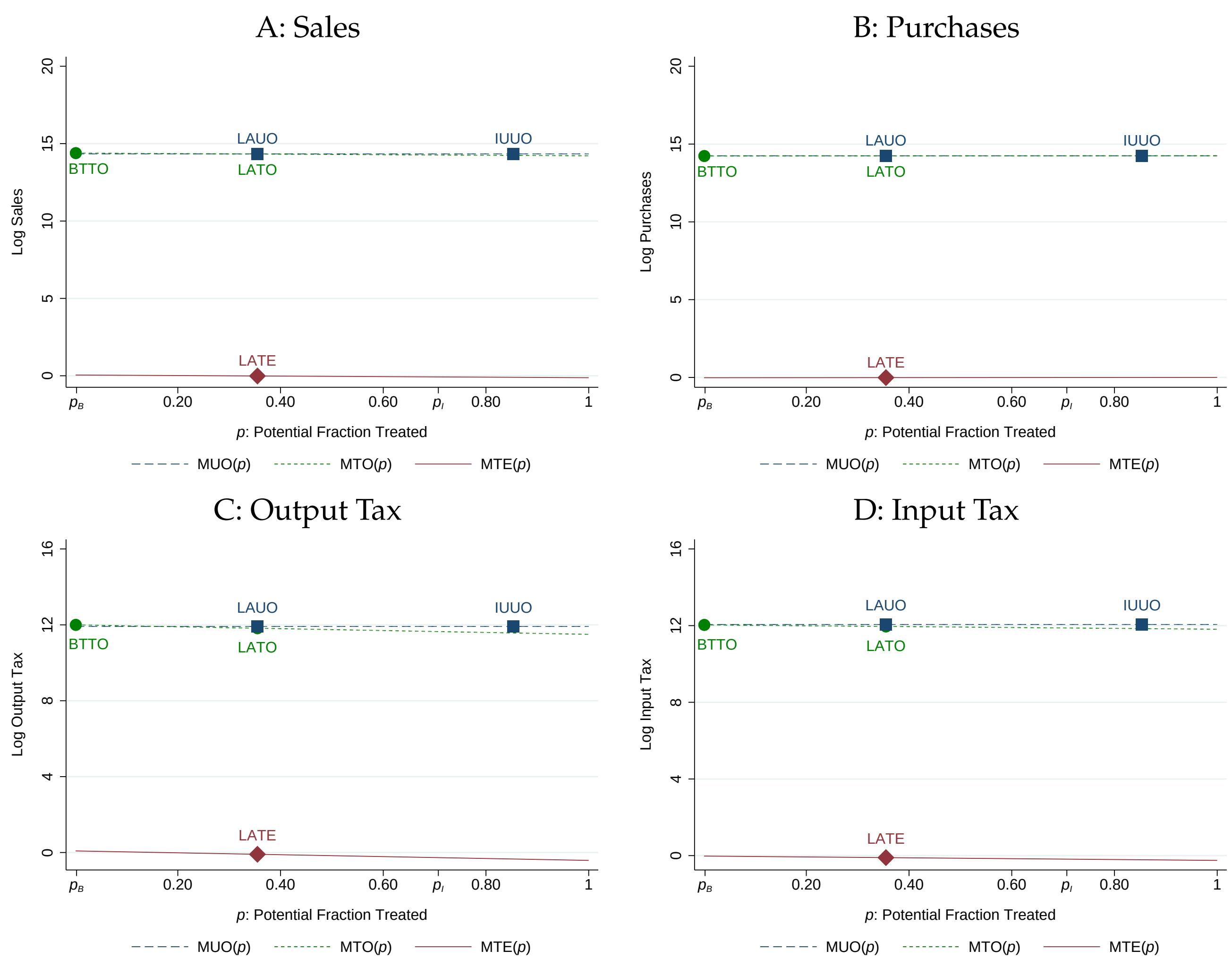
Notes: The figure compares the evolution of outcomes across audited and unaudited firms. To construct these charts, we regress the log of the outcome variable shown in the title of each panel on the full set of firm, month, and month×audit dummies, dropping the dummies for July 2008. We then plot the coefficients on the month×audit dummies from these regressions. The gray surface plot shows the 95% confidence interval around the coefficient. The audit dummy indicates firms whose audit was conducted during the first wave. These includes firms whose audit was assigned through the random computer ballot ($Z_i = 1$) and firms whose audit was initiated by the local tax office on their own accord ($Z_i = 0$). The unaudited firms are all other firms in the population of VAT filers. We cluster standard errors at the firm level. Year t on the horizontal axis indicates July of the corresponding year. Vertical dashed lines denotes September 13, 2013—the date first random computer ballot was held on.

FIGURE VI: **AUDITED VS. UNAUDITED FIRMS – SECOND AUDIT WAVE**



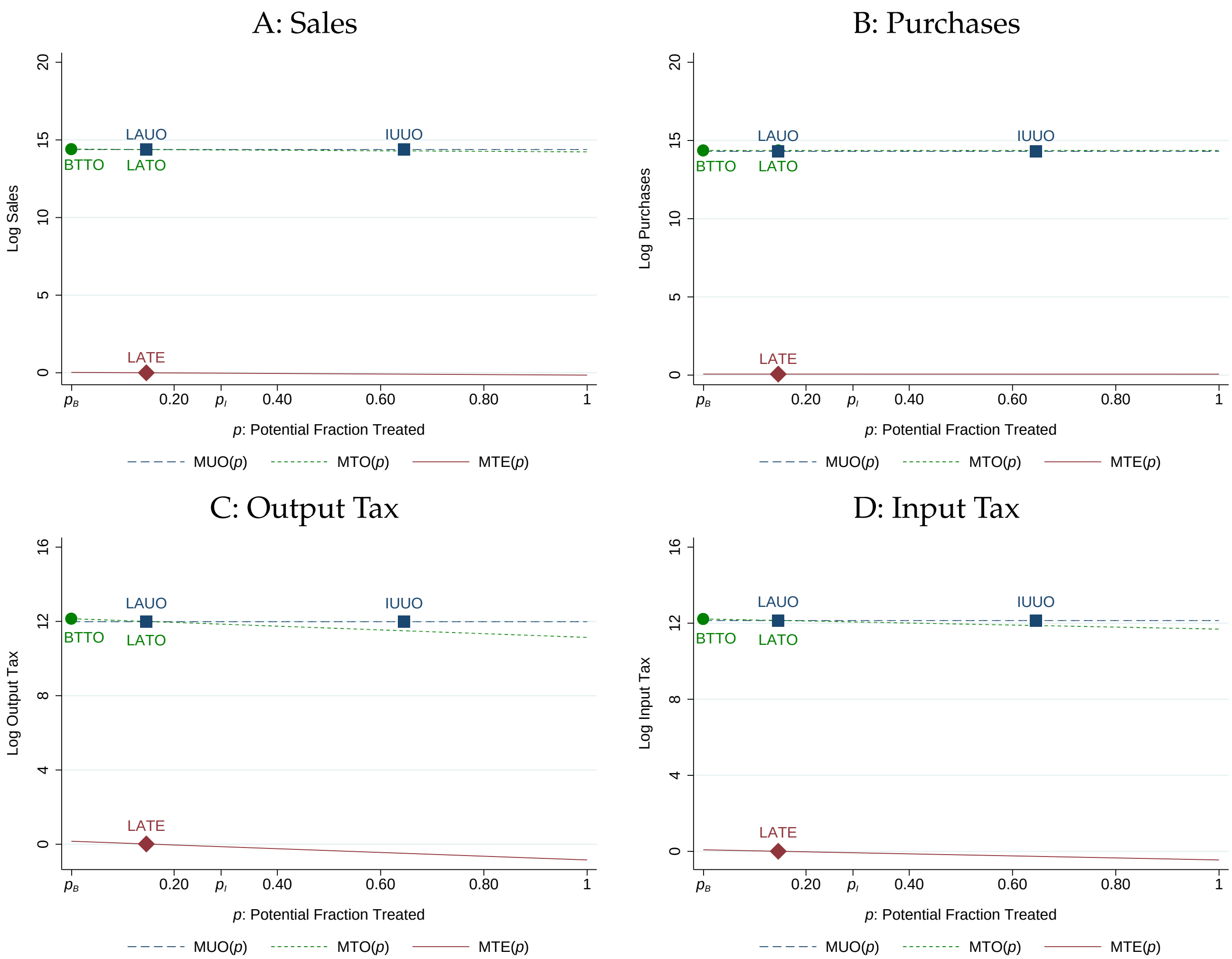
Notes: The figure compares the evolution of outcomes across audited and unaudited firms. To construct these charts, we regress the log of the outcome variable shown in the title of each panel on the full set of firm, month, and month×audit dummies, dropping the dummies for July 2008. We then plot the coefficients on the month×audit dummies from these regressions. The gray surface plot shows the 95% confidence interval around the coefficient. The audit dummy indicates firms whose audit was conducted during the second wave. These includes firms whose audit was assigned through the random computer ballot ($Z_i = 1$) and firms whose audit was initiated by the local tax office on their own accord ($Z_i = 0$). The unaudited firms are all other firms in the population of VAT filers. We cluster standard errors at the firm level. Year t on the horizontal axis indicates July of the corresponding year. Vertical dashed lines denotes September 25, 2014—the date first random computer ballot was held on.

FIGURE VII: **MARGINAL TREATMENT EFFECTS – FIRST AUDIT WAVE**



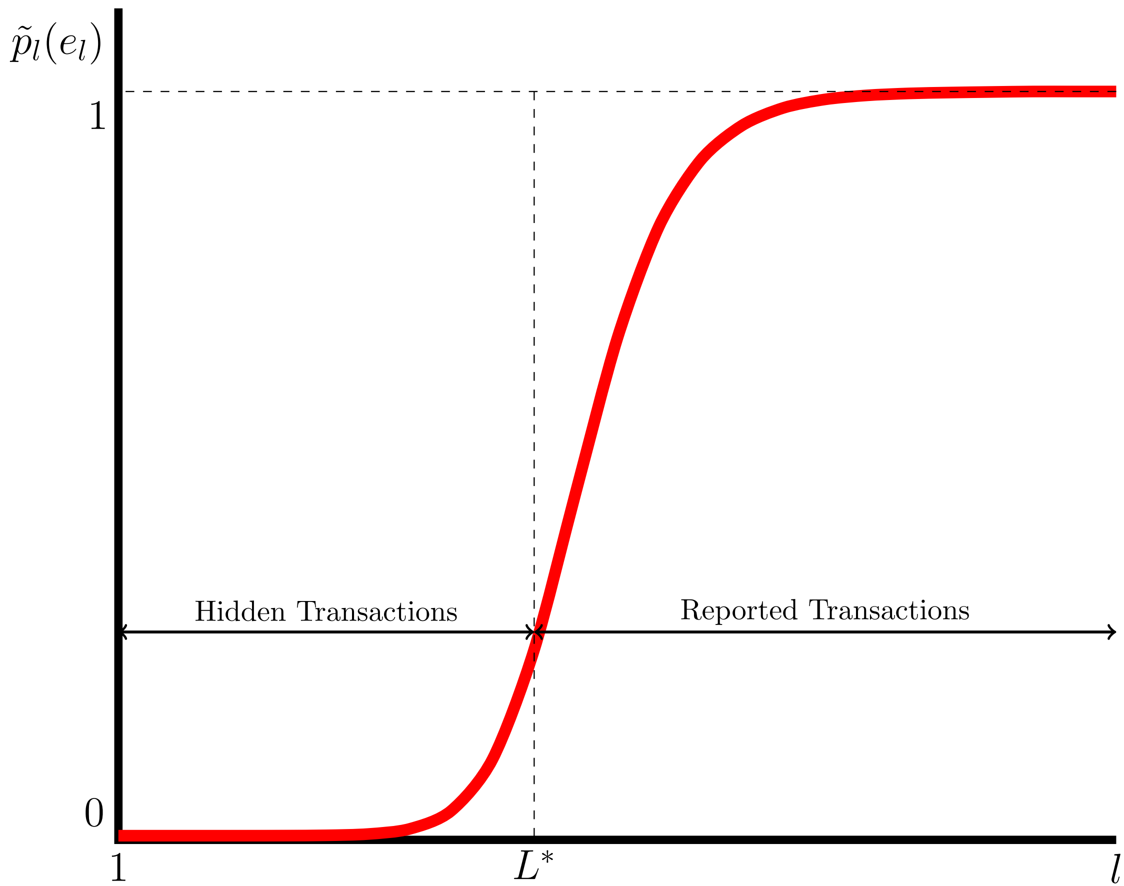
Notes: The figure plots the $MTE(p)$ curve for four outcomes using random assignment in the first audit wave as instrument. Please see Appendix A.2 for technical details. The fraction treated $p \equiv P(D = 1|Z)$ is shown along the horizontal axis. It increases from 0 (no treatment) to 1 (full treatment). We also indicate the baseline treatment probability $p_B \equiv P(D = 1|Z = 0)$ and the intervention treatment probability $p_I \equiv P(D = 1|Z = 1)$ along this axis. The green solid curve shows the marginal treated outcomes curve $MTO(p)$. It is identified at two points indicated in the plot by circular markers. The blue, dashed curve depicts the marginal untreated outcomes curve $MUO(p)$. It is also identified at two points indicated in the plot with square markers. For both curves, we extrapolate between the two points using linearity assumption. The difference between the two curves represents the $MTE(p)$. Since in our setting all three curves sit above each other, we lift both $MTO(p)$ and $MUO(p)$ up by adding the constant from the corresponding regression to distinguish them from the primary object of our interest $MTE(p)$.

FIGURE VIII: **MARGINAL TREATMENT EFFECTS – SECOND AUDIT WAVE**



Notes: The figure plots the $MTE(p)$ curve for four outcomes using random assignment in the second audit wave as instrument. Please see Appendix A.2 for technical details. The fraction treated $p \equiv P(D = 1|Z)$ is shown along the horizontal axis. It increases from 0 (no treatment) to 1 (full treatment). We also indicate the baseline treatment probability $p_B \equiv P(D = 1|Z = 0)$ and the intervention treatment probability $p_I \equiv P(D = 1|Z = 1)$ along this axis. The green solid curve shows the marginal treated outcomes curve $MTO(p)$. It is identified at two points indicated in the plot by circular markers. The blue, dashed curve depicts the marginal untreated outcomes curve $MUO(p)$. It is also identified at two points indicated in the plot with square markers. For both curves, we extrapolate between the two points using linearity assumption. The difference between the two curves represents the $MTE(p)$. Since in our setting all three curves sit above each other, we lift both $MTO(p)$ and $MUO(p)$ up by adding the constant from the corresponding regression to distinguish them from the primary object of our interest $MTE(p)$.

FIGURE IX: **PROBABILITY OF DETECTION**



Notes: The figure plots the probability of detection faced by a typical firm. We arrange L transactions carried out by the firm in term of the detection probability they entail $p_l(e_l)$ in ascending order. The probability of transaction is low if the other party to the transaction is (1) a consumer, (2) an unregistered firm, or (3) a firm willing to collude. In all these case, the transaction does not create any third-party information for the government. The probability of detection is high otherwise. The curve accordingly turns sharply once transactions between arm-length parties unwilling to collude begin. The transaction L^*represents the first transaction for which the detection probability is so high that inequality (8) fails. The firm would accordingly report transactions $[L^*, L]$, hiding the rest. Note that the threshold L^*would vary across firms depending among other things on their size, industry, and trading network.

TABLE I: **DESCRIPTIVE STATISTICS OF AUDIT I**

Audit	Tax	Ballot	Audits Assigned			Audits Conducted	
Wave	Year	Date	Mode	Number	Percent	Assigned	Unassigned
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	2013	Sep 13, 2013	Random	4,926	5%	3,482	521
2	2014	Sep 25, 2014	Random	12,447	12%	3,612	293
3	2015	Sep 14, 2015	Random	8,372	7.5%	1,122	164
4	2016	Jan 05, 2017	Parametric	8,935	7.5%	884	332
5	2017	Dec 04, 2018	Parametric	8,785	7.5%	852	352

Notes*:* The table reports some descriptive statistics of the five audit waves in our sample. Column (2) reports the tax year during which the computer ballot to draw audit cases was held. Column (3) reports the exact ballot date. The ballot was random for the first three waves and parametric for the next two. The volume of cases picked during the ballot is mentioned in Column (5) in numbers and in Column (6) as the proportion of population. Column (7) reports the number of audits completed out of those assigned through the computer ballot. Column (8), on the other hand, reports the number of audits initiated by the local tax office on their own accord. During the five audit waves, a total of 43,625 cases were picked for audit through computer ballots. Out of these, the tax identifiers of 218 were inaccurate. We were therefore unable to merge these 218 cases with VAT and audit records. We accordingly drop these 218 cases from the sample and focus instead on the 43,465 audits assigned through the computer ballot as reported in Column (5).

TABLE II: **DESCRIPTIVE STATISTICS OF AUDIT II**

Audit Wave	Audits Initiated			Amount Detected		
	Within 1 Month	Within 3 Months	Within 6 Months	Mean	Median	90th Percentile
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	0.646	0.942	0.950	617	0	165
2	0.925	0.993	0.998	619	0	100
3	0.852	0.945	0.964	4,098	0	158

Notes*:* The table presents a few descriptive statistics of randomly assigned audits during the first three audit waves. Columns (2)-(4) report the time lag between the assignment and initiation of audit. Column (2), for example, shows that around 65% of audits assigned in the first random ballot were initiated with the first month of assignment. This ratio was 93% and 85% for the next two audit waves. Columns (5)-(7) report the amount detected during each wave of audit. Column (5) reports the mean amount detected in PKR thousands. The US$-PKR exchange rate during this time (2013) was around 100. The next columns of the table report the median and the 90th percentile of the amount detected, illustrating that it is highly skewed toward right with the mean significantly larger than the median for all three audit waves.

TABLE III: **RANDOMIZATION TEST**

	First Wave				Second Wave				Third Wave			
	Mean ($Z_i = 0$)	Mean ($Z_i = 1$)	Diff. in Means	SE	Mean ($Z_i = 0$)	Mean ($Z_i = 1$)	Diff. in Means	SE	Mean ($Z_i = 0$)	Mean ($Z_i = 1$)	Diff. in Means	SE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A: VAT Outcomes												
1. Sales	14.251	14.282	0.031	0.043	14.278	14.298	0.020	0.026	14.335	14.831	0.496	0.029
2. Purchases	14.081	14.095	0.014	0.047	14.234	14.186	-0.048	0.029	14.264	14.248	-0.015	0.035
3. Output Tax	11.671	11.707	0.036	0.049	11.791	11.768	-0.024	0.030	11.969	11.953	-0.017	0.035
4. Input Tax	11.768	11.802	0.033	0.052	11.990	11.911	-0.079	0.031	12.149	11.886	-0.263	0.037
5. Tax Payable	10.200	10.300	0.100	0.063	10.392	10.360	-0.032	0.041	10.570	10.830	0.260	0.045
6. Tax Paid	9.532	9.607	0.076	0.058	9.805	9.785	-0.020	0.034	9.850	10.338	0.488	0.039
7. Exports	15.288	15.169	-0.119	0.114	14.904	15.145	0.241	0.068	14.619	15.655	1.036	0.064
8. Imports	14.905	14.887	-0.018	0.078	14.858	14.843	-0.015	0.048	14.878	15.902	1.024	0.076
9. Refund	12.037	11.884	-0.153	0.152	12.214	12.188	-0.026	0.089	12.086	12.424	0.338	0.093
10. Carry Forward	11.642	11.667	0.026	0.078	12.010	12.160	0.150	0.046	12.162	12.248	0.086	0.050
B: Firm Characteristics												
11. Manufacturer	0.339	0.350	0.010	0.010	0.314	0.339	0.025	0.006	0.215	0.786	0.572	0.006
12. Importer	0.111	0.109	-0.003	0.006	0.124	0.118	-0.006	0.004	0.159	0.019	-0.140	0.002
13. Exporter	0.025	0.019	-0.005	0.003	0.040	0.025	-0.016	0.002	0.050	0.021	-0.029	0.002
14. Distributor	0.028	0.030	0.001	0.003	0.031	0.034	0.003	0.002	0.036	0.011	-0.025	0.002
15. Wholesaler	0.240	0.241	0.001	0.008	0.229	0.240	0.011	0.005	0.251	0.046	-0.205	0.003
16. Service Provider	0.193	0.192	-0.002	0.008	0.193	0.185	-0.009	0.005	0.208	0.099	-0.110	0.005
17. Major City	0.640	0.636	-0.004	0.010	0.631	0.639	0.008	0.006	0.625	0.650	0.024	0.007
18. LTU	0.013	0.013	0.000	0.004	0.012	0.008	-0.004	0.002	0.005	0.042	0.037	0.003
19. Years Registered	12.987	13.680	0.694	0.109	11.745	12.967	1.222	0.070	10.496	13.607	3.111	0.091
20. Textile	0.162	0.163	0.001	0.008	0.143	0.152	0.009	0.005	0.108	0.266	0.157	0.006

Notes: The table runs balance tests on the three randomization waves in our sample. For each outcome, we estimate model (5) restricting the sample to the baseline period only. The baseline period is June 2012 for the first, June 2013 for the second, and June 2014 for the third randomization wave. The last two columns for each randomization wave report the coefficient $\hat{\beta}$ and its standard error from the model. The details of the variables used here are provided in Appendix A.1.

TABLE IV: **AUDIT FINDINGS**

	# Audits	Sales	Amount Detected		VAT Paid at the Baseline		Evasion Rate
			PKR	% of Sales	PKR	% of Sales	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<u>A: First Audit Wave</u>							
All Audited Firms	3,482	498.4	2.15	0.43	28.16	5.65	7.6
Amount Detected > 0	986	137.0	2.15	1.57	3.20	2.33	67.1
Size Quartile 1	1,057	0.1	0.06	684.76	0.00	8.78	7801.8
Size Quartile 2	824	1.7	0.07	3.94	0.04	2.52	156.2
Size Quartile 3	809	12.3	0.22	1.75	0.21	1.67	104.6
Size Quartile 4	792	484.3	1.80	0.37	27.91	5.76	6.5
<u>B: Second Audit Wave</u>							
All Audited Firms	3,612	2200.0	2.24	0.10	88.37	4.02	2.5
Amount Detected > 0	1,220	264.6	2.24	0.84	7.52	2.84	29.7
Size Quartile 1	1,007	0.4	0.04	10.21	0.02	3.81	268.1
Size Quartile 2	892	4.9	0.17	3.37	0.11	2.15	156.4
Size Quartile 3	862	24.4	0.22	0.89	0.30	1.24	71.8
Size Quartile 4	851	2170.2	1.81	0.08	87.95	4.05	2.1

Notes*:* The table presents descriptive statistics of audit outcomes. The first column reports the number of audits conducted for each group of firms indicated in the corresponding row. Aggregate turnover of this group for the baseline year in PKR billions is reported in the next column. The next two columns report the amount detected by audit, in PKR billions in column 3 and as a percent of aggregate sales in column 4. Columns 5-6 report the VAT paid at the baseline by the group of firms indicated in the corresponding row, in PKR billions in column 5 and as a percent of aggregate sales in column 6. The last column presents the evasion rate implied by the detected amount. It is calculated as the ratio of columns 4 and 6 (alternatively columns 3 and 5).

TABLE V: **IMPACT OF AUDIT ON FIRM BEHAVIOR – FIRST WAVE**

	Impacts After One Year					Impacts After Three Years				
	Sales	Purchases	Output Tax	Input Tax	Tax Payable	Sales	Purchases	Output Tax	Input Tax	Tax Payable
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A: ITT Estimates										
assign × after	-0.010	-0.010	-0.016	-0.017	-0.037	-0.007	-0.021	-0.025	-0.036	-0.016
	(0.016)	(0.018)	(0.021)	(0.023)	(0.028)	(0.017)	(0.019)	(0.022)	(0.023)	(0.028)
Observations	2,831,140	2,468,502	2,086,889	2,099,210	1,415,795	3,839,502	3,328,628	2,884,225	2,906,045	1,913,096
B: LATE Estimates										
audit × after	-0.014	-0.014	-0.023	-0.024	-0.051	-0.010	-0.030	-0.035	-0.051	-0.022
	(0.023)	(0.027)	(0.030)	(0.033)	(0.039)	(0.024)	(0.028)	(0.031)	(0.033)	(0.039)
Observations	2,831,140	2,468,502	2,086,889	2,099,210	1,415,795	3,839,502	3,328,628	2,884,225	2,906,045	1,913,096
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table estimates the impact of audit on firms' future behavior. In the top panel, the coefficient assign × after shows $\hat{\gamma}$ from model (6), where the dummy variable assign$_i$ denotes that firm i's audit was assigned through the first random ballot held on September 13, 2013. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The dummy variable after$_t$ indicates that month t falls after the date of the ballot. The sample includes periods up to October 2014 for the first five columns and periods up to October 2016 for the rest. Panel B shows the corresponding results from 2sls regressions, where the endogenous variable audit$_i$ is instrumented by the initial random assignment. Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE VI: **IMPACT OF AUDIT ON FIRM BEHAVIOR – SECOND WAVE**

	Impacts After One Year					Impacts After Three Years				
	Sales	Purchases	Output Tax	Input Tax	Tax Payable	Sales	Purchases	Output Tax	Input Tax	Tax Payable
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A: ITT Estimates										
assign × after	-0.021	-0.021	-0.030	-0.026	-0.022	-0.010	-0.009	-0.013	-0.007	0.006
	(0.010)	(0.012)	(0.012)	(0.013)	(0.016)	(0.010)	(0.012)	(0.013)	(0.013)	(0.016)
Observations	3,133,061	2,725,243	2,343,583	2,357,343	1,568,363	4,159,404	3,587,740	3,088,403	3,137,794	2,034,932
B: LATE Estimates										
audit × after	-0.071	-0.073	-0.109	-0.091	-0.081	-0.032	-0.033	-0.044	-0.025	0.022
	(0.033)	(0.043)	(0.043)	(0.046)	(0.058)	(0.035)	(0.043)	(0.044)	(0.045)	(0.057)
Observations	3,133,061	2,725,243	2,343,583	2,357,343	1,568,363	4,159,404	3,587,740	3,088,403	3,137,794	2,034,932
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table estimates the impact of audit on firms' future behavior. In the top panel, the coefficient assign × after shows $\hat{\gamma}$ from model (6), where the dummy variable assign$_i$ denotes that firm i's audit was assigned through the first random ballot held on September 25, 2014. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The dummy variable after$_t$ indicates that month t falls after the date of the ballot. The sample includes periods up to October 2015 for the first five columns and periods up to October 2017 for the rest. Panel B shows the corresponding results from 2sls regressions, where the endogenous variable audit$_i$ is instrumented by the initial random assignment. Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE VII: **IMPACTS OF RANDOM AUDITS ASSIGNED IN THE FIRST WAVE – OTHER OUTCOMES**

	Impacts After One Year					Impacts After Three Years				
	Exports	Imports	Tax Paid	Refund	Carry Forward	Exports	Imports	Tax Paid	Refund	Carry Forward
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A: ITT Estimates										
assign × after	0.013	0.047	-0.052	-0.116	-0.049	0.027	0.035	-0.025	-0.070	-0.085
	(0.037)	(0.028)	(0.031)	(0.092)	(0.040)	(0.038)	(0.027)	(0.033)	(0.091)	(0.040)
Observations	317,130	570,949	1,161,513	234,207	1,594,740	450,661	838,590	1,723,448	287,241	2,490,894
B: LATE Estimates										
audit × after	0.018	0.073	-0.072	-0.175	-0.071	0.037	0.054	-0.035	-0.102	-0.124
	(0.051)	(0.043)	(0.043)	(0.138)	(0.058)	(0.053)	(0.042)	(0.046)	(0.134)	(0.059)
Observations	317,130	570,949	1,161,513	234,207	1,594,740	450,661	838,590	1,723,448	287,241	2,490,894
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table estimates the impact of audit on firms' future behavior. In the top panel, the coefficient assign × after shows $\hat{\gamma}$ from model (6), where the dummy variable assign$_i$ denotes that firm i's audit was assigned through the first random ballot held on September 13, 2013. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The dummy variable after$_t$ indicates that month t falls after the date of the ballot. The sample includes periods up to October 2014 for the first five columns and periods up to October 2016 for the rest. Panel B shows the corresponding results from 2sls regressions, where the endogenous variable audit$_i$ is instrumented by the initial random assignment. Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE VIII: **EXTENSIVE MARGIN IMPACT OF RANDOM AUDITS**

Outcome:	$1(Return\ Filed_{it})$					
Random Draw Held On:	September 13, 2013		September 25, 2014		September 14, 2015	
Impacts After:	One Year	Three Years	One Year	Three Years	One Year	Three Years
	(1)	(2)	(3)	(4)	(5)	(6)
<u>A: ITT Estimates</u>						
assign × after	0.002	0.004	0.008	0.009	0.010	0.008
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	7,097,120	9,852,941	8,129,498	11,062,795	8,502,891	11,171,180
<u>B: LATE Estimates</u>						
audit × after	0.002	0.006	0.027	0.029	0.075	0.058
	(0.002)	(0.003)	(0.004)	(0.004)	(0.010)	(0.009)
Observations	7,097,120	9,852,941	8,129,498	11,062,795	8,502,891	11,171,180
Mean of the Dependent Variable	0.955	0.955	0.956	0.956	0.956	0.956
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes
Period FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes*:* The table estimates the impact of audit on firms' extensive margin behavior. We estimate model (6) using an indicator that the firm filed its VAT return for the period (month) t as the outcome variable. In the top panel, the coefficient assign × after shows $\hat{\gamma}$ from the model. The dummy variable assign$_i$ denotes that firm i's audit was assigned through the random ballot held on the date indicated in the heading of each column. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The dummy variable after$_t$ indicates that month t falls after the date of the ballot. The sample for odd columns includes periods up to one year after the ballot and for even columns up to three years after the ballot. Panel B shows the corresponding results from 2sls regressions, where the endogenous variable audit$_i$ is instrumented by the initial random assignment. Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE IX: **SELECTION IN COMPLIANCE? AUDITED VS. NON-AUDITED FIRMS**

	First Wave				Second Wave			
	Mean ($D_i = 0$)	Mean ($D_i = 1$)	Difference in Means	Standard Error	Mean ($D_i = 0$)	Mean ($D_i = 1$)	Difference in Means	Standard Error
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A: VAT Outcomes								
1. Sales	14.547	14.816	0.269	0.044	14.553	14.776	0.222	0.043
2. Purchases	14.311	14.567	0.255	0.048	14.438	14.519	0.080	0.049
3. Output Tax	11.936	12.214	0.279	0.050	12.031	12.199	0.168	0.051
4. Input Tax	12.006	12.328	0.322	0.051	12.196	12.230	0.033	0.051
5. Tax Payable	10.537	10.866	0.328	0.068	10.698	10.870	0.172	0.075
6. Tax Paid	10.039	10.435	0.397	0.062	10.221	10.359	0.137	0.063
7. Exports	15.752	15.705	-0.047	0.114	15.353	15.793	0.440	0.096
8. Imports	15.183	15.261	0.078	0.075	15.096	15.235	0.139	0.074
9. Refund	12.410	12.673	0.263	0.139	12.578	12.667	0.089	0.130
10. Carry Forward	11.926	12.192	0.266	0.081	12.276	12.446	0.170	0.083
B: Firm Characteristics								
11. Manufacturer	0.383	0.448	0.064	0.010	0.361	0.418	0.056	0.009
12. Importer	0.105	0.087	-0.018	0.006	0.116	0.111	-0.005	0.006
13. Exporter	0.023	0.016	-0.007	0.003	0.036	0.013	-0.023	0.003
14. Distributor	0.027	0.026	-0.001	0.004	0.029	0.028	-0.001	0.004
15. Wholesaler	0.214	0.196	-0.018	0.008	0.206	0.219	0.012	0.008
16. Service Provider	0.190	0.174	-0.016	0.008	0.189	0.166	-0.023	0.008
17. Major City	0.661	0.661	0.000	0.000	0.654	0.654	0.000	0.000
18. LTU	0.045	0.045	-0.000	0.000	0.039	0.039	-0.000	0.000
19. Years Registered	13.499	14.729	1.230	0.117	12.388	14.221	1.833	0.119
20. Textile	0.165	0.171	0.005	0.007	0.148	0.160	0.012	0.006

Notes: The table explores selection in audit, comparing audited and unaudited firms. We estimate a version of model (5), regressing the outcome in each row on two dummy variables (D_i and $corporate_i$) and tax office fixed effects. We restrict the sample to the baseline period only. The dummy variable D_i takes the value 1 for all audited firms including those whose audit was assigned through the random ballot and those whose audit was taken up by the local tax office of its own accord. The unaudited firms ($D_i = 0$) include all other firms in the eligible sample. The baseline period is June 2012 for the first and June 2013 for the second audit wave. The last two columns for each wave report the coefficient $\hat{\beta}$ and its standard error from the model. The details of the variables used here are provided in Appendix A.1.

TABLE X: **SELECTION IN COMPLIANCE? AUDITED VS. NON-AUDITED FIRMS (WITHIN $Z_i = 1$ GROUP)**

	2013 Draw				2014 Draw			
	Mean ($D_i = 0$)	Mean ($D_i = 1$)	Difference in Means	Standard Error	Mean ($D_i = 0$)	Mean ($D_i = 1$)	Difference in Means	Standard Error
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A: VAT Outcomes								
1. Sales	14.567	14.569	0.001	0.095	14.560	14.633	0.073	0.061
2. Purchases	14.360	14.312	-0.048	0.108	14.393	14.410	0.017	0.066
3. Output Tax	11.885	12.005	0.120	0.104	11.982	12.094	0.112	0.069
4. Input Tax	11.944	12.075	0.131	0.117	12.131	12.115	-0.017	0.070
5. Tax Payable	10.524	10.666	0.142	0.132	10.648	10.715	0.067	0.101
6. Tax Paid	9.935	10.175	0.240	0.129	10.206	10.173	-0.033	0.083
7. Exports	15.602	15.678	0.076	0.285	15.476	15.897	0.422	0.226
8. Imports	15.131	15.150	0.018	0.178	15.057	15.154	0.097	0.101
9. Refund	11.650	12.482	0.832	0.385	12.502	12.681	0.179	0.257
10. Carry Forward	11.833	12.023	0.190	0.173	12.424	12.331	-0.093	0.108
B: Firm Characteristics								
11. Manufacturer	0.364	0.406	0.042	0.022	0.378	0.397	0.019	0.013
12. Importer	0.115	0.096	-0.019	0.016	0.107	0.120	0.013	0.010
13. Exporter	0.018	0.017	-0.001	0.006	0.022	0.022	0.000	0.004
14. Distributor	0.030	0.027	-0.003	0.008	0.033	0.029	-0.003	0.005
15. Wholesaler	0.228	0.210	-0.018	0.020	0.218	0.215	-0.003	0.012
16. Service Provider	0.186	0.188	0.001	0.017	0.185	0.170	-0.015	0.011
17. Major City	0.655	0.655	-0.000	0.000	0.659	0.659	-0.000	0.000
18. LTU	0.043	0.043	0.000	0.000	0.035	0.035	0.000	0.000
19. Years Registered	13.865	14.313	0.448	0.258	13.175	14.222	1.047	0.167
20. Textile	0.163	0.167	0.005	0.017	0.158	0.154	-0.004	0.009

Notes: The table explores selection in audit, comparing audited and unaudited firms within the sample drawn for audit in the corresponding random ballot. We estimate a version of model (5), regressing the outcome in each row on two dummy variables (D_i and $corporate_i$) and tax office fixed effects. We restrict the sample to the baseline period only. The dummy variable D_i takes the value 1 for firms whose audit was conducted. The unaudited firms ($D_i = 0$) include all other firms in the randomly drawn sample. The baseline period is June 2012 for the first and June 2013 for the second audit wave. The last two columns for each wave report the coefficient $\hat{\beta}$ and its standard error from the model. The details of the variables used here are provided in Appendix A.1.

TABLE XI: **SELECTION IN COMPLIANCE? AUDITED VS. NON-AUDITED FIRMS (WITHIN $Z_i = 0$ GROUP)**

	2013 Draw				2014 Draw			
	Mean ($D = 0$)	Mean ($D = 1$)	Difference in Means	Standard Error	Mean ($D = 0$)	Mean ($D = 1$)	Difference in Means	Standard Error
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<u>A: VAT Outcomes</u>								
1. Sales	14.548	15.693	1.145	0.086	14.556	15.776	1.220	0.149
2. Purchases	14.312	15.330	1.018	0.095	14.446	15.549	1.103	0.169
3. Output Tax	11.936	12.958	1.022	0.098	12.040	13.046	1.006	0.170
4. Input Tax	12.008	13.045	1.037	0.097	12.208	13.053	0.844	0.171
5. Tax Payable	10.537	11.704	1.167	0.148	10.708	11.979	1.270	0.235
6. Tax Paid	10.040	11.220	1.180	0.132	10.227	11.390	1.163	0.190
7. Exports	15.750	16.009	0.258	0.216	15.330	16.019	0.689	0.372
8. Imports	15.183	15.473	0.290	0.129	15.101	15.689	0.588	0.198
9. Refund	12.425	13.168	0.743	0.291	12.585	13.168	0.583	0.381
10. Carry Forward	11.927	12.857	0.930	0.164	12.255	13.138	0.883	0.279
<u>B: Firm Characteristics</u>								
11. Manufacturer	0.384	0.622	0.239	0.021	0.359	0.531	0.172	0.030
12. Importer	0.105	0.049	-0.055	0.010	0.117	0.110	-0.007	0.019
13. Exporter	0.023	0.017	-0.006	0.004	0.037	0.026	-0.011	0.002
14. Distributor	0.026	0.020	-0.006	0.007	0.028	0.017	-0.012	0.014
15. Wholesaler	0.214	0.133	-0.081	0.012	0.205	0.182	-0.023	0.022
16. Service Provider	0.190	0.111	-0.079	0.016	0.190	0.103	-0.087	0.025
17. Major City	0.661	0.661	0.000	0.000	0.652	0.652	-0.000	0.000
18. LTU	0.045	0.045	-0.000	0.000	0.040	0.040	0.000	0.000
19. Years Registered	13.493	16.379	2.886	0.288	12.266	14.597	2.331	0.437
20. Textile	0.165	0.187	0.021	0.017	0.147	0.208	0.061	0.025

Notes: The table explores selection in audit, comparing audited and unaudited firms excluding from the sample firms drawn for audit in the corresponding random ballot. We estimate a version of model (5), regressing the outcome in each row on two dummy variables (D_i and $corporate_i$) and tax office fixed effects. We restrict the sample to the baseline period only. The dummy variable D_i takes the value 1 for firms whose audit was conducted. The baseline period is June 2012 for the first and June 2013 for the second audit wave. The last two columns for each wave report the coefficient $\hat{\beta}$ and its standard error from the model. The details of the variables used here are provided in Appendix A.1.

A Online Appendix

A.1 Definition of Variables

(*i*) **Sales.** The value of all goods and services supplied by the firm in the given tax period (month) including exports.

(*ii*) **Purchases.** The value of all taxable intermediates acquired by the firm in the given tax period (month).

(*iii*) **Output Tax.** The value of VAT charged on sales made by the firm in the given tax period (month). It equals $\tau.\,(\hat{s}_{it} - \hat{s}_{E,it})$, where τ is the applicable VAT rate and $(\hat{s}_{it} - \hat{s}_{E,it})$ is the value of non-export sales reported by firm i in period t. Because exports are zero-rated, they do not appear in the output tax.

(*iv*) **Input Tax.** The value of VAT credit claimed on intermediates acquired by the firm in the given tax period (month). It equals $\tau.\hat{c}_{it}$, where τ is the applicable VAT rate and $\hat{c}_{it}$ is the value of purchases of intermediates claimed by firm i in period t.

(*v*) **Tax Payable.** The VAT payable by the firm in the given tax period (month). By definition, it equals the output tax minus the input tax.

(*vi*) **Tax Paid** The VAT actually paid by the firm in the given tax period (month). It may differ from Tax Payable if the firm has any carry-forward from previous months.

(*vii*) **Exports.** The value of all goods and services exported by the firm in the given tax period (month).

(*viii*) **Imports.** The value of all goods and services imported by the firm in the given tax period (month).

(*ix*) **Refund.** The amount of refund claimed by the firm in the given tax period (month). The refund arises when the firm's input tax exceeds its output tax. In this case, the firm has the option to carry forward the balance amount or seek its refund. Because exports are zero-rated, firms the majority of whose output is exported are likely to claim refund every tax period.

(*x*) **Carry Forward.** The amount of carry forward claimed by a firm. The carry forward arises when the firm's input tax exceeds its output tax and it does not opt to seek the refund of the balance amount.

(*xi*) **Manufacturer.** A firm whose principal business activity is the manufacture of goods. Manufacturing is the process whereby a firm converts inputs into a distinct article capable of being put to use differently than inputs and includes any process incidental or ancillary to it.

(*xii*) **Importer.** A firm whose principal business activity is the import of goods for sale in the local market without carrying out any manufacturing process on them.

(*xiii*) **Exporter.** A firm whose principal business activity is the export of goods. These firms may supply in the local market, but a majority of their output is exported out of country.

(*xiv*) **Distributor.** Distributor means a person appointed by a manufacturer, importer or any other person for a specified area to purchase goods from him for further supply and includes a person who in addition to being a distributor is also engaged in supply of goods as a wholesaler or a retailer.

(*xv*) **Wholesaler.** Wholesaler' includes a dealer and means any person who carries on, whether regularly or otherwise, the business of buying and selling goods by wholesale or of supplying or distributing goods, directly or indirectly, by wholesale for cash or deferred payment or for commission or other valuable consideration or stores such goods belonging to others as an agent for the purpose of sale; and includes a person supplying taxable goods to a person who deducts income tax at source under the Income Tax Ordinance, 2001.

(*xvi*) **Retailer.** A person, supplying goods to general public for the purpose of consumption.

(*xvii*) **Industry.** The Pakistani tax administration uses 4-digit Harmonized Commodity Description and Coding System (HS code) to classify firms into industry. The code, used by customs administrations throughout the world, divides all goods and services into 99 chapters (the first two digits in the code) and 21 sections. The sections broadly correspond to major industries in the country. I take the section a firm falls in as its industry.

(*xviii*) **Major City** The dummy variable takes the value 1 if the firm's head office is in one of the three major cities of Pakistan—Karachi, Lahore, and Islamabad.

(*xix*) **LTU** The dummy variable takes the value 1 if the firm is administered by on of the four Large Taxpayer Centers in the country located in Karachi, Lahore, and Islamabad.

A.2 Marginal Treatment Effects

In this section, we describe how we estimate the $MTE(p)$ curves shown in Figures VII and VIII. Because we have access to a binary instrument only, full nonparametric identification (see Heckman & Vytlacil, 2005, 2007) is not feasible in our setup, and instead we identify MTEs under a functional structure following the approach developed in Kowalski (2016) and Brinch *et al.* (2017).

As in the paper, Z here denotes the instrument (random assignment) and D the treatment (actual audit). Following the standard terminology in this literature, we refer to $p \equiv P(D = 1|Z)$ as the potential fraction treated. For any outcome Y, The $MTE(p)$ is defined as

$$MTE(p) \equiv \mathbb{E}(Y_T - Y_U|U_D = p)$$

where Y_T represents the potential outcome in the audited state $(D = 1)$ and Y_U the potential outcome in the unaudited state $(D = 0)$. The unobserved cost and benefit of audit are represented by U_D and p. The MTE therefore captures the treatment effect on a unit marginal to selecting into treatment. Using the above definition, it can be written as the difference between the marginal treated outcome (MTO) and the marginal untreated outcome (MUO)

$$\begin{aligned} MTO(p) \equiv & \quad \mathbb{E}(Y_T|U_D = p) \\ MUO(p) \equiv & \quad \mathbb{E}(Y_U|U_D = p) \end{aligned}$$

These curves are defined for every value of $p(Z)$ but given our binary instrument only two values of p are observed: the baseline treatment probability $p_B \equiv P(D = 1|Z = 0)$ and the intervention treatment probability $p_I \equiv P(D = 1|Z = 1)$. We therefore assume that both these curves are linear. The $MTO(p)$ is identified at two points

$$\begin{aligned} BTTO = & \quad \mathbb{E}(Y|X = x, D = 1, Z = 0) \\ LATO = & \quad \frac{1}{p_I - p_B} \left[p_I ITTO - p_B BTTO \right], \end{aligned}$$

where $ITTO = \mathbb{E}(Y|X = x, D = 1, Z = 1)$. We use the linearity assumption to extrapolate between these two points. Similarly, the $MUO(p)$ is identified at

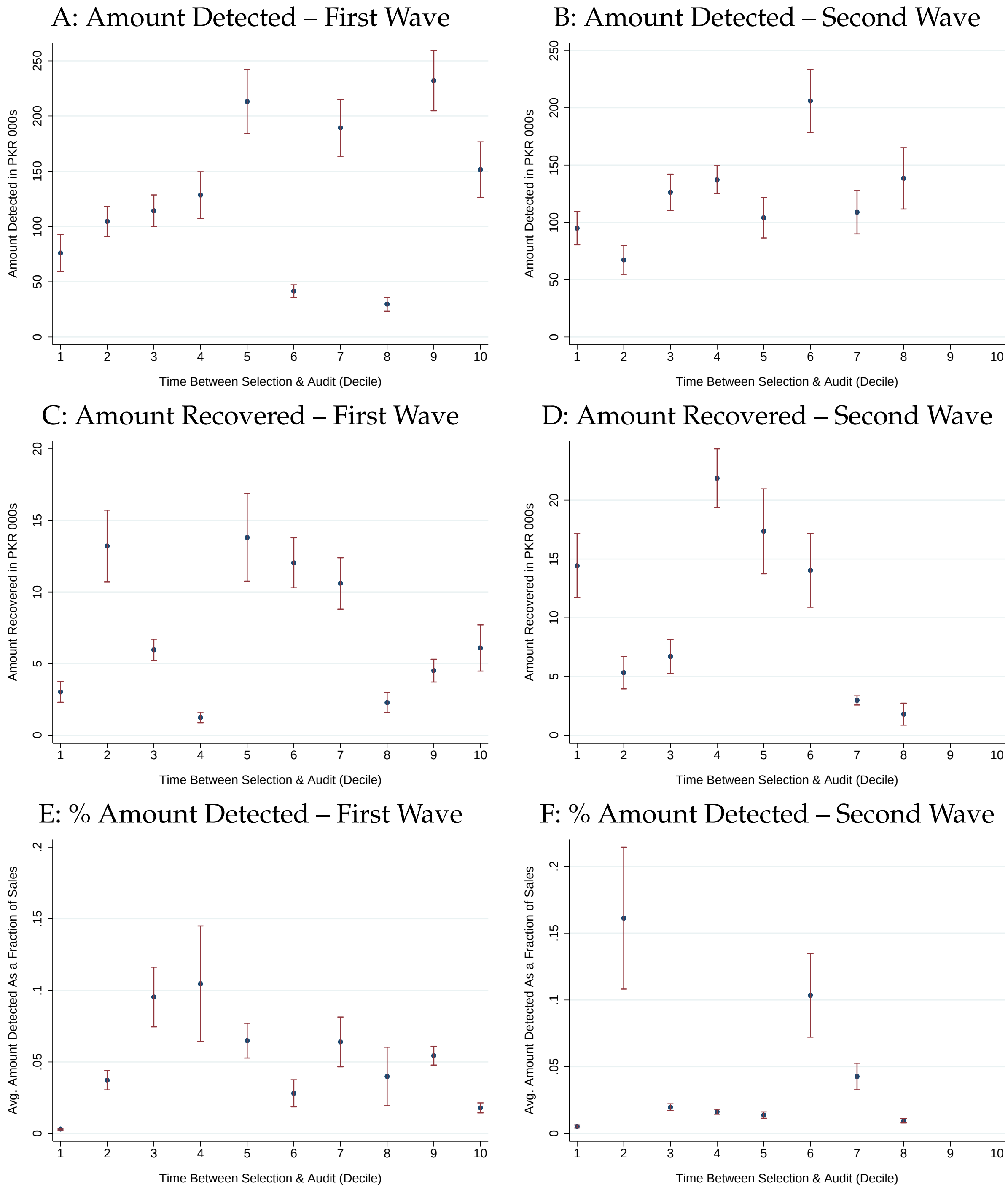
$$\begin{aligned} IUUO = & \qquad \mathbb{E}(Y|X = x, D = 0, Z = 1) \\ LAUO = & \quad \frac{1}{p_I - p_B} \left[(1 - p_B) BUUO - (1 - p_I) IUUO \right] \end{aligned}$$

where $BUUO = \mathbb{E}(Y|X = x, D = 0, Z = 0)$.[20]

To plot the $MTO(p)$ curve, we regress the outcome variable on a full set of firm and period fixed fixed effects and an interaction term of the audit (D) and post dummies, restricting the sample to firms randomly selected for audit ($Z = 1$). The regression gives us estimates of ITTO and IUUO. Running a similar regression on a sample of firms not drawn in the random ballot ($Z = 0$) delivers the estimates of BTTO and BUUO. We then find LATO and LAUO using the definitions above. The $MTO(p)$ curve is identified at two points $(BTTO, \frac{p_B}{2})$ and $(LATO, \frac{p_B + p_I}{2})$. We extrapolate between the two using the linearity assumption. Similarly, $MUO(p)$ is identified at $(LAUO, \frac{p_B + p_I}{2})$ and $(IUUO, \frac{p_I + 1}{2})$, and we extrapolate using linearity. The $MTO(p)$ curve is the difference between the two. We draw these curves for four outcomes and two audit waves separately. Since in our setting all these curves sit above each other, we lift both $MTO(p)$ and $MUO(p)$ up by adding the constant from the corresponding regression to distinguish them from the primary object of our interest $MTE(p)$.

[20] In all these definitions, O stands for outcomes, T for treated, U for untreated, B for baseline, I for intervention, and LA for local average. Please see Kowalski (2016) for detail of these terms.

FIGURE A.I: **AMOUNT DETECTED BY TIMING OF AUDIT**



Notes: The figure examines if the order in which audits were taken up is correlated with audit outcomes, exploring thereby if audits were systematically targeted toward specific firms. We divide the time between assignment and initiation of audit into ten deciles and then plot the average audit outcome and the 95% confidence interval around it for each decile. The top panels look at the average amount detected by audit in PKR thousands, the middle panels at the average amount recovered in PKR thousands, and the bottom panels at the average amount detected as a ratio of annual baseline turnover of the firm. To take care of outliers, we drop observations where the amount detected is more than the 99th percentile of the distribution. This affects the top and bottom panels only. The LHS panels plot outcomes for the first randomized ballot and the RHS for the second.

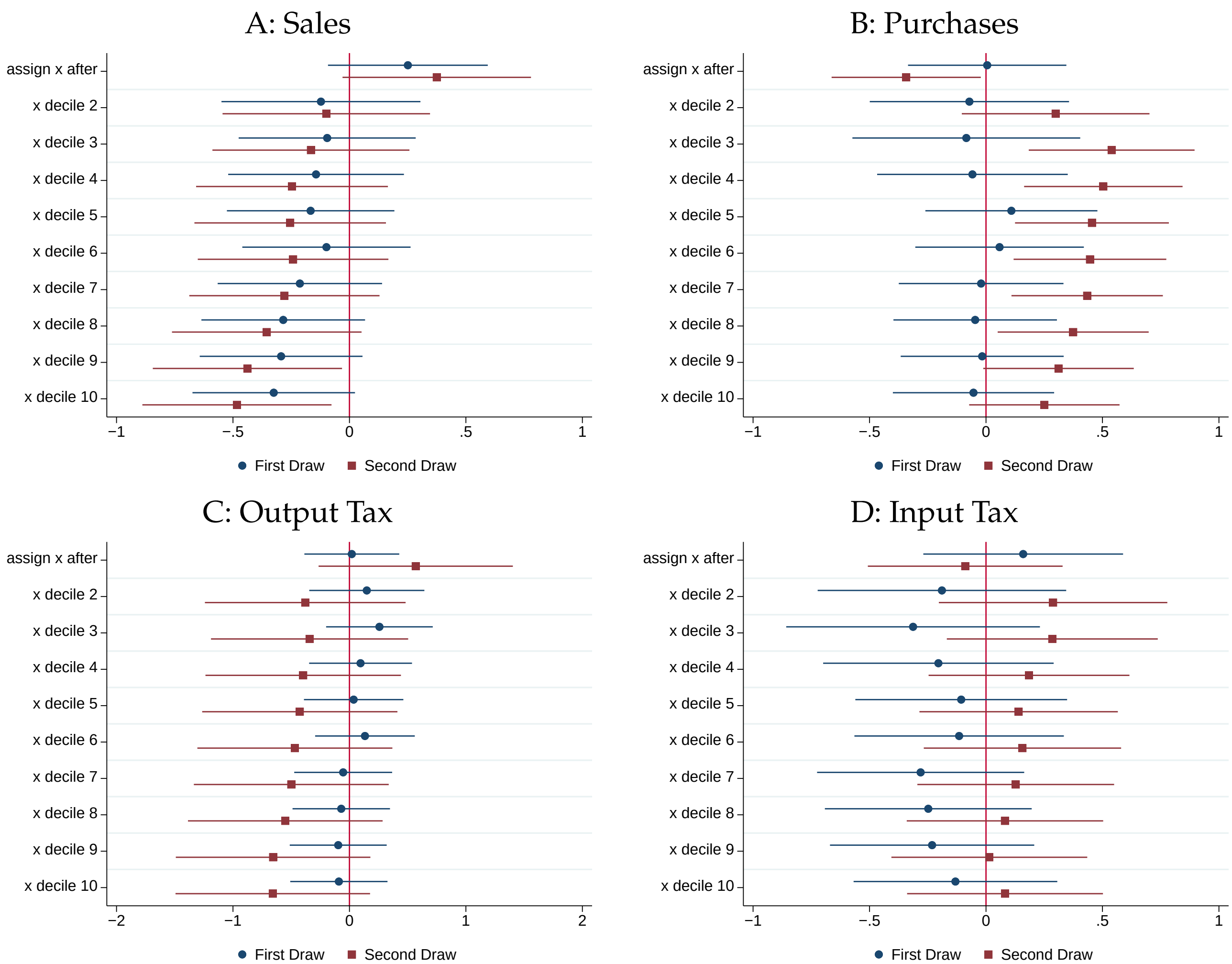
FIGURE A.II: **INTENTION TO TREAT EFFECTS OF THIRD AUDIT WAVE**

Notes: The figure explores the impacts of audit on future firm behavior. We compare the evolution of four VAT outcomes across the treatment and control groups. The treatment groups consists of firms whose audit was assigned through the first random ballot held on September 14, 2015. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments, firms already under audit, and firms subject to fixed and withholding tax regimes. We do not identify the last type of firms and therefore are unable to exclude them from the eligible sample. To construct these charts, we regress the log of the outcome variable shown in the title of each panel on the full set of firm and month fixed effects, dropping the dummy for July 2008. We then plot the coefficients on the time dummies of these regressions. The sample includes all tax periods from July 2008 to June 2018. The regressions are run separately for the two groups of firms. Year t on the horizontal axis indicates July of the corresponding year. Vertical dashed lines demarcate the date the random computer ballot was held on.

FIGURE A.III: **INTENTION TO TREAT EFFECTS OF THIRD AUDIT WAVE**

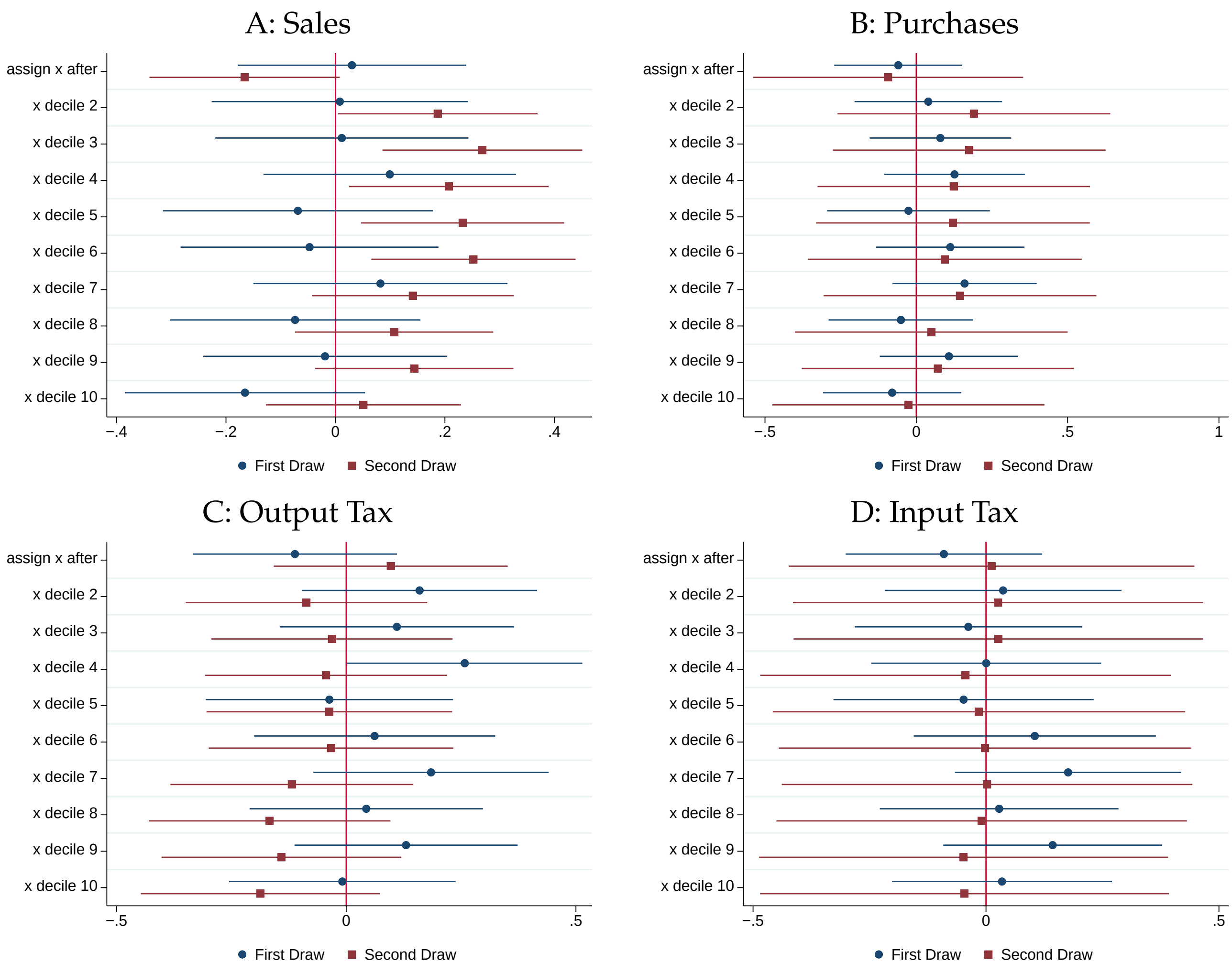
Notes: The figure shows the difference-in-differences version of the plots in Figure A.II. To construct these charts, we regress the log of the outcome variable shown in the title of each panel on the full set of firm, month, and month$\times$treat dummies, dropping the dummies for July 2008. We then plot the coefficients on the month$\times$treat dummies from these regressions. The gray surface plot shows the 95% confidence interval around the coefficient. The treatment groups consists of firms whose audit was assigned through the first random ballot held on September 14, 2015. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments, firms already under audit, and firms subject to fixed and withholding tax regimes. We do not identify the last type of firms and therefore are unable to exclude them from the eligible sample. We cluster standard errors at the firm level. Year t on the horizontal axis indicates July of the corresponding year. Vertical dashed lines demarcate the date the random computer ballot was held on.

FIGURE A.IV: **HETEROGENEITY IN RESPONSE BY FIRM SIZE**



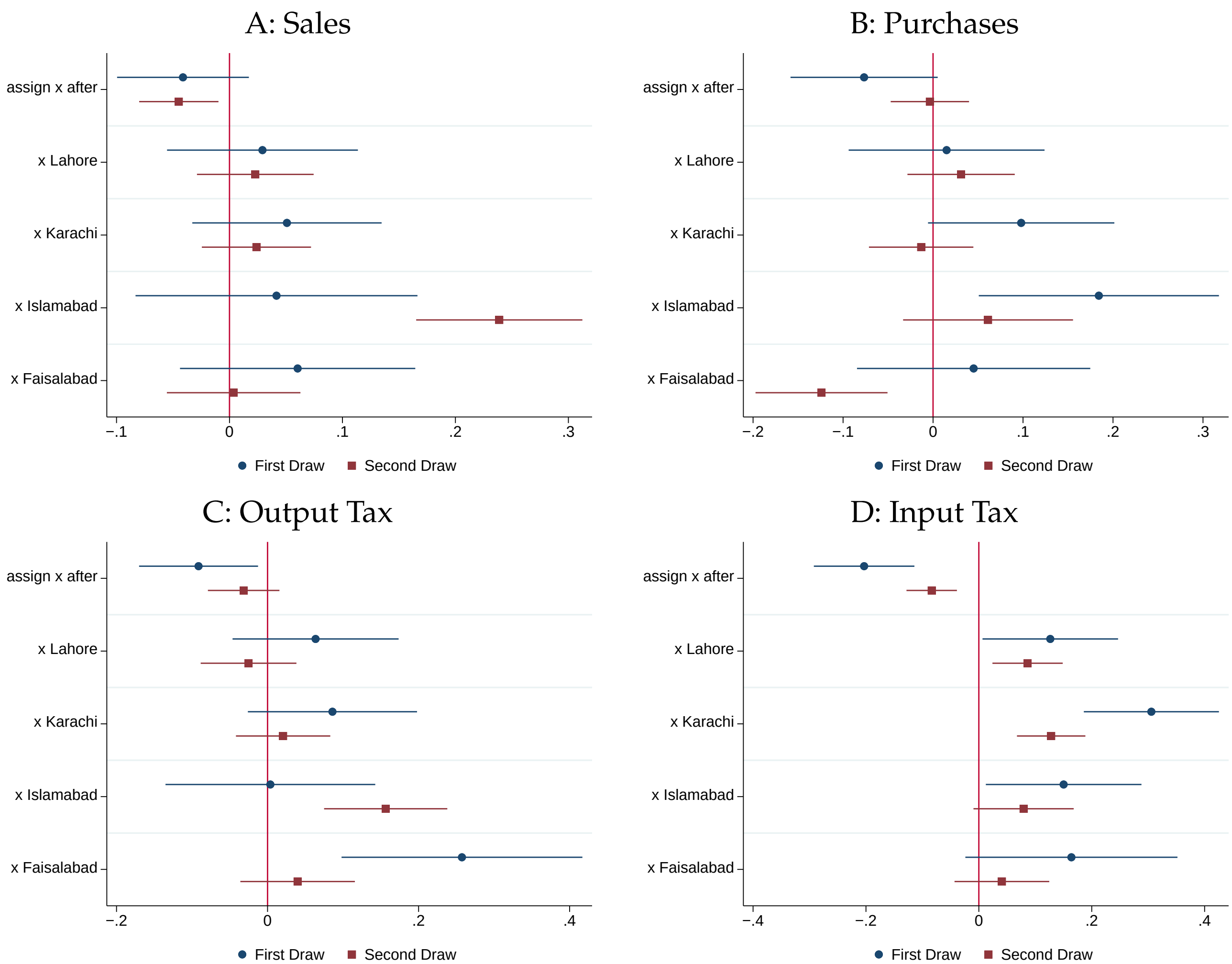
Notes: The figure explores heterogeneity in the audit effect. We divide firms into ten deciles based on their annual turnover in the baseline year. We then estimate a triple-difference version of model (6). The model includes interactions of the firm decile dummy with the $assign \times after_{it}$ dummy. The $assign_i$ dummies takes the value 1 if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The dummy variable $after_t$ indicates that month t falls after the date of the ballot. We drop the triple-interaction term involving the first decile. The coefficients and the 95% confidence intervals on the double and triple-interaction terms from these regressions are plotted. Regressions are run separately for the first and the second audit waves. The first wave results are in blue and the second wave results are in red. Standard errors are clustered at the firm level.

FIGURE A.V: **HETEROGENEITY IN RESPONSE BY FIRM AGE**



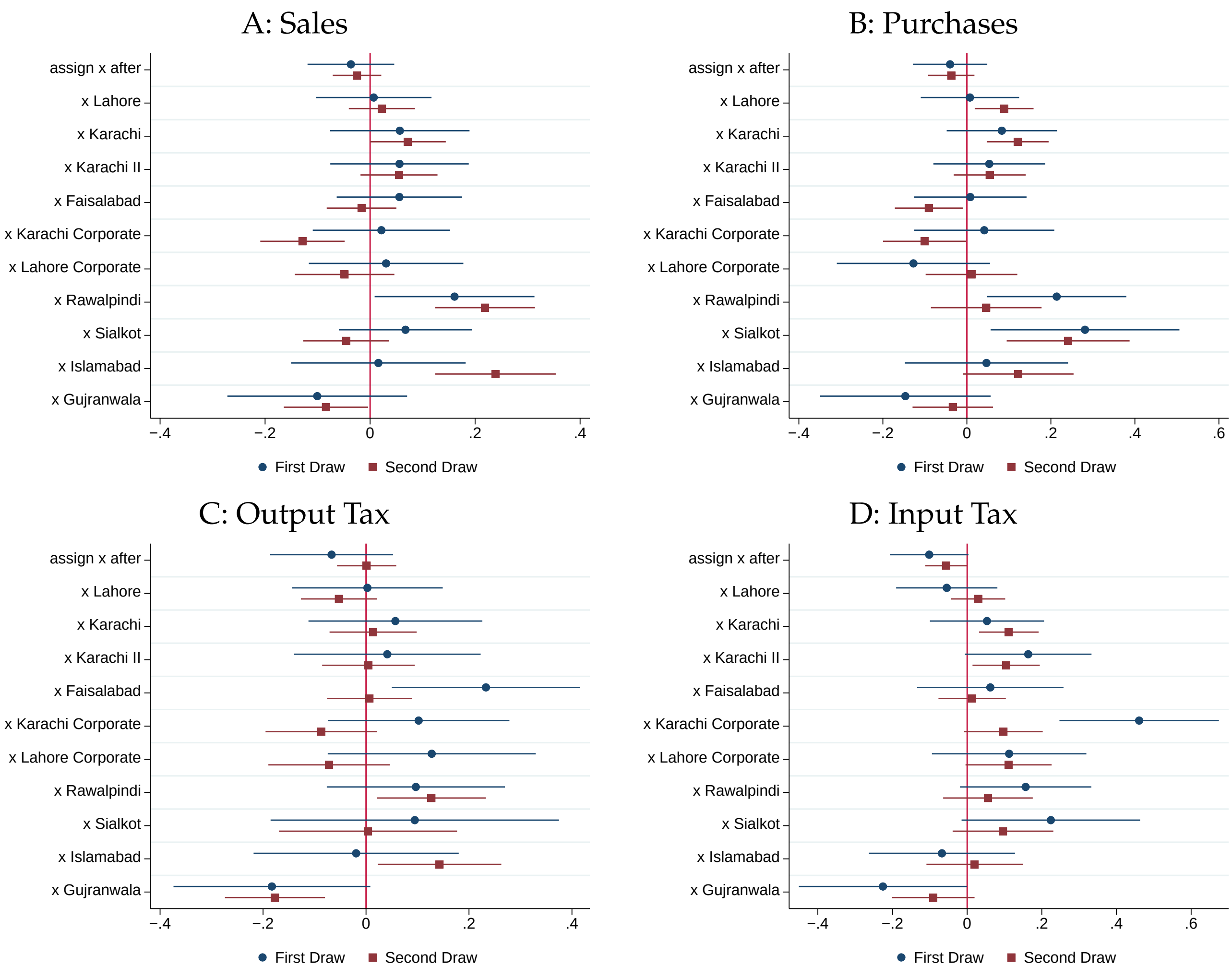
Notes: The figure explores heterogeneity in the audit effect. We divide firms into ten deciles based on their age, defining age as the number of days between July 1, 2013 and the date of registration of the firm. We then estimate a triple-difference version of model (6). The model includes interactions of the firm decile dummies with the $assign \times after_{it}$ dummy. The $assign_i$ dummy takes the value 1 if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The dummy variable $after_t$ indicates that month t falls after the date of the ballot. We drop the triple-interaction term involving the first decile. The coefficients and the 95% confidence intervals on the double and triple-interaction terms from these regressions are plotted. Regressions are run separately for the first and the second audit waves. The first wave results are in blue and the second wave results are in red. Standard errors are clustered at the firm level.

FIGURE A.VI: **HETEROGENEITY IN RESPONSE BY LOCATION**



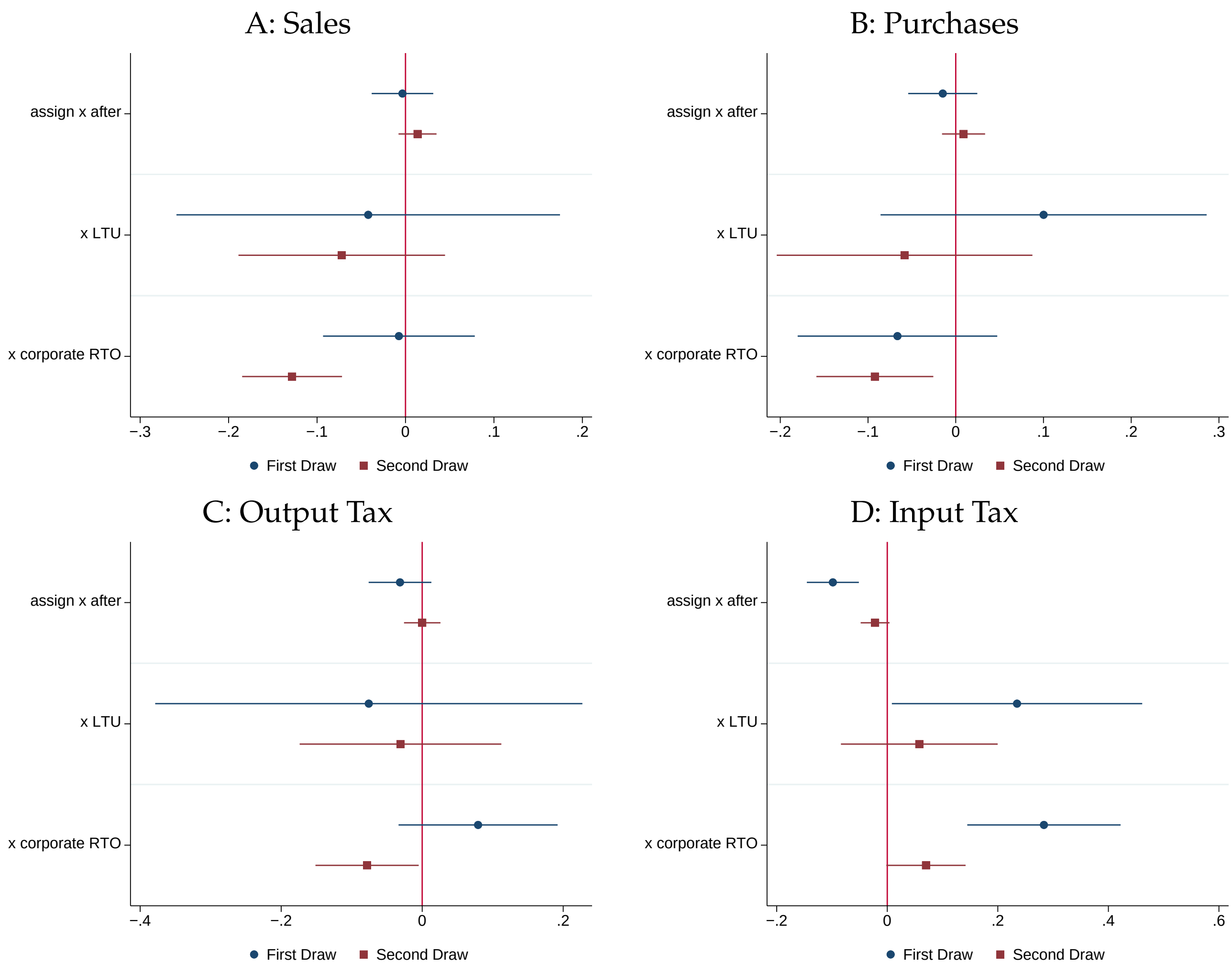
Notes: The figure explores heterogeneity in the audit effect. We divide firms into five groups depending upon the city their head office is located in. Firms not located in the four major cities of the country—Lahore, Karachi, Islamabad, and Faisalabad— are included in the baseline category. We then estimate a triple-difference version of model (6). The model includes interactions of the firm location dummies with the $assign \times after_{it}$ dummy. The $assign_i$ dummy takes the value 1 if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The dummy variable $after_t$ indicates that month t falls after the date of the ballot. The coefficients and the 95% confidence intervals on the double and triple-interaction terms from these regressions are plotted. Regressions are run separately for the first and the second audit waves. The first wave results are in blue and the second wave results are in red. Standard errors are clustered at the firm level.

FIGURE A.VII: **HETEROGENEITY IN RESPONSE BY TAX OFFICE**



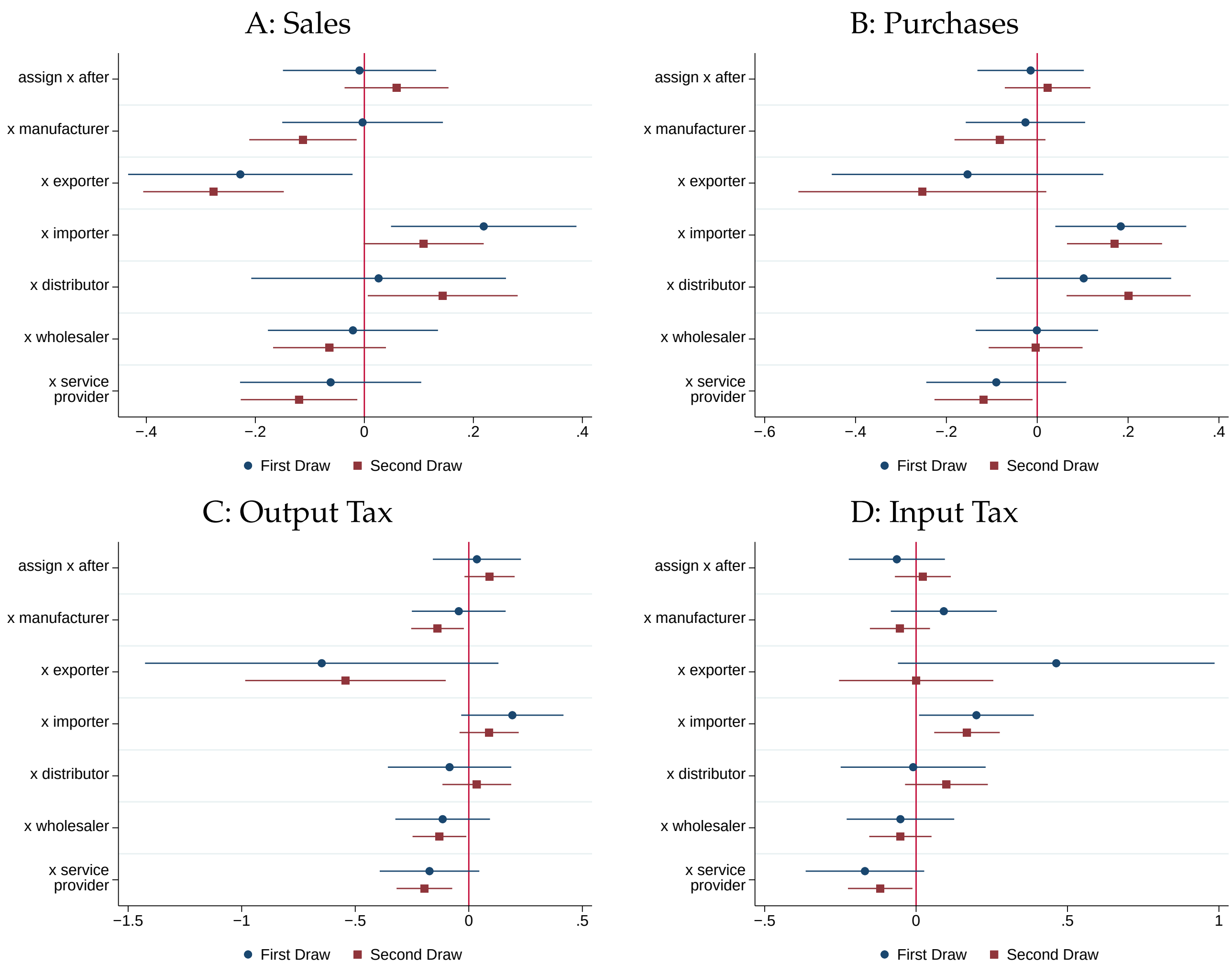
Notes: The figure explores heterogeneity in the audit effect. We divide firms into eleven groups based on the local tax office they are subject to. Firms not in the ten major tax offices are included in the baseline category. We then estimate a triple-difference version of model (6). The model includes interactions of the tax office dummies with the $assign \times after_{it}$ dummy. The $assign_i$ dummy takes the value 1 if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The dummy variable $after_t$ indicates that month t falls after the date of the ballot. The coefficients and the 95% confidence intervals on the double and triple-interaction terms from these regressions are plotted. Regressions are run separately for the first and the second audit waves. The first wave results are in blue and the second wave results are in red. Standard errors are clustered at the firm level.

FIGURE A.VIII: **HETEROGENEITY IN RESPONSE BY TAX OFFICE TYPE**



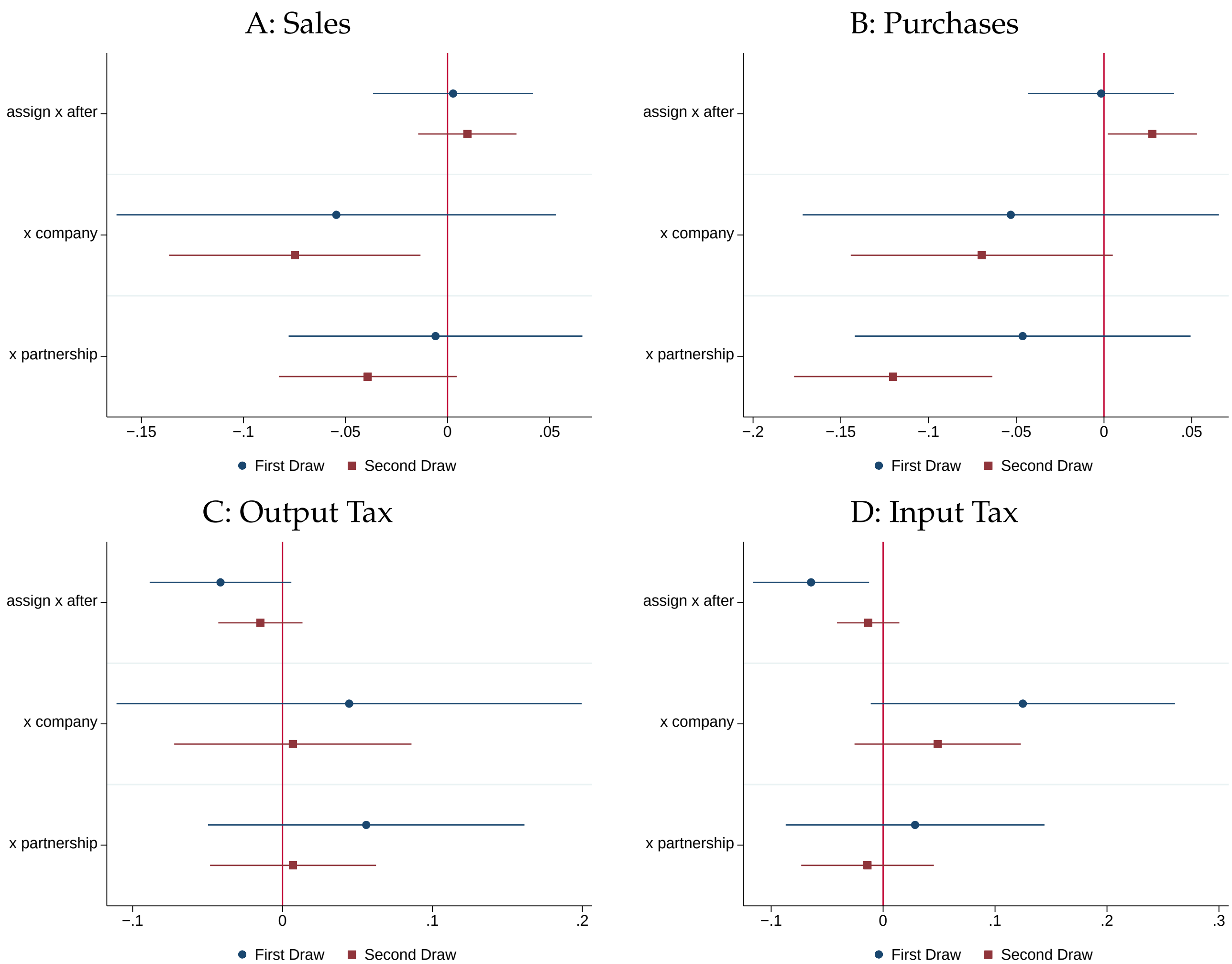
Notes: The figure explores heterogeneity in the audit effect. We divide firms into three groups based on the type of tax office they are subject to. Firms in four Large Taxpayer Units of the country are included in the first group (LTU), firms in the two Corporate Regional Tax Offices are included in the second group, and the rest of the firms are included in the baseline category. These firms are subject to a normal Regional Tax Office. We then estimate a triple-difference version of model (6). The model includes interactions of the tax office type dummies with the $assign \times after_{it}$ dummy. The $assign_i$ dummy takes the value 1 if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The dummy variable $after_t$ indicates that month t falls after the date of the ballot. The coefficients and the 95% confidence intervals on the double and triple-interaction terms from these regressions are plotted. Regressions are run separately for the first and the second audit waves. The first wave results are in blue and the second wave results are in red. Standard errors are clustered at the firm level.

FIGURE A.IX: **HETEROGENEITY IN RESPONSE BY PRODUCTION STAGE**



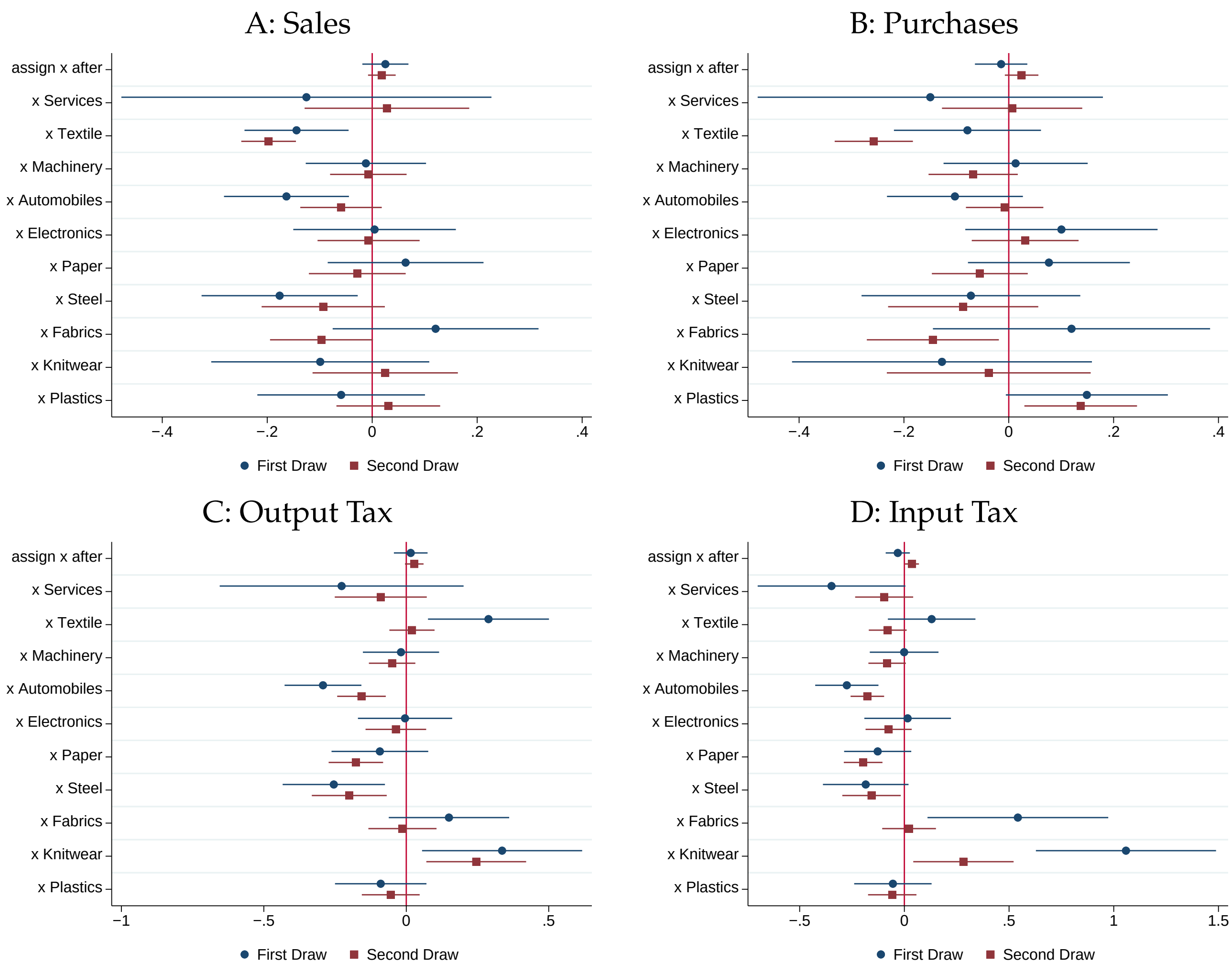
Notes: The figure explores heterogeneity in the audit effect. We divide firms into seven groups based on their principle business activity. The baseline category are retailers. These activities roughly capture the position of the firm in the supply chain. We then estimate a triple-difference version of model (6). The model includes interactions of the production stage dummies with the $assign \times after_{it}$ dummy. The $assign_i$ dummy takes the value 1 if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The dummy variable $after_t$ indicates that month t falls after the date of the ballot. The coefficients and the 95% confidence intervals on the double and triple-interaction terms from these regressions are plotted. Regressions are run separately for the first and the second audit waves. The first wave results are in blue and the second wave results are in red. Standard errors are clustered at the firm level.

FIGURE A.X: **HETEROGENEITY IN RESPONSE BY BUSINESS ORGANIZATION**



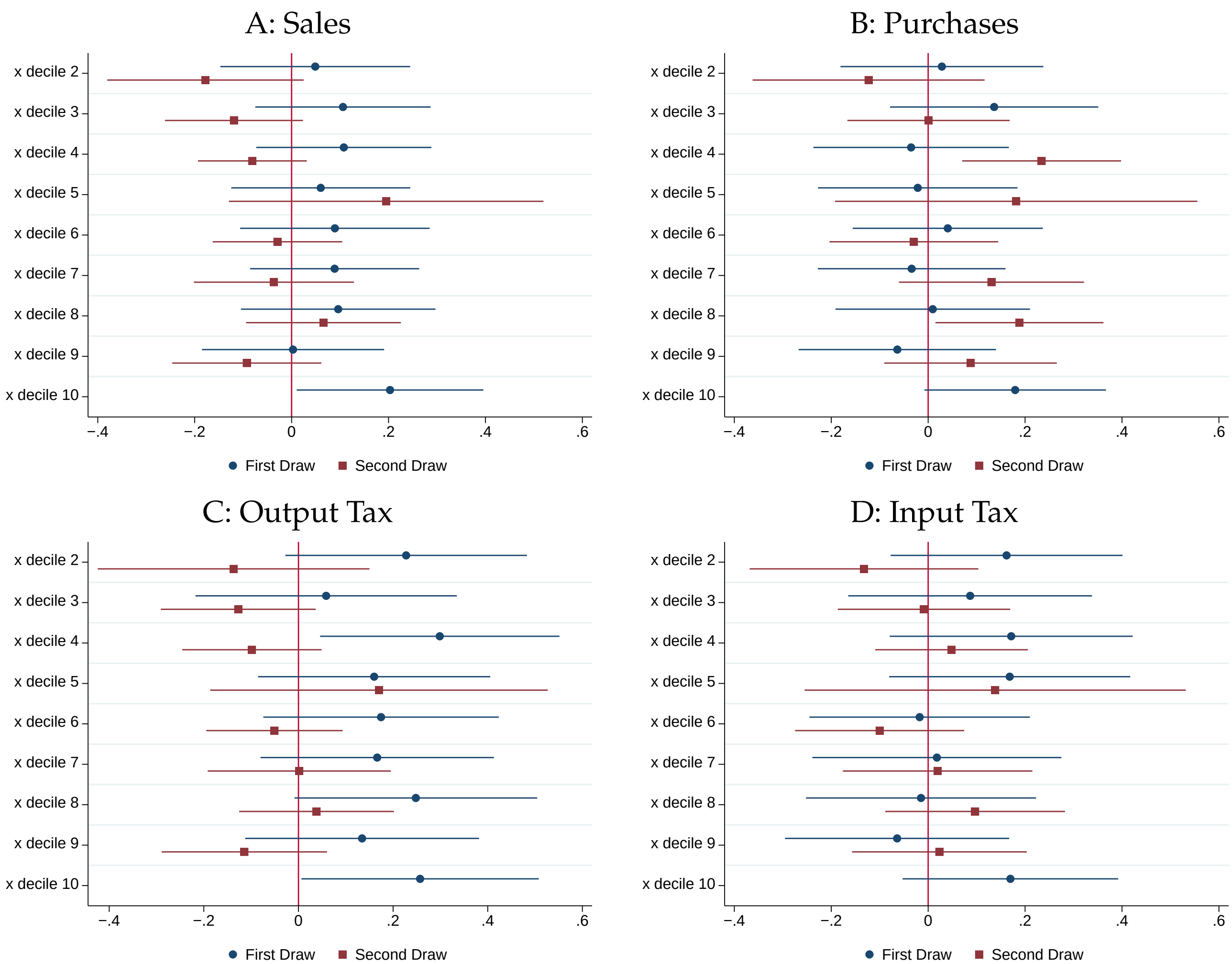
Notes: The figure explores heterogeneity in the audit effect. We divide firms into three groups based on their business organization. The baseline category are sole proprietors. We then estimate a triple-difference version of model (6). The model includes interactions of the business organization dummies with the $assign \times after_{it}$ dummy. The $assign_i$ dummy takes the value 1 if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The dummy variable $after_t$ indicates that month t falls after the date of the ballot. The coefficients and the 95% confidence intervals on the double and triple-interaction terms from these regressions are plotted. Regressions are run separately for the first and the second audit waves. The first wave results are in blue and the second wave results are in red. Standard errors are clustered at the firm level.

FIGURE A.XI: **HETEROGENEITY IN RESPONSE BY INDUSTRY**



Notes: The figure explores heterogeneity in the audit effect. We divide firms into 12 groups based on the industry they operate in. We separate firms in 11 major industries of the country and club the rest into the baseline category. We then estimate a triple-difference version of model (6). The model includes interactions of the industry dummies with the $assign \times after_{it}$ dummy. The $assign_i$ dummy takes the value 1 if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The dummy variable $after_t$ indicates that month t falls after the date of the ballot. The coefficients and the 95% confidence intervals on the double and triple-interaction terms from these regressions are plotted. Regressions are run separately for the first and the second audit waves. The first wave results are in blue and the second wave results are in red. Standard errors are clustered at the firm level.

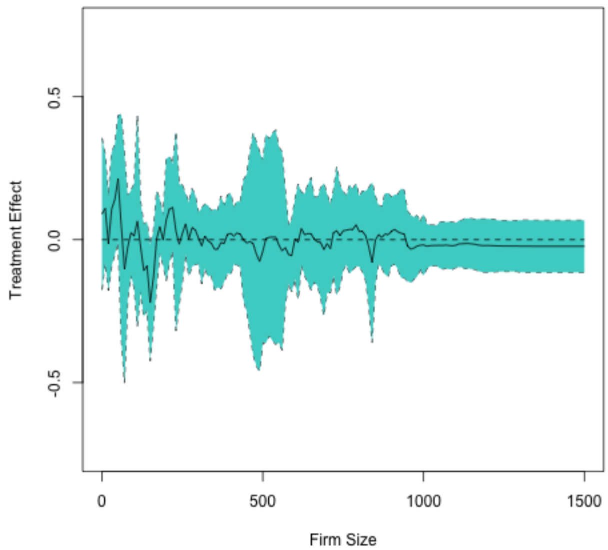
FIGURE A.XII: **HETEROGENEITY IN RESPONSE BY TIMING OF AUDIT**



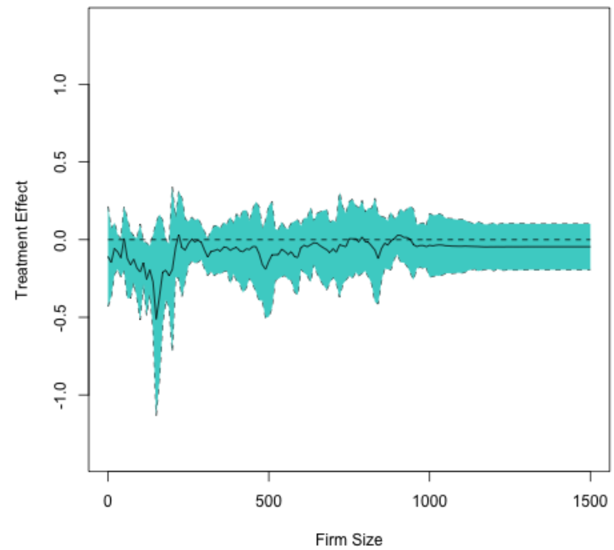
Notes: The figure explores heterogeneity in the audit effect. We divide firms into ten deciles based on the time lag between the assignment and initiation of audit in days. We then estimate a triple-difference version of model (6). The model includes interactions of the firm decile dummies with the $assign \times after_{it}$ dummy. The $assign_i$ dummy takes the value 1 if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The dummy variable $after_t$ indicates that month t falls after the date of the ballot. We drop the triple-interaction term involving the first decile. The coefficients and the 95% confidence intervals on the double and triple-interaction terms from these regressions are plotted. Regressions are run separately for the first and the second audit waves. The first wave results are in blue and the second wave results are in red. Standard errors are clustered at the firm level.

FIGURE A.XIII: **HETEROGENEITY IN RESPONSE BY FIRM SIZE (FIRST WAVE)**

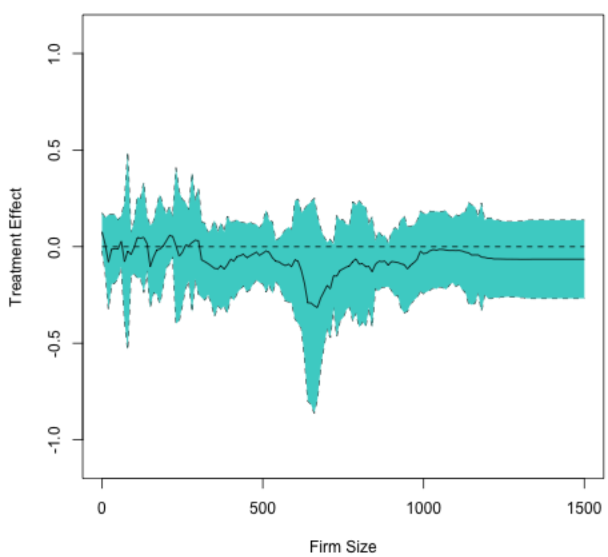
A: Sales



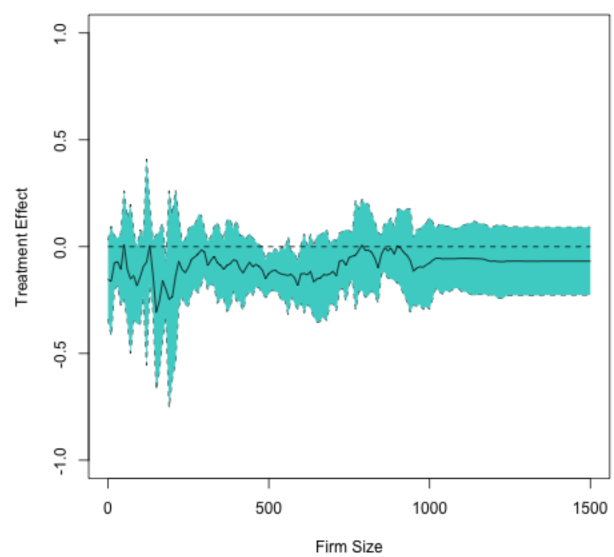
B: Purchases



C: Output Tax

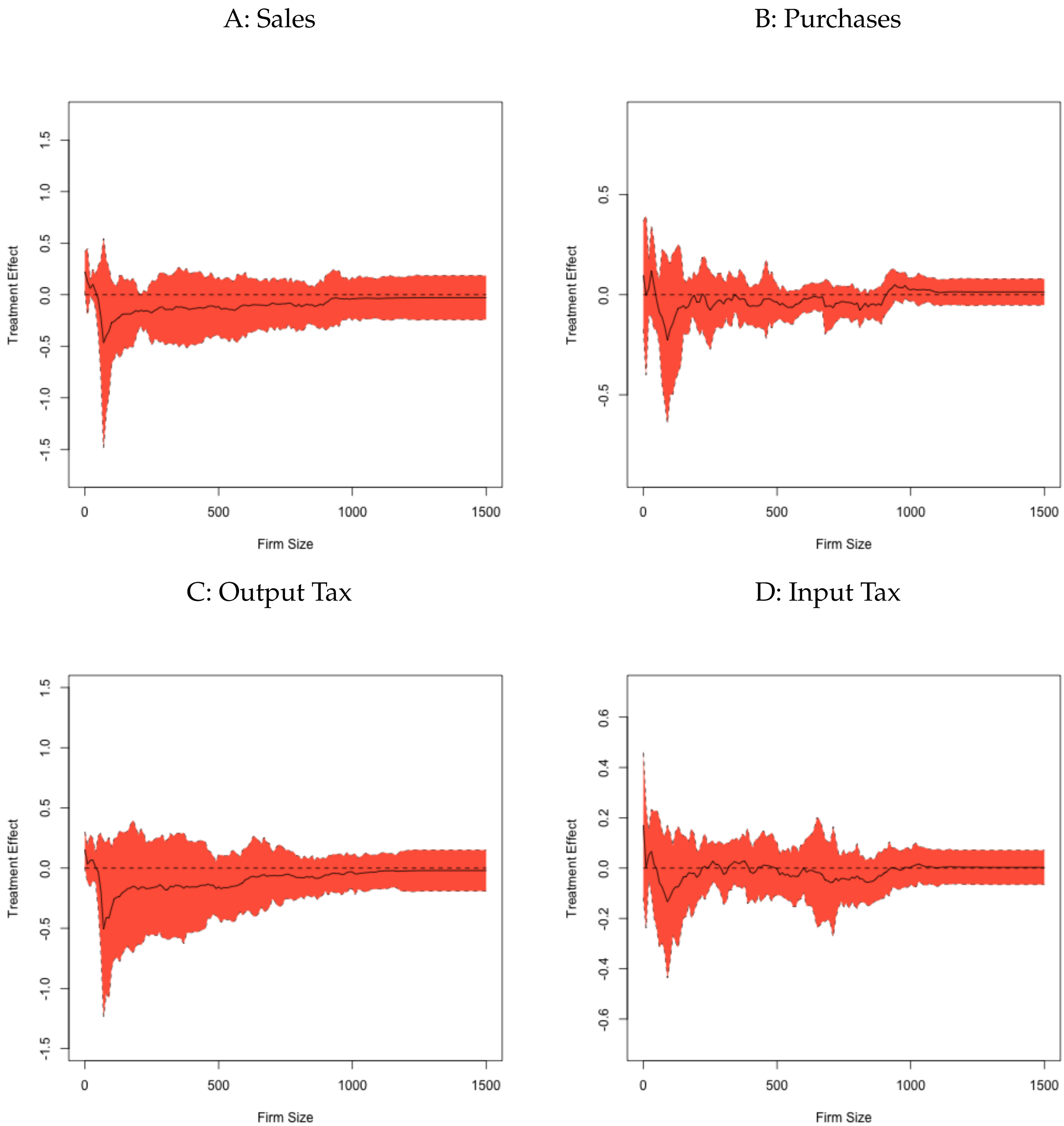


D: Input Tax



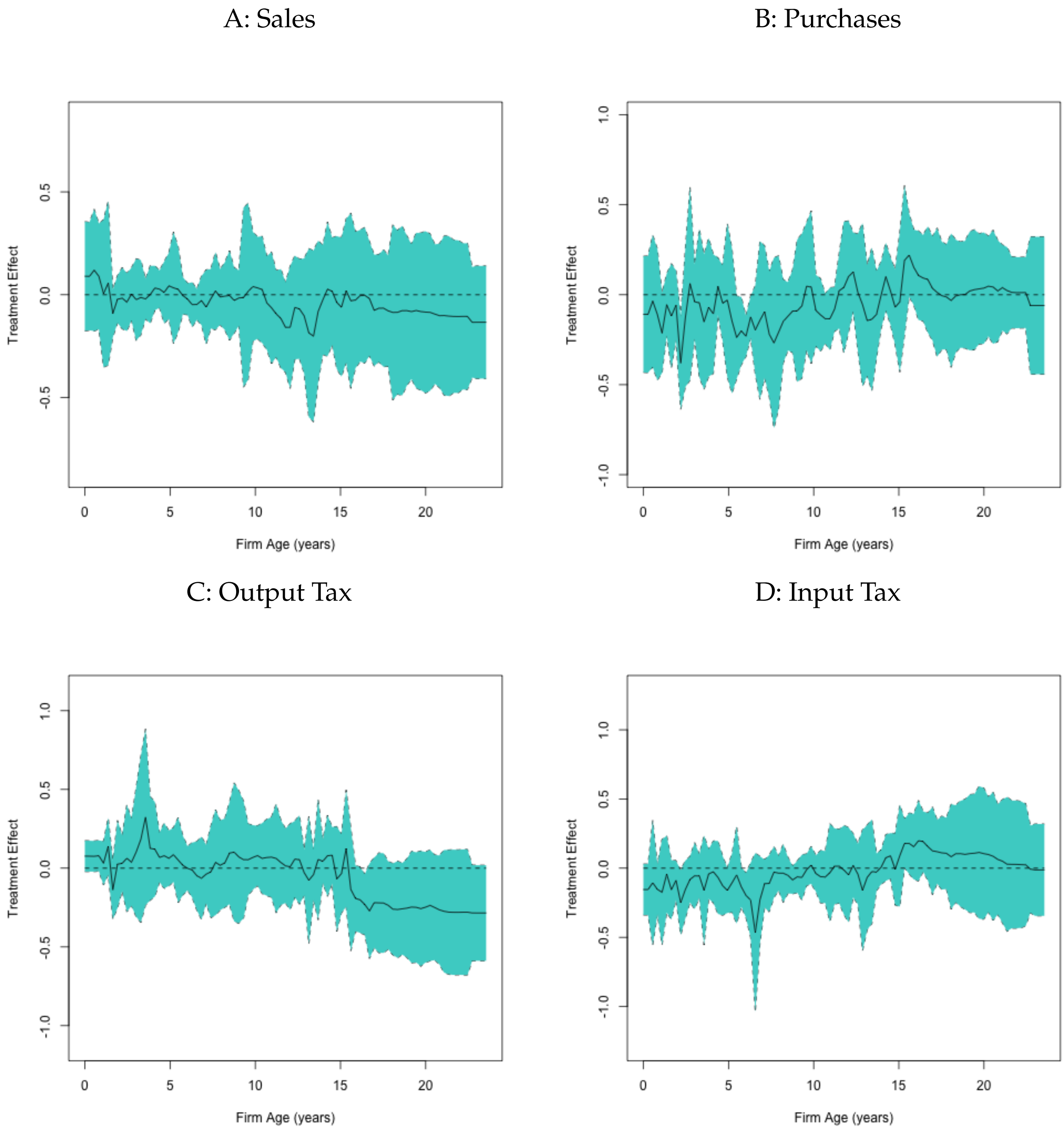
Notes: The figure explores heterogeneity in the audit effect. We use firm-size as a continuous variable. We then use a generalized random forest model to estimate the treatment effects of the audit for all values within the feasible range based on the available data. We consider a firm as treated (audited) if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The estimated treatment effects and 95% confidence intervals on the estimated treatment effects are plotted. Models are estimated separately for each outcome variable.

FIGURE A.XIV: **HETEROGENEITY IN RESPONSE BY FIRM SIZE (SECOND WAVE)**



Notes: The figure explores heterogeneity in the audit effect. We use firm-size as a continuous variable. We then use a generalized random forest model to estimate the treatment effects of the audit for all values within the feasible range based on the available data. We consider a firm as treated (audited) if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The estimated treatment effects and 95% confidence intervals on the estimated treatment effects are plotted. Models are estimated separately for each outcome variable.

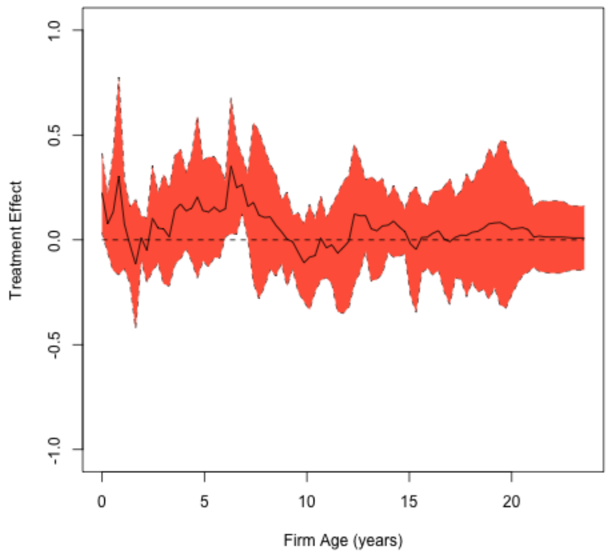
FIGURE A.XV: **HETEROGENEITY IN RESPONSE BY FIRM AGE (FIRST WAVE)**



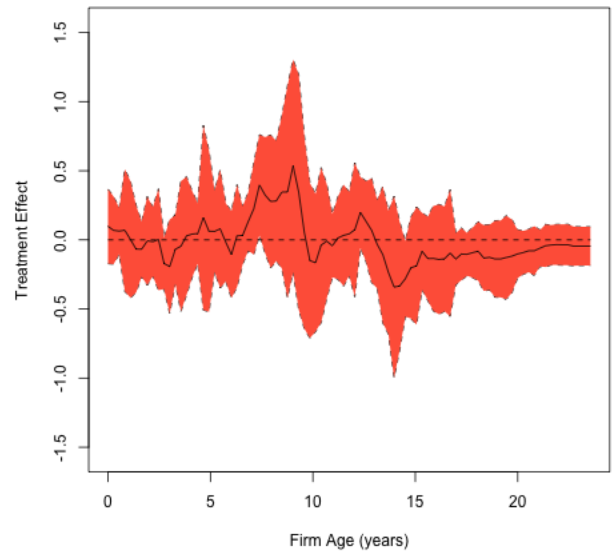
Notes: The figure explores heterogeneity in the audit effect. We use firm-age as a continuous variable. We then use a generalized random forest model to estimate the treatment effects of the audit for all values within the feasible range based on the available data. We consider a firm as treated (audited) if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The estimated treatment effects and 95% confidence intervals on the estimated treatment effects are plotted. Models are estimated separately for each outcome variable.

FIGURE A.XVI: **HETEROGENEITY IN RESPONSE BY FIRM AGE (SECOND WAVE)**

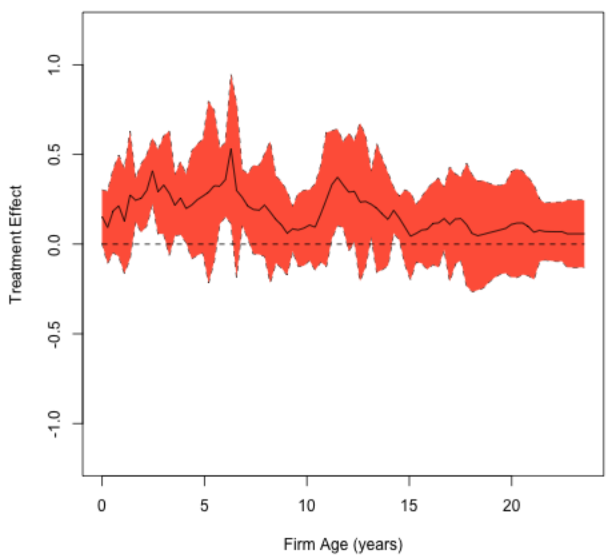
A: Sales



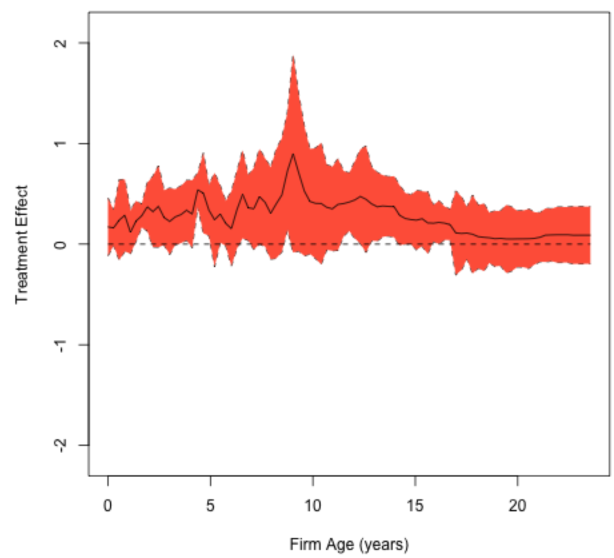
B: Purchases



C: Output Tax

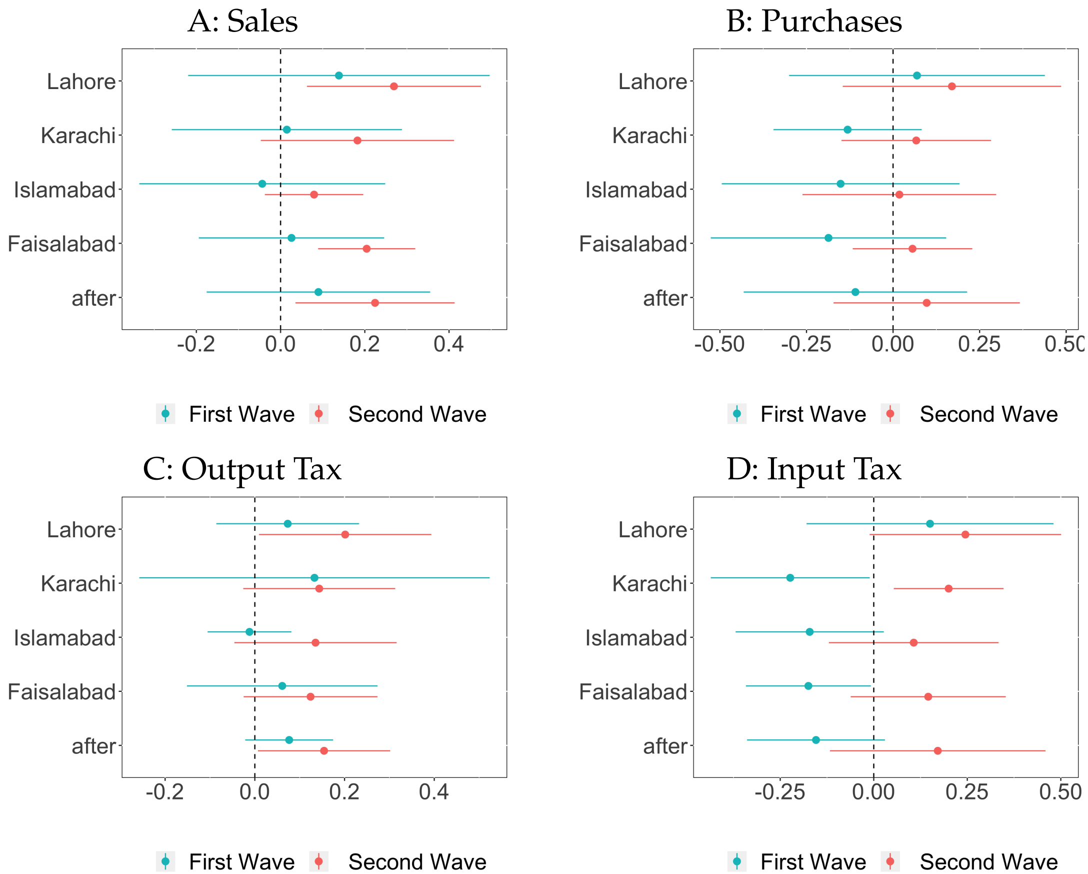


D: Input Tax



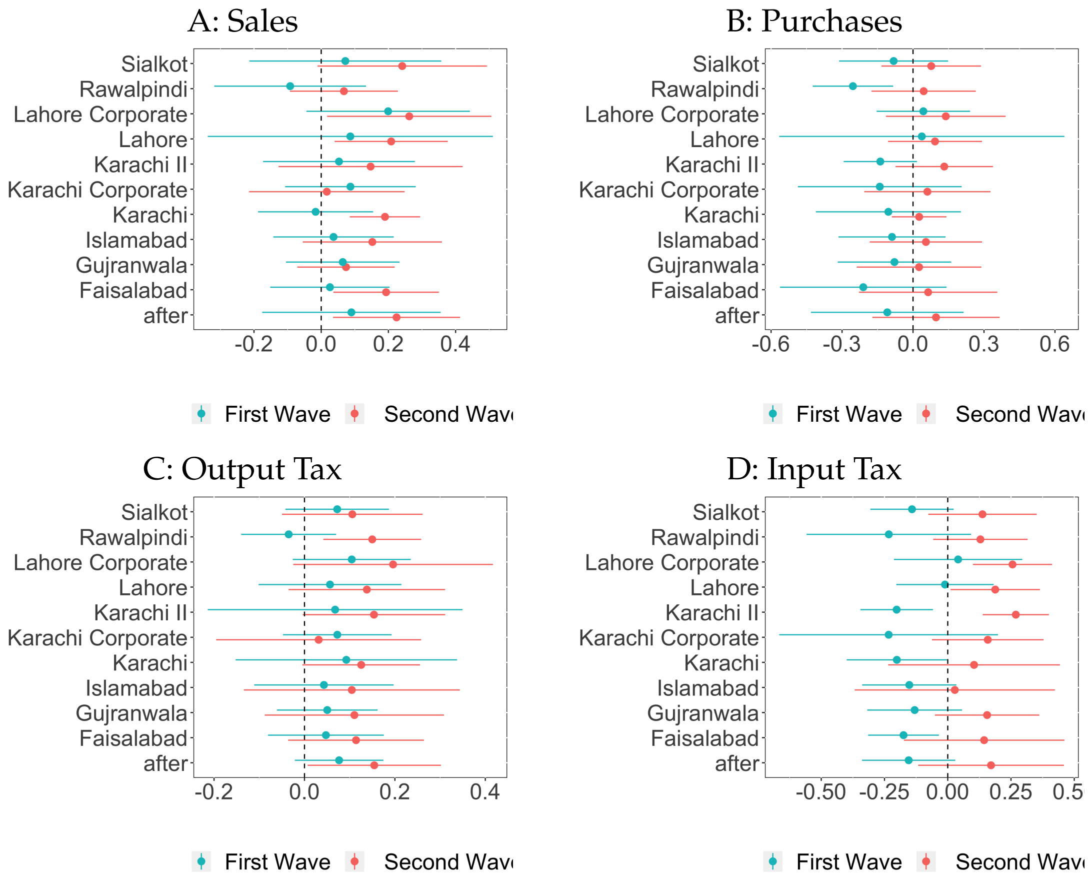
Notes: The figure explores heterogeneity in the audit effect. We use firm-age as a continuous variable. We then use a generalized random forest model to estimate the treatment effects of the audit for all values within the feasible range based on the available data. We consider a firm as treated (audited) if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The estimated treatment effects and 95% confidence intervals on the estimated treatment effects are plotted. Models are estimated separately for each outcome variable.

FIGURE A.XVII: **HETEROGENEITY IN RESPONSE BY LOCATION**



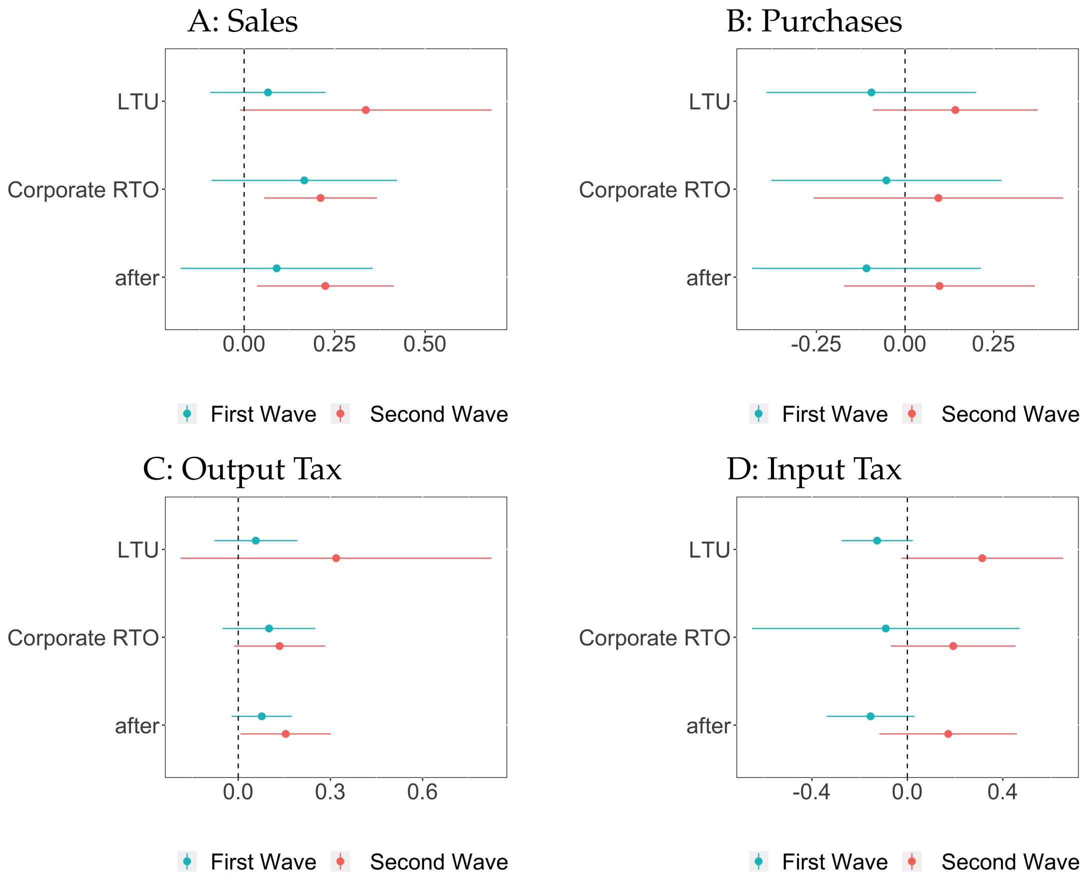
Notes: The figure explores heterogeneity in the audit effect. We divide firms into five groups depending upon the city their head office is located in. Firms not located in the four major cities of the country— Lahore, Karachi, Islamabad, and Faisalabad— are included in the baseline category. We then use a generalized random forest model to estimate the treatment effects of the audit. The model includes dummy variables for each group along with a dummy variable for "after" - indicating the time period after the date of the ballot. We consider a firm as treated (audited) if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The coefficients and the 95% confidence intervals on the estimated treatment effects are plotted. Models are estimated separately for the first and the second audit waves and for each outcome variable. The first wave results are in blue and the second wave results are in red.

FIGURE A.XVIII: **HETEROGENEITY IN RESPONSE BY TAX OFFICE**



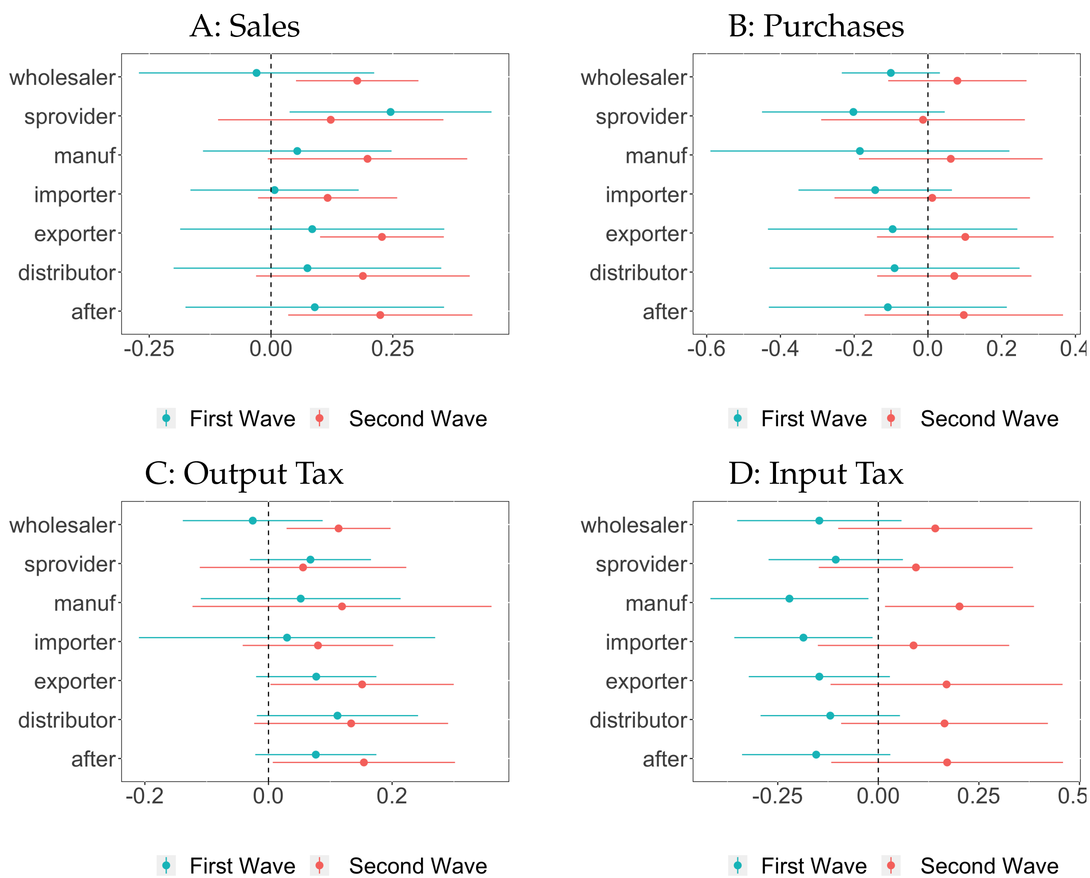
Notes: The figure explores heterogeneity in the audit effect. We divide firms into eleven groups based on the local tax office they are subject to. Firms not in the ten major tax offices are included in the baseline category. We then use a generalized random forest model to estimate the treatment effects of the audit. The model includes dummy variables for each group along with a dummy variable for "after" - indicating the time period after the date of the ballot. We consider a firm as treated (audited) if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The coefficients and the 95% confidence intervals on the estimated treatment effects are plotted. Models are estimated separately for the first and the second audit waves and for each outcome variable. The first wave results are in blue and the second wave results are in red.

FIGURE A.XIX: **HETEROGENEITY IN RESPONSE BY TAX OFFICE TYPE**



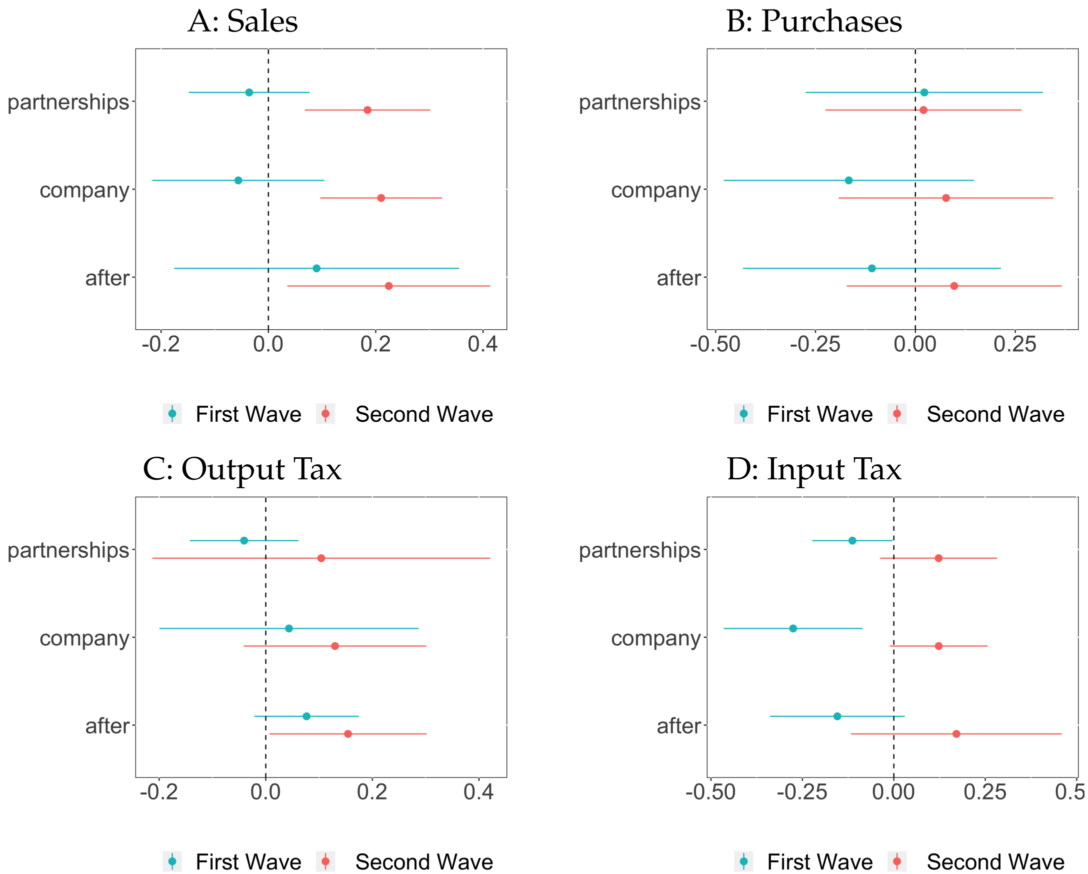
Notes: The figure explores heterogeneity in the audit effect. We divide firms into three groups based on the type of tax office they are subject to. Firms in four Large Taxpayer Units of the country are included in the first group (LTU), firms in the two Corporate Regional Tax Offices are included in the second group, and the rest of the firms are included in the baseline category. These firms are subject to a normal Regional Tax Office. We then use a generalized random forest model to estimate the treatment effects of the audit. The model includes dummy variables for each tax office type along with a dummy variable for "after" - indicating the time period after the date of the ballot. We consider a firm as treated (audited) if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The coefficients and the 95% confidence intervals on the estimated treatment effects are plotted. Models are estimated separately for the first and the second audit waves and for each outcome variable. The first wave results are in blue and the second wave results are in red.

FIGURE A.XX: **HETEROGENEITY IN RESPONSE BY PRODUCTION STAGE**



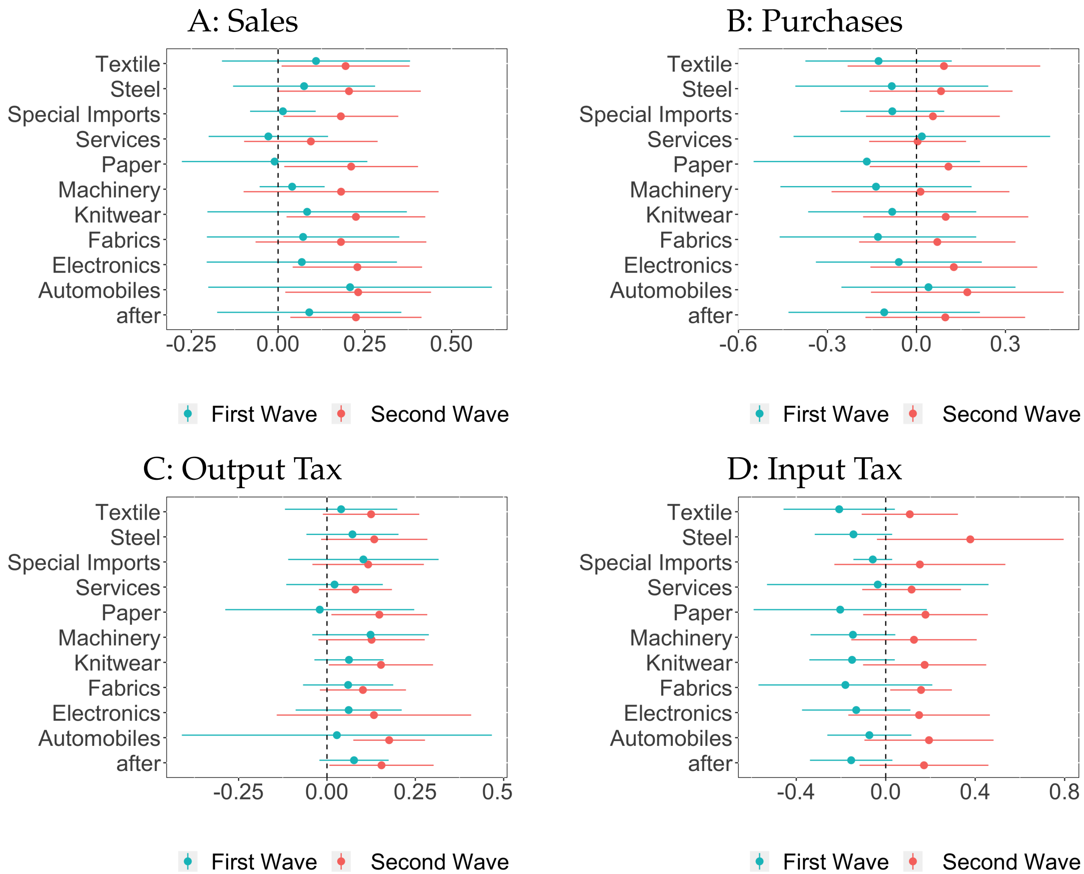
Notes: The figure explores heterogeneity in the audit effect. We divide firms into seven groups based on their principle business activity. The baseline category are retailers. These activities roughly capture the position of the firm in the supply chain. We then use a generalized random forest model to estimate the treatment effects of the audit. The model includes dummy variables for each group along with a dummy variable for "after" - indicating the time period after the date of the ballot. We consider a firm as treated (audited) if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The coefficients and the 95% confidence intervals on the estimated treatment effects are plotted. Models are estimated separately for the first and the second audit waves and for each outcome variable. The first wave results are in blue and the second wave results are in red.

FIGURE A.XXI: **HETEROGENEITY IN RESPONSE BY BUSINESS ORGANIZATION**



Notes: The figure explores heterogeneity in the audit effect. We divide firms into three groups based on their business organization. The baseline category are sole proprietors We then use a generalized random forest model to estimate the treatment effects of the audit. The model includes dummy variables for each group along with a dummy variable for "after" - indicating the time period after the date of the ballot. We consider a firm as treated (audited) if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The coefficients and the 95% confidence intervals on the estimated treatment effects are plotted. Models are estimated separately for the first and the second audit waves and for each outcome variable. The first wave results are in blue and the second wave results are in red.

FIGURE A.XXII: **HETEROGENEITY IN RESPONSE BY INDUSTRY**



Notes: The figure explores heterogeneity in the audit effect. We divide firms into 11 groups based on the industry they operate in. We separate firms in 10 major industries of the country and club the rest into the baseline category. We then use a generalized random forest model to estimate the treatment effects of the audit. The model includes dummy variables for each group along with a dummy variable for "after" - indicating the time period after the date of the ballot. We consider a firm as treated (audited) if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The coefficients and the 95% confidence intervals on the estimated treatment effects are plotted. Models are estimated separately for the first and the second audit waves and for each outcome variable. The first wave results are in blue and the second wave results are in red.

TABLE A.I: **BREAKDOWN OF THE DETECTED AMOUNT**

	Amt. Detected		Amt. Recovered		Amt. Recoverable		Refund Curtailed	
	PKR	%	PKR	%	PKR	%	PKR	%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A: First Audit Wave								
All Audited Firms	2.147	0.431	0.023	0.005	2.118	0.425	0.004	0.001
Amount Detected > 0	2.147	1.567	0.023	0.017	2.118	1.546	0.004	0.003
Size Quartile 1	0.062	684.756	0.001	11.221	0.061	673.534	0.000	0.000
Size Quartile 2	0.067	3.936	0.003	0.186	0.064	3.750	0.000	0.000
Size Quartile 3	0.215	1.746	0.008	0.067	0.203	1.648	0.003	0.021
Size Quartile 4	1.802	0.372	0.011	0.002	1.790	0.370	0.002	0.000
B: Second Audit Wave								
All Audited Firms	2.235	0.102	0.040	0.002	2.191	0.100	0.003	0.000
Amount Detected > 0	2.235	0.845	0.040	0.015	2.191	0.828	0.003	0.001
Size Quartile 1	0.045	10.205	0.002	0.473	0.042	9.649	0.000	0.000
Size Quartile 2	0.166	3.367	0.009	0.179	0.157	3.188	0.000	0.000
Size Quartile 3	0.217	0.889	0.009	0.036	0.205	0.840	0.003	0.012
Size Quartile 4	1.808	0.083	0.020	0.001	1.786	0.082	0.000	0.000

Notes: The table breaks down the total amount detected by audit (columns 1-2) into its three major components (columns 3-8). The odd-number columns report the amounts in PKR billions and the even-number columns the amount as a ratio of the aggregate annual turnover of the corresponding group of firm. Amount Recovered is the amount paid by the taxpayer as a result of audit. Amount Recoverable, on the other hand, is unpaid amount out of the total detected by audit. This amount is subject to quasi-judicial determination and appeal processes. Refund Curtailed indicates the amount by which the firm agreed to reduce its refund claim pending with the department.

TABLE A.II: **SELECTION IN SEQUENCING OF AUDITS**

	Outcome: Days between assignment and initiation			
	(1)	(2)	(3)	(4)
Sales	-1.785	-4.301	2.542	2.679
	(7.492)	(7.489)	(2.749)	(2.657)
Purchases	-0.727	-3.636	-2.583	0.588
	(8.569)	(8.568)	(5.626)	(5.433)
Output Tax	9.936	8.030	-2.929	0.718
	(30.624)	(30.012)	(12.057)	(11.651)
Input Tax	-4.050	1.030	3.229	-2.713
	(14.118)	(13.982)	(11.034)	(10.648)
Tax Paid	-6.673	-3.513	-1.108	-2.919
	(23.011)	(22.538)	(4.718)	(4.554)
Exports	-0.550	-0.126	1.836	2.399
	(1.560)	(1.540)	(1.002)	(0.974)
Imports	-0.201	-0.223	-0.370	-0.264
	(1.884)	(1.916)	(0.643)	(0.624)
Refund	1.382	1.662	-1.847	-2.325
	(1.395)	(1.377)	(0.866)	(0.840)
Carry Forward	1.734	1.132	-0.143	-0.300
	(3.374)	(3.355)	(0.569)	(0.549)
Manufacturer	-13.271	-11.003	-1.860	-1.986
	(5.331)	(5.298)	(1.615)	(1.581)
Importer	-0.785	-0.614	-3.302	0.310
	(6.230)	(6.190)	(1.833)	(1.791)
Exporter	1.834	6.001	-1.649	-1.134
	(9.390)	(9.301)	(2.282)	(2.295)
Distributor	7.098	9.746	-0.251	-1.645
	(9.143)	(8.977)	(2.469)	(2.395)
Wholesaler	-5.548	-2.847	-1.848	0.958
	(5.391)	(5.315)	(1.669)	(1.624)
Service Provider	-7.959	-4.111	0.109	1.141
	(5.332)	(5.247)	(1.661)	(1.606)
Constant	46.995	44.436	18.961	17.830
	(4.843)	(4.768)	(1.490)	(1.443)
Observations	3,482	3,481	3,612	3,611
Corporation FEs	Yes	Yes	Yes	Yes
Tax Office FEs	No	Yes	No	Yes

Notes*:* The table explores selection in audit. We regress the time lag measured in number of days between the assignment and initiation of audit on baseline firm characteristics. We standardize the first nine variables in this table by subtracting the mean and dividing by the standard deviation of the variable. Since audits were taken up by local tax offices, we include the tax office fixed effects in even-numbered columns. The first two columns report results for the first audit wave and the last two for the second audit wave. Standard errors are in parenthesis.

TABLE A.III: **PREEXISTING TRENDS**

	First Wave					Second Wave				
	Sales	Purchases	Output	Input	Tax Payable	Sales	Purchases	Output	Input	Tax Payable
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
assign $\times$ year $\in [s-1, s]$	-0.018	-0.005	-0.039	-0.004	-0.033	-0.016	-0.018	-0.027	-0.030	0.002
	(0.015)	(0.017)	(0.020)	(0.021)	(0.025)	(0.010)	(0.011)	(0.013)	(0.014)	(0.017)
assign $\times$ year $\in [s-3, s]$	0.001	0.021	-0.031	0.021	-0.006	-0.006	-0.014	-0.005	-0.020	0.012
	(0.014)	(0.016)	(0.018)	(0.020)	(0.022)	(0.010)	(0.012)	(0.014)	(0.014)	(0.017)
assign $\times$ year $\in [s-5, s]$	-0.004	0.042	-0.019	0.040	0.051	0.028	0.007	0.020	-0.002	0.056
	(0.021)	(0.022)	(0.022)	(0.024)	(0.033)	(0.011)	(0.012)	(0.013)	(0.014)	(0.017)
Observations	2,324,186	2,025,380	1,672,095	1,681,583	1,154,574	2,628,878	2,290,848	1,934,273	1,945,733	1,312,928
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes*:* The table explores if the preexisting trends for the five outcomes indicated in the heading of each column were parallel between firms who were picked for audit in a random ballot and other firms in the eligible sample. We estimate a model similar to (6) replacing the $assign \times after_{it}$ dummy with three dummies shown in the top three rows. The dummy variable $assign_i$ denotes that firm i's was picked for audit in the random ballot indicated in the heading of the column. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The sample for these regressions include the baseline periods only, from July 2008 to August 2013 for the first wave and from July 2008 to August 2014 for the second. The dummy variable $year \in [s-1, s]$ indicates that the period is one of the last twelve months included in the regression and so on. Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE A.IV: **IMPACTS OF RANDOM AUDITS ASSIGNED IN THE FIRST WAVE**

	Impacts After One Year					Impacts After Three Years				
	Sales	Purchases	Output Tax	Input Tax	Tax Payable	Sales	Purchases	Output Tax	Input Tax	Tax Payable
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A: ITT Estimates										
assign × after	-0.009	-0.009	-0.016	-0.017	-0.037	-0.007	-0.021	-0.025	-0.036	-0.015
	(0.016)	(0.019)	(0.021)	(0.026)	(0.027)	(0.014)	(0.019)	(0.019)	(0.023)	(0.030)
Observations	2,802,387	2,456,864	2,061,472	2,089,489	1,393,541	3,809,614	3,315,994	2,857,885	2,895,330	1,890,220
B: LATE Estimates										
treat × after	-0.013	-0.014	-0.022	-0.024	-0.051	-0.010	-0.030	-0.035	-0.051	-0.021
	(0.022)	(0.027)	(0.029)	(0.037)	(0.036)	(0.019)	(0.027)	(0.026)	(0.031)	(0.041)
Observations	2,802,387	2,456,864	2,061,472	2,089,489	1,393,541	3,809,614	3,315,994	2,857,885	2,895,330	1,890,220
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes*:* The table estimates the impact of audit on firms' future behavior. In the top panel, the coefficient assign × after shows $\hat{\gamma}$ from model (6), where the dummy variable assign$_i$ denotes that firm i's audit was assigned through the first random ballot held on September 13, 2013. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The dummy variable after$_t$ indicates that month t falls after the date of the ballot. The sample includes periods up to October 2014 for the first five columns and periods up to October 2016 for the rest. Panel B shows the corresponding results from 2sls regressions, where the endogenous variable audit$_i$ is instrumented by the initial random assignment. Standard errors are in parenthesis, which have been clustered at the tax office level.

TABLE A.V: **AUDIT IMPACTS – FIRST STAGE**

Outcome:	$audit \times after_{it}$					
Random Draw Held On:	September 13, 2013		September 25, 2014		September 14, 2015	
Post Sample:	One Year	Three Years	One Year	Three Years	One Year	Three Years
	(1)	(2)	(3)	(4)	(5)	(6)
assign $\times$ after	0.704	0.703	0.294	0.296	0.133	0.134
	(0.007)	(0.007)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	6,893,186	9,681,146	7,894,004	10721371	8,241,185	10829729
F Statistic	10,353	10,071	4,751	4,658	1,120	1,102

Notes: The table reports the first stage of our 2sls models. We estimate model (6) using the dummy $treat \times after_{it}$ as the outcome variable, where $treat_i$ takes the value 1 if firm i was audited in the corresponding audit wave indicated in the heading of each column. The coefficient assign $\times$ after shows $\hat{\gamma}$ from these regressions. The dummy variable assign$_i$ denotes that firm i's audit was assigned through the random ballot indicated in the heading of each column. The sample includes the population of VAT filers excluding government departments and firms already under audit. The dummy variable after$_t$ indicates that month t falls after the date of the ballot. We report results for two Post Samples: One Year specifications include twelve $after_t$ periods and Three Years specifications include 36 $after_t$ periods. In each case, the samples includes all months from July 2008 to the last $after_t$ period. Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE A.VI: **IMPACTS OF RANDOM AUDITS ASSIGNED IN THE THIRD WAVE**

	Impacts After One Year					Impacts After Three Years				
	Sales	Purchases	Output Tax	Input Tax	Tax Payable	Sales	Purchases	Output Tax	Input Tax	Tax Payable
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
assign × after	-0.034	-0.024	-0.039	-0.009	0.004	-0.050	-0.040	-0.071	-0.076	-0.093
	(0.011)	(0.013)	(0.014)	(0.014)	(0.014)	(0.011)	(0.014)	(0.015)	(0.015)	(0.015)
Observations	3,007,568	2,590,734	2,256,294	2,265,080	2,758,303	3,910,133	3,341,025	2,879,242	2,930,477	3,577,794
B: LATE Estimates										
treat × after	-0.261	-0.185	-0.296	-0.063	0.033	-0.376	-0.297	-0.487	-0.527	-0.652
	(0.083)	(0.102)	(0.106)	(0.105)	(0.108)	(0.087)	(0.106)	(0.110)	(0.108)	(0.112)
Observations	3,007,568	2,590,734	2,256,294	2,265,080	2,758,303	3,910,133	3,341,025	2,879,242	2,930,477	3,577,794
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes*:* The table estimates the impact of audit on firms' future behavior. In the top panel, the coefficient assign × after shows $\hat{\gamma}$ from model (6), where the dummy variable assign$_i$ denotes that firm i's audit was assigned through the first random ballot held on September 14, 2015. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The dummy variable after$_t$ indicates that month t falls after the date of the ballot. The sample includes periods up to October 2016 for the first five columns and periods up to October 2018 for the rest. Panel B shows the corresponding results from 2sls regressions, where the endogenous variable audit$_i$ is instrumented by the initial random assignment. Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE A.VII: **PREEXISTING TRENDS – AUDITED VS. NOT AUDITED**

	First Wave					Second Wave				
	Sales	Purchases	Output	Input	Tax Payable	Sales	Purchases	Output	Input	Tax Payable
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
treat $\times$ year $\in [s-1, s]$	0.019	0.038	-0.016	0.022	-0.046	0.001	0.020	0.022	-0.024	-0.007
	(0.016)	(0.018)	(0.021)	(0.020)	(0.026)	(0.015)	(0.019)	(0.023)	(0.022)	(0.028)
treat $\times$ year $\in [s-3, s]$	0.070	0.074	0.006	0.071	0.029	0.003	0.011	0.037	0.029	-0.006
	(0.016)	(0.017)	(0.019)	(0.019)	(0.024)	(0.015)	(0.019)	(0.023)	(0.022)	(0.027)
treat $\times$ year $\in [s-5, s]$	0.089	0.066	0.011	0.066	0.098	0.034	0.028	0.054	0.064	0.025
	(0.024)	(0.022)	(0.022)	(0.024)	(0.033)	(0.018)	(0.019)	(0.022)	(0.022)	(0.028)
Observations	2,324,186	2,025,380	1,672,095	1,681,583	1,154,574	2,628,878	2,290,848	1,934,273	1,945,733	1,312,928
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes*:* The table explores if the preexisting trends for the five outcomes indicated in the heading of each column were parallel between audited and unaudited firms. We estimate a model similar to (6) replacing the $assign \times after_{it}$ dummy with three dummies shown in the top three rows. The dummy variable $treat_i$ denotes that firm i was audited in the wave indicated in the heading of the column. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The sample for these regressions include the baseline periods only, from July 2008 to August 2013 for the first wave and from July 2008 to August 2014 for the second. The dummy variable $year \in [s-1, s]$ indicates that the period is one of the last twelve months included in the regression and so on. Standard errors are in parenthesis, which have been clustered at the firm level.

TABLE A.VIII: **HETEROGENEITY IN RESPONSE WITH RESPECT TO AMOUNT DETECTED**

	Sales	Purchases	Output Tax	Input Tax	Tax Payable
	(1)	(2)	(3)	(4)	(5)
A: First Wave					
assign × after	-0.009	-0.016	-0.020	-0.029	0.004
	(0.019)	(0.021)	(0.025)	(0.026)	(0.031)
assign × after × trait	0.009	-0.023	-0.022	-0.031	-0.089
	(0.040)	(0.048)	(0.052)	(0.054)	(0.070)
Observations	3,839,502	3,328,628	2,884,225	2,906,045	1,913,096
B: Second Wave					
assign × after	-0.014	-0.019	-0.016	-0.009	0.005
	(0.011)	(0.013)	(0.013)	(0.013)	(0.017)
assign × after × trait	0.040	0.119	0.053	0.038	0.010
	(0.031)	(0.041)	(0.042)	(0.039)	(0.048)
Observations	4,390,478	3,791,277	3,262,221	3,313,664	2,151,912
Firm FEs	Yes	Yes	Yes	Yes	Yes
Period FEs	Yes	Yes	Yes	Yes	Yes

Notes*:* The table explores heterogeneity in the audit effect. We divide firms into two groups. Firms against whom a positive amount was detected by audit are included in one group (indicated by the dummy variable $trait_i$); the rest of the firms are included in the baseline category. We then estimate a triple-difference version of model (6). The model includes interactions of the $trait_i$ dummy with the $assign \times after_{it}$ dummy. The $assign_i$ dummy takes the value 1 if the firm's audit was assigned in the corresponding random computer ballot. The control group comprises the rest of the firms in the eligible sample. The eligible sample consists of the population of VAT filers excluding government departments and firms already under audit. The dummy variable $after_t$ indicates that month t falls after the date of the ballot. The coefficients and the 95% confidence intervals on the double and triple-interaction terms from these regressions are plotted. Regressions are run separately for the first and the second audit waves. Standard errors are clustered at the firm level.

TABLE A.IX: **HETEROGENEITY IN AMOUNT DETECTED BY SHARE FINAL SALES**

	Outcome: Amount Detected (Std. Deviations)					
	(1)	(2)	(3)	(4)	(5)	(6)
A: Share Final Sales						
2nd Quartile	-0.100*	-0.099**	-0.097*	-0.098*	-0.105**	-0.096**
	(0.051)	(0.051)	(0.050)	(0.050)	(0.052)	(0.048)
3rd Quartile	-0.094*	-0.091*	-0.085*	-0.090*	-0.098*	-0.086*
	(0.051)	(0.050)	(0.047)	(0.050)	(0.052)	(0.046)
4th Quartile	-0.101**	-0.097*	-0.090**	-0.085*	-0.108*	-0.085*
	(0.051)	(0.049)	(0.046)	(0.044)	(0.056)	(0.045)
Observations	6,561	6,561	6,561	6,560	6,548	6,547
B: Share (Final Sales + Purchases from Unregistered Sector)						
2nd Quartile	-0.085	-0.082	-0.076	-0.081	-0.088	-0.074
	(0.052)	(0.051)	(0.048)	(0.051)	(0.053)	(0.046)
3rd Quartile	-0.108**	-0.087*	-0.094**	-0.083*	-0.113**	-0.074*
	(0.052)	(0.045)	(0.045)	(0.043)	(0.057)	(0.042)
4th Quartile	-0.113**	-0.086**	-0.095**	-0.086**	-0.118**	-0.076*
	(0.052)	(0.044)	(0.043)	(0.043)	(0.059)	(0.043)
Observations	6,561	6,561	6,561	6,560	6,548	6,547
Size FEs	No	Yes	No	No	No	Yes
Production Stage FEs	No	No	Yes	No	No	Yes
Tax Office FEs	No	No	No	Yes	No	Yes
Industry FEs	No	No	No	No	Yes	Yes

Notes*:* The table examines if the amount detected by audit changes with the share of final sales reported by a firm at the baseline. The outcome variable here is the amount detected by audit, normalized by its standard deviation. To maximize power, we pool together the audits conducted in the first two waves. Final sales are defined as sales where the other party to the transaction does not possess a national tax number: they are either consumers or informal firms. We divide firms into four quartiles based on the share of final sales in their turnover at the baseline. We regress the outcome variable on the three quartile dummies, omitting the first quartile as the reference group. We successively introduce the controls indicated in the last four lines. ***, **, and * denote significance at the 1%, 5%, and 10% levels.